

Tesis Doctoral:

CONTRASTES DE NO INVERTIBILIDAD  
Y COINTEGRACIÓN EN MODELOS VARIMA

DOCTORANDO: Carlos Díaz Vela.

DIRECTOR: Dr. D. José Luis Gallego Gómez.

Departamento de Economía.

Universidad de Cantabria.

Santander, febrero de 2012.





# Agradecimientos

En primer lugar, me gustaría agradecer a mi director, José Luis Gallego, todo el tiempo y el trabajo dedicados a la dirección de esta tesis. Gracias también por sus consejos, su apoyo y su ánimo que tanto me han ayudado. Sin sus orientaciones, su ayuda y su paciencia me habría sido imposible llegar a este punto. De José Luis he aprendido muchas cosas, pero le estoy especialmente agradecido por haberme mostrado el valor del trabajo constante y riguroso y la importancia de la precisión en el uso del lenguaje. Espero haber sido buen alumno. También le agradezco su apoyo personal. En los periodos de desánimo ha sabido mantener mi interés por el tema y no desfallecer, gracias por ello también.

Gracias a Cristina por su amistad y por todos los buenos y malos ratos que hemos pasado; también a ella debo en parte el inicio y fin de esta tesis. Agradezco a todos aquellos que, mediante sus comentarios en diversos congresos y foros de discusión han hecho que este trabajo mejore. Gracias también a la Universidad de Cantabria por la oportunidad que me ha dado, y a los alumnos que me ayudan cada día a expresarme un poco mejor.

Gracias de forma muy especial a mis padres, que me han dado una gran educación y la posibilidad de llegar hasta aquí, que siempre me han apoyado y a los que sé que siempre voy a tener conmigo, igual que yo voy a estar siempre con ellos. Gracias por su cariño incondicional y su comprensión. A Paula, por todo el apoyo que me ha dado durante estos años, por lo mucho que me ha soportado, por haberme ayudado a llevar los momentos duros y por estar siempre ahí. Sin su comprensión y consejos las cosas hubieran sido muy distintas. A mi hermana, por su buen humor que muchas veces me ha ayudado a desconectar. A mis abuelos, mis tíos, mis primos y el resto mi familia en general, que siempre han estado pendientes de mí y con los que he crecido en un entorno agradable que me han ayudado a ser la persona que soy. A Ingrid por su apoyo y sus consejos. A Pedro Álvarez, Marijose, Pedro Casares, Virginia y el resto del Departamento de Economía de la UC.



# Índice general

<b>1. Introducción</b>	<b>9</b>
<b>2. Conceptos básicos de inferencia estadística</b>	<b>17</b>
2.1. Introducción . . . . .	17
2.2. Contrastes uniformemente más potentes . . . . .	19
2.2.1. El teorema de Neymann-Pearson . . . . .	19
2.2.2. Contrastes uniformemente más potentes . . . . .	20
2.2.3. Contrastes uniformemente más potentes e insesgados . . . . .	22
2.3. Contrastes localmente óptimos . . . . .	24
2.3.1. Motivación . . . . .	24
2.3.2. Contrastes localmente más potentes . . . . .	24
2.3.3. Contrastes localmente más potentes e insesgados . . . . .	26
2.4. Invarianza . . . . .	28
2.5. Aplicaciones . . . . .	29
2.5.1. El modelo de regresión univariante . . . . .	29
2.5.2. El modelo de regresión multivariante . . . . .	31
<b>3. Contrastes de no invertibilidad en el modelos VARIMA</b>	<b>35</b>
3.1. Introducción . . . . .	35
3.2. El contraste de Nyblom y Harvey . . . . .	37
3.2.1. Modelo y estadístico de contraste . . . . .	37
3.2.2. Distribución del estadístico bajo la hipótesis nula . . . . .	39
3.2.3. Corrección no paramétrica por correlación serial . . . . .	41
3.3. El contraste básico de no invertibilidad . . . . .	41
3.3.1. Modelo y derivación del estadístico . . . . .	41
3.3.2. Derivación alternativa del estadístico de Nyblom y Harvey . . . . .	44

3.3.3.	Distribución en muestras finitas del estadístico . . . . .	46
3.3.4.	Distribución asintótica del estadístico . . . . .	49
3.4.	El modelo de Nyblom y Harvey corregido . . . . .	51
3.4.1.	Formulación del y derivación del estadístico . . . . .	51
3.4.2.	Estudio Monte Carlo . . . . .	54
3.5.	Extensiones del contraste básico de no invertibilidad . . . . .	55
3.5.1.	Raíces unitarias y corrección paramétrica . . . . .	55
3.5.2.	Casos particulares: El modelo diagonal . . . . .	57
<b>4.</b>	<b>Contrastes de no invertibilidad en modelos estacionales</b>	<b>65</b>
4.1.	Introducción . . . . .	65
4.2.	Contrastes invariantes para estructuras de covarianzas . . . . .	66
4.3.	Contrastes de componentes deterministas . . . . .	68
4.3.1.	Nivel local estacional . . . . .	68
4.3.2.	Tendencia lineal estacional . . . . .	69
4.3.3.	Tendencia lineal regular . . . . .	70
4.3.4.	Modelos VIMA(1,1) <sub>s</sub> estacionales . . . . .	71
4.4.	Modelos dinámicos . . . . .	72
4.4.1.	Ciclos estacionales deterministas . . . . .	72
4.4.2.	Ficticias estacionales . . . . .	74
4.4.3.	Estacionalidad determinista en una frecuencia . . . . .	75
4.5.	La aproximación IG . . . . .	76
4.6.	Aplicación . . . . .	78
<b>5.</b>	<b>Cointegración en modelos VARIMA</b>	<b>83</b>
5.1.	Introducción . . . . .	83
5.2.	La aproximación basada en el modelo de regresión . . . . .	84
5.2.1.	Formulación VIMA(1,1) y contraste de cointegración . . . . .	84
5.2.2.	Extensión para modelos con correlación serial . . . . .	88
5.2.3.	El modelo de tendencias comunes . . . . .	90
5.2.4.	Ejemplo ilustrativo: El modelo bivalente cointegrado . . . . .	93
5.3.	Cointegración en modelos VIMA(1,q) . . . . .	97
5.3.1.	El modelo VIMA(1,1) cointegrado . . . . .	97
5.3.2.	El modelo VIMA(1,q) cointegrado . . . . .	100
5.3.3.	Estudio Monte Carlo . . . . .	101
5.4.	Aplicación . . . . .	107

---

<b>6. Conclusiones</b>	<b>111</b>
<b>A. Contrastes de cointegración</b>	<b>115</b>
A.1. La aproximación basada en regresión . . . . .	115
A.1.1. Representación y estimación . . . . .	115
A.1.2. Métodos de estimación en procesos con correlación serial . . . . .	118
A.1.3. Contrastes basados en regresión . . . . .	121
A.2. Modelo de tendencias comunes . . . . .	123
A.2.1. Representación y forma estructural . . . . .	123
A.2.2. Contraste de tendencias comunes . . . . .	124
A.3. Cointegración en modelos VAR . . . . .	125
A.3.1. Modelo VAR(p) y Modelo de corrección de error . . . . .	125
A.3.2. Regresión de rango reducido y contraste de cointegración . . . . .	126
A.3.3. Cointegración y tendencias comunes . . . . .	127



# Capítulo 1

## Introducción

La clase general de modelos autorregresivos-de medias móviles vectoriales (VARMA), que contiene como caso especial a su análoga univariante, puede contemplarse como la representación *parsimoniosa* del proceso estocástico lineal general multivariante y puede usarse para describir, predecir y corregir un vector de series temporales interrelacionadas. Las propiedades estadísticas de este modelo multivariante, así como la metodología de modelización siguiendo el procedimiento iterativo popularizado por Box y Jenkins, se discuten en Jenkins y Alavi (1981) y Tiao y Box (1981) y se han incorporado como un nuevo capítulo en el libro clásico de análisis de series temporales de Box et al. (2008). Esta metodología de análisis multivariante, sin embargo, no ha alcanzado la aceptación y popularidad de su versión univariante. Esto puede explicarse por las dificultades que plantean en el análisis aplicado tanto la denominada *maldición de la dimensionalidad*, que conlleva la estimación de un gran número de parámetros y requiere de la disponibilidad de un *software* muy especializado, como el tratamiento de la no estacionariedad en media. Tiao y Box (1981) advirtieron de que combinaciones lineales de un vector de series temporales no estacionarias pueden ser estacionarias, de manera que la diferenciación simultánea de todas ellas puede conducir a complicaciones innecesarias en el ajuste del modelo. El problema de cómo manejar estas combinaciones lineales en las etapas de identificación, estimación y diagnóstico de modelos VARMA ha recibido mucha menos atención que en otras metodologías rivales, centrándose las principales contribuciones en el problema de reducir la dimensión (Peña y Box, 1987; Escribano y Peña, 1993; Peña y Poncela, 2006) o en la estimación de modelos parcialmente no estacionarios (Johansen, 1988, 1991; Ahn y Reinsel, 1988, 1990; Yap y Reinsel, 1995; Lüktephl y Claessen, 1997).

La idea de que una combinación lineal de un vector de series temporales no estacionarias puede ser estacionaria es la que subyace tras el concepto de cointegración propuesto por Granger (1981) e incorporado en el modelo de corrección de error por Engle y Granger (1987), laureados ambos con el premio Nobel en Economía de 2003

por su contribución, entre otras, al desarrollo de una vasta teoría estadística para abordar los problemas de cómo identificar, representar y estimar relaciones de cointegración. Johansen (2004) revisa esta teoría distinguiendo tres aproximaciones diferentes a la representación de variables cointegradas según se use un modelo de regresión, un modelo autorregresivo o un modelo estructural.

En la aproximación basada en el modelo de regresión, se considera un vector de  $m$  series temporales  $\mathbf{z}_t = (\mathbf{z}'_{1t} \mathbf{z}'_{2t})'$  compuesto de  $m_1$  variables dependientes y  $m_2$  variables explicativas,

$$\mathbf{z}_{1t} = \mathbf{B}\mathbf{z}_{2t} + \mathbf{u}_{1t}, \quad \nabla\mathbf{z}_{2t} = \mathbf{u}_{2t},$$

en donde el término de error es un proceso lineal invertible. Como  $\mathbf{z}_{1t}$  y  $\mathbf{z}_{2t}$  son dos vectores no estacionarios, la ecuación de regresión definirá una relación de cointegración siempre que su error,  $\mathbf{u}_{1t}$ , sea estacionario. La aproximación tradicional para realizar el contraste de cointegración consiste en aplicar contrastes de no estacionariedad tipo Dickey y Fuller (1979) a los residuos de la relación de cointegración tentativa, obtenidos por mínimos cuadrados u otros métodos asintóticamente libres del sesgo de endogeneidad que afecta a aquél (por ejemplo, los propuestos por Phillips y Hansen, 1990; Saikkonen, 1991; Park, 1992, entre otros). Por su parte Phillips y Ouliaris (1990), citando a Engle, resaltan la conveniencia de que la hipótesis nula a contrastar sea la de cointegración, lo que implica contrastar la estacionariedad de los residuos en lugar de su no estacionariedad. Diversos autores estudian las propiedades de los contrastes de estacionariedad univariantes tipo Tanaka (1990) o Kwiatkowski et al. (1992) cuando el contraste se realiza sobre los residuos de las relaciones de cointegración. Entre ellos caben destacar los contrastes de Leybourne y McCabe (1993), Harris y Inder (1994) o Shin (1994), los contrastes localmente óptimos de Tanaka (1993), Jansson (2005) o Kurozumi y Arai (2008) o los contrastes multivariantes de estacionariedad de Choi y Ahn (1995) y Choi y Ahn (1999) y de cointegración de Nyblom y Harvey (2000).

En la aproximación basada en el modelo autorregresivo, el vector de variables endógenas  $\mathbf{z}_t$  se describe por el proceso lineal general multivariante

$$\mathbf{\Pi}(B)\mathbf{z}_t = \mathbf{a}_t,$$

en donde  $\mathbf{\Pi}(B) = \mathbf{I} - \sum_{j=1}^{\infty} \mathbf{\Pi}_j B^j$  es un polinomio matricial en el operador retardo  $B$  cuyos coeficientes  $\mathbf{\Pi}_j$  son matrices de orden  $m \times m$ . Usando la denominada descomposición de Dickey-Fuller del polinomio  $\mathbf{\Pi}(B)$  (véase, p. ej. Patterson, 2011),

$\mathbf{\Pi}(B) = \mathbf{\Pi}(1)B + \mathbf{\Pi}^*(B)\nabla$ , se obtiene que

$$\mathbf{\Pi}^*(B)\nabla\mathbf{z}_t = -\mathbf{\Pi}(1)\mathbf{z}_{t-1} + \mathbf{a}_t.$$

Si el sistema presenta al menos una raíz unitaria, entonces  $|\mathbf{\Pi}(1)| = 0$ , con lo que el rango de  $\mathbf{\Pi}(1)$  no será completo. El teorema de la representación de Granger (Engle y Granger, 1987; Johansen, 1988) demuestra que, si el rango de esta matriz de ganancias es  $r < m$ , ésta puede descomponerse como  $\mathbf{\Pi}(1) = \boldsymbol{\alpha}\boldsymbol{\beta}'$ , donde  $\boldsymbol{\alpha}$  y  $\boldsymbol{\beta}$  son dos matrices  $m \times r$  de rango pleno, y  $\boldsymbol{\beta}'\mathbf{z}_{t-1}$  representa el vector de relaciones de cointegración. En la práctica, el polinomio matricial de orden infinito  $\mathbf{\Pi}(B)$  se sustituye por una aproximación finita de orden  $p$  suficientemente largo para recoger la dinámica del proceso,  $\boldsymbol{\Phi}(B) = \mathbf{I}_m - \sum_{j=1}^p \boldsymbol{\Phi}_j B^j$ , y que admite la misma descomposición. En este modelo VAR(p), Johansen (1988, 1991) estudia la estimación de máxima verosimilitud del proceso cointegrado y proporciona un contraste para determinar el rango de la matriz de ganancias,  $\boldsymbol{\Phi}(1)$ , o rango de cointegración. Métodos de estimación y contrastes de cointegración alternativos han sido desarrollados en este marco, entre otros, por Ahn y Reinsel (1990) y Reinsel y Ahn (1992). Más recientemente Yap y Reinsel (1995), Lütkepohl y Claessen (1997) y Mauricio (2006) extienden el análisis al modelo VARMA parcialmente no estacionario, donde el estudio de cointegración sigue centrándose en el polinomio autorregresivo.

En la aproximación basada en el modelo estructural, se parte de la forma MA( $\infty$ ) para la primera diferencia de  $\mathbf{z}_t$ ,

$$\nabla\mathbf{z}_t = \boldsymbol{\Psi}(B)\mathbf{a}_t,$$

en donde  $\boldsymbol{\Psi}(B) = \mathbf{I} + \sum_{j=1}^{\infty} \boldsymbol{\Psi}_j B^j$  es un polinomio matricial en el operador retardo  $B$  cuyos coeficientes  $\boldsymbol{\Psi}_j$  son matrices de orden  $m \times m$ . Usando la descomposición de Beveridge-Nelson del polinomio  $\boldsymbol{\Psi}(B)$  (véase, p. ej. Patterson, 2011),  $\boldsymbol{\Psi}(B) = \boldsymbol{\Psi}(1) + \boldsymbol{\Psi}^*(B)\nabla$ , se obtiene que

$$\nabla\mathbf{z}_t = \boldsymbol{\Psi}(1)\mathbf{a}_t + \boldsymbol{\Psi}^*(B)\nabla\mathbf{a}_t,$$

que, por sustituciones sucesivas en la expresión anterior, se puede escribir en forma integrada como

$$\mathbf{z}_t = \mathbf{z}_0 + \boldsymbol{\Psi}(1) \sum_{j=1}^t \mathbf{a}_j + \boldsymbol{\Psi}^*(B)\mathbf{a}_t,$$

en donde  $\mathbf{z}_0$  es una constante que se anula al diferenciar, el segundo término es el com-

ponente permanente y el tercero, el componente transitorio. De forma análoga al caso anterior, la presencia de raíces unitarias en el polinomio  $\Psi(B)$  implica  $|\Psi(1)| = 0$ , por lo que el rango de esta matriz de ganancias será  $r < m$ , y admitirá una descomposición  $\Psi(1) = \mathbf{h}\mathbf{b}'$  con  $\mathbf{h}$  y  $\mathbf{b}$  dos matrices  $m \times r$ . Con esta restricción, se demuestra que  $\mathbf{z}_t$  puede descomponerse en una parte estacionaria  $\Psi^*(B)\mathbf{a}_t$ , y  $r$  tendencias estocásticas correspondientes a la acumulación de las  $r$  combinaciones lineales de las perturbaciones del modelo,  $\mathbf{b}'\mathbf{a}_t$ . La forma en que éstas afectan a cada una de las series viene definida por las filas de la matriz de difusión,  $\mathbf{h}$ . Las relaciones de cointegración se definen en este marco como aquellas combinaciones lineales de  $\mathbf{z}_t$  que eliminan este componente permanente. Stock y Watson (1988) demuestran que  $\mathbf{z}_t$  puede representarse alternativamente como un modelo de tendencias comunes

$$\mathbf{z}_t = \mathbf{z}_0 + \mathbf{h}\boldsymbol{\mu}_t + \mathbf{u}_t, \quad \boldsymbol{\mu}_t = \boldsymbol{\mu}_{t-1} + \mathbf{v}_t,$$

en donde  $\mathbf{u}_t = \sum_{j=1}^{\infty} \Psi_j^* \mathbf{a}_t$ ,  $\mathbf{z}_0$  es el vector de condiciones iniciales. Esta representación de tendencias comunes ha sido también estudiada y ampliada por Vahid y Engle (1993) o más recientemente por Wagner (2010) y guarda relación los modelos de factores comunes. Nyblom y Harvey (2000) y Wagner (2010) proponen un análisis de cointegración en modelos de espacio de los estados similar al anterior.

Resulta instructivo asimilar las dos aproximaciones anteriores con las dos estrategias comunes que se siguen en la identificación del orden de integración de una serie temporal univariante. En los contrastes tipo Dickey y Fuller (1979) se pone de manifiesto una preferencia por la infradiferenciación de series al ajustar un proceso AR(1) para comprobar su estacionariedad; por el contrario, los contrastes tipo Nyblom y Mäkeläinen (1983) revelan una preferencia por la sobrediferenciación al ajustar un proceso IMA(1,1) para detectar su no invertibilidad. La aproximación VARIMA al análisis de cointegración propuesta en esta tesis sigue esta segunda estrategia y comparte el punto de partida de la aproximación estructural, el modelo VIMA(1, $\infty$ ), pero escrito en forma de VARIMA(p,1,q+1), en el que se incorpora la advertencia de Tiao y Box (1981): la diferenciación simultánea de un vector de series temporales cointegradas conduce a una representación no invertible,  $|\Psi(1)| = 0$ . A diferencia de lo que ocurre en la aproximación estructural, se usa la propia estructura VARMA para capturar la correlación serial multivariante.

Los contrastes de cointegración que se derivan en esta tesis requieren el desarrollo de contrastes de no invertibilidad multivariante análogos a los contrastes univariantes de raíces MA unitarias de Tanaka (1990) o Saikkonen y Luukkonen (1993a). Estos contrastes univariantes son equivalentes al contraste de componentes deterministas desarrollado por Nyblom y Mäkeläinen (1983), extendido al caso multivariante por Nyblom y Harvey

(2000) y cuyo estadístico de contraste, por tanto, puede aplicarse para detectar no invertibilidad en un modelo VIMA(1,1) cuya matriz de parámetros media móvil tiene estructura escalar. Para realizar el contraste sobre modelos VARIMA(p,1,q+1), se proporciona una corrección paramétrica del estadístico, alternativa a la corrección no paramétrica propuesta por Nyblom y Harvey (2000), y que puede contemplarse como la extensión multivariante de las propuestas por Saikkonen y Luukkonen (1993a), Tam y Reinsel (1997) y Leybourne y McCabe (1994). En la tesis doctoral de Mazas (2012) se revisan e interrelacionan todos estos estadísticos siguiendo una aproximación común y se proporciona una forma del estadístico basada en los residuos exactos del modelo ARIMA, por lo que su implementación práctica es relativamente sencilla. La corrección propuesta en esta tesis es la generalización multivariante de esta última. El procedimiento para la derivación del estadístico permite extender fácilmente los contrastes de no invertibilidad al caso estacional y a modelos con componentes deterministas, lo que permitirá en posteriores desarrollos del trabajo la extensión de los contrastes de cointegración al caso estacional en la línea de Ahn y Reinsel (1994) o Johansen y Schaumburg (1999).

Se muestra también un procedimiento para contrastar la presencia de cointegración en un modelo VARIMA, que se reduce a detectar raíces unitarias en el polinomio media móvil. Se demuestra cómo las tres representaciones anteriores para un proceso cointegrado pueden acomodarse en un modelo VARIMA. La representación basada en el modelo regresión recibe una especial atención dado el gran interés que ha suscitado en la literatura. Para este caso, es posible especificar un modelo ARIMA multivariante que, sujeto a unas determinadas restricciones sobre el polinomio media móvil, permite tanto la estimación consistente de la relación de cointegración como la derivación de un contraste de cointegración alternativo a los habitualmente disponibles y que comparte las propiedades de optimalidad de los contrastes de no invertibilidad. Para modelos más generales, se muestra cómo la corrección paramétrica por correlación serial es también útil para contrastar la presencia de cointegración.

Todos los estadísticos de contraste derivados en la tesis tienen una distribución asintótica no estándar que puede ser evaluada por integración numérica; sin embargo, su análoga en muestras finitas debe ser evaluada por simulación. El procedimiento habitual en estos casos es aproximar la distribución de los estadísticos mediante regresiones de superficie de respuesta (MacKinnon, 1994) o mediante aproximaciones paramétricas basadas en distribuciones conocidas (Doornik, 1998). Utilizando este segundo enfoque, en esta tesis se presenta una aproximación a la distribución de los estadísticos, tanto asintótica como en muestras finitas, basada en la función de distribución inversa gaussiana. Los parámetros que caracterizan esta distribución se obtienen a partir de los dos primeros momentos, por lo que para cada estadístico derivado se obtienen expresiones

cerradas para su media y varianza. Los contrastes desarrollados en esta tesis, esta aproximación a su distribución, y el algoritmo de Gallego (2009) para la estimación por máxima verosimilitud exacta de modelos VARIMA y la obtención de los residuos han sido programados en C++ y se han incorporado al software estadístico **Empiricus**, disponible en [www.empiricus.unican.es](http://www.empiricus.unican.es).

La tesis se estructura en seis capítulos más un apéndice. En el segundo capítulo, que sigue a esta introducción, se resumen los principales resultados de inferencia estadística utilizados a lo largo de la tesis. La derivación de los estadísticos de contraste de no invertibilidad exige el uso de una metodología no estándar basada en conceptos de optimalidad local. En este capítulo se definen los conceptos de contraste localmente óptimo y contraste localmente óptimo e insesgado, ampliamente utilizados a lo largo de la tesis. Asimismo, se resume la metodología de King y Hillier (1985) para la obtención de los estadísticos localmente óptimos en el marco de un modelo de regresión y su extensión multivariante desarrollada por Nyblom (2001).

En el tercer capítulo se deriva el contraste localmente óptimo, insesgado e invariante para contrastar invertibilidad en un modelo multivariante. Se obtiene su distribución tanto asintótica como en muestras finitas bajo la hipótesis nula y bajo una secuencia de alternativas y se ilustra su relación con el contraste multivariante de nivel local determinista de Nyblom y Harvey (2000). Además, se propone la extensión multivariante de la corrección paramétrica de Mazas (2012) para realizar el contraste sobre modelos VARIMA( $p,1,q+1$ ). Asimismo, se ilustra cómo esta corrección paramétrica es útil en determinadas situaciones de especial interés.

En el capítulo 4 se derivan los contrastes de no invertibilidad estacional. Se contempla además una extensión para considerar la presencia de determinados componentes deterministas en el proceso generador de datos. Para cada estadístico se obtienen expresiones cerradas que permiten el cálculo sencillo de su media y varianza. Además, se propone y evalúa una aproximación a la distribución en muestras finitas de estos estadísticos basada en la distribución Inversa Gaussiana. Un ejemplo práctico ilustra el procedimiento de contraste.

El quinto capítulo ilustra el procedimiento de estimación y contraste de cointegración en un modelo VARIMA. Por su relevancia en la literatura se considera la forma VARIMA de la representación basada en el modelo de regresión de un sistema cointegrado. Se muestra cómo en este caso el polinomio media móvil matricial cumple una serie de restricciones que permiten derivar un contraste de cointegración con las mismas propiedades de optimalidad que los contrastes de no invertibilidad derivados en el capítulo 3. Se ilustra también el procedimiento para la detección y el contraste de cointegración en situaciones más generales y se muestra cómo la corrección paramétrica propuesta en el

---

capítulo 3 es útil también para contrastar cointegración en este caso. Varios ejemplos ilustran la metodología de contraste en ambos casos, y estudios de potencia evalúan la potencia del contraste de cointegración en ambas situaciones. En el apéndice se revisan los principales contrastes de cointegración.



## Capítulo 2

# Conceptos básicos de inferencia estadística

### 2.1. Introducción

En este primer capítulo de carácter introductorio se resume el conjunto de conceptos básicos de inferencia estadística que se manejarán a lo largo de la tesis, tomando como referencia los trabajos de Ferguson (1967), Lehmann y Romano (2005) y Gouriéroux y Monfort (1995).

Los problemas de contraste analizados en esta tesis se basan en un modelo paramétrico del tipo  $(X, F(x|\theta), \Omega)$  donde  $X$  es una variable aleatoria,  $F(x|\theta)$  su función de distribución conocida y caracterizada por el parámetro  $\theta$  sobre el que se centra el interés del contraste. Este parámetro pertenece a un espacio paramétrico  $\Omega$  y, si el modelo está identificado, cada valor de  $\theta \in \Omega$  da lugar a un valor distinto de la función de distribución. El tipo de problemas que se plantean en esta tesis consisten en determinar si este parámetro  $\theta$  pertenece a un determinado subespacio  $\Omega_0$  de  $\Omega$  (hipótesis nula,  $H_0$ ) o a su espacio complementario  $\Omega_1$  (hipótesis alternativa,  $H_1$ ). El espacio paramétrico  $\Omega$  se particiona así de tal forma que no exista ningún elemento  $\theta$  que pertenezca simultáneamente a ambos espacios, por lo que sólo una de las dos hipótesis será cierta. Cuando  $\Omega_0$  (u  $\Omega_1$ ) contiene un solo elemento  $\theta_0$  (o  $\theta_1$ ), la hipótesis nula (o alternativa) se denomina simple; en caso contrario la hipótesis se denomina compuesta. Igualmente, asociada a cada hipótesis existe un conjunto de funciones de distribución  $\mathcal{F}_j = \{F(\theta)|\theta \in \Omega_j\}$  ( $j = 0, 1$ ), la clase de distribuciones bajo  $H_0$  o bajo  $H_1$ .

Un contraste se puede definir como una regla de decisión que indica si, dada una realización particular  $x$  de la variable aleatoria  $X$ , se acepta o se rechaza la hipótesis nula. Todo contraste lleva asociada una función crítica,  $\phi(X)$ , tal que  $\phi(x) = 1$  implica el rechazo de la hipótesis nula (o la aceptación de la alternativa) y  $\phi(x) = 0$  implica el

no rechazo de la hipótesis nula (o el rechazo de la alternativa). De esta forma se podría dividir el espacio muestral de  $X$  ( $S$ ) entre  $S_0$ , que contiene las observaciones de  $X$  para las que se acepta la hipótesis nula y  $S_1$  que contiene aquellas para las que se rechaza y que se denomina habitualmente región crítica. Asimismo, en este problema de decisión se pueden cometer dos tipos de errores: rechazar la hipótesis nula cuando ésta es cierta (Error de Tipo I) y aceptarla cuando es falsa (Error de Tipo II), definidos formalmente de la siguiente manera,

$$\begin{array}{ll} \text{Error Tipo I} & \text{Prob}(\phi(X) = 1|\theta) = \text{Prob}(X \in S_1|\theta), \quad \forall \theta \in \Omega_0, \\ \text{Error Tipo II} & \text{Prob}(\phi(X) = 0|\theta) = \text{Prob}(X \in S_0|\theta), \quad \forall \theta \in \Omega_1. \end{array}$$

A la probabilidad de Error tipo I se le denomina *nivel de significación* o *tamaño* del contraste, mientras que a la probabilidad  $\beta(\theta) = 1 - \text{Prob}(\phi(X) = 1|\theta)$  evaluada como función de  $\theta \in \Omega_1$  se le denomina *función potencia* del contraste. La regla de decisión óptima será aquella que haga mínima la probabilidad de incurrir en cualquiera de ambos errores, sin embargo, ambas probabilidades no se pueden controlar simultáneamente. El procedimiento habitual consiste en fijar una probabilidad de Error Tipo I tolerable (nivel de significación,  $\alpha$ ) y minimizar la probabilidad de Error Tipo II, lo que equivale a maximizar la función potencia del contraste sujeto a un determinado nivel de significación.

En la discusión anterior se ha supuesto que el espacio muestral de  $X$  se puede particionar en dos conjuntos disjuntos  $S_0$  y  $S_1$ , sin embargo, esta situación no es la habitual sino que, dada una realización particular de la variable aleatoria  $x$ , existirá una determinada probabilidad de rechazar la hipótesis nula y viceversa. En este caso, que es el considerado en esta tesis, para una función crítica  $\phi(X)$  definida sobre una variable aleatoria  $X$  con función de distribución continua tal que  $dF(x|\theta) = f(x|\theta)dx$ , la función potencia se define en adelante como

$$E_\theta[\phi(X)] = \int \phi(x)f(x|\theta)dx, \quad (2.1)$$

con lo que el problema a la hora de construir un contraste consiste en obtener la función  $\phi(X)$  que maximice la probabilidad de rechazo,

$$\beta(\theta) = E_\theta[\phi(X)], \quad \forall \theta \in \Omega_1, \quad \text{sujeto a } E_\theta[\phi(X)] \leq \alpha, \quad \forall \theta \in \Omega_0.$$

Es decir, una vez fijado un nivel de significación (o probabilidad de error tipo I tolerable)

se busca un contraste cuya función potencia sea mayor que la de cualquier otro contraste con igual nivel de significación en cualquier punto del espacio paramétrico bajo la hipótesis alternativa. Si tal contraste existe, se dice que es uniformemente más potente (que cualquier otro). En aquellos casos en los que no exista un contraste con estas características, se buscará uno cuya función potencia sea mayor que la de cualquier otro en un entorno de  $H_0$ . A estos contrastes se les denomina localmente óptimos. El resto del capítulo se desarrolla de la siguiente manera: en el apartado 2 se presentan los contrastes uniformemente más potentes, en el apartado 3, los contrastes localmente óptimos y en el 4 varias aplicaciones al modelo de regresión univariante y multivariante.

## 2.2. Contrastes uniformemente más potentes

### 2.2.1. El teorema de Neymann-Pearson

El Teorema fundamental de Neymann-Pearson define el contraste uniformemente más potente para contrastar una hipótesis nula simple  $H_0 : \theta = \theta_0$  frente a una alternativa también simple  $H_1 : \theta = \theta_1$ . Si  $F(x|\theta_0)$  y  $F(x|\theta_1)$  son dos funciones de distribución con funciones de densidad  $f(x|\theta_0)$  y  $f(x|\theta_1)$  respectivamente, existe un contraste  $\phi(X)$  y una constante  $c$  tal que, para un nivel de significación determinado,  $E_{\theta_0}[\phi(X)] = \alpha$ ,

$$\phi(x) = \begin{cases} 1 & \text{si } f(x|\theta_1) > cf(x|\theta_0) \\ \gamma(x) & \text{si } f(x|\theta_1) = cf(x|\theta_0) \\ 0 & \text{si } f(x|\theta_1) < cf(x|\theta_0) \end{cases} \quad (2.2)$$

donde  $\gamma(x)$  es una constante. Se puede comprobar que  $\phi(X)$  es uniformemente más potente que cualquier otro contraste  $\phi'(X)$  con igual nivel de significación,  $E_{\theta_0}[\phi'(X)] = \alpha$ . Si ambos tienen la forma (2.2) entonces las diferencias entre sus funciones potencia,  $E_{\theta}[\phi(X)] - E_{\theta}[\phi'(X)]$ , definidas en (2.1) evaluadas  $\forall \theta$  se puede expresar como

$$\int [\phi(x) - \phi'(x)][f(x|\theta_1) - cf(x|\theta_0)]dx \geq 0, \quad (2.3)$$

Que el signo de esta integral es positivo se puede demostrar fácilmente, teniendo en cuenta que cuando  $f(x|\theta_1) > cf(x|\theta_0)$ ,  $\phi(x) = 1$ , mientras que  $\phi'(x)$  puede ser igual a 1 ó 0, con lo que el primer factor en (2.3) será 0 ó 1, mientras que el segundo será siempre positivo. Por el contrario, si  $f(x|\theta_1) < cf(x|\theta_0)$ , entonces  $\phi(x) = 0$  e igualmente  $\phi'(x)$  puede ser igual a 0 ó 1, por lo que el primer factor será 0 ó -1, siendo el segundo siempre

negativo en este caso. Dada la definición (2.1), la expresión (2.3) implica que

$$E_{\theta_1}[\phi(X)] - E_{\theta_1}[\phi'(X)] \geq c(E_{\theta_0}[\phi(X)] - E_{\theta_0}[\phi'(X)]).$$

De las condiciones anteriores, y suponiendo  $c > 0$ , se sabe que  $E_{\theta_0}[\phi(X)] - E_{\theta_0}[\phi'(X)] = 0$ , porque ambos están contruidos al mismo nivel de significación,  $\alpha$ . Esto implica que  $E_{\theta_1}[\phi(X)] - E_{\theta_1}[\phi'(X)] \geq 0$ , es decir, que la función potencia de  $\phi(X)$  es mayor o igual que la de cualquier otro test  $\phi'(X)$  de igual nivel de significación, es decir, que es el contraste uniformemente más potente para este problema.

Los resultados del teorema no se ven alterados si en lugar de utilizar las funciones de densidad se aplica una transformación logarítmica de forma que

$$\phi(x) = \begin{cases} 1 & \text{si } \log f(x|\theta_1) - \log f(x|\theta_0) > c^* \\ 0 & \text{si } \log f(x|\theta_1) - \log f(x|\theta_0) < c^* \end{cases}$$

con  $c^* = \log(c)$ . La interpretación intuitiva del teorema es sencilla: se aceptará la hipótesis nula cuando el logaritmo de la función de verosimilitud de  $X$  bajo  $H_0$  y bajo  $H_1$  estén suficientemente cercanas.

### 2.2.2. Contrastes uniformemente más potentes

Aunque el problema de contrastar una hipótesis nula simple frente a una alternativa también simple tiene relevancia teórica, lo más común en la práctica es encontrar problemas en los que tanto la hipótesis nula como la alternativa, o al menos una de las dos, sean compuestas. Así, se considera en este apartado el problema general de contrastar  $H_0 : \theta \in \Omega_0$  frente a  $H_1 : \theta \in \Omega_1$ , de forma que un contraste  $\phi(X)$  tiene nivel de significación  $\alpha$  si

$$E_{\theta}[\phi(X)] \leq \alpha, \quad \forall \theta \in \Omega_0,$$

y es uniformemente más potente (UMP) si, para cualquier otro contraste  $\phi'(X)$  de igual nivel de significación se cumple que

$$E_{\theta}[\phi(X)] \geq E_{\theta}[\phi'(X)], \quad \forall \theta \in \Omega_1.$$

A continuación se estudian los dos tipos de contrastes hipótesis compuestas más habituales en la práctica y que se utilizarán a lo largo de la tesis.

### Contrastes de una cola

En primer lugar, se considera el problema de contraste  $H_0 : \theta \leq \theta_0$  frente a  $H_1 : \theta > \theta_0$ . El teorema fundamental de Neymann-Pearson proporciona un contraste que maximiza la función potencia para problemas del tipo  $H_0 : \theta = \theta_0$  frente a  $H_1 : \theta = \theta_1$  (con, por ejemplo,  $\theta_1 > \theta_0$ ), pero nada asegura que este contraste maximice también la potencia para cualquier otro punto  $\theta_2 > \theta_0$  ( $\theta_2 \neq \theta_1$ ). El hecho de que el contraste  $\phi(X)$  definido en (2.2) dependa de un punto del espacio paramétrico bajo la alternativa implica que el contraste no es uniformemente más potente para este tipo de contrastes de una cola.

Existe, sin embargo, un caso en el que sí se pueden obtener contrastes UMP por aplicación del teorema de Neymann-Pearson, aquel en el que la familia de distribuciones  $F(x|\theta)$  tiene ratio de verosimilitudes monótona, esto es, si para cualquier  $\theta_1 < \theta_2$  el cociente de sus funciones de densidad

$$r(x|\theta_2, \theta_1) = \frac{f(x|\theta_2)}{f(x|\theta_1)},$$

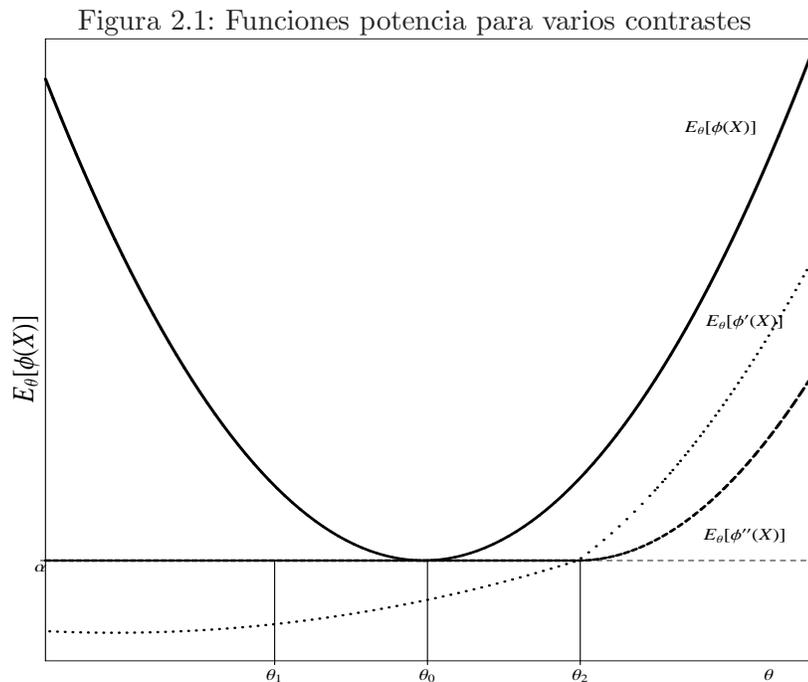
es una función no decreciente. En este caso, suponiendo  $\theta_1 > \theta_0$ , se puede obtener el contraste uniformemente más potente con nivel de significación  $\alpha$  como

$$\phi(x) = \begin{cases} 1 & \text{si } f(x|\theta_1) > cf(x|\theta_0) \\ \gamma(x) & \text{si } f(x|\theta_1) = cf(x|\theta_0) \\ 0 & \text{si } f(x|\theta_1) < cf(x|\theta_0) \end{cases}$$

con  $E_{\theta_0}[\phi(X)] = \alpha$ . Para comprobar que  $\phi(X)$  no depende de la alternativa considerada  $\theta_1$ , y que por lo tanto es uniformemente más potente, basta con observar que se rechaza  $H_0$  para valores de  $r(x|\theta_1, \theta_0) = f(x|\theta_1)/f(x|\theta_0) > c$  con  $\theta_1 > \theta_0$ . Dado que  $r(x|\theta_1, \theta_0)$  es no decreciente en  $\theta_1$ , si el contraste rechaza para valor de  $\theta_1$  arbitrariamente cercano a  $\theta_0$ , también se rechazará para cualquier  $\theta > \theta_1$ . Cambiando el signo de la desigualdad en  $\phi(x)$  se obtiene la expresión del contraste uniformemente más potente para el problema  $H_0 : \theta \geq \theta_0$  frente a  $H_1 : \theta < \theta_0$ .

### Contrastes de dos colas

Otro tipo de problemas relevantes a contrastar son los basados en hipótesis nulas a dos colas del tipo  $H_0 : \theta_1 \leq \theta \leq \theta_2$  frente a  $H_1 : \theta < \theta_1$  ó  $\theta > \theta_2$  o, considerando  $\theta_1$  arbitrariamente cercano a  $\theta_2$ ,  $H_0 : \theta = \theta_0$  frente a  $H_1 : \theta \neq \theta_0$ . Es habitual que



para este tipo de problemas no exista contrastes uniformemente más potentes. Como ilustración, la figura 2.1 representa las funciones potencia de tres contrastes.  $E_\theta[\phi'(X)]$  representa la función potencia del contraste uniformemente más potente para el problema  $H_0 : \theta \leq \theta_2$  frente a  $H_1 : \theta > \theta_2$  con función de verosimilitud monótona. Bajo  $H_0$ , la función potencia es menor que el nivel de significación,  $\alpha$ , cota que alcanza en  $\theta_2$ , y su función potencia es creciente bajo  $H_1$ . Suponiendo que esta función potencia es mayor que la de cualquier otro,  $\phi'(X)$  sería el contraste uniformemente más potente para este problema. Si el propósito es contrastar  $H_0 : \theta_1 \leq \theta \leq \theta_2$  frente a  $H_1 : \theta < \theta_1$  o  $\theta > \theta_2$ ,  $\phi'(X)$  no es un contraste óptimo ya que para  $\theta < \theta_1$  su función potencia es menor que el nivel de significación. Contrastes con funciones potencia de este tipo son claramente poco deseables para este tipo de problemas de contraste a dos colas, ya que implicaría que bajo un subespacio de la alternativa, en este caso aquellos valores de  $\theta < \theta_1$ , la probabilidad de rechazar  $H_1$  es menor que el nivel de significación, lo que no es compatible con la estructura para un contraste óptimo. Para este caso se definen los procedimientos de contraste insesgados.

### 2.2.3. Contrastes uniformemente más potentes e insesgados

En ocasiones, como en el caso de los contrastes a dos colas y ratio de verosimilitudes monótona, es difícil encontrar contrastes uniformemente más potentes y que además

tengan la propiedad deseable de que su función potencia nunca sea menor que su nivel de significación. A esta segunda condición deseable para cualquier contraste se la denomina propiedad de insesgadez. Formalmente, un contraste  $\phi(X)$  se define insesgado si

$$E_{\theta}[\phi(X)] \geq \alpha, \quad \forall \theta \in \Omega_1,$$

es decir, si la probabilidad de rechazar la hipótesis nula para cualquier alternativa nunca es menor que la probabilidad de aceptar la hipótesis nula cuando es cierta. Por lo tanto, un contraste uniformemente más potente e insesgado (uniformly most powerful and unbiased, UMPU) con nivel de significación  $\alpha$  es aquel contraste,  $\phi(X)$ , cuya función potencia es la mayor de entre todos los contrastes insesgados y con nivel de significación  $\alpha$ . En la figura 2.1 se representa la función potencia de  $\phi''(X)$ , un contraste insesgado y óptimo para el problema  $H_0 : \theta_1 \leq \theta \leq \theta_2$  frente a  $H_1 : \theta < \theta_1$  ó  $\theta > \theta_2$ . Su función potencia es igual al nivel de significación,  $\alpha$ , para  $\theta < \theta_1$  y estrictamente creciente para  $\theta > \theta_2$ . La propiedad de insesgadez implica que el contraste  $\phi''(X)$  es preferible a  $\phi'(X)$  para este problema, aunque la función potencia de este último sea mayor para  $\theta > \theta_2$ , ya que ésta cae por debajo del nivel de significación  $\alpha$  para  $\theta < \theta_1$ .

Aplicando el teorema generalizado de Neymann-Pearson (Lehmann y Romano, 2005, p.77) y suponiendo que la variable aleatoria  $X$  tiene ratio de verosimilitudes monótona, se define el contraste uniformemente más potente e insesgados para contrastar  $H_0 : \theta_1 \leq \theta \leq \theta_2$  frente a  $H_1 : \theta < \theta_1$  ó  $\theta > \theta_2$  como aquel contraste  $\phi(X)$  tal que  $E_{\theta_i}[\phi(X)] = \alpha$  ( $i = 1, 2$ ) cuya estructura es

$$\phi(x) = \begin{cases} 1 & \text{si } f(x|\theta) < c_1 f(x|\theta_1) + c_2 f(x|\theta_2) \\ \gamma(x) & \text{si } f(x|\theta) = c_1 f(x|\theta_1) + c_2 f(x|\theta_2) \\ 0 & \text{si } f(x|\theta) > c_1 f(x|\theta_1) + c_2 f(x|\theta_2) \end{cases} \quad (2.4)$$

donde  $c_1, c_2$  y  $0 \leq \gamma(x) \leq 1$  se eligen tal que

$$E_{\theta_1}[\phi(X)] = E_{\theta_2}[\phi(X)] = \alpha. \quad (2.5)$$

Esta última es la condición de insesgadez, que asegura que la potencia del contraste no caiga por debajo de su nivel de significación. Nótese que si la ratio de verosimilitudes de la variable sobre la que se realiza el contraste es monótona, la única forma que puede tener la función potencia de un contraste para cumplir (2.5) es la representada para el test  $\phi''(X)$  en el gráfico 2.1.

Haciendo arbitrariamente pequeña la distancia entre  $\theta_1$  y  $\theta_2$  se llegaría al problema de contraste a dos colas  $H_0 : \theta = \theta_0$  frente a  $H_1 : \theta \neq \theta_0$ . En este caso la forma del contraste óptimo es también (2.4) y la condición de insesgadez (2.5) implica que, si  $\theta_0 = \theta_1$  y  $\theta_2 = \theta_0 + \epsilon$  ( $\epsilon > 0$ )

$$E_{\theta_0}[\phi(X)] = \alpha, \quad \text{y} \quad \lim_{\epsilon \rightarrow 0} \frac{E_{\theta_0}[\phi(X)] - E_{\theta_0 + \epsilon}[\phi(X)]}{\theta_0 - (\theta_0 + \epsilon)} \equiv \left. \frac{\partial E_{\theta}[\phi(X)]}{\partial \theta} \right|_{\theta = \theta_0} = 0,$$

es decir, que su función potencia, dada la propiedad de insesgadez, debe alcanzar un mínimo bajo la hipótesis nula, siendo el valor de la función en este punto igual al nivel de significación,  $\alpha$ . Un ejemplo de contraste insesgado en este caso es el contraste  $\phi(X)$ , cuya función potencia de muestra en la figura 2.1. Este contraste se define para a una variable aleatoria con función de densidad no monótona y su función potencia tiene un mínimo en  $\theta_0$  con valor igual a  $\alpha$ .

## 2.3. Contrastes localmente óptimos

### 2.3.1. Motivación

Hasta el momento la exposición se ha centrado en derivar contrastes óptimos para situaciones en las que estos se definen sobre variables aleatorias con ratio de verosimilitudes monótona. En la práctica esta situación no es habitual, por lo que rara vez existen contrastes uniformemente más potentes. En este caso la estrategia consiste en encontrar contrastes cuya función potencia, si bien no sea la máxima en todo el espacio paramétrico bajo la hipótesis alternativa, al menos lo sea en un entorno de  $H_0$ . A este tipo de contrastes se les denomina localmente localmente más potentes o localmente óptimos.

### 2.3.2. Contrastes localmente más potentes

Si el problema de contraste de interés es  $H_0 : \theta \leq \theta_0$  frente a la alternativa de una cola  $H_1 : \theta > \theta_0$ , donde la distribución de la variable aleatoria  $X$  sobre la que se define no tiene ratio de verosimilitudes monótona, es habitual no disponer de un contraste uniformemente más potente, por lo que la estrategia consiste en encontrar el contraste localmente más potente. Se dice que el contraste  $\phi(X)$ , con nivel de significación  $\alpha$  tal que  $E_{\theta_0}[\phi(X)] = \alpha$ , es localmente óptimo (Locally best, LB) si la pendiente de su función potencia es mayor que la de cualquier otro contraste  $\phi'(X)$  de igual nivel de significación en un entorno de  $\theta_0$ . Una interpretación intuitiva se obtiene a partir de la expansión

en serie de Taylor de primer orden de la función potencia de  $\phi(X)$  alrededor de  $\theta_0$  y la de cualquier otro contraste  $\phi'(X)$  con igual nivel de significación. De esta forma, la diferencia de ambas funciones potencia en un entorno de  $\theta_0$  se puede aproximar como

$$E_{\theta}[\phi(X)] - E_{\theta}[\phi'(X)] \simeq E_{\theta_0}[\phi(X)] - E_{\theta_0}[\phi'(X)] + (\theta - \theta_0) \left( \frac{\partial E_{\theta}[\phi(X)]}{\partial \theta} \Big|_{\theta=\theta_0} - \frac{\partial E_{\theta}[\phi'(X)]}{\partial \theta} \Big|_{\theta=\theta_0} \right).$$

Si ambos contrastes tienen el mismo nivel de significación,  $E_{\theta_0}[\phi(X)] = E_{\theta_0}[\phi'(X)] = \alpha$ , y dado que bajo la alternativa  $\theta - \theta_0 > 0$ , entonces la función potencia del primero será mayor que la de cualquier otra sólo en el caso en que la pendiente de la función potencia de  $\phi(X)$  evaluada en  $\theta_0$  sea mayor que la de cualquier otro contraste  $\phi'(X)$ . Esto implica que el contraste localmente óptimo es aquel cuya función potencia tiene la mayor pendiente en un entorno  $\theta = \theta_0 + \epsilon$  ( $\epsilon > 0$ ) de entre todos los contrastes de igual nivel de significación.

A partir del teorema de Neymann-Pearson, reemplazando  $f(x|\theta_1)$  por su aproximación en serie de Taylor de primer orden alrededor de  $\theta_0$ , se define el contraste localmente óptimo como (Ferguson, 1967),

$$\phi(x) = \begin{cases} 1 & \text{si } \frac{\partial f(x|\theta)}{\partial \theta} \Big|_{\theta=\theta_0} > c f(x|\theta_0) \\ \gamma(x) & \text{si } = \\ 0 & \text{si } < \end{cases} \quad (2.6)$$

donde, nótese que

$$\frac{1}{f(x|\theta_0)} \frac{\partial f(x|\theta)}{\partial \theta} \Big|_{\theta=\theta_0} = \frac{\partial \log f(x|\theta)}{\partial \theta} \Big|_{\theta=\theta_0},$$

por lo que el contraste localmente óptimo implica que se rechazará la hipótesis nula si la primera derivada del logaritmo de la función de densidad de la variable sobre la que se realiza el contraste es mayor que un valor crítico  $c$ , elegido de tal forma que el nivel de significación del contraste sea  $\alpha$ .

Si  $\phi(X)$  es un contraste localmente más potente que cualquier otro  $\phi'(X)$  de igual nivel de significación, entonces la diferencia entre sus funciones potencia será positiva con  $\theta_1$  arbitrariamente cercano por la derecha a  $\theta_0$  ( $\theta_1 = \theta_0 + \epsilon$  con  $\epsilon > 0$ ). La forma (2.6) del contraste óptimo se puede demostrar sustituyendo  $f(x|\theta_1)$  en (2.3) por su aproximación

en serie de Taylor de primer orden alrededor de  $\theta_0$ ,

$$\int [\phi(x) - \phi'(x)] \left[ f(x|\theta_0) + (\theta - \theta_0) \frac{\partial f(x|\theta)}{\partial \theta} \Big|_{\theta=\theta_0} - cf(x|\theta_0) \right] dx \geq 0.$$

Definiendo la primera derivada de la función potencia como

$$\frac{\partial E_\theta[\phi(X)]}{\partial \theta} = \int \phi(x) \frac{\partial f(x|\theta)}{\partial \theta} dx,$$

y como  $E_{\theta_0}[\phi(X)] - E_{\theta_0}[\phi'(X)] = 0$ , dado que ambos contrastes tiene nivel de significación  $\alpha$ , entonces  $\int [\phi(x) - \phi'(x)] f(x|\theta_0) dx = 0$ , con lo que la integral anterior será positiva siempre que

$$(\theta - \theta_0) \int [\phi(x) - \phi'(x)] \frac{\partial f(x|\theta)}{\partial \theta} \Big|_{\theta=\theta_0} dx \geq 0 \Rightarrow \frac{\partial E_\theta[\phi(X)]}{\partial \theta} \geq \frac{\partial E_\theta[\phi'(X)]}{\partial \theta},$$

ya que  $\theta - \theta_0 > 0$  bajo la hipótesis alternativa. Es decir, si la pendiente de la función potencia de  $\phi(X)$  es mayor que la de cualquier otro contraste  $\phi'(X)$  de igual nivel de significación.

Una derivación más intuitiva de (2.6) se obtiene observando que para contrastar  $H_0 : \theta = \theta_0$  frente a  $H_1 : \theta = \theta_1$  el teorema de Neymann-Pearson proporciona un contraste óptimo  $\phi(X)$  tal que se rechaza la hipótesis nula,  $\phi(x) = 1$ , si  $\log f(x|\theta_1) - \log f(x|\theta_0) > c^*$ , que se puede expresar alternativamente como  $[\log f(x|\theta_1) - \log f(x|\theta_0)]/(\theta_1 - \theta_0) > c^{**}$ . Considerando  $\theta_1$  arbitrariamente cercano a  $\theta_0$ ,  $\phi(x) = 1$  si

$$\lim_{\theta_1 \rightarrow \theta_0} \frac{\log f(x|\theta_1) - \log f(x|\theta_0)}{(\theta_1 - \theta_0)} \equiv \frac{\partial \log f(x|\theta)}{\partial \theta} \Big|_{\theta=\theta_0} > c,$$

que es la expresión inicial. En cualquiera de los dos casos, queda claro que la región de rechazo se define como  $\partial \log f(x|\theta)/\partial \theta|_{\theta=\theta_0} > c$  donde la constante  $c$  se elige de tal forma que la probabilidad de error tipo I sea igual al nivel de significación  $\alpha$  definido.

### 2.3.3. Contrastes localmente más potentes e insesgados

Para el caso del contraste a dos colas  $H_0 : \theta = \theta_0$  frente a  $H_1 : \theta \neq \theta_0$  existe un contraste localmente óptimo e insesgado (locally best unbiased, LBU). De forma similar al caso anterior, se puede utilizar una expansión en serie de Taylor de segundo orden alrededor de  $\theta_0$  para ver que cualquier contraste  $\phi(X)$  con nivel de significación  $\alpha$ , tal

que cumpla la condición de insesgadez, es localmente óptimo e insesgado si la segunda derivada de su función potencia es la mayor de entre todos los contrastes  $\phi'(X)$  insesgados y con nivel de significación  $\alpha$ ,

$$\begin{aligned} E_{\theta}[\phi(X)] - E_{\theta}[\phi'(X)] &\simeq E_{\theta_0}[\phi(X)] - E_{\theta_0}[\phi'(X)] \\ &+ (\theta - \theta_0) \left( \frac{\partial E_{\theta}[\phi(X)]}{\partial \theta} \Big|_{\theta=\theta_0} - \frac{\partial E_{\theta}[\phi'(X)]}{\partial \theta} \Big|_{\theta=\theta_0} \right) \\ &+ \frac{(\theta - \theta_0)^2}{2} \left( \frac{\partial^2 E_{\theta}[\phi(X)]}{\partial \theta^2} \Big|_{\theta=\theta_0} - \frac{\partial^2 E_{\theta}[\phi'(X)]}{\partial \theta^2} \Big|_{\theta=\theta_0} \right). \end{aligned}$$

Si ambos contrastes se construyen al nivel de significación  $\alpha$  y se cumple la condición de insesgadez para ambos, de forma que las pendientes de sus funciones potencia sean igual a cero evaluadas en  $\theta_0$ , entonces  $E_{\theta}[\phi(X)] - E'_{\theta}[\phi(X)] > 0$  sólo en el caso en que la segunda derivada de la función potencia de  $\phi(X)$  sea mayor que la de cualquier otra  $\phi'(X)$ . Así,  $\phi(X)$  tiene la siguiente expresión (Ferguson, 1967),

$$\phi(x) = \begin{cases} 1 & \text{si } \frac{\partial^2 f(x|\theta)}{\partial \theta^2} \Big|_{\theta=\theta_0} > c_1 \frac{\partial f(x|\theta)}{\partial \theta} \Big|_{\theta=\theta_0} + c_2 f(x|\theta_0) \\ \gamma(x) & \text{si } = \\ 0 & \text{si } < \end{cases} \quad (2.7)$$

que se obtiene reemplazando  $f(x|\theta_1)$  en (2.2) por su aproximación en serie de Taylor de segundo orden alrededor de  $\theta_0$ . Al igual que en el caso de los contrastes localmente óptimos, se puede demostrar (2.7) sustituyendo  $f(x|\theta_1)$  por su aproximación en serie de Taylor de segundo orden en un entorno de  $\theta_0$  en la expresión en (2.3) que define la diferencia entre la función potencia de dos contrastes. Entonces,  $\phi(X)$  es localmente óptimo e insesgado con respecto a cualquier otro contraste insesgado y del mismo nivel de significación,  $\phi'(X)$ , si

$$\int [\phi(x) - \phi'(x)] \left[ \frac{\partial^2 f(x|\theta)}{\partial \theta^2} \Big|_{\theta=\theta_0} - c_1 \frac{\partial f(x|\theta)}{\partial \theta} \Big|_{\theta=\theta_0} - c_2 f(x|\theta_0) \right] dx \geq 0.$$

Definiendo la segunda derivada de la función potencia como

$$\frac{\partial^2 E_{\theta}[\phi(X)]}{\partial \theta^2} = \int \phi(x) \frac{\partial^2 f(x|\theta)}{\partial \theta^2} dx,$$

la anterior expresión se puede expresar de la siguiente manera

$$\left( \frac{\partial^2 E_\theta[\phi(X)]}{\partial \theta^2} \Big|_{\theta=\theta_0} - \frac{\partial^2 E_\theta[\phi'(X)]}{\partial \theta^2} \Big|_{\theta=\theta_0} \right) - c_1 \left( \frac{\partial E_\theta[\phi(X)]}{\partial \theta} \Big|_{\theta=\theta_0} - \frac{\partial E_\theta[\phi'(X)]}{\partial \theta} \Big|_{\theta=\theta_0} \right) - c_1 (E_{\theta_0}[\phi(X)] - E_{\theta_0}[\phi'(X)]) \geq 0.$$

Ambos contrastes están definidos para el mismo nivel de significación, por lo que  $E_{\theta_0}[\phi(X)] - E_{\theta_0}[\phi'(X)] = 0$  y el segundo término de la expresión es cero por la propiedad de insesgaredad de ambos. Por lo tanto,  $\phi(X)$  es localmente óptimo e insesgado con respecto a cualquier otro test insesgado y de igual tamaño, si la segunda derivada de su función potencia es la mayor en un entorno de  $\theta_0$ .

## 2.4. Invarianza

Los modelos estadísticos con los que se trabaja en la práctica pueden contener una gran cantidad de parámetros distintos de aquél sobre el que se realiza el contraste,  $\theta$ . Por lo tanto, resultaría útil definir algún tipo de transformación sobre el modelo que elimine estos parámetros *de ruido* que pueden complicar la derivación del contraste. Una vez definida esta transformación, se debería comprobar que tanto el modelo estadístico como la hipótesis nula no han cambiado sustancialmente.

Un modelo estadístico  $(X, F(x|\theta), \Omega)$  es invariante ante un grupo de transformaciones  $G$  si  $F(g(X)|\theta) = F(X|\theta')$  con  $\theta' \in \Omega$  y  $g \in G$ . Es decir, la transformación conserva la misma distribución que la variable original, pero implica un cambio en el parámetro que la caracteriza. De igual forma, si la transformación actúa sobre el parámetro de interés, se podría recuperar  $\theta' = \bar{g}(\theta)$  con  $\theta' \in \Omega$  y  $\bar{g} \in G$ . Por lo tanto, un problema de contraste  $H_0 : \theta \in \Omega_0$  frente a  $H_1 : \theta \in \Omega_1$  es invariante ante transformaciones del tipo  $G$  si el modelo estadístico es invariante y si  $\theta' \in \Omega_0 \forall \bar{g} \in G$ . Esto es, si el espacio paramétrico de la hipótesis nula no cambia bajo la transformación considerada.

En esta tesis el interés se centra en encontrar contrastes invariantes. Un contraste  $\phi(X)$  es invariante ante transformaciones del tipo  $G$  si

$$\phi(g(x)) = \phi(x), \quad \forall x \in X, \forall g \in G.$$

Es decir, que la probabilidad de rechazo no debería cambiar ante transformaciones del modelo definidas por  $G$ , de forma que daría igual realizar el contraste sobre  $x$  que sobre  $g(x)$ . Cualquier contraste invariante se define a partir de un *maximal invariante*, definido

como una función  $T(\cdot)$  tal que  $T(x) = T(x') \Leftrightarrow x' = g(x)$ ,  $\forall g \in G$ . Es decir, un maximal invariante es una función invariante ante el grupo de transformaciones  $G$ , pero con un valor distinto para cada miembro del grupo. Cualquier contraste invariante es función de un maximal invariante, por lo que el procedimiento para obtener contrastes óptimos e invariantes consiste en definir un maximal invariante ante el grupo de transformaciones de interés y, una vez definida su función de densidad, aplicar los resultados anteriores sobre ella. Cuando, como en los casos considerados en la tesis, exista más de un maximal invariante, el contraste se define a partir de uno de ellos.

## 2.5. Aplicaciones

### 2.5.1. El modelo de regresión univariante

King (1980) o King y Hillier (1985) derivan una metodología para derivar contrastes localmente óptimos para detectar determinadas estructuras de la matriz de covarianzas en un modelo de regresión

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}, \quad \mathbf{u} \sim N(\mathbf{0}, \boldsymbol{\Omega}(\theta)), \quad (2.8)$$

donde  $\mathbf{y}$  es un vector  $n \times 1$ ,  $\mathbf{X}$  es una matriz fija de dimensiones  $n \times k$ ,  $\boldsymbol{\beta}$  es un vector de parámetros  $k \times 1$  y  $\mathbf{u}$  el vector de perturbaciones. Se define por  $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$  el estimador de mínimos cuadrados ordinarios de  $\boldsymbol{\beta}$  y el correspondiente vector de residuos  $\hat{\mathbf{u}} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} \equiv \mathbf{M}\mathbf{y}$ , siendo  $\mathbf{M} = \mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$  la matriz generadora de residuos, simétrica e idempotente. Si el vector de perturbaciones del modelo tiene una matriz de covarianzas que depende del escalar  $\theta$  de tal manera que  $\boldsymbol{\Omega}(0) = \mathbf{I}_n$  y  $\boldsymbol{\Omega} \equiv \boldsymbol{\Omega}(\theta_1)$  es una matriz fija y definida positiva, una hipótesis de interés consiste de determinar si  $\theta = 0$ , lo que implicaría que las perturbaciones del modelo son *iid*.

Kadiyala (1970) deriva un maximal invariante ante cambios de origen y escala del tipo  $\mathbf{y} \mapsto a\mathbf{y} + \mathbf{X}\mathbf{b}$  donde  $a$  es un escalar positivo y  $\mathbf{b}$  un vector fijo de dimensiones  $k \times 1$ . Si  $\mathbf{P}$  es la matriz de autovectores de  $\mathbf{M}$ , entonces, al ser ésta simétrica e idempotente de rango  $n-k$ ,  $\mathbf{P}\mathbf{M}\mathbf{P}' = \text{diag}(\mathbf{I}_{n-k}, \mathbf{0}_k)$  con  $\mathbf{P}\mathbf{P}' = \mathbf{P}'\mathbf{P} = \mathbf{I}_n$ . Se define la partición  $\mathbf{P} = [\mathbf{P}'_1 | \mathbf{P}'_2]'$ , donde  $\mathbf{P}_1$  es una matriz  $(n-k) \times n$  que contiene los autovectores correspondientes a los autovalores no nulos de  $\mathbf{M}$ , tal que  $\mathbf{P}_1\mathbf{P}'_1 = \mathbf{I}_{n-k}$ , y  $\mathbf{P}_2$  es una matriz  $k \times n$  tal que  $\mathbf{P}'_1\mathbf{P}_2 = \mathbf{0}$ . De lo anterior se deduce que  $\mathbf{P}\mathbf{M} = [\mathbf{P}'_1 | \mathbf{0}]'$  y, dado que  $\mathbf{P}'_1\mathbf{P}_1 = \mathbf{I}_n - \mathbf{P}'_2\mathbf{P}_2$ , premultiplicando por  $\mathbf{M}$  y teniendo en cuenta que  $\mathbf{P}_2\mathbf{M} = \mathbf{0}$ , entonces  $\mathbf{M} = \mathbf{P}'_1\mathbf{P}_1$ . De esta forma, King (1980) y Kadiyala (1970) obtienen el maximal invariante ante cambios

de origen y escala definidas anteriormente como

$$\mathbf{w} = \mathbf{P}_1 \mathbf{u} / (\mathbf{u}' \mathbf{P}'_1 \mathbf{P}_1 \mathbf{u})^{1/2}. \quad (2.9)$$

La función de densidad de (2.9) se derivará por partes. En primer lugar,  $\mathbf{z} \equiv \mathbf{P}_1 \mathbf{u}$  es un maximal invariante ante cambios de origen cuya función de densidad es, asumiendo normalidad en las perturbaciones,

$$f(\mathbf{z}) = \frac{1}{\sqrt{2\pi}} |\mathbf{P}_1 \boldsymbol{\Omega}(\theta) \mathbf{P}'_1|^{-1/2} \exp(\mathbf{z}' [\mathbf{P}_1 \boldsymbol{\Omega}(\theta) \mathbf{P}'_1]^{-1} \mathbf{z}).$$

Haciendo el cambio de variable de  $\mathbf{z}$  a  $\mathbf{w} = \mathbf{z} / (\mathbf{z}' \mathbf{z})^{1/2}$ , cuyo jacobiano es  $(\mathbf{w}' \mathbf{w})^{n-k-1}$ , se obtiene su función de densidad como

$$f(\mathbf{w}) = \frac{1}{\sqrt{2\pi}} |\mathbf{P}_1 \boldsymbol{\Omega}(\theta) \mathbf{P}'_1|^{-1/2} (\mathbf{w}' \mathbf{w})^{(n-k-1)/2} \exp(\mathbf{w}' [\mathbf{P}_1 \boldsymbol{\Omega}(\theta) \mathbf{P}'_1]^{-1} \mathbf{w}).$$

Bajo  $H_0$ ,  $\mathbf{w}$  sigue una distribución uniforme tal que  $f(\mathbf{w}|0) = 0,5\Gamma((n-k)/2)\pi^{-(n-k)/2}$ . Para cualquier otro valor de  $\theta$  la función de densidad de (2.9) se puede expresar con respecto a la medida uniforme,

$$f(\mathbf{w}|\theta) = \frac{1}{2} \Gamma((n-k)/2) \pi^{-(n-k)/2} |\mathbf{P}_1 \boldsymbol{\Omega}(\theta_1) \mathbf{P}'_1|^{-1/2} [\mathbf{w}' (\mathbf{P}_1 \boldsymbol{\Omega}(\theta) \mathbf{P}'_1)^{-1} \mathbf{w}]^{-(n-k)/2}, \quad (2.10)$$

detalles adicionales sobre su derivación e interpretación se dan en Kadiyala (1970). La principal ventaja de esta expresión es que no depende de ningún supuesto sobre la distribución de los errores del modelo, con lo que los contrastes pueden derivarse a partir de (2.10) en situaciones generales.

Para contrastar  $H_0 : \theta = 0$  frente a una alternativa simple del tipo  $H_1 : \theta = \theta_1 > 0$ , el teorema de Neymann-Pearson define la siguiente región crítica

$$\mathbf{w}' (\mathbf{P}'_1 \boldsymbol{\Omega}(\theta_1) \mathbf{P}_1)^{-1} \mathbf{w} \equiv \frac{\tilde{\mathbf{u}}' \boldsymbol{\Omega}(\theta_1) \tilde{\mathbf{u}}}{\hat{\mathbf{u}}' \hat{\mathbf{u}}} < c_1,$$

donde  $c_1$  es una constante y  $\tilde{\mathbf{u}}$  es el vector de residuos de mínimos cuadrados generalizados para la matriz de covarianzas  $\boldsymbol{\Omega}(\theta_1)$  (ver King, 1980). El valor del estadístico depende de  $\theta$  a través de  $\boldsymbol{\Omega}(\theta)$ , por lo que no existe un contraste uniformemente más potente e

invariante para problemas del tipo

$$H_0 : \theta = 0, \quad \text{frente a} \quad H_1 : \theta > 0. \quad (2.11)$$

En este caso sí existe un test localmente óptimo e invariante (LBI), cuya región crítica se obtiene aplicando (2.6) a (2.10):

$$-\frac{1}{2} \frac{\partial \log |\mathbf{P}_1 \boldsymbol{\Omega}(\theta_1) \mathbf{P}'_1|}{\partial \theta} \Big|_{\theta=0} - \frac{(n-k)}{2} \frac{\partial \log [\mathbf{w}'(\mathbf{P}_1 \boldsymbol{\Omega}(\theta) \mathbf{P}'_1)^{-1} \mathbf{w}]}{\partial \theta} \Big|_{\theta=0} > c.$$

El primer término es constante, y del segundo se obtiene la región crítica como

$$s \equiv \frac{\hat{\mathbf{u}}' \mathbf{A} \hat{\mathbf{u}}}{\hat{\mathbf{u}}' \hat{\mathbf{u}}} > c \quad \mathbf{A} = \frac{\partial \boldsymbol{\Omega}(\theta)^{-1}}{\partial \theta} \Big|_{\theta=\theta_0}.$$

Cuando  $\boldsymbol{\Omega}(\theta)$  es definida positiva para valores positivos y negativos de  $\theta$  el problema de interés será el contraste a dos colas

$$H_0 : \theta = 0, \quad \text{frente a} \quad H_1 : \theta \neq 0. \quad (2.12)$$

En este caso existe un test localmente óptimo, invariante e insesgado (LBIU) cuya región crítica, aplicando (2.7) a (2.10) se obtiene como

$$\frac{\partial^2 f(\mathbf{w}|\theta)}{\partial \theta^2} \Big|_{\theta=0} > c_1 f(\mathbf{w}|0) + c_2 \frac{\partial f(\mathbf{w}|\theta)}{\partial \theta} \Big|_{\theta=0},$$

Una expresión concreta del estadístico y diversos casos particulares pueden obtenerse en King y Hillier (1985), quien también demuestra que en este caso, el estadístico del contraste se puede derivar como un contraste de multiplicadores de Lagrange.

### 2.5.2. El modelo de regresión multivariante

Una extensión natural de (2.8) es el modelo de regresión multivariante considerado por Nyblom (2001),

$$\mathbf{Y} = \mathbf{X}\mathbf{B} + \mathbf{U}, \quad \mathbf{U} \sim N(\mathbf{0}, \boldsymbol{\Omega}(\theta) \otimes \boldsymbol{\Sigma}), \quad (2.13)$$

donde las matrices  $\mathbf{Y}$  y  $\mathbf{E}$  son de dimensiones  $n \times m$ ,  $\mathbf{X}$  es de dimensiones  $n \times k$  y  $\mathbf{B}$  es  $k \times m$ . Cada una de las filas de  $\mathbf{Y}$ ,  $\mathbf{y}_t = (y_{1t}, \dots, y_{mt})'$ , se interpreta como el vector de  $m$  variables aleatorias en el momento  $t$ . Las matrices  $\mathbf{\Omega}(\theta)$  y  $\mathbf{\Sigma}$  son matrices definidas positivas de dimensiones  $n \times n$  y  $m \times m$  respectivamente y  $\otimes$  denota el producto de Kronecker. El parámetro  $\theta$  es tal que si  $\theta = 0$ ,  $\mathbf{\Omega}(0) = \mathbf{I}_n$ , luego los errores son *iid*, donde la única correlación presente es la contemporánea. Nótese que para  $m = 1$  el modelo se reduce a (2.8), por lo que los procedimientos que se expondrán a continuación podrían verse como la extensión multivariante de King (1980) y King y Hillier (1985). En este caso, la invarianza se considera con respecto a transformaciones del tipo

$$\mathbf{Y} \mapsto \mathbf{Y}\mathbf{P} + \mathbf{X}\mathbf{A}, \quad (2.14)$$

donde  $\mathbf{P}$  es una matriz  $m \times m$  no singular y  $\mathbf{A}$  es una matriz arbitraria de dimensiones  $k \times m$ . Siguiendo a Nyblom (2001), se deriva primero el maximal invariante y su función de densidad para discutir posteriormente la derivación de los contrastes óptimos. De la sección anterior, se sabe que existe una matriz  $\mathbf{P}_1$  de dimensiones  $(n - k) \times n$  tal que  $\mathbf{P}_1\mathbf{P}'_1 = \mathbf{I}_{n-k}$  y  $\mathbf{P}'_1\mathbf{P}_1 = \mathbf{M}$ , de forma que la matriz  $\mathbf{Z} = \mathbf{P}_1\mathbf{Y}$  es un maximal invariante ante transformaciones del tipo  $\mathbf{Y} \mapsto \mathbf{Y} + \mathbf{X}\mathbf{B}$ . Por lo tanto, un maximal invariante ante transformaciones  $\mathbf{Z} \mapsto \mathbf{Z}\mathbf{P}$  será un maximal invariante ante (2.14). En este caso Nyblom (2001) define  $\mathbf{W} = [\mathbf{z}_1 \dots \mathbf{z}_m]$  y propone como maximal invariante

$$\mathbf{v}_t = \mathbf{W}^{-1}\mathbf{z}_t, \quad t = m + 1, \dots, n - k. \quad (2.15)$$

La derivación de la distribución de  $\mathbf{v}_t$  también se realiza por partes. En primer lugar, se deriva la distribución de  $\mathbf{z}_t$ , y mediante un cambio de variable, la del maximal invariante. Denotando el vector  $((n - k)m \times 1)$ ,  $\mathbf{z} \equiv (\mathbf{P}_1 \otimes \mathbf{I}_m)\text{vec}(\mathbf{Y}')$ , cuya matriz de covarianzas es  $\text{cov}(\mathbf{z}) \equiv \mathbf{F} = [(\mathbf{P}_1\mathbf{\Omega}(\theta)\mathbf{P}'_1) \otimes \mathbf{\Sigma}]$ , bajo normalidad de las perturbaciones su función de densidad es

$$f(\mathbf{z}) = (2\pi)^{-m(n-k)/2} |\mathbf{P}_1\mathbf{\Omega}(\theta)\mathbf{P}'_1|^{-m/2} |\mathbf{\Sigma}|^{-(n-k)/2} \exp\left(-\frac{1}{2}\mathbf{z}'[(\mathbf{P}_1\mathbf{\Omega}(\theta)\mathbf{P}'_1)^{-1} \otimes \mathbf{\Sigma}^{-1}]\mathbf{z}\right).$$

Haciendo el cambio de variable  $\mathbf{y} \mapsto \mathbf{u} = (\mathbf{w}'_1, \dots, \mathbf{w}'_m, \mathbf{v}'_{m+1}, \dots, \mathbf{v}'_{n-k})'$ , donde  $\mathbf{w}_j$  son las filas de  $\mathbf{W}$  y  $\mathbf{v}_t$  el maximal invariante (2.15). Dado que el determinante del jacobiano de la transformación es  $|\mathbf{W}|^{n-k-m} = |\mathbf{W}'\mathbf{W}|^{(n-k-m)/2}$  y que, por la propiedad de invarianza se puede suponer  $\mathbf{\Sigma} = \mathbf{I}_m$ , la función de densidad de  $\mathbf{v}_t = (\mathbf{v}'_{m+1}, \dots, \mathbf{v}'_{n-k})'$

condicionado a  $\mathbf{w} = (\mathbf{w}'_1, \dots, \mathbf{w}'_m)'$  es

$$\begin{aligned}
f(\mathbf{u}|\mathbf{v}) &= (2\pi)^{-m(n-k)/2} |\mathbf{P}_1 \boldsymbol{\Omega}(\theta) \mathbf{P}'_1|^{-m/2} \times \\
&\quad \times \int |\mathbf{W}' \mathbf{W}|^{(n-k-m)/2} \exp\left(-\frac{1}{2} \mathbf{u}' [(\mathbf{P}_1 \boldsymbol{\Omega}(\theta) \mathbf{P}'_1)^{-1} \otimes \mathbf{W}' \mathbf{W}] \mathbf{u}\right) d\mathbf{w}, \\
&= (2\pi)^{-m(n-k)/2} |\mathbf{P}_1 \boldsymbol{\Omega}(\theta) \mathbf{P}'_1|^{-m/2} \times \\
&\quad \times \int |\mathbf{W}' \mathbf{W}|^{(n-k-m)/2} \exp\left(-\frac{1}{2} \text{tr}[(\mathbf{U}' (\mathbf{P}_1 \boldsymbol{\Omega}(\theta) \mathbf{P}'_1)^{-1} \mathbf{U}) \mathbf{W}' \mathbf{W}]\right) d\mathbf{w}, \\
&\propto |\mathbf{U}' (\mathbf{P}_1 \boldsymbol{\Omega}(\theta) \mathbf{P}'_1)^{-1} \mathbf{U}|^{(n-k)/2}. \tag{2.16}
\end{aligned}$$

Esta última expresión se deriva de la siguiente propiedad de la distribución Wishart (Anderson, 1951, problema 6, página 176). Sea  $\mathbf{A}$  una matriz con distribución Wishart  $W_p(\boldsymbol{\Sigma}, n)$ , entonces

$$\int \frac{|\mathbf{A}|^{(n-p-1)/2} \exp\left[-\frac{1}{2} \text{tr}(\mathbf{B}^{-1} \mathbf{A})\right] d\mathbf{A}}{2^{np/2} \pi^{p(p-1)/4} \prod_{i=1}^p \Gamma((n+1-i)/2)} = |\mathbf{B}|^{n/2}.$$

Por lo tanto, el teorema de Neymann-Pearson aplicado con (2.16) proporciona la siguiente región de rechazo para contrastar  $H_0 : \theta = 0$  frente a la alternativa simple de  $H_1 : \theta = \theta_1 > 0$ ,

$$S = \frac{|\mathbf{U}' (\mathbf{P}_1 \boldsymbol{\Omega}(\theta_1) \mathbf{P}'_1)^{-1} \mathbf{U}|}{|\mathbf{U}' \mathbf{U}|} \equiv \frac{|\tilde{\mathbf{E}}' \boldsymbol{\Omega}(\theta_1)^{-1} \tilde{\mathbf{E}}|}{|\hat{\mathbf{E}}' \hat{\mathbf{E}}|} < c,$$

donde  $\hat{\mathbf{E}} = [\mathbf{I}_n - \mathbf{X}(\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}'] \mathbf{Y}$  y  $\tilde{\mathbf{E}} = [\mathbf{I}_n - \mathbf{X}(\mathbf{X}' \boldsymbol{\Omega}(\theta_1)^{-1} \mathbf{X})^{-1} \mathbf{X}' \boldsymbol{\Omega}(\theta_1)^{-1}] \mathbf{Y}$ . Como en el caso univariante la forma de  $\boldsymbol{\Omega}(\theta)$  determinará si  $S$  es UMPI. Cuando no sea así, el estadístico localmente óptimo e invariante (LBI) se puede obtener utilizando la primera derivada del logaritmo de (2.16). Sabemos que  $d \log |\mathbf{A}|/dx = \text{tr}(\mathbf{A}^{-1} d\mathbf{A}/dx)$ , luego el estadístico se obtiene como

$$\begin{aligned}
&\frac{\partial \log |\mathbf{U}' (\mathbf{P}_1 \boldsymbol{\Omega}(\theta) \mathbf{P}'_1)^{-1} \mathbf{U}|}{\partial \theta} \Big|_{\theta=0} = \\
&= -\frac{(n-k)}{2} \text{tr} \left[ (\mathbf{U}' (\mathbf{P}_1 \boldsymbol{\Omega}(0) \mathbf{P}'_1)^{-1} \mathbf{U})^{-1} \mathbf{U}' \left( \mathbf{P}_1 \frac{\partial \boldsymbol{\Omega}(\theta)}{\partial \theta} \Big|_{\theta=0} \mathbf{P}'_1 \right)^{-1} \mathbf{U} \right]. \tag{2.17}
\end{aligned}$$

Por su parte, para el mismo problema, Forchini (2005) propone utilizar como maximal invariante,  $\mathbf{e}_t = (\mathbf{Y}' \mathbf{M} \mathbf{Y})^{-1/2} \mathbf{P}_1 \mathbf{Y}$ , donde la analogía con el maximal invariante del caso univariante es evidente. Sin embargo, la forma de su función de densidad es bastante más complicada, lo que dificulta la derivación de los estadísticos de contraste.



## Capítulo 3

# Contrastes de no invertibilidad en el modelos VARIMA

### 3.1. Introducción

Nyblom y Mäkeläinen (1983) derivaron el contraste localmente óptimo e invariante (LBI) ante cambios de origen y escala para detectar la presencia de un nivel determinista el modelo de paseo aleatorio más ruido univariante,

$$y_t = \mu_t + u_t, \quad \mu_t = \mu_{t-1} + v_t, \quad t = 1, \dots, n,$$

siendo  $u_t \sim N(0, \sigma_u^2)$  y  $v_t \sim N(0, \sigma_v^2)$  dos ruidos blancos independientes. El problema de contraste se plantea como  $H_0 : \rho = \sigma_v^2 / \sigma_u^2 = 0$  frente a  $H_1 : \rho > 0$ , y el estadístico de contraste se obtiene utilizando la metodología de King y Hillier (1985) descrita en el capítulo anterior. Mazas (2012) demuestra que este estadístico coincide con el derivado por Tanaka (1990) para el contraste localmente óptimo, insesgado e invariante (LBIU) de no invertibilidad o raíz unitaria media móvil en el marco de un modelo IMA(1,1),

$$(1 - B)y_t = (1 - \theta B)a_t, \quad a_t \sim N(0, \sigma_a^2),$$

donde  $By_t = y_{t-1}$ , siendo el problema de contraste en este modelo  $H_0 : \theta = 1$  frente a  $H_1 : \theta < 1$ . Es conocido que el modelo de nivel local admite una representación IMA(1,1) con  $\rho = (1 - \theta)^2 / \theta$ , por lo que un nivel determinista ( $\rho = 0$ ) en la representación estructural de  $y_t$  implica una raíz unitaria en el polinomio media móvil de su forma reducida IMA(1,1). En ambos casos, la aceptación de la hipótesis nula implica que  $y_t$  es una serie estacionaria alrededor de un nivel fijo  $\mu_0$ . Dado que el propósito de la tesis es

derivar contrastes cuya hipótesis nula sea la de cointegración, resulta necesario extender el contraste de no invertibilidad de Tanaka al caso multivariante.

El contraste de Nyblom y Mäkeläinen (1983) ha sido extendido por Nyblom y Harvey (2000) para detectar la presencia de un vector de niveles constantes en el modelo de nivel local multivariante. En este capítulo se obtiene la extensión multivariante del contraste de Tanaka, se muestra su relación con el contraste de Nyblom y Harvey y la equivalencia entre los estadísticos de ambos. Para derivar su estadístico de contraste, Nyblom y Harvey (2000) proponen una extensión multivariante de la metodología de King y Hillier (1985), desarrollada posteriormente en Nyblom (2001). En este capítulo se muestra cómo este mismo estadístico puede derivarse de forma más sencilla a partir de la función de verosimilitud del modelo de nivel local, sin necesidad de definir un maximal invariante. Por su parte, la expresión del estadístico derivado a partir de la forma IMA del proceso resulta más conveniente para caracterizar su distribución en muestras finitas. Por lo tanto, dada la equivalencia entre ambas representaciones, la metodología que se propone en esta tesis para derivar los contrastes de no invertibilidad consiste en utilizar la representación estructural para derivar el estadístico de contraste y la forma ARIMA para caracterizar su distribución. La equivalencia entre los estadísticos derivados en ambas representaciones se demuestra fácilmente utilizando una serie de equivalencias matriciales descritas en Mazas (2012).

Si el contraste de no invertibilidad univariante se realiza sobre un modelo  $ARIMA(p,d,q)$ , Mazas (2012) propone una corrección paramétrica del estadístico de Tanaka basada en los residuos exactos del modelo evaluados bajo la hipótesis nula de no invertibilidad, en la línea de las propuestas de Saikkonen y Luukkonen (1993a) o Tam y Reinsel (1997). En este capítulo se propone también la extensión multivariante de esta corrección paramétrica y se evalúa la potencia del contraste en muestras finitas. Esta corrección paramétrica no sólo es útil cuando el contraste se realiza sobre un modelo  $VARIMA(p,d,q)$ , sino también para una serie de casos de especial interés práctico que se estudiarán en el apartado final.

La estructura del capítulo es la siguiente: En la segunda sección se resumen los principales resultados del contraste de Nyblom y Harvey (2000), con especial atención al proceso de derivación del estadístico. En el tercer apartado se deriva el contraste de no invertibilidad y se caracteriza su distribución bajo la hipótesis nula y bajo una secuencia de alternativas y se propone un procedimiento alternativo para la derivación del estadístico de Nyblom y Harvey (2000). En el cuarto apartado se extiende al caso multivariante la corrección paramétrica de Mazas (2012) y se evalúa su potencia en muestras finitas mediante una serie de experimentos Monte Carlo. Por último, se considera el contraste de no invertibilidad en un conjunto de modelos de especial interés práctico.

## 3.2. El contraste de Nyblom y Harvey

### 3.2.1. Modelo y estadístico de contraste

El modelo de nivel local o paseo aleatorio más ruido blanco para el vector  $m$ -dimensional de series temporales  $\mathbf{y}_t = (y_{1t}, \dots, y_{mt})'$  se define como,

$$\mathbf{y}_t = \boldsymbol{\mu}_t + \mathbf{u}_t, \quad \mathbf{u}_t \sim N(\mathbf{0}, \boldsymbol{\Sigma}_u), \quad (3.1)$$

$$\boldsymbol{\mu}_t = \boldsymbol{\mu}_{t-1} + \mathbf{v}_t, \quad \mathbf{v}_t \sim N(\mathbf{0}, \boldsymbol{\Sigma}_v) \quad t = 1, \dots, n, \quad (3.2)$$

donde  $\boldsymbol{\mu}_t = (\mu_{1t}, \dots, \mu_{mt})'$  es un paseo aleatorio,  $\mathbf{u}_t$  y  $\mathbf{v}_t$  son dos vectores ruido blanco independientes y  $\boldsymbol{\Sigma}_u$  es una matriz  $m \times m$  definida positiva. Nyblom y Harvey (2000) consideraron este modelo para derivar el contraste localmente óptimo e invariante (LBI) para detectar la presencia de un vector de niveles determinista, esto es,  $\boldsymbol{\Sigma}_v = \mathbf{0}$  frente a la alternativa de que los niveles son estocásticos. Dado que las matrices  $\boldsymbol{\Sigma}_v$  y  $\boldsymbol{\Sigma}_u$  son reales y simétricas, ambas se pueden diagonalizar simultáneamente (e.g. Rao, 2002, p. 41), de forma que existe una matriz  $m \times m$  no singular  $\mathbf{P}$  tal que  $\mathbf{P}\boldsymbol{\Sigma}_u\mathbf{P}' = \mathbf{I}_m$  y  $\mathbf{P}\boldsymbol{\Sigma}_v\mathbf{P}' = \mathbf{R} = \text{diag}\{\rho_1, \dots, \rho_m\}$ , donde  $\rho_j$  son las raíces de  $|\boldsymbol{\Sigma}_u - \rho_j\boldsymbol{\Sigma}_v| = 0$  ( $j = 1, \dots, m$ ) y la  $j$ -ésima columna de  $\mathbf{P}$  el correspondiente autovector. El nivel local será determinista cuando  $\rho_1 = \dots = \rho_m = 0$  y estocástico cuando  $\rho_1 = \dots = \rho_m = \rho > 0$ , supuesto que permite expresar  $\boldsymbol{\Sigma}_v = \rho\boldsymbol{\Sigma}_u$ . El parámetro  $\rho$  controla el grado de estocasticidad de  $\boldsymbol{\mu}_t$ , de forma que si  $\rho = 0$  el proceso  $\mathbf{y}_t$  es estacionario alrededor de un nivel fijo  $\boldsymbol{\mu}_0$  y si  $\rho > 0$  todos los  $m$  niveles son estocásticos. Por lo tanto, el problema de decisión se especifica como

$$H_0 : \rho = 0, \quad \text{frente a} \quad H_1 : \rho > 0. \quad (3.3)$$

Para derivar el estadístico de contraste conviene escribir el modelo en forma matricial como  $\mathbf{Y} = \mathbf{i}_n\boldsymbol{\mu}'_0 + \mathbf{E}$  en donde se supone que el vector de condiciones iniciales  $\boldsymbol{\mu}_0 = (\mu_{10}, \dots, \mu_{m0})'$  es fijo,  $\mathbf{i}_n$  es un vector  $n \times 1$  de unos,  $\mathbf{Y} = [\mathbf{y}_1, \dots, \mathbf{y}_n]'$  y  $\mathbf{E} = [\mathbf{e}_1, \dots, \mathbf{e}_n]'$  tal que  $\mathbf{e}_t = \mathbf{u}_t + \sum_{i=1}^t \mathbf{v}_i$ . Por lo tanto,  $\mathbf{Y} \sim N(\mathbf{i}_n\boldsymbol{\mu}'_0, \boldsymbol{\Omega}_y(\rho))$ , donde la matriz de covarianzas de la matriz  $\mathbf{Y}$  es  $\boldsymbol{\Omega}_y(\rho) = (\mathbf{I}_n + \rho\mathbf{C}_n\mathbf{C}'_n) \otimes \boldsymbol{\Sigma}_u$ , denotando  $\otimes$  el producto de Kronecker y siendo  $\mathbf{C}_n$  una matriz triangular inferior de unos de orden  $n$ . Por simplicidad se denotará en adelante  $\boldsymbol{\Omega}(\rho) = \mathbf{I}_n + \rho\mathbf{C}_n\mathbf{C}'_n$ . El problema de contraste es invariante ante transformaciones del tipo  $\mathbf{y}_t \mapsto \mathbf{P}\mathbf{y}_t + \mathbf{a}$ , siendo  $\mathbf{P}$  cualquier matriz cuadrada no singular y  $\mathbf{a} \neq \mathbf{0}$  un vector constante. La hipótesis nula sigue implicando la presencia un vector de niveles determinista si se transforma el vector de series temporales de esta manera. Para derivar el estadístico óptimo ante transformaciones de este tipo, Nyblom y Harvey

(2000) obtienen un maximal invariante y su función de densidad de forma secuencial. Si  $\mathbf{P}_1$  es cualquier matriz  $(n-1) \times n$  tal que  $\mathbf{P}_1 \mathbf{i}_n = \mathbf{0}$ ,  $\mathbf{P}_1 \mathbf{P}'_1 = \mathbf{I}_{n-1}$  y  $\mathbf{P}'_1 \mathbf{P}_1 = \mathbf{M}$  siendo  $\mathbf{M} = \mathbf{I}_n - \mathbf{i}_n (\mathbf{i}'_n \mathbf{i}_n)^{-1} \mathbf{i}'_n$  la matriz de proyección sobre el espacio generado por un vector de unos, entonces  $\mathbf{U} = \mathbf{P}_1 \mathbf{Y}$  es un maximal invariante ante cambios de origen,  $\mathbf{y}_t \mapsto \mathbf{y}_t + \mathbf{a}$ . Posteriormente, definiendo  $\mathbf{U}_0 = [\mathbf{u}_1, \dots, \mathbf{u}_m]'$  como una matriz  $m \times m$  que contiene las primeras filas de  $\mathbf{U}$ , entonces  $\mathbf{z}_t = \mathbf{U}_0^{-1} \mathbf{u}_t$  ( $t = 1, \dots, n-1$ ) es un maximal invariante ante  $\mathbf{u}_t \mapsto \mathbf{P} \mathbf{u}_t$ . Dado que  $\mathbf{z}_t$  está construido a partir de un maximal invariante ante cambios de escala,  $\mathbf{z}_t$  también es invariante ante este tipo de transformaciones, por lo que  $\mathbf{z}_t$  es un maximal invariante ante las transformaciones definidas inicialmente. Dada la propiedad de invarianza se puede considerar, sin pérdida de generalidad,  $\Sigma_u = \mathbf{I}_m$  con lo que la matriz de covarianzas de  $\mathbf{Z} = [\mathbf{z}_1, \dots, \mathbf{z}_{n-1}]'$  es  $\mathbf{U}_0^{-1} (\mathbf{P}_1 \Omega(\rho) \mathbf{P}'_1) \mathbf{U}_0^{-1'}$ . De los resultados de Nyblom (2001) resumidos en el capítulo anterior, se sabe que la función de densidad de  $\mathbf{Z} = [\mathbf{z}_1, \dots, \mathbf{z}_{n-1}]'$  es

$$f(\mathbf{Z}|\rho) \propto |\mathbf{Z}'(\mathbf{P}_1 \Omega(\rho) \mathbf{P}'_1)^{-1} \mathbf{Z}|^{(n-1)/2}.$$

Dado que  $\mathbf{P}_1 \Omega(0) \mathbf{P}'_1 = \mathbf{P}_1 \mathbf{P}'_1 = \mathbf{I}_{n-1}$ , y que

$$\left. \frac{d(\mathbf{P}_1 \Omega(\rho) \mathbf{P}'_1)^{-1}}{d\rho} \right|_{\rho=0} = (\mathbf{P}_1 \Omega(0) \mathbf{P}'_1)^{-1} \mathbf{P}_1 \left. \frac{d\Omega(\rho)}{d\rho} \right|_{\rho=0} \mathbf{P}'_1 (\mathbf{P}_1 \Omega(0) \mathbf{P}'_1)^{-1} = \mathbf{P}_1 \mathbf{C}_n \mathbf{C}'_n \mathbf{P}'_1,$$

el estadístico de contraste LBI se obtiene a partir de la primera derivada del logaritmo de la función de densidad del maximal invariante evaluada bajo  $H_0$ , cuya expresión es

$$\left. \frac{d \log f(\rho|\mathbf{Z})}{d\rho} \right|_{\rho=0} = \frac{(n-1)}{2} \text{tr} [(\mathbf{Z}'(\mathbf{P}_1 \Omega(0) \mathbf{P}'_1)^{-1} \mathbf{Z})^{-1} \mathbf{Z}' \mathbf{P}_1 \mathbf{C}_n \mathbf{C}'_n \mathbf{P}'_1 \mathbf{Z}].$$

La propiedad de invarianza ante premultiplicaciones de matrices no singulares, que se hace evidente en la expresión anterior, también permite reemplazar  $\mathbf{Z}$  por  $\mathbf{U}$  dada la relación entre ambas matrices antes definida, con lo que el primer término de la traza se puede expresar como  $\mathbf{Z}'(\mathbf{P}_1 \Omega(0) \mathbf{P}'_1)^{-1} \mathbf{Z} = \mathbf{U}' \mathbf{U} = \mathbf{Y}' \mathbf{P}'_1 \mathbf{P}_1 \mathbf{Y}$ . De forma análoga, el segundo término de la traza se puede expresar como  $\mathbf{Z}' \mathbf{P}_1 \mathbf{C}_n \mathbf{C}'_n \mathbf{P}'_1 \mathbf{Z} = \mathbf{Y}' \mathbf{P}'_1 \mathbf{P}_1 \mathbf{C}_n \mathbf{C}'_n \mathbf{P}'_1 \mathbf{P}_1 \mathbf{Y}$ . Teniendo en cuenta que  $\mathbf{P}'_1 \mathbf{P}_1 = \mathbf{M}$  y dividiendo la primera derivada por  $(n-1)n$ , se obtiene el estadístico del contraste LBI como

$$NH = \text{tr}[(\mathbf{Y}' \mathbf{M} \mathbf{Y})^{-1} \mathbf{Y}' \mathbf{M} \mathbf{C}_n \mathbf{C}'_n \mathbf{M} \mathbf{Y}] / n, \quad (3.4)$$

donde  $n^{-1}\mathbf{Y}'\mathbf{M}\mathbf{Y}$  es la matriz de covarianzas de los residuos de mínimos cuadrados multivariantes de  $\mathbf{y}_t$  sobre un término constante, por lo que (3.4) puede expresarse alternativamente como

$$NH = \frac{1}{n} \text{tr} \left[ \left( \sum_{t=1}^n (\mathbf{y}_t - \bar{\mathbf{y}})(\mathbf{y}_t - \bar{\mathbf{y}})' \right)^{-1} \sum_{t=1}^n \left( \sum_{i=1}^t (\mathbf{y}_i - \bar{\mathbf{y}}) \right) \left( \sum_{i=1}^t (\mathbf{y}_i - \bar{\mathbf{y}}) \right)' \right],$$

con  $\bar{\mathbf{y}} = n^{-1} \sum_{t=1}^n \mathbf{y}_t$ , el vector de medias de  $\mathbf{y}_t$ . Tanto el problema de contraste como el estadístico son la extensión multivariante de los del contraste de Nyblom y Mäkeläinen (1983). La extensión del procedimiento de contraste hacia modelos más complejos, como por ejemplo la introducción de una deriva en (3.2), en la línea de Nyblom (1986), se puede realizar utilizando un procedimiento similar. Ésta y otras extensiones del modelo de nivel local multivariante se considerarán en el capítulo siguiente.

### 3.2.2. Distribución del estadístico bajo la hipótesis nula

Aunque Nyblom y Harvey (2000) no derivaron la distribución en muestras finitas de (3.4), ésta se puede caracterizar utilizando las siguientes identidades matriciales obtenidas por Mazas (2012)

$$\begin{aligned} \mathbf{M} &= \mathbf{D}'(\mathbf{D}\mathbf{D}')^{-1}\mathbf{D}, \\ \mathbf{M}\mathbf{C}_n\mathbf{C}_n'\mathbf{M} &= \mathbf{D}'(\mathbf{D}\mathbf{D}')^{-2}\mathbf{D}, \end{aligned}$$

donde  $\mathbf{D} = [d_{ij}]$  es la matriz de primeras diferencias que se define como una matriz rectangular de dimensiones  $(n-1) \times n$  tal que  $d_{ij} = -1$  si  $i = j$ ,  $d_{ij} = 1$  si  $i = j-1$  y  $d_{ij} = 0$  en cualquier otro caso. De esta forma el estadístico se puede expresar alternativamente como

$$NH = \text{tr}[(\mathbf{W}'(\mathbf{D}\mathbf{D}')^{-1}\mathbf{W})^{-1}\mathbf{W}'(\mathbf{D}\mathbf{D}')^{-2}\mathbf{W}]/n,$$

donde las columnas de  $\mathbf{W} = \mathbf{D}\mathbf{Y}$  contienen las primeras diferencias de cada una de las  $m$  series temporales. De nuevo, la propiedad de invarianza del estadístico permite suponer sin pérdida de generalidad que  $\Sigma_u = \mathbf{I}_m$ , de forma que bajo  $H_0$ ,  $\mathbf{W} \sim N(\mathbf{0}, (\mathbf{D}\mathbf{D}') \otimes \mathbf{I}_m)$  y de la descomposición espectral de la matriz tridiagonal  $\mathbf{D}\mathbf{D}' = \mathbf{P}\mathbf{\Lambda}\mathbf{P}'$  con  $\mathbf{\Lambda} = \text{diag}\{\lambda_1, \dots, \lambda_{n-1}\}$  y  $\lambda_k = 4\sin^2(2\pi k/n)$  (Anderson, 1971). Todos estos resultados permiten expresar  $\mathbf{W} = \mathbf{P}\mathbf{\Lambda}^{1/2}\mathbf{E}$  donde las filas de  $\mathbf{E} = [\epsilon_1, \dots, \epsilon_{n-1}]$  son procesos

ruido blanco vectoriales tales que  $\boldsymbol{\epsilon}_t \sim N(\mathbf{0}, \mathbf{I}_m)$ , de tal forma que si  $\mathbf{P}'\mathbf{P} = \mathbf{I}_{n-1}$ , (3.4) se puede expresar bajo  $H_0$  como

$$NH = \frac{1}{n} \text{tr} \left[ \left( \sum_{t=1}^{n-1} \boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_t' \right)^{-1} \sum_{t=1}^{n-1} \lambda_t^{-1} \boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_t' \right].$$

Para el caso univariante ( $m = 1$ ) esta expresión se reduce a una ratio de formas cuadráticas en variables normales, cuya distribución se puede evaluar numéricamente mediante los procedimientos de Imhof (1961) o Davies (1973) implementados ambos en **Empiricus**. Para el caso multivariante ( $m > 1$ ) la distribución de  $NH$  deberá ser evaluada por simulación. A partir de esta expresión del estadístico se puede derivar su distribución asintótica. En primer lugar, es inmediato comprobar que  $n^{-1} \sum_{t=1}^{n-1} \boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_t' \xrightarrow{p} \mathbf{I}_m$  y  $\lim_{n \rightarrow \infty} n^{-1} \lambda_k^{-1} \rightarrow 1/(\pi k)^2$  de forma que

$$NH \xrightarrow{d} \sum_{k=1}^{\infty} \frac{1}{(\pi k)^2} \zeta_k \quad \zeta_k = \text{tr}(\boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_t') \sim \chi_m^2.$$

Nótese que para  $m = 1$  esta distribución coincide con la del estadístico de bondad de ajuste de Cramèr-von Mises, caracterizada y tabulados sus percentiles por Anderson y Darling (1952) y por Hansen (1990) para  $m > 1$ . Por esta razón la distribución límite de  $L$  se suele denominar como distribución de Cramèr-von Mises con  $m$  grados de libertad o  $CvM(m)$ . Una expresión alternativa de la distribución asintótica se obtiene a partir de la expresión en términos de desviaciones con respecto a la media, teniendo en cuenta que

$$n^{-1/2} \sum_{i=1}^{[nu]} (\mathbf{y}_i - \bar{\mathbf{y}}) = n^{-1/2} \sum_{i=1}^{[nu]} \boldsymbol{\epsilon}_i - n^{-3/2} \sum_{t=1}^n \boldsymbol{\epsilon}_t.$$

De Phillips y Durlauf (1986),  $n^{-1/2} \sum_{t=1}^{[nu]} \boldsymbol{\epsilon}_t \xrightarrow{d} \mathbf{W}(u)$ ,  $n^{-1/2} \sum_{t=1}^n \boldsymbol{\epsilon}_t \xrightarrow{d} \mathbf{W}(1)$ ,  $[\cdot]$  denota la parte entera y  $[nu]/n = u$  donde  $\mathbf{W}(u) = (W_1(u), \dots, W_m(u))'$  es un proceso browniano vectorial, donde cada uno de sus elementos  $W_j(u)$  es independiente del resto. Por lo tanto, la distribución bajo la hipótesis nula de (3.4) se puede expresar como

$$NH \xrightarrow{d} \int_0^1 \text{tr} [\mathbf{B}(u)\mathbf{B}(u)'] du = \sum_{j=1}^m \int_0^1 B_j^2(u) du,$$

donde  $\mathbf{B}(u) = \mathbf{W}(u) - u\mathbf{W}(1)$  es un puente browniano y  $B_j(u)$  su elemento genérico.

### 3.2.3. Corrección no paramétrica por correlación serial

El estadístico de contraste derivado anteriormente no sería de utilidad para contrastar la presencia de un nivel determinista si el error de la ecuación de observación (3.1) presenta correlación serial. En este caso, si  $\mathbf{u}_t = \Psi(B)\epsilon_t$  donde el polinomio matricial  $\text{MA}(\infty)$   $\Psi(B)$  es invertible, la distribución tanto asintótica como en muestras finitas del estadístico se ven afectadas, dependiendo de las covarianzas de  $\mathbf{u}_t$  con lo que los valores críticos tabulados no se podrían utilizar. En la línea de Kwiatkowski et al. (1992), Nyblom y Harvey (2000) proponen corregir (3.4) sustituyendo la matriz de covarianzas de los residuos  $\hat{\mathbf{u}}_t = (\mathbf{y}_t - \bar{\mathbf{y}})$  por una estimación de la matriz de covarianzas de largo plazo. Estos autores proponen utilizar una estimación no paramétrica de la misma como

$$\Gamma(l) = \sum_{\tau=-l}^l w_{\tau l} \hat{\Gamma}_{\tau},$$

donde  $\hat{\Gamma}_{\tau}$  es la matriz de covarianzas de  $\hat{\mathbf{u}}_t$  en el retardo  $\tau$  y  $w_t = 1 - \tau/(l + 1)$ . De esta forma, el estadístico a utilizar en el caso de presencia de correlación serial sería  $NH(l) = \text{tr}[\Gamma(l)^{-1} \mathbf{Y}' \mathbf{M} \mathbf{C}_n \mathbf{C}'_n \mathbf{M} \mathbf{Y}] / n^2$ . Con respecto al parámetro de truncamiento  $l$ , Andrews (1991) recomienda que  $l \rightarrow \infty$  pero a una tasa menor que  $n$ , y sugiere  $l = o(n^{1/4})$ . En este caso, Nyblom y Harvey (2000) demuestran que la distribución límite del estadístico no se altera.

## 3.3. El contraste básico de no invertibilidad

### 3.3.1. Modelo y derivación del estadístico

Tanaka (1990) deriva el estadístico del contraste localmente óptimo, insesgado e invariante (LBIU) de no invertibilidad univariante a partir de la segunda derivada del logaritmo de la función de verosimilitud del proceso estacionario  $\nabla y_t$ , y demuestra que ésta tiene un mínimo en  $\theta = 1$ . Utilizando este procedimiento se puede derivar un contraste de no invertibilidad estricta para el modelo IMA(1,1) multivariante, cuya expresión es

$$\nabla \mathbf{y}_t = (\mathbf{I}_m - \Theta B) \mathbf{a}_t, \quad t = 2, \dots, n, \quad (3.5)$$

donde  $\mathbf{y}_t = (y_{1t}, \dots, y_{mt})'$  es un vector  $m$ -dimensional de series temporales,  $B$  es el operador retardo tal que  $B \mathbf{y}_t = \mathbf{y}_{t-1}$ ,  $\nabla = (1 - B)$  es el operador diferencia,  $\mathbf{a}_t = (a_{1t}, \dots, a_{mt})' \sim N(\mathbf{0}, \Omega_a)$  es un vector aleatorio ruido blanco con matriz de covarianzas

definida positiva  $\mathbf{\Omega}_a$  y  $\mathbf{\Theta}$  una matriz  $m \times m$  de parámetros. Se dice que el polinomio media móvil matricial es estrictamente no invertible si las  $m$  raíces de  $|\mathbf{I}_m - \mathbf{\Theta}B| = 0$  caen fuera del círculo unitario o, equivalentemente, si todos los autovalores de  $\mathbf{\Theta}$  son menores que 1 en valor absoluto. Para derivar un contraste de no invertibilidad estricta sobre (3.5), se puede suponer sin pérdida de generalidad que  $\mathbf{\Theta}$  tiene una estructura escalar tal que  $\mathbf{\Theta} = \theta \mathbf{I}_m$ , de forma que si el parámetro  $\theta = 1$ , el polinomio MA(1) matricial es estrictamente no invertible. Por lo tanto, el problema a contrastar será

$$H_0 : \theta = 1, \quad \text{frente a} \quad H_1 : \theta \neq 1.$$

Nótese que bajo la hipótesis nula los  $m$  componentes del vector  $\mathbf{y}_t$  están sobrediferenciados, por lo que el contraste puede también interpretarse como una prueba de estacionariedad conjunta de todas las series consideradas. El modelo IMA(1,1) multivariante (3.5) con matriz de coeficientes escalar  $\mathbf{\Theta} = \theta \mathbf{I}_m$  es la forma reducida del modelo homogéneo (3.1)-(3.2) con  $\mathbf{\Sigma}_v = \rho \mathbf{\Sigma}_u$  y que, igual que en el caso univariante,  $\rho = (1 - \theta)/\theta$ . Por lo tanto, la hipótesis nula de no invertibilidad en el marco VIMA escalar equivale a la de nivel determinista en el marco estructural obtenido en el capítulo anterior. Para derivar el estadístico del contraste localmente óptimo, insesgado e invariante (LBIU), donde la invarianza se define ante el tipo de transformaciones consideradas en el apartado anterior ( $\mathbf{y}_t \mapsto \mathbf{P}\mathbf{y}_t + \mathbf{a}$ ), conviene expresar el proceso (3.5) en forma vectorial que, suponiendo una estructura escalar para  $\mathbf{\Theta}$ , es

$$(\mathbf{D} \otimes \mathbf{I}_m)\mathbf{y} = (\mathbf{D}(\theta) \otimes \mathbf{I}_m)\mathbf{a}, \quad (3.6)$$

donde  $\mathbf{y} = (\mathbf{y}'_1, \dots, \mathbf{y}'_n)'$ ,  $\mathbf{a} = (\mathbf{a}'_1, \dots, \mathbf{a}'_n)'$ ,  $\mathbf{D}(\theta) = [d_{ij}]$  es una matriz rectangular de dimensiones  $(n-1) \times n$  con  $d_{ij} = -\theta$  si  $i = j$ ,  $d_{ij} = 1$  si  $i = j - 1$  y  $d_{ij} = 0$  en cualquier otro caso y  $\mathbf{D} = \mathbf{D}(1)$ , de forma que la matriz de covarianzas de  $\mathbf{w} = (\mathbf{D} \otimes \mathbf{I}_m)\mathbf{y}$  se puede expresar como  $\mathbf{\Omega}_w(\theta) = (\mathbf{\Omega}(\theta) \otimes \mathbf{\Omega}_a)$  con  $\mathbf{\Omega}(\theta) = \mathbf{D}(\theta)\mathbf{D}(\theta)'$ , y su función de densidad

$$f(\mathbf{w}|\theta, \mathbf{\Omega}_a) = (2\pi)^{-(n-1)m/2} |\mathbf{\Omega}(\theta)|^{-m/2} |\mathbf{\Omega}_a|^{-(n-1)/2} \exp \left[ -\frac{1}{2} \text{tr}(\mathbf{\Omega}_a^{-1} \mathbf{W}' \mathbf{\Omega}^{-1}(\theta) \mathbf{W}) \right],$$

donde la matriz  $\mathbf{W} = [\mathbf{w}_2, \dots, \mathbf{w}_n]'$  de dimensiones  $(n-1) \times m$  se puede expresar también como  $\mathbf{W} = \mathbf{D}\mathbf{Y}$ , siendo  $\mathbf{Y} = [\mathbf{y}_1, \dots, \mathbf{y}_n]'$ . En la línea de los resultados de King y Hillier (1985), se puede derivar el estadístico de contraste LBIU a partir de la segunda derivada de la función de verosimilitud de  $\mathbf{w}$ , evitando la utilización de un maximal invariante que en este caso multivariante resulta complejo (Nyblom, 2001; Forchini, 2005). Nótese

que  $\mathbf{w}_t$  es un maximal invariante ante cambios de origen del tipo  $\mathbf{y}_t \mapsto \mathbf{y}_t + \mathbf{a}$ , y que la estructura de  $\mathbf{\Omega}_w(\theta)$  permite concentrar el logaritmo de la función de verosimilitud con respecto a la matriz de covarianzas contemporáneas del vector de perturbaciones, cuyo estimador de máxima verosimilitud es  $\tilde{\mathbf{\Omega}}_a(\theta) = \mathbf{W}'\mathbf{\Omega}^{-1}(\theta)\mathbf{W}/(n-1)$ . Sustituyendo  $\tilde{\mathbf{\Omega}}_a(\theta)$  en  $f(\mathbf{w}|\theta, \mathbf{\Omega}_a)$  y tomando logaritmos se obtiene el logaritmo de la función de verosimilitud concentrada de  $\mathbf{w}$  como

$$\log f_*(\theta|\mathbf{w}, \tilde{\mathbf{\Omega}}_a) \propto -\frac{m}{2} \log |\mathbf{\Omega}(\theta)| - \frac{(n-1)}{2} \log |\tilde{\mathbf{\Omega}}_a(\theta)| \quad (3.7)$$

Transformaciones del tipo  $\mathbf{w}_t \mapsto \mathbf{P}\mathbf{w}_t$  no cambian el valor de esta función, por lo que cualquier estadístico derivado a partir de ella será invariante ante este tipo de transformaciones. El estadístico del contraste LBIU, siguiendo la aproximación de Tanaka (1990), se obtiene a partir de la segunda derivada de (3.7). Teniendo en cuenta las siguientes propiedades:

$$\begin{aligned} \frac{d\mathbf{\Omega}(\theta)^{-1}}{d\theta} &= -\mathbf{\Omega}(\theta)^{-1} \frac{d\mathbf{\Omega}(\theta)}{d\theta} \mathbf{\Omega}(\theta)^{-1}, \\ \frac{d^2\mathbf{\Omega}(\theta)^{-1}}{d\theta^2} &= 2\mathbf{\Omega}(\theta)^{-1} \frac{d\mathbf{\Omega}(\theta)}{d\theta} \mathbf{\Omega}(\theta)^{-1} \frac{d\mathbf{\Omega}(\theta)}{d\theta} \mathbf{\Omega}(\theta)^{-1} - \mathbf{\Omega}(\theta)^{-1} \frac{d^2\mathbf{\Omega}(\theta)}{d\theta^2} \mathbf{\Omega}(\theta)^{-1}, \end{aligned}$$

las dos primeras derivadas de (3.7) con respecto a  $\theta$  son

$$\begin{aligned} \frac{d \log f_*(\theta|\mathbf{w})}{d\theta} &= -\frac{m}{2} \text{tr} \left[ \mathbf{\Omega}(\theta)^{-1} \frac{d\mathbf{\Omega}(\theta)}{d\theta} \right] - \\ &\quad - \frac{(n-1)}{2} \text{tr} \left[ (\mathbf{W}'\mathbf{\Omega}(\theta)^{-1}\mathbf{W})^{-1} \mathbf{W}' \left( \frac{d\mathbf{\Omega}(\theta)^{-1}}{d\theta} \right) \mathbf{W} \right], \\ \frac{d^2 \log f_*(\theta|\mathbf{w})}{d\theta^2} &\propto \frac{(n-1)}{2} \text{tr} \left[ \frac{d(\mathbf{W}'\mathbf{\Omega}(\theta)^{-1}\mathbf{W})^{-1}}{d\theta} \mathbf{W}' \left( \frac{d\mathbf{\Omega}(\theta)^{-1}}{d\theta} \right) \mathbf{W} \right] + \\ &\quad + \frac{(n-1)}{2} \text{tr} \left[ (\mathbf{W}'\mathbf{\Omega}(\theta)^{-1}\mathbf{W})^{-1} \mathbf{W}' \left( \frac{d^2\mathbf{\Omega}(\theta)^{-1}}{d\theta^2} \right) \mathbf{W} \right], \end{aligned}$$

donde, en la segunda derivada, se muestran sólo los elementos correspondientes al segundo término de la primera derivada, ya que los del primero son constantes y  $o(n^2)$ . Dado que la matriz de covarianzas  $\mathbf{\Omega}(\theta)$  se puede expresar como  $\mathbf{\Omega}(\theta) = (1 + \theta^2)\mathbf{I}_{n-1} - \theta(\mathbf{L} + \mathbf{L}')$ , siendo  $\mathbf{L}$  una matriz cuadrada de orden  $n-1$  con unos en la primera subdiagonal inferior y ceros en el resto,  $d\mathbf{\Omega}(\theta)/d\theta|_{\theta=1} = \mathbf{\Omega}(1)$  y  $d^2\mathbf{\Omega}(\theta)/d\theta^2 = 2\mathbf{I}_{n-1}$ . Evaluada bajo  $H_0$ , la primera derivada de la función de log-verosimilitud concentrada es igual a cero, lo que indica que ésta tiene un máximo o un mínimo en  $\theta = 1$ . Por lo tanto, el estadístico LBIU

se obtiene a partir del segundo término de la segunda derivada ya que el primero es igual a  $m$ , y que se puede expresar utilizando las relaciones anteriores como

$$T = \text{tr} [(\mathbf{W}'\boldsymbol{\Omega}(1)^{-1}\mathbf{W})^{-1}\mathbf{W}'\boldsymbol{\Omega}(1)^{-2}\mathbf{W}] / (n-1)m, \quad (3.8)$$

donde el factor de corrección se define como  $(n-1)m$  para que converja la primera traza de la segunda derivada. Nótese que, teniendo en cuenta las relaciones matriciales anteriormente planteadas y que  $\boldsymbol{\Omega}(1) = \mathbf{D}\mathbf{D}'$ , se obtiene que  $\mathbf{D}'(\mathbf{D}\mathbf{D}')^{-1}\mathbf{D} = \mathbf{M}$ , donde  $\mathbf{M} = \mathbf{I}_n - \mathbf{i}_n(\mathbf{i}'_n\mathbf{i}_n)^{-1}\mathbf{i}'_n$ , con  $\mathbf{i}_n$  un vector de unos, y que  $\mathbf{D}'(\mathbf{D}\mathbf{D}')^{-2}\mathbf{D} = \mathbf{M}\mathbf{C}_n\mathbf{C}'_n\mathbf{M}$  con  $\mathbf{C}_n$  una matriz triangular inferior de unos, el estadístico (3.8) se puede expresar alternativamente también como  $T = \text{tr}[(\mathbf{Y}'\mathbf{M}\mathbf{Y})^{-1}\mathbf{Y}'\mathbf{M}\mathbf{C}_n\mathbf{C}'_n\mathbf{M}\mathbf{Y}] / (n-1)m$ , o

$$T = \frac{1}{(n-1)m} \text{tr} \left[ \left( \sum_{t=1}^n \hat{\mathbf{u}}_t \hat{\mathbf{u}}_t' \right)^{-1} \sum_{t=1}^n \left( \sum_{i=1}^t \hat{\mathbf{u}}_i \right) \left( \sum_{i=1}^t \hat{\mathbf{u}}_i \right)' \right], \quad (3.9)$$

donde  $\hat{\mathbf{u}}_t = \mathbf{y}_t - \bar{\mathbf{y}}$  son los residuos mínimo-cuadráticos de  $\mathbf{y}_t$  sobre un vector de unos. De esta forma, se muestra la equivalencia entre la extensión multivariante del estadístico de Tanaka y el del contraste de Nyblom y Harvey. Por último, nótese que residuos  $\hat{\mathbf{u}}_t$  coinciden con los denominados residuos exactos del modelo (3.5) evaluados bajo  $H_0$ , que se obtienen como  $E[\mathbf{a}_t|\mathbf{w}] \equiv \text{cov}(\mathbf{w}, \mathbf{a})[\text{var}(\mathbf{w})]^{-1}\mathbf{w} = [\mathbf{D}'(\mathbf{D}\mathbf{D}')^{-1}\mathbf{D} \otimes \mathbf{I}_m]\mathbf{y}$ , donde la igualdad con los residuos mínimo cuadráticos surge de las relaciones antes descritas.

### 3.3.2. Derivación alternativa del estadístico de Nyblom y Harvey

La expresión del estadístico de contraste de no invertibilidad multivariante (3.8) se ha obtenido a partir del logaritmo de la función de verosimilitud del proceso estacionario, concentrada con respecto a la matriz de covarianzas contemporáneas de los errores, sin necesidad de definir un maximal invariante. En este apartado se demuestra que este procedimiento también se puede utilizar para derivar el estadístico de contraste de Nyblom y Harvey. El modelo (3.1)-(3.2) se puede expresar alternativamente en forma reducida como

$$(\mathbf{D} \otimes \mathbf{I}_m)\mathbf{y} = \mathbf{v} + (\mathbf{D} \otimes \mathbf{I}_m)\mathbf{u},$$

donde  $\mathbf{y} = (\mathbf{y}'_1, \dots, \mathbf{y}'_n)'$ ,  $\mathbf{u} = (\mathbf{u}'_1, \dots, \mathbf{u}'_n)'$ ,  $\mathbf{v} = (\mathbf{v}'_2, \dots, \mathbf{v}'_n)'$  y  $\mathbf{D}$  la matriz de primeras diferencias definida anteriormente. Nótese que  $\mathbf{w} = (\mathbf{D} \otimes \mathbf{I}_m)\mathbf{y}$  es un maximal invariante ante cambios de origen en el vector  $\mathbf{y}_t$ , y que  $\mathbf{w} \sim N(\mathbf{0}, \boldsymbol{\Omega}_w(\rho))$ , donde  $\boldsymbol{\Omega}_w(\rho) = [(\rho\mathbf{I}_n + \mathbf{D}\mathbf{D}') \otimes \boldsymbol{\Sigma}_u]$  y por simplicidad se denota  $\boldsymbol{\Omega}(\rho) = (\rho\mathbf{I}_n + \mathbf{D}\mathbf{D}')$ . Defi-

niendo  $\mathbf{W} = [\mathbf{w}_2, \dots, \mathbf{w}_n]'$ , el logaritmo de su función de verosimilitud tiene la siguiente expresión

$$\log f(\rho, \boldsymbol{\Sigma}_u | \mathbf{w}) \propto -\frac{m}{2} \log |\boldsymbol{\Omega}(\rho)| - \frac{(n-1)}{2} \log |\boldsymbol{\Sigma}_u| - \frac{1}{2} \text{tr} [\boldsymbol{\Sigma}_u^{-1} \mathbf{W}' \boldsymbol{\Omega}(\rho)^{-1} \mathbf{W}].$$

De la primera derivada de esta expresión con respecto a  $\boldsymbol{\Sigma}_u$  se obtiene la expresión del estimador de máxima verosimilitud de la matriz de covarianzas contemporáneas de  $\mathbf{u}_t$  como  $\tilde{\boldsymbol{\Sigma}}_u = \mathbf{W}' \boldsymbol{\Omega}(\rho)^{-1} \mathbf{W} / (n-1)$ , lo que permite concentrar el logaritmo de la función de verosimilitud con respecto a ella de tal forma que

$$\log f_*(\rho | \mathbf{w}) \propto -\frac{m}{2} \log |\boldsymbol{\Omega}(\rho)| - \frac{(n-1)}{2} \log |\mathbf{W}' \boldsymbol{\Omega}(\rho)^{-1} \mathbf{W}|.$$

El estadístico de contraste se obtiene a partir de la primera derivada de esta función con respecto a  $\rho$ , que tiene la siguiente expresión

$$\begin{aligned} \frac{d \log f_*(\rho | \mathbf{w})}{d\rho} &= -\frac{m}{2} \text{tr} \left[ \boldsymbol{\Omega}(\rho)^{-1} \frac{d\boldsymbol{\Omega}(\rho)}{d\rho} \right] - \\ &\quad - \frac{(n-1)}{2} \text{tr} \left[ (\mathbf{W}' \boldsymbol{\Omega}(\rho)^{-1} \mathbf{W})^{-1} \mathbf{W}' \frac{d\boldsymbol{\Omega}(\rho)^{-1}}{d\rho} \mathbf{W} \right]. \end{aligned}$$

Bajo  $H_0$ ,  $\boldsymbol{\Omega}(0) = \mathbf{D}\mathbf{D}'$  y  $d\boldsymbol{\Omega}(\rho)/d\rho|_{\rho=0} = \mathbf{I}_{n-1}$ , por lo que la primera traza en la expresión anterior es igual a  $n(n+1)/2$ . En cuanto al segundo término, nótese que

$$\left. \frac{d\boldsymbol{\Omega}(\rho)^{-1}}{d\rho} \right|_{\rho=0} = -\boldsymbol{\Omega}(0)^{-1} \left. \frac{d\boldsymbol{\Omega}(\rho)}{d\rho} \right|_{\rho=0} \boldsymbol{\Omega}(0)^{-1},$$

con lo que, dadas las expresiones anteriores, el segundo término se puede expresar como  $(n-1) \text{tr} [(\mathbf{W}'(\mathbf{D}\mathbf{D}')^{-1}\mathbf{W})^{-1} \mathbf{W}'(\mathbf{D}\mathbf{D}')^{-2} \mathbf{W}] / 2$ . Finalmente, dado que  $\mathbf{W} = \mathbf{D}\mathbf{Y}$ , y teniendo en cuenta las relaciones definidas en Mazas (2012),  $\mathbf{D}'(\mathbf{D}\mathbf{D}')^{-1}\mathbf{D} = \mathbf{M}$  y  $\mathbf{D}'(\mathbf{D}\mathbf{D}')^{-2}\mathbf{D} = \mathbf{M}\mathbf{C}_n\mathbf{C}'_n\mathbf{M}$ , normalizando la primera derivada por un factor dependiente en  $n^2$ , como por ejemplo  $mn(n-1)$ , el estadístico se puede expresar como

$$NH = \text{tr}[(\mathbf{Y}'\mathbf{M}\mathbf{Y})^{-1} \mathbf{Y}'\mathbf{M}\mathbf{C}_n\mathbf{C}'_n\mathbf{M}\mathbf{Y}] / nm,$$

que es el estadístico Nyblom y Harvey (2000). Conviene volver a resaltar que, aunque la derivación del estadístico se ha realizado a partir de un maximal invariante ante sólo transformaciones de origen, éste es también invariante ante premultiplicaciones de  $\mathbf{y}_t$

por matrices definidas positivas, lo que se consigue al concentrar la función de densidad con respecto a  $\Sigma_u$ . Este procedimiento simplifica la derivación del estadístico y será el utilizado en la derivación de los estadísticos de contraste para modelos más complejos en el capítulo siguiente.

### 3.3.3. Distribución en muestras finitas del estadístico

Aunque la expresión (3.9) es útil para el cálculo valor del estadístico, su distribución en muestras finitas se puede obtener más fácilmente a partir de la expresión (3.8). Nótese que las columnas de la matriz  $\mathbf{W} = [\mathbf{w}_j]$  ( $j = 1, \dots, m$ ) se corresponden con las observaciones de cada una de las  $m$  series temporales estacionarias. Bajo la hipótesis alternativa,  $\theta < 1$ , y suponiendo  $\Omega_a = \mathbf{I}_m$  sin pérdida de generalidad ya que el contraste es invariante,  $\mathbf{w}_j \sim N(\mathbf{0}, \Omega(\theta))$ . De la descomposición espectral de la matriz de covarianzas,  $\Omega(\theta) = \mathbf{P}\Lambda(\theta)\mathbf{P}'$  donde  $\Lambda(\theta) = \text{diag}\{\lambda_1(\theta), \dots, \lambda_{n-1}(\theta)\}$  es una matriz diagonal que contiene los autovalores de la matriz tridiagonal  $\mathbf{D}(\theta)\mathbf{D}(\theta)'$ ,  $\lambda_t(\theta) = (1 - \theta)^2 + 4\theta\sin^2(t\pi/2n)$ , (Anderson, 1971), y  $\mathbf{P} = [p_{ij}]$  es la matriz de autovectores correspondientes,  $p_{ij} = \sqrt{2/n}\sin(2\pi ij/n)$  ( $i, j = 1, \dots, n-1$ ), de forma que  $\mathbf{e}_j \equiv \Lambda(\theta)^{-1/2}\mathbf{P}^{-1}\mathbf{w}_j \sim N(\mathbf{0}, \mathbf{I}_{n-1})$ . Teniendo en cuenta que  $\Omega(1)^{-1} = \mathbf{P}\Lambda(1)^{-1}\mathbf{P}'$ ,  $\Omega(1)^{-2} = \mathbf{P}\Lambda(1)^{-2}\mathbf{P}'$  y  $\mathbf{P}\mathbf{P}' = \mathbf{I}_{n-1}$ , el estadístico bajo la hipótesis alternativa se puede expresar como

$$T(\theta) = \text{tr} \left[ (\mathbf{E}'\Lambda(\theta)^{1/2}\Lambda(1)^{-1}\Lambda(\theta)^{1/2}\mathbf{E})^{-1}\mathbf{E}'\Lambda(\theta)^{1/2}\Lambda(1)^{-2}\Lambda(\theta)^{1/2}\mathbf{E} \right] / (n-1)m,$$

con  $\mathbf{E} = [\mathbf{e}_1, \dots, \mathbf{e}_m]$ , o bien

$$T(\theta) = \frac{1}{(n-1)m} \text{tr} \left[ \left( \sum_{t=1}^{n-1} \frac{\lambda_t(\theta)}{\lambda_t} \boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_t' \right)^{-1} \left( \sum_{t=1}^{n-1} \frac{\lambda_t(\theta)}{\lambda_t^2} \boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_t' \right) \right], \quad (3.10)$$

donde  $\lambda_t = \lambda_t(1) = 4\sin^2(t\pi/2n)$  y  $\boldsymbol{\epsilon}_t$  es la  $t$ -ésima fila de  $\mathbf{E}$ . Nótese que para  $m = 1$ ,  $T(\theta)$  se reduce a la ratio de dos formas cuadráticas en variables normales, cuya distribución en muestras finitas puede ser tabulada a partir de los procedimientos de Imhof (1961) o Davies (1973). Sin embargo, en el caso multivariante ( $m > 1$ ) el uso de estos algoritmos no es posible, por lo que la distribución debe ser tabulada por simulación a partir de

$$T = \frac{1}{(n-1)m} \text{tr} \left[ \left( \sum_{t=1}^{n-1} \boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_t' \right)^{-1} \left( \sum_{t=1}^{n-1} \lambda_t^{-1} \boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_t' \right) \right]. \quad (3.11)$$

Esta expresión del estadístico es útil para obtener los dos primeros momentos de la distribución del estadístico bajo  $H_0$ . Teniendo en cuenta que  $T = \sum_{t=1}^{n-1} \lambda_t^{-1} B_t$  donde  $B_t = \epsilon'_t \left( \sum_{t=1}^{n-1} \epsilon_t \epsilon'_t \right)^{-1} \epsilon_t \sim \text{Beta}(m/2, (n-1-m)/2)$  (ver Rao, 2002, p.53), Nyblom (2001) deriva los dos primeros momentos de esta combinación lineal de variables Beta como

$$E(T) = m\bar{\lambda}_1, \quad V(T) = c(\bar{\lambda}_2 - \bar{\lambda}_1^2), \quad \bar{\lambda}_r = \frac{1}{(n-1)} \sum_{t=1}^n \lambda_t^{-r},$$

con  $c = 2m(n-1-m)/[(n-2)(n+1)]$ . Dada la expresión conocida de los autovalores, se obtiene para este caso que  $\bar{\lambda}_1 = (n+1)/6$  y  $\bar{\lambda}_2 = \bar{\lambda}_1(2n^2+7)/30$  con lo que los dos primeros momentos de  $T/(n-1)m$  son

$$E(T) = \frac{n+1}{6(n-1)}, \quad V(T) = \frac{(n-1-m)(2n-1)}{90m(n-1)^2}. \quad (3.12)$$

En el caso univariante, los momentos asintóticos del estadístico se pueden calcular fácilmente como  $E(T) \rightarrow 1/6$  y  $V(T) \rightarrow 1/45$ , que coinciden con los del estadístico de Nyblom y Mäkeläinen (1983).

En el caso especial de que la matriz de covarianzas contemporáneas  $\Omega_a$  fuera conocida, el modelo VIMA(1,1) podría expresarse como un proceso univariante estacional con periodo  $m$ . Sin pérdida de generalidad se puede considerar  $\Omega_a^* = \sigma_a^2 \Omega_a$  con  $\sigma_a^2$  una constante arbitraria y, definiendo los procesos normalizados  $\mathbf{w}_t^* = \Omega_a^{-1/2} \mathbf{w}_t$  y  $\mathbf{a}_t^* = \Omega_a^{-1/2} \mathbf{a}_t$ , el proceso (3.5) normalizado puede escribirse como

$$w_{it}^* = (1 - \theta B^m) a_{it}^*, \quad a_{it}^* \sim N(0, \sigma_a^2), \quad i = 1, \dots, m.$$

Siguiendo a Tam y Reinsel (1997), y denotando por  $\mathbf{w}_* = (\mathbf{I}_{n-1} \otimes \Omega^{-1/2}) \mathbf{w}$ , el estadístico del contraste LBIU para la hipótesis nula de no invertibilidad  $H_0 : \theta = 1$  frente a  $H_1 : \theta \neq 1$  en este modelo univariante estacional tiene la siguiente expresión:

$$TR = \frac{1}{(n-1)} \frac{\mathbf{w}'_* [(\mathbf{D}\mathbf{D}')^{-2} \otimes \mathbf{I}_m] \mathbf{w}_*}{\mathbf{w}'_* [(\mathbf{D}\mathbf{D}')^{-1} \otimes \mathbf{I}_m] \mathbf{w}_*}.$$

Así expresado, el estadístico es una ratio de formas cuadráticas en variables normales cuya distribución puede evaluarse por el procedimiento de Imhof (1961) utilizando los inversos de los autovalores de la matriz tridiagonal  $\mathbf{D}\mathbf{D}'$  cada uno de ellos con multiplicidad  $m$ . La tabla 3.1 muestra los percentiles del estadístico de contraste de no invertibilidad calculados de ambas formas, por simulación y utilizando el procedimiento de Imhof.

Ambos métodos proporcionan percentiles similares, si bien en muestras finitas existen distorsiones no superiores a una centésima en la mayoría de los casos.

Si la matriz de covarianzas no fuera conocida, la normalización utilizada para derivar  $TR$  se llevaría a cabo usando una estimación consistente de ella, como por ejemplo la de mínimos cuadrados multivariantes,  $\hat{\mathbf{\Omega}}_a = (n-1)^{-1} \mathbf{Y}' \mathbf{M} \mathbf{Y} = (n-1)^{-1} \mathbf{W}' (\mathbf{D} \mathbf{D}')^{-1} \mathbf{W}$ . Sustituyéndola en la expresión del estadístico, éste se puede expresar como la siguiente ratio de trazas,

$$TR = \frac{1}{(n-1)} \frac{\text{tr}[(\mathbf{W}' (\mathbf{D} \mathbf{D}')^{-1} \mathbf{W})^{-1} \mathbf{W}' (\mathbf{D} \mathbf{D}')^{-2} \mathbf{W}]}{\text{tr}[(\mathbf{W}' (\mathbf{D} \mathbf{D}')^{-1} \mathbf{W})^{-1} \mathbf{W}' (\mathbf{D} \mathbf{D}')^{-1} \mathbf{W}]},$$

donde la traza del denominador es igual a  $m$ , con lo que el estadístico  $TR$  factible es igual a (3.8).

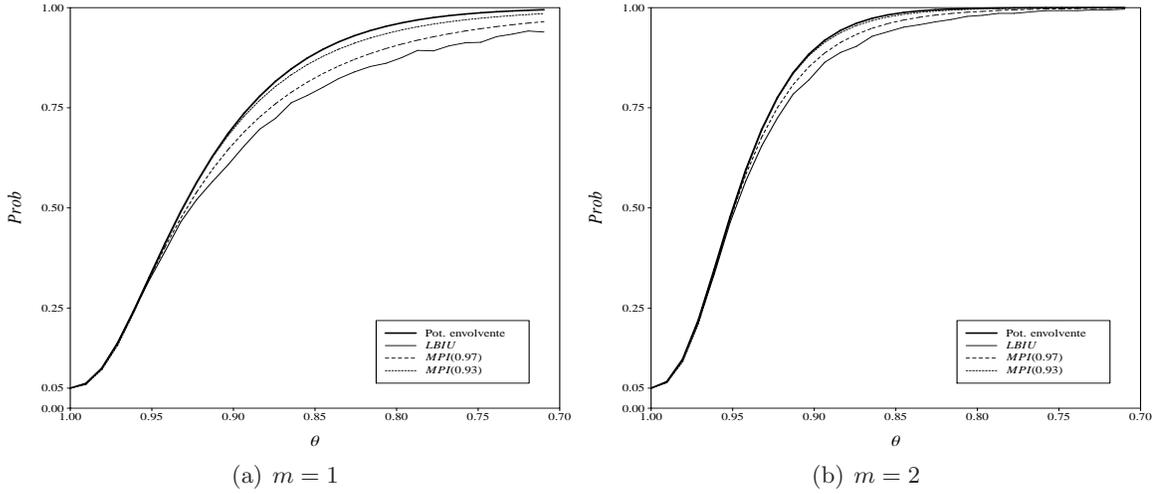
### Función potencia envolvente

El contraste más potente e invariante (MPI) ante transformaciones del tipo  $\mathbf{y}_t \mapsto \mathbf{a} + \mathbf{P} \mathbf{y}_t$  para contrastar  $H_0 : \theta = 1$  frente a una determinada alternativa  $H_1 : \theta = \theta_1$  se puede obtener, aplicando el teorema de Neymann-Pearson, como el cociente de las funciones de verosimilitud en ambos casos. A partir de la expresión del logaritmo de la función de verosimilitud de (3.6), y teniendo en cuenta que la propiedad de invarianza permite considerar  $\mathbf{\Omega}_a = \mathbf{I}_m$ , el estadístico más potente e invariante (MPI) para este problema tiene la siguiente expresión

$$MPI(\theta_1) = \text{tr} \left( \sum_{t=1}^{n-1} \frac{\lambda_t(\theta)}{\lambda_t(\theta_1)} \boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_t' \right) / \text{tr} \left( \sum_{t=1}^{n-1} \frac{\lambda_t(\theta)}{\lambda_t} \boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_t' \right),$$

rechazándose  $H_0$  para valores pequeños del estadístico. El estadístico  $MPI(\theta_1)$  se puede expresar como un cociente de dos formas cuadráticas del tipo  $\mathbf{e}' \mathbf{L}_1 \mathbf{e} / \mathbf{e}' \mathbf{L}_2 \mathbf{e}$  donde  $\mathbf{e}$  es un vector de dimensión  $(n-1)m$  de variables aleatorias normales estándar y  $\mathbf{L}_1$  y  $\mathbf{L}_2$  son dos matrices diagonales que contienen los cocientes de autovalores que figuran en  $MPI(\theta_1)$  cada uno de ellos con multiplicidad  $m$ , pudiendo ser tabulada su distribución de forma exacta utilizando el procedimiento de Imhof (1961). La Figura 3.1 muestra la función potencia envolvente del contraste, obtenida como la probabilidad de rechazo de  $H_0 : \theta = 1$  frente a  $H_1 : \theta_1 = \theta$  cuando el verdadero valor del parámetro media móvil coincide con la alternativa considerada  $\theta$ , y que indica la máxima potencia que puede alcanzar cualquier estadístico con las características del estadístico localmente óptimo  $T$ . También se muestra la función potencia del contraste para  $m = 1$  y  $m = 2$ , así como

Figura 3.1: Función potencia y envolvente



la función potencia de MPI(0,97) y MPI(0,93).

### 3.3.4. Distribución asintótica del estadístico

A partir de las expresiones anteriores se puede derivar la distribución límite del estadístico de contraste. Por simplicidad en la exposición, y sin pérdida de generalidad se considera la corrección  $nm$  para el estadístico  $T$ , en lugar de  $(n-1)m$  con la que está definida. A continuación se deriva la distribución del estadístico bajo una secuencia de alternativas del tipo  $\theta = 1 - c/n$  ( $c > 0$ ). Partiendo de (3.10), nótese que  $n^{-1} \sum_{t=1}^{n-1} \epsilon_t \epsilon_t' \xrightarrow{p} \mathbf{I}_m$  y  $\lim_{n \rightarrow \infty} n^{-2} \lambda_t^{-1} = 1/(\pi k)^2$ . Además,  $\lambda_t(\theta)/\lambda_t = n^{-2} c^2 \lambda_t^{-1} + (1 - n^{-1}c)$  y  $\lambda_t(\theta)/\lambda_t^2 = n^{-2} c^2 \lambda_t^{-2} + (1 - n^{-1}c) \lambda_t^{-1}$ . Por lo tanto, la matriz dentro de la inversa,  $n^{-1} \sum_{t=1}^{n-1} [n^{-2} c^2 \lambda_t^{-1} + (1 - n^{-1}c)] \epsilon_t \epsilon_t' \xrightarrow{p} \mathbf{I}_m$ , con lo que, expresando la segunda matriz como  $n^{-2} \sum_{t=1}^{n-1} [n^{-2} c^2 \lambda_t^{-2} + (1 - n^{-1}c) \lambda_t^{-1}] \epsilon_t \epsilon_t'$ , la distribución límite del estadístico bajo una secuencia de alternativas  $\theta = 1 - c/n$  es

$$T(\theta) \xrightarrow{d} \frac{1}{m} \sum_{k=1}^{\infty} \left[ \frac{1}{(\pi k)^2} + \frac{c^2}{(\pi k)^4} \right] tr(\epsilon_t \epsilon_t') \quad tr(\epsilon_t \epsilon_t') \equiv \zeta_k \sim \chi_m^2 \quad (3.13)$$

Bajo  $H_0$  ( $c = 0$ ), la distribución límite de  $T$  es la denominada Cramèr-von Mises con  $m$  grados de libertad,  $CvM(m)$ , ampliamente descrita por Anderson y Darling (1952) para el caso univariante  $m = 1$  y tabulada por Hansen (1990) para  $m > 1$ . Una expresión alternativa para la distribución límite del estadístico se puede obtener a partir de (3.9).

Bajo una secuencia de alternativas  $\theta = 1 - c/n$  se expresa el modelo como  $\nabla \mathbf{y}_i = \nabla \mathbf{a}_i + n^{-1}c\mathbf{a}_{i-1}$ , con lo que, suponiendo  $\mathbf{y}_0 = \mathbf{0}$ ,  $\mathbf{y}_i = \mathbf{a}_i + n^{-1}c \sum_{j=1}^{i-1} \mathbf{a}_j$  y los residuos MCO,  $\hat{\mathbf{u}}_t = \mathbf{y}_t - \bar{\mathbf{y}}$ , acumulados tienen la siguiente expresión

$$n^{-1/2} \sum_{i=1}^t \hat{\mathbf{u}}_i = n^{-1/2} \sum_{i=1}^t \mathbf{a}_i + \frac{c}{n^{3/2}} \sum_{i=1}^t \boldsymbol{\varepsilon}_i - \frac{t}{n^{3/2}} \sum_{t=1}^n \mathbf{a}_t - \frac{c}{n^{5/2}} \sum_{t=1}^n t\boldsymbol{\varepsilon}_t$$

donde  $\boldsymbol{\varepsilon}_t = \sum_{j=1}^{t-1} \mathbf{a}_j$  es un paseo aleatorio. De Phillips (1986), Phillips (1987), Park y Phillips (1988), Park y Phillips (1989) y Phillips y Durlauf (1986), se sabe que  $n^{-1/2} \sum_{i=1}^t \mathbf{a}_i \xrightarrow{d} \mathbf{W}(t)$ ;  $n^{-3/2} \sum_{i=1}^t \boldsymbol{\varepsilon}_i \xrightarrow{d} \int_0^t \mathbf{W}(r)dr$ ,  $n^{-3/2} \sum_{t=1}^n t\mathbf{a}_t \xrightarrow{d} r\mathbf{W}(1)$  y  $n^{-5/2} \sum_{t=1}^n t\boldsymbol{\varepsilon}_t \xrightarrow{d} \int_0^1 r\mathbf{W}(r)dr$ , donde  $\mathbf{W}(r)$  es un proceso browniano  $m$ -dimensional estándar. De igual manera que en el caso anterior,  $n^{-1} \sum_{t=1}^n \hat{\mathbf{u}}_t \hat{\mathbf{u}}_t' \xrightarrow{p} \mathbf{I}_m$ , de forma que

$$T(\theta) \xrightarrow{d} \frac{1}{m} \text{tr} \left( \int_0^1 \mathbf{B}(r)\mathbf{B}(r)' dr \right) = \frac{1}{m} \sum_{j=1}^m \int_0^1 B_j^2(r) dr$$

$$\mathbf{B}(r) = \mathbf{W}(r) - r\mathbf{W}(1) + c \int_0^t (\mathbf{W}(u) - u\mathbf{W}(1)) du$$

y  $B_j(r)$  es el componente genérico de  $\mathbf{B}(r)$ .

En cuanto a la distribución asintótica de  $MPI(\theta)$ , con los resultados anteriores y  $\lambda_k(\theta) = c^2 n^{-2} \lambda_k + (1 - n^{-1}c)$ ,

$$n^{-1} \sum_{k=1}^n \frac{\lambda_t(\theta)}{\lambda_t} \zeta_k \xrightarrow{p} m \quad \text{y} \quad n^{-1} \sum_{k=1}^n \frac{\lambda_t(\theta)}{\lambda_t(\theta_1)} \zeta_t = n^{-1} \sum_{k=1}^n \frac{(\pi k)^{-2} c^2 + 1}{(\pi k)^{-2} c_1^2 + 1} \zeta_t \xrightarrow{p} m$$

con lo que  $MPI(\theta) \xrightarrow{p} 1$ , ya que  $n^{-1}(\pi k)^{-2} = o_p(1)$ . Dado que  $MPI(\theta)$  tiene una distribución degenerada y que la tasa a la que converge a su valor asintótico 1 es  $O_p(n)$ , conviene trabajar, como en Saikkonen y Luukkonen (1993b) y Tam y Reinsel (1997) en el caso univariante, con el estadístico  $n[1 - MPI(\theta)]$ , que se puede expresar como

$$n[1 - MPI(\theta_1)] = n \frac{\sum_{t=1}^n \lambda_t^{*-1}(\theta_1) \lambda_t^*(\theta) [\lambda_t^*(\theta_1) - 1] \zeta_t}{\sum_{t=1}^n \lambda_t^*(\theta) \zeta_t}$$

donde  $\lambda_t^*(\theta) = \lambda_t^{-1} \lambda_t^*(\theta) = c^2 n^{-2} \lambda_t^{-1} + (1 - cn^{-1})$ . Al igual que en el caso anterior, el denominador de esta expresión corregido por  $n^{-1}$  converge a  $m$  y el numerador se puede

expresar como

$$\sum_{t=1}^n \left[ c^2 n^{-2} \lambda_t^{-1} - n^{-1} c - \frac{[(c^2 - c_1^2)n^{-2} \lambda_t^{-1} - (c - c_1)n^{-1}]}{(c_1^2 n^{-2} \lambda_t^{-1} + 1 - c_1 n^{-1})} \right] \zeta_t$$

de donde, utilizando los resultados anteriores y que  $(c - c_1)n^{-1} / (c_1^2 n^{-2} \lambda_t^{-1} + 1 - c_1 n^{-1}) \xrightarrow{p} (c - c_1)n^{-1}$ ,

$$n[1 - MPI(\theta_1)] \xrightarrow{d} c^2 \sum_{k=1}^{\infty} \frac{1}{(\pi k)} \zeta_k + (c_1^2 - c^2) \sum_{k=1}^{\infty} \frac{1}{c_1^2 + (\pi k)^2} \zeta_k - c_1$$

Esta expresión resulta útil para derivar el óptimo puntual y la función potencia envolvente asintótica.

### 3.4. El modelo de Nyblom y Harvey corregido

#### 3.4.1. Formulación del y derivación del estadístico

Mazas (2012) propone una corrección paramétrica para el estadístico de contraste de no invertibilidad en un modelo ARIMA(p,d,q) univariante. Esta corrección se puede generalizar al caso multivariante para contrastar no invertibilidad en un modelo VARIMA. Como en apartados anteriores, la derivación del estadístico de contraste es más sencilla utilizando la representación estructural, por lo que el contraste se deriva a partir del modelo de nivel local generalizado de Mazas (2012), cuya extensión multivariante es

$$\mathbf{y}_t = \boldsymbol{\mu}_t + \mathbf{e}_t, \quad \Phi(B)\mathbf{e}_t = \Theta(B)\mathbf{u}_t, \quad (3.14)$$

$$\boldsymbol{\mu}_t = \boldsymbol{\mu}_{t-1} + \mathbf{r}_t, \quad \Phi(B)\mathbf{r}_t = \Theta(B)\mathbf{v}_t, \quad (3.15)$$

donde, al igual que en otros casos,  $\mathbf{u}_t \sim N(\mathbf{0}, \boldsymbol{\Sigma}_u)$ ,  $\mathbf{v}_t \sim N(\mathbf{0}, \rho \boldsymbol{\Sigma}_u)$  y  $\Phi(B)$  y  $\Theta(B)$  son dos polinomios matriciales de órdenes  $p$  y  $q$  respectivamente sin raíces comunes. Aunque el supuesto de que la correlación de ambos errores sea la misma puede parecer fuerte, dada la independencia entre los errores  $\mathbf{u}_t$  y  $\mathbf{v}_t$ , la forma reducida de (3.14)-(3.15) es un modelo VARMA(p,1,q+1) tal que

$$\Phi(B)\nabla \mathbf{y}_t = (\mathbf{I}_m - \theta B)\Theta(B)\mathbf{a}_t, \quad \mathbf{a}_t \sim N(\mathbf{0}, \boldsymbol{\Omega}_a). \quad (3.16)$$

El estadístico del contraste LBI para la hipótesis nula de nivel local determinista  $H_0 : \rho = 0$  frente a  $H_1 : \rho > 0$  se obtiene a partir de la primera derivada del logaritmo de la función de verosimilitud de  $\mathbf{y}_t$  expresado en forma estructural. En este caso  $\mathbf{e} = (\mathbf{e}'_1, \dots, \mathbf{e}'_n)' \sim N(\mathbf{0}, \mathbf{\Omega}_{(n)})$ , siendo  $\mathbf{\Omega}_{(n)}$  la matriz  $mn \times mn$  de covarianzas del proceso VARMA(p,q) que sigue  $\mathbf{e}_t$ . Por su parte  $\mathbf{r}_t = (\mathbf{r}'_2, \dots, \mathbf{r}'_n)' \sim N(\mathbf{0}, \rho \mathbf{\Omega}_{(n-1)})$  donde  $\mathbf{\Omega}_{(n-1)}$  tiene la misma estructura que  $\mathbf{\Omega}_{(n)}$  pero de dimensiones  $(n-1)m \times (n-1)m$ . De esta forma, la matriz de covarianzas de la forma reducida  $\mathbf{w}_t = \nabla \mathbf{y}_t$  se expresa como  $\mathbf{\Omega}_w(\rho) = \rho \mathbf{\Omega}_{(n-1)} + \mathbf{D}_m \mathbf{\Omega}_{(n)} \mathbf{D}'_m$ , donde  $\mathbf{D}_m = (\mathbf{D} \otimes \mathbf{I}_m)$  y  $\mathbf{D}$  es la matriz  $(n-1) \times n$  de diferencias definida en apartados anteriores. De esta forma, el logaritmo de la función de verosimilitud de  $\mathbf{w}_t$  se define como

$$\log f(\mathbf{x}, \rho | \mathbf{w}) \propto -\frac{1}{2} \log |\mathbf{\Omega}_w(\rho)| - \frac{1}{2} \mathbf{w}' \mathbf{\Omega}_w^{-1}(\rho) \mathbf{w},$$

donde el vector  $\mathbf{x}$  contiene todos los parámetros del modelo distintos de  $\rho$ . Teniendo en cuenta que  $\mathbf{\Omega}_w(0) = \mathbf{D}_m \mathbf{\Omega}_{(n)} \mathbf{D}'_m$  y  $d\mathbf{\Omega}_w^{-1}(\rho)/d\rho|_{\rho=0} = -\mathbf{\Omega}_w^{-1}(0) d\mathbf{\Omega}_w(\rho)/d\rho|_{\rho=0} \mathbf{\Omega}_w^{-1}(0)$ , la primera derivada de esta función con respecto a  $\rho$  es

$$\left. \frac{d \log f(\mathbf{x}, \rho | \mathbf{w})}{d\rho} \right|_{\rho=0} = -\frac{1}{2} \text{tr} \left[ \mathbf{\Omega}_w^{-1}(0) \left. \frac{d\mathbf{\Omega}_w^{-1}(\rho)}{d\rho} \right|_{\rho=0} \right] + \frac{1}{2} \mathbf{w}' \mathbf{\Omega}_w^{-1}(0) \mathbf{\Omega}_{(n-1)} \mathbf{\Omega}_w^{-1}(0) \mathbf{w}$$

Dado que el primer término de la derivada es constante, el estadístico de contraste surge del segundo. Esta forma cuadrática puede expresarse en términos de los residuos exactos del modelo. Para comprobarlo, se define la matriz  $\mathbf{D}_* = \mathbf{D}_m \mathbf{\Omega}_{(n)}^{1/2}$ , donde  $\mathbf{\Omega}_{(n)}^{1/2}$  es el factor de Cholesky de la matriz de covarianzas de  $\mathbf{e}$ , y teniendo en cuenta que  $\mathbf{\Omega}_w(0) = (\mathbf{D}_* \mathbf{D}'_*)$  es la matriz de covarianzas de  $\nabla \mathbf{y}_t$  en (3.16) bajo la hipótesis nula, los residuos exactos de este modelo se pueden expresar como

$$\hat{\mathbf{a}} = \text{cov}(\mathbf{a}, \mathbf{w}) [\text{var}(\mathbf{w})]^{-1} \mathbf{w} = \mathbf{D}'_* (\mathbf{D}_* \mathbf{D}'_*)^{-1} \mathbf{w}.$$

Se puede comprobar que  $\mathbf{D}'_* (\mathbf{D}_* \mathbf{D}'_*)^{-1} \mathbf{D}_* = \mathbf{M}_*$ , es la matriz generadora de residuos en la regresión de mínimos cuadrados generalizados de  $\mathbf{y}_t$  sobre una constante y con correlación tipo VARMA(p,q). Por lo tanto,  $\hat{\mathbf{a}} = \mathbf{M}_* \mathbf{\Omega}_{(n)}^{-1/2} \mathbf{y}$ , esto es, los residuos exactos del modelo VARIMA (3.16) evaluados bajo la hipótesis nula coinciden con los residuos de mínimos cuadrados generalizados del vector  $nm \times 1$ ,  $\mathbf{y}$  sobre un término constante, cuando la correlación viene definida por el modelo VARMA(p,q) del término de error  $\mathbf{e}_t$ . Con todo esto, el estadístico de contraste puede expresarse, corregido por el tamaño del

vector  $\mathbf{w}$ , alternativamente de las dos siguientes maneras:

$$\begin{aligned} T_1^* &= \frac{1}{(n-1)^2 m} \hat{\mathbf{a}}' \mathbf{D}'_* \boldsymbol{\Omega}_w^{-1}(0) \boldsymbol{\Omega}_{(n-1)} \boldsymbol{\Omega}_w^{-1}(0) \mathbf{D}_* \hat{\mathbf{a}}, \\ &= \frac{1}{(n-1)^2 m} \sum_{i=1}^m \sum_{j=1}^m \sum_{r=1}^{n-1} \sum_{l=1}^{n-1} \gamma_{i,j}(r-s) e_{ir} e_{js}, \end{aligned} \quad (3.17)$$

donde  $e_{ir}$  es el elemento  $(r-1)m+i$  del vector  $\mathbf{e} = \boldsymbol{\Omega}_w^{-1}(0) \mathbf{w}$  y  $\gamma_{i,j}(r-s)$  es el elemento  $i, j$  de la matriz de covarianzas de  $\mathbf{u}_t$  en el retardo  $r-s$ ,  $\boldsymbol{\Gamma}(r-s)$  contenidas en  $\boldsymbol{\Omega}_{(n-1)}$ . La primera expresión es útil para derivar la distribución del estadístico, mientras que la segunda refleja el procedimiento computacional para el cálculo del estadístico. Una expresión alternativa del estadístico en términos de los residuos de mínimos cuadrados generalizados se obtiene notando que

$$\boldsymbol{\Omega}_{(n-1)} = \mathbf{D}_* \boldsymbol{\Omega}_{(n)}^{-1/2} \mathbf{C}_n \boldsymbol{\Omega}_{(n)} \mathbf{C}_n' \boldsymbol{\Omega}_{(n)}^{-1/2'} \mathbf{D}_*'$$

con  $\mathbf{C}_n$  una matriz triangular inferior de unos. Insertando esta relación en la primera expresión de  $T_1^*$  en (3.17), recordando que  $\hat{\mathbf{a}} = \hat{\mathbf{u}}$  donde  $\hat{\mathbf{u}}$  es el vector de residuos de mínimos cuadrados generalizados,  $\hat{\mathbf{u}} = \mathbf{M}_* \boldsymbol{\Omega}_{(n)}^{-1/2} \mathbf{y}$ , y que  $\mathbf{M}_*$  es una matriz simétrica e idempotente, una expresión alternativa para  $T_1^*$  es

$$T_1^* = \frac{1}{(n-1)^2 m} \hat{\mathbf{u}}' \boldsymbol{\Omega}_{(n)}^{-1/2} \mathbf{C}_n \boldsymbol{\Omega}_{(n)} \mathbf{C}_n' \boldsymbol{\Omega}_{(n)}^{-1/2'} \hat{\mathbf{u}}, \quad (3.18)$$

que puede ser más conveniente para el cálculo del estadístico. Nótese por último que, aunque el estadístico de contraste se ha construido para ser invariante ante cambios de origen del tipo  $\mathbf{y}_t \mapsto \mathbf{y}_t + \mathbf{a}$  ( $\mathbf{a} \neq \mathbf{0}$ ), donde  $\mathbf{D}_m \mathbf{y}$  es un maximal invariante, también lo es ante cambios de escala tales como  $\mathbf{y}_t \mapsto \mathbf{P} \mathbf{y}_t$  al estar calculado sobre los residuos exactos que están normalizados.

Por último, si las estimaciones de los parámetros son consistentes, entonces  $\hat{\mathbf{a}}$  converge a un vector  $nm \times 1$  de variables aleatorias normales estándar con lo que, a partir de la primera expresión de (3.17), se puede expresar la distribución en muestras finitas de  $T_1^*$  como

$$T_1^* = \frac{1}{(n-1)^2 m} \sum_{t=1}^{(n-1)m} \lambda_{1t}^* e_t^2 \quad \lambda_{1t}^* = \text{eig}[\boldsymbol{\Omega}_{(n-1)} (\mathbf{D}_* \mathbf{D}_*')^{-1}]$$

donde  $\lambda_{1t}^*$  son los autovalores no nulos de  $\mathbf{D}_* \boldsymbol{\Omega}_w^{-1}(0) \boldsymbol{\Omega}_{(n-1)} \boldsymbol{\Omega}_w^{-1}(0) \mathbf{D}_*$  que, dado que  $\boldsymbol{\Omega}_w(0) = \mathbf{D}_* \mathbf{D}_*'$ , son iguales a los autovalores de  $\boldsymbol{\Omega}_{(n-1)} \boldsymbol{\Omega}_w^{-1}(0)$ . Esta matriz puede expre-

sarse alternativamente como  $[\mathbf{\Omega}_{(n-1)}^{-1/2} \mathbf{D}_m \mathbf{\Omega}_{(n-1)}^{1/2} \mathbf{\Omega}_{(n-1)}^{1/2'} \mathbf{D}_m' \mathbf{\Omega}_{(n-1)}^{-1/2'}]^{-1}$ , donde las matrices conmutan asintóticamente, de forma que para muestras grandes, los autovalores de esta matriz son equivalentes a los de la matriz  $(\mathbf{D}\mathbf{D}')^{-1}$  con multiplicidad  $m$ , que coinciden con los que se utilizaron para tabular la distribución del estadístico en el modelo sin correlación. Por lo tanto, para muestras moderadas o grandes, no habría pérdida sensible de potencia si se calcula el estadístico como en (3.17) y utilizan los valores críticos habituales para realizar el contraste.

De igual forma, se puede considerar que para muestras moderadas o grandes, las matrices centrales de (3.18) conmutan de forma que la forma cuadrática puede aproximarse como  $\hat{\mathbf{u}}' \mathbf{C}_n \mathbf{C}_n' \hat{\mathbf{u}}$ . Sin pérdida de generalidad, dado que la distribución para muestras moderadas o grandes de  $T_1^*$  es igual a la de (3.9) se puede utilizar este estadístico para contrastar la hipótesis nula de no invertibilidad en un modelo VARMA(p,1,q+1) utilizando los residuos exactos del modelo evaluados bajo  $H_0$  como

$$T_2^* = \text{tr} \left[ (\tilde{\mathbf{A}}' \tilde{\mathbf{A}})^{-1} \tilde{\mathbf{A}}' \mathbf{C}_n \mathbf{C}_n' \tilde{\mathbf{A}} \right] / (n-1)m,$$

donde la  $t$ -ésima fila de la matriz  $\tilde{\mathbf{A}}$  es  $\hat{\mathbf{a}}_t - \bar{\mathbf{a}}$ . Si las estimaciones de los parámetros del modelo VARMA(p,1,q+1) son consistentes de forma que el vector de residuos  $\hat{\mathbf{a}}_t$  converge a  $\boldsymbol{\epsilon}_t$  con  $\boldsymbol{\epsilon}_t \sim N(\mathbf{0}, \mathbf{I}_m)$ , entonces, por los resultados anteriores,  $T_2^* \xrightarrow{d} CvM(m)/m$ . La potencia de esta aproximación se evalúa en los siguientes experimentos Monte Carlo.

### 3.4.2. Estudio Monte Carlo

Para estudiar la potencia del contraste y de la corrección paramétrica se proponen dos estudios Monte Carlo basados en un modelo bivalente, donde el estadístico de contraste se ha calculado a la manera de  $T_2^*$ . Para el primero se utiliza el siguiente proceso generador de datos

$$\nabla \mathbf{y}_t = (\mathbf{I}_2 - \theta \mathbf{I}_2 B) \mathbf{a}_t, \quad \mathbf{a}_t \sim N(\mathbf{0}, \mathbf{\Omega}_a),$$

donde  $\mathbf{\Omega}_a = [\sigma_{ij}]$  con  $\sigma_{11} = \sigma_{22} = 1, 0$ . En el primero se evalúa la función potencia del contraste ante distintos valores de  $\theta = \{1, 0,99, \dots, 6\}$  para un nivel de significación del 5%. La tabla 3.3 muestra función potencia del contraste para distintos tamaños muestrales ( $n = 20, 30, 50, 100$ ) y distintos valores del coeficiente de correlación entre las perturbaciones  $\rho = \sigma_{12} = \sigma_{21} = \{-0,8, -0,5, -0,2, 0,0, 0,2, 0,5, 0,8\}$ . Como es de esperar, ya que el contraste se ha construido para maximizar su función potencia ante este tipo de alternativas, la probabilidad de rechazar la hipótesis nula de no invertibilidad aumenta a medida que el parámetro  $\theta$  se aleja de 1. La pendiente de la función potencia aumenta

con el tamaño muestral, de forma que el contraste es más potente cuanto mayor sea el tamaño muestral. La correlación entre los errores no parece afectar a la potencia del contraste, lo que es de esperar dada la definición de invarianza sobre la que se construye el estadístico.

En el segundo experimento se comprueban las propiedades de la corrección paramétrica. Para ello se consideran los siguientes procesos generadores de datos en el modelo bivalente con estructura escalar y dos tamaños muestrales,  $n = 50$  y  $n = 100$ .

$$\begin{aligned}(\mathbf{I}_2 - \phi_1 \mathbf{I}_m B) \nabla \mathbf{y}_t &= (1 - \theta_1 B) \mathbf{a}_t, \\ \nabla \mathbf{y}_t &= (1 - \theta_1 B) (\mathbf{I}_2 - \theta_2 \mathbf{I}_m B) \mathbf{a}_t.\end{aligned}$$

Se considera el parámetro interés tal que  $\theta_1 = \{1, 0,99, \dots, 0,8\}$ . En cuanto a los polinomios adicionales AR y MA se considera que ambos tienen raíces reales e iguales a  $\phi_1, \theta_2 = \{-0,8, ,0,5, 0, 0,5, 0,8\}$ . Los resultados se muestran en las tablas 3.5 y 3.6. Si la correlación serial es de tipo AR no existen distorsiones importantes del tamaño empírico ni de la probabilidad de rechazo. La corrección paramétrica funciona bien, incluso en aquellas situaciones en que se da una casi cancelación entre los polinomios autorregresivo y media móvil. Similares resultados se dan para el caso en que la correlación es de tipo MA, salvo cuando el parámetro del polinomio adicional se acerca a uno. En este caso, que coincide con una segunda raíz unitaria media móvil, el tamaño empírico está por debajo de su valor nominal y la probabilidad de rechazo cuando el parámetro de interés,  $\theta_1$ , se aleja de 1 es ligeramente menor que en otros caso.

### 3.5. Extensiones del contraste básico de no invertibilidad

#### 3.5.1. Raíces unitarias y corrección paramétrica

Los contrastes derivados en apartados anteriores están contruidos para contrastar la presencia de un polinomio media móvil  $(\mathbf{I}_m - \Theta B)$  estrictamente no invertible cuando la matriz  $\Theta$  tiene una estructura escalar, de forma que las  $m$  raíces del polinomio son reales e iguales a uno. Dado que en la práctica esta matriz puede ser una matriz cuadrada de coeficientes reales, un polinomio MA(1) matricial es no invertible cuando al menos una raíz de la ecuación característica  $|\mathbf{I}_m - \Theta B| = 0$  sea igual a uno. Utilizando la descomposición espectral de la matriz media móvil,  $\Theta = \mathbf{P} \mathbf{\Lambda} \mathbf{P}^{-1}$ , con  $\mathbf{\Lambda}$  una matriz diagonal de autovalores, la ecuación característica del polinomio MA(1) se puede expresarse como  $|\mathbf{I}_m - \mathbf{\Lambda} B| = \prod_{j=1}^m (1 - \lambda_j B) = 0$ , por lo que el polinomio tendrá tantas raíces unitarias

como autovalores unitarios tenga la matriz  $\Theta$ . Por tanto, sería deseable disponer de un procedimiento que permita detectar la presencia de no invertibilidad en un caso más general como éste. Expresando el polinomio media móvil como  $(\mathbf{I} - \Theta B) = (\mathbf{I}_m - \Theta) + \Theta \nabla$ , el modelo se puede reescribir como  $\nabla \mathbf{y}_t - \Theta(1)\mathbf{a}_t = \nabla \Theta \mathbf{a}_t$ , por lo que el contraste de no invertibilidad calculando el estadístico de contraste (3.8) sobre los residuos exactos del modelo también es potente bajo la alternativa de  $m_1 < m$  raíces unitarias.

Para comprobar que la corrección paramétrica permite contrastar la hipótesis nula de no invertibilidad estricta en situaciones más generales que las consideradas en este capítulo, se diseñan dos estudios Monte Carlo basados en el siguiente proceso generador de datos bivariante

$$\nabla \mathbf{y}_t = (\mathbf{I}_2 - \Theta B)\mathbf{a}_t, \quad \mathbf{a}_t \sim N(\mathbf{0}, \Omega_a),$$

donde  $\Theta = [\theta_{ij}]$  y  $\Omega_a = [\sigma_{ij}]$ . Un primer experimento consiste en evaluar la función potencia de estadístico cuando  $\theta_{11} = 1$ ,  $\theta_{22} \leq 1$  y  $\theta_{12} = \theta_{21} = 0$ . En este caso, en la secuencia de alternativas considerada sólo uno de los parámetros se aleja de 1 de forma que, tanto bajo la hipótesis nula como bajo la alternativa el modelo es no invertible. Sin embargo, el contraste se ha construido para detectar si ambos parámetros son iguales a 1, por lo que éste será potente si se rechaza la hipótesis nula cuando uno solo de los parámetros se aleje de su valor bajo  $H_0$ . La tabla 3.2 muestra las probabilidades de rechazo calculadas para un nivel de significación del 5% utilizando el estadístico de contraste calculado con los residuos exactos evaluados bajo la hipótesis nula. Al igual que en el caso anterior, la probabilidad de rechazo aumenta a medida que uno sólo de los parámetros se aleja de 1, si bien la pendiente de la función potencia es menor que en el caso anterior. De nuevo, la probabilidad de rechazo aumenta con el tamaño muestral y la correlación entre los errores no parece introducir distorsiones graves.

En segundo lugar, si la matriz  $\Theta$  tiene una estructura triangular tal que  $\theta_{11} \neq 0$ ,  $\theta_{22} \neq 0$ ,  $\theta_{21} = 0,2$  y  $\theta_{12} = 0$  con  $\rho = 0,5$ , el contraste utilizando el estadístico de no invertibilidad calculado sobre los residuos exactos, también tiene potencia. El cuadro 3.4 muestra las probabilidades de rechazo en este caso. Como en el caso anterior, esta probabilidad aumenta con el tamaño muestral y, en la línea de los resultados anteriores, la pendiente de la función potencia aumenta a medida que ambos parámetros se acercan a uno.

El principal resultado que se obtiene de estos experimentos es que la corrección paramétrica del estadístico del contraste de no invertibilidad hace que éste sea potente para detectar un polinomio media móvil no invertible, incluso en casos en que la matriz de parámetros es no escalar. Un cuarto caso de especial interés consiste en considerar una estructura cuadrada para  $\Theta$ . Dada la implicación que tiene esta estructura sobre la pre-

sencia de relaciones de cointegración, este caso se estudia separadamente en el capítulo 5.

### 3.5.2. Casos particulares: El modelo diagonal

Existen dos casos particulares de especial interés práctico basados en el modelo IMA(1,1) con estructura diagonal

$$\nabla \mathbf{y}_t = (\mathbf{I}_m - \Theta B) \mathbf{a}_t, \quad \mathbf{a}_t \sim N(\mathbf{0}, \Sigma_a),$$

con  $\Theta = \text{diag}(\theta_1, \dots, \theta_m)$ . En primer lugar, si  $\theta_i = 1$  ( $\forall i \neq j$ ), contrastar la hipótesis nula  $H_0 : \theta_j = 1$  frente a  $H_1 : \theta_j \neq 1$  supone detectar la presencia de una raíz unitaria adicional, sabiendo que existen ya  $m-1$ , esto es, que el polinomio MA es estrictamente no invertible. Este contraste puede interpretarse como un contraste de estacionariedad sobre la variable  $y_j$  cuando ya se sabe que el resto de variables son estacionarias. La diferencia con respecto al contraste de estacionariedad univariante sobre  $y_j$  radica en que en este caso se considera la posible existencia de correlación contemporánea entre las series. En el marco estructural se puede expresar el problema de contraste como  $H_0 : \rho = 0$  frente a  $H_1 : \rho > 0$  sobre el modelo

$$\begin{aligned} \mathbf{y}_t &= \boldsymbol{\mu}_t + \mathbf{u}_t & \mathbf{u}_t &\sim N(\mathbf{0}, \Sigma_u), \\ \mu_{jt} &= \mu_{j,t-1} + v_{jt} & v_{jt} &\sim N(0, \rho\sigma_{jj}) \end{aligned}$$

con  $\boldsymbol{\mu}_t = (\mu_{i0}, \dots, \mu_{jt}, \dots, \mu_{m0})'$  y  $\mu_{i0}$  ( $i \neq j$ ) constantes.

Un segundo caso de interés consiste en considerar  $\theta_i = 0$  ( $\forall i \neq j$ ) y contrastar  $H_0 : \theta_j = 1$  frente a  $H_1 : \theta \neq 1$ . Este contraste se puede interpretar como una prueba de estacionariedad sobre  $y_j$  cuando el resto de variables son no estacionarias. En el marco estructural el problema a contrastar sería  $H_0 : \rho = 0$  frente a  $H_1 : \rho > 0$  en el modelo

$$\begin{aligned} \mathbf{y}_t &= \boldsymbol{\mu}_t + \mathbf{u}_t & \mathbf{u}_t &\sim N(\mathbf{0}, \Sigma_u) \\ \boldsymbol{\mu}_t &= \boldsymbol{\mu}_{t-1} + \mathbf{v}_t & \mathbf{u}_t &\sim N(\mathbf{0}, \Sigma_v(\rho)) \end{aligned}$$

donde  $\Sigma_v(\rho) = \mathbf{c}_\rho \odot \Sigma \odot \mathbf{c}'_\rho$  donde por  $\odot$  se denota el operador producto de Haddamard y  $\mathbf{c}_\rho$  es un vector de unos con  $\rho$  en la  $j$ -ésima posición. Bajo la hipótesis nula  $\Sigma_v$  es una matriz con ceros en la  $i$ -ésima fila y columna, con lo cual el rango será  $m-1$ , y el estadístico sería asimilable al desarrollado por Nyblom y Harvey (2000). Ambos

Cuadro 3.1: Percentiles del estadísticos  $T/(n-1)$ 

$m$	$\alpha$	n=30		n=40		n=50		n=100		$\infty$
		sim	Imhof	sim	Imhof	sim	Imhof	sim	Imhof	
1	50	0.1313	0.1303	0.1268	0.1274	0.1258	0.1257	0.1223	0.1223	0.1196
	20	0.2572	0.2579	0.2509	0.2537	0.2495	0.2512	0.2459	0.2462	0.2422
	10	0.3657	0.3656	0.3589	0.3611	0.3521	0.3583	0.3496	0.3528	0.3484
	5	0.4814	0.4775	0.4715	0.4736	0.4643	0.4712	0.4623	0.4664	0.4623
	1	0.7393	0.7354	0.7260	0.7378	0.7371	0.7393	0.7422	0.7417	0.7432
2	50	0.1517	0.1502	0.1488	0.1474	0.1456	0.1456	0.1420	0.1422	0.1395
	20	0.2482	0.2477	0.2439	0.2440	0.2397	0.2419	0.2366	0.2375	0.2341
	10	0.3162	0.3186	0.3134	0.3148	0.3098	0.3126	0.3059	0.3081	0.3044
	5	0.3794	0.3878	0.3778	0.3843	0.3801	0.3823	0.3742	0.3781	0.3746
	1	0.5171	0.5422	0.5177	0.5410	0.5299	0.5403	0.5214	0.5386	0.5371
3	50	0.1617	0.1585	0.1569	0.1557	0.1552	0.1540	0.1514	0.1505	0.1478
	20	0.2391	0.2392	0.2358	0.2358	0.2337	0.2338	0.2290	0.2297	0.2264
	10	0.2908	0.2944	0.2856	0.2909	0.2860	0.2888	0.2840	0.2846	0.2812
	5	0.3381	0.3469	0.3333	0.3436	0.3361	0.3416	0.3360	0.3375	0.3342
	1	0.4338	0.4618	0.4340	0.4597	0.4405	0.4584	0.4534	0.4557	0.4534
4	50	0.1664	0.1631	0.1625	0.1602	0.1602	0.1585	0.1560	0.1551	0.1524
	20	0.2333	0.2331	0.2299	0.2298	0.2273	0.2278	0.2245	0.2239	0.2207
	10	0.2754	0.2793	0.2729	0.2759	0.2706	0.2739	0.2691	0.2699	0.2666
	5	0.3129	0.3225	0.3123	0.3193	0.3101	0.3173	0.3117	0.3133	0.3101
	1	0.3839	0.4161	0.3912	0.4136	0.3935	0.4120	0.3996	0.4089	0.4063
5	50	0.1691	0.1659	0.1661	0.1630	0.1631	0.1613	0.1583	0.1579	0.1552
	20	0.2269	0.2284	0.2253	0.2252	0.2223	0.2233	0.2194	0.2194	0.2164
	10	0.2612	0.2687	0.2611	0.2654	0.2588	0.2635	0.2575	0.2595	0.2563
	5	0.2923	0.3061	0.2935	0.3028	0.2922	0.3009	0.2921	0.2970	0.2938
	1	0.3494	0.3860	0.3595	0.3833	0.3599	0.3816	0.3678	0.3782	0.3755

contrastos pueden realizarse utilizando la corrección paramétrica fijando los parámetros de la forma descrita.

Cuadro 3.2: Función potencia del contraste para el modelo escalar ( $\alpha = 5\%$ )

$\theta$	$\rho = -0,8$				$\rho = -0,5$				$\rho = -0,2$				$\rho = 0,0$			
	$n = 20$	$n = 30$	$n = 50$	$n = 100$	$n = 20$	$n = 30$	$n = 50$	$n = 100$	$n = 20$	$n = 30$	$n = 50$	$n = 100$	$n = 20$	$n = 30$	$n = 50$	$n = 100$
1.00	0.0422	0.0486	0.0490	0.0482	0.0428	0.0456	0.0518	0.0448	0.0434	0.0406	0.0490	0.0540	0.0456	0.0462	0.0554	0.0566
0.99	0.0436	0.0416	0.0582	0.0598	0.0488	0.0490	0.0526	0.0650	0.0494	0.0496	0.0492	0.0622	0.0484	0.0484	0.0490	0.0680
0.98	0.0458	0.0508	0.0604	0.1220	0.0462	0.0514	0.0636	0.1134	0.0504	0.0516	0.0704	0.1204	0.0456	0.0530	0.0560	0.1186
0.97	0.0470	0.0608	0.0802	0.2138	0.0442	0.0534	0.0816	0.2116	0.0508	0.0546	0.0818	0.2134	0.0460	0.0588	0.0850	0.2064
0.96	0.0496	0.0672	0.1178	0.3342	0.0470	0.0654	0.1098	0.3432	0.0504	0.0710	0.1196	0.3236	0.0508	0.0662	0.1158	0.3308
0.95	0.0600	0.0748	0.1574	0.4394	0.0606	0.0772	0.1544	0.4472	0.0544	0.0828	0.1574	0.4350	0.0556	0.0836	0.1582	0.4460
0.94	0.0578	0.0972	0.2046	0.5596	0.0670	0.0898	0.2142	0.5478	0.0606	0.0924	0.2002	0.5588	0.0642	0.0982	0.2018	0.5596
0.93	0.0696	0.1118	0.2686	0.6412	0.0672	0.1154	0.2578	0.6482	0.0676	0.1104	0.2604	0.6476	0.0782	0.1266	0.2580	0.6558
0.92	0.0758	0.1388	0.3214	0.7092	0.0764	0.1352	0.3268	0.7020	0.0842	0.1392	0.3410	0.7098	0.0718	0.1342	0.3208	0.7030
0.91	0.0922	0.1796	0.3818	0.7722	0.0940	0.1682	0.3690	0.7776	0.0918	0.1722	0.3842	0.7692	0.0908	0.1690	0.3824	0.7804
0.90	0.1030	0.2056	0.4314	0.8178	0.1160	0.1996	0.4558	0.8174	0.1088	0.2024	0.4466	0.8142	0.1038	0.2086	0.4380	0.8114
0.85	0.1892	0.3810	0.6812	0.9400	0.1916	0.3786	0.6740	0.9398	0.1942	0.3820	0.6714	0.9410	0.1936	0.3854	0.6796	0.9414
0.80	0.3068	0.5422	0.8128	0.9784	0.2900	0.5534	0.8208	0.9762	0.3004	0.5440	0.8038	0.9776	0.2982	0.5410	0.8158	0.9764
0.75	0.4322	0.6692	0.8840	0.9900	0.4268	0.6650	0.8808	0.9912	0.4284	0.6798	0.8844	0.9898	0.4242	0.6790	0.8868	0.9928
0.70	0.5398	0.7586	0.9348	0.9946	0.5458	0.7568	0.9334	0.9940	0.5304	0.7516	0.9278	0.9952	0.5380	0.7604	0.9360	0.9962
0.65	0.6330	0.8256	0.9540	0.9978	0.6194	0.8266	0.9602	0.9976	0.6264	0.8348	0.9558	0.9988	0.6244	0.8270	0.9522	0.9972
0.60	0.6864	0.8642	0.9698	0.9990	0.6926	0.8776	0.9710	0.9984	0.7020	0.8642	0.9698	0.9986	0.7060	0.8602	0.9714	0.9992

$\theta$	$\rho = 0,2$				$\rho = 0,5$				$\rho = 0,8$			
	$n = 20$	$n = 30$	$n = 50$	$n = 100$	$n = 20$	$n = 30$	$n = 50$	$n = 100$	$n = 20$	$n = 30$	$n = 50$	$n = 100$
1.00	0.0428	0.0450	0.0468	0.0508	0.0460	0.0472	0.0498	0.0492	0.0498	0.0480	0.0460	0.0442
0.99	0.0470	0.0510	0.0470	0.0600	0.0434	0.0440	0.0544	0.0664	0.0468	0.0422	0.0548	0.0618
0.98	0.0450	0.0550	0.0620	0.1238	0.0440	0.0466	0.0618	0.1156	0.0436	0.0518	0.0592	0.1160
0.97	0.0534	0.0556	0.0774	0.2056	0.0478	0.0558	0.0810	0.2066	0.0522	0.0542	0.0894	0.2048
0.96	0.0568	0.0716	0.1100	0.3254	0.0566	0.0612	0.1096	0.3252	0.0504	0.0636	0.1228	0.3344
0.95	0.0580	0.0770	0.1636	0.4414	0.0570	0.0814	0.1520	0.4524	0.0586	0.0772	0.1598	0.4540
0.94	0.0636	0.0956	0.2042	0.5356	0.0634	0.0964	0.2018	0.5618	0.0532	0.0926	0.2036	0.5514
0.93	0.0700	0.1160	0.2652	0.6542	0.0700	0.1138	0.2694	0.6398	0.0700	0.1152	0.2670	0.6312
0.92	0.0766	0.1382	0.3140	0.7128	0.0764	0.1354	0.3256	0.7246	0.0834	0.1402	0.3418	0.7140
0.91	0.0890	0.1756	0.3868	0.7684	0.0866	0.1718	0.3784	0.7740	0.0916	0.1672	0.3826	0.7668
0.90	0.1076	0.2048	0.4462	0.8232	0.0972	0.2058	0.4452	0.8226	0.0986	0.1902	0.4342	0.8102
0.85	0.2046	0.3766	0.6814	0.9366	0.1804	0.3742	0.6780	0.9414	0.1948	0.3852	0.6646	0.9328
0.80	0.3018	0.5412	0.8070	0.9782	0.3150	0.5404	0.8152	0.9774	0.3098	0.5466	0.8108	0.9778
0.75	0.4392	0.6726	0.8906	0.9904	0.4266	0.6716	0.8850	0.9904	0.4204	0.6716	0.8944	0.9884
0.70	0.5356	0.7678	0.9366	0.9952	0.5394	0.7698	0.9320	0.9948	0.5202	0.7558	0.9330	0.9962
0.65	0.6146	0.8254	0.9548	0.9976	0.6236	0.8252	0.9578	0.9978	0.6248	0.8218	0.9560	0.9970
0.60	0.7072	0.8634	0.9690	0.9992	0.6978	0.8658	0.9700	0.9996	0.7078	0.8718	0.9716	0.9982

Cuadro 3.3: Función potencia del contraste para el modelo diagonal ( $\alpha = 5\%$ )

$\theta_{22}$	$\rho = -0,8$				$\rho = -0,5$				$\rho = -0,2$				$\rho = 0,0$			
	$n = 20$	$n = 30$	$n = 50$	$n = 100$	$n = 20$	$n = 30$	$n = 50$	$n = 100$	$n = 20$	$n = 30$	$n = 50$	$n = 100$	$n = 20$	$n = 30$	$n = 50$	$n = 100$
1.00	0.0392	0.0454	0.0484	0.0492	0.0438	0.0474	0.0492	0.0448	0.0414	0.0416	0.0490	0.0454	0.0466	0.0472	0.0506	0.0484
0.99	0.0508	0.0498	0.0528	0.0666	0.0386	0.0546	0.0572	0.0594	0.0434	0.0450	0.0472	0.0544	0.0420	0.0492	0.0482	0.0592
0.98	0.0534	0.0530	0.0684	0.1430	0.0472	0.0524	0.0546	0.1004	0.0440	0.0564	0.0578	0.0808	0.0416	0.0488	0.0556	0.0792
0.97	0.0470	0.0612	0.1012	0.2516	0.0466	0.0590	0.0722	0.1468	0.0420	0.0518	0.0718	0.1338	0.0472	0.0524	0.0638	0.1278
0.96	0.0608	0.0856	0.1390	0.3636	0.0598	0.0640	0.0874	0.2292	0.0472	0.0546	0.0798	0.1936	0.0484	0.0562	0.0822	0.1964
0.95	0.0632	0.0972	0.1968	0.4670	0.0478	0.0724	0.1178	0.3222	0.0566	0.0650	0.1056	0.2608	0.0516	0.0618	0.1044	0.2560
0.94	0.0744	0.1202	0.2536	0.5500	0.0522	0.0776	0.1572	0.3892	0.0606	0.0814	0.1300	0.3318	0.0566	0.0670	0.1256	0.3258
0.93	0.0882	0.1536	0.3066	0.6156	0.0634	0.1000	0.1922	0.4526	0.0570	0.0818	0.1552	0.4040	0.0576	0.0796	0.1562	0.3896
0.92	0.0948	0.1774	0.3598	0.6564	0.0670	0.1108	0.2362	0.5196	0.0628	0.0850	0.1868	0.4620	0.0634	0.0888	0.1892	0.4626
0.91	0.1090	0.2224	0.4052	0.7042	0.0824	0.1406	0.2644	0.5582	0.0750	0.1082	0.2300	0.5164	0.0600	0.1056	0.2162	0.5010
0.90	0.1358	0.2402	0.4676	0.7386	0.0888	0.1348	0.3062	0.6064	0.0636	0.1198	0.2616	0.5642	0.0740	0.1140	0.2530	0.5346
0.85	0.2306	0.3984	0.6208	0.8560	0.1394	0.2680	0.4840	0.7724	0.1192	0.2084	0.4238	0.7172	0.1144	0.2230	0.4112	0.7190
0.80	0.3296	0.5132	0.7246	0.9126	0.2124	0.3754	0.6150	0.8560	0.1750	0.3258	0.5566	0.8112	0.1764	0.3254	0.5396	0.8134
0.75	0.4142	0.6032	0.7868	0.9354	0.2982	0.4664	0.6884	0.8946	0.2542	0.4210	0.6528	0.8908	0.2454	0.4130	0.6352	0.8718
0.70	0.4892	0.6568	0.8230	0.9606	0.3614	0.5382	0.7518	0.9300	0.3122	0.4920	0.7188	0.9086	0.3096	0.4920	0.7052	0.9126
0.65	0.5492	0.7030	0.8606	0.9674	0.4278	0.5946	0.7910	0.9444	0.3770	0.5600	0.7670	0.9408	0.3734	0.5490	0.7578	0.9326
0.60	0.5844	0.7466	0.8812	0.9700	0.4742	0.6558	0.8338	0.9574	0.4508	0.6180	0.8062	0.9524	0.4410	0.6066	0.8018	0.9574

$\theta_{22}$	$\rho = 0,2$				$\rho = 0,5$				$\rho = 0,8$			
	$n = 20$	$n = 30$	$n = 50$	$n = 100$	$n = 20$	$n = 30$	$n = 50$	$n = 100$	$n = 20$	$n = 30$	$n = 50$	$n = 100$
1.00	0.0414	0.0412	0.0498	0.0498	0.0410	0.0490	0.0442	0.0488	0.0442	0.0446	0.0474	0.0456
0.99	0.0416	0.0506	0.0482	0.0482	0.0362	0.0480	0.0484	0.0632	0.0432	0.0464	0.0598	0.0728
0.98	0.0426	0.0458	0.0572	0.0874	0.0480	0.0510	0.0590	0.0962	0.0454	0.0546	0.0652	0.1546
0.97	0.0438	0.0468	0.0672	0.1386	0.0502	0.0552	0.0718	0.1586	0.0540	0.0678	0.1008	0.2600
0.96	0.0420	0.0604	0.0876	0.1962	0.0532	0.0636	0.0876	0.2382	0.0586	0.0766	0.1530	0.3626
0.95	0.0558	0.0630	0.0960	0.2578	0.0556	0.0672	0.1220	0.3070	0.0638	0.0934	0.2066	0.4582
0.94	0.0546	0.0736	0.1306	0.3316	0.0560	0.0782	0.1562	0.3802	0.0760	0.1222	0.2576	0.5496
0.93	0.0536	0.0812	0.1540	0.4058	0.0604	0.0980	0.1980	0.4426	0.0858	0.1564	0.3128	0.6072
0.92	0.0668	0.0968	0.1980	0.4590	0.0620	0.1084	0.2154	0.5182	0.0958	0.1808	0.3696	0.6612
0.91	0.0754	0.1004	0.2254	0.5128	0.0774	0.1196	0.2744	0.5582	0.1138	0.2166	0.4052	0.6942
0.90	0.0682	0.1246	0.2620	0.5640	0.0776	0.1420	0.2988	0.5964	0.1412	0.2362	0.4568	0.7374
0.85	0.1076	0.2152	0.4240	0.7228	0.1402	0.2616	0.4702	0.7616	0.2216	0.3944	0.6300	0.8710
0.80	0.1834	0.3298	0.5558	0.8174	0.2084	0.3770	0.6076	0.8546	0.3286	0.5192	0.7180	0.9120
0.75	0.2536	0.4192	0.6414	0.8784	0.2806	0.4764	0.6990	0.8984	0.4192	0.6084	0.7848	0.9376
0.70	0.3240	0.5034	0.7050	0.9170	0.3710	0.5442	0.7536	0.9254	0.4918	0.6556	0.8254	0.9524
0.65	0.3804	0.5512	0.7556	0.9404	0.4300	0.6122	0.7954	0.9458	0.5494	0.6932	0.8558	0.9698
0.60	0.4334	0.6198	0.8020	0.9494	0.4860	0.6626	0.8344	0.9592	0.5826	0.7338	0.8870	0.9698

Cuadro 3.4: Función potencia del contraste para el modelo triangular ( $\alpha = 5\%$ )

$\theta_1$	$\theta_2 = -0,8$				$\theta_2 = -0,5$				$\theta_2 = -0,2$			
	$n = 20$	$n = 30$	$n = 50$	$n = 100$	$n = 20$	$n = 30$	$n = 50$	$n = 100$	$n = 20$	$n = 30$	$n = 50$	$n = 100$
1.00	0.0472	0.0462	0.0484	0.0500	0.0508	0.0452	0.0492	0.0518	0.0410	0.0428	0.0492	0.0452
0.99	0.0440	0.0508	0.0572	0.0604	0.0440	0.0484	0.0478	0.0568	0.0474	0.0478	0.0480	0.0642
0.98	0.0492	0.0454	0.0572	0.0856	0.0494	0.0534	0.0514	0.0896	0.0434	0.0522	0.0518	0.0856
0.97	0.0514	0.0530	0.0668	0.1388	0.0444	0.0588	0.0614	0.1388	0.0496	0.0528	0.0692	0.1420
0.96	0.0510	0.0558	0.0848	0.2262	0.0502	0.0602	0.0894	0.2188	0.0556	0.0594	0.0790	0.2096
0.95	0.0574	0.0616	0.1158	0.2914	0.0576	0.0616	0.1122	0.2828	0.0466	0.0630	0.1096	0.2844
0.94	0.0594	0.0700	0.1398	0.3594	0.0544	0.0814	0.1342	0.3520	0.0532	0.0738	0.1346	0.3496
0.93	0.0634	0.0938	0.1780	0.4254	0.0626	0.0882	0.1722	0.4194	0.0636	0.0824	0.1692	0.4288
0.92	0.0698	0.1016	0.2172	0.4808	0.0586	0.1018	0.2056	0.4868	0.0698	0.0996	0.2210	0.4754
0.91	0.0766	0.1086	0.2446	0.5424	0.0734	0.1224	0.2458	0.5368	0.0762	0.1176	0.2300	0.5274
0.90	0.0776	0.1338	0.2842	0.5902	0.0792	0.1364	0.2778	0.5768	0.0834	0.1318	0.2938	0.5856
0.85	0.1332	0.2536	0.4514	0.7396	0.1272	0.2456	0.4508	0.7386	0.1278	0.2450	0.4498	0.7456
0.80	0.1954	0.3514	0.5748	0.8404	0.2084	0.3518	0.5820	0.8236	0.1736	0.3478	0.5712	0.8324
0.75	0.2640	0.4496	0.6646	0.8930	0.2702	0.4346	0.6628	0.8940	0.2560	0.4278	0.6524	0.8842
0.70	0.3482	0.5138	0.7348	0.9190	0.3364	0.5164	0.7214	0.9190	0.3324	0.5026	0.7328	0.9216
0.65	0.4160	0.5844	0.7828	0.9344	0.3988	0.5740	0.7774	0.9374	0.3996	0.5708	0.7638	0.9348
0.60	0.4642	0.6134	0.8200	0.9494	0.4594	0.6296	0.8036	0.9588	0.4532	0.6172	0.7954	0.9494

$\theta_1$	$\theta_2 = 0,2$				$\theta_2 = 0,5$				$\theta_2 = 0,8$			
	$n = 20$	$n = 30$	$n = 50$	$n = 100$	$n = 20$	$n = 30$	$n = 50$	$n = 100$	$n = 20$	$n = 30$	$n = 50$	$n = 100$
1.00	0.0442	0.0460	0.0468	0.0452	0.0458	0.0470	0.0454	0.0516	0.0326	0.0338	0.0366	0.0440
0.99	0.0460	0.0510	0.0504	0.0614	0.0450	0.0486	0.0434	0.0584	0.0348	0.0336	0.0418	0.0490
0.98	0.0476	0.0512	0.0556	0.0870	0.0412	0.0456	0.0604	0.0860	0.0360	0.0348	0.0446	0.0838
0.97	0.0426	0.0502	0.0654	0.1408	0.0480	0.0512	0.0612	0.1238	0.0382	0.0388	0.0600	0.1320
0.96	0.0486	0.0554	0.0878	0.1966	0.0480	0.0528	0.0748	0.1924	0.0424	0.0418	0.0754	0.1922
0.95	0.0554	0.0666	0.1052	0.2646	0.0452	0.0582	0.1010	0.2526	0.0444	0.0508	0.0904	0.2792
0.94	0.0548	0.0750	0.1322	0.3512	0.0498	0.0706	0.1312	0.3358	0.0456	0.0624	0.1204	0.3540
0.93	0.0576	0.0812	0.1588	0.4018	0.0564	0.0880	0.1630	0.3930	0.0428	0.0786	0.1534	0.4266
0.92	0.0648	0.0986	0.1988	0.4668	0.0580	0.0960	0.1880	0.4636	0.0510	0.0852	0.1814	0.4840
0.91	0.0678	0.1112	0.2314	0.5160	0.0668	0.1000	0.2226	0.5070	0.0636	0.0888	0.2298	0.5430
0.90	0.0812	0.1264	0.2660	0.5690	0.0750	0.1262	0.2566	0.5556	0.0678	0.1040	0.2398	0.5924
0.85	0.1174	0.2292	0.4354	0.7256	0.1174	0.2110	0.4158	0.7146	0.1096	0.2014	0.4334	0.7556
0.80	0.1832	0.3270	0.5632	0.8264	0.1632	0.3186	0.5480	0.8234	0.1762	0.3278	0.5762	0.8518
0.75	0.2552	0.4218	0.6518	0.8712	0.2438	0.4050	0.6342	0.8758	0.2484	0.4186	0.6772	0.9080
0.70	0.3188	0.5040	0.7150	0.9124	0.3088	0.4972	0.7070	0.9036	0.3082	0.5028	0.7420	0.9292
0.65	0.3922	0.5742	0.7740	0.9398	0.3740	0.5646	0.7648	0.9396	0.3810	0.5686	0.7906	0.9522
0.60	0.4450	0.6192	0.8020	0.9526	0.4282	0.6038	0.7894	0.9448	0.4436	0.6358	0.8250	0.9642

Cuadro 3.5: Función potencia para un VARIMA(1,1,1) ( $\alpha = 5\%$ )

$\theta_1, \phi$	$n = 50$					$n = 100$				
	-0,8	-0,5	0	0,5	0,8	-0,8	-0,5	0	0,5	0,8
1.00	0.043	0.040	0.039	0.049	0.064	0.065	0.050	0.048	0.051	0.054
0.99	0.037	0.045	0.052	0.056	0.073	0.071	0.067	0.058	0.088	0.072
0.98	0.068	0.062	0.064	0.069	0.079	0.108	0.106	0.115	0.104	0.122
0.97	0.078	0.091	0.088	0.091	0.080	0.195	0.192	0.216	0.206	0.179
0.96	0.092	0.096	0.126	0.108	0.117	0.363	0.326	0.320	0.326	0.297
0.95	0.132	0.150	0.152	0.142	0.160	0.418	0.435	0.441	0.461	0.422
0.94	0.185	0.188	0.204	0.173	0.185	0.526	0.564	0.542	0.544	0.526
0.93	0.233	0.252	0.248	0.219	0.228	0.648	0.642	0.665	0.614	0.617
0.92	0.319	0.304	0.331	0.316	0.275	0.723	0.681	0.719	0.711	0.701
0.91	0.373	0.377	0.424	0.383	0.364	0.748	0.755	0.768	0.760	0.760
0.90	0.439	0.431	0.464	0.428	0.407	0.800	0.814	0.823	0.818	0.820
0.89	0.468	0.497	0.502	0.492	0.439	0.850	0.835	0.841	0.840	0.840
0.88	0.532	0.540	0.550	0.526	0.519	0.877	0.892	0.899	0.882	0.894
0.87	0.574	0.572	0.604	0.589	0.564	0.906	0.902	0.913	0.906	0.917
0.86	0.629	0.631	0.637	0.633	0.615	0.927	0.919	0.934	0.914	0.927
0.85	0.664	0.666	0.676	0.660	0.649	0.927	0.933	0.945	0.943	0.951
0.84	0.706	0.724	0.690	0.717	0.715	0.940	0.943	0.945	0.948	0.951
0.83	0.755	0.734	0.750	0.728	0.731	0.965	0.953	0.958	0.969	0.956
0.82	0.751	0.750	0.789	0.773	0.753	0.962	0.963	0.977	0.963	0.958
0.81	0.771	0.802	0.817	0.790	0.771	0.972	0.969	0.980	0.968	0.972
0.80	0.805	0.816	0.797	0.795	0.825	0.969	0.969	0.981	0.983	0.977

Cuadro 3.6: Función potencia para un VIMA(1,2) ( $\alpha = 5\%$ )

$\theta_1, \theta_2$	$n = 50$					$n = 100$				
	-0,8	-0,5	0	0,5	0,8	-0,8	-0,5	0	0,5	0,8
1.00	0.050	0.041	0.039	0.035	0.014	0.047	0.044	0.048	0.042	0.027
0.99	0.059	0.061	0.052	0.042	0.008	0.071	0.061	0.058	0.059	0.027
0.98	0.084	0.083	0.064	0.038	0.011	0.110	0.135	0.115	0.090	0.065
0.97	0.106	0.107	0.088	0.057	0.017	0.207	0.214	0.216	0.179	0.120
0.96	0.138	0.123	0.126	0.099	0.025	0.366	0.328	0.320	0.311	0.221
0.95	0.154	0.191	0.152	0.112	0.034	0.468	0.467	0.441	0.434	0.329
0.94	0.227	0.218	0.204	0.162	0.061	0.582	0.581	0.542	0.521	0.428
0.93	0.297	0.278	0.248	0.218	0.073	0.632	0.685	0.665	0.614	0.532
0.92	0.354	0.363	0.331	0.273	0.133	0.687	0.708	0.719	0.695	0.634
0.91	0.404	0.408	0.424	0.314	0.168	0.798	0.782	0.768	0.757	0.694
0.90	0.490	0.475	0.464	0.391	0.187	0.839	0.826	0.823	0.801	0.718
0.89	0.513	0.529	0.502	0.419	0.248	0.867	0.843	0.841	0.845	0.774
0.88	0.572	0.616	0.550	0.478	0.277	0.897	0.890	0.899	0.876	0.810
0.87	0.625	0.627	0.604	0.526	0.355	0.914	0.912	0.913	0.892	0.860
0.86	0.669	0.651	0.637	0.584	0.378	0.925	0.941	0.934	0.926	0.861
0.85	0.691	0.677	0.676	0.622	0.463	0.955	0.941	0.945	0.929	0.898
0.84	0.758	0.727	0.690	0.660	0.477	0.953	0.943	0.945	0.948	0.905
0.83	0.745	0.767	0.750	0.683	0.491	0.964	0.963	0.958	0.953	0.931
0.82	0.787	0.803	0.789	0.734	0.518	0.968	0.975	0.977	0.957	0.938
0.81	0.813	0.827	0.817	0.759	0.600	0.980	0.972	0.980	0.967	0.960
0.80	0.825	0.823	0.797	0.806	0.628	0.989	0.976	0.981	0.971	0.951



## Capítulo 4

# Contrastes de no invertibilidad en modelos estacionales

### 4.1. Introducción

La metodología de King y Hillier (1985) permite derivar una amplia clase de contrastes para detectar la presencia de componentes deterministas en modelos de series temporales univariantes. Miembros de esta clase son los contrastes localmente óptimos e invariantes (LBI) de Nyblom y Mäkeläinen (1983), Nyblom (1986), Kwiatkowski et al. (1992) o los contrastes estacionales de Nyblom y Harvey (2001), Canova y Hansen (1995), Caner (1998), Buseti y Harvey (2003), Taylor (2003a) o Taylor (2003b). Igualmente, esta metodología permite derivar los contrastes localmente óptimos, insesgados e invariantes (LBIU) para contrastar la presencia de una raíz unitaria en el polinomio media móvil de un modelo ARIMA. En este marco, las principales contribuciones son las desarrolladas por Tanaka (1990), Saikkonen y Luukkonen (1993a) o los contrastes estacionales de Tam y Reinsel (1997) o Tam y Reinsel (1998). Existen extensiones multivariantes de estos contrastes sólo en el marco estructural (Nyblom y Harvey, 2000, 2001; Buseti, 2006) cuya derivación se basa en la extensión multivariante de la metodología King y Hillier (1985) propuesta por Nyblom (2001). En este capítulo se muestra la relación entre estos contrastes de componentes deterministas y los contrastes de no invertibilidad para la forma reducida ARIMA de cada modelo. Para cada estadístico de contraste se derivan además expresiones cerradas para sus dos primeros momentos utilizando los resultados de Nyblom (2001).

Una vez obtenidos los estadísticos de contraste para cada modelo, es necesaria la caracterización de su distribución en muestras finitas para implementar el contraste. En el caso univariante, los estadísticos de los contrastes listados anteriormente se formulan como ratios de formas cuadráticas en variables normales, cuya función de distribución se puede evaluar mediante la inversión numérica de su función característica utilizando los procedimientos de Imhof (1961) o Davies (1973). Sin embargo, en el caso multivariante los estadísticos se expresan como combinaciones lineales de variables Beta, cuya distribución debe ser tabulada por simulación. En estos casos en que la distribución en muestras finitas del estadístico es complicada de evaluar resulta de utilidad disponer de un procedimiento sencillo y preciso para calcular p-valores y cuantiles. En un problema similar, MacKinnon (1994) o Doornik (1998) proponen aproximaciones a la distribución en muestras finitas de estadísticos tipo Dickey-Fuller y de cointegración basadas en regresiones de superficie de respuesta o la función de distribución Gamma. Utilizando la segunda aproximación, se comprueba que la función de distribución Inversa Gaussiana (IG) proporciona un buen ajuste de la distribución de los estadísticos considerados en el presente capítulo, por lo que puede utilizarse para calcular de forma sencilla sus p-valores.

Los resultados de este capítulo se organizan de la siguiente manera: En el segundo apartado se resume el procedimiento utilizado para derivar los estadísticos de contraste considerados, así como la forma de obtener la media y varianza de los mismos. En el apartado 3 se obtiene el contraste multivariante de nivel estacional determinista, sus extensiones considerando la presencia de un conjunto de componentes deterministas y su relación con los contrastes de no invertibilidad en los correspondientes modelos reducidos ARIMA. En cada caso, se obtienen expresiones cerradas para la media y la varianza del estadístico de contraste. En el cuarto apartado se obtienen los contrastes para detectar la presencia de uno o varios ciclos estacionales deterministas y los dos primeros momentos del estadístico de contraste. En el quinto apartado se describe y evalúa la aproximación IG a la distribución en muestras finitas de los estadísticos considerados y en el último apartado se ilustra el procedimiento de contraste con un ejemplo. Los resultados de este capítulo aparecen detallados también en Gallego y Díaz (2011).

## 4.2. Contrastes invariantes para estructuras de covarianzas

Como se ha mostrado en el capítulo anterior los estadísticos de contraste de no invertibilidad en modelos VARIMA se pueden derivar de una forma sencilla utilizando su representación estructural equivalente. Nyblom (2001) proporciona una extensión multivariante de la metodología de King (1980) y King y Hillier (1985) cuyos resultados

principales se resumen en esta sección. Partiendo del modelo de regresión multivariante

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\Pi} + \mathbf{E}, \quad \mathbf{E} \sim N(\mathbf{0}, \boldsymbol{\Omega}(\delta) \otimes \boldsymbol{\Sigma}), \quad (4.1)$$

donde  $\mathbf{Y}$  es una matriz  $T \times m$  que contiene por columnas las  $m$  series temporales, de forma que la  $t$ -ésima fila de  $\mathbf{Y}$  es  $\mathbf{y}_t$ .  $\mathbf{X}$  es una matriz fija de dimensiones  $T \times k$  y  $\boldsymbol{\Pi}$  es la correspondiente matriz  $k \times m$  de coeficientes. La matriz  $T \times m$  de perturbaciones  $\mathbf{E} = [\mathbf{e}_1, \dots, \mathbf{e}_T]'$  es tal que  $E(\mathbf{E}) = \mathbf{0}$  y  $\text{var}(\text{vec}(\mathbf{E}')) = \boldsymbol{\Omega}(\delta) \otimes \boldsymbol{\Sigma}$  donde  $\boldsymbol{\Sigma}$  es una matriz  $m \times m$  definida positiva y la matriz  $T \times T$  y  $\boldsymbol{\Omega}(\rho)$  depende del parámetro  $\rho$  de tal forma que  $\boldsymbol{\Omega}(0) = \mathbf{I}_T$ , luego la única correlación entre las series es la contemporánea. Nyblom (2001) demuestra que en este marco, el contraste localmente óptimo e invariante (LBI) para contrastar  $H_0 : \rho = 0$  frente a  $H_1 : \rho > 0$ , definiendo la invarianza ante transformaciones del tipo  $\mathbf{y}_t \mapsto \mathbf{P}\mathbf{y}_t + \mathbf{X}\mathbf{A}$ , con  $\mathbf{A}$  cualquier matriz arbitraria  $k \times m$  y  $\mathbf{P}$  cuadrada y definida positiva, tiene la siguiente expresión general

$$L = \text{tr} \left[ (\hat{\mathbf{E}}' \hat{\mathbf{E}})^{-1} \hat{\mathbf{E}}' \mathbf{K} \hat{\mathbf{E}} \right], \quad (4.2)$$

donde  $\hat{\mathbf{E}} = \mathbf{M}\mathbf{Y}$  es la matriz de residuos de mínimos cuadrados multivariantes de  $\mathbf{Y}$  sobre la matriz  $\mathbf{X}$ , de forma que  $\mathbf{M} = \mathbf{I}_{T-k} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$  y  $\mathbf{K} = d\boldsymbol{\Omega}(\rho)/d\rho|_{\rho=0}$ . De King y Hillier (1985) se puede demostrar que el estadístico es también localmente óptimo, insesgado e invariante (LBIU) para contrastar  $H_0 : \rho = 0$  frente a la alternativa de dos colas  $H_1 : \rho \neq 1$ . Bajo la hipótesis nula, el estadístico se puede expresar como

$$L = \left( \sum_{t=1}^{T-k} \boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_t' \right)^{-1} \left( \sum_{t=1}^{T-k} \lambda_t \boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_t' \right) = \sum_{t=1}^{T-k} \lambda_t \boldsymbol{\epsilon}_t' \underbrace{\left( \sum_{t=1}^{T-k} \boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_t' \right)^{-1}}_{B_t} \boldsymbol{\epsilon}_t, \quad (4.3)$$

donde  $\lambda_t$  ( $t = 1, \dots, T-k$ ) son los autovalores no nulos de la matriz  $\mathbf{M}\mathbf{K}$  y  $\boldsymbol{\epsilon}_t \sim N(\mathbf{0}, \mathbf{I}_m)$ , de forma que  $(\sum_{t=1}^{T-k} \boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_t') \sim W_m(T-k, \mathbf{I}_m)$  es una distribución Wishart con  $T-k$  grados de libertad. El segundo término de la igualdad en (4.3) donde  $B_t \sim \text{Beta}(\frac{m}{2}, \frac{T-k-m}{2})$  con  $B_1 = 1 - B_2 - \dots - B_{T-k-1}$  (ver Rao, 2002, pp. 540) permite expresar el estadístico como una suma ponderada de variables aleatorias Beta. A partir de esta última formulación

del estadístico, Nyblom (2001) obtiene sus dos primeros momentos como

$$E(L) = m\bar{\lambda}_1, \quad \bar{\lambda}_1 = \frac{1}{T-k} \sum_{t=1}^{T-k} \lambda_t, \quad (4.4)$$

$$V(L) = \frac{2m(T-k-m)}{(T-k-1)(T-k+2)} [\bar{\lambda}_2 - \bar{\lambda}_1^2], \quad \bar{\lambda}_2 = \frac{1}{T-k} \sum_{t=1}^{T-k} \lambda_t^2. \quad (4.5)$$

Para cada modelo derivado en este capítulo se derivan las expresiones de los estadísticos de contraste y las expresiones de  $\bar{\lambda}_1$  y  $\bar{\lambda}_2$  que permiten calcular sus dos primeros momentos.

### 4.3. Contrastes de componentes deterministas

#### 4.3.1. Nivel local estacional

El modelo de nivel local multivariante (3.1)-(3.2) tratado ampliamente en el capítulo anterior se puede extender al caso estacional (con periodo  $s$ ) como

$$\begin{aligned} \mathbf{y}_t &= \boldsymbol{\mu}_t + \mathbf{u}_t, & \mathbf{u}_t &\sim N(\mathbf{0}, \boldsymbol{\Sigma}_u), \\ \boldsymbol{\mu}_t &= \boldsymbol{\mu}_{t-s} + \mathbf{v}_t, & \mathbf{v}_t &\sim N(\mathbf{0}, \rho \boldsymbol{\Sigma}_u), \end{aligned} \quad (4.6)$$

con  $t = 1, \dots, T$  donde por simplicidad, y sin pérdida de generalidad, se supone que  $T = ns$  siendo  $n$  el número de años completos. De esta forma, el vector  $\mathbf{y}_t$  se descompone como la suma de un paseo aleatorio multivariante  $\boldsymbol{\mu}_t$  y un proceso ruido blanco Gaussiano  $\mathbf{u}_t$ . El vector  $\mathbf{v}_t$  es también un ruido blanco multivariante independiente de  $\mathbf{u}_t$  y donde el parámetro  $\rho \geq 0$  mide el grado de estocasticidad de  $\mathbf{v}_t$ , de forma que si  $\rho = 0$ , cada una de las  $m$  series de  $\mathbf{y}_t$  es estacionaria alrededor de  $s$  variables ficticias estacionales. Definiendo las matrices  $\mathbf{Y} = [\mathbf{y}_1, \dots, \mathbf{y}_T]'$ ,  $\mathbf{A} = [\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_T]'$ ,  $\mathbf{U} = [\mathbf{u}_1, \dots, \mathbf{u}_T]'$  y  $\mathbf{V} = [\mathbf{v}_1, \dots, \mathbf{v}_T]'$ , (4.6) se puede expresar en forma matricial como

$$\begin{aligned} \mathbf{Y} &= \mathbf{A} + \mathbf{U}, & \mathbf{U} &\sim N(\mathbf{0}, \mathbf{I}_T \otimes \boldsymbol{\Sigma}), \\ (\nabla_n \otimes \mathbf{I}_s) \mathbf{A} &= (\mathbf{d}_n \otimes \mathbf{I}_s) \mathbf{A}_0 + \mathbf{V}, & \mathbf{V} &\sim N(\mathbf{0}, \mathbf{I}_T \otimes \rho \boldsymbol{\Sigma}), \end{aligned} \quad (4.7)$$

donde  $\mathbf{A}_0 = [\boldsymbol{\mu}_{-s+1}, \dots, \boldsymbol{\mu}_0]'$  es una matriz  $s \times m$  de condiciones iniciales,  $\otimes$  denota el producto de Kronecker,  $\nabla_n$  es una matriz  $n \times n$  bidiagonal inferior con unos en la diagonal principal y -1 en la primera subdiagonal inferior, que se puede expresar

como  $\nabla_n = [\mathbf{d}_n, \mathbf{D}'_n]'$  con  $\mathbf{d}_n = (1, 0, \dots, 0)'$  y  $\mathbf{D}_n$  la matriz  $(n-1) \times n$  generadora de primeras diferencias. Si se supone  $\mathbf{A}_0$  fija, (4.7) es un caso especial del modelo de regresión multivariante (4.1) con

$$\mathbf{X} = \mathbf{i}_n \otimes \mathbf{I}_s, \quad \mathbf{\Pi} = \mathbf{A}_0, \quad \mathbf{\Omega}(\rho) = \mathbf{I}_T + \rho(\mathbf{C}_n \mathbf{C}'_n \otimes \mathbf{I}_s),$$

donde  $\mathbf{C}_n = \nabla_n^{-1}$ ,  $\mathbf{i}_n = \mathbf{C}_n \mathbf{d}_n$  y  $\mathbf{X}$  es una matriz  $T \times s$  de variables ficticias estacionales.

La expresión del estadístico del contraste LBI para la hipótesis nula de estacionalidad determinista ( $H_0 : \rho = 0$ ) frente a la alternativa de paseo aleatorio estacional más ruido ( $H_1 : \rho > 0$ ), que se denotará por  $RW_{m,s,n}$  viene dada por (4.2) con  $\mathbf{K} = \mathbf{C}_n \mathbf{C}'_n \otimes \mathbf{I}_m$  y que se puede calcular como

$$RW_{m,s,n} = \text{tr} \left\{ \left( \sum_{t=1}^T \hat{\mathbf{u}}_t \hat{\mathbf{u}}'_t \right)^{-1} \sum_{j=1}^s \left[ \sum_{t=1}^n \left( \sum_{\tau=1}^t \hat{\mathbf{u}}_{(\tau-1)s+j} \right) \left( \sum_{\tau=1}^t \hat{\mathbf{u}}_{(\tau-1)s+j} \right)' \right] \right\}, \quad (4.8)$$

donde  $\hat{\mathbf{u}}_t$  son los residuos en la regresión multivariante de  $\mathbf{y}_t$  sobre  $s$  variables ficticias estacionales. Los dos primeros momentos de  $RW_{m,s,n}/(n-1)m$  se pueden obtener a partir de (4.4) y (4.5), donde las expresiones cerradas para el autovalor medio y la media cuadrática de los autovalores de  $\mathbf{MK}$  son

$$\bar{\lambda}_1 = \frac{n+1}{6} \quad \text{and} \quad \bar{\lambda}_2 = \bar{\lambda}_1 \left( \frac{2n^2+7}{30} \right)$$

Estas expresiones sugieren corregir el estadístico  $RW_{m,s,n}$  por un factor dependiente del tamaño muestral para que sus momentos no diverjan y la distribución no sea degenerada. Algunos candidatos son  $n$ ,  $nm$  o  $(n-1)m$ , donde el factor  $m$  es arbitrario. Nótese que  $RW_{m,s,n}/nm$  tiene media asintótica igual a  $1/6$  y varianza  $1/45ms$ , que para el caso regular univariante ( $s = m = 1$ ) coincide con los momentos del estadístico de Nyblom y Mäkeläinen (1983) y para  $s = 1$  y para  $m > 1$  con los del estadístico multivariante de no invertibilidad derivado en el capítulo anterior.

#### 4.3.2. Tendencia lineal estacional

Incluyendo una deriva estacional en (4.6) se tiene el modelo

$$\mathbf{y}_t = \boldsymbol{\mu}_t + \mathbf{u}_t, \quad \boldsymbol{\mu}_t = \boldsymbol{\mu}_{t-s} + \boldsymbol{\beta}_{t-s} + \mathbf{v}_t, \quad \boldsymbol{\beta}_t = \boldsymbol{\beta}_{t-s}, \quad (4.9)$$

con  $\mathbf{u}_t \sim N(\mathbf{0}, \boldsymbol{\Sigma}_u)$  y  $\mathbf{v}_t \sim N(\mathbf{0}, \rho\boldsymbol{\Sigma}_u)$ , que en forma matricial se puede expresar como

$$\begin{aligned} \mathbf{Y} &= \mathbf{A} + \mathbf{U}, \\ (\nabla_n \otimes \mathbf{I}_s)\mathbf{A} &= (\mathbf{d} \otimes \mathbf{I}_s)\mathbf{A}_0 + (\mathbf{i}_n \otimes \mathbf{I}_s)\mathbf{B}_0 + \mathbf{V}, \end{aligned}$$

donde  $\mathbf{B}_0 = [\beta_{-s+1}, \dots, \beta_0]'$  es una matriz fija  $s \times m$  de condiciones iniciales para  $\beta_t$ . Si  $\mathbf{A}_0$  es fija, (4.9) se puede expresar como un caso especial de (4.1) con  $\mathbf{X} = [\mathbf{i}_n \otimes \mathbf{I}_s, \mathbf{t}_n \otimes \mathbf{I}_s]$ , donde  $\mathbf{i}_n$  es un vector  $n$ -dimensional de unos y  $\mathbf{t}_n = (1, 2, \dots, n)'$ ,  $\boldsymbol{\Pi} = [\mathbf{A}_0, \mathbf{B}_0]$  y el término de error  $\mathbf{E} = \mathbf{U} + (\mathbf{C}_n \otimes \mathbf{I}_s)\mathbf{V}$ , de forma que su matriz de covarianzas será  $\boldsymbol{\Omega}(\rho) = \mathbf{I}_T + \rho(\mathbf{C}_n \mathbf{C}'_n \otimes \mathbf{I}_s)$ . Es claro que la inclusión de una deriva en  $\boldsymbol{\mu}_t$  solamente afecta a la matriz de medias de  $\mathbf{Y}$ , pero no a su matriz de covarianzas. Por lo tanto, es estadístico LBI para contrastar la hipótesis nula de nivel estacional determinista bajo la presencia de tendencia lineal estacional determinista ( $H_0 : \rho = 0$ ), frente a la alternativa de paseo aleatorio estacional con deriva ( $H_1 : \rho > 0$ ),  $DRW_{m,s,n}$ , se puede calcular como (4.8), donde  $\hat{\mathbf{u}}_t$  son ahora los residuos de mínimos cuadrados ordinarios de  $\mathbf{y}_t$  sobre  $s$  variables ficticias estacionales y  $s$  tendencias lineales estacionales. En este caso las expresiones de  $\bar{\lambda}_1$  y  $\bar{\lambda}_2$  para el estadístico  $DRW_{m,s,n}/(n-1)m$  son

$$\bar{\lambda}_1 = \frac{n+2}{15}, \quad \bar{\lambda}_2 = \bar{\lambda}_1 \left( \frac{11n^2 + 181}{840} \right),$$

por lo que  $DRW_{m,s,n}/nm$  tiene media y varianza asintótica  $1/15$  y  $11/6300ms$  respectivamente. Nótese que para  $s = m = 1$  el estadístico coincide con el derivado por Nyblom (1986) para el mismo problema en el caso univariante y los momentos asintóticos coinciden, si bien aquí se han derivado de una forma más sencilla.

### 4.3.3. Tendencia lineal regular

Una extensión adicional del modelo (4.6) consiste en incluir una tendencia lineal en la parte regular del modelo,

$$\mathbf{y}_t = \boldsymbol{\mu}_t + \beta t + \mathbf{u}_t, \quad \boldsymbol{\mu}_t = \boldsymbol{\mu}_{t-s} + \mathbf{v}_t, \quad (4.10)$$

con  $\mathbf{u}_t \sim N(\mathbf{0}, \boldsymbol{\Sigma}_u)$  y  $\mathbf{v}_t \sim N(\mathbf{0}, \rho\boldsymbol{\Sigma}_u)$ , o en forma matricial,

$$\mathbf{Y} = \boldsymbol{\alpha} + \beta \mathbf{t}_T + \mathbf{U}, \quad (\nabla_n \otimes \mathbf{I}_s)\mathbf{A} = (\mathbf{d}_n \otimes \mathbf{I}_s)\mathbf{A}_0 + \mathbf{V},$$

que es un caso especial de (4.1) con  $\mathbf{X} = [(\mathbf{i}_n \otimes \mathbf{I}_k), \mathbf{t}_T]$ ,  $\mathbf{t}_T = (1, 2, \dots, T)'$ ,  $\mathbf{\Pi} = [\mathbf{A}_0, \boldsymbol{\beta}]$ ,  $\mathbf{E} = \mathbf{U} + \mathbf{C}_T \mathbf{V}$ , luego  $\boldsymbol{\Omega}(\rho) = \mathbf{I}_T + \rho \mathbf{C}_T \mathbf{C}'_T$  donde  $\mathbf{C}_T$  es una matriz triangular inferior de unos de orden  $T$ . El estadístico del contraste LBI para el problema  $H_0 : \rho = 0$  frente a  $H_1 : \rho > 0$  en (4.10), que se denotará por  $TRW_{m,s,n}$ , y que se puede calcular como

$$TRW_{m,s,n} = \text{tr} \left\{ \left( \sum_{t=1}^T \hat{\mathbf{u}}_t \hat{\mathbf{u}}_t' \right)^{-1} \left[ \sum_{t=1}^T \left( \sum_{\tau=1}^t \hat{\mathbf{u}}_t \right) \left( \sum_{\tau=1}^t \hat{\mathbf{u}}_t \right)' \right] \right\},$$

donde  $\hat{\mathbf{u}}_t$  es ahora el vector de residuos de la regresión de mínimos cuadrados multivariantes de  $\mathbf{y}_t$  sobre  $s$  variables ficticias estacionales y una tendencia regular. Las expresiones de la media y la media cuadrática de los autovalores de  $\mathbf{MK}$  en este caso tienen las siguientes expresiones:

$$\begin{aligned} \bar{\lambda}_1 &= \frac{(5s-3)n^2 - (5s+3)}{30(T-s-1)}, \\ \bar{\lambda}_2 &= \frac{(140s-129)n^4 + (350s-213)n^2 - (490s+234)}{12600(T-s-1)}. \end{aligned}$$

Nótese que  $TRW_{1,1,n} = DRW_{1,1,n}$  es el estadístico derivado por Nyblom (1986) para contrastar la hipótesis nula de tendencia lineal determinista en el modelo univariante, y que la media y varianza asintóticas de  $TRW_{1,1,n}/n$  son  $1/15$  y  $11/6300$ , que coinciden con las derivadas por Nyblom (1986) en una prueba bastante complicada.

#### 4.3.4. Modelos VIMA(1,1)<sub>s</sub> estacionales

Al igual que en el capítulo anterior, todos los modelos estructurales admiten una forma reducida ARIMA, de forma que los contrastes de estabilidad paramétrica se pueden interpretar también como contrastes de raíces unitarias MA. Así, el modelo de nivel local estacional (4.6) se puede expresar como un proceso IMA(1,1)<sub>s</sub> vectorial

$$(\mathbf{I}_m - \mathbf{I}_m B^s) \mathbf{y}_t = (\mathbf{I}_m - \theta \mathbf{I}_m B^s) \mathbf{a}_t, \quad \mathbf{a}_t \sim N(\mathbf{0}, \boldsymbol{\Omega}_a), \quad t = s+1, \dots, T, \quad (4.11)$$

donde  $\mathbf{a}_t = (a_{1t}, \dots, a_{mt})'$ ,  $\boldsymbol{\Omega}_a$  es una matriz  $m \times m$  definida positiva y el parámetro  $\theta$  está restringido a ser positivo, de forma que  $\rho = (1-\theta)^2/\theta \geq 0$ . Al igual que en el modelo regular ( $s = 1$ ), el proceso (4.11) es invertible si  $\theta < 1$  y estrictamente no invertible cuando  $\theta = 1$ . En este último caso, la cancelación de los polinomios matriciales a ambos lados de la igualdad revelan la presencia de estacionalidad determinista. Teniendo en

cuenta que  $\rho(\theta) = \rho(1/\theta)$ , el problema de contraste de una cola  $H_0 : \rho = 0$  frente a  $H_1 : \rho > 0$  es equivalente al problema de contraste de dos colas  $H_0 : \theta = 1$  frente a  $H_0 : \theta \neq 1$ . Por lo tanto,  $RW_{m,s,n}$  se puede interpretar también como el estadístico LBIU para el contraste de no invertibilidad estricta en (4.11). Una forma alternativa de probarlo se obtiene escribiendo el modelo en forma matricial como

$$(\nabla_n \otimes \mathbf{I}_s)\mathbf{Y} = (\mathbf{d}_n \otimes \mathbf{I}_s)(\mathbf{Y}_0 - \theta\bar{\mathbf{A}}_0) + (\nabla_n(\theta) \otimes \mathbf{I}_s)\bar{\mathbf{A}},$$

donde  $\bar{\mathbf{A}} = (\mathbf{a}_1, \dots, \mathbf{a}_T)'$  y  $\mathbf{Y}_0 - \theta\bar{\mathbf{A}}_0 = (\mathbf{y}_{-s+1} - \theta\mathbf{a}_{-s+1}, \dots, \mathbf{y}_0 - \theta\mathbf{a}_0)'$  es una matriz  $s \times m$  de condiciones iniciales. Por lo tanto, (4.11) es un caso especial del modelo de regresión multivariante (4.1) con

$$\mathbf{X} = \mathbf{i}_n \otimes \mathbf{I}_s, \quad \mathbf{\Pi} = \mathbf{Y}_0 - \theta\bar{\mathbf{A}}_0, \quad \mathbf{\Omega}(\theta) = \mathbf{C}_n \nabla_n(\theta) \nabla_n(\theta)' \mathbf{C}_n' \otimes \mathbf{I}_s,$$

y  $d^2\mathbf{\Omega}/d\theta^2|_{\theta=1} = 2\mathbf{C}_n\mathbf{C}_n' \otimes \mathbf{I}_s$  dadas las expresiones de las derivadas de  $\nabla(\theta)\nabla(\theta)'$  con respecto a  $\theta$  obtenidas en el capítulo anterior. De esta forma,  $RW_{m,s,n}$  se puede interpretar como la extensión multivariante del estadístico de Tam y Reinsel (1997) para contrastar la presencia de una raíz unitaria  $MA$ .

De forma análoga,  $DRW_{m,s,n}$  y  $TRW_{m,s,n}$  son los estadísticos LBIU para contrastar  $H_0 : \theta = 1$  frente a  $H_1 : \theta \neq 1$  en las formas reducidas de (4.9),

$$(\mathbf{I}_m - \mathbf{I}_m B^s)^2 \mathbf{y}_t = (\mathbf{I}_m - \mathbf{I}_m B^s)(\mathbf{I}_m - \theta \mathbf{I}_m B^s) \mathbf{a}_t,$$

y (4.10),

$$(\mathbf{I}_m - \mathbf{I}_m B)(\mathbf{I}_m - \mathbf{I}_m B^s) \mathbf{y}_t = (\mathbf{I}_m - \mathbf{I}_m B)(\mathbf{I}_m - \theta \mathbf{I}_m B^s) \mathbf{a}_t,$$

respectivamente. Nótese que  $DRW_{1,s,m}$  está relacionado con el estadístico de Taylor (2003a), mientras que  $TRW_{1,s,m}$  es el estadístico de Tam y Reinsel (1998) de raíz unitaria estacional frente a una raíz en la parte regular.

## 4.4. Modelos dinámicos

### 4.4.1. Ciclos estacionales deterministas

El polinomio estacional  $(1 - \theta B^s)$  contiene  $s$  raíces (en adelante se considera  $s$  par), de forma que al contrastar la hipótesis nula  $H_0 : \theta = 1$  se está contrastando que todas

ellas sean iguales a uno, lo cual es una crítica habitual a los contrastes estacionales del tipo Tam y Reinsel (1997) o Dickey et al. (1984). Nótese que el polinomio diferencia estacional se puede descomponer como

$$(1 - B^s) = (1 - B)(1 + B) \prod_{k=1}^{s/2-1} (1 - e^{i\omega_k} B)(1 - e^{-i\omega_k} B), \quad \omega_k = \frac{2\pi k}{s},$$

donde  $i = \sqrt{-1}$ , de forma que el polinomio  $(1 - B^s)$  tiene dos raíces reales, 1 y  $-1$ , y  $s/2 - 1$  pares de raíces complejas conjugadas  $(e^{i\omega_k}, e^{-i\omega_k})$ . De la ecuación de Euler,  $e^{i\omega_k} = c_k + is_k$  donde  $c_k = \cos(\omega_k)$  y  $s_k = \sin(\omega_k)$ , de forma que  $(1 - e^{i\omega_k} B)(1 - e^{-i\omega_k} B) = (1 - 2c_k B + B^2)$ , con lo que el par de raíces complejas  $(e^{i\omega_k}, e^{-i\omega_k})$  implica la presencia de un ciclo estacional de frecuencia  $\omega_k$ . Por su parte, la raíz  $-1$  indica un ciclo estacional de frecuencia  $k = s/2$ , y da lugar al polinomio  $(1 - c_{s/2} B)$ . La diferenciación estacional  $(1 - B^s)$  implica, por tanto, la eliminación de los ciclos estacionales en todas las frecuencias, de forma que si no existiera un ciclo en una determinada frecuencia estacional, la serie estaría sobrediferenciada en esta frecuencia. Para detectar este tipo de sobrediferenciaciones se puede extender el procedimiento de contraste anterior descomponiendo el filtro media móvil estacional como

$$(1 - \theta B^s) = (1 - \theta_0 B)(1 + \theta_{s/2} B) \prod_{k=1}^{s/2-1} (1 - 2\sqrt{\theta_k} c_k B + \theta_k B^2).$$

donde cada factor controla la dinámica de cada ciclo estacional. El modelo  $VIMA(1, 1)_s$  admite, por tanto, la siguiente representación alternativa

$$(1 - B)(1 + B) \prod_{k=1}^{s/2-1} (1 - 2c_k B + B^2) \mathbf{y}_t = (1 - \theta_0 B)(1 + \theta_{s/2} B) \prod_{k=1}^{s/2-1} (1 - 2\sqrt{\theta_k} c_k B + \theta_k B^2) \mathbf{a}_t,$$

de forma que, como en casos anteriores, si  $\theta_k = 1$ , la cancelación con el operador  $(1 - 2c_k B + B^2)$  implica la presencia de un ciclo estacional determinista en la frecuencia  $\omega_k$ . Si las  $s$  raíces del polinomio  $(1 - \theta B^s)$  tienen módulo igual a la unidad, el proceso  $\mathbf{y}_t$  se puede expresar como la suma de  $s/2$  ciclos estacionales (ver Abraham y Box, 1978)

$$\mathbf{y}_t = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_{s/2} (-1)^t + \sum_{k=1}^{s/2-1} [\boldsymbol{\beta}_{1k} c_{kt} + \boldsymbol{\beta}_{2k} s_{kt}] + \mathbf{a}_t, \quad (4.12)$$

donde  $c_{kt} = \cos(2\pi kt/s)$  y  $s_{kt} = \sin(2\pi kt/s)$  que serán deterministas cuando lo sean los coeficientes  $\beta_{1k}$  y  $\beta_{1k}$  (ver, p. ej. Harvey, 1989). De igual forma, suponiendo en la descomposición del filtro MA que todos los parámetros menos  $\theta_0$  son iguales, se puede expresar este polinomio como  $(1 - \theta_0 B)(1 + \theta_1 B + \dots + \theta_1^{s-1} B^{s-1})$ , donde el segundo factor es la suma anual que recoge la dinámica estacional y el primero la de la parte regular. Esta última expresión es útil para derivar el contraste en situaciones habituales en la práctica. Por ejemplo, permite contrastar la hipótesis de estacionalidad determinista ( $\theta_1 = 1$ ) cuando se sabe que el nivel es determinista ( $\theta_0 = 1$ ), o su problema dual de contraste de nivel determinista en un modelo con estacionalidad determinista. En los siguientes apartados se muestra los contrastes que permiten detectar la presencia de ciclos estacionales deterministas en determinadas frecuencias o de estacionalidad frente en presencia de nivel determinista.

Nótese por último que si sólo una de las raíces de  $(1 - \theta B^s)$  tiene módulo unitario, por ejemplo en la frecuencia  $k \neq \{0, \pi/2\}$ , entonces la serie se puede expresar como  $\nabla_s^* \mathbf{y}_t = \beta_{1k} c_{kt} + \beta_{2k} s_{kt} + \theta_s^*(B) \mathbf{a}_t$  donde  $\nabla_s^*$  es el operador diferencia para todas las frecuencias distintas de  $k$  y  $\theta_s^*(B)$  es el operador media móvil para todas las frecuencias distintas de  $k$ . El contraste de estacionalidad determinista en esa frecuencia se puede realizar utilizando el derivado en esta sección, y la correlación serial que implica la presencia de estacionalidad estocástica en el resto de frecuencias podría corregirse utilizando la corrección paramétrica del estadístico de no invertibilidad.

#### 4.4.2. Ficticias estacionales

El modelo (4.12) se puede expresar en términos de  $s$  variables ficticias estacionales:  $\mathbf{y}_t = \sum_{j=1}^s \beta_j d_{jt} + \mathbf{a}_t$ , donde  $d_{jt} = 1$  si el vector de series temporales  $\mathbf{y}_t$  pertenece a la estación  $j$  y cero en otro caso. Para contrastar la presencia de estacionalidad determinista en la estación  $j$  se considera el modelo

$$\begin{aligned} \mathbf{y}_t &= \sum_{j=1}^s \beta_{jt} d_{jt} + \mathbf{u}_t, \\ \beta_{j,t} &= \beta_{j,t-1} + \delta_j \mathbf{v}_{jt}, \quad j = 1, \dots, s, \end{aligned} \tag{4.13}$$

donde  $\mathbf{u}_t \sim N(\mathbf{0}, \Sigma_u)$  y  $\mathbf{v}_{jt} \sim N(\mathbf{0}, \rho_j \Sigma_u)$  son dos errores mutua y serialmente independientes. El parámetro  $\delta_j$  toma sólo los valores 0 y 1, de forma que el vector  $\beta_{jt}$  es constante si  $\delta_j = 0$ , indicando así que la media de la  $j$ -ésima estación es determinista. Si  $\sum_{j=1}^s \delta_j = r \leq s$ ,  $r$  de las  $s$  estaciones son potencialmente estocásticas, y el estadístico contraste de la hipótesis nula de estacionalidad determinista  $H_0 : \rho_{j_1} = \dots = \rho_{j_r} = \rho' = 0$

frente a la alternativa de estacionalidad estocástica en  $r$  de las estaciones,  $H_1 : \rho_{j_1} = \dots = \rho_{j_r} = \rho' > 0$ , se puede derivar siguiendo el procedimiento expuesto en apartados anteriores.

El modelo (4.13) se puede expresar como un modelo de regresión multivariante (4.1) donde  $\mathbf{X} = [\mathbf{d}_1, \dots, \mathbf{d}_s]$ ,  $\mathbf{d}_j = (d_{j1}, \dots, d_{jT})'$ ,  $\mathbf{\Pi} = [\boldsymbol{\beta}_{1,0}, \dots, \boldsymbol{\beta}_{s,0}]'$  y  $\boldsymbol{\Omega}_E(\rho') = \mathbf{I}_T + \rho' \sum_{j=1}^s \delta_j \mathbf{A}_j$ , siendo  $\mathbf{A}_j = \mathbf{d}_j \odot \mathbf{C}_T \mathbf{C}'_T \odot \mathbf{d}'_j$  y  $\odot$  el producto de Hadamard o producto elemento a elemento. A efectos de comparación conviene definir  $\rho' = \rho/s$  de forma que el estadístico LBI para el problema de contraste planteado,  $SD_{m,s,n}(r)$ , viene dado por (4.3) con  $\mathbf{K} = \sum_{j=1}^s \delta_j \mathbf{A}_j/s$ , que coincide con (4.8) cuando  $r = s$ . Teniendo en cuenta que  $\mathbf{M}\mathbf{A}_i\mathbf{M}\mathbf{A}_j = \mathbf{0}$  para  $i \neq j$ , la expresión del autovalor medio de  $\mathbf{M}\mathbf{K}$  y su media cuadrática es

$$\bar{\lambda}_1 = \frac{(n+1)r}{6s}, \quad \bar{\lambda}_2 = \bar{\lambda}_1 \left[ \frac{2n^2 + 7}{30} \right],$$

lo que permite calcular los momentos de  $SD_{m,s,n}(r)/nm$ .

#### 4.4.3. Estacionalidad determinista en una frecuencia

De forma análoga, este mismo problema se puede tratar utilizando la representación trigonométrica estacional de (4.12), de forma que se pueda contrastar la presencia de un ciclo estacional en  $\mathbf{y}_t$ . Sea el modelo

$$\mathbf{y}_t = \sum_{j=1}^s x_{jt} \boldsymbol{\beta}_{jt} + \mathbf{u}_t, \quad \mathbf{u}_t \sim N(\mathbf{0}, \boldsymbol{\Sigma}), \quad (4.14)$$

$$\boldsymbol{\beta}_{jt} = \boldsymbol{\beta}_{j,t-1} + \delta_j \mathbf{v}_{jt}, \quad \mathbf{v}_{jt} \sim N(\mathbf{0}, \rho_j \boldsymbol{\Sigma}), \quad t = 1, \dots, T,$$

donde  $x_{jt}$  son variables trigonométricas estacionales ( $x_{1t} = 1$ ,  $x_{jt} = \cos(\pi jt/s)$  si  $j$  es par y  $x_{jt} = \sin(\pi jt/s)$  si  $j$  es impar y  $j > 1$ ). A efectos de comparación, conviene suponer que  $\rho = a_j \rho/s$  donde  $a_j = 1$  si  $j = 1, s$  y  $a_j = 2$  si  $j = 2, \dots, s/2 - 1$ . A diferencia del caso anterior, este marco permite contrastar la naturaleza determinista o estocástica del nivel local  $\boldsymbol{\beta}_{1t}$ , un ciclo estacional en la  $(j/2)$ -ésima frecuencia armónica  $\boldsymbol{\beta}_{jt} \cos(\pi jt/s) + \boldsymbol{\beta}_{j+1,t} \sin(\pi jt/s)$  ( $j = 2, 4, \dots, s$ ) o una combinación de varias frecuencias. Si, por ejemplo se desea contrastar la presencia de un ciclo determinista en la  $(j/2)$ -ésima frecuencia armónica,  $H_0 : \rho_j = \rho_{j+1} = \rho = 0$  frente a  $H_1 : \rho_j = \rho_{j+1} = \rho > 0$ , (4.14) se puede escribir como un modelo de regresión multivariante donde  $\mathbf{X} = [\mathbf{c}_0, \mathbf{c}_1, \mathbf{s}_1, \dots, \mathbf{c}_{s/2}]$ ,  $\mathbf{c}_k = (c_{k1}, \dots, c_{kT})'$  con  $c_{kt} = \cos(2\pi kt/s)$ ,  $\mathbf{s}_k = (s_{k1}, \dots, s_{kT})'$  con  $s_{kt} = \sin(2\pi kt/s)$ ,  $\mathbf{\Pi} = [\boldsymbol{\beta}_0, \boldsymbol{\beta}_{1,1}, \boldsymbol{\beta}_{2,1}, \dots, \boldsymbol{\beta}_{1,s/2}]$  y  $\boldsymbol{\Omega}_E = \mathbf{I}_T + a_j \rho (\mathbf{C}_k \mathbf{C}'_k + \mathbf{S}_k \mathbf{S}'_k)/s$  siendo  $\mathbf{C}_k = \mathbf{c}_k \odot \mathbf{C}_T$  y  $\mathbf{S}_k = \mathbf{s}_k \odot \mathbf{C}_T$ , de forma que el estadístico LBI para este problema,  $TV_{m,s,n}(j)$ , viene

dado por (4.3) donde  $\mathbf{K} = a_j(\mathbf{A}_j + \mathbf{A}_{j+1})/s^2$  con  $\mathbf{A}_j = \mathbf{C}_k \mathbf{C}'_k$  y  $\mathbf{A}_{j+1} = \mathbf{S}_k \mathbf{S}'_k$ . Este estadístico se corresponde con el derivado por Buseti (2006). Los valores de  $\bar{\lambda}_1$  y  $\bar{\lambda}_2$  para el estadístico de contraste basado en esta frecuencia frecuencia son

$$\begin{aligned}\bar{\lambda}_1 &= a_j \frac{n+1}{6s}, \\ \bar{\lambda}_2 &= \bar{\lambda}_1 \left[ \frac{(4n^2-1)s^2+15}{60s^2} + b_j \frac{3}{4s^2 \sin^2(j\pi/s)} \right], \quad b_j = \begin{cases} 0 & \text{si } j = \{0, s/2\} \\ 1 & \text{si } j \in (0, s/2) \end{cases}\end{aligned}$$

Estas expresiones permiten obtener la media y la varianza del estadístico del contraste para detectar la presencia de un nivel determinista en presencia de estacionalidad determinista,  $H_0 : \rho_1 = 0$  frente a  $H_1 : \rho_1 > 0$ , que está relacionado con la extensión del estadístico de Kwiatkowski et al. (1992) con variables ficticias estacionales de Phillips y Jin (2002). Caner (1998) y Buseti y Harvey (2003) derivaron el estadístico de contraste LBI,  $TB_{1,s,m}$  para el problema dual de estacionalidad determinista en presencia de un nivel determinista,  $H_0 : \rho_2 = \dots = \rho_k = \rho = 0$  frente a  $H_1 : \rho_2 = \dots = \rho_k = \rho > 0$ , generalizado al caso multivariante por Buseti (2006). Este estadístico se puede calcular como (4.3) con  $\mathbf{K} = \sum_{j=2}^s a_j \mathbf{A}_j/s^2$ , siendo el autovalor medio y el autovalor medio cuadrático de  $\mathbf{MK}$

$$\bar{\lambda}_1 = \frac{(n+1)(s-1)}{6s}, \quad \bar{\lambda}_2 = \bar{\lambda}_1 \left[ \frac{2n^2+7}{30} - \frac{(s-1)}{4s^2} \right].$$

El estadístico  $TB_{1,s,m}/Tm$  sigue asintóticamente una distribución  $CvM((s-1)m)/(s-1)m$ , dada la aditividad de la distribución Cramér-von Mises.

## 4.5. La aproximación IG

A partir de (3.13), y sabiendo que la función característica de una variable aleatoria chi cuadrado con  $m$  grados de libertad es  $\phi_\chi(t) = (1-2it)^{-m/2}$ , se puede obtener la expresión de la función característica de la distribución límite de  $RW_{m,1,n}/(n-1)m$  bajo  $H_0$  como

$$\phi_R W(t) = \left[ \prod_{k=1}^{\infty} \left( 1 - \frac{2it}{\pi^2 k^2} \right) \right]^{-m/2} = \left[ \frac{\sqrt{2it}}{\sin \sqrt{2it}} \right]^{m/2}$$

donde la segunda expresión se obtiene dado que  $x \prod_{k=1}^{\infty} [1 - (x^2/(\pi k)^2)] = \sin(x)$ . Nótese que para  $m = 1$  esta expresión coincide con la función característica del estadístico de bondad de ajuste de Cramér-von Mises caracterizado por Anderson y Darling (1952),

Cuadro 4.1: Valores críticos y aproximación IG para  $TV_{1,s,n}/(n-1)$ 

s	n-1	$\alpha = 0,5$		$\alpha = 0,2$		$\alpha = 0,1$		$\alpha = 0,05$		$\alpha = 0,01$	
		$c_\alpha$	$\hat{\alpha}_\alpha$	$c_\alpha$	$\hat{\alpha}_\alpha$	$c_\alpha$	$\hat{\alpha}_\alpha$	$c_\alpha$	$\hat{\alpha}_\alpha$	$c_\alpha$	$\hat{\alpha}_\alpha$
4	10	0.135	50.25	0.201	19.25	0.244	9.51	0.284	4.83	0.369	1.14
	20	0.123	50.22	0.185	19.50	0.227	9.66	0.268	4.88	0.355	1.08
	30	0.119	50.19	0.180	19.58	0.222	9.71	0.262	4.90	0.351	1.06
	$\infty$	<i>0.111</i>		<i>0.168</i>		<i>0.209</i>		<i>0.249</i>		<i>0.342</i>	
12	10	0.178	47.87	0.218	16.64	0.242	12.23	0.263	6.50	0.306	1.57
	20	0.163	49.86	0.202	19.75	0.225	9.89	0.246	5.01	0.289	1.06
	30	0.158	49.84	0.196	19.76	0.219	9.93	0.240	5.03	0.283	1.05
	$\infty$	<i>0.147</i>		<i>0.185</i>		<i>0.207</i>		<i>0.228</i>		<i>0.273</i>	

En cursiva aparecen los valores críticos asintóticos calculados utilizando la aproximación *IG*.

razón por la cual la distribución límite del estadístico  $RW_{m,1,n}/(n-1)m$  es a menudo conocida como distribución Cramér-von Mises con grados de libertad igual a la dimensión del vector de variables,  $CvM(m)$ . Brown (1982) derivó los cuatro primeros cumulantes de la distribución  $CvM(m)$ , cuyas expresiones son

$$\kappa_1 = \frac{m}{6}, \quad \kappa_2 = \frac{m}{45}, \quad \kappa_3 = \frac{8m}{945}, \quad \kappa_4 = \frac{8m}{1575}$$

de forma que la distribución del estadístico, además de tener soporte positivo, es asimétrica hacia la derecha y leptocúrtica. De entre todas las distribuciones con estas características y con soporte positivo, la distribución Inversa Gaussiana con parámetros  $\mu$  y  $\lambda$ ,  $IG(\mu, \lambda)$  (Tweedie, 1957a,b; Johnson y Kotz, 1970) proporciona un buen ajuste para la distribución límite del estadístico de contraste. Las funciones de densidad y distribución de la distribución *IG* son

$$f(x; \mu, \lambda) = \sqrt{\frac{\lambda}{2\pi x^3}} \exp\left[-\frac{\lambda(x-\mu)^2}{2\mu^2 x}\right], \quad x > 0, \quad (4.15)$$

$$F(x; \mu, \lambda) = \Phi\left[\sqrt{\frac{\lambda}{x}}\left(\frac{x}{\mu} - 1\right)\right] + e^{2\lambda/\mu} \Phi\left[-\sqrt{\frac{\lambda}{x}}\left(\frac{x}{\mu} + 1\right)\right], \quad x > 0, \quad (4.16)$$

donde  $\Phi(\cdot)$  denota la función de distribución acumulada Normal estándar. Sus cuatro primeros cumulantes, expresados en términos de los parámetros, tiene las siguientes expresiones

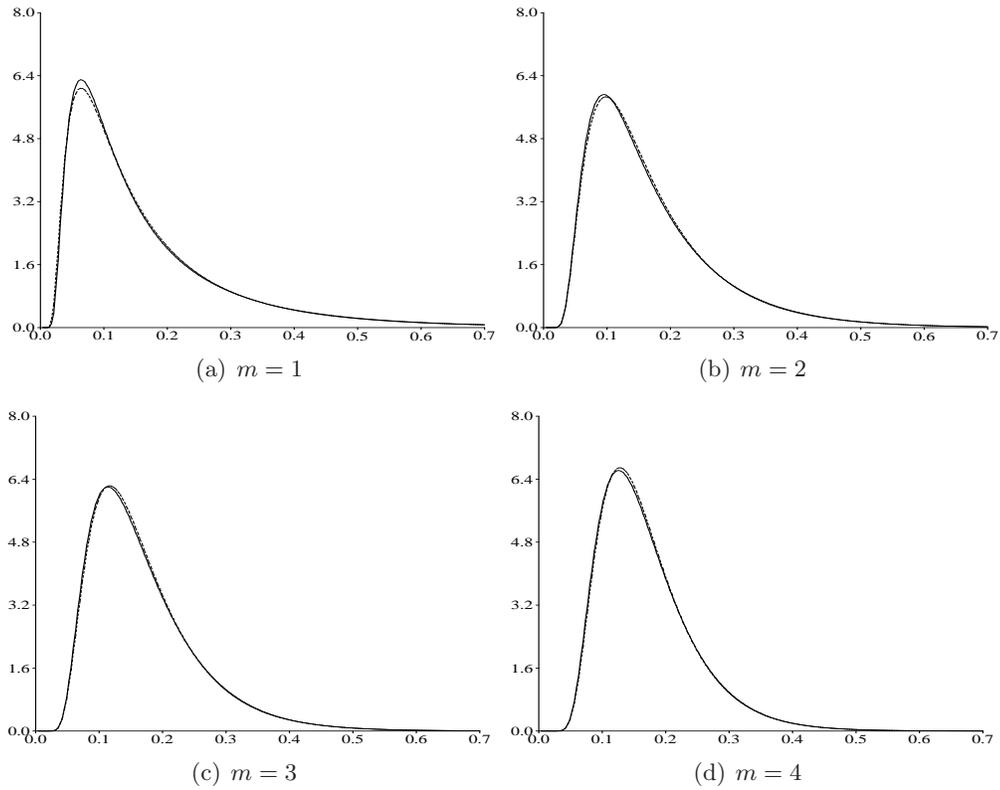
$$\kappa_1 = \mu, \quad \kappa_2 = \frac{\mu^3}{\lambda}, \quad \kappa_3 = 3\frac{\mu^5}{\lambda^2}, \quad \kappa_4 = \frac{15\mu^7}{\lambda^3}$$

igualando los dos primeros cumulantes se obtienen los parámetros de la distribución IG como  $\mu = m/6$  y  $\mu^3/\lambda = m/45$ , siendo el tercer y cuarto cumulante de la distribución  $IG(m/6, 45m^2/6^3)$  iguales a  $\kappa_3 = 8m/900$  y  $\kappa_4 = 8m/1350$ , similares a los de la distribución  $CvM(m)$ . Por lo tanto, dado que la distribución asintótica del estadístico (3.9) es  $CvM(m)/m$ , podríamos aproximarla utilizando la distribución IG.

La figura (4.1) muestra la comparación entre la distribución  $CvM(m)/m$  evaluada siguiendo el procedimiento de Imhof (1961) y la distribución  $IG(\mu, \lambda)$  con (4.15) para  $m = \{1, 2, 3, 4\}$ , de forma que  $\mu = 1/6$  y  $\lambda = 45m/6^3$ . Como se puede observar, la distribución IG proporciona un buen ajuste en todo el soporte de la distribución, lo que sugiere utilizar (4.16) para calcular los p-valores asintóticos del estadístico de contraste también en muestras finitas. Teniendo en cuenta que los parámetros de la distribución IG se pueden obtener a partir de sus dos primeros momentos como  $E(IG) = \mu$  y  $V(IG) = \mu^3/\lambda$ , igualándolos a los momentos derivados para  $RW_{m,s,n}/(n-1)m$  se pueden obtener los parámetros  $\mu$  y  $\lambda$  para aproximar la distribución en muestras finitas del estadístico utilizando (4.16). La Tabla 4.3 muestra los percentiles de para distintos valores de  $m$ ,  $s$  y  $n$  calculados por simulación para  $m > 1$  y utilizando el procedimiento de Imhof (1961) en los casos univariantes. Cada percentil lleva asociado el p-valor correspondiente calculado utilizando la aproximación IG. Como se muestra en la tabla, en general los p-valores aproximados están bastante próximos a su valor nominal incluso en muestras pequeñas, siendo la desviación absoluta media de los errores menor de 0.003 puntos porcentuales. Estas desviaciones pueden considerarse no relevantes en aplicaciones prácticas, por lo que que esta aproximación a la distribución en muestras finitas de  $RW_{m,1,n}/nm$  funciona bien. Este procedimiento se puede extender a cada uno de los contrastes derivados en este capítulo. Resultados similares se encuentran para los estadísticos  $DRW_{m,s,n}/(n-1)m$  y  $TV_{m,s,n}/(n-1)m$  y se muestran en las tablas 4.4 y 4.1 respectivamente.

## 4.6. Aplicación

Como ilustración del procedimiento de contraste de estacionalidad determinista, se aplican los contrastes estudiados a los datos mensuales de casas comenzadas ( $y_1$ ) y vendidas ( $y_2$ ) en EE.UU. entre enero de 1965 y diciembre de 1974 (Hillmer y Tiao, 1979). Reinsel (1997) propone ajustar el modelo  $(\mathbf{I}_2 - \Phi B)(1 - B^{12})\mathbf{y}_t = (\mathbf{I}_2 - \Theta B^{12})\mathbf{a}_t$  donde  $\mathbf{a}_t \sim N(\mathbf{0}, \Omega_a)$ . La estimación de máxima verosimilitud exacta proporciona las siguientes

Figura 4.1: IG approximation to the asymptotic pdf of  $RW_{m,1,n}/nm$ 

estimaciones de los parámetros del modelo (entre paréntesis figuran los errores estándar).

$$\hat{\Phi} = \begin{pmatrix} 0,4531 & 0,9551 \\ (0,0966) & (0,1795) \\ 0,1368 & 0,7100 \\ (0,0566) & (0,1065) \end{pmatrix} \quad \hat{\Theta} = \begin{pmatrix} 0,9569 & -0,1298 \\ (0,1109) & (0,1922) \\ 0,0142 & 1,0428 \\ (0,0699) & (0,1173) \end{pmatrix} \quad \hat{\Omega}_a = \begin{pmatrix} 29,44 \\ (7,2040 & 8,9615) \end{pmatrix}$$

Nótese que los elementos de fuera de la diagonal principal de la matriz  $\hat{\Theta}$  son estadísticamente cero, y que los elementos de la diagonal están muy próximos a uno. Resulta interesante comprobar si  $\Theta = \mathbf{I}_2$ , lo que implicaría una cancelación del polinomio media móvil y el operador diferencia estacional, indicando que el componente estacional presente en la serie es determinista. Dado que el modelo presenta correlación serial de tipo VAR(1), el contraste de la hipótesis nula  $H_0 : \Theta = \mathbf{I}_2$  frente a la alternativa  $H_1 : \Theta = \theta \mathbf{I}_2$  con  $\theta \neq 1$  se realiza utilizando la corrección paramétrica propuesta en el capítulo anterior y que consiste en calcular el estadístico utilizando los residuos exactos del modelo bajo la hipótesis nula. La segunda columna de las dos últimas filas de la tabla 4.2 muestran

Cuadro 4.2: Estadísticos de contraste

	RW	0	1	2	3	4	5	6
u1	0.1376	0.0274	0.0345	0.0073	0.0158	0.0345	0.0239	0.0095
p-val	0.8812	0.1638	0.3752	0.9947	0.8493	0.3758	0.6206	0.6290
u2	0.1440	0.0131	0.0430	0.0137	0.0115	0.0321	0.0280	0.0184
p-val	0.8360	0.4689	0.2458	0.9018	0.9484	0.4224	0.5149	0.3122
conjunto	0.1904	0.0231	0.0368	0.0111	0.0148	0.0348	0.0228	0.0122
p-val	0.4292	0.2084	0.3455	0.9948	0.9638	0.3935	0.7676	0.6025

el valor del estadístico de contraste y el p-valor asociado calculado utilizando la aproximación IG. Éste indica que existe un componente estacional determinista en ambas series. Las columnas siguientes muestran los estadísticos y p-valores para el contraste a cada una de las frecuencias estacionales. Como se observa, existen ciclos deterministas en cada frecuencia. El contraste de estacionalidad determinista también se ha calculado para cada una de las series individualmente. En este caso se han eliminado los elementos no significativos en  $\hat{\Theta}$  y se han realizado los contrastes individuales a partir del vector de residuos exactos del modelo imponiendo alternativamente un valor 1 al elemento de la matriz media móvil correspondiente a cada variable, dejando inalterado el otro. Los resultados que muestra la tabla 4.2 confirman los resultados del contraste multivariante.

Cuadro 4.3: Valores críticos y aproximación IG para  $RW_{m,s,n}/(n-1)m$ 

(m,s)	n	$c_{50}$	$\hat{\alpha}_{50}$	$c_{20}$	$\hat{\alpha}_{20}$	$c_{10}$	$\hat{\alpha}_{10}$	$c_5$	$\hat{\alpha}_5$	$c_1$	$\hat{\alpha}_1$
(1,1)	30	.130	51.17	.258	19.64	.366	9.42	.477	4.63	.735	1.03
	50	.126	50.94	.251	19.76	.358	9.56	.471	4.71	.739	1.02
	100	.119	52.18	.241	20.55	.347	10.01	.461	4.92	.743	1.00
	$\infty$	.119	50.58	.241	19.94	.346	9.81	.460	4.88	.738	1.04
(1,4)	10	.186	50.07	.259	19.22	.305	9.54	.348	4.88	.434	1.16
	20	.169	50.08	.240	19.50	.286	9.69	.329	4.91	.421	1.09
	30	.163	50.07	.233	19.58	.279	9.74	.323	4.92	.416	1.07
	$\infty$	.151	50.15	.219	19.86	.265	9.95	.308	5.04	.404	1.08
(1,12)	10	.195	49.75	.236	19.63	.261	9.86	.282	5.03	.325	1.12
	20	.178	49.83	.219	19.74	.242	9.90	.264	5.02	.308	1.07
	30	.173	49.84	.212	19.77	.236	9.92	.258	5.02	.302	1.07
	$\infty$	.161	49.86	.200	19.84	.224	9.95	.245	5.02	.290	1.05
(2,1)	30	.152	50.52	.248	19.12	.316	9.39	.379	4.86	.517	1.18
	50	.146	50.88	.240	19.88	.310	9.68	.380	4.73	.530	1.06
	100	.142	50.69	.237	19.78	.306	9.80	.374	4.95	.521	1.17
(2,4)	10	.194	49.51	.244	19.20	.273	9.64	.299	4.99	.353	1.15
	20	.176	49.77	.225	19.51	.255	9.92	.281	5.13	.336	1.17
	30	.170	49.86	.219	19.86	.249	9.96	.277	4.95	.332	1.12
(2,12)	100	.162	49.84	.211	19.80	.241	9.91	.267	5.09	.323	1.16
	10	.198	49.58	.226	19.88	.242	9.97	.256	5.05	.283	1.16
	20	.181	50.29	.209	19.73	.225	9.93	.239	5.01	.268	1.02
(3,1)	30	.162	49.50	.239	18.94	.291	9.34	.338	4.79	.434	1.21
	50	.155	50.12	.234	19.28	.286	9.62	.336	4.87	.441	1.15
	100	.151	50.00	.229	19.67	.284	9.62	.336	4.82	.453	1.00
(3,4)	10	.196	49.73	.235	19.54	.258	9.99	.278	5.15	.315	1.29
	20	.179	49.87	.218	19.92	.241	9.92	.262	5.08	.304	1.09
	30	.173	50.09	.212	19.88	.235	10.03	.256	5.04	.299	1.09
(3,12)	100	.165	50.07	.204	19.76	.227	9.94	.249	5.01	.292	1.10
	10	.198	49.98	.221	20.03	.234	10.11	.245	5.20	.266	1.15
	20	.182	49.91	.204	19.77	.217	9.98	.228	5.11	.249	1.14
(4,1)	30	.166	49.21	.233	18.51	.275	9.16	.313	4.74	.384	1.30
	50	.160	49.74	.227	19.24	.271	9.61	.310	4.96	.394	1.17
	100	.156	49.84	.225	19.24	.269	9.59	.312	4.80	.400	1.10
(4,4)	10	.197	49.82	.230	19.73	.249	10.10	.265	5.22	.298	1.17
	20	.180	49.37	.214	19.42	.233	9.80	.251	4.94	.285	1.10
	30	.174	49.61	.208	19.76	.228	9.92	.245	5.05	.281	1.05
(4,12)	100	.166	49.92	.200	19.80	.220	9.99	.237	5.06	.273	1.09
	10	.199	49.41	.219	19.48	.230	9.74	.239	4.90	.258	1.01
	20	.1820	50.21	.201	20.25	.212	10.10	.221	5.11	.240	1.00
(5,1)	30	.169	49.14	.227	18.76	.261	9.59	.292	5.02	.349	1.43
	50	.163	49.60	.222	19.23	.259	9.76	.292	5.05	.360	1.25
	100	.158	50.12	.219	19.35	.256	9.76	.292	5.07	.368	1.13

Cuadro 4.4: Valores críticos y aproximación IG para  $DRW_{m,s,n}/(n-2)m$ 

(m,s)	n	$c_{50}$	$\hat{\alpha}_{50}$	$c_{20}$	$\hat{\alpha}_{20}$	$c_{10}$	$\hat{\alpha}_{10}$	$c_5$	$\hat{\alpha}_5$	$c_1$	$\hat{\alpha}_1$
(1,1)	30	.063	51.06	.099	20.38	.127	9.91	.154	4.80	.217	.92
	50	.060	50.83	.096	20.48	.124	10.03	.152	4.88	.217	.92
	100	.058	50.65	.094	20.52	.121	10.13	.150	4.94	.218	.92
	$\infty$	.056	49.92	.093	.	.119	10.26	.148	4.98	.218	.92
(1,4)	10	.159	50.93	.225	19.72	.269	9.60	.310	4.75	.396	1.01
	20	.143	50.78	.206	19.96	.249	9.78	.289	4.81	.377	.99
	30	.138	50.71	.200	20.02	.242	9.83	.282	4.84	.371	.97
	100	.131	50.61	.191	20.10	.232	9.91	.272	4.89	.361	.97
$\infty$	.128	.	.187	.	.228	.	.267	.	.356	.	
(1,12)	10	.187	49.52	.227	19.53	.250	9.99	.271	5.07	.314	1.09
	20	.170	49.79	.209	19.63	.232	9.91	.253	4.97	.296	1.05
	30	.164	50.22	.203	19.64	.226	9.86	.247	4.93	.290	1.03
	100	.156	50.46	.194	19.99	.217	9.98	.238	4.96	.281	1.03
$\infty$	.1531	.	.190	.	.213	.	.234	.	.278	.	
(2,1)	30	.068	50.91	.094	19.74	.112	9.63	.128	4.81	.163	.99
	50	.065	51.07	.091	20.15	.109	10.02	.127	4.77	.165	.91
	100	.063	50.74	.090	20.00	.108	9.88	.125	4.78	.165	.91
(2,4)	10	.167	50.27	.212	19.76	.240	9.83	.264	4.99	.313	1.14
	20	.150	50.39	.194	19.97	.222	9.87	.246	4.98	.299	1.04
	30	.145	50.43	.189	19.63	.216	9.76	.241	4.92	.292	1.06
100	.135	50.41	.180	20.07	.207	9.86	.233	4.81	.285	.99	
(2,12)	10	.189	49.98	.216	20.12	.232	10.13	.246	5.20	.2731	1.11
	20	.173	49.48	.199	19.92	.215	10.01	.228	5.04	.256	1.03
	30	.167	49.74	.194	19.80	.209	9.99	.223	4.97	.250	1.06
	100	.159	49.93	.185	20.02	.200	10.08	.214	5.13	.241	1.06
(3,1)	30	.070	50.41	.091	19.56	.104	9.65	.116	4.87	.141	1.05
	50	.067	50.67	.088	19.97	.102	9.74	.115	4.85	.142	.98
	100	.064	50.88	.086	20.52	.100	10.12	.114	4.79	.142	.96
(3,4)	10	.170	49.97	.205	19.72	.226	9.93	.245	4.98	.283	1.12
	20	.153	50.31	.188	20.07	.209	10.13	.228	5.02	.269	.99
	30	.148	50.18	.182	19.92	.203	9.87	.223	4.91	.263	1.00
100	.140	49.82	.175	19.84	.195	9.95	.215	4.97	.255	1.00	
(3,12)	10	.190	49.86	.212	19.73	.225	9.89	.236	5.03	.257	1.04
	20	.173	49.94	.195	19.89	.207	9.99	.218	5.06	.240	1.00
	30	.168	50.05	.189	20.24	.202	9.93	.213	4.80	.234	.96
	100	.160	49.69	.181	19.89	.194	9.82	.204	4.93	.225	1.02
(4,1)	30	.071	50.31	.089	19.74	.099	9.71	.109	4.88	.128	1.10
	50	.068	49.97	.086	19.54	.098	9.78	.108	4.84	.129	.99
	100	.066	50.15	.084	20.18	.096	10.04	.107	4.94	.130	.92
(4,4)	10	.171	49.85	.201	19.60	.219	9.91	.234	5.07	.264	1.17
	20	.154	50.13	.184	19.99	.202	9.93	.218	4.92	.250	1.02
	30	.149	50.02	.178	20.19	.196	10.05	.213	4.91	.245	1.00
100	.141	50.18	.171	20.04	.188	10.07	.204	5.06	.238	.99	
(4,12)	10	.190	49.72	.210	19.74	.220	10.03	.229	5.07	.247	1.11
	20	.174	49.66	.193	19.59	.203	9.82	.212	4.94	.230	1.05
	30	.168	49.82	.186	20.20	.197	10.21	.206	5.08	.224	.98
	100	.160	50.23	.178	20.34	.189	10.29	.197	5.23	.216	.97
(5,1)	30	.072	49.68	.087	19.23	.096	9.55	.104	4.88	.120	1.16
	50	.069	49.99	.085	19.85	.094	9.86	.103	4.92	.121	1.01
	100	.066	50.35	.083	20.25	.093	9.96	.102	5.04	.121	1.05

## Capítulo 5

# Cointegración en modelos VARIMA

### 5.1. Introducción

La advertencia de Tiao y Box (1981) de que la diferenciación simultánea de un vector de series temporales no estacionarias puede conducir a un modelo más complejo cuando existen combinaciones lineales estacionarias entre ellas se usa aquí como estrategia para detectar cointegración en un marco VARIMA. La complejidad asociada a la sobrediferenciación se concreta en la presencia de una estructura VIMA(1,1) adicional no invertible que puede ser detectada usando los contrastes LBIU propuestos en el capítulo 3. Aunque no se mencione explícitamente, esta idea es la que subyace tras los contrastes de Stock y Watson (1988) y Nyblom y Harvey (2000), pero a diferencia de estos autores aquí se usa el propio modelo de interés para realizar el contraste de cointegración. Una ventaja adicional de la aproximación VARIMA es que permite acomodar también la denominada aproximación basada en el modelo de regresión. Por tanto, las tres aproximaciones señaladas por Johansen (2004) pueden reducirse a dos: la que contrasta la hipótesis nula de no-cointegración frente a la alternativa de cointegración y, la opuesta, que contrasta la hipótesis nula de cointegración frente a la alternativa de no-cointegración. Como se ha señalado en la introducción, estas son las dos aproximaciones comúnmente seguidas en la identificación del orden de integración de una serie temporal individual y que se vinculan a los trabajos pioneros de Dickey y Fuller (1979), por un lado, y Nyblom y Mäkeläinen (1983), por otro.

En este capítulo se propone un procedimiento para detectar cointegración en un modelo VARIMA general usando los contrastes de no-invertibilidad propuestos en el capítulo 3. En primer lugar, se deriva la forma VIMA de la representación basada en el modelo de regresión de un proceso cointegrado y se deriva el contraste óptimo de cointegración en

este caso. En este modelo, las matrices de parámetros cumplen una serie de restricciones que permiten identificar fácilmente las relaciones de cointegración dentro del polinomio MA. Se ilustra también cómo un sistema con tendencias comunes puede expresarse en términos de un modelo VARIMA no invertible. La potencia del contraste y la consistencia del estimador de máxima verosimilitud exacta de los parámetros del modelo se evalúan mediante un estudio Monte Carlo. En el tercer apartado se muestra cómo la presencia de raíces unitarias en un polinomio media móvil con estructura general puede implicar la presencia de relaciones de cointegración sin imponer restricciones sobre los parámetros. Como contraste de cointegración se propone el contraste básico de no invertibilidad con la corrección paramétrica derivada en el capítulo 3. La potencia del contraste en este caso se estudia en una serie de experimentos Monte Carlo y una aplicación práctica en el apartado 4 ilustra el procedimiento de análisis de cointegración.

## 5.2. La aproximación basada en el modelo de regresión

### 5.2.1. Formulación VIMA(1,1) y contraste de cointegración

Una relación de cointegración entre  $y_t$  y el vector  $\mathbf{x}_t$  de variables no estacionarias,  $y_t = \beta' \mathbf{x}_t + u_t$ , puede interpretarse como un modelo de regresión múltiple donde las variables explicativas son no estacionarias. Engle y Granger (1987) proponen estimar la relación de cointegración  $\beta$  en este modelo por mínimos cuadrados ordinarios, que proporciona un estimador superconsistente de éste (Stock, 1987), y aplicar contrastes de no estacionariedad tipo Dickey y Fuller (1979) sobre los residuos de esta regresión para contrastar la hipótesis nula de no cointegración. Por su parte, Phillips (1991) y Phillips y Loretan (1991) extienden esta representación al caso multivariante. Un vector  $m \times 1$  de series temporales  $\mathbf{y}_t = (\mathbf{y}'_{1t}, \mathbf{y}'_{2t})'$ , expresado en forma triangular tiene la siguiente estructura:

$$\mathbf{y}_{1t} = \mathbf{B}\mathbf{y}_{2t} + \mathbf{a}_{1t}, \quad \mathbf{a}_{1t} \sim N(\mathbf{0}, \mathbf{\Omega}_{11}), \quad (5.1)$$

$$\nabla \mathbf{y}_{2t} = \mathbf{a}_{2t}, \quad \mathbf{a}_{2t} \sim N(\mathbf{0}, \mathbf{\Omega}_{22}), \quad (5.2)$$

donde  $\mathbf{y}_{1t}$  y  $\mathbf{y}_{2t}$  son dos vectores de dimensiones  $m_1$  y  $m_2$  respectivamente ( $m = m_1 + m_2$ ), con errores  $\mathbf{a}_{1t}$  y  $\mathbf{a}_{2t}$ , tales que  $E(\mathbf{a}_{1t}\mathbf{a}'_{2t}) = \mathbf{\Omega}_{12} \neq \mathbf{0}$  y  $\mathbf{B}$  una matriz  $m_1 \times m_2$  de coeficientes. Tanto  $\mathbf{y}_{1t}$  como  $\mathbf{y}_{2t}$  contienen variables  $I(1)$ , con lo que, a partir de (5.1), el vector  $\mathbf{z}_t = \mathbf{y}_{1t} - \mathbf{B}\mathbf{y}_{2t}$  definirá un conjunto de  $m_1$  relaciones de cointegración siempre que el error  $\mathbf{a}_{1t}$ , que se puede interpretar como el vector de desviaciones con respecto

a las relaciones de largo plazo, sea estacionario. De esta forma, cada una de las filas de la matriz  $[\mathbf{I}_{m_1} | -\mathbf{B}]$  se interpretan como los  $m_1$  vectores de cointegración del sistema, debidamente normalizados. Esta representación es especialmente útil cuando se conoce la dirección de las relaciones de cointegración o se pretende contrastar la validez de una determinada relación de cointegración en un conjunto de datos.

Con respecto a la estimación de la matriz  $\mathbf{B}$ , se sabe que su estimador de mínimos cuadrados ordinarios en (5.1), aunque es consistente, es ineficiente dado que el vector de regresores es no estacionario y que los errores  $\mathbf{a}_{1t}$  y  $\mathbf{a}_{2t}$  pueden presentar correlación tanto contemporánea como serial. Esto introduce sesgos asintóticos en la distribución mínimo cuadrática de  $\mathbf{B}$  (Phillips, 1991) lo que complicaría la distribución límite de cualquier estadístico de contraste de cointegración basado en los residuos MCO de (5.1). Para corregir este problema, se han sugerido distintos métodos de estimación asintóticamente eficientes (Phillips y Hansen, 1990; Park, 1992; Saikkonen, 1991; Kurozumi y Hayakawa, 2009, entre otros), por lo que la distribución de los estadísticos de contraste de cointegración derivados a partir de los residuos construidos utilizando estos métodos estaría asintóticamente libre de sesgos. Alternativamente, nótese que el sistema (5.1)-(5.2) se puede expresar conjuntamente como un modelo VAR(1) en forma estructural que se podría estimar por máxima verosimilitud a partir de su forma reducida. Si existe correlación entre los errores del sistema, éste podría expresarse como un proceso VARMA, donde la estacionariedad del sistema dependería de forma crucial tanto de la partición de  $\mathbf{y}_t$ , como del valor de la matriz de relaciones de cointegración,  $\mathbf{B}$ . Por lo tanto, si se estimara el modelo usando los algoritmos iterativos de estimación habituales, tanto una mala partición del vector de variables, como la asignación de un valor inicial de  $\mathbf{B}$  muy alejado de su verdadero valor, pueden provocar fallos de convergencia en el proceso de estimación. Para evitar estos problemas, se propone modelizar conjuntamente el proceso estacionario  $\nabla \mathbf{y}_t = (\nabla \mathbf{y}'_{1t}, \nabla \mathbf{y}'_{2t})'$  sin necesidad de realizar ninguna partición. Diferenciando (5.1) y sumando y restando  $\mathbf{B}\mathbf{a}_{2,t-1}$ , el modelo (5.1)-(5.2) se puede expresar alternativamente como un proceso VIMA(1,1) no invertible

$$\begin{pmatrix} \nabla \mathbf{y}_{1t} \\ \nabla \mathbf{y}_{2t} \end{pmatrix} = \begin{pmatrix} \mathbf{a}_{1t}^* \\ \mathbf{a}_{2t}^* \end{pmatrix} - \underbrace{\begin{pmatrix} \mathbf{I}_{m_1} & -\mathbf{B} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}}_{\Theta_*} \begin{pmatrix} \mathbf{a}_{1,t-1}^* \\ \mathbf{a}_{2,t-1}^* \end{pmatrix}, \quad (5.3)$$

con  $\mathbf{a}_{1t}^* = \mathbf{a}_{1t} + \mathbf{B}\mathbf{a}_{2t}$  y  $\mathbf{a}_{2t}^* = \mathbf{a}_{2t}$ , y donde el polinomio media móvil  $\Theta_*(B) = (\mathbf{I}_m - \Theta_* B)$  contiene  $m_1$  raíces unitarias. Si no existiera cointegración, la ecuación (5.1) se podría expresar como  $\mathbf{y}_{1t} = \mathbf{B}\mathbf{y}_{2t} + \boldsymbol{\epsilon}_t$  donde el error es un proceso integrado como, por ejemplo,  $\boldsymbol{\epsilon}_t = \mathbf{a}_{1t} + b \sum_{i=1}^{t-1} \mathbf{a}_{1i}$  y  $b$  una constante arbitraria que, sin pérdida de generalidad se puede

definir como  $b = (1 - \theta)$ . De esta forma, el sistema (5.1)-(5.2) no cointegrado se puede expresar, siguiendo el procedimiento anterior, como un proceso VIMA(1,1) invertible

$$\begin{pmatrix} \nabla \mathbf{y}_{1t} \\ \nabla \mathbf{y}_{2t} \end{pmatrix} = \begin{pmatrix} \mathbf{a}_{1t}^* \\ \mathbf{a}_{2t}^* \end{pmatrix} - \begin{pmatrix} \theta \mathbf{I}_{m_1} & \boldsymbol{\Theta}_{12} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{a}_{1,t-1}^* \\ \mathbf{a}_{2,t-1}^* \end{pmatrix}, \quad t = 2, \dots, n, \quad (5.4)$$

donde  $\boldsymbol{\Theta}_{12} = -\mathbf{B}$ . Esta representación permite parametrizar el problema de contraste, de forma que el sistema estará cointegrado si  $\theta = 1$ , con lo que el problema de decisión se plantea como

$$H_0 : \theta = 1, \quad \text{frente a} \quad H_1 : \theta \neq 1,$$

Al igual que en los casos estudiados en capítulos anteriores, se puede derivar un contraste localmente óptimo, insesgado e invariante para el problema planteado, cuyo estadístico de contraste se obtiene a partir de la segunda derivada del logaritmo de la función de verosimilitud de  $\mathbf{w}_t = \nabla \mathbf{y}_t$  con respecto a  $\theta$ . En forma matricial, el modelo (5.4) se puede expresar como  $\mathbf{w} = \boldsymbol{\Theta}(\theta)\mathbf{a}$  con  $\boldsymbol{\Theta}(\theta)$  una matriz por bloques de dimensiones  $(n-1)m \times nm$  cuyo bloque  $\boldsymbol{\Theta}_{ii} = -\boldsymbol{\Theta}_*$ ,  $\boldsymbol{\Theta}_{i,i+1} = \mathbf{I}_m$  y  $\boldsymbol{\Theta}_{ij} = \mathbf{0}$  en otro caso. El vector de perturbaciones  $\mathbf{a} = (\mathbf{a}_1^*, \dots, \mathbf{a}_n^*)' \sim N(\mathbf{0}, \mathbf{I}_n \otimes \boldsymbol{\Omega}_a)$ , y la matriz de covarianzas de  $\mathbf{a}_t$  se puede recuperar a partir de  $\boldsymbol{\Omega}_a$ . La matriz de covarianzas de  $\mathbf{w} = (\mathbf{w}'_2, \dots, \mathbf{w}'_n)'$  es  $\boldsymbol{\Omega}_w(\theta) = \boldsymbol{\Theta}(\theta)(\mathbf{I}_n \otimes \boldsymbol{\Omega}_a)\boldsymbol{\Theta}(\theta)'$ , con lo que el logaritmo de su función de verosimilitud tiene la siguiente expresión

$$\log f(\theta|\mathbf{w}) \propto -\frac{1}{2} \log |\boldsymbol{\Omega}_w(\theta)| - \frac{1}{2} \mathbf{w}' \boldsymbol{\Omega}_w^{-1}(\theta) \mathbf{w}. \quad (5.5)$$

Denotando por  $\mathbf{D} = [d_{ij}]$  la matriz de primeras diferencias con dimensiones  $(n-1) \times n$  tal que  $d_{ii} = -1$ ,  $d_{i,i+1} = 1$  y  $d_{ij} = 0$  en otro caso, se puede expresar el proceso estacionario como  $\mathbf{w} = (\mathbf{D} \otimes \mathbf{I}_m)\mathbf{y}$ . Antes de obtener la primera derivada conviene tener en cuenta los siguientes resultados:  $d\boldsymbol{\Theta}_*/d\theta = \text{diag}(\mathbf{I}_{m_1}, \mathbf{0})$ ,  $d^2\boldsymbol{\Theta}_*/d\theta^2 = \mathbf{0}$  y  $d\boldsymbol{\Theta}(\theta)/d\theta = [-(\mathbf{I}_{n-1} \otimes \mathbf{I}_{(m_1)})|\mathbf{0}_{(n-1) \times m}]$  donde  $\mathbf{I}_{(m_1)}$  es una matriz diagonal  $m \times m$  con unos en las posiciones correspondientes a  $\mathbf{w}_{1t}$ , de forma que se puede expresar alternativamente  $d\boldsymbol{\Theta}(\theta)/d\theta = [(\mathbf{D} \otimes \mathbf{I}_{(m_1)})|\mathbf{0}_{(n-1) \times m}] - [\mathbf{0}_{(n-1) \times m} | (\mathbf{I}_{(n-1)} \otimes \mathbf{I}_{(m_1)})]$ . Nótese además que el vector de residuos exactos de (5.4) se obtiene como  $\hat{\mathbf{a}}_0 = (\mathbf{I}_n \otimes \mathbf{P}'_a)\boldsymbol{\Theta}(\theta)'\boldsymbol{\Omega}_w^{-1}\mathbf{w}$ . Así, la primera derivada de (5.5) con respecto al parámetro de interés  $\theta$  evaluada bajo la

hipótesis nula tiene la siguiente expresión

$$\begin{aligned} \frac{d \log f(\theta|\mathbf{w})}{d\theta} \Big|_{\theta=1} &= -\frac{1}{2} \text{tr} \left[ \boldsymbol{\Omega}_w^{-1} \frac{d\boldsymbol{\Omega}_w(\theta)}{d\theta} \Big|_{\theta=1} \right] + \frac{1}{2} \mathbf{w}' \boldsymbol{\Omega}_w^{-1} \frac{d\boldsymbol{\Omega}_w(\theta)}{d\theta} \Big|_{\theta=1} \boldsymbol{\Omega}_w^{-1} \mathbf{w}, \\ &= -\frac{1}{2} \text{tr} \left[ \boldsymbol{\Omega}_w^{-1} \frac{d\boldsymbol{\Omega}_w(\theta)}{d\theta} \Big|_{\theta=1} \right] + \text{tr} \left[ \sum_{t=1}^n \hat{\mathbf{a}}_t \hat{\mathbf{a}}_t^{(m_1)'} \right] - \text{tr} \left[ \sum_{t=2}^n \hat{\mathbf{a}}_t \hat{\mathbf{e}}_t^{(m_1)'} \right], \end{aligned}$$

donde para simplificar la notación, se escribe  $\boldsymbol{\Omega}_w^{-1} = \boldsymbol{\Omega}_w^{-1}(1)$ . Por su parte, los vectores  $\hat{\mathbf{a}}_t$  contiene los residuos exactos, el vector  $\hat{\mathbf{a}}_t^{(m_1)}$  contiene los residuos exactos correspondientes a  $\mathbf{y}_{1t}$  en sus posiciones y ceros en otro caso, y  $\hat{\mathbf{e}}_t^{(m_1)}$  tiene esta misma estructura con los elementos de la serie filtrada  $\hat{\mathbf{e}} = \boldsymbol{\Omega}_w^{-1} \mathbf{w}$ . Se puede comprobar numéricamente que el valor esperado del tercer término de expresión de la traza es igual a la mitad de la segunda, cuya esperanza es  $nm_1$ . Dado que  $\text{tr}[\boldsymbol{\Omega}_w^{-1} d\boldsymbol{\Omega}_w] \simeq nm_1$ , se comprueba que el valor esperado de la primera derivada es cero, por lo que el estadístico de contraste se obtiene a partir de la segunda, cuya expresión es

$$\begin{aligned} \frac{d^2 \log f(\theta|\mathbf{w})}{d\theta^2} \Big|_{\theta=1} &= -\frac{1}{2} \text{tr} \left[ - \left( \boldsymbol{\Omega}_w^{-1} \frac{d\boldsymbol{\Omega}_w(\theta)}{d\theta} \Big|_{\theta=1} \right)^2 + \boldsymbol{\Omega}_w^{-1} \frac{d^2 \boldsymbol{\Omega}_w(\theta)}{d\theta^2} \Big|_{\theta=1} \right] - \\ &\quad - \mathbf{w}' \boldsymbol{\Omega}_w^{-1} \frac{d\boldsymbol{\Omega}_w(\theta)}{d\theta} \Big|_{\theta=1} \boldsymbol{\Omega}_w^{-1} \frac{d\boldsymbol{\Omega}_w(\theta)}{d\theta} \Big|_{\theta=1} \boldsymbol{\Omega}_w^{-1} \mathbf{w} + \\ &\quad + \frac{1}{2} \mathbf{w}' \boldsymbol{\Omega}_w^{-1} \frac{d^2 \boldsymbol{\Omega}_w(\theta)}{d\theta^2} \Big|_{\theta=1} \boldsymbol{\Omega}_w^{-1} \mathbf{w} \end{aligned}$$

donde el estadístico surge del tercer término de la segunda derivada. De las relaciones anteriores,  $d^2 \boldsymbol{\Omega}_w(\theta)/d\theta^2|_{\theta=1} = d\boldsymbol{\Theta}(1)(\mathbf{I}_n \otimes \boldsymbol{\Omega}_a)d\boldsymbol{\Theta}(1)' = (\mathbf{I}_{n-1} \otimes \mathbf{I}_{(m_1)} \boldsymbol{\Omega}_a \mathbf{I}_{(m_1)})$ . Por último, teniendo en cuenta que  $(\mathbf{I}_n \otimes \mathbf{I}_{(m_1)}) \boldsymbol{\Theta}'(1) = (\mathbf{D}' \otimes \mathbf{I}_{(m_1)})$ , insertando a ambos lados de la segunda derivada de  $\boldsymbol{\Omega}_w(\theta)$  la matriz  $[(\mathbf{D}\mathbf{D}')^{-1} \mathbf{D}\mathbf{D}' \otimes \mathbf{I}_{(m_1)}]$ , ese tercer término de la derivada puede expresarse como  $\hat{\mathbf{a}}' [\mathbf{D}' (\mathbf{D}\mathbf{D}')^{-1} \mathbf{D} \otimes \mathbf{I}_{(m_1)}] \hat{\mathbf{a}}$ . De Mazas (2012),  $\mathbf{D}' (\mathbf{D}\mathbf{D}')^{-1} \mathbf{D} = \mathbf{M}\mathbf{C}_n \mathbf{C}'_n \mathbf{M}$ , donde  $\mathbf{M}$  es la matriz generadora de observaciones en desviaciones con respecto a la media y  $\mathbf{C}_n$  una matriz triangular inferior de unos. De esta forma, el estadístico de contraste puede expresarse de la siguiente manera

$$L = \text{tr} \left[ \left( \sum_{t=1}^n \tilde{\mathbf{a}}_{1t}^0 \tilde{\mathbf{a}}_{1t}^{0'} \right)^{-1} \sum_{t=1}^n \left( \sum_{\tau=1}^t \tilde{\mathbf{a}}_{1\tau}^0 \right) \left( \sum_{\tau=1}^t \tilde{\mathbf{a}}_{1\tau}^0 \right)' \right], \quad (5.6)$$

donde  $\tilde{\mathbf{a}}_{1t}^0 = \hat{\mathbf{a}}_{1t}^0 - \bar{\mathbf{a}}_1^0$ , es el vector de residuos exactos del modelo (5.4) correspondientes a  $\mathbf{a}_1$ , evaluados bajo la hipótesis nula y centrados. La matriz inversa se ha añadido,

sin pérdida de generalidad ya que los residuos están estandarizados, por analogía con los estadísticos derivados en capítulos anteriores. De Pötscher (1991) y Poskitt (2006), el estimador de máxima verosimilitud exacta es consistente incluso bajo la presencia de raíces unitarias en el polinomio media móvil, por lo que los residuos exactos  $\hat{\mathbf{a}}_{1t}^0$  convergen a  $\mathbf{a}_{1t}$ , un proceso ruido blanco vectorial de dimensiones  $m_1 \times 1$  con matriz de covarianzas  $\mathbf{\Omega}_{11}$ . Por lo tanto, de Phillips y Durlauf (1986),  $n^{-1} \sum_{t=1}^n \hat{\mathbf{a}}_{1t}^0 \hat{\mathbf{a}}_{1t}^{0'} \xrightarrow{p} \mathbf{\Omega}_{11}$  y de los resultados obtenidos en el capítulo 3

$$n^{-1} \sum_{\tau=1}^t (\hat{\mathbf{a}}_{1\tau} - \bar{\mathbf{a}}_1) \xrightarrow{d} \mathbf{\Omega}_{11}^{1/2} [\mathbf{W}(r) - r\mathbf{W}(1)],$$

donde  $\mathbf{W}(r)$  es un proceso browniano vectorial de dimensiones  $m_1 \times 1$ . Por lo tanto, el estadístico de contraste tiene una distribución Cramèr-von Mises con  $m_1$  grados de libertad,  $L/(n-1)m_1 \xrightarrow{d} CvM(m_1)/m_1$ . La distribución en muestras finitas del estadístico obtenida por simulación coincide con la derivada para el contraste de no invertibilidad estricta cuando la dimensión del vector es  $m_1$ , por lo que la aproximación IG se puede aplicar también a este caso.

### 5.2.2. Extensión para modelos con correlación serial

El modelo simple (5.1)-(5.2) se puede extender permitiendo que los errores  $\mathbf{a}_{1t}$  y  $\mathbf{a}_{2t}$  presenten correlación serial. Una primera extensión consiste en suponer que sólo el error del vector de variables integradas,  $\mathbf{y}_{2t}$ , presenta correlación de tipo VARMA(p,q), tal que  $\mathbf{y}_{1t} = \mathbf{B}\mathbf{y}_{2t} + \mathbf{u}_{1t}$  y  $\mathbf{\Phi}_{22}(B)\nabla\mathbf{y}_{2t} = \mathbf{\Theta}_{22}(B)\mathbf{u}_{2t}$  donde  $\mathbf{u}_t = (\mathbf{u}'_{1t}, \mathbf{u}'_{2t})' \sim N(\mathbf{0}, \mathbf{\Sigma})$  con  $\text{cov}(\mathbf{u}_{1t}, \mathbf{u}_{2t}) \neq 0$ , y los polinomios  $\mathbf{\Phi}_{22}(B)$  y  $\mathbf{\Theta}_{22}(B)$  de órdenes  $p$  y  $q$  respectivamente tienen todas sus raíces fuera del círculo unitario y no presentan raíces comunes. De forma conjunta el sistema puede expresarse como

$$\begin{bmatrix} \mathbf{I}_{m_1} \nabla^{-1} & -\mathbf{B} \nabla^{-1} \\ \mathbf{0} & \mathbf{\Phi}_{22}(B) \end{bmatrix} \begin{pmatrix} \nabla \mathbf{y}_{1t} \\ \nabla \mathbf{y}_{2t} \end{pmatrix} = \begin{bmatrix} \mathbf{I}_{m_1} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Theta}_{22}(B) \end{bmatrix} \begin{pmatrix} \mathbf{u}_{1t} \\ \mathbf{u}_{2t} \end{pmatrix}, \quad \mathbf{H} = \begin{bmatrix} \mathbf{I}_{m_1} \nabla^{-1} & -\mathbf{B} \nabla^{-1} \\ \mathbf{0} & \mathbf{I}_{m_2} \end{bmatrix},$$

donde  $\nabla^{-1} = (1 - B)^{-1}$ . En este modelo el conjunto de variables  $\mathbf{y}_{2t}$  presentan una tendencia estocástica, que se transmite al vector  $\mathbf{y}_{1t}$  a través del conjunto de relaciones de cointegración, la matriz  $\mathbf{B}$ . Premultiplicando el modelo por  $\mathbf{H}^{-1}$  se obtiene su forma

reducida como un modelo VARIMA(p,1,q+1) no invertible

$$\begin{bmatrix} \mathbf{I}_{m_1} & \Phi_{12}^*(B) \\ \mathbf{0} & \Phi_{22}(B) \end{bmatrix} \begin{pmatrix} \nabla \mathbf{y}_{1t} \\ \nabla \mathbf{y}_{2t} \end{pmatrix} = \begin{bmatrix} \mathbf{I}_{m_1} & \Theta_{12}^*(B) \\ \mathbf{0} & \Theta_{22}(B) \end{bmatrix} \left[ \begin{pmatrix} \mathbf{I}_{m_1} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{m_2} \end{pmatrix} - \begin{pmatrix} \mathbf{I}_{m_1} & -\mathbf{B} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} B \right] \begin{pmatrix} \mathbf{u}_{1t}^* \\ \mathbf{u}_{2t} \end{pmatrix},$$

donde  $\mathbf{u}_{1t}^* = \mathbf{u}_{1t} + \mathbf{B}\mathbf{u}_{2t}$ ,  $\Phi_{12}^* = \mathbf{B}\Phi_{22}(B) - \mathbf{B}$  es un polinomio matricial de orden  $p$  tal que  $\Phi_{12}^* = -\Phi_{12,1}^*B - \dots - \Phi_{12,p}^*B^p$ , siendo las matrices  $\Phi_{12,j}^* = \mathbf{B}\Phi_{22,j}$  y  $\Theta_{22}^*(B)$  se define de manera similar. Las raíces unitarias del modelo se concentran en el segundo factor del polinomio media móvil, donde existen  $m_1$  raíces iguales a uno correspondientes al primer subvector del sistema original. De igual forma que en el modelo simple, la presencia de cointegración se manifiesta en un bloque (1,1) de la matriz de coeficientes del segundo factor media móvil igual a la matriz identidad. A partir del modelo así expresado y utilizando la corrección paramétrica descrita en el capítulo 3, el contraste de cointegración se puede realizar utilizando el estadístico (5.6) sobre los residuos exactos del modelo VARMA(p,1,q+1) correspondientes a  $\mathbf{a}_1$ , evaluados bajo la hipótesis nula de cointegración. La derivación de este resultado se omite por ser similar a la realizada en el apartado anterior, utilizando en este caso el proceso filtrado  $\Theta^{*-1}(B)\Phi^*(B)\nabla \mathbf{y}_t$ .

Una segunda extensión del modelo simple consiste en considerar que las perturbaciones de ambas ecuaciones, (5.1) y (5.2), presentan correlación serial. En este modelo  $\mathbf{y}_{2t}$  presenta una tendencia estocástica, mientras que en la primera ecuación se permite que las desviaciones de corto plazo con respecto a la relación de cointegración tengan su propia dinámica estacionaria de tipo VARMA. De esta forma, el modelo conjunto puede expresarse como

$$\begin{bmatrix} \nabla^{-1}\mathbf{I}_{m_1} & -\mathbf{B}\nabla^{-1} \\ \mathbf{0} & \mathbf{I}_{m_2} \end{bmatrix} \begin{pmatrix} \nabla \mathbf{y}_{1t} \\ \nabla \mathbf{y}_{2t} \end{pmatrix} = \begin{bmatrix} \Psi_{11}(B) & \mathbf{0} \\ \mathbf{0} & \Psi_{22}(B) \end{bmatrix} \begin{pmatrix} \mathbf{u}_{1t} \\ \mathbf{u}_{2t} \end{pmatrix},$$

donde  $\mathbf{u}_t = (\mathbf{u}'_{1t}, \mathbf{u}'_{2t})' \sim N(\mathbf{0}, \Sigma)$ , y  $\Psi_{ii}(B) = \Phi_{ii}^{-1}(B)\Theta_{ii}(B)$  ( $i = 1, 2$ ) son dos polinomios matriciales tales que las raíces de  $\Phi_{ii}(B)$  y  $\Theta_{ii}(B)$  son distintas y caen estrictamente fuera del círculo unitario. La matriz de covarianzas  $\Sigma$  del vector de perturbaciones del modelo es definida positiva y  $\text{cov}(\mathbf{u}_{1t}, \mathbf{u}'_{2t}) = \Sigma_{12} \neq \mathbf{0}$ , de forma que la única correlación posible entre las perturbaciones de ambas ecuaciones es la contemporánea. Igual que en el caso anterior, premultiplicando por la inversa de la matriz que acompaña a  $\nabla \mathbf{y}_t$  se obtiene la forma reducida del modelo como un VARIMA(p,1,q+1) con una estructura

triangular por bloques

$$\begin{bmatrix} \Phi_{11}(B) & \Phi_{12}^*(B) \\ \mathbf{0} & \Phi_{22}(B) \end{bmatrix} \begin{pmatrix} \nabla \mathbf{y}_{1t} \\ \nabla \mathbf{y}_{2t} \end{pmatrix} = \begin{bmatrix} \Theta_{11}(B) & \Theta_{12}^*(B) \\ \mathbf{0} & \Theta_{22}(B) \end{bmatrix} (\mathbf{I}_m - \Theta_* B) \begin{pmatrix} \mathbf{u}_{1t}^* \\ \mathbf{u}_{2t} \end{pmatrix},$$

donde  $\mathbf{u}_{1t}^* = \mathbf{u}_{1t} + \mathbf{B}\mathbf{u}_{2t}$ ,  $\Phi_{12}^*(B) = -\Phi_{12,1}^*B - \dots - \Phi_{12,p}^*B^p$  con  $\Phi_{12,j}^* = \Phi_{11,j}\mathbf{B} - \mathbf{B}\Phi_{22,j}$ ,  $\Theta_{12}^*(B)$  se define de forma similar y  $\Theta_*$  es la matriz definida en la ecuación (5.3) del apartado anterior. Al igual que en el modelo simple, la matriz del segundo factor concentra las raíces unitarias del polinomio MA y permite obtener una estimación directa de las relaciones de cointegración a partir del bloque (1,2) de la matriz media móvil y plantear el contraste de cointegración a partir del bloque (1,1) de ésta. De igual forma que en el caso anterior, el contraste de cointegración se puede realizar utilizando (5.6) sobre los residuos exactos del modelo evaluados bajo la hipótesis nula. En ambos casos la matriz de relaciones de cointegración,  $\mathbf{B}$  se puede estimar consistentemente utilizando los algoritmos habituales de máxima verosimilitud exacta incluso bajo la presencia de raíces unitarias MA (Pötscher, 1991; Poskitt, 2006). El algoritmo de Gallego (2009) está preparado para estimar modelos multivariantes multiplicativos como los propuestos aquí.

### 5.2.3. El modelo de tendencias comunes

El modelo de tendencias comunes de Stock y Watson (1988) puede expresarse en forma VARIMA, de forma que el estadístico (5.6) para el contraste de cointegración puede utilizarse también para detectar el número de tendencias comunes en un vector de series. Un sistema con  $m_2 < m$  tendencias comunes y sin otro componente determinista tiene la siguiente estructura considerando la misma partición del vector  $\mathbf{y}_t$  que en modelos anteriores

$$\begin{bmatrix} \mathbf{y}_{1t} \\ \mathbf{y}_{2t} \end{bmatrix} = \begin{bmatrix} \mathbf{H}_1 \\ \mathbf{H}_2 \end{bmatrix} \boldsymbol{\mu}_t + \begin{bmatrix} \mathbf{u}_{1t} \\ \mathbf{u}_{2t} \end{bmatrix}, \quad \boldsymbol{\mu}_t = \boldsymbol{\mu}_{t-1} + \mathbf{v}_t,$$

donde  $\boldsymbol{\mu}_t$  es un vector  $m_2 \times 1$  que contiene las  $m_2 < m$  tendencias que afectan al vector de series  $\mathbf{y}_t$  y  $\mathbf{v}_t \sim N(\mathbf{0}, \boldsymbol{\Sigma}_v)$  con  $\boldsymbol{\Sigma}_v$  una matriz  $m_2 \times m_2$  definida positiva de rango completo. Para identificar el sistema se normaliza la matriz  $\mathbf{H} = [\mathbf{H}'_1 | \mathbf{H}'_2]'$  tal que  $\mathbf{H} = [\mathbf{A}' | \mathbf{I}_{m_2}]'$  donde  $\mathbf{A} = \mathbf{H}_2^{-1}\mathbf{H}_1$ . Esta normalización implica que  $\mathbf{y}_{2t} = \boldsymbol{\mu}_t + \mathbf{u}_{2t}$  es un proceso I(1) donde  $\boldsymbol{\mu}_t$  es el nivel local propio, por lo que se puede considerar  $\mathbf{y}_{2t}$  como el vector de variables exógenas. Por su parte, el subvector  $\mathbf{y}_{1t}$  también será un proceso integrado pero su nivel,  $\mathbf{A}\boldsymbol{\mu}_t$ , se expresa como combinación lineal del nivel estocástico propio de  $\mathbf{y}_{2t}$ . Así normalizado el modelo de tendencias comunes coincide con el descrito por

Nyblom y Harvey (2000) bajo la hipótesis nula de que el modelo presenta  $m_2$  tendencias comunes. Obviamente, esta normalización no es trivial, por lo que debe ser realizada con cuidado. Harris (1997) propone un método basado en componentes principales para realizar esta partición.

Es bien conocido que la forma reducida de un modelo de nivel local admite una representación IMA(1,1),  $\nabla \mathbf{y}_t = (\mathbf{I}_m - \Theta B)\mathbf{a}_t$ , con  $\mathbf{a}_t \sim N(\mathbf{0}, \Omega_a)$ , donde la matriz de covarianzas  $\Omega_a$  tiene rango completo y la estructura de tendencias comunes implica una serie de restricciones sobre la matriz  $\Theta$ . Utilizando ambas representaciones, las dos únicas matrices de covarianzas no nulas de  $\nabla \mathbf{y}_t$  son  $\Gamma(0) = E(\nabla \mathbf{y}_t \nabla \mathbf{y}_t') = \mathbf{H}\Sigma_v \mathbf{H}' + 2\Sigma_v \equiv \Omega_a + \Theta \Omega_a \Theta'$  y  $\Gamma(1) = E(\nabla \mathbf{y}_t \nabla \mathbf{y}_{t-1}') = -\Sigma_u \equiv -\Theta \Omega_a$ , lo que permite expresar su matriz de covarianzas de largo plazo como

$$\Gamma_{lp} = \Gamma(0) + \Gamma(1) + \Gamma(1)' = \mathbf{H}\Sigma_v \mathbf{H}' = (\mathbf{I}_m - \Theta)\Omega_a(\mathbf{I}_m - \Theta)',$$

La matriz  $\mathbf{H}\Sigma_v \mathbf{H}'$  tiene rango  $m_2 < m$ , igual al número de columnas de  $\mathbf{H}$ , lo que implica que el último término tendrá igual rango. Teniendo en cuenta que  $\Omega_a$  tiene rango  $m$  completo, entonces el rango de la matriz de ganancias del polinomio,  $(\mathbf{I}_m - \Theta)$ , es necesariamente igual a  $m_2$ . Dado que  $\Theta$  es una matriz real, admite la siguiente descomposición espectral,  $\Theta = \mathbf{P}\mathbf{\Lambda}\mathbf{P}^{-1}$  donde la matriz diagonal  $\mathbf{\Lambda}$  contiene sus autovalores y las columnas de  $\mathbf{P}$  sus correspondientes autovectores, lo que permite reexpresar la matriz de ganancias como  $(\mathbf{I}_m - \Theta) = \mathbf{P}(\mathbf{I}_m - \mathbf{\Lambda})\mathbf{P}^{-1}$ . Si el rango de esta matriz es igual a  $m_2$ , entonces  $m_1$  de los autovalores de  $\Theta$  serán iguales a uno lo que, como se indicó en el capítulo 3, implica que existen  $m_1$  raíces unitarias en el polinomio media móvil. Por lo tanto, una condición necesaria para la presencia de  $m_2$  tendencias comunes en un modelo VIMA(1,1) es la existencia de  $m_1$  raíces unitarias.

Además, la existencia de tendencias comunes introduce una serie de restricciones adicionales sobre la matriz media móvil. Combinando ambas representaciones se puede expresar el modelo como  $\mathbf{y}_t = \boldsymbol{\mu}_t + \Theta \mathbf{a}_t$  con  $\boldsymbol{\mu}_t = \boldsymbol{\mu}_{t-1} + (\mathbf{I}_m - \Theta)\mathbf{a}_t$ . Convenientemente particionado e imponiendo la normalización de la matriz de difusión del vector de tendencias,  $\mathbf{H}$ , definida anteriormente,

$$\begin{aligned} \begin{pmatrix} \mathbf{y}_{1t} \\ \mathbf{y}_{2t} \end{pmatrix} &= \begin{pmatrix} \mathbf{A} \\ \mathbf{I}_{m_2} \end{pmatrix} \boldsymbol{\mu}_{1t} + \begin{pmatrix} \Theta_{11} & \Theta_{12} \\ \Theta_{21} & \Theta_{22} \end{pmatrix} \begin{pmatrix} \mathbf{a}_{1t} \\ \mathbf{a}_{2t} \end{pmatrix}, \\ \boldsymbol{\mu}_{1t} &= \boldsymbol{\mu}_{1t-1} + (\mathbf{I}_{m_1} - \Theta_{11})\mathbf{a}_{1t} - \Theta_{12}\mathbf{a}_{2t}, \\ \boldsymbol{\mu}_{2t} &= \boldsymbol{\mu}_{2t-1} + (\mathbf{I}_{m_2} - \Theta_{22})\mathbf{a}_{2t} - \Theta_{21}\mathbf{a}_{1t}. \end{aligned}$$

Tal y como está definido el modelo,  $\mathbf{y}_{2t}$  tiene su propia tendencia estocástica, por lo que las perturbaciones propias de  $\mathbf{y}_{1t}$  no deberían afectar a  $\boldsymbol{\mu}_{2t}$ , lo que justifica imponer la restricción  $\boldsymbol{\Theta}_{21} = \mathbf{0}$ . Por otra parte, dado que  $\boldsymbol{\mu}_{1t} = \mathbf{A}\boldsymbol{\mu}_{2t}$ , reemplazando en las expresiones anteriores se obtiene que  $(\mathbf{I}_{m_1} - \boldsymbol{\Theta}_{11})\mathbf{a}_{1t} - \boldsymbol{\Theta}_{12}\mathbf{a}_{2t} = \mathbf{A}(\mathbf{I}_{m_2} - \boldsymbol{\Theta}_{22})\mathbf{a}_{2t}$ , donde la igualdad se cumple si  $\boldsymbol{\Theta}_{11} = \mathbf{I}_{m_1}$  y  $\mathbf{A} = -\boldsymbol{\Theta}_{12}(\mathbf{I}_{m_2} - \boldsymbol{\Theta}_{22})^{-1}$ . Con todo esto, la presencia de  $m_2 < m$  tendencias estocásticas comunes en el vector de series temporales  $\mathbf{y}_t$  se puede describir mediante un modelo VIMA(1,1) con la siguiente estructura:

$$\begin{pmatrix} \nabla \mathbf{y}_{1t} \\ \nabla \mathbf{y}_{2t} \end{pmatrix} = \begin{pmatrix} \mathbf{a}_{1t} \\ \mathbf{a}_{2t} \end{pmatrix} - \begin{pmatrix} \mathbf{I}_{m_1} & \boldsymbol{\Theta}_{12} \\ \mathbf{0} & \boldsymbol{\Theta}_{22} \end{pmatrix} \begin{pmatrix} \mathbf{a}_{1t-1} \\ \mathbf{a}_{2t-1} \end{pmatrix}. \quad (5.7)$$

Puesto que la matriz media móvil es diagonal por bloques, las raíces del polinomio pueden obtenerse a partir de los autovalores de las matrices que componen su diagonal principal. Tal y como se había comentado anteriormente, la presencia de  $m_2$  tendencias comunes implica que el modelo tiene  $m_1 = m - m_2$  raíces unitarias, que se manifiestan en el bloque (1,1) de la matriz media móvil. En cuanto al resto de raíces, éstas dependerán de los autovalores de  $\boldsymbol{\Theta}_{22}$ . Por sustituciones sucesivas en (5.7) se puede obtener la forma MA( $\infty$ ) de  $\mathbf{y}_t$  como

$$\begin{pmatrix} \mathbf{y}_{1t} \\ \mathbf{y}_{2t} \end{pmatrix} = \mathbf{y}_0 + \begin{pmatrix} \mathbf{0} & -\boldsymbol{\Theta}_{12} \\ \mathbf{0} & (\mathbf{I}_{m_2} - \boldsymbol{\Theta}_{22}) \end{pmatrix} \begin{pmatrix} \sum_{i=1}^{t-1} \mathbf{a}_{1t} \\ \sum_{i=1}^{t-1} \mathbf{a}_{2t} \end{pmatrix} + \begin{pmatrix} \mathbf{a}_{1t} \\ \mathbf{a}_{2t} \end{pmatrix}.$$

Así, el vector  $\mathbf{y}_t$  se descompone en una parte estacionaria, el vector  $\mathbf{a}_t$  que recoge la correlación contemporánea entre las series modelizadas, un vector de condiciones iniciales,  $\mathbf{y}_0$  y un vector que recoge las perturbaciones acumuladas del sistema o tendencias estocásticas. La matriz que acompaña a este último componente muestra cómo se difunde la tendencia estocástica a ambos conjuntos de series  $\mathbf{y}$ , como se ve, la única tendencia que afecta al sistema es la propia de  $\mathbf{y}_{2t}$ , cuya dimensión es  $m_2 < m$ . Nótese por último que esta tendencia estocástica se puede eliminar premultiplicando el sistema por la matriz  $[\mathbf{I}_{m_1} | \boldsymbol{\Theta}_{12}(\mathbf{I}_{m_2} - \boldsymbol{\Theta}_{22})^{-1}]$ , de forma que la combinación lineal  $\mathbf{y}_{1t} - \mathbf{A}\mathbf{y}_{2t}$  con  $\mathbf{A} = \boldsymbol{\Theta}_{12}(\mathbf{I}_{m_2} - \boldsymbol{\Theta}_{22})^{-1}$  es un proceso estacionario. Por lo tanto, las filas de la matriz  $\mathbf{A}$  se pueden interpretar como las  $m_1$  relaciones de cointegración que implican la presencia de  $m_2$  tendencias comunes. Igualmente, suponiendo  $\mathbf{y}_0 = \mathbf{0}$ , el modelo (5.7) admite la siguiente representación triangular

$$\mathbf{y}_{1t} = \mathbf{A}\mathbf{y}_{2t} + \mathbf{a}_{1t} - \mathbf{A}\mathbf{a}_{2t}, \quad \nabla \mathbf{y}_{2t} = (\mathbf{I}_{m_2} - \boldsymbol{\Theta}_{22}B)\mathbf{a}_{2t}.$$

Donde el vector  $\mathbf{y}_{2t}$  contiene el conjunto de variables exógenas e  $I(1)$ , y cuya dinámica se corresponde con un modelo IMA(1,1) multivariante. Por su parte  $\mathbf{y}_{1t}$  recoge las relaciones de cointegración más un ruido blanco, por lo que el modelo de tendencias comunes puede contemplarse como un caso particular la primera extensión del modelo triangular con correlación serial, considerada en el apartado anterior. El contraste de cointegración o de tendencias comunes se puede realizar utilizando el procedimiento descrito en el apartado anterior, ya que (5.7) se puede reescribir como

$$\begin{pmatrix} \nabla \mathbf{y}_{1t} \\ \nabla \mathbf{y}_{2t} \end{pmatrix} = \left[ \begin{pmatrix} \mathbf{I}_{m_1} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{m_2} \end{pmatrix} - \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \Theta_{22} \end{pmatrix} B \right] \left[ \begin{pmatrix} \mathbf{I}_{m_1} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{m_2} \end{pmatrix} - \begin{pmatrix} \theta \mathbf{I}_{m_1} & \Theta_{12} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} B \right] \begin{pmatrix} \mathbf{a}_{1t} \\ \mathbf{a}_{2t} \end{pmatrix}$$

donde bajo la hipótesis nula de cointegración,  $H_0 : \theta = 1$ , existen  $m_2$  tendencias comunes y, por lo tanto,  $m_1$  relaciones de cointegración que se pueden estimar consistentemente a partir de la estimación de máxima verosimilitud exacta de (5.7). Este modelo de tendencias comunes guarda relación con los modelos factores comunes de Escribano y Peña (1993) o más recientemente Peña y Poncela (2006) o de cointegración en modelos estructurales de Wagner (2010), en posteriores desarrollos del trabajo se estudiará la relación de estos con los contraste aquí descritos.

#### 5.2.4. Ejemplo ilustrativo: El modelo bivariante cointegrado

##### El contraste de cointegración

Para evaluar la potencia del contraste de cointegración derivado este apartado, se diseñan dos experimentos Monte-Carlo, utilizando en ambos casos el mismo proceso generador de datos

$$\begin{pmatrix} \nabla y_{1t} \\ \nabla y_{2t} \end{pmatrix} = \begin{pmatrix} a_{1t}^* \\ a_{2t}^* \end{pmatrix} - \begin{pmatrix} \theta & -\beta \\ 0 & 0 \end{pmatrix} \begin{pmatrix} a_{1,t-1}^* \\ a_{2,t-1}^* \end{pmatrix}, \quad t = 2, \dots, n.$$

Este modelo es la forma reducida del modelo triangular  $y_{1t} = \beta y_{2t} + a_{1t}$  y  $\nabla y_{2t} = a_{2t}$ , con  $a_{1t}^* = a_{1t} + \beta a_{2t}$  y  $a_{2t}^* = a_{2t}$ . Las varianzas de ambos errores son unitarias,  $\sigma_{a_1}^2 = \sigma_{a_2}^2 = 1,0$ , y su coeficiente de correlación es de  $\rho = 0,5$ . En el primer experimento se estudia la potencia del contraste, conociendo el valor exacto de la relación de cointegración ( $\beta = 1,5$ ). Todos los experimentos se han realizado utilizando el software estadístico *Empiricus* y el número de replicaciones para cada caso es de 5000. El cuadro 5.1 muestra la probabilidad de rechazo de la hipótesis nula cuando el estadístico de contraste se calcula

Cuadro 5.1: Potencia de  $L/(n-1)$  para  $\hat{a}_{1t}^*$  ( $\alpha = 5\%$ ).

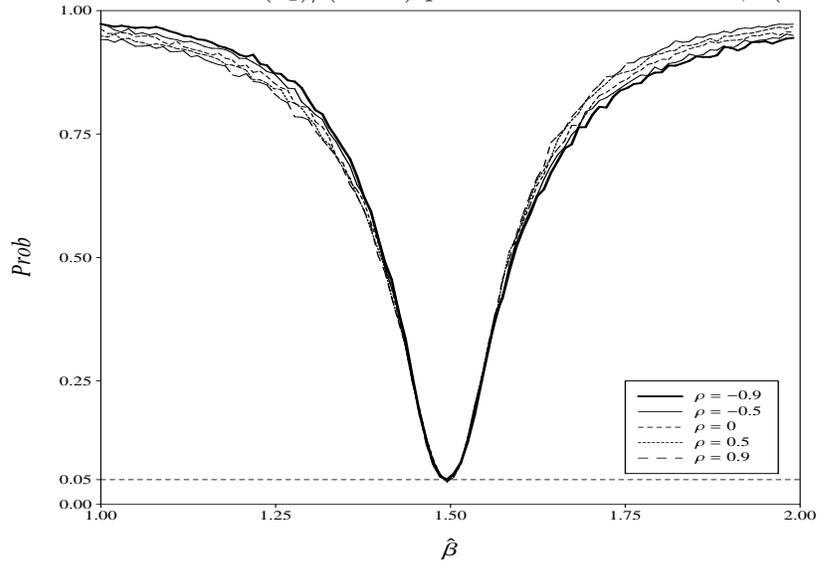
$\theta$	$n = 10$	$n = 20$	$n = 30$	$n = 40$	$n = 50$	$n = 100$	$n = 200$
1.00	0.0552	0.0566	0.0488	0.0560	0.0502	0.0462	0.0560
0.99	0.0538	0.0524	0.0468	0.0470	0.0508	0.0634	0.0990
0.98	0.0562	0.0514	0.0520	0.0568	0.0660	0.0942	0.2386
0.97	0.0488	0.0516	0.0600	0.0686	0.0766	0.1546	0.3620
0.96	0.0514	0.0512	0.0708	0.0750	0.1038	0.2442	0.5206
0.95	0.0594	0.0674	0.0724	0.0958	0.1224	0.3072	0.6112
0.90	0.0632	0.0930	0.1474	0.2144	0.3002	0.6114	0.8622
0.85	0.0722	0.1432	0.2464	0.3698	0.4742	0.7554	0.9416
0.80	0.0892	0.2254	0.3748	0.4946	0.5984	0.8554	0.9756

Cuadro 5.2: Potencia de  $L/(n-1)$  para  $\hat{a}_{2t}^*$  ( $\alpha = 5\%$ ).

$\theta$	$n = 10$	$n = 20$	$n = 30$	$n = 40$	$n = 50$	$n = 100$	$n = 200$
1.00	0.0540	0.0518	0.0530	0.0516	0.0466	0.0496	0.0490
0.99	0.0548	0.0484	0.0476	0.0464	0.0528	0.0556	0.0508
0.98	0.0528	0.0510	0.0458	0.0486	0.0538	0.0520	0.0532
0.97	0.0490	0.0520	0.0496	0.0538	0.0532	0.0530	0.0496
0.96	0.0490	0.0476	0.0540	0.0482	0.0556	0.0454	0.0500
0.95	0.0516	0.0518	0.0482	0.0530	0.0510	0.0448	0.0500
0.90	0.0546	0.0490	0.0482	0.0548	0.0464	0.0542	0.0502
0.85	0.0494	0.0572	0.0486	0.0492	0.0430	0.0466	0.0496
0.80	0.0584	0.0546	0.0488	0.0516	0.0470	0.0460	0.0462

Cuadro 5.3: Potencia del test conjunto para  $\hat{a}_{1t}^*$  y  $\hat{a}_{2t}^*$  ( $\alpha = 5\%$ ).

$\theta$	$n = 10$	$n = 20$	$n = 30$	$n = 40$	$n = 50$	$n = 100$	$n = 200$
1.00	0.0320	0.0438	0.0466	0.0500	0.0472	0.0476	0.0536
0.99	0.0336	0.0464	0.0492	0.0496	0.0586	0.1084	0.2588
0.98	0.0372	0.0496	0.0612	0.0772	0.1034	0.2494	0.5562
0.97	0.0360	0.0642	0.0856	0.1266	0.1644	0.4190	0.7046
0.96	0.0330	0.0688	0.1216	0.1720	0.2568	0.5450	0.8222
0.95	0.0388	0.0892	0.1610	0.2428	0.3288	0.6350	0.8646
0.90	0.0602	0.2052	0.3772	0.5068	0.6084	0.8580	0.9684
0.85	0.0830	0.3504	0.5356	0.6474	0.7386	0.9140	0.9862
0.80	0.1302	0.4564	0.6448	0.7504	0.8084	0.9512	0.9942

Figura 5.1: Potencia de  $L(\hat{a}_1)/(n-1)$  para distintos valores de  $\beta$  ( $\alpha = 5\%$ )

sobre el residuo correspondiente a  $y_1$ ,  $\hat{a}_{1t}^*$ , obtenido del modelo evaluado bajo la hipótesis nula,  $H_0 : \theta = 1$ . Como es de esperar, a medida que el verdadero valor parámetro  $\theta$  se aleja de su valor bajo la hipótesis nula, la probabilidad de rechazarla tiende a uno tanto más rápido cuanto mayor sea el tamaño muestral. En cuanto a la potencia del contraste realizado sobre  $\hat{a}_{2t}$ , que se corresponde con la variable  $y_2$  en el modelo, la probabilidad de rechazo prácticamente no se mueve de su valor nominal del 5%, como se observa en el cuadro 5.2. Por último, la tabla 5.3 muestra que el contraste conjunto también es potente cuando el estadístico se calcula sobre los dos residuos del modelo.

En el segundo experimento se estudia la potencia del contraste ante un error en la especificación de la relación de cointegración. Nótese que el estadístico (5.6) puede utilizarse también para contrastar si un determinado valor de la relación de cointegración es válido, de forma que sería deseable que la probabilidad de rechazar la hipótesis nula de cointegración fuera alta cuando este valor impuesto difiera del verdadero. En el caso bivalente considerado, sabiendo que existe una relación de cointegración entre  $y_1$  e  $y_2$ , siendo  $\beta$  el verdadero valor de ésta, se quiere contrastar si  $\alpha = \beta + \delta$  ( $\delta \neq 0$ ) puede considerarse también como relación de cointegración. De (5.3) se obtiene que el residuo exacto,  $\hat{a}_{1t}$ , se puede expresar como  $\hat{a}_{1t} = a_{1t} - \delta y_{2t} + \alpha a_{2t}$  donde  $y_2$  es una variable  $I(1)$ , de forma que la no estacionaridad de  $\hat{a}_{1t}$  se hará más evidente cuanto mayor sea el valor absoluto del sesgo,  $\delta$ . La figura 5.1 muestra la función potencia de  $L/(n-1)$  construido sobre los residuos exactos  $\hat{a}_{1t}$  imponiendo distintos valores de la relación de cointegración y de la correlación entre las perturbaciones del modelo. Como era de esperar, cuanto más se aleje el valor de  $\hat{\beta}$  de su valor verdadero, mayor será la probabilidad de rechazar la

Cuadro 5.4: Estimaciones MVE del modelo VIMA(1,1) cointegrado

	$\theta_{11} = 1$	$\theta_{12} = -1,5$	$\theta_{21} = 0$	$\theta_{22} = 0$
Media	0.9974	-1.5017	0.0064	0.0068
Err. est.	(0.0664)	(0.0900)	(0.0432)	(0.0567)
	$\sigma_{11} = 1$	$\sigma_{12} = -0,5714$	$\sigma_{21} = -0,5714$	$\sigma_{22} = 0,5714$
Media	0.9930	-0.5756	-0.5756	0.5717
Err. est.	(0.1432)	(0.0950)	(0.0950)	(0.0812)

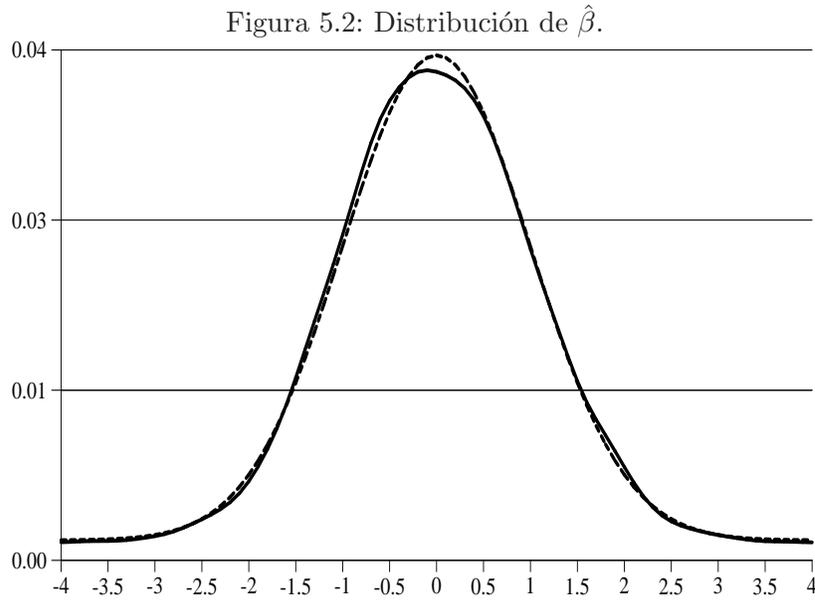
hipótesis nula de cointegración. La correlación entre las perturbaciones parece no afectar a la potencia del contraste, lo que es de esperar teniendo en cuenta la propiedad de invarianza sobre la que está construido.

### Estimación de un modelo VIMA(1,1) cointegrado

En la identificación de un proceso cointegrado se pueden considerar dos etapas definidas implícitamente en la ecuación (5.4) que define la estructura del proceso VIMA(1,1) cointegrado. En primer lugar, partiendo de un VIMA(1,1) general tal que  $\nabla \mathbf{y}_t = \mathbf{a}_t - \Theta \mathbf{a}_{t-1}$ , se deberían de encontrar aquellas filas de la matriz  $\Theta$  estadísticamente iguales a cero que indiquen la partición del vector de series temporales  $\mathbf{y}_t$ . Este primer paso se puede realizar mediante un contraste de significación sobre los parámetros del modelo VIMA de la forma habitual, por lo que es importante cerciorarse de que las estimaciones de los parámetros del modelo son consistentes y asintóticamente normales incluso bajo la presencia de raíces MA unitarias. Como muestra el cuadro 5.4, el método de máxima verosimilitud exacta proporciona estimaciones consistentes de los parámetros.

La figura 5.3 muestra la función de densidad de los estimadores de máxima verosimilitud estandarizados de  $\theta_{21}$  y  $\theta_{22}$ , utilizando como proceso generador de datos el descrito en el apartado anterior a partir de 5000 (línea continua). La línea discontinua muestra la función de densidad normal estándar. Por lo tanto, la presencia de raíces unitarias en el polinomio media móvil no implica la pérdida de normalidad en los estimadores del resto de parámetros del modelo, con lo que estadísticos t de significación habituales siguen siendo aplicables.

Una vez definida la partición del vector de series en  $\mathbf{y}_{1t}$  e  $\mathbf{y}_{2t}$  el segundo paso consiste, una vez reorganizado el vector a la manera (5.4), en contrastar si el bloque (1,1) de la matriz  $\Theta$  es igual a la matriz identidad. En caso de aceptar la hipótesis nula de cointegración, el bloque (1,2) de la matriz media móvil contendrá las relaciones de cointegración que serán igualmente normales. La figura 5.2 muestra la función de densidad de  $\hat{\beta}$  estan-



darizada (el elemento 1,2 de la matriz MA) en el ejemplo anterior (línea continua) y la distribución normal estándar (línea discontinua). El estimador de máxima verosimilitud de este parámetro es consistente y asintóticamente normal, por lo que se puede realizar inferencia sobre la relación de cointegración utilizando los procedimientos tradicionales.

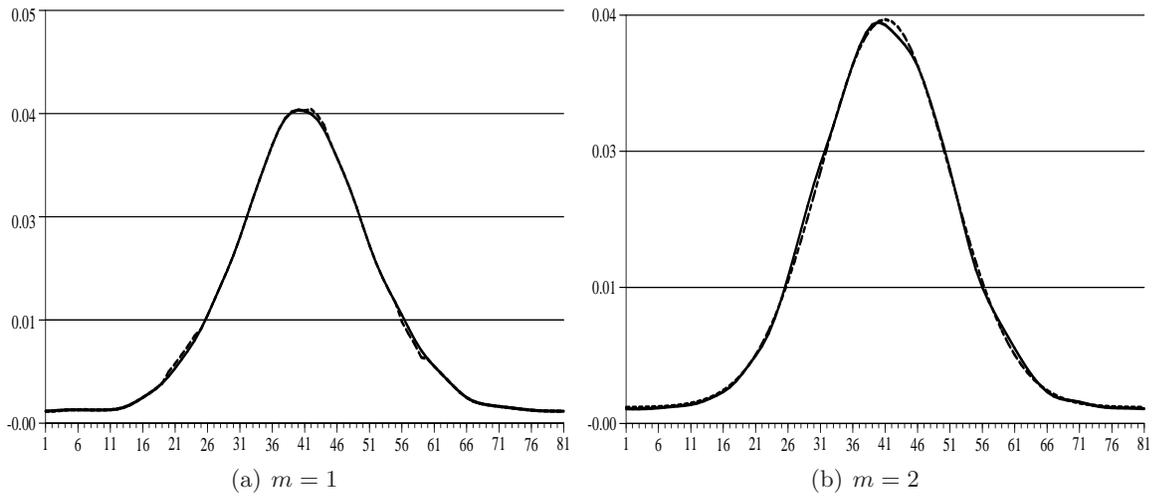
### 5.3. Cointegración en modelos VIMA(1,q)

#### 5.3.1. El modelo VIMA(1,1) cointegrado

El modelo triangular descrito en el apartado anterior precisa de una partición del vector de variables en dependientes y explicativas para detectar la presencia de cointegración. Sin embargo, es posible encontrar relaciones de cointegración cuando todas las variables están interrelacionadas entre sí. Como ilustración, se considera el modelo VIMA(1,1),

$$\nabla \mathbf{y}_t = (\mathbf{I}_m - \Theta B) \mathbf{a}_t,$$

en el que existirá cointegración si todas las series son integradas y existe al menos una raíz unitaria en el polinomio media móvil. Diversas estructuras de la matriz de coeficientes  $\Theta$  dan lugar a raíces unitarias, pero no todas indican la presencia de cointegración. Si esta matriz tiene estructura escalar  $\Theta = \theta \mathbf{I}_m$ , como la considerada en el capítulo 3, el

Figura 5.3: Distribución de  $\hat{\theta}_{21}$  y  $\hat{\theta}_{22}$ .

polinomio media móvil tendrá  $m$  raíces reales e iguales a  $\theta^{-1}$ , por lo que el polinomio será no invertible si  $\theta = 1$ , lo que implica una sobrediferenciación de todo el vector  $\mathbf{y}_t$  y no cointegración. Igualmente, si la matriz es diagonal,  $\Theta = \text{diag}(\theta_1, \dots, \theta_m)$ , el polinomio tendrá  $m$  raíces reales y distintas, con lo que una raíz unitaria implicaría  $\theta_j = 1$  o la sobrediferenciación de la  $j$ -ésima variable. Por lo tanto, existirá cointegración en un modelo VIMA(1,1) cuando el polinomio media móvil sea no invertible y la matriz de parámetros sea triangular por bloques o cuadrada. El primer caso se corresponde con la estructura triangular considerada en el apartado anterior, mientras que en este se considera el segundo.

En un modelo VIMA(1,1) con matriz de parámetros cuadrada existirán tantas relaciones de cointegración como raíces unitarias reales tenga el polinomio MA matricial. Si existen  $m_1 < m$  raíces unitarias reales, entonces  $m_1$  de los autovalores de  $\Theta$  serán iguales a uno, suponiendo el resto menores que uno en valor absoluto. Agrupando los autovalores iguales a uno en la matriz diagonal  $\Lambda_1$  y los  $m_2 = m - m_1$  restantes en  $\Lambda_2$ , se puede expresar la matriz de parámetros MA como  $\Theta = \mathbf{P}\Lambda\mathbf{Q}$  donde  $\Lambda = \text{diag}\{\Lambda_1, \Lambda_2\}$ ,  $\mathbf{P} = [\mathbf{p}_1|\mathbf{p}_2]$  es la matriz  $m \times m$  de autovectores tal que las  $m_1$  columnas de  $\mathbf{p}_1$  contienen los autovectores correspondientes a los autovalores de  $\Theta$  recogidos en  $\Lambda_1$  y  $\mathbf{p}_2$  el resto, y  $\mathbf{Q} = [\mathbf{q}_1|\mathbf{q}_2]'$  es una matriz tal que  $\mathbf{Q} = \mathbf{P}^{-1}$ . Utilizando la descomposición de Beveridge-Nelson, el modelo VIMA(1,1) se puede reescribir también de la siguiente manera

$$\nabla \mathbf{y}_t = \Theta(1)\mathbf{a}_t + \nabla \Theta \mathbf{a}_t, \quad \Theta(1) = \mathbf{p}_1(\mathbf{I}_{m_1} - \Lambda_1)\mathbf{q}'_1 + \mathbf{p}_2(\mathbf{I}_{m_2} - \Lambda_2)\mathbf{q}'_2, \quad (5.8)$$

donde la matriz de ganancias  $\Theta(1)$  tendrá rango  $m_2 < m$  en el caso en que el modelo tenga  $m_1$  raíces unitarias ( $\Lambda_1 = \mathbf{I}_{m_1}$ ), lo que implica que  $\Theta(1) = \mathbf{p}_2(\mathbf{I}_{m_2} - \Lambda_2)\mathbf{q}'_2$ . Así, integrando en la expresión anterior, el vector  $\mathbf{y}_t$  se descompone en una parte permanente que acompaña a la matriz de ganancias, la tendencia estocástica, y una parte transitoria. De esta forma, si existen  $m_1$  raíces unitarias, el sistema tiene  $m_1$  relaciones de cointegración definidas como aquel conjunto de combinaciones lineales de  $\mathbf{y}_t$  que anulan la tendencia estocástica, y que vendrán definidas por las columnas de  $\mathbf{q}_1$ . Nótese que premultiplicando (5.8) por  $\mathbf{q}'_1$ , y teniendo en cuenta que  $\mathbf{q}'_1\mathbf{p}_2 = \mathbf{0}$ , que  $\mathbf{q}'_1\mathbf{p}_1 = \mathbf{I}_{m_1}$  y que  $\mathbf{q}'_1\Theta = \mathbf{q}'_1$ , entonces  $\nabla\mathbf{q}'_1\mathbf{y}_t = \nabla\mathbf{q}'_1\mathbf{a}_t$ , siendo por lo tanto  $\mathbf{q}'_1\mathbf{y}_t$  un vector de variables  $I(0)$  y  $\mathbf{q}_1$ , las filas de  $\mathbf{Q}$  correspondientes a los autovalores unitarios de  $\Theta$ , forman el conjunto de relaciones de cointegración. Al igual que en el modelo de tendencias comunes, de (5.8) se puede comprobar que la presencia de  $m_1$  raíces unitarias (o relaciones de cointegración) implica la existencia de  $m_2$  tendencias comunes, definidas como  $\sum_{i=1}^{t-1} \mathbf{q}'_2\mathbf{a}_i$ . En este caso, cada tendencia estocástica se define a partir de una combinación lineal de las perturbaciones de todas las variables del modelo, mientras que en el caso anterior, es la acumulación sólo de los errores de la parte exógena del modelo la que da lugar a las tendencias que afectan a todo el sistema.

Si no existen raíces unitarias, de forma que los elementos de  $\Lambda_1$  son estrictamente menores que 1 en valor absoluto, entonces la matriz de ganancias se puede descomponer como  $\Theta(1) = \mathbf{p}_1(\mathbf{I}_{m_1} - \Lambda_1)\mathbf{q}'_1 + \mathbf{p}_2(\mathbf{I}_{m_2} - \Lambda_2)\mathbf{q}'_2$  y  $\Theta = \mathbf{p}_1\Lambda_1\mathbf{q}'_1 + \mathbf{p}_2\Lambda_2\mathbf{q}'_2$  tal que, dada esta estructura y premultiplicando el modelo (5.8) por  $\mathbf{q}'_1$  se obtiene

$$\nabla\mathbf{q}'_1\mathbf{y}_t = (\mathbf{I}_{m_1} - \Lambda_1 B)\mathbf{q}'_1\mathbf{a}_t,$$

y el vector de combinaciones lineales  $\mathbf{q}'_1\mathbf{y}_t$  se expresa como un proceso VIMA(1,1) invertible. A partir de la expresión anterior, se puede derivar un contraste para la hipótesis nula de cointegración, donde el problema a contrastar es  $H_0 : \Lambda_1 = \mathbf{I}_{m_1}$  frente a  $H_1 : \Lambda_1 \neq \mathbf{I}_{m_1}$ . El contraste se puede realizarse aplicando el contraste de no invertibilidad derivado en el capítulo 3 sobre  $\hat{\mathbf{q}}'_1\mathbf{y}_t$ . Una forma alternativa y más directa de realizar el contraste consiste en utilizar la corrección paramétrica propuesta en el capítulo 3. Esta corrección implica calcular el estadístico básico de no invertibilidad sobre el vector de  $m$  residuos exactos del modelo VIMA(1,1), evaluados bajo la hipótesis nula, esto es, a partir de

$$\nabla\mathbf{y}_t = (\mathbf{I}_m - \hat{\Theta}B)\mathbf{a}_t, \quad \hat{\Theta} = \hat{\mathbf{p}}_1\hat{\mathbf{q}}'_1 + \hat{\mathbf{p}}_2\hat{\Lambda}_2\hat{\mathbf{q}}'_2,$$

donde la matriz  $\hat{\Theta}$  se ha reconstruido a partir de los autovalores y autovectores de su estimación de máxima verosimilitud bajo la hipótesis alternativa. De esta forma, el pro-

cedimiento para realizar el contraste de cointegración es el siguiente: En primer lugar se estima el modelo VIMA(1,1) sin imponer restricciones y se observan sus raíces. Si existen  $m_1$  raíces reales potencialmente unitarias, éstas se deberían manifestarse como autovalores unitarios de la matriz media móvil estimada. Se eligen los  $m_1$  mayores autovalores de la matriz, se impone su valor a uno y se reconstruye la matriz  $\hat{\Theta}$  de la manera propuesta. A partir de ella se obtienen los residuos exactos del modelo con los que se calcula el estadístico de contraste de la manera habitual.

### 5.3.2. El modelo VIMA(1,q) cointegrado

El procedimiento descrito en el apartado anterior se puede extender a modelos más generales VARIMA(p,1,q+1). Sin pérdida de generalidad se considera el modelo VIMA(1,q+1) definido como

$$\nabla \mathbf{y}_t = \Theta(B)\mathbf{a}_t, \quad \mathbf{a}_t \sim N(\mathbf{0}, \Omega_a),$$

donde  $\Theta(B) = (\mathbf{I}_m - \Theta_1 B - \dots - \Theta_{q+1} B^{q+1})$  es un polinomio matricial de orden  $q+1$ . Al igual que en el caso anterior, la existencia de cointegración exige un tipo especial de no invertibilidad, por ejemplo  $\Theta(B) \neq (\mathbf{I} - \Theta_* B)\Theta^*(B)$  con  $\Theta_* = \text{diag}(\theta_1^*, \dots, \theta_m^*)$  de tal forma que  $\theta_j^*$  implique estacionariedad de algunas variables. Un caso extremo de esta situación se daría si  $\Theta(B) = \nabla \Theta^*(B)$ , implicando que todas las variables serían estacionarias. Utilizando la descomposición de Beveridge-Nelson el polinomio media móvil se puede escribir como  $\Theta(B) = \Theta(1) + \Theta^*(B)\nabla$  donde se define el polinomio matricial  $\Theta^*(B) = \Theta_0^* + \Theta_1^* B + \dots + \Theta_q^* B^q$  con las matrices  $\Theta_j^* = (\Theta_{j+1} + \dots + \Theta_{q+1})$ . De esta forma el modelo VIMA se puede expresar alternativamente como

$$\nabla \mathbf{y}_t = \Theta(1)\mathbf{a}_t + \Theta_0^* \nabla \mathbf{a}_t + \dots + \Theta_q^* \nabla \mathbf{a}_{t-q}. \quad (5.9)$$

Utilizando una argumentación análoga a la del apartado anterior, si existe al menos una raíz unitaria, la matriz de ganancias,  $\Theta(1) = \mathbf{I}_m - \Theta_0^*$ , no tendrá rango pleno, con lo que  $m_1 < m$  de sus autovalores serán iguales a cero. De forma similar, se puede descomponer esta matriz como  $\Theta(1) = \mathbf{p}_1(\mathbf{I}_{m_1} - \Lambda_1)\mathbf{q}'_1 + \mathbf{p}_2(\mathbf{I}_{m_2} - \Lambda_2)\mathbf{q}'_2$ , lo que implica que  $\Theta_0^* = \mathbf{p}_1\Lambda_1\mathbf{q}'_1 + \mathbf{p}_2\Lambda_2\mathbf{q}'_2$  donde  $\Lambda_1$  contiene los autovalores potencialmente unitarios. Por lo tanto, si  $\Lambda_1 = \mathbf{I}_{m_1}$ ,  $\mathbf{q}_1$  se interpreta como la matriz de  $m_1$  vectores de cointegración que anula las  $m_2$  tendencia estocástica que acompañan a la matriz de ganancias. Premultiplicando (5.9) por  $\mathbf{q}'_1$ , y teniendo en cuenta que  $\mathbf{q}'_1\Theta_0^* = \mathbf{q}'_1$  se

obtiene el siguiente modelo para la relación de cointegración

$$\nabla \mathbf{q}'_1 \mathbf{y}_t = \nabla \mathbf{q}'_1 \mathbf{a}_t + \nabla \mathbf{q}'_1 \Theta_1^* \mathbf{a}_{t-1} + \dots + \nabla \mathbf{q}'_1 \Theta_q^* \mathbf{a}_{t-q}.$$

Si no existe cointegración, de forma que  $\Lambda_1 \neq \mathbf{I}_{m_1}$ , entonces  $\mathbf{q}'_1 \Theta(1) = (\mathbf{I}_{m_1} - \Lambda_1) \mathbf{q}'_1$  y  $\mathbf{q}'_1 \Theta_0^* = \Lambda_1 \mathbf{q}'_1$  con lo que, de (5.9) se obtiene el siguiente modelo bajo el supuesto de no cointegración:

$$\nabla \mathbf{q}'_1 \mathbf{y}_t = (\mathbf{I}_{m_1} - \Lambda_1 B) \mathbf{q}'_1 \mathbf{a}_t + \nabla \mathbf{q}'_1 \Theta_1^* \mathbf{a}_{t-1} + \dots + \nabla \mathbf{q}'_1 \Theta_{q-1}^* \mathbf{a}_{t-q+1}.$$

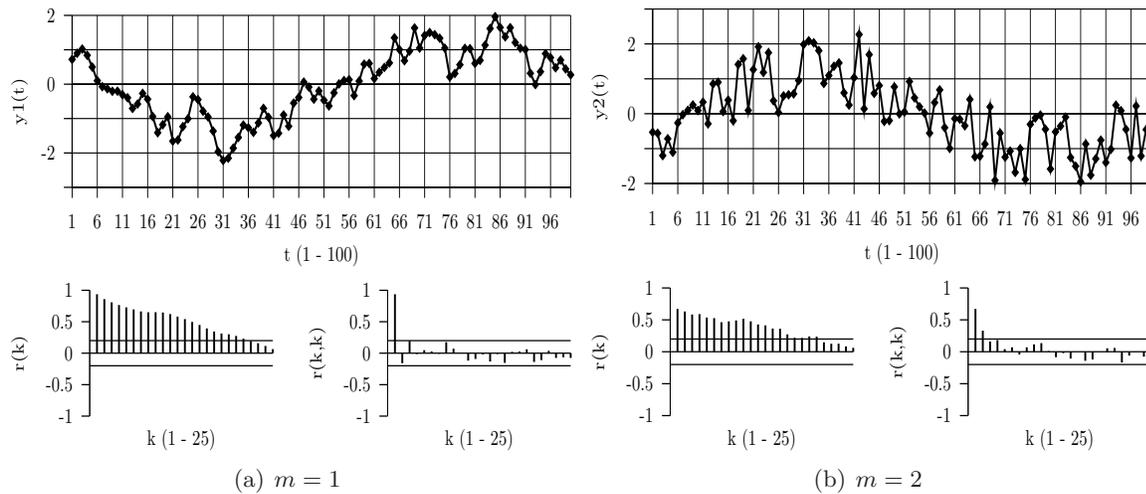
De aquí surge de forma natural el contraste de cointegración donde el problema a contrastar es  $H_0 : \Lambda_1 = \mathbf{I}_{m_1}$  frente a  $H_1 : \Lambda_1 \neq \mathbf{I}_{m_1}$ . Siguiendo la aproximación univariante de Tam y Reinsel (1997) o Leybourne y McCabe (1994), utilizando estimaciones consistentes de los parámetros, se podría filtrar la serie para reducir el modelo anterior a un VIMA(1,1) sobre el que contrastar la hipótesis nula utilizando el contraste básico de no invertibilidad. Sin embargo, un procedimiento más directo consiste en realizar el contraste sobre los  $m$  residuos exactos del modelo, evaluados bajo la hipótesis nula. Dado que el problema de contraste no está planteado directamente sobre ningún parámetro del modelo, en este caso es más complicado evaluar el modelo bajo la hipótesis nula. El procedimiento sugerido consiste en identificar aquellos autovalores de la matriz de ganancias potencialmente iguales a cero y, una vez elegidos, imponer su valor en cero y reconstruir  $\Theta(1) = \mathbf{p}_2(\mathbf{I}_{m_2} - \Lambda_2) \mathbf{q}'_2$ . El resto de matrices del polinomio MA(q) se recuperan de forma recursiva a partir de la definición de las matrices  $\Theta_j^*$ . Si el modelo contiene parte autorregresiva estacionaria, el contraste se realizaría igualmente sobre los residuos exactos del modelo. Tanto la hipótesis nula a contrastar como la forma de obtener los residuos bajo la hipótesis nula no varían con respecto al procedimiento descrito.

### 5.3.3. Estudio Monte Carlo

#### Ejemplo ilustrativo: El modelo VIMA(1,1) cointegrado

Como ejemplo ilustrativo la figura 5.4 muestra la realización del siguiente proceso IMA(1,1) bivalente con  $n = 100$ ,

$$\begin{pmatrix} \nabla y_{1t} \\ \nabla y_{2t} \end{pmatrix} = \begin{pmatrix} a_{1t} \\ a_{2t} \end{pmatrix} - \begin{pmatrix} -0,2311 & -1,5 \\ 0,6 & 1,7311 \end{pmatrix} \begin{pmatrix} a_{1,t-1} \\ a_{2,t-2} \end{pmatrix}, \quad \begin{pmatrix} a_{1t} \\ a_{2t} \end{pmatrix} \sim N \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0,5 \\ 0,5 & 1 \end{pmatrix} \right).$$

Figura 5.4: Gráficos, acf y pacf de  $y_1$  e  $y_2$ .

Las raíces del polinomio media móvil del modelo son 1 y 2 respectivamente y los autovalores de la matriz media móvil  $\Theta$  son, por lo tanto, 1 y 0,5, lo que indica que existe una relación de cointegración.

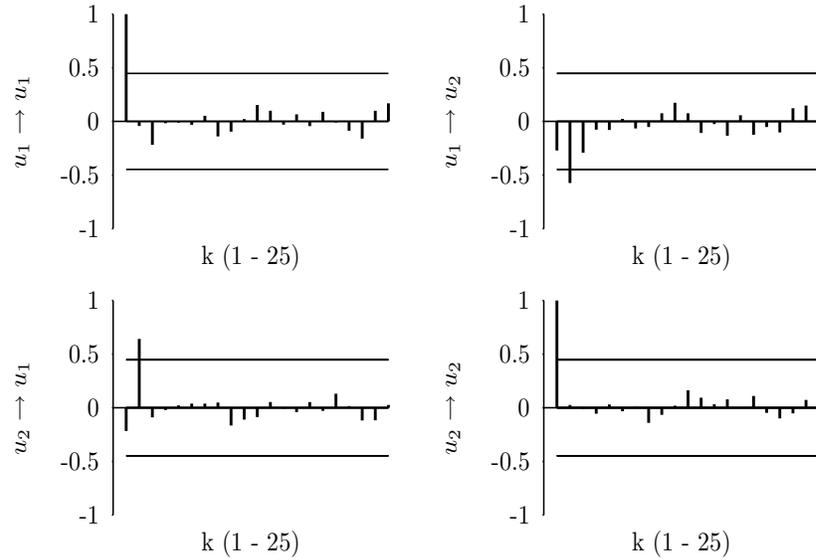
Cuadro 5.5: Modelos univariantes para  $Y_1$  e  $Y_2$ .

Name	$\hat{\theta}$	$\hat{\sigma}_a$	llik	Q(24)	$\hat{\mu}_a$
$Y_1$	-0.2080248 (0.098275)	4.034312	-209.54	24.21 (0.550508)	-0.021298 (1.998378)
$Y_2$	0.6806293 (0.073777)	5.052351	-220.97	15.29 (0.087949)	0.003491 (2.236474)

Errores estándar entre paréntesis.

El modelo IMA(1,1) univariante ajusta bien la dinámica univariante de ambas series, como muestran los resultados de las estimaciones de estos modelos de recogidas en el cuadro 5.5, si bien los coeficientes de correlación cruzada entre los residuos de ambos modelos univariantes sugiere que existe algún tipo de correlación entre ambos. Dado que el modelo IMA(1,1) proporciona un buena ajuste para ambas series, se prueba la estimación conjunta de un modelo IMA(1,1) bivalente. Los resultados de la estimación de máxima verosimilitud exacta son los siguientes (entre paréntesis figuran los errores

Figura 5.5: Correlaciones cruzadas entre los residuos de los modelos univariantes.



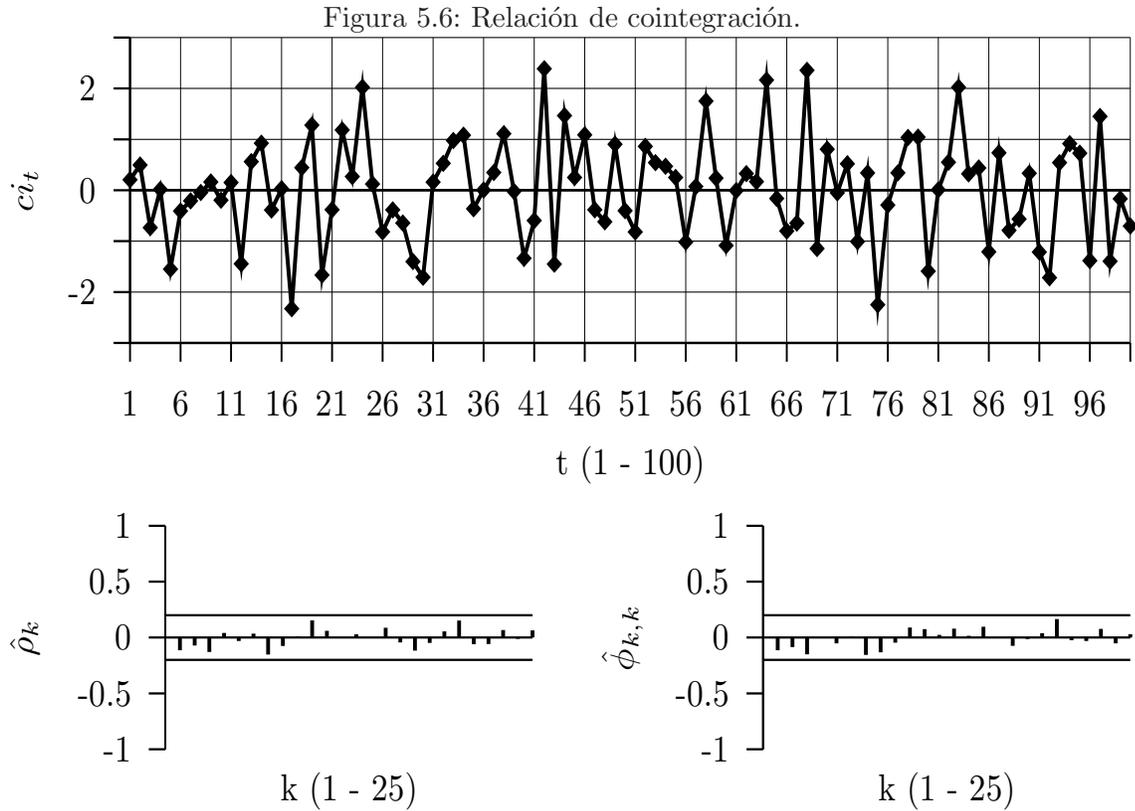
estándar):

$$\hat{\Theta}_1 = \begin{pmatrix} -0,1436 & -1,5283 \\ 0,0657 & 0,0634 \\ 0,5563 & 1,7434 \\ 0,0331 & 0,0320 \end{pmatrix} \quad \hat{\Omega}_a = \begin{pmatrix} 0,9808 & \\ 0,6773 & 1,2049 \end{pmatrix} \quad \hat{\Theta}(1) = \begin{pmatrix} 1,1436 & 1,5283 \\ -0,5563 & -0,7434 \end{pmatrix}$$

$$\text{llik} = -270,02 \quad Q(24) = 87,16(0,20) \quad \hat{\mu}_{a_1} = -0,015(0,986) \quad \hat{\mu}_{a_2} = -0,018(1,093)$$

El polinomio estimado tiene dos raíces reales e iguales a 1 y 1,6671 respectivamente. Además, los autovalores de la matriz de ganancias estimada son 0,4 y 0,0, lo que indica la presencia de una relación cointegración en el sistema. La fila de la inversa de la matriz de autovectores de  $\hat{\Theta}(1)$  correspondiente al autovalor cero, convenientemente normalizada, proporciona la siguiente relación de cointegración  $ci_t = Y_{2t} + 0,4864420Y_{1t}$ . Esta combinación lineal entre las variables del sistema será una relación de cointegración si es estacionaria, lo que muestran tanto sus funciones de autocorrelación (gráfico 5.6), como el estadístico de contraste de no invertibilidad sobre  $\nabla ci_t$ , cuyo valor es  $L(ci) = 0,0747$  con p-valor igual a 0,7398, no rechazando la hipótesis nula de no invertibilidad tanto al nivel de significación del 5% como del 1%. Este mismo resultado se puede obtener realizando el contraste conjuntamente sobre los dos residuos exactos del modelo VIMA(1,1) evaluados bajo la hipótesis nula imponiendo el valor del mayor autovalor de  $\Theta$  igual a 1. Así calculado, el valor del estadístico bivalente es de 0,0880016 y su p-valor

es igual a 0,7974, con el mismo resultado, la presencia de una raíz unitaria y la validez de esta relación de cointegración.



### Experimentos Monte Carlo

Para evaluar la potencia del contraste de cointegración se ha utilizado el siguiente proceso generador de datos:

$$\begin{pmatrix} \nabla y_{1t} \\ \nabla y_{2t} \end{pmatrix} = \begin{pmatrix} a_{1t} \\ a_{2t} \end{pmatrix} - \begin{pmatrix} \theta_{11} & \theta_{12} \\ \theta_{21} & \theta_{22} \end{pmatrix} \begin{pmatrix} a_{1,t-1} \\ a_{2,t-2} \end{pmatrix} \quad \theta_{ii} = 0,5[(\lambda_1 + \lambda_2) \pm \sqrt{(\lambda_1 - \lambda_2)^2 - 4\theta_{12}\theta_{21}}]$$

donde  $\lambda_1$  y  $\lambda_2$  son las dos raíces del polinomio, que se suponen reales. Fijando  $\theta_{12} = -1,5$  y  $\theta_{21} = 0,6$  y la matriz de covarianzas entre los errores con varianzas unitarias y coeficiente de correlación entre ambas igual a  $\text{corr}(a_1, a_2) = 0,5$ , el experimento consiste en estudiar la potencia del contraste de cointegración, o de la presencia de una raíz unitaria,

ante distintos valores de la segunda. La tabla 5.6 muestra los resultados del experimento, calculando el estadístico con la corrección paramétrica descrita en el apartado 3. En primer lugar, la probabilidad de rechazo cuando una raíz es exactamente 1 está cercana a su valor nominal del 5% en todos los casos, y crece a medida que una de las raíces se aleja de uno. Sin embargo, la pendiente de la función potencia se va haciendo menor a medida que la segunda raíz ( $\lambda_2$ ) se acerca a 1. Al igual que en otros experimentos, la probabilidad de rechazo aumenta con el tamaño muestral.

En el segundo experimento Monte Carlo se evalúa la potencia del contraste en un modelo IMA(1,2) bivalente. Para ello, el proceso generador de datos que se utiliza es

$$\begin{pmatrix} \nabla y_{1t} \\ \nabla y_{2t} \end{pmatrix} = \left( \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \begin{pmatrix} -0,7593 & 0,7 \\ -2 & 1,6593 \end{pmatrix} B \right) \left( \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \begin{pmatrix} \theta_{11}^{(2)} & \theta_{12}^{(2)} \\ \theta_{21}^{(2)} & \theta_{22}^{(2)} \end{pmatrix} B \right) \begin{pmatrix} a_{1t} \\ a_{2t} \end{pmatrix}$$

donde se consideran únicamente raíces reales que se controlan de la misma manera que en el caso anterior para cada uno de los polinomios. El experimento centrará su interés en el segundo polinomio, siendo el primero estrictamente invertible con raíces reales e iguales a 5 y 1,4. La tabla 5.7 muestra los resultados de este experimento para  $n = 100$ . Los resultados son similares a los del experimento anterior. Bajo la hipótesis nula la probabilidad de rechazo es del 5% y aumenta a medida que una de las raíces se aleja de 1. El valor de la segunda raíz parece no afectar a la potencia del contraste siempre y cuando no esté muy cerca de 1, lo que daría lugar a una segunda raíz unitaria.

Cuadro 5.6: Experimento 1: Potencia del contraste de cointegración bivariente ( $\alpha = 5\%$ )

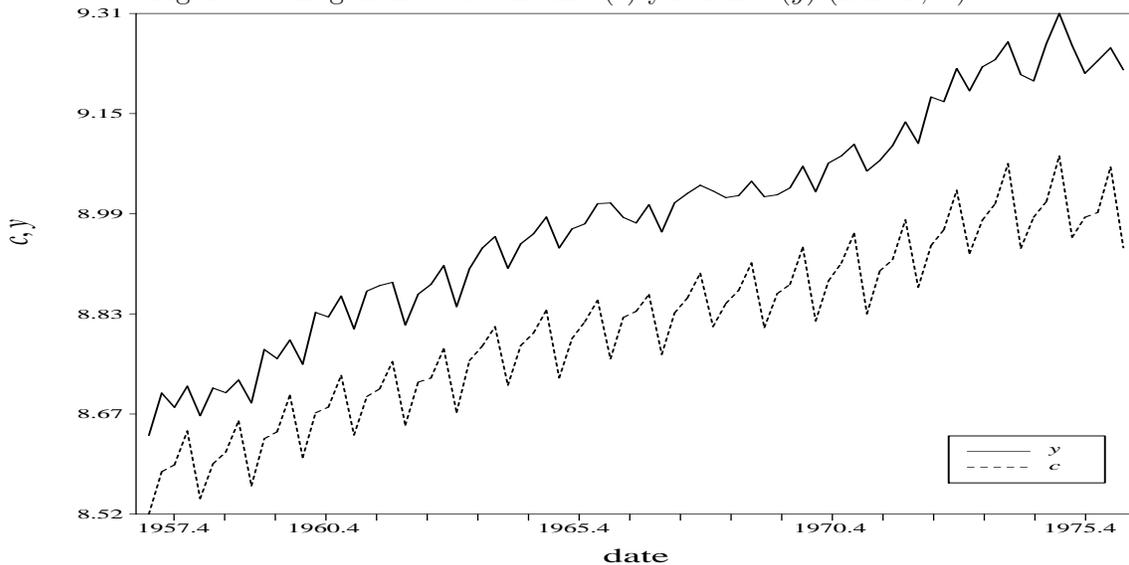
$\lambda_1, \lambda_2$	n=20							n=30						
	-0.9	-0.8	-0.5	0	0.5	0.8	0.9	-0.9	-0.8	-0.5	0	0.5	0.8	0.9
1.00	0.0470	0.0496	0.0458	0.0452	0.0414	0.0322	0.0390	0.0552	0.0446	0.0444	0.0432	0.0476	0.0424	0.0378
0.99	0.0470	0.0474	0.0460	0.0474	0.0446	0.0426	0.0402	0.0510	0.0548	0.0506	0.0468	0.0448	0.0380	0.0416
0.98	0.0560	0.0502	0.0512	0.0502	0.0434	0.0378	0.0404	0.0602	0.0616	0.0560	0.0640	0.0466	0.0438	0.0382
0.97	0.0542	0.0574	0.0514	0.0580	0.0468	0.0352	0.0392	0.0726	0.0620	0.0760	0.0818	0.0538	0.0388	0.0392
0.96	0.0632	0.0572	0.0604	0.0694	0.0498	0.0392	0.0412	0.0882	0.0868	0.0908	0.1160	0.0572	0.0462	0.0394
0.95	0.0658	0.0668	0.0736	0.0858	0.0532	0.0394	0.0370	0.1114	0.1044	0.1158	0.1496	0.0698	0.0506	0.0444
0.94	0.0712	0.0734	0.0864	0.1036	0.0604	0.0460	0.0448	0.1246	0.1400	0.1572	0.1972	0.0784	0.0634	0.0478
0.93	0.0974	0.0866	0.1022	0.1310	0.0562	0.0428	0.0422	0.1588	0.1586	0.1920	0.2460	0.0916	0.0686	0.0484
0.92	0.1016	0.1134	0.1120	0.1528	0.0642	0.0496	0.0440	0.1922	0.2060	0.2226	0.2882	0.1130	0.0906	0.0598
0.91	0.1166	0.1238	0.1362	0.1848	0.0714	0.0566	0.0520	0.2148	0.2252	0.2594	0.3296	0.1306	0.0946	0.0634
0.90	0.1386	0.1436	0.1614	0.2094	0.0788	0.0702	0.0566	0.2594	0.2614	0.2934	0.3816	0.1426	0.1104	0.0688
0.85	0.2272	0.2374	0.2782	0.3528	0.1378	0.1068	0.0676	0.4152	0.4194	0.4582	0.5316	0.2600	0.2152	0.1178
0.80	0.3396	0.3530	0.3826	0.4702	0.1994	0.1406	0.0992	0.5116	0.5368	0.5692	0.6444	0.3606	0.2986	0.1600
0.75	0.4174	0.4394	0.4692	0.5404	0.2748	0.1762	0.1388	0.5766	0.6008	0.6416	0.6914	0.4700	0.3438	0.2194
0.70	0.4714	0.4852	0.5280	0.6000	0.3676	0.1988	0.1856	0.6394	0.6474	0.6948	0.7418	0.5502	0.3942	0.2890
0.65	0.5128	0.5312	0.5654	0.6552	0.4300	0.2214	0.2310	0.6744	0.7110	0.7258	0.7766	0.6078	0.4320	0.3542
0.60	0.5690	0.5752	0.6226	0.6940	0.4706	0.2542	0.2692	0.7092	0.7322	0.7684	0.8138	0.6494	0.4608	0.4192

$\lambda_1, \lambda_2$	n=50							n=100						
	-0.9	-0.8	-0.5	0	0.5	0.8	0.9	-0.9	-0.8	-0.5	0	0.5	0.8	0.9
1.00	0.0526	0.0528	0.0512	0.0512	0.0508	0.0372	0.0342	0.0560	0.0534	0.0500	0.0514	0.0482	0.0434	0.0396
0.99	0.0574	0.0508	0.0582	0.0662	0.0506	0.0442	0.0352	0.0750	0.0774	0.0818	0.1058	0.0604	0.0530	0.0418
0.98	0.0756	0.0744	0.0818	0.0986	0.0558	0.0442	0.0364	0.1552	0.1640	0.1826	0.2516	0.0996	0.0876	0.0604
0.97	0.1128	0.1102	0.1224	0.1602	0.0690	0.0554	0.0416	0.2758	0.2830	0.3096	0.4146	0.1576	0.1564	0.1168
0.96	0.1534	0.1634	0.1774	0.2408	0.0896	0.0730	0.0486	0.3896	0.4012	0.4202	0.5234	0.2416	0.2218	0.1606
0.95	0.2020	0.2230	0.2422	0.3202	0.1248	0.1000	0.0660	0.4794	0.4930	0.5340	0.6076	0.3118	0.3018	0.2486
0.94	0.2646	0.2868	0.2988	0.3898	0.1454	0.1288	0.0776	0.5598	0.5712	0.6034	0.6800	0.4094	0.3812	0.3214
0.93	0.3148	0.3230	0.3672	0.4658	0.1934	0.1590	0.1070	0.6104	0.6354	0.6776	0.7442	0.4446	0.4468	0.3760
0.92	0.3794	0.3830	0.4258	0.5070	0.2420	0.2026	0.1320	0.6670	0.6808	0.7138	0.7776	0.5246	0.4944	0.4380
0.91	0.4252	0.4376	0.4782	0.5538	0.2622	0.2298	0.1428	0.7168	0.7246	0.7526	0.8138	0.5792	0.5688	0.4886
0.90	0.4662	0.4802	0.5168	0.6106	0.2962	0.2810	0.1692	0.7494	0.7646	0.7902	0.8356	0.6174	0.6048	0.5284
0.85	0.6398	0.6502	0.6680	0.7418	0.4834	0.4364	0.2806	0.8584	0.8608	0.8804	0.9156	0.7860	0.7596	0.6576
0.80	0.7214	0.7392	0.7550	0.8218	0.6106	0.5520	0.3772	0.9098	0.9112	0.9266	0.9538	0.8512	0.8274	0.7478
0.75	0.7746	0.7770	0.8164	0.8612	0.6820	0.6120	0.4630	0.9360	0.9318	0.9468	0.9604	0.8972	0.8650	0.8132
0.70	0.8204	0.8270	0.8428	0.8796	0.7468	0.6676	0.5500	0.9526	0.9576	0.9594	0.9714	0.9290	0.8894	0.8618
0.65	0.8274	0.8550	0.8800	0.8982	0.8004	0.6866	0.6344	0.9572	0.9636	0.9694	0.9778	0.9490	0.9130	0.9018
0.60	0.8482	0.8708	0.8834	0.9174	0.8228	0.7202	0.6826	0.9610	0.9712	0.9760	0.9820	0.9592	0.9254	0.9202

Cuadro 5.7: Potencia del contraste de cointegración para un MA(2)  $\alpha = 5\%$ 

$\lambda_1$	$\lambda_2 = -0,9$	$\lambda_2 = -0,5$	$\lambda_2 = 0,0$	$\lambda_2 = 0,5$	$\lambda_2 = 0,9$
1.00	0.0480	0.0536	0.0488	0.0432	0.0408
0.99	0.0604	0.0584	0.0516	0.0512	0.0348
0.98	0.0968	0.0926	0.0902	0.0716	0.0660
0.97	0.1446	0.1468	0.1378	0.1134	0.1050
0.96	0.2118	0.2132	0.1992	0.1750	0.1664
0.95	0.2836	0.2766	0.2740	0.2402	0.2290
0.90	0.5430	0.5652	0.5452	0.4748	0.4074
0.85	0.6960	0.7050	0.6948	0.6314	0.5886
0.80	0.7648	0.7790	0.7738	0.7150	0.6794
0.75	0.8104	0.8468	0.8304	0.7734	0.7100
0.70	0.8176	0.8766	0.8632	0.8114	0.7594
0.65	0.8048	0.8906	0.8854	0.8346	0.8086
0.60	0.8062	0.9032	0.9024	0.8540	0.8230

## 5.4. Aplicación

Figura 5.7: Logaritmo del consumo ( $c$ ) y la renta ( $y$ ) (DHSY,78).

En este apartado se ilustra el procedimiento de contraste de cointegración utilizando los datos de consumo de bienes no duraderos ( $c$ ) y renta personal disponible ( $y$ ) en el Reino Unido a precios constantes analizados por Davidson et al. (1978). Como se observa en la figura 5.7, los logaritmos de ambas variables entre el primer trimestre de 1957 y el

Cuadro 5.8: Modelos univariantes

Variable	$\hat{\theta}$	$\hat{\Theta}$	$\hat{\sigma}_a$	llik	Q(17)
$y$	0.2021 (0.1198)	0.5586 (0.1083)	0.0212	164.67	16.19 (0.4899)
$c$	0.1614 (0.1176)	0.4388 (0.1129)	0.0101	216.25	12.72 (0.2449)

segundo trimestre de 1976 muestran un claro componente estacional y un crecimiento a largo plazo que, salvo en la última parte de la muestra, parece seguir una evolución conjunta. La tabla 5.8 muestra los resultados de la estimación del modelo univariante de líneas aéreas,  $\nabla\nabla_4 z_t = (1 - \theta B)(1 - \Theta B^4)a_t$ , para ambas series. En esta tabla se muestra, además de las estimaciones de los parámetros media móvil regular y estacional junto con sus errores estándar en paréntesis, el error estándar de la estimación, el valor del logaritmo de la función de verosimilitud y el estadístico de autocorrelación de Ljung-Box con el p-valor asociado. Dado que el modelo de líneas aéreas parece ofrecer un buen ajuste para ambas series individualmente, se prueba con este mismo modelo en la estimación conjunta, obteniendo las siguientes estimaciones de los parámetros del modelo multiplicativo  $\nabla\nabla_4 \mathbf{y}_t = (\mathbf{I}_2 - \Theta_1 B)(\mathbf{I}_2 - \Theta_4 B^4)\mathbf{a}_t$ ,

$$\hat{\Theta}_1 = \begin{pmatrix} 0,5194 & -0,8804 \\ (0,2357) & (0,4730) \\ -0,1890 & 0,3942 \\ (0,1242) & (0,2527) \end{pmatrix} \quad \hat{\Theta}_4 = \begin{pmatrix} 0,5614 & -0,0120 \\ (0,2416) & (0,5036) \\ 0,0906 & 0,2634 \\ (0,1235) & (0,2587) \end{pmatrix}$$

$$\hat{\Omega}_a = \begin{pmatrix} 4,07 & \\ 1,33 & 0,96 \end{pmatrix} \times 10^{-4} \quad \text{llik} = 435,21 \quad Q(18) = 65,76(0,29)$$

Entre paréntesis figuran los errores estándar excepto para el estadístico de correlación de Ljung-Box que indica el p-valor asociado. Aunque la utilización de un modelo multiplicativo en el análisis multivariante parecería desaconsejable, ya que ambos polinomios matriciales generalmente no son conmutativos, en este caso proporciona un buen ajuste y una representación parsimoniosa de ambas series. Se ha preferido trabajar con la estructura típica del modelo de líneas aéreas conservando como primer factor el polinomio regular, aunque los resultados de la estimación del modelo invirtiendo el orden de los polinomios no difiere sustancialmente. Como se puede observar, algunos de los coeficientes son no significativos por lo que, eliminando secuencialmente los parámetros estadística-

Cuadro 5.9: Contrastes de no invertibilidad.

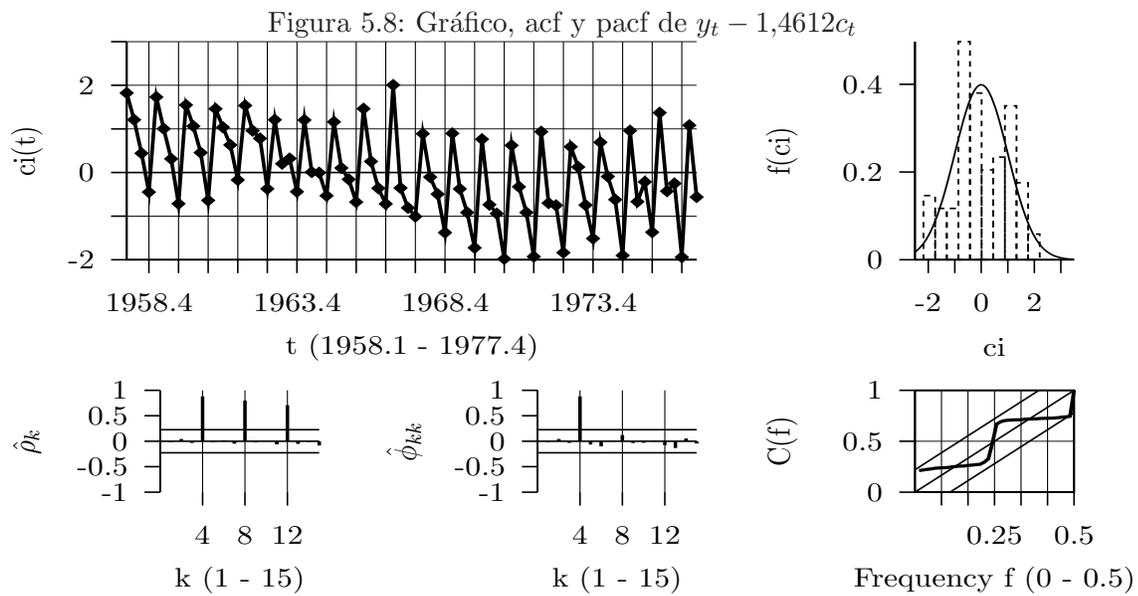
	$\hat{a}_1$	$\hat{a}_2$	conjunto
test	0.2380	0.083	0.1479
p-valor	0.2077	0.695	0.4776

mente iguales a cero, queda el modelo restringido  $\nabla \nabla_4 \mathbf{y}_t = (\mathbf{I}_2 - \Theta_1^R B)(\mathbf{I}_2 - \Theta_4^R B^4) \mathbf{a}_t$

$$\hat{\Theta}_1^R = \begin{pmatrix} 0,8330 & -1,4612 \\ (0,1220) & (0,2975) \\ 0 & 0 \\ (-) & (-) \end{pmatrix} \quad \hat{\Theta}_4^R = \begin{pmatrix} 0,4689 & 0 \\ (0,1598) & (-) \\ 0 & 0,3723 \\ (-) & (0,1605) \end{pmatrix}$$

$$\hat{\Omega}_a = \begin{pmatrix} 4,30 \\ 1,346 & 0,897 \end{pmatrix} \times 10^{-4} \quad \text{llik} = 435,34 \quad Q(18) = 78,31(0,69)$$

Nótese que la segunda fila de la matriz media móvil del polinomio regular es igual a cero y que el elemento (1,1) tiene un valor elevado. Por su parte, la matriz del polinomio estacional es diagonal con sus elementos inferiores a uno, con lo que el polinomio media móvil estacional es invertible. A la vista de los resultados descritos en los apartados anteriores, la estructura de la parte regular parece indicar la presencia de una relación de cointegración definida por  $c_i t = y_t - 1,4612c_t$  que, normalizada a la inversa, implica un coeficiente de cointegración de 0,684369 cercano al valor 0,74 que obtienen Davidson et al. (1978). Como muestra el gráfico 5.8 esta relación de cointegración tiene un componente estacional claro. El contraste de cointegración se puede realizar a partir de los residuos exactos del modelo evaluados bajo la hipótesis nula de cointegración,  $H_0 : \theta_{11}^{(1)} = 1$  frente a  $H_1 : \theta_{11}^{(1)} \neq 1$  donde  $\theta_{11}^{(1)}$  es el elemento (1,1) de  $\Theta_1^R$ . La tabla 5.9 muestra los estadísticos de contrastes de no invertibilidad para los dos residuos exactos del modelo, donde el contraste de cointegración tal y como se ha definido se plantea como el contraste de estacionariedad sobre  $\hat{a}_1$ , para el que se acepta la hipótesis nula, indicando la presencia de cointegración del tipo anteriormente descrito.



## Capítulo 6

# Conclusiones

La clase de modelos VARIMA es la extensión natural de su análoga univariante, y puede utilizarse para describir, predecir y corregir variables interrelacionadas. Esta representación parsimoniosa contiene como casos especiales a los modelos VAR, que se han utilizados extensamente para representar y contrastar relaciones de cointegración, y a la representación estructural. El análisis de cointegración en modelos VARIMA ha recibido escasa atención en la literatura, centrándose los procedimientos de estimación y contraste en la modelización VARMA parcialmente no estacionaria. Sin embargo, resulta razonable pensar que si el analista elabora un modelo VARIMA útil para los propósitos descritos, también se debería utilizar este modelo para realizar el análisis de cointegración, sin perjuicio de que se usen otros contrastes complementarios.

En esta tesis se ha presentado un procedimiento para contrastar la presencia de relaciones de cointegración en un modelo VARIMA general, que se basa en una extensión multivariante de los contrastes de no invertibilidad desarrollados en modelos ARIMA univariantes. Hay dos diferencias fundamentales entre el procedimiento propuesto y el contraste estructural análogo de nivel local multivariante determinista derivado por Nyblom y Harvey (2000). La primera se encuentra en la representación usada; mientras estos autores trabajan con la representación estructural, el contraste presentado se deriva a partir de su forma reducida VIMA con estructura escalar. Por otro lado, frente a la corrección no paramétrica propuesta por Nyblom y Harvey (2000), se propone una corrección paramétrica que aprovecha toda la información que proporciona el modelo VARIMA con el que se trabaja. Además, en el marco presentado el contraste puede expresarse de forma aproximada como un contraste univariante estacional, lo que permite utilizar los resultados bien establecidos en Tam y Reinsel (1997) tanto para derivar el contraste como para aproximar la distribución en muestras finitas del estadístico utilizando los procedimientos habituales de Imhof (1961) o Davies (1973).

Las principales contribuciones de esta tesis son las siguientes:

1. Se ha derivado el contraste de Nyblom y Harvey (2000) en el marco de un modelo VIMA(1,1) escalar a partir de la segunda derivada de su función de verosimilitud. Se han obtenido dos formas equivalentes del estadístico, una basada en la serie no estacionaria y otra en la serie diferenciada.
2. Se ha extendido el contraste para permitir correlación serial multivariante. La corrección paramétrica propuesta se basa en los residuos exactos del modelo VARMA y puede aplicarse también a modelos estructurales.
3. Se ha generalizado el contraste de no invertibilidad para detectar la presencia de raíces unitarias en el polinomio MA(1) matricial permitiendo estructuras más generales para la matriz de coeficientes. La corrección paramétrica propuesta se muestra de utilidad en este caso.
4. Se ha obtenido la extensión multivariante del contraste de no invertibilidad estacional de Tam y Reinsel (1998) y se han considerado otras formas de no invertibilidad estacional.
5. Se ha propuesto una estrategia para contrastar cointegración, que encajaría en la etapa de diagnóstico del procedimiento iterativo de elaboración de modelos de Box y Jenkins. Para la representación de cointegración basada en el modelo de regresión se obtiene no sólo un contraste de cointegración alternativo a los propuestos en la literatura, sino también un método de estimación consistente de la relación de cointegración.
6. Todos los estadísticos derivados siguen distribuciones no estándar que pueden aproximarse por la distribución IG. Para cada estadístico derivado se obtienen expresiones cerradas de sus dos primeros momentos.
7. Finalmente, se han desarrollado rutinas incorporadas en **Empiricus** para implementar la metodología de análisis propuesta que requiere de la identificación, estimación y diagnóstico de modelos VARMA generales.

Todas estas contribuciones se resumen en un artículo publicado, basado en el capítulo 4 (Gallego y Díaz, 2011) y dos más en elaboración. En el primero, que se corresponde con el capítulo 3 de la tesis, se describe la derivación del contraste de no invertibilidad, de su distribución y su potencia en muestras finitas y se propone una corrección paramétrica para contrastar no invertibilidad en modelos VARMA(p,1,q+1). En el segundo, se describe el procedimiento de identificación, estimación y contraste de cointegración en

modelos VARIMA descritos en el capítulo 5. Se ilustra cómo las tres representaciones habituales de un proceso cointegrado pueden acomodarse a una estructura VARIMA y el procedimiento para contrastar cointegración en una variedad de casos de interés.

Entre las líneas posibles de investigación a desarrollar en el futuro cabe destacar las siguientes: Mediante un estudio Monte Carlo se compararán las tres aproximaciones al análisis de cointegración resumidas por Johansen (2004) y la propuesta en esta tesis. También se comparará la potencia de los contrastes más utilizados en cada una de estas representaciones y el contraste de cointegración propuesto aquí. También se considerará la extensión del análisis de cointegración al caso estacional en la línea de las contribuciones de Ahn y Reinsel (1990), Johansen y Schaumburg (1999) o Busetti (2006). En cuanto a los contrastes de no invertibilidad, se estudiará el efecto que sobre ellos tiene la inclusión de variables de intervención. También se evaluará la capacidad predictiva de un modelo VARIMA cointegrado y se estudiará el efecto que sobre ésta pueda tener el reconocimiento de la presencia de cointegración en el modelo. Una última línea de investigación consistiría en estudiar los métodos de ajuste estacional basados en modelos VARIMA cointegrados y no cointegrados.



## Apéndice A

# Contrastes de cointegración

### A.1. La aproximación basada en regresión

#### A.1.1. Representación y estimación

Phillips (1991) y Phillips y Loretan (1991) estudian la estimación y el contraste de cointegración en un modelo triangular para el vector  $m \times 1$  de series temporales  $\mathbf{y}_t$  cuya estructura es la siguiente:

$$\mathbf{y}_{1t} = \mathbf{B}\mathbf{y}_{2t} + \mathbf{a}_{1t}, \quad \mathbf{a}_{1t} \sim N(\mathbf{0}, \boldsymbol{\Omega}_{11}), \quad (\text{A.1})$$

$$\nabla \mathbf{y}_{2t} = \mathbf{a}_{2t}, \quad \mathbf{a}_{2t} \sim N(\mathbf{0}, \boldsymbol{\Omega}_{22}), \quad (\text{A.2})$$

donde  $\mathbf{y}_{1t}$  y  $\mathbf{y}_{2t}$  son dos vectores de dimensiones  $m_1$  y  $m_2$  respectivamente ( $m = m_1 + m_2$ ) con errores  $\mathbf{a}_{1t}$  y  $\mathbf{a}_{2t}$  tal que  $E(\mathbf{a}_{1t}\mathbf{a}'_{2t}) = \boldsymbol{\Omega}_{12} \neq \mathbf{0}$  y  $\mathbf{B}$  una matriz  $m_1 \times m_2$  de parámetros a estimar. Tanto  $\mathbf{y}_{1t}$  como  $\mathbf{y}_{2t}$  son procesos integrados, de que la combinación lineal entre ellos  $\mathbf{z}_t = \mathbf{y}_{1t} - \mathbf{B}\mathbf{y}_{2t}$  definirá un conjunto de relaciones de cointegración si el vector de errores  $\mathbf{a}_{1t}$  es estacionario. En este caso, cada una de las filas de la matriz  $[\mathbf{I}_{m_1} \quad -\mathbf{B}]$  se interpretan como los  $m_1$  vectores de cointegración debidamente normalizados y  $\mathbf{a}_{1t}$  como las desviaciones de corto plazo con respecto a ellas. Este modelo sencillo se puede expresar como un proceso VAR(1) tanto en forma estructural como en forma reducida

$$\begin{bmatrix} \mathbf{I}_{m_1} & -\mathbf{B} \\ \mathbf{0} & \mathbf{I}_{m_2} \end{bmatrix} \begin{pmatrix} \mathbf{y}_{1t} \\ \mathbf{y}_{2t} \end{pmatrix} = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{m_2} \end{bmatrix} \begin{pmatrix} \mathbf{y}_{1,t-1} \\ \mathbf{y}_{2,t-1} \end{pmatrix} + \begin{pmatrix} \mathbf{a}_{1t} \\ \mathbf{a}_{2t} \end{pmatrix}, \quad (\text{A.3})$$

$$\begin{pmatrix} \mathbf{y}_{1t} \\ \mathbf{y}_{2t} \end{pmatrix} = \begin{bmatrix} \mathbf{0} & \mathbf{B} \\ \mathbf{0} & \mathbf{I}_{m_2} \end{bmatrix} \begin{pmatrix} \mathbf{y}_{1,t-1} \\ \mathbf{y}_{2,t-1} \end{pmatrix} + \begin{bmatrix} \mathbf{I}_{m_1} & \mathbf{B} \\ \mathbf{0} & \mathbf{I}_{m_2} \end{bmatrix} \begin{pmatrix} \mathbf{a}_{1t} \\ \mathbf{a}_{2t} \end{pmatrix}, \quad (\text{A.4})$$

denotando por  $\Phi_0$  la matriz de coeficientes de  $\mathbf{y}_t$  en la forma estructural (A.3) y  $\Phi_1$  a la matriz de coeficientes de  $\mathbf{y}_{t-1}$  en la forma reducida (A.4). En esta última, el error que afecta a  $\mathbf{y}_{1t}$  viene definido por la combinación lineal  $\mathbf{a}_{1t} + \mathbf{B}\mathbf{a}_{2t}$  que sigue siendo estacionario. Nótese que  $m_2$  de las  $m$  raíces de del polinomio VAR en forma reducida,  $|\mathbf{I}_m - \Phi_1 B| = 0$ , son iguales a uno lo que implica que  $\Phi(1) = \mathbf{I} - \Phi$  tiene rango  $m_1 < m$ . Esto último es inmediato de comprobar, ya que las  $m_2$  últimas filas de  $\Phi(1)$  correspondientes a  $\mathbf{y}_{2t}$  son exactamente cero. Utilizando una descomposición tipo Dickey-Fuller, este polinomio se puede expresar también como  $\Phi(B) = (1 - B)\mathbf{I}_m - \Pi B$ , donde  $\Pi = -\Phi(1)$  es la matriz de ganancias que, al tener rango  $m_1$ , puede descomponerse como producto de dos matrices de rango completo,  $\Pi = \mathbf{A}\mathbf{C}$ . En este caso sencillo  $\mathbf{A} = [-\mathbf{I}_{m_1} \mathbf{0}']'$  una matriz fija y  $\mathbf{C} = [\mathbf{I}_{m_1} \ -\mathbf{B}]$ , la matriz que contiene los vectores de cointegración. Esta descomposición del polinomio autorregresivo permite la representación de (5.1)-(5.2) como un modelo de corrección de error

$$\begin{pmatrix} \nabla \mathbf{y}_{1t} \\ \nabla \mathbf{y}_{2t} \end{pmatrix} = \begin{bmatrix} -\mathbf{I}_{m_1} \\ \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{I}_{m_1} & -\mathbf{B} \end{bmatrix} \begin{pmatrix} \mathbf{y}_{1,t-1} \\ \mathbf{y}_{2,t-1} \end{pmatrix} + \begin{pmatrix} \mathbf{v}_{1t} \\ \mathbf{v}_{2t} \end{pmatrix} \quad (\text{A.5})$$

con  $\mathbf{v}_{1t} = \mathbf{a}_{1t} + \mathbf{B}\mathbf{a}_{2t}$  y  $\mathbf{v}_{2t} = \mathbf{a}_{2t}$  y  $\mathbf{v}_t \sim N(\mathbf{0}, \Sigma)$ . A diferencia del modelo de corrección de error tradicional, la matriz  $\mathbf{A}$  no es un parámetro más a estimar, sino que está fijada en la definición de la partición de  $\mathbf{y}_t$ , de forma que la matriz  $\mathbf{B}$  a estimar entra de forma lineal en el modelo.

Estas tres representaciones alternativas dan lugar a tres estimadores distintos de la matriz de relaciones de cointegración. En primer lugar, el estimador de mínimos cuadrados multivariantes de  $\mathbf{B}$  en (A.1)

$$\hat{\mathbf{B}}' = \left( \sum_{t=1}^n \mathbf{y}_{2t} \mathbf{y}'_{2t} \right)^{-1} \left( \sum_{t=1}^n \mathbf{y}_{2t} \mathbf{y}'_{1t} \right) \equiv \mathbf{B}' + \left( \sum_{t=1}^n \mathbf{y}_{2t} \mathbf{y}'_{2t} \right)^{-1} \left( \sum_{t=1}^n \mathbf{y}_{2t} \mathbf{u}'_{1t} \right)$$

es consistente, pero su distribución límite obtenida por Phillips y Hansen (1990) tiene la siguiente expresión

$$n(\hat{\mathbf{B}}' - \mathbf{B}') \xrightarrow{d} \left( \int_0^1 \mathbf{W}_{m_2}(u) \mathbf{W}'_{m_2}(u) du \right)^{-1} \left( \int_0^1 \mathbf{W}_{m_2}(u) d\mathbf{W}'_{m_1}(u) + \Omega_{21} \right)$$

donde  $\mathbf{W}(u) = [\mathbf{W}'_{m_1}(u), \mathbf{W}'_{m_2}(u)]'$  es un proceso Browniano  $m$ -dimensional con matriz de covarianzas  $\Omega$ . Esta distribución asintótica de  $\hat{\mathbf{B}}$  se obtiene de las relaciones detalladas en Park y Phillips (1989) notando que  $\sum_{t=1}^n \mathbf{y}_{2t} \mathbf{u}'_{1t} = \sum_{t=1}^n [\mathbf{y}_{2,t-1} + \mathbf{u}_{2t}] \mathbf{u}'_{1t}$ . Phillips

(1991) obtiene una expresión alternativa para esta distribución definiendo  $\mathbf{W}_{1.2}(u) = \mathbf{W}_{m_1}(u) - \boldsymbol{\Omega}_{12}\boldsymbol{\Omega}_{22}^{-1}\mathbf{W}_{m_2}(u)$  como lo que el proceso browniano  $\mathbf{W}_{1.2}(u)$  es independiente de  $\mathbf{W}_{m_2}(u)$ , y cuya matriz de covarianzas es  $\boldsymbol{\Omega}_{1.2} = \boldsymbol{\Omega}_{11} - \boldsymbol{\Omega}_{12}\boldsymbol{\Omega}_{22}^{-1}\boldsymbol{\Omega}_{21}$ . Esto permite expresar la distribución límite de  $\hat{\mathbf{B}}$  de la siguiente manera

$$\begin{aligned} n(\hat{\mathbf{B}}' - \mathbf{B}') &\xrightarrow{d} \left( \int_0^1 \mathbf{W}_{m_2}(u)\mathbf{W}'_{m_2}(u)du \right)^{-1} \int_0^1 \mathbf{W}_{m_2}(u)d\mathbf{W}'_{1.2}(u) + \\ &+ \left( \int_0^1 \mathbf{W}_{m_2}(u)\mathbf{W}'_{m_2}(u)du \right)^{-1} \int_0^1 \mathbf{W}_{m_2}(u)d\mathbf{W}'_{m_2}(u)\boldsymbol{\Omega}_{22}^{-1}\boldsymbol{\Omega}_{21} + \\ &+ \left( \int_0^1 \mathbf{W}_{m_2}(u)\mathbf{W}'_{m_2}(u)du \right)^{-1} \boldsymbol{\Omega}_{21} \end{aligned} \quad (\text{A.6})$$

donde el segundo término aparece por debido a que los  $m_2$  regresores en (A.1) son no estacionarios y el tercer término surge por la correlación contemporánea entre  $\mathbf{a}_{1t}$  y  $\mathbf{a}_{2t}$ . Nótese que si  $\mathbf{y}_{2t}$  fuera un vector de regresores estrictamente exógenos en (A.1) tal que  $\boldsymbol{\Omega}_{12} = \mathbf{0}$ , estos dos términos desaparecería, de forma que la distribución límite en este caso sería solamente el primer término, sin parámetros *de ruido* y sobre el cual se puede realizar inferencia utilizando resultados bien establecidos. Phillips (1991) ilustra también cómo la incorporación de información sobre el modelo en la estimación reduce estos sesgos en la distribución asintótica. En particular, la estimación conjunta del sistema, esto es, la estimación de  $\boldsymbol{\Phi}_1$  en (A.4) estimando las raíces unitarias, esto es, sin imponer un bloque (2,2) igual a la matriz identidad, elimina el tercer sesgo de (A.6) que surge precisamente por no reconocer la naturaleza endógena de ambos vectores de series. Si en lugar de estimarse las raíces se imponen, el estimador de máxima verosimilitud con información completa de  $\mathbf{B}$  está libre de ambos sesgos y se puede obtener, bien a partir de (A.4) con la presencia de  $m_2$  raíces unitarias impuestas o bien a partir de la estimación de máxima verosimilitud de (A.5), cuyo logaritmo de la función de verosimilitud tiene la siguiente expresión

$$\log f(\mathbf{B}, \boldsymbol{\Sigma} | \mathbf{Y}) \propto -(n/2) \log |\boldsymbol{\Sigma}| - (1/2) \sum_{t=2}^n (\nabla \mathbf{y}_t - \mathbf{A}\mathbf{C}\mathbf{y}_{t-1})' \boldsymbol{\Sigma}^{-1} (\nabla \mathbf{y}_t - \mathbf{A}\mathbf{C}\mathbf{y}_{t-1})$$

Particionando  $\boldsymbol{\Sigma}$  de acuerdo con  $\mathbf{y}_t$  y dada la estructura triangular del modelo, definiendo  $\mathbf{a}_{1.2t} = \mathbf{a}_{1t} - \boldsymbol{\Omega}_{12}\boldsymbol{\Omega}_{22}^{-1}\mathbf{a}_{2t}$  cuya matriz de covarianzas es  $\text{var}(\mathbf{a}_{1t} | \mathbf{a}_{2t}) \equiv \boldsymbol{\Sigma}_{11.2} = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}$ , de forma que la función de verosimilitud de  $\mathbf{y}_t$  se puede descomponer

alternativamente de la siguiente manera

$$\begin{aligned} \log f(\mathbf{B}, \boldsymbol{\Sigma} | \mathbf{Y}) &\propto -(n/2) \log |\boldsymbol{\Sigma}_{11.2}| - \\ &\quad -(1/2) \sum_{t=2}^n (\mathbf{y}_{1t} - \mathbf{B}\mathbf{y}_{2,t-1} - \mathbf{C}\nabla\mathbf{y}_{2t})' \boldsymbol{\Sigma}_{11.2}^{-1} (\mathbf{y}_t - \mathbf{B}\mathbf{y}_{2,t-1} - \mathbf{C}\nabla\mathbf{y}_{2t}) - \\ &\quad -(n/2) \log |\boldsymbol{\Sigma}_{22}| - (1/2) \sum_{t=2}^n \nabla\mathbf{y}'_{2t} \boldsymbol{\Sigma}_{22}^{-1} \nabla\mathbf{y}_{2t} \end{aligned}$$

donde la primera parte se corresponde con el logaritmo de la función de verosimilitud de  $\mathbf{y}_{1t}$  condicional a  $\nabla\mathbf{y}_{2t}$ , el segundo término es la función de densidad marginal de  $\nabla\mathbf{y}_{2t}$  y  $\mathbf{C} = \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}$ . Dado que la matriz  $\mathbf{B}$  no aparece en la densidad marginal, la estimación de máxima verosimilitud de  $\mathbf{B}$  se puede obtener a partir de la densidad condicional, que implica la estimación de mínimos cuadrados en la regresión,

$$\mathbf{y}_t = \mathbf{B}\mathbf{y}_{2t} + \mathbf{C}\nabla\mathbf{y}_{2t} + \mathbf{a}_{1.2t} \quad (\text{A.7})$$

Nótese que el término  $\mathbf{C}\nabla\mathbf{y}_{2t}$  añadido a la relación de cointegración está corrigiendo la correlación entre los errores de (A.1)-(A.2), de forma que el error del modelo de regresión sobre el que se estima  $\mathbf{B}$  es  $\mathbf{a}_{1.2t} \equiv E(\mathbf{a}_{1t} | \mathbf{a}_{2t}) = \mathbf{a}_{1t} - \boldsymbol{\Omega}_{12}\boldsymbol{\Omega}_{22}^{-1}\mathbf{a}_{2t}$  de forma que los regresores del modelo y el error están incorrelados, eliminándose así la fuente de sesgos en la distribución asintótica de  $\hat{\mathbf{B}}$ .

### A.1.2. Métodos de estimación en procesos con correlación serial

Es habitual considerar que la dinámica de corto plazo del sistema presenta dinámica tal que  $\mathbf{u}_t = \sum_{j=0}^{\infty} \boldsymbol{\Psi}_j \mathbf{a}_{t-j}$  donde  $\mathbf{a}_t \sim N(\mathbf{0}, \boldsymbol{\Omega}_a)$  con el polinomio MA( $\infty$ ) invertible. En este caso, se definen las siguientes matrices de largo plazo

$$\boldsymbol{\Omega}_a = \lim_{n \rightarrow \infty} n^{-1} \sum_{t=1}^n E(\mathbf{u}_t \mathbf{u}_t') \quad \boldsymbol{\Omega}_1 = \lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^{\infty} \sum_{t=1}^n E(\mathbf{u}_t \mathbf{u}_{t-i}')$$

de forma que la matriz de covarianzas de largo plazo se puede expresar como  $\boldsymbol{\Omega} = \boldsymbol{\Omega}_a + \boldsymbol{\Omega}_1 + \boldsymbol{\Omega}_1'$ .

Andrews (1991) propone un procedimiento, refinado en Andrews y Monahan (1992), para obtener estimaciones consistentes de las matrices de covarianza de largo plazo de un vector  $\mathbf{x}_t$  de dimensiones  $m \times 1$ . El procedimiento consiste en ajustar un VAR(1) a

vector de variables de interés  $\mathbf{x}_t$

$$\mathbf{x}_t = \Phi_1 \mathbf{x}_{t-1} + \mathbf{e}_t, \quad t = 2, \dots, n.$$

Utilizando los residuos  $\hat{\mathbf{e}}_t$  se obtendrían las estimaciones no paramétricas de  $\Omega_e$  y  $\Lambda_e$  como

$$\hat{\Omega}_e = \sum_{j=-n}^n w\left(\frac{j}{K}\right) \hat{\Gamma}_e(j), \quad \hat{\Lambda}_e = \sum_{j=0}^n w\left(\frac{j}{K}\right) \hat{\Gamma}_e(j),$$

donde  $w(\cdot)$  es una función kernel y  $\hat{\Gamma}_e(j)$  la matriz de autocovarianzas de  $\hat{\mathbf{e}}_t$  en el retardo  $j$ . Andrews (1991) proporciona también un procedimiento automático de selección del punto de truncamiento óptimo ( $k^*$ ). Las expresiones de las funciones kernel más utilizadas y su punto de truncamiento óptimo son las siguientes:

$$\text{Bartlett} \quad w\left(\frac{j}{k^*}\right) = \begin{cases} 1 - |j/k^*| & \text{si } |j/k^*| \leq 1 \\ 0 & \text{en otro caso} \end{cases} \quad k^* = 1,1147(\hat{\alpha}_1 n)^{1/3}$$

$$\text{Parzen} \quad w\left(\frac{j}{k^*}\right) = \begin{cases} 1 - 6(j/k^*)^2 + 6|j/k^*|^3 & \text{si } 0 \leq |j/k^*| \leq 0,5 \\ 2(1 - |j/k^*|)^3 & \text{si } 0,5 \leq |j/k^*| \leq 1 \\ 0 & \text{en otro caso} \end{cases} \quad k^* = 2,6614(\hat{\alpha}_2 n)^{1/5}$$

$$\text{QS} \quad w\left(\frac{j}{k^*}\right) = \frac{25}{12\pi^2(j/k^*)^2} \left( \frac{\sin(6\pi(j/k^*)/5)}{6\pi(j/k^*)/5} - \cos(6\pi(j/k^*)/5) \right) \quad k^* = 1,3221(\hat{\alpha}_2 n)^{1/5}$$

donde

$$\hat{\alpha}_1 = \sum_{i=1}^m \frac{4\hat{\rho}_i^2 \hat{\sigma}_i^2}{(1 - \hat{\rho}_i)^6 (1 + \hat{\rho}_i)^2} / \sum_{j=1}^m \frac{\hat{\sigma}_j^2}{(1 - \hat{\rho}_j)^4}$$

$$, \hat{\alpha}_2 = \sum_{i=1}^m \frac{4\hat{\rho}_i^2 \hat{\sigma}_i^2}{(1 - \hat{\rho}_i)^8} / \sum_{j=1}^m \frac{\hat{\sigma}_j^2}{(1 - \hat{\rho}_j)^4},$$

con  $\hat{\rho}_j$  y  $\hat{\sigma}_j^2$  son las estimaciones del parámetro autoregresivo y la varianza de un proceso AR(1) univariante ajustado a cada elemento de  $\hat{\mathbf{e}}_t$ . Por último, las estimaciones

consistentes de  $\mathbf{\Omega}$  y  $\mathbf{\Lambda}$  se obtienen como

$$\begin{aligned}\hat{\mathbf{\Omega}} &= (\mathbf{I}_m - \hat{\mathbf{\Phi}})^{-1'} \hat{\mathbf{\Omega}}_e (\mathbf{I}_m - \hat{\mathbf{\Phi}})^{-1}, \\ \hat{\mathbf{\Lambda}} &= (\mathbf{I}_m - \hat{\mathbf{\Phi}})^{-1'} \hat{\mathbf{\Lambda}}_e (\mathbf{I}_m - \hat{\mathbf{\Phi}})^{-1} - (\mathbf{I}_m - \hat{\mathbf{\Phi}})^{-1'} \hat{\mathbf{\Phi}}' \hat{\mathbf{\Sigma}}_u,\end{aligned}$$

con  $\hat{\mathbf{\Sigma}}_u$  la matriz de covarianzas de  $\mathbf{x}_t$ . Cuando estas correcciones no paramétricas se utilizan para contrastar estacionariedad, Choi y Ahn (1995) advierten que el parámetro de truncamiento óptimo es  $O_p(n)$ , por lo que, siguiendo a Gabriel (2003), se fija el parámetro de truncamiento en  $k^{**} = \min(k^*, 12(n/100)^{1/5})$ . Este procedimiento se utiliza para obtener estimaciones consistentes de las matrices de covarianzas de largo plazo que se utilizan en los métodos de estimación asintóticamente eficientes listados a continuación. El propósito de estos métodos es obtener estimaciones consistentes de la matriz de relaciones de cointegración,  $\mathbf{B}$  cuya distribución no tenga los sesgos asintóticos que presenta su estimador de mínimo-cuadrático.

### Mínimos cuadrados modificados (FM-OLS)

Phillips y Hansen (1990) y Phillips (1995) proponen una corrección no paramétrica del estimador de mínimos cuadrados ordinarios de  $\mathbf{B}$  para eliminar de su distribución los sesgos asintóticos. Esquemáticamente, los pasos a dar para la estimación de  $\mathbf{A}$  por el método son los siguientes:

1. obtener los residuos de la regresión (A.1) y construir el vector  $\mathbf{w}_t = (\hat{\mathbf{a}}'_{1t}, \nabla \mathbf{y}'_{2t})'$ ,
2. obtener las estimaciones de las matrices  $\hat{\mathbf{\Omega}}$  y  $\hat{\mathbf{\Lambda}}$  utilizando  $\mathbf{w}_t$  mediante el procedimiento anterior,
3. modificar el vector  $\mathbf{y}_t$  como

$$\mathbf{y}_{1t}^+ = \mathbf{y}_{1t} - \hat{\mathbf{\Omega}}_{12} \hat{\mathbf{\Omega}}_{22}^{-1} \nabla \hat{\mathbf{u}}_{2t},$$

4. y calcular el estimador FM-OLS de  $\mathbf{B}$  tiene la siguiente expresión

$$\hat{\mathbf{B}}^+ = (\mathbf{Y}'_2 \mathbf{Y}_2)^{-1} \left( \mathbf{Y}'_2 \mathbf{Y}_1^+ - n \begin{bmatrix} \mathbf{0} \\ \mathbf{\Delta}_{12}^+ \end{bmatrix} \right), \quad \hat{\mathbf{\Delta}}_{12}^+ = \hat{\mathbf{\Delta}}_{12} - \hat{\mathbf{\Omega}}_{12} \hat{\mathbf{\Omega}}_{22}^{-1} \hat{\mathbf{\Delta}}_{22},$$

donde las filas de  $\mathbf{Y}_1^+$  y  $\mathbf{Y}_2$  contienen las observaciones de  $\mathbf{y}_{1t}^+$  e  $\mathbf{y}_{2t}$  respectivamente.

Phillips (1995) demuestra que el estimador FM-OLS de  $\mathbf{B}$  es consistente y su distribución carece de sesgos asintóticos. Los residuos de la regresión se obtienen de la manera habitual como  $\hat{\mathbf{a}}_{1t}^+ = \hat{\mathbf{y}}_{1t}^+ - \hat{\mathbf{A}}^+ \mathbf{y}_{2t}$ .

### Regresión de Correlación Canónica

Park (1992) desarrolla un método de estimación del vector de cointegración en un modelo triangular, donde la corrección de los sesgos se realiza de una manera similar la anterior de Phillips y Hansen (1990). El estimador  $\hat{\mathbf{B}}^*$  se obtiene a partir de la regresión de mínimos cuadrados multivariantes similar a (A.1) donde las variables se modifican de la siguiente manera

$$\begin{aligned} \mathbf{y}_{1t}^* &= \mathbf{y}_{1t}^+ - \hat{\mathbf{B}} \hat{\mathbf{\Delta}}_2 \hat{\mathbf{\Sigma}}^{-1} \hat{\mathbf{w}}_t \\ \mathbf{y}_{2t}^* &= \mathbf{y}_{2t} - \hat{\mathbf{\Delta}}_2 \hat{\mathbf{\Sigma}}^{-1} \hat{\mathbf{w}}_t \end{aligned}$$

donde  $\mathbf{y}_{1t}^+$  es la modificación de  $\mathbf{y}_{1t}$  descrita para el método anterior. Los residuos de la regresión de correlación canónica se obtienen fácilmente como  $\hat{\mathbf{a}}_t^* = \mathbf{y}_{1t}^* - \hat{\mathbf{B}}^* \mathbf{y}_{2t}^*$ .

### Mínimos Cuadrados Dinámicos

Un tercer método es el desarrollado por Saikkonen (1991) y Stock y Watson (1993) que proponen estimar  $\mathbf{B}$  en la siguiente regresión aumentada con valores pasados y futuros de  $\nabla \mathbf{y}_{2t}$ ,

$$\mathbf{y}_{1t} = \mathbf{B} \mathbf{y}_{2t} + \sum_{j=-k}^k \mathbf{B}_j \nabla \mathbf{y}_{2t-j} + \boldsymbol{\epsilon}_t \quad (\text{A.8})$$

Alternativamente se puede obtener el estimador en dos etapas, utilizando regresando los residuos de la regresión de  $\mathbf{y}_{1t}$  sobre  $\{\nabla \mathbf{y}_{2t-j}\}_{j=-k}^k$  sobre los de la regresión de  $\mathbf{y}_{2t}$  sobre los mismos regresores. Gabriel (2003) propone utilizar  $k = \text{int}(n^{1/3})$ .

#### A.1.3. Contrastes basados en regresión

Dada su forma intuitiva la relativa facilidad para estimar la relación de cointegración, el modelo (A.1)-(A.2) ha sido ampliamente utilizado para derivar contrastes de cointegración. Engle y Granger (1987), Phillips y Ouliaris (1988) o Phillips y Ouliaris (1990) proponen utilizar contrastes de no estacionariedad basados en los residuos de (A.1) para  $m_1 = 1$ . Sin embargo, como destacan Phillips y Ouliaris citando a Engle, resultaría más conveniente que la hipótesis nula a contrastar fuera la de cointegración, esto es,

contrastar si el error correspondiente a la ecuación regresión,  $\mathbf{a}_{1t}$ , es estacionaria. Utilizando esta aproximación Hansen (1992) o Shin (1994) derivan contrastes de cointegración tipo multiplicadores de Lagrange, análogos a los contrastes de estacionariedad o componentes deterministas univariantes (Kwiatkowski et al., 1992; Tanaka, 1990) construidas sobre los residuos de los métodos de estimación asintóticamente eficientes descritos anteriormente. Otros contrastes propuestos en este marco son los de Harris y Inder (1994), Leybourne y McCabe (1993), McCabe et al. (1997) o Saikkonen y Luukkonen (1997), donde la diferencia entre ellos radica en la forma paramétrica de la corrección de los sesgos de endogeneidad. Por su parte, Tanaka (1993), Jansson (2005) y Kurozumi y Arai (2008) derivan contrastes localmente óptimos para la hipótesis nula de cointegración cuando (A.1) contiene sólo una relación de cointegración. Por su parte, Choi y Ahn (1995) proponen varios estadísticos para el contraste de cointegración en el caso multivariante ( $m_1 > 1$ ),

$$LM_I = \text{tr} \left( n^{-1} \sum_{t=2}^n \nabla \mathbf{S}_t^* \mathbf{S}_{t-1}^{*'} - \hat{\boldsymbol{\kappa}} \hat{\boldsymbol{\Sigma}} \hat{\boldsymbol{\kappa}}' \right) \hat{\boldsymbol{\Omega}}_{11.2}^{*-1} \times \quad (\text{A.9})$$

$$\times \left( n^{-1} \sum_{t=2}^n \nabla \mathbf{S}_t^* \mathbf{S}_{t-1}^{*'} - \hat{\boldsymbol{\kappa}} \hat{\boldsymbol{\Sigma}} \hat{\boldsymbol{\kappa}}' \right) \hat{\boldsymbol{\Omega}}_{11.2}^{*-1},$$

$$LM_{II} = \text{tr} \left( \sum_{t=2}^n \nabla \mathbf{S}_t^* \mathbf{S}_{t-1}^{*'} - n \hat{\boldsymbol{\kappa}} \hat{\boldsymbol{\Sigma}} \hat{\boldsymbol{\kappa}}' \right) \left( \sum_{t=2}^T \mathbf{S}_{t-1}^* \mathbf{S}_{t-1}^{*'} \right)^{-1} \times \quad (\text{A.10})$$

$$\times \left( n^{-1} \sum_{t=2}^n \nabla \mathbf{S}_t^* \mathbf{S}_{t-1}^{*'} - n \hat{\boldsymbol{\kappa}} \hat{\boldsymbol{\Sigma}} \hat{\boldsymbol{\kappa}}' \right) \hat{\boldsymbol{\Omega}}_{11.2}^{*-1},$$

$$SBDH = \text{tr} \left[ \left( n^{-2} \sum_{t=1}^n \mathbf{S}_t^* \mathbf{S}_t^{*'} \right) \hat{\boldsymbol{\Omega}}_{11.2}^{*-1} \right], \quad (\text{A.11})$$

donde  $\mathbf{S}_t^* = \sum_{i=1}^t \hat{\mathbf{u}}_i^*$  con  $\hat{\mathbf{u}}_i^*$  los residuos de la regresión de correlación canónica de sistema cointegrado, y  $\hat{\boldsymbol{\kappa}} = [\mathbf{I}_{m_1} \quad -\hat{\boldsymbol{\Omega}}_{12}^* \hat{\boldsymbol{\Omega}}_{22}^{-1*}]$  y  $\hat{\boldsymbol{\Omega}}_{1.2}^* = \hat{\boldsymbol{\Omega}}_{11}^* - \hat{\boldsymbol{\Omega}}_{12}^* \hat{\boldsymbol{\Omega}}_{22}^{-1*} \hat{\boldsymbol{\Omega}}_{21}^*$  se obtienen a partir de la estimación no paramétrica de  $\boldsymbol{\Omega}_a$  descrita anteriormente. Nótese en el caso de ausencia de correlación serial, el estadístico SBDH es equivalente al desarrollado por

Nyblom y Harvey (2000). Las distribuciones límite de estos estadísticos son

$$\begin{aligned}
 LM_I &\xrightarrow{d} \text{tr} \left[ \left( \int_0^1 d\mathbf{B}^*(u)\mathbf{B}^*(u)' \right) \left( \int_0^1 \mathbf{B}^*(u)d\mathbf{B}^*(u)' \right) \right] \\
 LM_{II} &\xrightarrow{d} \text{tr} \left[ \left( \int_0^1 d\mathbf{b}^*(u)\mathbf{B}^*(u)' \right) \left( \int_0^1 \mathbf{B}^*(u)\mathbf{B}^*(u)' du \right)^{-1} \left( \int_0^1 \mathbf{B}^*(u)d\mathbf{B}^*(u)' \right) \right] \\
 SBDH &\xrightarrow{d} \text{tr} \left[ \int_0^1 \mathbf{B}^*(u)\mathbf{B}^*(u)' du \right]
 \end{aligned}$$

donde se definen

$$\begin{aligned}
 \mathbf{B}^*(u) &= \mathbf{W}_{m_1}(u) - \left( \int_0^1 d\mathbf{W}_{m_1}(s)\mathbf{W}_{m_2}(s)' \right) \left( \int_0^1 \mathbf{W}_{m_2}(s)\mathbf{W}_{m_2}(s)' ds \right)^{-1} \int_0^r \mathbf{W}_{m_2}(s) ds, \\
 d\mathbf{B}^*(u) &= d\mathbf{W}_{m_1}(u) - \left( \int_0^1 d\mathbf{W}_{m_1}(s)\mathbf{W}_{m_2}(s)' \right) \left( \int_0^1 \mathbf{W}_{m_2}(s)\mathbf{W}_{m_2}(s)' ds \right)^{-1} \mathbf{W}_{m_2}(r) du,
 \end{aligned}$$

y  $\mathbf{W}(u)$  es un proceso browniano como particionado de acuerdo con el vector  $\mathbf{y}_t$ .

## A.2. Modelo de tendencias comunes

### A.2.1. Representación y forma estructural

Stock y Watson (1988) derivan la representación en forma de tendencias comunes a partir de la forma MA( $\infty$ ) del proceso en primeras diferencias,

$$\nabla \mathbf{y}_t = \boldsymbol{\Psi}(B)\mathbf{a}_t \quad \mathbf{a}_t, \sim N(\mathbf{0}, \boldsymbol{\Omega}_a).$$

Mediante sustituciones sucesivas, se obtiene la descomposición de Beveridge y Nelson para  $\mathbf{y}_t$  como

$$\mathbf{y}_t = \mathbf{y}_0 + \boldsymbol{\Psi}(1) \sum_{i=1}^t \mathbf{a}_i + \boldsymbol{\Psi}^*(B)\mathbf{a}_t, \quad (\text{A.12})$$

con  $\mathbf{y}_0$  como valor inicial fijo, donde el segundo término es el componente permanente o tendencia estocástica y las matrices de la parte transitoria se obtienen como  $\boldsymbol{\Psi}^*(B) = (1 - B)^{-1}[\boldsymbol{\Psi}(B) - \boldsymbol{\Psi}(1)]$  tal que  $\boldsymbol{\Psi}_j^* = -\sum_{i=j+1}^{\infty} \boldsymbol{\Psi}_i$ . Si existe al menos una raíz unitaria en este polinomio MA( $\infty$ ), entonces  $|\boldsymbol{\Psi}(1)| = 0$ . Suponiendo que el rango de esta matriz de ganancias es  $m - r$  ( $r > 0$ ), existe una matriz  $m \times r$ ,  $\boldsymbol{\beta}$ , que elimina la tendencia

estocástica de tal forma que  $\beta' \Psi(1) = \mathbf{0}$  con lo que el conjunto de combinaciones lineales entre variables integradas

$$\beta' \mathbf{y}_t = \beta' \mathbf{y}_0 + \beta' \Psi^*(B) \mathbf{a}_t,$$

es un proceso estacionario de relaciones de cointegración. Dado que  $\Psi(1)$  tiene rango  $r$ , existe una matriz  $\mathbf{H} = [\mathbf{H}_1 | \mathbf{H}_2]$ , con  $\mathbf{H}_1$  y  $\mathbf{H}_2$  dos matrices de dimensiones  $m \times r$  y  $m \times (m - r)$  respectivamente ortogonales entre sí, tales que  $\Psi(1)[\mathbf{H}_1 | \mathbf{H}_2] = [\mathbf{0}_{m \times (m-r)} | \mathbf{A}] = \mathbf{A} \mathbf{S}_r$  con  $\mathbf{A}$  una matriz  $m \times (m - r)$  de rango completo y la matriz de selección  $\mathbf{S}_{(m-r)} = [\mathbf{0}_{(m-r), r} | \mathbf{I}_{(m-r)}]$ . Insertando  $\mathbf{H} \mathbf{H}^{-1}$  a la derecha de  $\Psi(1)$  en (A.12), esta ecuación se puede expresar alternativamente como

$$\mathbf{y}_t = \mathbf{y}_0 + \mathbf{A} \boldsymbol{\mu}_t + \mathbf{u}_t \quad \boldsymbol{\mu}_t = \boldsymbol{\mu}_{t-1} + \mathbf{a}_t, \quad (\text{A.13})$$

donde  $\boldsymbol{\mu}_t = \mathbf{S}_r \mathbf{H}^{-1} \sum_{i=1}^t \mathbf{a}_i$  y  $\mathbf{u}_t = \Psi^*(B) \mathbf{a}_t$ . Esta expresión indica que el proceso  $\mathbf{y}_t$  se compone de  $m - r$  tendencias que afectan a las  $m$  series consideradas de la forma definida por la matriz  $\mathbf{A}$ .

### A.2.2. Contraste de tendencias comunes

Stock y Watson (1988) proponen su contraste sobre el número de tendencias comunes que afectan al modelo a partir de una transformación lineal de  $\mathbf{y}_t$ , de forma que bajo la hipótesis nula de  $m - r$  tendencias comunes ésta sea estacionaria. Definiendo  $\mathbf{D} = [\boldsymbol{\beta} | \boldsymbol{\beta}_\perp]$  con  $\boldsymbol{\beta}_\perp$  una matriz  $m \times (m - r)$  tal que  $\boldsymbol{\beta}'_\perp \boldsymbol{\beta} = \mathbf{0}$  y  $\boldsymbol{\beta}'_\perp \boldsymbol{\beta}_\perp = \mathbf{I}_{m-r}$ , de forma que los  $r$  primeros componentes de  $\mathbf{D} \mathbf{y}_t$  serán las relaciones de cointegración,  $\mathbf{z}_t = \boldsymbol{\beta}' \mathbf{y}_t$  y el resto ( $\mathbf{x}_t = \boldsymbol{\beta}'_\perp \mathbf{y}_t$ ) contiene  $m - r$  combinaciones lineales no estacionarias, de tal forma que  $\nabla \mathbf{x}_t = \boldsymbol{\beta}'_\perp \Psi(B) \mathbf{a}_t$ . De esta forma, el proceso transformado se puede expresar conjuntamente de la siguiente manera:

$$\begin{bmatrix} \mathbf{z}_t \\ \nabla \mathbf{x}_t \end{bmatrix} = \mathbf{D}' \mathbf{y}_0 + \begin{bmatrix} \mathbf{0}_{r \times m} \\ \boldsymbol{\beta}'_\perp \end{bmatrix} \Psi(1) \sum_{i=1}^t \mathbf{a}_i + \mathbf{D}' \Psi^*(B) \mathbf{a}_t.$$

Bajo la hipótesis nula de  $m - r$  tendencias comunes ( $r$  relaciones de cointegración), el vector  $(m - r)$ -dimensional  $\mathbf{x}_t$  es I(1) de forma que  $\hat{\mathbf{R}}(1)$ , la matriz de autocorrelación en el primer retardo de  $\mathbf{x}_t$  definida como

$$\hat{\mathbf{R}}(1) = \left[ \sum_{t=2}^n \mathbf{x}_t \mathbf{x}'_{t-1} \right] \left[ \sum_{t=1}^n \mathbf{x}_t \mathbf{x}_t' \right]^{-1}, \quad (\text{A.14})$$

tendrá (asintóticamente)  $m - r$  autovalores unitarios. Si existe una tendencia común menos,  $\mathbf{R}(1)$ , el límite en probabilidad de  $\hat{\mathbf{R}}(1)$ , tendrá  $m - r - 1$  autovalores unitarios y uno con módulo menor que la unidad. Por lo tanto, denotando por  $\lambda_{min}$  el autovalor de  $\mathbf{R}(1)$  con menor parte real, el problema a contrastar será  $H_0 : Re(\lambda_{min}) = 1$ , esto es, la presencia de  $m - r$  tendencias comunes ( $r$  relaciones de cointegración) frente a  $H_1 : Re(\lambda_{min}) < 1$ , una tendencia común menos (una relación de cointegración más).

El estadístico de contraste, por tanto, se construye a partir de (A.14), para lo que se necesita una estimación consistente de la matriz  $\mathbf{D}$  que contiene los vectores de cointegración. Stock y Watson (1988) proponen utilizar los componentes principales de  $\mathbf{y}_t$  como estimación consistente de la matriz  $\mathbf{D}$ , de forma que  $\tilde{\boldsymbol{\beta}}$  serán las  $r$  combinaciones lineales de  $\mathbf{y}_t$  con menor varianza y  $\tilde{\boldsymbol{\beta}}_{\perp}$  el resto, asegurándose así la ortogonalidad entre las columnas de  $\mathbf{D}$ . Por último, una vez obtenido el vector  $\tilde{\mathbf{x}}_t = \tilde{\boldsymbol{\beta}}'_{\perp} \mathbf{y}_t$  se ajusta un modelo VAR(p)  $\boldsymbol{\Pi}(B)$  sobre  $\nabla \tilde{\mathbf{x}}_t$  para eliminar la posible correlación serial presente en el modelo. Una vez obtenidas estimaciones consistentes de los parámetros autorregresivos, se obtiene el vector filtrado

$$\hat{\mathbf{u}}_t = \hat{\boldsymbol{\Pi}}(B) \tilde{\mathbf{x}}_t$$

y el estadístico de contraste de la hipótesis nula de  $m - r$  tendencias comunes frente a  $m - r - 1$  de Stock y Watson (1988) se define como

$$SW = n(\hat{\lambda}_{min} - 1)$$

donde  $\hat{\lambda}_{min}$  es el autovalor con menor parte real de la matriz de autocorrelación en el primer retardo del proceso filtrado  $\hat{\mathbf{u}}_t$ . Stock y Watson derivaron los valores críticos del estadístico.

### A.3. Cointegración en modelos VAR

#### A.3.1. Modelo VAR(p) y Modelo de corrección de error

Un VAR( $p$ ) para el vector  $m$ -dimensional  $\mathbf{y}_t$  se define como

$$\mathbf{y}_t = \boldsymbol{\Phi}_1 \mathbf{y}_{t-1} + \dots + \boldsymbol{\Phi}_p \mathbf{y}_{t-p} + \mathbf{a}_t, \quad \mathbf{a}_t \sim N(\mathbf{0}, \boldsymbol{\Omega}_a),$$

con  $\boldsymbol{\Omega}_a$  una matriz definida positiva. Si el polinomio autorregresivo  $\boldsymbol{\Phi}(B) = \mathbf{I}_m - \boldsymbol{\Phi}_1 B - \dots - \boldsymbol{\Phi}_p B^p$  tiene  $r < m$  raíces unitarias, entonces  $|\boldsymbol{\Phi}(1)| = 0$  y el rango de esta matriz

de ganancias,  $\Phi(1)$ , es  $r < m$ , de forma que puede expresarse como el producto de dos matrices de dimensiones  $m \times r$  de rango completo tal que  $\Phi(1) = \alpha\beta'$ . El modelo VAR(p) también admite una representación en forma de modelo de corrección de error,

$$\nabla \mathbf{y}_t = \alpha\beta' \mathbf{y}_{t-1} + \sum_{i=1}^{p-1} \Phi_i^* \nabla \mathbf{y}_{t-i} + \mathbf{a}_t \quad t = p+1, \dots, n, \quad (\text{A.15})$$

donde las matrices que muestran la dinámica de corto plazo se definen como  $\Phi_j^* = -\sum_{i=j+1}^p \Phi_i$ . El conjunto de  $r$  combinaciones lineales  $\beta' \mathbf{y}_{t-1}$  es un vector estacionario de relaciones de cointegración o relaciones a largo plazo y la matriz  $\alpha$  es una matriz de difusión que indica de qué forma afecta cada relación de cointegración a cada serie.

### A.3.2. Regresión de rango reducido y contraste de cointegración

En sus trabajos, Johansen (1988, 1991), Johansen propone un procedimiento en dos etapas para obtener el estimador de máxima verosimilitud de un modelo VAR(p) con rango de cointegración  $r$ . En un primer paso se elimina del modelo (A.15) todos los parámetros de corto plazo, lo que se realiza mediante las siguientes regresiones auxiliares,

$$\nabla \mathbf{y}_t = \sum_{i=1}^{p-1} \mathbf{A}_i^* \nabla \mathbf{y}_{t-i} + \mathbf{r}_{1t}, \quad \mathbf{y}_{t-1} = \sum_{i=1}^{p-1} \mathbf{B}_i^* \nabla \mathbf{y}_{t-i} + \mathbf{r}_{2t},$$

donde el modelo concentrado con respecto a los parámetros de corto plazo es  $\mathbf{r}_{1t} = \alpha\beta' \mathbf{r}_{1t} + \mathbf{u}_t$ , de forma que el logaritmo de su función de verosimilitud puede expresarse como

$$\log f(\Omega_a, \alpha, \beta) \propto -\left(\frac{n}{2}\right) \log |\Omega_a| - \left(\frac{1}{2}\right) \sum_{t=1}^n [(\mathbf{r}_{1t} - \alpha\beta' \mathbf{r}_{2t})' \Omega_a^{-1} (\mathbf{r}_{1t} - \alpha\beta' \mathbf{r}_{2t})]. \quad (\text{A.16})$$

Definiendo los autovalores  $\lambda_j$  como las soluciones de la ecuación

$$|\lambda \mathbf{S}_{22} - \mathbf{S}_{21} \mathbf{S}_{11}^{-1} \mathbf{S}_{12}| = 0 \quad \mathbf{S}_{ij} = \sum_{t=1}^n \mathbf{r}_{it} \mathbf{r}'_{jt}, \quad i, j = 1, 2,$$

que pueden obtenerse también a partir de los autovalores de la matriz  $\mathbf{S}_{22}^{-1} \mathbf{S}_{21} \mathbf{S}_{11}^{-1} \mathbf{S}_{12}$ . Estos autovalores también pueden interpretarse como los coeficientes de correlación canónica parcial de  $\mathbf{y}_{t-1}$  y  $\nabla \mathbf{y}_t$  dado  $\{\nabla \mathbf{y}_{t-j}\}_{j=1}^{p-1}$  (ver, p. ej. Reinsel, 1997). Ordenados de forma

que  $\hat{\lambda}_1 > \dots > \hat{\lambda}_m > 0$ , la función de verosimilitud se puede expresar de forma alternativa como (p. ej. Juselius, 2006, p.118)

$$\ell^*(m) \propto -\left(\frac{n}{2}\right) \log \left[ |\mathbf{S}_{11}| \prod_{i=1}^m (1 - \hat{\lambda}_i) \right]. \quad (\text{A.17})$$

Los contrastes de cointegración de Johansen se plantean como contrastes de ratio de verosimilitudes. Así, el estadístico de contraste para la hipótesis nula  $H_0 : \text{rango}(\Phi(1)) \leq r$  frente a  $H_1 : \text{rango}(\Phi(1)) = m$  se plantea como

$$\begin{aligned} J_1 = & -n \sum_{j=r+1}^m \log(1 - \hat{\lambda}_j) \xrightarrow{d} \text{tr} \left( \int_0^1 \mathbf{W}_{m-r}(u) d\mathbf{W}'_{m-r}(u) \right)' \times \\ & \times \left( \int_0^1 \mathbf{W}_{m-r}(u) \mathbf{W}'_{m-r}(u) du \right)^{-1} \left( \int_0^1 \mathbf{W}_{m-r}(u) d\mathbf{W}'_{m-r}(u) \right), \quad (\text{A.18}) \end{aligned}$$

con  $\mathbf{W}_{m-r}(u)$  un proceso browniano de dimensiones  $(m-r) \times 1$ . Si no se rechaza la hipótesis nula, existirán al menor  $m-r$  raíces unitarias en el sistema. Por su parte, para contratar la hipótesis nula de que el sistema presente  $r-1$  relaciones de cointegración frente a éstas sean  $r$ , también se propone el estadístico de contraste

$$J_2 = -n \log(1 - \hat{\lambda}_r) \xrightarrow{d} \frac{\left( \int_0^1 W(u) dW(u) \right)^2}{\int_0^1 W(u) du},$$

con  $W(u)$  un proceso browniano univariante. Una aproximación a la distribución en muestras finitas de ambos estadísticos, basada en la distribución Gamma, ha sido propuesta por Doornik (1998).

### A.3.3. Cointegración y tendencias comunes

Si el polinomio VAR(p) presenta al menos una raíz unitaria tal que  $|\Phi(1)| = 0$ , esta matriz de ganancias tiene rango  $r < m$  y se puede descomponer como  $\Phi(1) = \alpha\beta'$  donde  $\alpha$  y  $\beta$  son dos matrices de dimensiones  $m \times r$  y rango completo  $r$ . Ambas matrices pueden obtenerse a partir de la descomposición espectral de la matriz de ganancias, tal que  $\Phi(1) = \mathbf{P}\mathbf{\Lambda}\mathbf{Q}$  donde  $\mathbf{\Lambda}$  es la matriz diagonal de autovalores con  $m-r$  de ellos iguales a cero y  $r$  distintos de cero, por ser la matriz de rango  $r$  y  $\mathbf{P}\mathbf{Q} = \mathbf{I}_m$ . Particionando las matrices de autovectores de acuerdo con los autovalores no nulos, se puede identificar  $\alpha = \mathbf{p}_1 \mathbf{\Lambda}_r^{1/2}$  y  $\beta = \mathbf{q}_1 \mathbf{\Lambda}_r^{1/2}$ , con  $\mathbf{p}_1$  ( $\mathbf{q}_1$ ) la matriz de columnas

(filas) de  $\mathbf{P}$  ( $\mathbf{Q}$ ) correspondientes a los autovalores no nulos de  $\Phi(1)$ . De igual forma, se pueden definir las matrices de dimensiones  $m \times (m - r)$ ,  $\beta_{\perp} = \mathbf{q}_2$  y  $\alpha_{\perp} = \mathbf{p}_2$  tal que  $\beta'_{\perp}\beta = \alpha'_{\perp}\alpha = \mathbf{0}$ . Con estas definiciones, el teorema de la representación de Granger proporciona la forma MA de un proceso VAR(p) cointegrado. Si  $\Phi(B)$  tiene alguna raíz unitaria y se cumple que

$$|\alpha'_{\perp}\Phi^*\beta_{\perp}| \neq 0, \quad \text{con} \quad \Phi^* = \mathbf{I}_m - \sum_{j=1}^{p-1} \Phi_j^*,$$

entonces,

$$(1 - B)\Phi^{-1}(B) = \mathbf{C}(1) + (1 - B)\mathbf{C}^*(B), \quad \mathbf{C}(1) = \beta_{\perp}(\alpha'_{\perp}\Phi^*\beta_{\perp})^{-1}\alpha'_{\perp},$$

donde  $\mathbf{C}^*(B) = \sum_{j=0}^{\infty} \mathbf{C}_j^* B^j$ , es un polinomio media móvil matricial invertible. De esta forma el proceso  $\mathbf{y}_t$  tiene la siguiente representación media móvil,

$$\mathbf{y}_t = \mathbf{y}_0^* + \beta_{\perp}(\alpha'_{\perp}\Phi^*\beta_{\perp})^{-1}\alpha'_{\perp} \sum_{i=1}^t \mathbf{a}_i + \sum_{j=0}^{\infty} \mathbf{C}_j^* \mathbf{a}_{t-j}.$$

Este teorema demuestra que un proceso VAR(p) cointegrado, con rango de cointegración  $r$ , se puede representar como un modelo con  $m - r$  tendencias comunes,  $\sum_{i=1}^t \alpha'_{\perp} \mathbf{a}_i$ , que afectan a cada serie temporal de acuerdo a la matriz de difusión  $\beta_{\perp}(\alpha'_{\perp}\Phi^*\beta_{\perp})^{-1}$ . Nótese que estas  $m - r$  tendencias comunes se eliminan al premultiplicar  $\mathbf{y}_t$  por  $\beta'$ , por lo que las relaciones de cointegración eliminan las tendencias comunes del modelo. Este teorema de la representación de Granger muestra la relación entre el modelo VAR(p) cointegrado y el modelo de tendencias comunes de Stock y Watson (1988).

# Bibliografía

- B. Abraham y G.E.P. Box. Deterministic and forecast-adaptive time-dependent models. *Applied Statistics*, 27(2):120–130, 1978.
- S.K. Ahn y G.C. Reinsel. Nested reduced rank autoregressive models for multiple time series. *Journal of the American Statistical Association*, 83:849–856, 1988.
- S.K. Ahn y G.C. Reinsel. Estimation of partially nonstationary multivariate autoregressive models. *Journal of the American Statistical Association*, 85:813–823, 1990.
- S.K. Ahn y G.C. Reinsel. Estimation of partially nonstationary multivariate autoregressive models with seasonal behavior. *Journal of Econometrics*, 62:317–350, 1994.
- T.W. Anderson. Estimating linear restrictions on regression coefficients for multivariate normal distributions. *Annals of Mathematical Statistics*, 22:327–351, 1951.
- T.W. Anderson. *The Statistical Analysis of Time Series*. Wiley, New York, 1971.
- T.W. Anderson y D.A. Darling. Asymptotic theory of certain 'goodness of fit' criteria based on stochastic processes. *Annals of Mathematical Statistics*, 23:193–212, 1952.
- D.W.K. Andrews. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, 59(3):817–858, 1991.
- D.W.K. Andrews y J.C. Monahan. An improved heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, 60(4):953–966, 1992.
- P. Billingsley. *Convergence of Probability Measures*. Wiley Series in Probability and Mathematical Statistics, 1968.
- G.E.P. Box, G.M. Jenkins, y G.C. Reinsel. *Time Series Analysis: Forecasting and Control*. John Wiley and sons, Inc., Hoboken, New Jersey, 4th edition, 2008.

- B.M. Brown. Cramèr-von Mises distributions and permutation tests. *Biometrika*, 69: 619–24, 1982.
- F. Busetti. Tests of seasonal integration and cointegration in multivariate unobserved component models. *Journal of Applied Econometrics*, 21:419–438, 2006.
- F. Busetti y A. Harvey. Seasonality tests. *Journal of Business and Economic Statistics*, 21(3):420–436, 2003.
- M. Caner. A locally optimal seasonal unit root test. *Journal of the American Statistical Association*, 16(3):349–356, 1998.
- F. Canova y B.E. Hansen. Are seasonal patterns constant over time? a test for seasonal stability. *Journal of Business and Economic Statistics*, 13(3):237–252, 1995.
- I. Choi y B.C. Ahn. Testing for cointegration in a system of equations. *Econometric Theory*, 11:952–983, 1995.
- I. Choi y B.C. Ahn. Testing the null of stationarity for multiple time series. *Journal of Econometrics*, 88:41–77, 1999.
- J.E.H. Davidson, D.F. Hendry, F. Srba, y S. Yeo. Econometric modelling of the aggregate time-series relationships between consumers' expenditure and income in the United Kingdom. *The Economic Journal*, 88(352):661–692, 1978.
- R.B. Davies. Numerical inversion of a characteristic function. *Biometrika*, 60(2):415–417, 1973.
- D.A. Dickey y W.A. Fuller. Distribution of the estimators for autoregressive time series with a unit root. *Journal of The American Statistical Association*, 74(366), 1979.
- D.A. Dickey, D.P. Hasza, y W.A. Fuller. Testing for unit roots in seasonal time series. *Journal of The American Statistical Association*, 79(386), 1984.
- J.A. Doornik. Approximation to the asymptotic distributions of cointegration tests. *Journal of Economic Surveys*, 12:573–593, 1998.
- R.F. Engle y C.W.J. Granger. Co-integration and error correction: Representation, estimation and testing. *Econometrica*, 35:143–159, 1987.
- A. Escribano y D. Peña. Cointegration and common factors. *Journal of Time Series Analysis*, 15(6):577–586, 1993.

- T.S. Ferguson. *Mathematical Statistics: A Decision Theoretic Approach*. Academic Press, New York, 1967.
- G. Forchini. Similar tests for covariance structures in multivariate linear models. *Journal of Multivariate Analysis*, 93:223–237, 2005.
- V. Gabriel. Testing for the null hypothesis of cointegration: A Monte Carlo comparison. *Econometric Reviews*, 22(4):411–435, 2003.
- J.L. Gallego. The exact likelihood function of a vector autoregressive moving average process. *Statistics and Probability Letters*, 79:711–714, 2009.
- J.L. Gallego y C. Díaz. Testing for deterministic components in vector seasonal time series. *Open Journal of Statistics*, 1(2):145–150, 2011.
- C. Gourieroux y A. Monfort. *Statistical and Econometric Models*, volume 2. Cambridge University Press, 1995. Testing, confidence regions, model selection and asymptotic theory.
- C.W.J. Granger. Some properties of time series data and their use in econometric model specification. *Journal of Econometrics*, 16:121–130, 1981.
- J.D. Hamilton. *Time Series Analysis*. Princeton, 1994.
- B.E. Hansen. Lagrange multiplier tests for parameter instability in non-linear models. *Unpublished manuscript, University of Rochester, Dept. of Economics*, 1990.
- B.E. Hansen. Tests for parameter instability in regressions with I(1) processes. *Journal of Business and Economics Statistics*, 10(3):321–335, 1992.
- D. Harris. Principal components analysis of cointegrated time series. *Econometric Theory*, 13:529–557, 1997.
- D. Harris y B. Inder. *Nonstationary Time Series Analysis and Cointegration*, chapter A Test of the Null Hypothesis of Cointegration, pages 133–152. Oxford University Press, Oxford, 1994.
- A.C. Harvey. *Forecasting, Structural Time Series and the Kalman Filter*. Cambridge University Press, 1989.
- S.C. Hillmer y G.C. Tiao. Likelihood function of stationary multiple autoregressive moving average models. *Journal of the American Statistical Association*, 74(367): 652–660, 1979.

- J.P. Imhof. Computing the distribution of quadratic forms in normal variables. *Biometrika*, 48:419–426, 1961.
- M. Jansson. Point optimal tests for the null hypothesis of cointegration. *Journal of Econometrics*, 124:187–201, 2005.
- G.M. Jenkins y A.S. Alavi. Some aspects of modelling and forecasting multivariate time series. *Journal of Time Series Analysis*, 2(1):1–47, 1981.
- S. Johansen. Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control*, 12:231–254, 1988.
- S. Johansen. Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica*, 59(6):1551–1580, 1991.
- S. Johansen. Cointegration: An overview. *Department of Applied Mathematics and Statistics working paper series. University of Copenhagen*, 2004.
- S. Johansen y E. Schaumburg. Likelihood analysis of seasonal cointegration. *Journal of Econometrics*, 88:301–339, 1999.
- N.L. Johnson y S. Kotz. *Continuous Univariate Distributions*, volume 2. Wiley, 1970.
- K. Juselius. *The Cointegrated VAR Model: Methodology and Applications*. Oxford University Press, 2006.
- K.R. Kadiyala. Testing for the independence of regression disturbances. *Econometrica*, 38(1):97–117, 1970.
- M.L. King. Robust tests for spherical symmetry and their application to least squares regression. *The Annals of Statistics*, 8(6):1265–1271, 1980.
- M.L. King y G.H. Hillier. Locally best invariant tests of the error covariance matrix of the linear regression model. *Journal of the Royal Statistical Society. Series B (Methodological)*, 47(1):98–102, 1985.
- E. Kurozumi y Y. Arai. Test for the null hypothesis of cointegration with reduced size distortion. *Journal of Time Series Analysis*, 29(3):476–500, 2008.
- E. Kurozumi y K. Hayakawa. Asymptotic properties of the efficient estimators for cointegrating regression models with serially dependent errors. *Journal of Econometrics*, 149:118–135, 2009.

- D. Kwiatkowski, P. Phillips, P. Schmidh, y Y. Shin. Testing the null hypothesis of stationarity against the alternative of a unit root: How sure we are that economic time series have a unit root? *Journal of Econometrics*, 54:159–178, 1992.
- E.L. Lehmann y J.P. Romano. *Testing Statistical Hypotheses*. Springer, third edition, 2005.
- S.J. Leybourne y B.P.M. McCabe. A simple test for cointegration. *Oxford Bulletin of Economics and Statistics*, 55(2):97–103, 1993.
- S.J. Leybourne y B.P.M. McCabe. A consistent test for a unit root. *Journal of Business Economics and Statistics*, 12(2):157–166, 1994.
- H. Lütkepohl. *New Introduction to Multiple Time Series Analysis*. Springer-Verlag, first edition, 2005.
- H. Lütkepohl y H. Claessen. Analysis of cointegrated VARMA processes. *Journal of Econometrics*, 80:223–239, 1997.
- J.G. MacKinnon. Approximate asymptotic distribution functions for unit root and cointegration tests. *Journal of Business and Economic Statistics*, 12(2):167–176, 1994.
- J.A. Mauricio. Exact maximum likelihood estimation of stationary vector ARMA models. *Journal of the American Statistical Association*, 90(429):282–291, 1995.
- J.A. Mauricio. Exact maximum likelihood estimation of partially nonstationary vector arma models. *Computational Statistics and Data Analysis*, 50:3644–3662, 2006.
- C. Mazas. *Modelos de series temporales no invertibles*. Tesis doctoral, 2012.
- B.P.M. McCabe, S.J. Leybourne, y Y. Shin. A parametric approach to testing the null of cointegration. *Journal of Time Series Analysis*, 18(4):395–413, 1997.
- J. Nyblom. Testing for deterministic linear trend in time series. *Journal of the American Statistical Association*, 81(394):545–549, 1986.
- J. Nyblom. Invariance tests for covariance structures in multivariate linear model. *J.Multivar. Anal.*, 76(2), 2001.
- J. Nyblom y A. Harvey. Tests of common stochastic trends. *Econometric Theory*, 16: 176–199, 2000.
- J. Nyblom y A. Harvey. Testing against smooth stochastic trends. *Journal of Applied Econometrics*, 16:415–429, 2001.

- J. Nyblom y T. Mäkeläinen. Comparisons of tests for the presence of random walk coefficients in a simple linear model. *Journal of the American Statistical Association*, 78(384):856–864, 1983.
- J.Y. Park. Canonical cointegration regressions. *Econometrica*, 60(1):119–143, 1992.
- J.Y. Park y P.C.B. Phillips. Statistical inference in regression with integrated processes: Part 1. *Econometric Theory*, 4:468–487, 1988.
- J.Y. Park y P.C.B. Phillips. Statistical inference in regression with integrated processes: Part 2. *Econometric Theory*, 5:85–131, 1989.
- K. Patterson. *Unit Root Tests in Time Series Volume 1: Concepts and Problems*. Palgrave Texts in Econometrics. Palgrave MacMillan, UK, 2011.
- D. Peña y G.E.P. Box. Identifying a simplifying structure in time series. *Journal of the American Statistical Association*, 82:836–843, 1987.
- D. Peña y P. Poncela. Nonstationary dynamic factor analysis. *Journal of Statistical Planning and Inference*, 136:1237–1257, 2006.
- P.C.B. Phillips. Understanding spurious regressions in econometrics. *Journal of Econometrics*, 33:311–340, 1986.
- P.C.B. Phillips. Towards a unified asymptotic theory for autoregression. *Biometrika*, 74:535–547, 1987.
- P.C.B. Phillips. Optimal inference in cointegrated systems. *Econometrica*, 59(2):283–306, 1991.
- P.C.B. Phillips. Fully modified least squares and vector autoregression. *Econometrica*, 63(5):1023–1078, 1995.
- P.C.B. Phillips y S. Durlauf. Multiple time series regression with integrated processes. *Review of Economic Studies*, 53(4):473–495, 1986.
- P.C.B. Phillips y B.E. Hansen. Statistical inference in instrumental variables regression with  $I(1)$  processes. *Review of Economic Studies*, 57:99–125, 1990.
- P.C.B. Phillips y S. Jin. The KPSS test with seasonal dummies. *Economics Letters*, 77:239–243, 2002.
- P.C.B. Phillips y M. Loretan. Estimating long-run economic equilibria. *Review of Economic Studies*, 58:473–495, 1991.

- P.C.B. Phillips y S. Ouliaris. Testing for cointegration using principal components methods. *Journal of Economics Dynamics and Control*, 12:205–230, 1988.
- P.C.B. Phillips y S. Ouliaris. Asymptotic properties of residual based tests for cointegration. *Econometrica*, 58(1):165–193, 1990.
- D.S. Poskitt. On the identification and estimation of nonstationary and cointegrated armax systems. *Econometric Theory*, 22:1138–1175, 2006.
- B.M. Pötscher. Noninvertibility and pseudo-maximum likelihood estimation of misspecified ARMA models. *Econometric Theory*, 7:435–449, 1991.
- C.R. Rao. *Linear Statistical Inference and its Applications*. Wiley, New York, 2nd. edition, 2002.
- G.C. Reinsel. *Elements of Multivariate Time Series Analysis*. Springer Series in Statistics. Springer, second edition, 1997.
- G.C. Reinsel y S.K. Ahn. Vector AR models with unit roots and reduced rank structure: Estimation, likelihood ratio test and forecasting. *Journal of Time Series Analysis*, 13: 353–375, 1992.
- P. Saikkonen. Asymptotically efficient estimation of cointegration regressions. *Econometric Theory*, 7:1–21, 1991.
- P. Saikkonen y R. Luukkonen. Testing for a moving average unit root in autoregressive integrated moving average models. *Journal of the American Statistical Association*, 88(422):596–601, 1993a.
- P. Saikkonen y R. Luukkonen. Point optimal tests for testing the order of differencing in arima models. *Econometric Theory*, 9:343–362, 1993b.
- P. Saikkonen y R. Luukkonen. Testing for cointegration in infinite order vector autoregressive processes. *Journal of Econometrics*, 81:93–126, 1997.
- Y. Shin. A residual-based test of the null of cointegration against the alternative of no cointegration. *Econometric Theory*, 10:91–115, 1994.
- J.H. Stock. Asymptotic properties of least-squares estimators of co-integrating vectors. *Econometrica*, 55:1035–1056, 1987.
- J.H. Stock y M.W. Watson. Testing for common trends. *Journal of the American Statistical Association*, 83:1097–1107, 1988.

- J.H. Stock y M.W. Watson. A simple estimator of cointegrating vectors in higher order integrated systems. *Econometrica*, 61(4):783–820, 1993.
- W.K. Tam y G.C. Reinsel. Tests for seasonal moving average unit root in ARIMA models. *Journal of the American Statistical Association*, 92(438):725–738, 1997.
- W.K. Tam y G.C. Reinsel. Seasonal moving-average unit root tests in the presence of a linear trend. *Journal of Time Series Analysis*, 19(5):609–625, 1998.
- K. Tanaka. Testing for a moving average unit root. *Econometric Theory*, 6:433–444, 1990.
- K. Tanaka. An alternative approach to the asymptotic theory of spurious regression, cointegration, and near cointegration. *Econometric Theory*, 9:36–61, 1993.
- A.M.R. Taylor. Locally optimal tests against unit roots in seasonal time series processes. *Journal of Time Series Analysis*, 24(5):591–612, 2003a.
- A.M.R. Taylor. Robust stationarity tests in seasonal time series processes. *Journal of Business and Economic Statistics*, 21(1):156–163, 2003b.
- G.C. Tiao y G.E.P. Box. Modelling multiple time series with applications. *Journal of the American Statistical Association*, 76(376):802–816, 1981.
- M.C.K. Tweedie. Statistical properties of Inverse Gaussian distributions i. *The Annals of Mathematical Statistics*, 28(2):362–377, 1957a.
- M.C.K. Tweedie. Statistical properties of Inverse Gaussian distributions ii. *The Annals of Mathematical Statistics*, 28(2):362–377, 1957b.
- F. Vahid y R.F. Engle. Common trends and common cycles. *Journal of Applied Econometrics*, 8:341–360, 1993.
- M. Wagner. Cointegration analysis with state space models. *AStA Advances in Statistical Analysis*, 94:273–305, 2010.
- S.F. Yap y G.C. Reinsel. Estimation and testing for unit roots in a partially nonstationary vector autoregressive moving average model. *Journal of the American Statistical Association*, 90(429):253–267, 1995.