

UNIVERSIDAD COMPLUTENSE DE MADRID

FACULTAD DE CIENCIAS ECONÓMICAS Y EMPRESARIALES

**Departamento de Fundamentos del Análisis Económico I
(Análisis Económico)**



**IDENTIFICACIÓN DE MODELOS PARA SERIES TEMPORALES
MEDIANTE MÉTODOS DE SUBESPACIOS**

MEMORIA PARA OPTAR AL GRADO DE DOCTOR

PRESENTADA POR

Alfredo García Hiernaux

Bajo la dirección de los doctores

Miguel Jérez Méndez
José Manuel Casals Carro

Madrid, 2005

ISBN: 978-84-669-2978-3

IDENTIFICACIÓN DE MODELOS PARA SERIES TEMPORALES MEDIANTE MÉTODOS DE SUBESPACIOS

Alfredo García Hiernaux¹

Dirigida por los doctores D. Miguel Jerez Méndez
y D. José Manuel Casals Carro

Septiembre 2004

¹Departamento de Fundamentos del Análisis Económico II. Facultad de Ciencias Económicas. Campus de Somosaguas. 28223 Madrid (SPAIN). email: *agarciah@ccee.ucm.es*

A Marta y a mis padres,

Índice general

1. Introducción	1
2. Revisión de Sistemas Lineales y Métodos de Subespacios	7
2.1. Modelos en Forma de Espacio de los Estados	7
2.2. Controlabilidad y Observabilidad	8
2.3. Definiciones de la Estructura Subespacial	10
2.4. Métodos de Subespacios	12
3. Raíces Unitarias y Cointegración	15
3.1. Correlaciones Canónicas y Raíces Unitarias	16
3.1.1. Procesos univariantes	18
3.1.2. Procesos multivariantes	24
3.2. Cointegración	33
3.2.1. Rango de cointegración	33
3.2.2. Estimación de la matriz de cointegración	47
3.3. Conclusiones del Capítulo	53
4. Orden del Sistema e Índices de Kronecker	54
4.1. Selección del Orden del Sistema	55
4.1.1. Procedimientos a priori	55
4.1.2. Procedimientos a posteriori	61
4.1.3. Ejercicios de simulación	64
4.2. Índices de Kronecker	79
4.2.1. Especificación de los índices de Kronecker	79
4.2.2. Análisis empírico de especificación automática	83
4.3. Conclusiones del Capítulo	87
5. Estimación Restringida Basada en Métodos de Subespacios	88
5.1. Algoritmo de Estimación	89
5.2. Ejercicios de Simulación	92
5.3. Conclusiones del Capítulo	100

6. Resultados Empíricos	101
6.1. Modelos Univariantes	102
6.1.1. Tráfico de Pasajeros de Líneas Aéreas (Box y Jenkins, 1976)	102
6.1.2. Identificación semi-automática de modelos para series tem- porales reales y simuladas	106
6.2. Modelos Multivariantes	112
6.2.1. Series de Precios de la Harina (Tiao y Tsay, 1989)	112
6.2.2. Series de tipos de interés de EEUU (Martín Manjón y Tread- way, 1997)	117
7. Conclusiones	128
A. Apéndice Matemático	134
A.1. Definiciones de Subespacios y Matrices Especiales	134
A.2. Descomposición en Valores Singulares (SVD)	136
A.3. Coeficientes de Correlación Canónica (CCC)	137
B. Apéndice de Algoritmos	138
B.1. Algoritmo NID	138
B.2. Algoritmo SUBEST1	141
B.3. Algoritmo SUBEST1SEAS	143
B.4. Algoritmo SUBEST2	144
Tabla de Símbolos	145

Capítulo 1

Introducción

Los modelos de series temporales tienen cuatro aplicaciones fundamentales en Economía: (i) la previsión de valores futuros, (ii) la contrastación empírica de hipótesis teóricas, (iii) la medición e interpretación de las relaciones dinámicas y (iv) el diseño de políticas para intervenir en el sistema dinámico, influyendo sobre los valores futuros de sus variables.

A mediados de los años 70, Box y Jenkins (1976) plantearon una metodología de construcción empírica de modelos lineales de coeficientes constantes para series temporales, fundamentada en tres ideas básicas que podemos resumir como: “dejar que los datos hablen”, “parametrizar con austeridad” y “permitir el error”.

“Dejar que los datos hablen” quiere decir que la especificación del modelo debe descansar en las propiedades observables de la muestra analizada. Es posible y deseable restringir las relaciones entre las variables, pero cualquier restricción que se imponga debe tener un sólido fundamento teórico y ser compatible con las características de los datos.

“Parametrizar con austeridad”, o “principio de parsimonia”, significa en la práctica que, si dos modelos poseen la misma capacidad descriptiva y predictiva, deberá preferirse aquél que sea más sencillo. Esta idea puede verse, por tanto, como una traducción del principio de la navaja de Occam a la modelización estadística.

Finalmente, “permitir el error” supone que toda especificación debe considerarse tentativa y someterse a un proceso estructurado de revisión crítica. Esta es la idea subyacente en el esquema clásico de identificación-estimación-diagnóstico que propusieron Box y Jenkins (1976) y que desde entonces, de una forma u otra, siguen todos los analistas de series temporales.

Existe un amplio consenso, prácticamente unánime, acerca de la validez de estas ideas. Hay un acuerdo similar acerca de cómo ponerlas en práctica cuando se está analizando una sola serie temporal (caso univariante). En cambio, no puede decirse que haya unanimidad acerca de la forma más adecuada de modelizar series múltiples (caso multivariante).

En el caso lineal univariante las decisiones básicas de modelización son dos: ¿cómo transformar los datos para inducir en ellos estacionariedad en media y varianza? y ¿cómo parametrizar de la forma más austera posible las autocorrelaciones muestrales de los datos? Una vez tomadas estas decisiones se dispone de una especificación preliminar que, concretada en la fase de estimación y refinada en la de diagnóstico y revisión, conduce al modelo final.

Frente a esta relativa simplicidad, el análisis de vectores de series temporales plantea un panorama mucho más complejo.

Por ejemplo, cuando se estudia cómo inducir estacionariedad en media en una sola serie temporal la decisión se limita a decidir si la serie ya es estacionaria o, por el contrario, requiere diferencias (generalmente una o dos) para convertirse en estacionaria. Nos enfrentamos, por tanto, a elegir una de entre tres alternativas. Si la serie presenta algún tipo de estacionalidad las opciones se vuelven algo más complejas, ya que debe decidirse si se aplica o no una diferencia estacional. Consecuentemente, en los casos más comunes, debemos elegir una opción de entre tres (si no hay estacionalidad) o seis (si la hay). Resulta inmediato que al modelizar m series, el analista debe elegir una de entre $3m$ ó $6m$ opciones. Además, si la opción elegida no coincide con la que se tomaría para cada serie por separado estaríamos en el caso de cointegración [Box y Tiao (1977) , Granger y Engle (1987), Banerjee et al. (1993)], en donde la transformación estacionaria más adecuada viene dada por una complicada matriz que se compone de diferencias y combinaciones lineales de las series.

Otra distinción clara entre el análisis univariante y multivariante se refiere a las propiedades de los modelos en forma reducida más habituales. En análisis univariante suelen usarse modelos de la familia ARIMA, o que pueden representarse en una forma ARIMA equivalente. Análogamente, los modelos de series múltiples más comunes pertenecen a la familia VARMAX, o admiten una representación VARMAX equivalente. Pues bien, en ausencia de raíces redundantes, los modelos ARIMA son formas canónicas, mientras que los VARMAX no lo son en general. Esto supone que en series múltiples la especificación de los órdenes de los polinomios autorregresivo y media móvil no garantiza que el proceso estocástico tiene el mínimo número de parámetros. Por tanto, en la práctica, es mucho más difícil

seguir el principio de austeridad paramétrica.

Esta Tesis propone una metodología para especificar modelos lineales para series temporales, tanto individuales como múltiples. En este sentido, enlaza con una amplia literatura previa, de la que cabe destacar aportaciones como la ya citada de Box y Jenkins (1976), o las de Sims (1980a,b), Tiao y Box (1981), Jenkins y Alavi (1981) y Reinsel (1997).

Como en muchos de estos casos, la primera fase del análisis tiene como objetivo determinar una transformación estacionaria tentativa de los datos. La posible existencia de raíces unitarias y/o relaciones de cointegración se estudia a partir de las propiedades de sus coeficientes de correlación canónica [Hotelling (1936), Akaike (1975, 1976), Tsay y Tiao (1985), Tiao y Tsay (1989)], en la línea de los trabajos previos de Vargas (1999), Poskitt (2000) o Bauer y Wagner (2002a).

Una vez transformadas las series, la segunda fase se centra en determinar el número de componentes dinámicos (orden del proceso o índice de McMillan) que caracteriza el mínimo orden de dinámica del sistema. Para ello se estima una secuencia de modelos que recorren un espacio de búsqueda predefinido para el índice de McMillan. Los modelos sucesivos se estiman rápida y eficientemente aplicando técnicas de subespacios [Van Overschee y De Moor (1996), Favoreel et al. (2000)], que se reducen básicamente a un problema de mínimos cuadrados generalizados con restricción de rango, y se comparan entre sí utilizando contrastes de razón de verosimilitud [Tsay (1989a) y Tiao y Tsay (1989)], criterios de información [Akaike (1974), Schwartz (1978) y Hannan y Quinn (1979)], criterios de correlación canónica (Bauer, 2001) y otros derivados en la Tesis.

Consecuentemente, la segunda fase requiere tan sólo decidir el valor tentativo de un escalar. Dada esta decisión, se dispone de: a) estimaciones eficientes de los parámetros de un modelo en forma de innovaciones que, si así se desea, admite una representación VARMAX alternativa y b) matrices de residuos que pueden usarse para revisar la decisión sobre el orden del sistema. En este momento, el econométra puede decidir si este modelo es suficiente para conseguir los objetivos de su análisis o si, por el contrario, debe continuar elaborándolo en fase posteriores.

En el caso de series múltiples (y sólo en éste) puede considerarse adecuado abordar una tercera fase, en la que se buscan ganancias en austeridad paramétrica imponiendo un vector de índices de observabilidad o “de Kronecker” al modelo. Estos índices pueden interpretarse como los órdenes de la dinámica de cada uno de los elementos del vector de variables endógenas. Para ello, se estiman modelos para todas las combinaciones posibles de índices de Kronecker que sumen el or-

den del sistema, especificado en la fase anterior. Los modelos no están anidados, por lo que se comparan entre sí mediante criterios de información. A diferencia de otros métodos [Hannan y Kavalieris (1984), Poskitt (1992), Lütkepohl (1993), Tsay (1989b) o Li y Tsay (1998)] nuestra propuesta utiliza nuevamente técnicas de subespacios y se restringe a los vectores de índices de Kronecker compatibles con las decisiones de las dos fases previas.

El resultado final de esta tercera fase es un modelo econométrico canónico, más austero por tanto que el de la fase anterior, que puede someterse nuevamente a diagnosis. Este modelo puede representarse en espacio de los estados en la forma canónica de Luenberger o, alternativamente, como un modelo VARMAX echelon [Hannan y Deistler (1988), Casals (1997) y Reinsel (1997)]. A partir de cualquiera de estas formas canónicas el modelo puede escribirse en cualquier otra representación que se considere oportuna como, por ejemplo, VARMAX.

La cuarta y última fase consiste en añadir al modelo canónico, univariante o multivariante, restricciones de exclusión que aportan parsimonia adicional mediante la eliminación de parámetros no significativos. Para estimar los sucesivos modelos y realizar los contrastes necesarios pueden usarse técnicas estándar de máxima verosimilitud o un método iterativo eficiente, que proponemos en esta Tesis, basado nuevamente en técnicas de subespacios.

Un aspecto peculiar de nuestra propuesta es, por tanto, que asume con matices el principio de “austeridad paramétrica”. En nuestra opinión, la austeridad no es una virtud absoluta por varias razones. En primer lugar, para conseguirla es necesario tomar decisiones de modelización más complejas y, por consiguiente, asumir un mayor riesgo de error y unos costes de modelización más elevados. En segundo lugar, la sobreparametrización no perjudica determinados usos del modelo econométrico como, por ejemplo, la previsión. Por último, la importancia de una parametrización austera era mayor en un período histórico en donde se disponía de relativamente pocas series temporales que, además, eran generalmente cortas. La aplicación intensa de las tecnologías de la información en todos los ámbitos de la Economía ha cambiado radicalmente este contexto. La cuestión no es, por tanto, ¿cómo conseguir austeridad? sino ¿cómo conseguir la austeridad necesaria?

La Tesis tiene la siguiente estructura:

El Capítulo 2 está dedicado a la revisión de sistemas lineales e introducción a los métodos de subespacios. En él se plantea el problema de identificación y estimación que resuelven estos procedimientos y se introduce la notación convencional usada en esta literatura y en la propia Tesis.

En el Capítulo 3 se propone un método para determinar el orden de integración y el rango de cointegración (si existe) de un proceso. Este paso se realiza proponiendo un criterio de decisión creado a partir de la estimación de sus funciones de penalización, que reposan en las propiedades estadísticas de los coeficientes de correlación canónica. Además, cuando se confirma la existencia de tendencias comunes, se propone un método para estimar consistente la matriz de cointegración. Finalmente, se presenta un apartado con ejercicios de Monte Carlo que contrasta los resultados teóricos.

El Capítulo 4 analiza el problema de la identificación del orden del sistema y de la especificación de los índices de Kronecker. En primer lugar, se presentan algunos criterios denominados “a priori”, ya que no requieren la estimación de modelos, y otros criterios “a posteriori”, cuya aplicación si requiere un paso previo de estimación. En el primero de los casos, se proponen dos criterios y se demuestra su consistencia asintótica.

En una segunda parte, se estudia la especificación de los índices de Kronecker. Para ello, se propone un procedimiento de búsqueda entre todos los juegos de índices de observabilidad que sumen el orden del sistema obtenido en la etapa anterior, utilizando métodos de subespacios para estimar los modelos y criterios de entropía para su comparación. Como en el Capítulo anterior se ofrecen los resultados de varios ejercicios de simulación.

En el Capítulo 5 se desarrolla un nuevo algoritmo iterativo de estimación, basado en los métodos de subespacios y el filtro de Kalman. Con este propósito, se deriva la función de máxima verosimilitud que implica la relación subsespacial entre la información futura y pasada. El procedimiento resulta útil para depurar el modelo obtenido en la fase anterior y generar condiciones iniciales adecuadas para la estimación máximo-verosímil o también para obtener estimaciones finales. Además, el uso de la representación en espacio de los estados implica un alto grado de independencia en cuanto a las formulaciones, pudiendo aplicarse a cualquier modelo lineal con coeficientes constantes que pueda representarse de esta forma. Por último, se contrasta el procedimiento de estimación en un estudio de Monte Carlo, en el que se compara en términos de precisión, eficiencia estadística y coste computacional con un algoritmo de máxima verosimilitud exacta.

El Capítulo 6 ilustra la aplicación práctica, con datos reales y series simuladas, de los algoritmos propuestos en la Tesis. Se desarrolla paso a paso la metodología de identificación, especificación y estimación presentada en los Capítulos anteriores, tanto en el caso univariante regular y/o estacional como en el multivariante,

con o sin cointegración.

La primera parte muestra la metodología completa de identificación y estimación para modelos univariantes con series de datos reales y simulados estudiadas en la literatura estadística y econométrica, para las que existen modelos de referencia. La segunda parte está dedicada al análisis multivariante, centrándose en tres conjuntos de datos con propiedades dispares.

En el Capítulo 7 se resumen las principales aportaciones de este trabajo y se proponen futuras líneas de investigación a partir de los resultados obtenidos. Estas incluyen: refinar el tratamiento de vectores de series estacionales, estudiar la posibilidad del análisis de cointegración estacional y adaptar la metodología para el estudio de modelos canónicos de heteroscedasticidad condicional.

Finalmente se incluyen dos apéndices. El primero define conceptos de subespacios y algunas matrices utilizadas en el trabajo y revisa dos instrumentos muy utilizados en la Tesis: la descomposición en valores singulares y los coeficientes de correlación canónica. En el segundo se describen los algoritmos derivados en este trabajo.

Capítulo 2

Revisión de Sistemas Lineales y Métodos de Subespacios

En este Capítulo se establece la notación usada en el trabajo y se resumen los principales resultados de los métodos de subespacios. Estos nacen como una mezcla entre la teoría y la geometría de sistemas lineales y el álgebra lineal numérica. Como artículos y manuales de referencia sobre identificación con métodos de subespacios cabe citar a Larimore (1990), Van Der Veen et al. (1993), Viberg (1995) o Van Overschee y De Moor (1996).

2.1. Modelos en Forma de Espacio de los Estados

Los métodos de subespacios están relacionados con modelos y sistemas en forma de Espacio de los Estados (EE). A lo largo de este trabajo, se considera \mathbf{z}_t como un vector de variables endógenas relacionado con su pasado y con los valores presentes de \mathbf{u}_t , un vector de inputs exógenos, a través de una dependencia lineal. Además, se asume que este modelo puede escribirse en forma de EE. La representación más general es la siguiente:

Definición 2.1.1 (Modelo en Espacio de los Estados)

$$\begin{aligned}\mathbf{x}_{t+1} &= \mathbf{\Phi}\mathbf{x}_t + \mathbf{\Gamma}\mathbf{u}_t + \mathbf{E}\mathbf{w}_t \\ \mathbf{z}_t &= \mathbf{H}\mathbf{x}_t + \mathbf{D}\mathbf{u}_t + \mathbf{C}\mathbf{v}_t\end{aligned}\tag{2.1.1}$$

donde $\mathbf{x}_t \in \mathbb{R}^n$ es el vector de estado, $\mathbf{u}_t \in \mathbb{R}^r$ es el vector de inputs exógenos, y $\mathbf{z}_t \in \mathbb{R}^m$ es el vector de outputs observables. Además, \mathbf{w}_t y \mathbf{v}_t son secuencias de perturbaciones tales que,

$$E(\mathbf{w}_t) = \mathbf{0}, E(\mathbf{v}_t) = \mathbf{0}, E(\mathbf{w}_t\mathbf{w}_t') = \mathbf{Q}, E(\mathbf{v}_t\mathbf{v}_t') = \mathbf{R} \text{ y } E(\mathbf{v}_t\mathbf{w}_t') = \mathbf{S}$$

La matriz \mathbf{Q} es semidefinida positiva y \mathbf{R} es definida positiva. Se supone además que ni las variables exógenas ni el vector de estado inicial \mathbf{x}_0 , están correlados con las perturbaciones, es decir,

$$E(\mathbf{w}_{t_1} \mathbf{u}'_{t_2}) = 0, \quad E(\mathbf{v}_{t_1} \mathbf{u}'_{t_2}) = 0 \quad \forall t_1, t_2, \quad t_1 \neq t_2$$

$$E(\mathbf{x}_0 \mathbf{w}'_t) = 0, \quad E(\mathbf{x}_0 \mathbf{v}'_t) = 0 \quad \forall t$$

Cuando $\mathbf{w}_t = \mathbf{v}_t = \boldsymbol{\epsilon}_t$ siendo $\boldsymbol{\epsilon}_t \in \mathbb{R}^m$, $\boldsymbol{\epsilon}_t \sim iidN(\mathbf{0}, \mathbf{Q})$ y $\mathbf{C} = \mathbf{I}$, se dice que el modelo (2.1.1) adopta la forma de innovaciones.

Definición 2.1.2 (Modelo en forma de innovaciones)

$$\begin{aligned} \mathbf{x}_{t+1} &= \boldsymbol{\Phi} \mathbf{x}_t + \boldsymbol{\Gamma} \mathbf{u}_t + \mathbf{E} \boldsymbol{\epsilon}_t \\ \mathbf{z}_t &= \mathbf{H} \mathbf{x}_t + \mathbf{D} \mathbf{u}_t + \boldsymbol{\epsilon}_t \end{aligned} \tag{2.1.2}$$

Un amplio abanico de modelos de series temporales que incluye los modelos de función de transferencia, modelos VAR y VARMAX, pueden representarse directamente en un modelo en EE en forma de innovaciones (ver Aoki, 1990; Terceiro, 1990). Además, como se demuestra en Casals et al. (1999), bajo supuestos relativamente débiles, cualquier modelo en la forma general (2.1.1) puede escribirse como (2.1.2) por medio de una transformación de sus parámetros. Por ello y debido a que la formulación (2.1.2) reduce considerablemente el espacio paramétrico, en este trabajo se utiliza principalmente la representación (2.1.2).

El problema que se plantea a partir de (2.1.2) consiste en estimar las matrices $\boldsymbol{\Phi} \in \mathbb{R}^{n \times n}$, $\boldsymbol{\Gamma} \in \mathbb{R}^{n \times r}$, $\mathbf{H} \in \mathbb{R}^{m \times n}$, $\mathbf{D} \in \mathbb{R}^{m \times r}$, $\mathbf{E} \in \mathbb{R}^{n \times m}$ y $\mathbf{Q} \in \mathbb{R}^{m \times m}$ usando únicamente las series temporales del output y los inputs observables.

La teoría de identificación de sistemas lineales trata de encontrar las matrices, ya sea del modelo (2.1.1) o de (2.1.2), que generen realizaciones de dimensión mínima, es decir, que usen el mínimo número de estados necesarios para representar el sistema dinámico que produce el output observado. En concreto, los métodos de subespacios buscan la identificación de una realización de dimensión mínima. Para que esto se cumpla, el sistema representado debe ser controlable y observable.

2.2. Controlabilidad y Observabilidad

A partir del modelo (2.1.2), el vector de estado en el instante $t = n$ puede formularse como una función lineal del vector de estado inicial \mathbf{x}_0 , los inputs

observables \mathbf{u}_t , y el input no observable $\boldsymbol{\epsilon}_t$, entre los instantes $t = 0$ y $t = n - 1$. Esto es,

$$\mathbf{x}_n = \Phi^n \mathbf{x}_0 + \begin{pmatrix} \Phi^{n-1}(\Gamma, \mathbf{E}) & \dots & \Phi(\Gamma, \mathbf{E}) & (\Gamma, \mathbf{E}) \end{pmatrix} \begin{pmatrix} \mathbf{u}_0 \\ \boldsymbol{\epsilon}_0 \\ \vdots \\ \mathbf{u}_{n-2} \\ \boldsymbol{\epsilon}_{n-2} \\ \mathbf{u}_{n-1} \\ \boldsymbol{\epsilon}_{n-1} \end{pmatrix} \quad (2.2.1)$$

donde $\mathbf{C}_n = \begin{pmatrix} \Phi^{n-1}(\Gamma, \mathbf{E}) & \dots & \Phi(\Gamma, \mathbf{E}) & (\Gamma, \mathbf{E}) \end{pmatrix}$ es la matriz de controlabilidad del sistema. No obstante, debe tenerse en cuenta que el rango de la matriz $(\boldsymbol{\epsilon}'_0 \ \boldsymbol{\epsilon}'_1 \ \dots \ \boldsymbol{\epsilon}'_{n-2} \ \boldsymbol{\epsilon}'_{n-1})'$ está relacionado con el rango de la matriz de covarianzas \mathbf{Q} y entonces:

Definición 2.2.1 (Controlabilidad) *Dada la expresión (2.2.1), el sistema (2.1.2) es controlable desde el origen si y sólo si,*

$$rg\left(\Phi^{n-1}(\Gamma, \mathbf{E}\mathbf{Q}^{\frac{1}{2}}) \ \dots \ \Phi(\Gamma, \mathbf{E}\mathbf{Q}^{\frac{1}{2}}) \ (\Gamma, \mathbf{E}\mathbf{Q}^{\frac{1}{2}})\right) = n \quad (2.2.2)$$

donde $rg(\mathbf{A})$ es el rango de la matriz \mathbf{A} .

Por otro lado, partiendo de nuevo del sistema (2.1.2), la relación entre el vector de estado inicial \mathbf{x}_0 y la información del output y de los inputs observables y no observables entre los instantes $t = 0$ y $t = n - 1$ puede representarse (sustituyendo recursivamente en la ecuación de observación la expresión del vector de estado) como:

$$\begin{pmatrix} z_0 \\ z_1 \\ \vdots \\ z_{n-1} \end{pmatrix} = \begin{pmatrix} \mathbf{H} \\ \mathbf{H}\Phi \\ \vdots \\ \mathbf{H}\Phi^{n-1} \end{pmatrix} \mathbf{x}_0 + \begin{pmatrix} \mathbf{D} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{H}\Gamma & \mathbf{D} & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{H}\Phi^{n-2}\Gamma & \mathbf{H}\Phi^{n-3}\Gamma & \dots & \mathbf{D} \end{pmatrix} \begin{pmatrix} \mathbf{u}_0 \\ \mathbf{u}_1 \\ \vdots \\ \mathbf{u}_{n-1} \end{pmatrix} + \begin{pmatrix} \mathbf{I}_m & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{H}\mathbf{E} & \mathbf{I}_m & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{H}\Phi^{n-2}\mathbf{E} & \mathbf{H}\Phi^{n-3}\mathbf{E} & \dots & \mathbf{I}_m \end{pmatrix} \begin{pmatrix} \boldsymbol{\epsilon}_0 \\ \boldsymbol{\epsilon}_1 \\ \vdots \\ \boldsymbol{\epsilon}_{n-1} \end{pmatrix} \quad (2.2.3)$$

donde $\mathbf{O}_n = \begin{pmatrix} \mathbf{H}' & (\mathbf{H}\Phi)' & \dots & (\mathbf{H}\Phi^{n-1})' \end{pmatrix}'$ es la matriz de observabilidad del sistema. Así:

Definición 2.2.2 (Observabilidad) Dada la expresión (2.2.3), el sistema (2.1.2) es observable si y sólo si,

$$rg(\mathbf{O}_n) = n \quad (2.2.4)$$

Por tanto, se dice que el sistema (2.1.2) es controlable desde el origen si, partiendo de $\mathbf{x}_0 = \mathbf{0}$, es posible llegar a cualquier valor de \mathbf{x}_n seleccionando adecuadamente el valor de los inputs. Asimismo, se dice que (2.1.2) es observable si, a partir de la información del output y los inputs entre los instantes $t = 0$ y $t = n-1$, puede determinarse de forma única el vector de estado inicial, \mathbf{x}_0 .

Rosenbrock (1970) muestra que una realización de (2.1.2) es de dimensión mínima, si y sólo si, se cumplen las condiciones de controlabilidad y observabilidad (2.2.2) y (2.2.4).

2.3. Definiciones de la Estructura Subespacial

A continuación, se definen las matrices más empleadas en los algoritmos de subespacios [en el Apéndice A.1 se encuentran algunas definiciones básicas de subespacios y se introducen algunas matrices que no son de uso habitual en Econometría]. Estos parten de la relación expresada en la ecuación (2.2.3) y cumplen las condiciones de controlabilidad y observabilidad. Denotando por i la dimensión de un subespacio o el espacio de filas de una matriz, es posible estructurar los datos en matrices Block Hankel de la siguiente forma,

Definición 2.3.1 (Matrices Block Hankel)

$$\mathbf{Y}_{0:2i-1} = \begin{pmatrix} \mathbf{Y}_p \\ \mathbf{Y}_f \end{pmatrix} = \begin{pmatrix} \mathbf{Y}_{0:i-1} \\ \mathbf{Y}_{i:2i-1} \end{pmatrix} = \begin{pmatrix} \mathbf{y}_0 & \mathbf{y}_1 & \dots & \mathbf{y}_{T-2i} \\ \mathbf{y}_1 & \mathbf{y}_2 & \dots & \mathbf{y}_{T-2i+1} \\ \vdots & \vdots & & \vdots \\ \mathbf{y}_{i-1} & \mathbf{y}_i & \dots & \mathbf{y}_{T-i-1} \\ \mathbf{y}_i & \mathbf{y}_{i+1} & \dots & \mathbf{y}_{T-i} \\ \mathbf{y}_{i+1} & \mathbf{y}_{i+2} & \dots & \mathbf{y}_{T-i+1} \\ \vdots & \vdots & & \vdots \\ \mathbf{y}_{2i-1} & \mathbf{y}_{2i} & \dots & \mathbf{y}_{T-1} \end{pmatrix} \quad (2.3.1)$$

donde $\mathbf{Y}_{0:2i-1}$ puede representar a $\mathbf{Z}_{0:2i-1}$ (matriz Output Block Hankel) o $\mathbf{U}_{0:2i-1}$ y $\mathbf{\epsilon}_{0:2i-1}$ (matrices Input Block Hankel). Obsérvese que i es la dimensión de un subespacio concreto. La elección de la dimensión de los subespacios de información

pasada y futura, p y f respectivamente, se estudia, entre otros, en Viberg (1995), Chui (1997) o Bauer (1998). No obstante, es habitual en la literatura definir ambos subespacios con la misma dimensión (ver Deistler et al., 1995; Peternell et al., 1996; Van Overschee y De Moor, 1996), es decir $p = f = i$. Sin pérdida de generalidad, todos los algoritmos propuestos en esta tesis mantienen esta igualdad.

Definición 2.3.2 (Secuencia de estados)

$$\mathbf{X}_t = (\mathbf{x}_t \quad \mathbf{x}_{t+1} \quad \mathbf{x}_{t+2} \quad \dots \quad \mathbf{x}_{t+T-2i})$$

A partir de esta definición, las secuencias de estados pasados y futuros, que se inician en $t = 0$ y $t = i$ respectivamente, pueden asimismo expresarse como $\mathbf{X}_p = \mathbf{X}_0$ y $\mathbf{X}_f = \mathbf{X}_i$.

Definición 2.3.3 (Matriz de Observabilidad)

$$\mathbf{O}_i = \begin{pmatrix} \mathbf{H} \\ \mathbf{H}\Phi \\ \mathbf{H}\Phi^2 \\ \vdots \\ \mathbf{H}\Phi^{i-1} \end{pmatrix} \in \mathbb{R}^{im \times n} \quad (2.3.2)$$

Definición 2.3.4 (Matriz Determinista Toeplitz triangular inferior por bloques)

$$\mathbf{H}_i^d = \begin{pmatrix} \mathbf{D} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{H}\Gamma & \mathbf{D} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{H}\Phi\Gamma & \mathbf{H}\Gamma & \mathbf{D} & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{H}\Phi^{i-2}\Gamma & \mathbf{H}\Phi^{i-3}\Gamma & \mathbf{H}\Phi^{i-4}\Gamma & \dots & \mathbf{D} \end{pmatrix} \in \mathbb{R}^{im \times im} \quad (2.3.3)$$

Definición 2.3.5 (Matriz Estocástica Toeplitz triangular inferior por bloques)

$$\mathbf{H}_i^s = \begin{pmatrix} \mathbf{I}_m & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{H}\mathbf{E} & \mathbf{I}_m & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{H}\Phi\mathbf{E} & \mathbf{H}\mathbf{E} & \mathbf{I}_m & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{H}\Phi^{i-2}\mathbf{E} & \mathbf{H}\Phi^{i-3}\mathbf{E} & \mathbf{H}\Phi^{i-4}\mathbf{E} & \dots & \mathbf{I}_m \end{pmatrix} \in \mathbb{R}^{im \times im} \quad (2.3.4)$$

Definición 2.3.6 (Matriz de proyección ortogonal) *Dada una matriz $\mathbf{A} \in \mathbb{R}^{m \times n}$ con $\text{rg}(\mathbf{A}) = m \leq n$, la matriz de proyección ortogonal del espacio de filas de \mathbf{A} , se denota como $\mathbf{\Pi}_\mathbf{A}$ y se define: $\mathbf{\Pi}_\mathbf{A} = \mathbf{A}^+ \mathbf{A}$ donde \mathbf{A}^+ es la pseudo inversa de Moore-Penrose de \mathbf{A} . De la misma forma, $\mathbf{\Pi}_\mathbf{A}^\perp = \mathbf{I} - \mathbf{A}^+ \mathbf{A}$ es la matriz de proyección en el espacio nulo de \mathbf{A} [ver Apéndices A.1 y A.2].*

2.4. Métodos de Subespacios

Utilizando las definiciones 2.3.1-2.3.5, la expresión (2.2.3) puede escribirse en estructura subespacial como,

$$\mathbf{Z}_p = \mathbf{O}_i \mathbf{X}_0 + \mathbf{H}_i^d \mathbf{U}_p + \mathbf{H}_i^s \boldsymbol{\epsilon}_p \quad (2.4.1)$$

donde la matriz de observabilidad revela la relación lineal entre \mathbf{Z}_p y la secuencia de estados inicial, mientras que las matrices \mathbf{H}_i^d y \mathbf{H}_i^s recogen el efecto que tienen \mathbf{U}_p y $\boldsymbol{\epsilon}_p$ sobre \mathbf{Z}_p , tanto en la ecuación de observación como a través del vector de estado.

Obsérvese que i tan sólo tiene que cumplir la restricción de que $i \geq n$ (si no es así, $rg(\mathbf{O}_i) = i < n$ que no cumple la condición de rango (2.2.4)). Siempre y cuando se cumpla la condición anterior, los algoritmos de subespacios estiman consistentemente las matrices del sistema independientemente del valor de i . Algunos autores como Akaike (1976), Larimore (1990) o Deistler et al. (1995) proponen seleccionar el valor de i como $\hat{\rho}_{AIC}$, que denota el orden estimado obtenido con un modelo VAR o VARX del sistema. Otros, como Bauer y Wagner (2002b) utilizan $i = 2\hat{\rho}_{AIC}$. Van Overschee y De Moor (1994), Casals (1997) o Ljung (1999) prefieren aplicar directamente criterios de información sobre el resultado que ofrecen los algoritmos de subespacios para un rango de valores de i .

Bajo la hipótesis de que la secuencia futura de vectores de estado es una función lineal de la información pasada, se define la secuencia de estados futuros como,

$$\mathbf{X}_f = \mathbf{A}_n \mathbf{M}_p \quad \text{donde} \quad \mathbf{M}_p = \begin{pmatrix} \mathbf{Z}_p \\ \mathbf{U}_p \end{pmatrix} \quad (2.4.2)$$

y \mathbf{A}_n es una matriz de rango n . Desplazando los subíndices temporales en (2.4.1), se obtiene,

$$\mathbf{Z}_f = \mathbf{O}_i \mathbf{X}_f + \mathbf{H}_i^d \mathbf{U}_f + \mathbf{H}_i^s \boldsymbol{\epsilon}_f \quad (2.4.3)$$

donde puede verse que el output futuro (\mathbf{Z}_f) sólo tiene relación con la información pasada (\mathbf{M}_p) a través del vector de estado (\mathbf{X}_f). Los algoritmos de subespacios resuelven estos *im* modelos de regresión con restricción de rango. Los más comunes, denotados como CVA (Larimore, 1983, 1990), MOESP (Verhaegen, 1994), N4SID (Van Overschee y De Moor, 1994) o la batería de algoritmos SID propuesta por Casals (1997), lo hacen resolviendo el siguiente problema de optimización,

$$\min_{\{\mathbf{O}_i, \mathbf{A}_n, \mathbf{H}_i^d\}} \left\| \mathbf{W}_1 \left(\mathbf{Z}_f - \mathbf{O}_i \mathbf{A}_n \mathbf{M}_p - \mathbf{H}_i^d \mathbf{U}_f \right) \right\|_F^2 \quad (2.4.4)$$

donde $\|\cdot\|_F$ es la norma Frobenius. Esto puede realizarse en dos fases: 1) resolviendo el problema de optimización con restricción de rango,

$$\min_{\{\hat{\mathbf{O}}_i, \hat{\mathbf{A}}_n\}} \left\| \mathbf{W}_1 \left(\mathbf{Z}_f - \hat{\mathbf{O}}_i \hat{\mathbf{A}}_n \mathbf{M}_p \right) \mathbf{W}_2 \right\|_F^2 \quad (2.4.5)$$

donde se obtienen las matrices $\hat{\mathbf{O}}_i$ y $\hat{\mathbf{A}}_n$ mediante la Descomposición en Valores Singulares (SVD, ver Eckart y Young, 1936, y Apéndice A.2) y 2) resolviendo el problema mínimo cuadrático,

$$\min_{\{\hat{\mathbf{H}}_i^d\}} \left\| \mathbf{W}_1 \left(\mathbf{Z}_f - \hat{\mathbf{O}}_i \hat{\mathbf{A}}_n \mathbf{M}_p - \hat{\mathbf{H}}_i^d \mathbf{U}_f \right) \right\|_F^2 \quad (2.4.6)$$

Como puede observarse en el Cuadro 2.4.1, \mathbf{W}_1 y \mathbf{W}_2 son matrices de ponderación que difieren según el algoritmo. Cuando $\mathbf{W}_2 = \mathbf{\Pi}_{\mathbf{U}_f}^\perp$ los problemas (2.4.5) y (2.4.6) conducen a una solución óptima. La expresión (2.4.4) puede interpretarse como un problema de Mínimos Cuadrados Generalizados (MCG) con una restricción de rango que afecta a una parte del modelo y cuyas soluciones son $\hat{\mathbf{O}}_i$, $\hat{\mathbf{A}}_n$ y $\hat{\mathbf{H}}_i^d$, con $rg(\hat{\mathbf{O}}_i) = n$.

Cuadro 2.4.1: Ponderaciones \mathbf{W}_1 y \mathbf{W}_2 en algoritmos de subespacios

Matrices	CVA	MOESP	N4SID	SID
\mathbf{W}_1	$(\mathbf{Z}_f \mathbf{\Pi}_{\mathbf{U}_f}^\perp \mathbf{Z}_f')^{-\frac{1}{2}}$	\mathbf{I}	\mathbf{I}	$(\mathbf{Z}_f \mathbf{\Pi}_{\mathbf{M}_p, \mathbf{U}_f}^\perp \mathbf{Z}_f')^{-\frac{1}{2}}$
\mathbf{W}_2	$\mathbf{\Pi}_{\mathbf{U}_f}^\perp$	$\mathbf{\Pi}_{\mathbf{U}_f}^\perp$	\mathbf{I}	$\mathbf{\Pi}_{\mathbf{U}_f}^\perp$

Resulta relevante para el resto del trabajo destacar que:

- Cuando $\mathbf{W}_1 = \mathbf{\Omega}_1^{-\frac{1}{2}}$ con $\mathbf{\Omega}_1 = \mathbf{Z}_f \mathbf{\Pi}_{\mathbf{U}_f}^\perp \mathbf{Z}_f'$ los Valores Singulares (VS) obtenidos de la descomposición SVD coinciden con los Coeficientes de Correlación Canónica (CCC) entre \mathbf{Z}_f y \mathbf{Z}_p .
- Bajo el supuesto de normalidad de las perturbaciones, para $\mathbf{W}_2 = \mathbf{\Omega}_2^{-\frac{1}{2}}$ y $\mathbf{\Omega}_2 = \mathbf{Z}_f \mathbf{\Pi}_{\mathbf{M}_p, \mathbf{U}_f}^\perp \mathbf{Z}_f'$, que no es más que tomar como matriz de ponderación la varianza del término de error, los estimadores obtenidos al resolver (2.4.4) son máximo-verosímiles.

Por tanto, la solución en dos fases del problema (2.4.4) genera estimaciones óptimas de las matrices $\hat{\mathbf{O}}_i$, $\hat{\mathbf{A}}_i$ y $\hat{\mathbf{H}}_i^d$ (donde $\hat{\mathbf{O}}_i$ cumple la condición de rango (2.2.4)) a partir de las cuales pueden recuperarse las matrices del sistema en EE. De esta forma, se resuelve el problema de identificación. Los procedimientos para obtener las matrices de la representación (2.1.2) varían según los distintos algoritmos de subespacios (ver Favoreel et al., 2000).

Estos métodos, que han tenido un marcado desarrollo en los últimos años en el campo de la identificación de sistemas lineales, están pasando relativamente inadvertidos en la literatura econométrica. Tan sólo algunos autores como Casals (1997), Bauer (1998), Vargas (1999) o Bauer y Wagner (2002b) han aplicado estos procedimientos a series temporales económicas. Con la excepción de los trabajos anteriores, estos métodos no tratan específicamente algunas características de los datos económicos como son la estacionalidad, los errores de observación, la cointegración o la reducida dimensión de las muestras. Asimismo, no prestan especial atención a la reducción de la complejidad o la interpretación de relaciones entre variables. En los Capítulos siguientes se presenta un metodología completa de análisis de series temporales económicas basada en los métodos de subespacios, que atiende a la mayoría de las cuestiones señaladas.

Capítulo 3

Raíces Unitarias y Cointegración

Una vez se ha decidido si es necesario transformar los datos para eliminar problemas de heteroscedasticidad (Box y Cox, 1964), resulta indispensable estudiar cuál es la transformación estacionaria adecuada. Esta cuestión se vuelve aún más relevante en el caso de modelos multivariantes donde entra en juego el concepto de cointegración es decir, la posible existencia de una combinación lineal de una o varias series que genera una nueva serie estacionaria (Box y Tiao, 1977; Granger y Engle, 1987).

En este Capítulo se propone una nueva metodología de detección de raíces unitarias basada en las propiedades asintóticas de los Coeficientes de Correlación Canónica (CCC). Para ello, se propone un procedimiento que permite adaptar empíricamente criterios de información estimando sus funciones de penalización. La metodología es válida tanto para procesos univariantes como multivariantes. En el segundo caso, bajo ciertos supuestos, el procedimiento se extiende fácilmente al cálculo del número de relaciones de cointegración existentes. Por último, partiendo de los resultados anteriores, se propone un método para estimar consistentemente las combinaciones lineales que producen series estacionarias, a partir de un modelo en forma de innovaciones.

Este Capítulo se divide en dos secciones. En la Sección 3.1 se relaciona los CCC con las raíces unitarias y se presenta la metodología de detección para modelos univariantes, desarrollando también un procedimiento para la identificación de la transformación estacionaria en caso de estacionalidad. Además, se extiende el método para procesos multivariantes. En la Sección 3.2 se presenta, primero, el cálculo del número de relaciones de cointegración y, en segundo lugar, un estimador consistente de la matriz de cointegración. En este apartado se ofrecen también los resultados de diferentes ejercicios de simulación. Por último, la Sección 3.3 ofrece las conclusiones más relevantes del Capítulo.

3.1. Correlaciones Canónicas y Raíces Unitarias

El trabajo pionero de Hotelling (1936) introdujo el análisis de correlación canónica. El objetivo es encontrar aquella combinación lineal de un conjunto de variables (\mathbf{x}) que maximice la correlación con otra combinación lineal de un segundo conjunto de variables (\mathbf{y}). Formalmente, consiste en encontrar unas matrices $\boldsymbol{\alpha}_{r \times m}$ y $\boldsymbol{\beta}_{r \times n}$ tal que los r -vectores $\boldsymbol{\epsilon} = \boldsymbol{\alpha}\mathbf{x}$ y $\boldsymbol{\omega} = \boldsymbol{\beta}\mathbf{y}$ capturen la mayor covarianza posible entre \mathbf{x} e \mathbf{y} . Las variaciones canónicas, denominadas $\boldsymbol{\epsilon}_i$ y $\boldsymbol{\omega}_i$ con $i = 1, 2, \dots, r$, poseen la propiedad de que $\boldsymbol{\epsilon}_1$ y $\boldsymbol{\omega}_1$ presentan la más alta correlación entre todas las combinaciones lineales posibles de \mathbf{x} e \mathbf{y} , $\boldsymbol{\epsilon}_2$ y $\boldsymbol{\omega}_2$ la más alta correlación entre todas las combinaciones lineales de \mathbf{x} e \mathbf{y} que están incorreladas con $\boldsymbol{\epsilon}_1$ y $\boldsymbol{\omega}_1$, y así sucesivamente [ver Apéndice A.3].

A partir de esta idea, Poskitt (2000) desarrolla un criterio de selección basándose en la superconsistencia de los CCC correspondientes a raíces unitarias. Esta propiedad se traduce en que los CCC convergen a su verdadero valor a una tasa superior a la de los correspondientes a raíces estacionarias. Esto permite discriminarlos decidiendo cuáles pueden ser iguales a la unidad.

Como señalamos en el Capítulo 2, los CCC entre la información pasada y futura, ordenados de mayor a menor, pueden obtenerse como los Valores Singulares (VS) de la descomposición SVD en la ecuación (2.4.5) cuando se utiliza una matriz de ponderación $\mathbf{W}_1 = (\mathbf{Z}_f \boldsymbol{\Pi}_{\mathbf{U}_f}^\perp \mathbf{Z}_f')^{-\frac{1}{2}}$. Bauer y Wagner (2002b) proponen un método rápido para poder detectar raíces unitarias, fundamentado en la idea de Poskitt (2000), pero adaptado a métodos de subespacios. Esto es, en lugar de calcular el CCC entre \mathbf{z}_t y \mathbf{z}_{t-1} , calculan los VS de (2.4.5), o lo que es lo mismo los CCC entre los bloques de información pasada \mathbf{Z}_p y de información futura \mathbf{Z}_f . Estos autores evalúan una transformación de los j primeros valores singulares estimados $\hat{\sigma}_j$ (donde $j \leq i$, puesto que el número de VS que devuelve la descomposición SVD de (2.4.5) es igual a la dimensión del subespacio de información pasada i), comparándola con una función de penalización $C(T)$, que depende del tamaño muestral. El criterio que presentan se define de la forma siguiente,

$$1 - \hat{\sigma}_j^2 - C(T) < 0 \quad (3.1.1)$$

El VS que cumple esta desigualdad puede considerarse igual a la unidad, correspondiendo a una raíz unitaria del sistema. La ecuación (3.1.1) compara la distancia a cero de $1 - \hat{\sigma}_j^2$ con una función de penalización. Esta representa una cota superior que implica que el valor $1 - \hat{\sigma}_j^2$ que se sitúe por debajo, será considerado igual a cero, con lo que el criterio devolverá al menos j raíces unitarias. Cabe destacar que el análisis de las correlaciones canónicas no puede de detectar directamente

procesos I(2), pues en estos casos tan sólo el primer CCC es igual a la unidad. Por ello es conveniente hablar de “al menos” una raíz unitaria, en este tipo de criterios. No obstante, esto no causa un problema en la práctica, puesto que al diferenciar la serie y analizar los CCC de la variable transformada se obtendrá la segunda raíz igual a la unidad. En cambio, el tratamiento de la estacionalidad sí resulta un problema más complejo, que se analizará posteriormente. La principal ventaja de este tipo de criterios es que se obtienen explorando las características de los VS o CCC y, por tanto, no hace falta estimar ningún modelo a priori. La función de penalización que especifica la literatura para la ecuación (3.1.1) es,

$$C(T) = \frac{\log(T)^2}{T} \quad (3.1.2)$$

que dibuja un pequeño crecimiento cuando está cerca de cero, pero comienza a decrecer lentamente para $T > 7$. Esto supone, en términos prácticos (para muestras con más de siete observaciones), un decaimiento de la penalización a medida que la muestra crece.

Existen, sin embargo, varios inconvenientes en la utilización de la función (3.1.2): 1) los propios autores la califican de “un tanto heurística”, 2) su uso sólo está fundado si el número de raíces unitarias (d) del sistema, es igual a la unidad y, 3) no depende de la dimensión del subespacio de información pasada, i . La segunda de estas cuestiones es muy relevante, ya que no permite asegurar buenos resultados en procesos multivariantes con dos o más raíces unitarias, cuando es en este tipo de modelos donde los métodos de subespacios presenta mayores ventajas relativas. La tercera cuestión también resulta de interés ya que, como se verá más adelante, cambios en la dimensión del subespacio de información pasada influyen en los valores calculados de los CCC.

Para resolver estas dificultades, en esta sección se formula un criterio de decisión de la familia del propuesto por Bauer y Wagner (2002b) que, 1) está basado tanto en la idea de Poskitt, como en un estudio de simulación y, 2) permite su uso en sistemas con un número de raíces unitarias superior a la unidad.

Para ello, se presenta una metodología de obtención de funciones de penalización basada en métodos de Monte Carlo y en el uso de las propiedades estadísticas, anteriormente citadas, de los CCC. La idea consiste en obtener una función que refleje el límite entre los CCC que se consideren iguales a uno y los inferiores a uno, a partir de la propiedad de convergencia a su verdadero valor a tasas diferentes, expuesta por Poskitt (2000). Esta función deberá cumplir dos condiciones:

- *Condición 3.1.1:* depender del número de observaciones T , de la dimensión

del subespacio de información pasada i y del número de raíces unitarias d , que se quiera contrastar, y

- *Condición 3.1.2:* tender a cero cuando T sea infinitamente grande, para asegurar la consistencia del estimador del número de raíces unitarias, \hat{d} .

La función de penalización depende de i debido a la relación positiva existente entre la dimensión del subespacio de información pasada y los CCC. Efectivamente, cuanto mayor es la información utilizada para el cálculo de los VS mayores serán los valores de los CCC. Así, sería un error comparar dos VS calculados a partir de diferentes valores de i , con una misma función de penalización, incluso si corresponden a un mismo conjunto de datos. La forma más general de la familia de criterios que se proponen puede formularse como:

$$f(\hat{\sigma}_j) - G_l(T, i, d) < 0 \quad (3.1.3)$$

donde $f(\hat{\sigma}_j)$ es una transformación que mide la distancia entre el j -ésimo VS o CCC y la unidad, y G_l hace referencia a diferentes funciones de penalización. Cada una de ellas dependerá del número de observaciones, de la dimensión del subespacio de información y del número de raíces unitarias que se desee evaluar. En este trabajo, se proponen como $f(\hat{\sigma}_j)$ las expresiones $1 - \hat{\sigma}_j^2$ y $1 - \hat{\sigma}_j$, aunque otras pueden ser igualmente válidas.

Para asegurar la consistencia del estimador del número de raíces unitarias \hat{d} , se imponen al criterio más general dos condiciones. La primera es una definición formal de la condición 3.1.2,

$$\lim_{T \rightarrow \infty} G_l(T, i, d) = 0 \quad (3.1.4)$$

y la segunda hace referencia al hecho de que si en el límite (cuando $T = \infty$) $f(\hat{\sigma}_j)$ y $G_l(T, i, d)$ son iguales a cero, entonces la expresión (3.1.3) debe escribirse,

$$f(\hat{\sigma}_j) - G_l(T, i, d) \leq 0 \quad (3.1.5)$$

para que el criterio sea consistente.

3.1.1. Procesos univariantes

Para darle una forma funcional concreta a $G_l(T, i, d)$, estudiaremos la distribución empírica de los CCC que no corresponden a raíces unitarias. Para ello, se simulan 1000 repeticiones de modelos estacionarios ($d = 0$) con CCC cercanos a la unidad (aunque diferentes de uno), para todos los instantes temporales t , desde $t = 20, 21, \dots, 500$. Al primer VS resultante de cada una de estas simulaciones se

le aplica la transformación $f(\hat{\sigma}_1)$, resumiendo toda esta información de cada instante temporal en un determinado percentil, que queda representado por la serie y_{1t} . De esta forma, se obtiene una función empírica que mide la distancia a cero, del primer VS que se corresponden a modelos situados en el límite de la no estacionariedad. El problema se reduce ahora a encontrar una función de penalización que cumpla las condiciones (3.1.1) y (3.1.2) anteriormente expuestas ajustándose correctamente a y_{1t} .

Para el caso en que $d = 0$, se considera el modelo loglineal siguiente:

$$\log(y_{1t}) = \alpha + \beta_1 \log(t) + \beta_2 \log(i_t) + \varepsilon_t \quad (3.1.6)$$

donde ε_t corresponde a un término de error e i_t hace referencia a la dimensión del subespacio de información pasada. Como comentamos en el Capítulo 2, en la literatura es habitual tomar i como *una* o *dos* veces el orden estimado obtenido con un modelo VAR o VARX del sistema (Deistler et al., 1995; Bauer y Wagner, 2002b). El objetivo es determinar el número suficiente de retardos que permitan capturar toda la estructura dinámica estadísticamente significativa. También es habitual aplicar criterios de información sobre el resultado que ofrecen los algoritmos de subespacios en un rango de valores de i (Van Overschee y De Moor, 1994; Ljung, 1999). Sin embargo, aunque estos procedimientos generalmente funcionan bien, presentan ciertas desventajas para implementarlos en la metodología que se propone en este trabajo. El uso del orden estimado por AIC en un VAR “largo”, puede mostrar problemas en muestras suficientemente cortas cuando el sistema presenta estructuras media móvil. Esto es así porque imponer un i grande cuando la muestra es pequeña perjudica la estimación mediante subespacios. Obviamente, estas deficiencias se acentúan cuando se utiliza $i = 2\hat{\rho}_{AIC}$. Por otro lado, aunque la utilización de criterios de información sobre modelos estimados, se revela como un método muy eficaz (Casals, 1997), su implementación en los algoritmos que se presentan en la Tesis podría generar, en muchos casos, un coste computacional excesivo. Por ello, para todos los procedimientos que se proponen en este trabajo se define $i_t = \max(4, h_t)$, siendo h_t el entero más cercano a $\log(T)$. Esta alternativa se basa en una regla heurística fundamentada en la experiencia empírica (Casals, 1997) que aconseja el uso de i como el entero más cercano a $\log(T)$. El hecho de imponer una cota inferior a i , disminuye el riesgo de que $i < n$ en muestras suficientemente cortas.

Una vez definidos y_{1t} e i_t , se tienen todas las variables necesarias para estimar (3.1.6). Esta ecuación genera un tipo particular de función de penalización dependiendo del valor de los parámetros estimados. Además, éstos variarán dependiendo del valor del CCC generador de las simulaciones ($\bar{\sigma}$) y del percentil elegido

(ψ) ambos indispensables para definir la variable y_{jt} . Por ello, se realizan múltiples simulaciones alterando estos dos valores. El Cuadro 3.1.1 muestra los parámetros estimados del modelo (3.1.6) para diferentes funciones $f(\hat{\sigma}_j)$ y valores de ψ , y $\bar{\sigma}$.

Cuadro 3.1.1: Funciones de penalización estimadas para contrastar $d > 0$

Distribución Empírica			Valores Estimados			Función de Penalización
$f(\hat{\sigma}_j) = y_{jt}$	$\bar{\sigma}$	ψ	α	β_1	β_2	
$1 - \hat{\sigma}_j^2$	0.975	95	1.09	-0.51	-0.17	$G_1(d > 0)$
$1 - \hat{\sigma}_j^2$	0.975	90	0.60	-0.50	-0.10	$G_2(d > 0)$
$1 - \hat{\sigma}_j^2$	0.95	80	0.10	-0.36	-0.10	$G_3(d > 0)$
$1 - \hat{\sigma}_j^2$	0.95	90	0.52	-0.40	-0.10	$G_4(d > 0)$
$1 - \hat{\sigma}_j$	0.95	90	0.10	-0.44	-0.05	$G_5(d > 0)$
$1 - \hat{\sigma}_j$	0.975	95	0.65	-0.56	-0.16	$G_6(d > 0)$
$\log(T)^2/T^*$			0.98	-0.52	-0.10	$C(T)$

$j = 1$

* Corresponde a la función de penalización de Bauer y Wagner (2002b)

En este cuadro se incorpora también, comparativamente, la función de penalización propuesta en Bauer y Wagner (2002b). Cuando se ajusta el modelo (3.1.6) a la serie generada por esta función, puede observarse que sus valores estimados no se distancian demasiado del resto. Por otro lado, como es lógico, los resultados muestran que diferencias en los valores que definen la distribución empírica conducen a distintas funciones de penalización.

La Figura 3.1.1 muestra una de las distribuciones empíricas seleccionadas y_{1t} , junto con su función de penalización ajustada, $G_4(d > 0)$. Además, a modo ilustrativo, se añade la distribución empírica generada por un CCC asociado a una raíz unitaria w_t , y la función de penalización propuesta en la literatura, $C(T)$. Obsérvese que las series y_{1t} y w_t , definidas por los pares ($\psi = 90, \bar{\sigma} = .95$) y ($\psi = 90, \bar{\sigma} = 1.0$) respectivamente, reflejan sensiblemente la diferencia entre las tasas de convergencia de los CCC a sus valores asintóticos. La primera está relacionada con un modelo estacionario y la segunda con uno no estacionario.

Si se considera que y_{1t} representa suficientemente bien la frontera entre los CCC iguales a la unidad, y los inferiores, entonces la función G_4 , que se ajusta correctamente a la distribución empírica seleccionada, puede establecerse como una función de penalización válida para evaluar si $d > 0$.

El problema se reduce entonces a decidir cuál de todas las posibles series y_{1t} refleja mejor la frontera entre los CCC inferiores e iguales a la unidad. En este

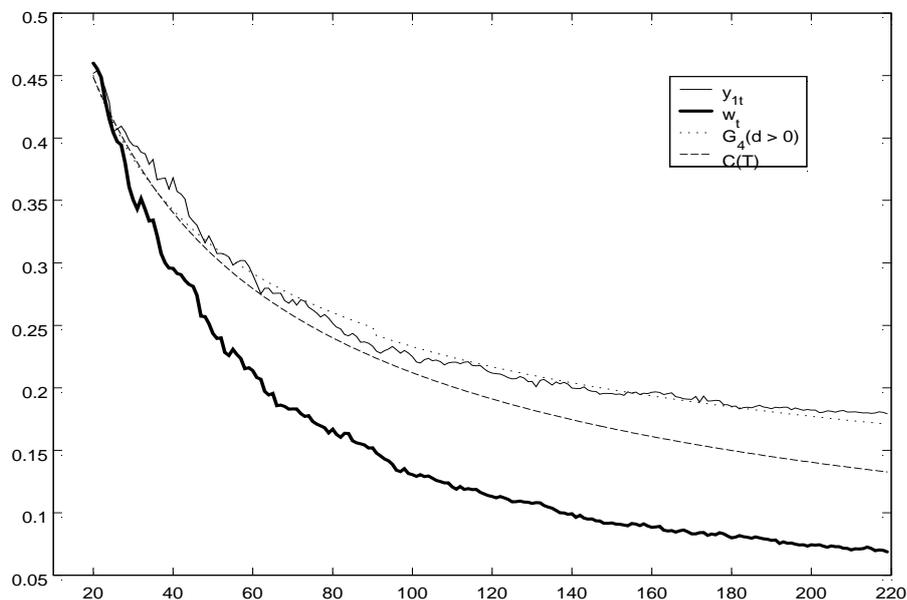


Figura 3.1.1: CCC simulados y funciones de penalización para un proceso univariante. y_{1t} corresponde al primer valor singular de un proceso univariante estacionario, w_t corresponde al primer valor singular de un proceso no estacionario, $G_4(d > 0)$ es una función de penalización propuesta para procesos univariantes y $C(T)$ es la función de penalización utilizada por Bauer y Wagner (2002b).

trabajo, se presenta una solución empírica a este problema. Para elegir una función de penalización, debería maximizarse tanto el tamaño como la potencia del criterio de decisión, es decir tanto la probabilidad de que el criterio acierte cuando se cumple la hipótesis nula, como la probabilidad de que el criterio acierte cuando no se cumple. Desafortunadamente, en este caso tamaño y potencia se mueven en sentido opuesto. Por ello, decidimos restringir las distribuciones empíricas a aquellas que cumplan alguna de las dos condiciones siguientes:

- *Condición 3.1.3:* maximizar el tamaño, con una potencia del criterio consistente, es decir que tiende a la unidad cuando aumenta la dimensión de la muestra.
- *Condición 3.1.4:* maximizar el tamaño, exigiendo una potencia mínima para un tamaño muestral determinado.

Cada una de ellas podrá ser utilizada por el usuario dependiendo del objetivo de su análisis y para valorar la proximidad de una serie a la no estacionariedad. Para el caso univariante, que se centra en decidir si $d > 0$, puede ser razonable exigir que la potencia mínima en la segunda condición sea igual al 50% para un

tamaño muestral $T = 50$. Para seleccionar las funciones de penalización y con ellas los criterios definitivos, se define el tamaño de un criterio en modelos univariantes como la probabilidad de estimar correctamente el número de raíces unitarias en un paseo aleatorio $\nabla z_t = a_t$. Asimismo, se define la potencia como la probabilidad de que $\hat{d} = d$ en un modelo estacionario autorregresivo del tipo $(1 - .9B)z_t = a_t$. Los resultados de los ejercicios de simulación de estos modelos con 1000 repeticiones se presentan en los Cuadros 3.1.2 y 3.1.3. Además de los criterios con las funciones de penalización propuestas ($G_l, l = 1, 2, \dots, 6$), se añaden los resultados obtenidos con el criterio de Bauer y Wagner (2002b), $C(T)$ y con el test aumentado de Dickey y Fuller (1981) al 95 %, calculado con un número de retardos (ρ) igual a 1.

*Cuadro 3.1.2: Tamaño de los criterios para $d > 0$
a partir del proceso $\phi(B)z_t = a_t$*

T	ADF ¹	C	G_1	G_2	G_3	G_4	G_5	G_6
50	0.950	0.953	0.960	0.879	0.888	0.953	0.970	0.956
100	0.951	0.984	0.984	0.935	0.958	0.989	0.995	0.984
500	0.939	0.996	0.997	0.996	0.997	1.0	1.0	0.997

¹ Test aumentado de Dickey y Fuller con $\rho = 1$; $\phi(B) = 1 - B$

*Cuadro 3.1.3: Potencia de los criterios para $d > 0$
a partir del proceso $\phi(B)z_t = a_t$*

T	ADF ¹	C	G_1	G_2	G_3	G_4	G_5	G_6
50	0.114	0.246	0.208	0.505	0.464	0.246	0.188	0.230
100	0.341	0.451	0.420	0.762	0.628	0.376	0.290	0.451
500	1.0	1.0	1.0	1.0	1.0	0.994	0.981	1.0

¹ $\rho = 1$; $\phi(B) = 1 - .9B$

Ambos Cuadros muestran la consistencia, tanto en tamaño como en potencia, de todos los criterios presentados. A la vista de los resultados, se elige como función de penalización que cumple la condición (3.1.3) a G_5 , que pasa a denominarse $G_a(d > 0)$. Además, se toma la función de penalización G_2 , de aquí en adelante $G_b(d > 0)$, que presenta una potencia del 50 % con un tamaño aceptable y consistente. El algoritmo de identificación (NID) desarrollado en este trabajo, comprende la detección de raíces unitarias con la familia de criterios expuestos para procesos univariantes y los que se proponen en el apartado siguiente para sistemas multivariantes.

Estacionalidad:

Es habitual que las series de interés económico con una frecuencia superior a la anual presenten un comportamiento estacional. Se han realizado diferentes modelizaciones de este tipo de series temporales como los modelos multiplicativos de Box et al. (1994) o los modelos de componentes no observables de Harvey y Shepard (1993). También se han derivado filtros estacionales que permiten trabajar con la parte no estacional de la serie como son los métodos X-11 (X-11, ARIMA X-11, ARIMA X-12) (ver, entre otros, Dagum, 1988) o el procedimiento SEATS/TRAMO de Gómez y Maravall (1996). No obstante, la atención recibida por los modelos estacionales en el caso univariante contrasta fuertemente con su pobre desarrollo en el caso multivariante. En este trabajo, tan sólo se trata la estacionalidad en el caso univariante.

Asimismo la estacionalidad, si es perceptible, es generalmente no estacionaria, con lo que el análisis de raíces unitarias se vuelve más complicado. Como comentamos anteriormente, el uso de los CCC no permite distinguir directamente más de una raíz unitaria en una misma serie, con lo que la utilización de los criterios propuestos para determinar el valor de d no genera buenos resultados con series estacionales. Ante este problema se presentan dos soluciones. La primera se reduce a utilizar alguno de los filtros estacionales nombrados anteriormente y analizar la serie corregida de estacionalidad con la metodología que se propone a lo largo de este Capítulo. La segunda, implementada en la batería de algoritmos que han resultado de este trabajo, consiste en utilizar un sencillo método para “filtrar” la estacionalidad más persistente de la serie y estudiar el orden de integración del resto. Para ello, se hace uso de la factorización de una diferencia estacional en una diferencia regular y una suma estacional, es decir $\nabla_s = \nabla S(B)$ donde $S(B) = 1 + B + B^2 + \dots + B^{s-1}$. El esquema siguiente muestra el procedimiento empleado con series estacionales.

1. Analizar gráficamente de la serie. Si la estacionalidad es perceptible, entonces...
2. ...filtrar la serie original por la suma estacional definida anteriormente,
$$S(B)z_t = w_t \tag{3.1.7}$$
3. y analizar el orden de integración de w_t tal y como se ha presentado en esta Sección.
4. a) Si se obtiene que w_t es $I(1)$, entonces decidimos que la transformación estacionaria para la serie original es $\nabla_s z_t$

- b) Si w_t es I(2), decidimos que la transformación estacionaria para la serie original es $\nabla\nabla_s z_t$

En el Capítulo 6 se ofrecen los resultados del análisis de varias series estacionales habituales en la literatura, reales y simuladas, con el método propuesto.

3.1.2. Procesos multivariantes

Si bien la metodología propuesta para la detección de raíces unitarias ofrece buenos resultados en modelos univariantes, en el tratamiento de modelos multivariantes es donde se encuentran sus principales ventajas. La posibilidad de aplicar el procedimiento anterior para estimar las funciones de penalización en sistemas con más de una raíz unitaria abre, como veremos más adelante, las puertas al análisis de cointegración.

Como se vio al inicio del Capítulo, las funciones de penalización estimadas no sólo dependen del tamaño muestral y de la dimensión del subespacio de información pasada, si no también del mínimo número de raíces unitarias que se desea contrastar. Esto implica que cada CCC candidato a ser igual a la unidad debe ser comparado con una función de penalización diferente. Hasta ahora, se ha presentado un criterio útil en procesos univariantes, que indica si la serie analizada es estacionaria o, si por el contrario $d > 0$. En esta subsección, se desarrollan los criterios para contrastar secuencialmente si $d > 1$, $d > 2$, $d > 3$ y $d > 4$, de forma que si los criterios aplicados al primer y segundo CCC, señalan que $d > 1$ y $d > 2$, pero rechazan $d > 3$ para el tercer CCC, entonces el algoritmo de detección devolverá la solución $\hat{d} = 2$. De nuevo, para estimar las funciones de penalización se simulan procesos bivariantes, trivariantes, tetravariantes y pentavariantes donde todos los VS, menos uno en cada caso, son iguales a la unidad. El CCC restante de cada sistema se genera cercano a la unidad aunque no unitario. Después, se estiman diferentes distribuciones empíricas para cada proceso, como en el apartado anterior, adaptando un modelo que se ajuste correctamente.

La Figura 3.1.2 ofrece la comparación de las distribuciones empíricas y_{1t} e y_{2t} correspondientes a los pares $(\psi = 90, \bar{\sigma} = 1.0)$ y $(\psi = 90, \bar{\sigma} = .95)$, de un sistema bivalente. Además, muestra la función de penalización ajustada a y_{1t} , $G_4(d > 1)$, y la función de penalización elegida que representaba la cota superior más baja de los dos criterios seleccionados en la sección anterior, $G_b(d > 0)$. La función de penalización con argumento $d > j - 1$ se usa para evaluar la existencia de al menos j raíces unitarias, con lo que $G_l(d > j - 1)$ es la función G_l que se aplica para decidir si el CCC j -ésimo se considera igual a la unidad. Al examinar el gráfico puede extraerse una conclusión relevante: el primer CCC, que corresponde a una

raíz unitaria del proceso bivalente, se mantiene muy por debajo de $G_b(d > 0)$, que se utiliza en el criterio con mayor potencia de los elegidos para contrastar $d > 0$. En efecto, si comparamos un proceso multivariante (\mathbf{A}_t) con d raíces unitarias con el mismo sistema que además incorpora una nueva serie no estacionaria (\mathbf{B}_t) (el proceso \mathbf{B}_t tendría al menos $d + 1$ raíces iguales a la unidad), entonces $\hat{\sigma}_j(\mathbf{B}_t) \geq \hat{\sigma}_j(\mathbf{A}_t)$ para $j = 1, \dots, d + 1$. De esta forma, aprovechando la ordenación de mayor a menor de los VS que devuelve la descomposición SVD, resulta válido utilizar los criterios expuestos de forma secuencial, para decidir el número de ellos iguales a la unidad.

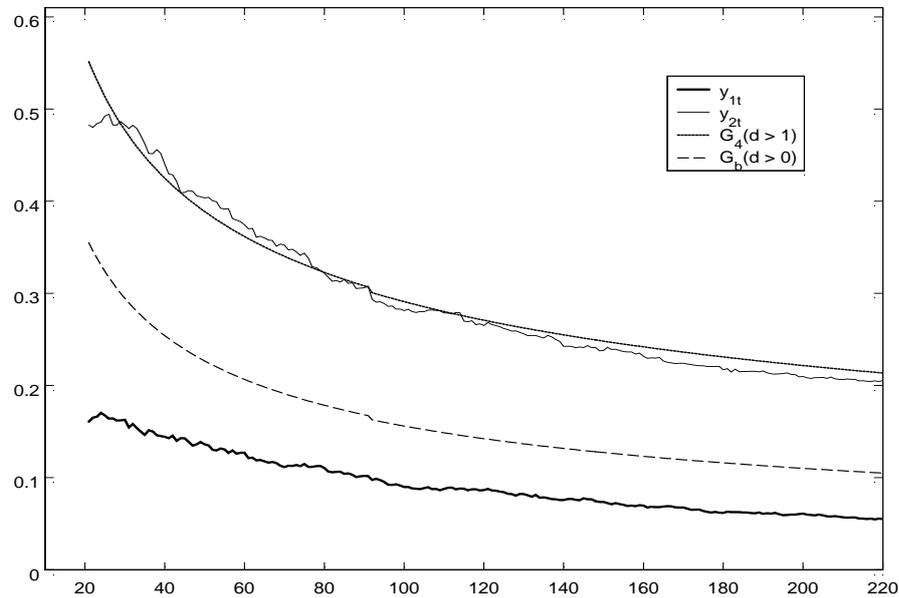


Figura 3.1.2: CCC simulados y funciones de penalización para un proceso bivalente. y_{1t} corresponde al primer valor singular de un sistema bivalente no estacionario (cuyos procesos univariantes son $I(1)$ e $I(0)$), y_{2t} corresponde al segundo valor singular del mismo sistema, $G_4(d > 1)$ es una de las funciones de penalización propuestas para detectar al menos dos raíces unitarias y $G_b(d > 0)$ es la función de penalización usada para detectar al menos una raíz unitaria.

El Cuadro 3.1.4 presenta los parámetros estimados del modelo (3.1.6) para cuatro posibles funciones de penalización propuestas para contrastar $d > 1$. Puede observarse en la Figura 3.1.2 que el modelo especificado para representar las distribuciones empíricas que se usan para contrastar $d > 0$, también se ajusta correctamente a las que se utilizan para decidir si $d > 1$, salvo quizá, en las muestras inferiores a 25 observaciones.

Cuadro 3.1.4: Funciones de penalización estimadas para $d > 1$

Distribución Empírica $f(\hat{\sigma}_j) = y_{jt}$	Valores Estimados					Función de Penalización
	$\bar{\sigma}$	ψ	α	β_1	β_2	
$1 - \hat{\sigma}_j^2$	0.95	75	0.43	-0.39	-0.07	G_1
$1 - \hat{\sigma}_j^2$	0.95	80	0.50	-0.39	-0.07	G_2
$1 - \hat{\sigma}_j^2$	0.95	85	0.57	-0.39	-0.07	G_3
$1 - \hat{\sigma}_j^2$	0.95	90	0.67	-0.39	-0.06	G_4

$j = 2$

Para discriminar entre las posibles funciones de penalización propuestas, se utilizan las condiciones (3.1.3) y (3.1.4), establecidos en la subsección anterior. Los resultados de las simulaciones realizadas para evaluar el tamaño y la potencia de los criterios formulados para $d > 1$ se exponen en los Cuadros 3.1.5 y 3.1.6. En ellos se refleja la probabilidad de que los diferentes criterios acierten el número de raíces unitarias del proceso generador, utilizando la función de penalización $G_b(d > 0)$ para decidir si $d > 0$.

Cuadro 3.1.5: Tamaño de los criterios para $d > 1$
a partir del proceso $\Phi(\mathbf{B})\mathbf{z}_t = \mathbf{a}_t$

T	G_1	G_2	G_3	G_4
50	0.708	0.779	0.837	0.901
100	0.854	0.891	0.924	0.953
500	0.997	0.998	0.999	0.999

$\Phi(\mathbf{B}) = \text{diag}(\nabla, \nabla)$

Cuadro 3.1.6: Potencia de los criterios para $d > 1$
a partir del proceso $\Phi(\mathbf{B})\mathbf{z}_t = \mathbf{a}_t$

T	G_1	G_2	G_3	G_4
50	0.452	0.373	0.293	0.204
100	0.516	0.428	0.337	0.239
500	0.993	0.989	0.968	0.947

$\Phi(\mathbf{B}) = \text{diag}(\nabla, (1 - .9B))$

Para el análisis de procesos bivariantes, se escogen las funciones de penalización G_4 y G_1 del Cuadro 3.1.4, que de aquí en adelante se nombran $G_a(d > 1)$ y $G_b(d > 1)$ respectivamente. La primera maximiza el tamaño dada una potencia

débil, pero consistente, y la segunda lo hace dada un potencia mínima del 45% para un tamaño muestral $T = 50$. En la decisión $d > 1$, se ha decidido disminuir la potencia mínima exigida con respecto de $d > 0$, al advertir que las propiedades de los criterios, tanto en tamaño como en potencia, se degradan ligeramente en muestras cortas cuando aumenta d . Los cuadros reflejan que todos los criterios estiman consistentemente.

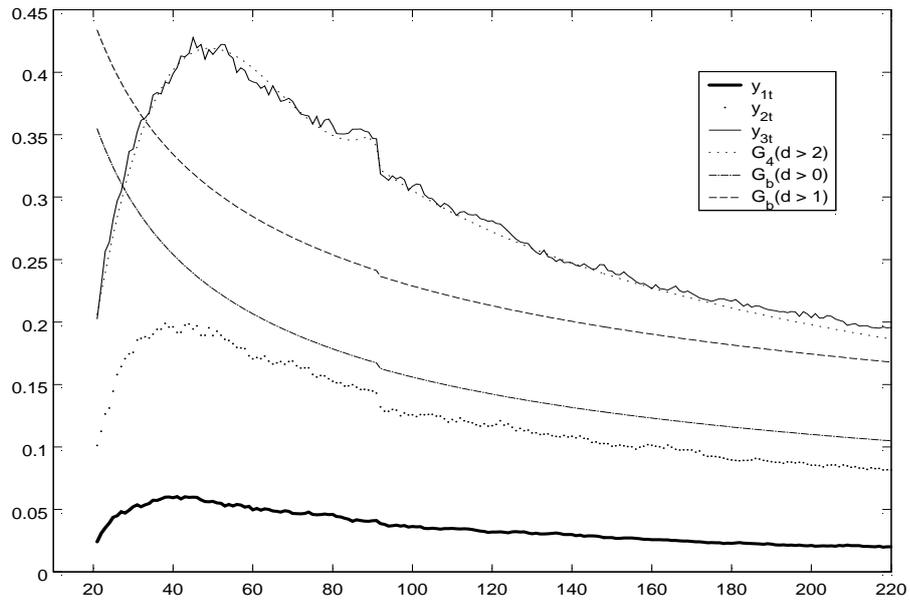


Figura 3.1.3: CCC simulados y funciones de penalización en un proceso trivariante. y_{1t} , y_{2t} e y_{3t} corresponden a los tres primeros valores singulares de un sistema no estacionario que contiene tres procesos univariantes, dos $I(1)$ y uno $I(0)$. $G_4(d > 2)$ es la función de penalización propuesta para detectar al menos tres raíces unitarias, $G_b(d > 0)$ y $G_b(d > 1)$ son las funciones de penalización usadas para detectar al menos una y dos raíces unitarias respectivamente.

Las Figuras 3.1.3, 3.1.4 y 3.1.5 presentan las distribuciones empíricas generadas con procesos trivariantes, tetravariantes y pentavariantes definidas por el par $(\psi = 90, \bar{\sigma} = .95)$. Estas representan la evolución, en función del tamaño muestral, de $f(\hat{\sigma}_j)$ para $j = 1, 2, 3$, para $j = 1, \dots, 4$ y para $j = 1, \dots, 5$, respectivamente. En los tres casos, tan sólo el último VS no corresponde a una raíz igual a la unidad. Estas distribuciones empíricas quedan representadas en cada figura por y_{jt} . Asimismo, los gráficos ofrecen las funciones de penalización estimadas que se ajustan adecuadamente al CCC considerado inferior a la unidad.

En la Figura 3.1.3, comparativamente, se añaden además las funciones de penalización usadas para contrastar $d > 0$ y $d > 1$. Como puede observarse, éstas son claramente superiores a y_{1t} e y_{2t} , respectivamente.

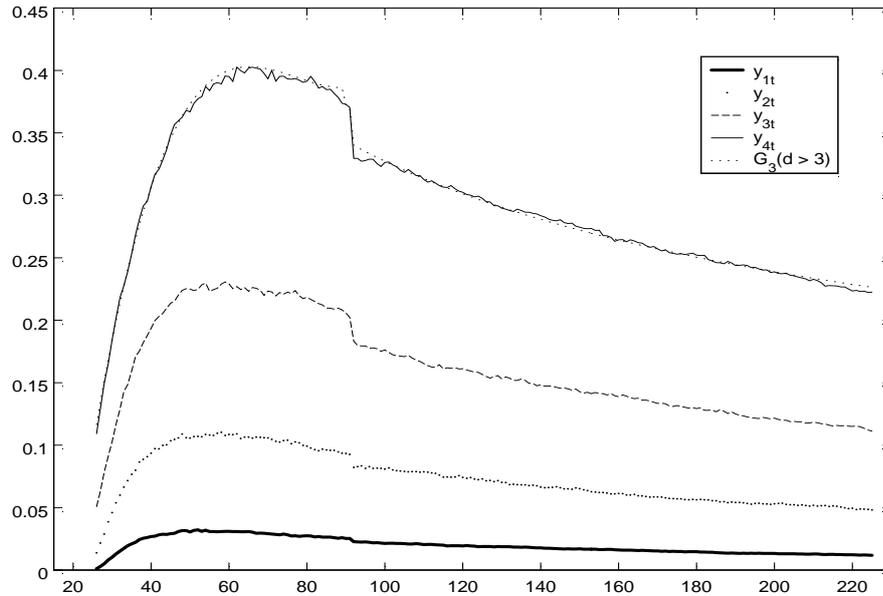


Figura 3.1.4: CCC simulados y funciones de penalización en un proceso tetravariante. y_{1t} , y_{2t} , y_{3t} e y_{4t} corresponden a los cuatro primeros valores singulares de un sistema no estacionario que contiene cuatro procesos univariantes, tres $I(1)$ y uno $I(0)$. $G_3(d > 3)$ es la función de penalización propuesta para detectar al menos cuatro raíces unitarias.

El método de obtención de las funciones de penalización para estos procesos es análogo al expuesto hasta ahora. Sin embargo, el modelo especificado para representarlas adecuadamente es diferente. De hecho, para obtener un mejor ajuste a las series y_{3t} , y_{4t} e y_{5t} , de las Figuras 3.1.3, 3.1.4 y 3.1.5, decidimos modelizar dos submuestras. Así, se propone una ecuación diferente para representar cada una de las dos submuestras escogidas. Los modelos ajustados son los siguientes:

$$\begin{aligned} t < t_j^* & \quad y_{jt} &= \alpha_1 + \delta_1 t + \delta_2 t^2 + \delta_3 t^3 + \delta_4 i_t + \varepsilon_{1t} \\ t \geq t_j^* & \quad \log(y_{jt}) &= \alpha_2 + \beta_1 \log(t) + \beta_2 \log(i_t) + \varepsilon_{2t} \end{aligned} \quad (3.1.8)$$

donde t_j^* se refiere a la primera realización de la segunda submuestra.

Los parámetros estimados de los modelos propuestos en (3.1.8) se ofrecen en el Cuadro 3.1.7. Allí se exponen los resultados de las estimaciones de las dos funciones de penalización que cumplen las condiciones 3.1.3 y 3.1.4, para contrastar la existencia de al menos $j - 1$ raíces unitarias, con $j = 3, 4, 5$. De nuevo, al igual que en el caso bivalente, se rebaja la potencia mínima exigida para el tamaño muestral $T = 50$, que hace referencia a la condición 3.1.4, al 40% para $d > 2$, $d > 3$ y $d > 4$. Obsérvese, en el Cuadro 3.1.7 y en las distintas Figuras que, como se comentó en el apartado 3.1.1, aumentar la dimensión del subespacio de información pasada i afecta negativamente a $f(\hat{\sigma}_j)$, para todo j . También, parece cumplirse que cuanto mayor es j mayor es la sensibilidad de y_{jt} ante cambios en i . Todas las figuras presentadas ilustran la influencia de los cambios en la dimensión del subespacio de información pasada sobre el cálculo de los VS.

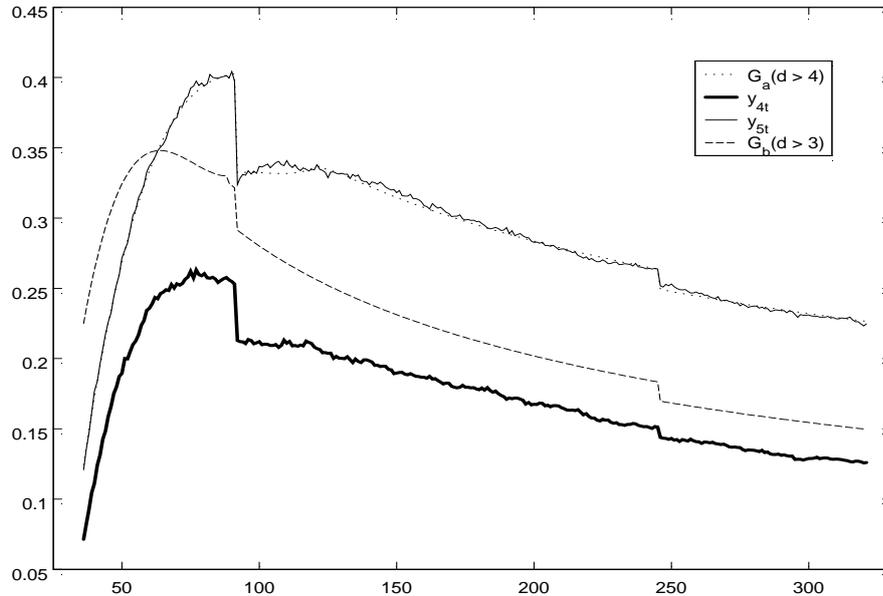


Figura 3.1.5: CCC simulados y funciones de penalización en un proceso pentavariante. y_{4t} e y_{5t} corresponden a los valores singulares cuarto y quinto de un sistema no estacionario cuyos procesos univariantes son $I(1)$ (cuatro) e $I(0)$ (uno). $G_a(d > 4)$ es la función de penalización propuesta para detectar al menos cinco raíces unitarias y $G_b(d > 3)$ es la función de penalización usada para detectar al menos cuatro raíces unitarias.

Los procesos generadores y las simulaciones realizadas para valorar el tamaño y la potencia de los criterios que usan las funciones de penalización del Cuadro 3.1.7 se presentan en los Cuadros 3.1.8 y 3.1.9. Todos los criterios estiman d consistentemente si bien, en general, los resultados revelan un deterioro en sus propiedades

Cuadro 3.1.7: Funciones de penalización propuestas para $d > 2$, $d > 3$ y $d > 4$

j	Distribución Empírica		$t < t^*$				$t \geq t^*$			Función de Penalización	
	$f(\hat{\sigma}_j) = y_{jt}$	$\bar{\sigma}$ ψ	α_1	δ_1	δ_2	δ_3	δ_4	α_0	β_1		β_2
3	$1 - \hat{\sigma}_{j}^2$	0.95 90	-0.305	0.040	-6.5×10^{-4}	3.3×10^{-6}	-	0.786	-0.328	-0.226	$G_a(d > 2)$
3	$1 - \hat{\sigma}_{j}^2$	0.95 60	-0.353	0.036	-5.9×10^{-4}	3.0×10^{-6}	-	0.188	-0.285	-0.172	$G_b(d > 2)$
4	$1 - \hat{\sigma}_{j}^2$	0.95 90	-0.635	0.044	-6.0×10^{-4}	2.7×10^{-6}	-	1.589	-0.437	-0.365	$G_a(d > 3)$
4	$1 - \hat{\sigma}_{j}^2$	0.95 60	-0.621	0.041	-5.6×10^{-4}	2.5×10^{-6}	-	1.557	-0.469	-0.418	$G_b(d > 3)$
5	$1 - \hat{\sigma}_{j}^2$	0.95 90	-0.317	0.032	-3.2×10^{-4}	1.1×10^{-6}	-0.076	1.313	-0.383	-0.280	$G_a(d > 4)$
5	$1 - \hat{\sigma}_{j}^2$	0.95 60	-0.366	0.030	-3.0×10^{-4}	1.0×10^{-6}	-0.063	1.131	-0.361	-0.378	$G_b(d > 4)$

Cuando $j = 3, 4$, entonces $t^* = 88$;

Cuando $j = 5$, entonces $t^* = 121$.

cuando aumenta el número de raíces unitarias que se desean contrastar. Por ejemplo, puede verse que tanto el tamaño como la potencia del criterio que usa la función de penalización $G_b(d > 3)$ son inferiores a los que corresponden al criterio que penaliza con $G_2(d > 0)$. En este punto se decide no estimar funciones de penalización para contrastar $d > 5$. Esto se debe a varios argumentos: 1) la utilización econométrica de modelos con más de cinco series temporales es realmente escasa, 2) el uso de estos criterios para procesos con una dimensión mayor pero con un número, a priori, no superior a cinco raíces unitarias es también válido, y 3) el deterioro que muestran los criterios con el aumento de d y la complejidad del análisis cuando aumenta la dimensión del sistema no motiva la búsqueda de funciones de penalización para procesos con un mayor número de raíces iguales a la unidad, aunque esto es posible.

Cuadro 3.1.8: Tamaño de los criterios para $d > 2$, $d > 3$ y $d > 4$ a partir de los procesos $\Phi_k(\mathbf{B})z_t = \mathbf{a}_t$

T	$d > 2$		$d > 3$		$d > 4$	
	G_a	G_b	G_a	G_b	G_a	G_b
50	0.897	0.601	0.903	0.603	0.882	0.600
100	0.919	0.669	0.921	0.644	0.934	0.626
500	0.998	0.996	0.997	0.926	0.990	0.966

$\Phi_1(\mathbf{B}) = \text{diag}(\nabla, \nabla, \nabla)$
 $\Phi_2(\mathbf{B}) = \text{diag}(\nabla, \nabla, \nabla, \nabla)$
 $\Phi_3(\mathbf{B}) = \text{diag}(\nabla, \nabla, \nabla, \nabla, \nabla)$

Cuadro 3.1.9: Potencia de los criterios para $d > 2$, $d > 3$ y $d > 4$ a partir del proceso $\Phi_k(\mathbf{B})z_t = \mathbf{a}_t$

T	$d > 2$		$d > 3$		$d > 4$	
	G_a	G_b	G_a	G_b	G_a	G_b
50	0.920	0.402	0.940	0.398	0.110	0.408
100	0.161	0.487	0.166	0.460	0.144	0.438
500	0.741	0.951	0.854	0.994	0.458	0.844

$\Phi_1(\mathbf{B}) = \text{diag}(\nabla, \nabla, (1 - .9B))$
 $\Phi_2(\mathbf{B}) = \text{diag}(\nabla, \nabla, \nabla, (1 - .9B))$
 $\Phi_3(\mathbf{B}) = \text{diag}(\nabla, \nabla, \nabla, \nabla, (1 - .9B))$

En resumen, a lo largo de esta sección se ha presentado una nueva familia de criterios de detección de raíces unitarias fundamentados en las propiedades asintóticas de los CCC. Todos ellos se basan en la estimación de sus funciones de

penalización. Así, se propone un primer criterio $C(G_a)$, que es el criterio definido en (3.1.5) con las funciones de penalización G_a derivadas en este Capítulo y que se aplica secuencialmente (con diferentes funciones de penalización para diferentes valores de d) maximizando el tamaño, con una potencia consistente. En segundo lugar, se presenta el criterio de la misma familia $C(G_b)$, que difiere del anterior en las funciones de penalización utilizadas (en este caso G_b) y que se aplica de forma equivalente pero que maximiza el tamaño, exigiendo una potencia mínima para un determinado tamaño muestral. En el siguiente apartado describiremos cómo aplicar estos métodos al estudio de relaciones de cointegración.

3.2. Cointegración

Una vez se ha determinado el número de raíces unitarias de un proceso multivariante, cabe preguntarse acerca de la existencia de relaciones de cointegración entre las series que forman el sistema. El concepto de cointegración (Box y Tiao, 1977; Granger y Engle, 1987) se ha desarrollado ampliamente, dando lugar a una extensa literatura y a potentes herramientas para su análisis, en series económicas y en series financieras. Los trabajos de Banerjee et al. (1993), Dhrymes (1998) o Maddala y In-Moo-Kim (1998) ofrecen una interesante revisión de esta literatura.

3.2.1. Rango de cointegración

En el marco de los modelos multivariantes, a partir del número de raíces unitarias del proceso y restringiendo el análisis a series $I(1)$ (o que pueden transformarse en $I(1)$) es posible obtener el número de relaciones de cointegración (c), como $m-d$, siendo m la dimensión del sistema y d el número de raíces unitarias del proceso. De esta forma, el análisis de relaciones de cointegración, a partir de la metodología presentada en la Sección 3.1 puede llevarse a cabo de la forma siguiente:

1. Analizar el orden de integración de los procesos univariantes con los criterios formulados para ello en la Sección 3.1.1, asegurándose de que éstos son $I(1)$.
2. Estimar el número de raíces unitarias d del proceso conjunto, a partir de la metodología presentada en la Sección 3.2.1.
3. Calcular \hat{c} , el estimador del rango de cointegración, como la diferencia entre la dimensión del sistema (m) y su número de raíces unitarias estimado (\hat{d}).

Es interesante, llegado este punto, analizar el uso de las dos funciones de penalización propuestas en la sección 3.1. Estas funciones se han obtenido a partir de las condiciones 3.1.3 y 3.1.4. Así, en el desarrollo de los pasos expuestos para estudiar el rango de cointegración, es coherente el uso de $C(G_a)$, que presenta un mayor tamaño que $C(G_b)$ en procesos univariantes, donde pueda considerarse “preferible” sobrediferenciar a infradiferenciar. Más aún, cuando el uso de la representación ARMA en esta metodología posibilita la utilización de un contraste de sobrediferenciación sobre los parámetros media móvil. En cambio, es conveniente la utilización de $C(G_b)$ con sistemas multivariantes, donde puede ser “preferible” sobreestimar a infraestimar el rango de cointegración. Esto es así porque en estos casos es posible emplear de nuevo el criterio de detección de raíces unitarias sobre las series supuestamente estacionarias, generadas a partir del producto de la matriz de cointegración y las series originales para detectar algún error en la etapa

anterior. De cualquier forma, ambos criterios están instrumentados en el algoritmo de identificación NID, para que el usuario los utilice en función de sus objetivos.

Para ilustrar las propiedades de los criterios propuestos, se presentan los resultados obtenidos mediante simulación. El mecanismo de generación de los procesos considerados son los tratados, para éste mismo propósito, por Riemers (1992), Toda (1995) y Poskitt (2000).

Como sugiere Toda (1995), se comienza considerando el caso más sencillo, es decir, un proceso bivalente. El Cuadro 3.2.1 presenta la frecuencia relativa observada para distintos rangos de cointegración estimados, \hat{c} . El modelo generador de los datos para este primer proceso es el siguiente:

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} .2 & .8 \\ .4 & .6 \end{pmatrix} \begin{pmatrix} x_{1t} \\ x_{2t} \end{pmatrix}, \quad (3.2.1)$$

donde

$$\begin{pmatrix} 1 & .0 \\ .0 & \nabla \end{pmatrix} \begin{pmatrix} x_{1t} \\ x_{2t} \end{pmatrix} = \begin{pmatrix} .0 \\ \delta \end{pmatrix} + \begin{pmatrix} \kappa(B) & .0 \\ .0 & 1 \end{pmatrix} \begin{pmatrix} \eta_{1t} \\ \eta_{2t} \end{pmatrix},$$

con $\kappa(B) = (1 - \alpha B)^{-1}$, y

$$\begin{pmatrix} \eta_{1t} \\ \eta_{2t} \end{pmatrix} = N \left(\begin{pmatrix} .0 \\ .0 \end{pmatrix}, \mathbf{R}_2 = \begin{pmatrix} 1.0 & \theta \\ \theta & 1.0 \end{pmatrix} \right)$$

Los resultados se han obtenido usando diferentes valores para α , δ y θ . Para cada combinación de parámetros y muestra, se han descartado los 50 primeros valores de $(y_{1t}, y_{2t})'$ evitando así problemas de inicialización. Los Cuadros presentados reflejan los resultados del uso secuencial de los criterios $C(G_a)$ y $C(G_b)$. Además, los resultados se comparan con los obtenidos por Poskitt (2000) usando el mismo proceso generador de los datos, también para 2000 repeticiones. Las columnas 5 a 8 muestran los resultados alcanzados con el criterio de Poskitt (2000), ∇_T , el procedimiento de Ratio de Verosimilitudes de Johansen (1988, 1991), LR_T , y criterios basados en una función de verosimilitud penalizada, PLR_T (para un examen más detallado sobre estos procedimientos ver Poskitt (2000)). Estos últimos necesitan una especificación del orden del modelo autorregresivo ajustado a los datos, a la que hacen referencia los superíndices SBC (Schwartz, 1978) y LP (Lütkepohl y Poskitt, 1998). Pese a no ser absolutamente comparables por no tratarse de la misma muestra, la inclusión de estos métodos alternativos puede ofrecer una visión relativa del comportamiento de los criterios propuestos y ayudar al análisis de los resultados obtenidos.

Los valores del Cuadro 3.2.1 muestran la consistencia de los criterios $C(G_a)$ y $C(G_b)$ para este caso. Además, no es necesario un tamaño muestral muy grande para que los resultados sean aceptablemente buenos. Los valores revelan que, para las dos funciones de penalización, los resultados se deterioran en muestras cortas cuando existe en el proceso un componente autorregresivo cercano a la no estacionariedad. Como es lógico, el criterio que utiliza las funciones de penalización G_a , presenta en muestras cortas peores resultados, puesto que por su propia definición (condiciones 3.1.3 y 3.1.4) tiene mayor propensión a atribuir la alta autocorrelación a una nueva tendencia estocástica y, por tanto, a infraestimar c . Este primer proceso generador de datos utilizado se basa en una estructura autorregresiva que es, probablemente, el mejor de los escenarios posibles.

Los resultados presentados en el Cuadro 3.2.2 están generados con un mecanismo similar al anterior, aunque aquí el vector $(x_{1t}, x_{2t})'$ viene definido por:

$$\begin{pmatrix} \nabla & .0 \\ .0 & \nabla \end{pmatrix} \begin{pmatrix} x_{1t} \\ x_{2t} \end{pmatrix} = \begin{pmatrix} .0 \\ \delta \end{pmatrix} + \begin{pmatrix} \kappa(B) & .0 \\ .0 & 1 \end{pmatrix} \begin{pmatrix} \eta_{1t} \\ \eta_{2t} \end{pmatrix}, \quad (3.2.2)$$

donde $\kappa(B) = 1 - \mu$. En este segundo caso, el criterio $C(G_a)$ bate a $C(G_b)$. De nuevo, los criterios se muestran consistentes y ofrecen resultados aceptables cuando μ se mantiene relativamente lejos de la unidad. Cuando $\mu = .9$ ambos criterios sobreestiman el valor de c . Esto se debe a que la raíz unitaria correspondiente al proceso x_{1t} se encuentra muy cerca de la cancelación con su estructura media móvil cercana a la no invertibilidad, $(1 - .9B)\eta_{1t}$.

Para más relaciones de cointegración, se estudian también los resultados obtenidos con el proceso generador trivariante siguiente:

$$\begin{pmatrix} y_{1t} \\ y_{2t} \\ y_{3t} \end{pmatrix} = \begin{pmatrix} .0 & .0 & 1.0 \\ .0 & .2 & .8 \\ .1 & .3 & .6 \end{pmatrix} \begin{pmatrix} x_{1t} \\ x_{2t} \\ x_{3t} \end{pmatrix}, \quad (3.2.3)$$

donde

$$\begin{pmatrix} d_1(B) & .0 & .0 \\ .0 & d_2(B) & .0 \\ .0 & .0 & \nabla \end{pmatrix} \begin{pmatrix} x_{1t} \\ x_{2t} \\ x_{3t} \end{pmatrix} = \begin{pmatrix} .0 \\ .0 \\ \delta \end{pmatrix} + \begin{pmatrix} \kappa_1(B) & .0 & .0 \\ .0 & \kappa_2(B) & .0 \\ .0 & .0 & 1.0 \end{pmatrix} \begin{pmatrix} \eta_{1t} \\ \eta_{2t} \\ \eta_{3t} \end{pmatrix}$$

con $(\eta_{1t}, \eta_{2t}, \eta_{3t})' = N(0, \mathbf{R}_3)$ y, $\mathbf{R}_3 = \begin{pmatrix} 1.0 & \theta & \theta \\ \theta & 1.0 & \theta \\ \theta & \theta & 1.0 \end{pmatrix}$

Los Cuadros 3.2.3, 3.2.4 y 3.2.5 presentan los resultados obtenidos por el proceso trivariante (3.2.3). En ellos, se reflejan diferentes escenarios para distintos

números de relaciones de cointegración: para $c = 0$, $d_1(B) = d_2(B) = \nabla$ y $\kappa_1(B) = \kappa_2(B) = 1$; para $c = 1$, $d_1(B) = 1$, $d_2(B) = \nabla$, $\kappa_1(B) = (1 - .\alpha B)^{-1}$ y $\kappa_2(B) = 1$; y por último, para $c = 2$, $d_1(B) = d_2(B) = 1$ y $\kappa_1(B) = \kappa_2(B) = (1 - .\alpha B)^{-1}$, con lo que los procesos trivariantes analizados contienen tres, dos y una raíz unitaria, respectivamente. El Cuadro 3.2.3 continúa mostrando la consistencia de ambos criterios junto con resultados favorables (superiores en $C(G_a)$ por la naturaleza de la función de penalización) para muestras no demasiado grandes, por ejemplo de 100 datos. Para los casos en que $c = 1, 2$ y $\alpha = .8$, $C(G_a)$ presenta indicios de infraestimación cuando $T = 100$, sin embargo estima razonablemente bien para muestras más grandes. Cuando $\alpha = .9$, aumenta el riesgo de infraestimar c con ambos criterios debido a la alta persistencia de los autorregresivos. En todos los casos del proceso trivariante, el criterio $C(G_b)$ presenta un comportamiento muy razonable, obteniendo una probabilidad mínima del 70% de identificaciones correctas cuando $T = 300$.

Para observar el comportamiento de los criterios con procesos diferentes y estudiar la influencia de la dimensión del sistema, se simula el proceso pentavariante:

$$\begin{pmatrix} y_{1t} \\ y_{2t} \\ y_{3t} \\ y_{4t} \\ y_{5t} \end{pmatrix} = \begin{pmatrix} .0 & .0 & .0 & .0 & 1.0 \\ .0 & .0 & .2 & .0 & .8 \\ .0 & .0 & .3 & .1 & .6 \\ .0 & .1 & .4 & .1 & .4 \\ .1 & .1 & .3 & .3 & .2 \end{pmatrix} \begin{pmatrix} x_{1t} \\ x_{2t} \\ x_{3t} \\ x_{4t} \\ x_{5t} \end{pmatrix} \quad (3.2.4)$$

donde el vector $\mathbf{X}_t = (x_{1t}, x_{2t}, \dots, x_{5t})'$ se genera a partir del modelo $\Phi(\mathbf{B})\mathbf{X}_t = \Delta + \Theta(\mathbf{B})\boldsymbol{\eta}_t$ siendo, $\Phi(\mathbf{B}) = \text{diag}(1, 1, 1, \nabla, \nabla)$ y $\Theta(\mathbf{B}) = \text{diag}((1 - \alpha B)^{-1}, (1 - \alpha B)^{-1}, (1 - \alpha B)^{-1}, 1, 1)$ para el proceso Pentavariante 1 y $\Phi(\mathbf{B}) = \text{diag}(1, 1, \nabla, \nabla, \nabla)$ y $\Theta(\mathbf{B}) = \text{diag}((1 - \alpha B)^{-1}, (1 - \alpha B)^{-1}, (1 - \mu B), 1, 1)$ para el proceso Pentavariante 2. Ambos procesos comparten $\Delta = (\delta, \delta, \dots, \delta)'$ y $\boldsymbol{\eta}_t = (\eta_{1t}, \eta_{2t}, \dots, \eta_{5t})'$ con $\boldsymbol{\eta}_t = N(0, \mathbf{R}_5)$, donde \mathbf{R}_5 denota una matriz con unos en la diagonal principal y θ en cualquier otra posición.

En el Cuadro 3.2.6 se ofrecen los primeros resultados del proceso Pentavariante 1. Aunque con $T = 100$ ambos criterios infraestiman el rango de cointegración (en general, todos los métodos reflejan dificultades para estimar correctamente c), tanto $C(G_a)$ como $C(G_b)$ presentan un buen comportamiento para $T = 300$, con lo que ambos criterios parecen no necesitar un gran número de observaciones para que el rango de cointegración estimado converja a su valor asintótico. El Cuadro 3.2.9 contiene los resultados para el mismo proceso pero con mayor persistencia en los procesos autorregresivos. Debido al deterioro manifestado por todos los métodos en estos casos con los procesos de menor dimensión y con el objetivo de seguir el análisis de Poskitt (2000), se incrementa el tamaño muestral en una secuencia de

Fibonacci hasta $T = 500, 800$ y 1300 . De nuevo, ambos criterios parecen llevar al verdadero rango de cointegración cuando la muestra crece. En tamaños muestrales de 500 y 800 observaciones, $C(G_b)$ domina claramente a $C(G_a)$.

Los resultados de las simulaciones a partir del proceso Pentavariante 2 se presentan en los Cuadros 3.2.8 y 3.2.9. Pese a la existencia de componentes autorregresivos con $\alpha = .8$ y lo cerca que se encuentra el modelo de la cancelación de una de las raíces unitarias con su parte media móvil, los dos criterios propuestos estiman razonablemente bien el rango de cointegración. Cuando aumenta la persistencia de los parámetros autorregresivos y el tamaño de la muestra, el Cuadro 3.2.9 revela el buen comportamiento de los dos criterios. Ambos presentan una probabilidad de seleccionar correctamente el rango de cointegración en torno al 90% para $T = 500$ (muy superior al resto de los métodos considerados).

En resumen, de este ejercicio de simulación se obtienen las siguientes conclusiones: 1) Los experimentos validan la capacidad de ambos criterios y por tanto, de la metodología propuesta para detectar y estimar el rango de cointegración en diferentes casos. 2) Ninguno de los dos criterios propuestos domina al otro (lo mismo ocurre con el resto de métodos presentados en los diferentes Cuadros) para todos los procesos y/o todas las muestras. Esto no es inesperado dadas las condiciones a partir de las cuales se han estimado las funciones de penalización de cada uno de ellos. 3) Los Cuadros que ofrecen los tamaños y potencias en las secciones 3.1.1 y 3.1.2 muestran que los resultados de los criterios se deterioran con el aumento del número de raíces unitarias de un proceso. En cambio, los resultados de las simulaciones de los Cuadros 3.2.1 a 3.2.9 revelan que el comportamiento de los criterios no se ve influido por la dimensión del sistema. Por ejemplo, los resultados no muestran un deterioro para un sistema pentavariante si $d = 2$, aunque probablemente lo muestren si $d = 5$. 4) Cuando las series presentan estructuras cercanas a la no estacionariedad o cuando se produce una cancelación aproximada de términos autorregresivos y medias móviles, parece claro que todos los métodos ofrecerán resultados relativamente pobres. 5) Por último, resulta relevante destacar que la metodología propuesta no precisa de la estimación ni de un modelo a priori, ni del orden del sistema, con lo que además de disminuir los posibles errores que se pueden producir en estos pasos intermedios, es computacionalmente eficiente.

Cuadro 3.2.1: Resultados para el proceso Bivariante 1^* .

T	\hat{c}	$C(G_a)$	$C(G_b)$	∇_T	LR_T	PLR_T^{SBC}	PLR_T^{LP}
$c = 1, \alpha = .7, \theta = .8, \delta = .0$							
100	0	0.0110	0.0075	0.0265	0.3250	0.0	0.1225
	1	0.9890	0.9925	0.9732	0.6280	1.0	0.8775
200	0	0.0	0.0	0.0005	0.0010	0.0	0.0050
	1	1.0	1.0	0.9995	0.9540	1.0	0.9950
300	0	0.0	0.0	0.0	0.0000	0.0	0.0
	1	1.0	1.0	1.0	0.9400	1.0	1.0
$c = 1, \alpha = .8, \theta = .8, \delta = .0$							
100	0	0.2320	0.0500	0.2275	0.6250	0.0020	0.4175
	1	0.7680	0.9500	0.7725	0.3360	0.9980	0.5825
200	0	0.0060	0.0015	0.0260	0.0720	0.0	0.1855
	1	0.9940	0.9985	0.9740	0.8840	1.0	0.8145
300	0	0.0	0.0005	0.0010	0.0010	0.0	0.0465
	1	1.0	0.9995	0.9990	0.9390	1.0	0.9535
$c = 1, \alpha = .9, \theta = .8, \delta = .0$							
100	0	0.7765	0.5050	0.6920	0.8150	0.0370	0.7395
	1	0.2235	0.4950	0.3080	0.1335	0.9630	0.2605
200	0	0.5645	0.2410	0.6250	0.6035	0.0020	0.8300
	1	0.4355	0.7590	0.3745	0.3630	0.9980	0.1700
300	0	0.3555	0.0800	0.4850	0.2390	0.0	0.7975
	1	0.6445	0.9195	0.5150	0.7070	1.0	0.2025

*El rango de cointegración estimado por cada procedimiento queda denotado por \hat{c} , $C(G_a)$ y $C(G_b)$ representan al criterio (3.1.5) con las funciones de penalización estimadas G_a y G_b , respectivamente. El resto de columnas muestran los criterios de Poskitt (2000) (∇_T), el Ratio de Verosimilitudes de Johansen (1988, 1991), (LR_T), y criterios basados en una función de verosimilitud penalizada (PLR_T) que contienen los superíndices SBC (Schwartz, 1978) o LP (Lütkepohl y Poskitt, 1998) en función del método usado para especificar el orden del modelo ajustado a los datos. Los resultados de las columnas 5 a 8 se toman de Poskitt (2000) y aunque fueron obtenidos con el mismo número de repeticiones (2000), tamaños muestrales ($T = 100, 200, 300$) y proceso generador de los datos, no se corresponden exactamente con la misma muestra.

Cuadro 3.2.2: Resultados para el proceso Bivariante 2^* .

T	\hat{c}	$C(G_a)$	$C(G_b)$	∇_T	LR_T	PLR_T^{SBC}	PLR_T^{LP}
$c = 0, \mu = .3, \theta = .8, \delta = .0$							
100	0	0.7880	0.6240	0.6620	0.9510	0.1510	0.8830
	1	0.2120	0.3760	0.3380	0.0465	0.8490	0.1170
200	0	0.8815	0.7765	0.8435	0.9445	0.2140	0.9845
	1	0.1185	0.2235	0.1565	0.0520	0.7860	0.0155
300	0	0.9415	0.8690	0.9260	0.9455	0.2565	0.9980
	1	0.0585	0.1310	0.0740	0.0485	0.7435	0.0020
$c = 0, \mu = .5, \theta = .8, \delta = .0$							
100	0	0.5795	0.4200	0.3185	0.9545	0.1480	0.8890
	1	0.4205	0.5800	0.6815	0.0405	0.8520	0.1110
200	0	0.7005	0.5545	0.4760	0.9465	0.2095	0.9845
	1	0.2995	0.4455	0.5240	0.0510	0.7905	0.0155
300	0	0.7870	0.6750	0.5925	0.9440	0.2600	0.9980
	1	0.2130	0.3250	0.4075	0.0505	0.7400	0.0020
$c = 0, \mu = .9, \theta = .8, \delta = .0$							
100	0	0.0230	0.0095	0.0	0.7015	0.0170	0.4840
	1	0.9770	0.9875	1.0	0.2780	0.9830	0.5160
200	0	0.0285	0.0130	0.0	0.4075	0.0095	0.5540
	1	0.9715	0.9870	1.0	0.5675	0.9905	0.4460
300	0	0.0535	0.0315	0.0	0.4305	0.0255	0.7650
	1	0.9465	0.9680	1.0	0.5415	0.9745	0.2350

*El rango de cointegración estimado por cada procedimiento queda denotado por \hat{c} , $C(G_a)$ y $C(G_b)$ representan al criterio (3.1.5) con las funciones de penalización estimadas G_a y G_b , respectivamente. El resto de columnas muestran los criterios de Poskitt (2000) (∇_T), el Ratio de Verosimilitudes de Johansen (1988, 1991), (LR_T), y criterios basados en una función de verosimilitud penalizada (PLR_T) que contienen los superíndices SBC (Schwartz, 1978) o LP (Lütkepohl y Poskitt, 1998) en función del método usado para especificar el orden del modelo ajustado a los datos. Los resultados de las columnas 5 a 8 se toman de Poskitt (2000) y aunque fueron obtenidos con el mismo número de repeticiones (2000), tamaños muestrales ($T = 100, 200, 300$) y proceso generador de los datos, no se corresponden exactamente con la misma muestra.

Cuadro 3.2.3: Resultados para el proceso Trivariante 1*.

T	\hat{c}	$C(G_a)$	$C(G_b)$	∇_T	LR_T	PLR_T^{SBC}	PLR_T^{LP}
$c = 0, \theta = .0, \delta = .0$							
100	0	0.9255	0.6470	0.7560	0.9395	0.4085	0.4230
	1	0.0740	0.3480	0.2385	0.0570	0.0890	0.5585
	2	0.0005	0.0050	0.0055	0.0030	0.5026	0.0185
200	0	0.9500	0.7735	0.9385	0.9410	0.6125	0.8555
	1	0.0500	0.2260	0.0615	0.0530	0.0740	0.1455
	2	0.0	0.0005	0.0	0.0060	0.3135	0.0
300	0	0.9595	0.8270	0.9860	0.9505	0.7300	0.9690
	1	0.0405	0.1725	0.0140	0.0465	0.0695	0.0310
	2	0.0	0.0005	0.0	0.0030	0.2005	0.0

*El rango de cointegración estimado por cada procedimiento queda denotado por \hat{c} , $C(G_a)$ y $C(G_b)$ representan al criterio (3.1.5) con las funciones de penalización estimadas G_a y G_b , respectivamente. El resto de columnas muestran los criterios de Poskitt (2000) (∇_T), el Ratio de Verosimilitudes de Johansen (1988, 1991), (LR_T), y criterios basados en una función de verosimilitud penalizada (PLR_T) que contienen los superíndices SBC (Schwartz, 1978) o LP (Lütkepohl y Poskitt, 1998) en función del método usado para especificar el orden del modelo ajustado a los datos. Los resultados de las columnas 5 a 8 se toman de Poskitt (2000) y aunque fueron obtenidos con el mismo número de repeticiones (2000), tamaños muestrales ($T = 100, 200, 300$) y proceso generador de los datos, no se corresponden exactamente con la misma muestra.

Cuadro 3.2.4: Resultados para el proceso Trivariante 2*.

T	\hat{c}	$C(G_a)$	$C(G_b)$	∇_T	LR_T	PLR_T^{SBC}	PLR_T^{LP}
$c = 1, \alpha = .8, \theta = .0, \delta = .0$							
100	0	0.4620	0.0615	0.2015	0.7760	0.1280	0.1515
	1	0.5290	0.8750	0.7485	0.1985	0.0990	0.7850
	2	0.0090	0.0630	0.0500	0.0215	0.7730	0.0635
200	0	0.0100	0.0	0.0340	0.3090	0.0140	0.0985
	1	0.9835	0.9645	0.9530	0.6455	0.2155	0.8935
	2	0.0065	0.0355	0.0130	0.0410	0.7705	0.0080
300	0	0.0	0.0	0.0030	0.0355	0.0	0.0200
	1	0.9975	0.9850	0.9920	0.9115	0.2785	0.9780
	2	0.0025	0.0150	0.0050	0.0500	0.7215	0.0020
$c = 1, \alpha = .9, \theta = .0, \delta = .0$							
100	0	0.8510	0.4180	0.5880	0.8925	0.3070	0.3310
	1	0.1470	0.5540	0.3970	0.0990	0.0740	0.6445
	2	0.0020	0.0280	0.0150	0.0070	0.6190	0.0245
200	0	0.5370	0.1450	0.6275	0.7810	0.2520	0.6090
	1	0.4605	0.8360	0.3705	0.1950	0.1200	0.3895
	2	0.0025	0.0190	0.0020	0.0210	0.6280	0.0015
300	0	0.2190	0.0175	0.5380	0.5525	0.1190	0.6300
	1	0.7810	0.9735	0.4610	0.4110	0.2245	0.3695
	2	0.0	0.0090	0.0010	0.0340	0.6565	0.0005

*El rango de cointegración estimado por cada procedimiento queda denotado por \hat{c} , $C(G_a)$ y $C(G_b)$ representan al criterio (3.1.5) con las funciones de penalización estimadas G_a y G_b , respectivamente. El resto de columnas muestran los criterios de Poskitt (2000) (∇_T), el Ratio de Verosimilitudes de Johansen (1988, 1991), (LR_T), y criterios basados en una función de verosimilitud penalizada (PLR_T) que contienen los superíndices SBC (Schwartz, 1978) o LP (Lütkepohl y Poskitt, 1998) en función del método usado para especificar el orden del modelo ajustado a los datos. Los resultados de las columnas 5 a 8 se toman de Poskitt (2000) y aunque fueron obtenidos con el mismo número de repeticiones (2000), tamaños muestrales ($T = 100, 200, 300$) y proceso generador de los datos, no se corresponden exactamente con la misma muestra.

Cuadro 3.2.5: Resultados para el proceso Trivariante \mathcal{I}^* .

T	\hat{c}	$C(G_a)$	$C(G_b)$	∇_T	LR_T	PLR_T^{SBC}	PLR_T^{LP}
$c = 2, \alpha = .8, \theta = .0, \delta = .0$							
100	0	0.2190	0.0070	0.1690	0.4520	0.0195	0.0485
	1	0.4855	0.2680	0.3485	0.4060	0.0020	0.6515
	2	0.2955	0.7250	0.4825	0.1175	0.9785	0.3000
200	0	0.0	0.0	0.0340	0.0035	0.0	0.0110
	1	0.0395	0.0100	0.0710	0.2085	0.0	0.4535
	2	0.9605	0.9900	0.8950	0.7345	1.0	0.5355
300	0	0.0	0.0	0.0020	0.0	0.0	0.0000
	1	0.0	0.0005	0.0080	0.0020	0.0	0.1510
	2	1.0	0.9995	0.9900	0.9480	1.0	0.8490
$c = 2, \alpha = .9, \theta = .0, \delta = .0$							
100	0	0.7500	0.2415	0.5255	0.8245	0.2100	0.2660
	1	0.2370	0.6545	0.4175	0.1525	0.0340	0.6795
	2	0.0130	0.1020	0.0570	0.0185	0.7560	0.0545
200	0	0.2495	0.0200	0.6405	0.4205	0.0525	0.4355
	1	0.6735	0.5770	0.2990	0.4175	0.0040	0.5465
	2	0.0770	0.3955	0.0650	0.1290	0.9435	0.1800
300	0	0.0415	0.0010	0.6310	0.0535	0.0015	0.4000
	1	0.7125	0.2830	0.2550	0.4565	0.0005	0.5740
	2	0.2460	0.7075	0.1140	0.4445	0.9980	0.0260

*El rango de cointegración estimado por cada procedimiento queda denotado por \hat{c} , $C(G_a)$ y $C(G_b)$ representan al criterio (3.1.5) con las funciones de penalización estimadas G_a y G_b , respectivamente. El resto de columnas muestran los criterios de Poskitt (2000) (∇_T), el Ratio de Verosimilitudes de Johansen (1988, 1991), (LR_T), y criterios basados en una función de verosimilitud penalizada (PLR_T) que contienen los superíndices SBC (Schwartz, 1978) o LP (Lütkepohl y Poskitt, 1998) en función del método usado para especificar el orden del modelo ajustado a los datos. Los resultados de las columnas 5 a 8 se toman de Poskitt (2000) y aunque fueron obtenidos con el mismo número de repeticiones (2000), tamaños muestrales ($T = 100, 200, 300$) y proceso generador de los datos, no se corresponden exactamente con la misma muestra.

Cuadro 3.2.6: Resultados para el proceso Pentavariante 1*.

T	\hat{c}	$C(G_a)$	$C(G_b)$	∇_T	LR_T	PLR_T^{LP}
$c = 3, \alpha = .8, \theta = .8, \delta = 1.0$						
100	0	0.0875	0.1040	0.0370	0.0850	0.0
	1	0.7750	0.4665	0.2685	0.5695	0.0830
	2	0.1370	0.3485	0.4090	0.2935	0.6705
	3	0.0005	0.0800	0.2835	0.0450	0.2435
	4	0.0	0.0005	0.0020	0.0060	0.0030
200	0	0.0005	0.0	0.0110	0.0	0.0
	1	0.0115	0.0	0.0585	0.0555	0.1035
	2	0.7060	0.1130	0.1365	0.5405	0.7075
	3	0.2820	0.8830	0.7925	0.4060	0.1890
	4	0.0	0.0040	0.0015	0.0320	0.0
300	0	0.0	0.0	0.0005	0.0	0.0
	1	0.0	0.0	0.0025	0.0	0.0925
	2	0.1390	0.0010	0.0095	0.0915	0.7310
	3	0.8610	0.9970	0.9875	0.8555	0.1765
	4	0.0	0.0020	0.0	0.0490	0.0

*El rango de cointegración estimado por cada procedimiento queda denotado por \hat{c} , $C(G_a)$ y $C(G_b)$ representan al criterio (3.1.5) con las funciones de penalización estimadas G_a y G_b , respectivamente. El resto de columnas muestran los criterios de Poskitt (2000) (∇_T), el Ratio de Verosimilitudes de Johansen (1988, 1991), (LR_T), y criterios basados en una función de verosimilitud penalizada (PLR_T) que contienen los superíndices SBC (Schwartz, 1978) o LP (Lütkepohl y Poskitt, 1998) en función del método usado para especificar el orden del modelo ajustado a los datos. Los resultados de las columnas 5 a 8 se toman de Poskitt (2000) y aunque fueron obtenidos con el mismo número de repeticiones (2000), tamaños muestrales ($T = 100, 200, 300$) y proceso generador de los datos, no se corresponden exactamente con la misma muestra.

Cuadro 3.2.7: Resultados para el proceso Pentavariante 1*.

T	\hat{c}	$C(G_a)$	$C(G_b)$	∇_T	LR_T	PLR_T^{LP}
$c = 3, \alpha = .9, \theta = .8, \delta = 1.0$						
500	0	0.0	0.0	0.4685	0.0	0.0
	1	0.0145	0.0	0.2880	0.0015	0.4715
	2	0.9530	0.5440	0.1180	0.2715	0.5120
	3	0.0325	0.4555	0.1255	0.6800	0.0165
	4	0.0	0.0005	0.0	0.0490	0.0
800	0	0.0	0.0	0.1625	0.0	0.0
	1	0.0	0.0	0.0825	0.0	0.3540
	2	0.5865	0.0650	0.0510	0.0005	0.6085
	3	0.4135	0.9350	0.7040	0.9535	0.0375
	4	0.0	0.0	0.0	0.0415	0.0
1300	0	0.0	0.0	0.0005	0.0	0.0
	1	0.0	0.0	0.0005	0.0	0.0645
	2	0.0290	0.0	0.0005	0.0	0.6665
	3	0.9710	1.0	0.9985	0.9475	0.2690
	4	0.0	0.0	0.0	0.0490	0.0

*El rango de cointegración estimado por cada procedimiento queda denotado por \hat{c} , $C(G_a)$ y $C(G_b)$ representan al criterio (3.1.5) con las funciones de penalización estimadas G_a y G_b , respectivamente. El resto de columnas muestran los criterios de Poskitt (2000) (∇_T), el Ratio de Verosimilitudes de Johansen (1988, 1991), (LR_T), y criterios basados en una función de verosimilitud penalizada (PLR_T) que contienen los superíndices SBC (Schwartz, 1978) o LP (Lütkepohl y Poskitt, 1998) en función del método usado para especificar el orden del modelo ajustado a los datos. Los resultados de las columnas 5 a 8 se toman de Poskitt (2000) y aunque fueron obtenidos con el mismo número de repeticiones (2000), tamaños muestrales ($T = 500, 800, 1300$) y proceso generador de los datos, no se corresponden exactamente con la misma muestra.

Cuadro 3.2.8: Resultados para el proceso Pentavariante \mathcal{Z}^* .

T	\hat{c}	$C(G_a)$	$C(G_b)$	∇_T	LR_T	PLR_T^{LP}
$c = 2, \alpha = .8, \mu = .9, \theta = .8, \delta = 1.0$						
100	0	0.0120	0.0160	0.0	0.0005	0.0005
	1	0.7735	0.5045	0.2990	0.1920	0.0695
	2	0.2145	0.4715	0.7335	0.6250	0.6550
	3	0.0	0.0080	0.0375	0.1625	0.2730
	4	0.0	0.0	0.0	0.0175	0.0020
200	0	0.0	0.0	0.0	0.0	0.0
	1	0.0635	0.0035	0.0750	0.0	0.0845
	2	0.9365	0.8805	0.9025	0.3190	0.6865
	3	0.0	0.1160	0.0225	0.6390	0.2280
	4	0.0	0.0	0.0	0.0395	0.0010
300	0	0.0	0.0	0.0	0.0	0.0050
	1	0.0170	0.0090	0.0085	0.0	0.1425
	2	0.9795	0.9290	0.9145	0.0315	0.7155
	3	0.0035	0.0615	0.0770	0.9250	0.1415
	4	0.0	0.0005	0.0	0.0395	0.0

*El rango de cointegración estimado por cada procedimiento queda denotado por \hat{c} , $C(G_a)$ y $C(G_b)$ representan al criterio (3.1.5) con las funciones de penalización estimadas G_a y G_b , respectivamente. El resto de columnas muestran los criterios de Poskitt (2000) (∇_T), el Ratio de Verosimilitudes de Johansen (1988, 1991), (LR_T), y criterios basados en una función de verosimilitud penalizada (PLR_T) que contienen los superíndices SBC (Schwartz, 1978) o LP (Lütkepohl y Poskitt, 1998) en función del método usado para especificar el orden del modelo ajustado a los datos. Los resultados de las columnas 5 a 8 se toman de Poskitt (2000) y aunque fueron obtenidos con el mismo número de repeticiones (2000), tamaños muestrales ($T = 100, 200, 300$) y proceso generador de los datos, no se corresponden exactamente con la misma muestra.

Cuadro 3.2.9: Resultados para el proceso Pentavariante \mathcal{P}^* .

T	\hat{c}	$C(G_a)$	$C(G_b)$	∇_T	LR_T	PLR_T^{LP}
$c = 2, \alpha = .9, \mu = .9, \theta = .8, \delta = 1.0$						
500	0	0.0060	0.0	0.0	0.0	0.0
	1	0.0985	0.0275	0.4225	0.0	0.4025
	2	0.8955	0.9725	0.5775	0.0965	0.5615
	3	0.0	0.0	0.0	0.8525	0.0360
	4	0.0	0.0	0.0	0.0480	0.0
800	0	0.0	0.0	0.0	0.0	0.0
	1	0.0590	0.0255	0.1580	0.0	0.5130
	2	0.9410	0.9740	0.8390	0.0	0.4685
	3	0.0	0.0005	0.0030	0.9605	0.0185
	4	0.0	0.0	0.0	0.0370	0.0
1300	0	0.0	0.0	0.0	0.0	0.0
	1	0.0595	0.0055	0.0055	0.0	0.2710
	2	0.9405	0.9945	0.9210	0.0	0.7080
	3	0.0	0.0	0.0735	0.9440	0.0210
	4	0.0	0.0	0.0	0.0515	0.0

*El rango de cointegración estimado por cada procedimiento queda denotado por \hat{c} , $C(G_a)$ y $C(G_b)$ representan al criterio (3.1.5) con las funciones de penalización estimadas G_a y G_b , respectivamente. El resto de columnas muestran los criterios de Poskitt (2000) (∇_T), el Ratio de Verosimilitudes de Johansen (1988, 1991), (LR_T), y criterios basados en una función de verosimilitud penalizada (PLR_T) que contienen los superíndices SBC (Schwartz, 1978) o LP (Lütkepohl y Poskitt, 1998) en función del método usado para especificar el orden del modelo ajustado a los datos. Los resultados de las columnas 5 a 8 se toman de Poskitt (2000) y aunque fueron obtenidos con el mismo número de repeticiones (2000), tamaños muestrales ($T = 500, 800, 1300$) y proceso generador de los datos, no se corresponden exactamente con la misma muestra.

3.2.2. Estimación de la matriz de cointegración

Como se ha mostrado en el apartado anterior, el método presentado permite obtener de forma rápida, una estimación consistente del número de relaciones de cointegración existentes en un sistema de variables. Para ello, es necesario imponer la condición de que los procesos univariantes que lo componen (o una transformación de los mismos) sean integrados de orden 1. Alcanzado este punto, resulta interesante preguntarse cuáles son las combinaciones lineales del sistema que generan series estacionarias. Para obtener una estimación de éstas es necesario, en un primer paso estimar, el modelo en EE, para lo que se precisa una estimación del orden del sistema (en el Capítulo 4 se prestará especial atención a este problema). A partir de las matrices obtenidas del modelo en forma de innovaciones y del rango de cointegración estimado con el método planteado anteriormente, se propone un estimador consistente de la matriz de cointegración.

Como se vio en el Capítulo 2, estimar el modelo en forma de innovaciones mediante métodos de subespacios es equivalente a resolver el problema dado en (2.4.4) identificando las matrices $\hat{\Phi}$, \hat{H} , $\hat{\Gamma}$, \hat{D} y en un segundo paso \hat{E} y \hat{Q} (Casals, 1997; Favoreel et al., 2000; Knudsen, 2001).

Una vez estimado el modelo, Bauer y Wagner (2002b) demuestran que cualquier proceso formado por variables endógenas I(1), puede escribirse como un modelo en forma de innovaciones que separa las secuencias de estados estacionarios de las de estados no estacionarios. Esto se puede realizar mediante un cambio de coordenadas en el vector de estado, que genera un modelo output-equivalente a (2.1.2) de la forma:

$$\mathbf{X}_{t+1} = \begin{pmatrix} \mathbf{X}_{1,t+1} \\ \mathbf{X}_{2,t+1} \end{pmatrix} = \begin{pmatrix} \mathbf{I}_d & \mathbf{0} \\ \mathbf{0} & \phi_{n-d} \end{pmatrix} \mathbf{X}_t + \begin{pmatrix} \mathbf{E}_1 \\ \mathbf{E}_2 \end{pmatrix} \epsilon_t \quad (3.2.5)$$

$$\mathbf{Z}_t = \mathbf{H} \mathbf{X}_t + \epsilon_t \quad (3.2.6)$$

donde, (3.2.5) es la ecuación de estado, n es el orden del sistema, $\mathbf{X}_{1,t+1}$ es el vector de estados del subsistema no estacionario con dimensión d (igual al número de raíces unitarias), $\mathbf{X}_{2,t+1}$ es el vector de estados del subsistema estacionario con dimensión $n - d$ y (3.2.6) es la ecuación de observación.

Como el vector de variables endógenas \mathbf{Z}_t está formado por series I(1), el número de relaciones de cointegración puede calcularse como $\hat{c} = m - \hat{d}$, aplicando al sistema uno de los criterios de detección de raíces unitarias de la Sección 3.1. A continuación, para conseguir una estimación de el (los) vector(es) de cointegración, es necesario encontrar en la nueva ecuación de observación (3.2.6), una matriz que anule la parte de \mathbf{H} que afecta a la secuencia de estados no estacionarios para las

series cointegradas. Esto es relativamente sencillo, puesto que el observador (3.2.6) puede escribirse como:

$$\begin{pmatrix} \mathbf{Z}_{1,t} \\ \mathbf{Z}_{2,t} \end{pmatrix} = \begin{pmatrix} \mathbf{H}_{11} & \mathbf{H}_{12} \\ \mathbf{H}_{21} & \mathbf{H}_{22} \end{pmatrix} \begin{pmatrix} \mathbf{X}_{1,t} \\ \mathbf{X}_{2,t} \end{pmatrix} + \begin{pmatrix} \boldsymbol{\epsilon}_{1,t} \\ \boldsymbol{\epsilon}_{2,t} \end{pmatrix} \quad (3.2.7)$$

donde $\mathbf{Z}_{1,t} \in \mathbb{R}^c$, $\mathbf{Z}_{2,t} \in \mathbb{R}^d$, $\mathbf{H}_{11} \in \mathbb{R}^{c \times d}$, $\mathbf{H}_{12} \in \mathbb{R}^{c \times (n-d)}$, $\mathbf{H}_{21} \in \mathbb{R}^{d \times d}$ y $\mathbf{H}_{22} \in \mathbb{R}^{d \times (n-d)}$.

Proposición 3.2.1 *La matriz $\boldsymbol{\Lambda} = (\mathbf{I}_c \quad -\mathbf{H}_{11}\mathbf{H}_{21}^{-1})$ es un estimador consistente de la matriz de cointegración.*

Demostración. Si premultiplicamos la ecuación de observación (3.2.7) por la matriz propuesta tenemos que,

$$\boldsymbol{\Lambda} \begin{pmatrix} \mathbf{Z}_{1,t} \\ \mathbf{Z}_{2,t} \end{pmatrix} = \boldsymbol{\Lambda} \begin{pmatrix} \mathbf{H}_{11} & \mathbf{H}_{12} \\ \mathbf{H}_{21} & \mathbf{H}_{22} \end{pmatrix} \begin{pmatrix} \mathbf{X}_{1,t} \\ \mathbf{X}_{2,t} \end{pmatrix} + \boldsymbol{\Lambda} \begin{pmatrix} \boldsymbol{\epsilon}_{1,t} \\ \boldsymbol{\epsilon}_{2,t} \end{pmatrix} \quad (3.2.8)$$

operando y tomando $\mathbf{H}_{11}\mathbf{H}_{21}^{-1} = \boldsymbol{\lambda}$, cuya dimensión es $c \times d$, se obtiene:

$$\mathbf{Z}_{1,t} - \boldsymbol{\lambda}\mathbf{Z}_{2,t} = (\mathbf{H}_{12} - \boldsymbol{\lambda}\mathbf{H}_{22})\mathbf{X}_{2,t} + \boldsymbol{\epsilon}_{1,t} - \boldsymbol{\lambda}\boldsymbol{\epsilon}_{2,t} \quad (3.2.9)$$

Como puede observarse, la ecuación (3.2.9) es una combinación lineal de las series del proceso \mathbf{Z}_t que no está influida por la secuencia de estados no estacionarios $\mathbf{X}_{1,t}$, y por tanto es estacionaria.

Por otra parte, cuando aumenta el valor de T (y por tanto también el valor de i), tanto \hat{n} (el estimador del orden del sistema, que se estudiará en el Capítulo 4) como las matrices de las ecuaciones (3.2.5) y (3.2.6) que representan el sistema, tienden a sus verdaderos valores (Casals, 1997). Por tanto, en particular se cumple que,

$$\lim_{T \rightarrow \infty} \hat{\boldsymbol{\Lambda}} = \boldsymbol{\Lambda} \quad (3.2.10)$$

□

Para evaluar el comportamiento del algoritmo de estimación del rango y la matriz de cointegración, se realizan dos nuevos ejercicios de simulación. Los Cuadros 3.2.10 y 3.2.11 presentan la probabilidad de estimar correctamente el número de relaciones de cointegración con el criterio $C(G_b)$ y, en los casos en que esta identificación es correcta, el vector de cointegración estimado. En el primer caso

se considera un sistema bivalente. El proceso que genera las realizaciones de este experimento es:

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 + \phi B \end{pmatrix} \begin{pmatrix} \nabla & 0 \\ \alpha & 1 \end{pmatrix} \begin{pmatrix} z_{1t} \\ z_{2t} \end{pmatrix} = \begin{pmatrix} \varepsilon_t \\ \eta_t \end{pmatrix}; \quad \Sigma_{\varepsilon\eta} = \begin{pmatrix} 1 & 0 \\ 0 & \sigma_\eta^2 \end{pmatrix} \quad (3.2.11)$$

donde α es el parámetro de cointegración. Los resultados se ofrecen para diferentes combinaciones de ϕ y σ_η^2 . Para cada tamaño muestral se han descartado las 50 primeras observaciones y se han realizado un total de 1000 repeticiones. En el Cuadro 3.2.10, $\alpha = -1$.

Los resultados revelan que el tamaño de la muestra no debe ser excesivamente grande para que las estimaciones del parámetro de cointegración sean razonablemente buenas. Este resultado es coherente con la teoría, puesto que las estimaciones de α son consistentes en todos los casos. Pese a que cuando $\phi = 0$, la estimación del parámetro de cointegración parece ser más exacta, no existen diferencias muy significativas entre los diferentes casos. Así, para $T = 50$, la estimación más precisa de α es de -0.999 y la menos precisa de -1.044 que apenas difieren en un 4.5% , mientras que para $T = 300$, las diferencias son aún menores: -1.000 y -0.988 , que representa un 1.2% . Por tanto, en este caso, la estimación del vector de cointegración no parece estar significativamente influida por la varianza de la transformación estacionaria o por la persistencia de los procesos autorregresivos.

El Cuadro 3.2.10 muestra también los resultados de la estimación del rango de cointegración. Estos sugieren que la varianza del error de la combinación lineal estacionaria no influye sobre el comportamiento de los criterios de detección de raíces iguales a la unidad. En cambio, como mostraban las simulaciones de la sección anterior, la estimación del rango de cointegración se ve directamente alterada por la persistencia de componentes autorregresivas. Como es lógico, el mejor de los casos es cuando $\phi = 0$, ya que el criterio de detección no tiene problemas para encontrar la única raíz unitaria del proceso. Cuando se añaden términos autorregresivos persistentes, en algunos casos el criterio los confunde con una segunda raíz unitaria. No obstante, el algoritmo se comporta suficientemente bien en estos modelos cuando el tamaño muestral no es demasiado pequeño.

Para ampliar el análisis de sensibilidad, modificamos el proceso generador de los datos. En este segundo caso, las simulaciones son generadas a partir del modelo:

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 + \phi B \end{pmatrix} \begin{pmatrix} \nabla & 0 & 0 \\ 0 & \nabla & 0 \\ \alpha_1 & \alpha_2 & 1 \end{pmatrix} \begin{pmatrix} z_{1t} \\ z_{2t} \\ z_{3t} \end{pmatrix} = \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \varepsilon_{3t} \end{pmatrix}; \quad \Sigma_\varepsilon = \begin{pmatrix} 1 & \theta & \theta \\ \theta & 1 & \theta \\ \theta & \theta & 1 \end{pmatrix} \quad (3.2.12)$$

Este segundo proceso tiene una única relación de cointegración, aunque dos factores comunes no estacionarios con los coeficientes de cointegración $\alpha_1 = -1.2$ y $\alpha_2 = .5$. Esto implica que existe una combinación lineal de las tres series que forman el proceso multivariante, que produce una serie estacionaria. De nuevo, se lleva a cabo un ejercicio de simulación con el mismo número de repeticiones aunque, en este caso, se obtienen los resultados para diferentes combinaciones de ϕ y de θ , con dos objetivos fundamentales: 1) analizar la estimación del vector de cointegración cuando existen dos factores comunes no estacionarios y 2) evaluar la influencia de la correlación contemporánea en los términos de error.

Los resultados del Cuadro 3.2.11 muestran que el valor del autorregresivo ϕ parece influir más en la estimación de los parámetros de cointegración que en el caso de que exista un sólo factor común no estacionario y las estimaciones parecen ser ligeramente menos precisas cuando $\phi = -.8$, que en los otros dos casos. Por otro lado, la existencia de correlación contemporánea del error parece influir en la estimación del vector de cointegración, aunque no queda claro si positiva o negativamente.

A su vez, el Cuadro 3.2.11 presenta la estimación del número de relaciones de cointegración. La probabilidad de estimar correctamente c disminuye considerablemente cuando $\phi = 0$ con respecto del caso bivalente, debido a que se está evaluando si $d > 2$, mientras que antes se evaluaba si $d > 1$. Esto no es sorprendente pues se encontró en el apartado anterior que, en muestras finitas, las propiedades de los criterios se deterioran cuando aumenta d . No ocurre lo mismo cuando el autorregresivo es más persistente debido a que este hecho facilita la correcta estimación de c . Esto se debe al uso del criterio $C(G_b)$ que presenta una mayor potencia en la estimación de raíces unitarias y, por tanto, es más “permisivo” al detectar relaciones de cointegración. Los resultados muestran además una relación positiva entre la correlación contemporánea de los términos de error y la probabilidad de estimar correctamente el número de relaciones de cointegración que se percibe levemente cuando $\phi = -.8$ y muy claramente para $\phi = 0$ y $\phi = .8$.

Cuadro 3.2.10: Rango y vector de cointegración para un proceso bivalente con $\alpha = -1^*$.

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 + \phi B \end{pmatrix} \begin{pmatrix} \nabla & 0 \\ \alpha & 1 \end{pmatrix} \begin{pmatrix} z_{1t} \\ z_{2t} \end{pmatrix} = \begin{pmatrix} \varepsilon_t \\ \eta_t \end{pmatrix}; \quad \Sigma_{\varepsilon\eta} = \begin{pmatrix} 1 & 0 \\ 0 & \sigma_\eta^2 \end{pmatrix}$$

σ_η^2	T	$\phi = -0.8$		$\phi = 0$		$\phi = 0.8$	
		$\hat{c} = 1$	$\hat{\alpha}$	$\hat{c} = 1$	$\hat{\alpha}$	$\hat{c} = 1$	$\hat{\alpha}$
0.5	50	0.709	-0.974	0.904	-0.996	0.631	-0.999
	100	0.923	-0.986	0.947	-1.001	0.873	-0.999
	300	0.990	-1.002	0.991	-0.999	0.989	-1.000
1	50	0.720	-1.044	0.922	-1.005	0.618	-1.003
	100	0.925	-1.029	0.948	-1.004	0.865	-1.001
	300	0.996	-0.998	0.988	-0.999	0.987	-1.000
2	50	0.712	-0.984	0.930	-1.013	0.608	-1.021
	100	0.907	-1.024	0.959	-1.002	0.870	-1.002
	300	0.991	-0.988	0.987	-1.000	0.984	-1.001
4	50	0.703	-0.993	0.935	-0.993	0.618	-0.988
	100	0.914	-1.058	0.958	-1.003	0.859	-1.001
	300	0.993	-1.006	0.985	-1.000	0.985	-1.000

* σ_η^2 es la varianza del error de la combinación lineal estacionaria, \hat{c} es el rango de cointegración estimado y $\hat{\alpha}$ es la estimación del coeficiente de cointegración. En alguna de las estimaciones del parámetro de cointegración con $T = 50$, se han detectado problemas debido a una mala estimación de la matriz \mathbf{H} del modelo de innovaciones. Por ello, se ha eliminado la influencia de las estimaciones 10 veces superiores al verdadero valor de α . En ningún caso, éstas suponen más del 1.0% de los casos.

Cuadro 3.2.11: Rango y vector de cointegración para un proceso trivariante con $\alpha_1 = -1.2$ y $\alpha_2 = .5^*$.

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 + \phi B \end{pmatrix} \begin{pmatrix} \nabla & 0 & 0 \\ 0 & \nabla & 0 \\ \alpha_1 & \alpha_2 & 1 \end{pmatrix} \begin{pmatrix} z_{1t} \\ z_{2t} \\ z_{3t} \end{pmatrix} = \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \varepsilon_{3t} \end{pmatrix}; \quad \Sigma_\varepsilon = \begin{pmatrix} 1 & \theta & \theta \\ \theta & 1 & \theta \\ \theta & \theta & 1 \end{pmatrix}$$

θ	T	$\phi = -0.8$			$\phi = 0$			$\phi = 0.8$		
		$\hat{c} = 1$	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{c} = 1$	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{c} = 1$	$\hat{\alpha}_1$	$\hat{\alpha}_2$
0.2	50	0.714	-1.213	0.433	0.696	-1.184	0.487	0.653	-1.197	0.505
	100	0.846	-1.198	0.509	0.843	-1.203	0.497	0.807	-1.202	0.504
	300	0.984	-1.198	0.505	0.974	-1.198	0.498	0.973	-1.201	0.502
0.4	50	0.705	-1.217	0.429	0.752	-1.193	0.520	0.695	-1.163	0.521
	100	0.875	-1.163	0.544	0.874	-1.199	0.504	0.866	-1.184	0.492
	300	0.986	-1.185	0.519	0.992	-1.198	0.504	0.992	-1.198	0.502
0.6	50	0.736	-1.084	0.446	0.817	-1.185	0.533	0.731	-1.183	0.535
	100	0.876	-1.109	0.535	0.913	-1.193	0.518	0.915	-1.194	0.508
	300	0.992	-1.172	0.523	0.994	-1.196	0.508	0.994	-1.197	0.502
0.8	50	0.725	-1.131	0.431	0.848	-1.173	0.526	0.788	-1.146	0.525
	100	0.879	-1.085	0.525	0.933	-1.174	0.488	0.918	-1.183	0.501
	300	0.992	-1.172	0.523	0.991	-1.194	0.505	0.995	-1.196	0.502

* θ es la correlación contemporánea en los términos de error, \hat{c} es el rango de cointegración estimado y $\hat{\alpha}_1$ y $\hat{\alpha}_2$ son los coeficientes de cointegración estimados. Al igual que en el caso anterior, se elimina la influencia de las estimaciones 10 veces superiores a los verdaderos valores de α_1 y α_2 producidas por una mala estimación de \mathbf{H} . Estas representan una frecuencia inferior al 2.0%.

3.3. Conclusiones del Capítulo

En este Capítulo se propone: 1) una metodología de detección de raíces unitarias (d) que se extiende al cálculo del rango de cointegración (c) cuando los procesos univariantes del sistema (o una transformación de éstos) presentan un orden de integración igual a la unidad y 2) un estimador consistente de la matriz de cointegración. El método se basa en las propiedades analizadas en la literatura de los coeficientes de correlación canónica entre los subespacios de información pasada y futura. Se ofrecen dos nuevos criterios de detección de d a partir de la estimación de sus funciones de penalización. Los resultados de los ejercicios de simulación reflejan la capacidad de los dos criterios propuestos, $C(G_a)$ y $C(G_b)$, para estimar el número de raíces unitarias y el rango de cointegración en diferentes situaciones. Asimismo, el rendimiento de estos criterios no parece estar influido significativamente por las varianzas de los términos de error, aunque sí por su correlación contemporánea. Las simulaciones no muestran indicios de que los criterios se deterioren con el aumento de la dimensión del sistema, aunque sí lo hacen si aumenta el número de raíces unitarias. Además la metodología, al contrario que otras en la literatura (Johansen, 1988, 1991), no precisa de la estimación de ningún modelo a priori y está basada en herramientas numéricas robustas (Golub y Van Loan, 1996).

Por otro lado, el cálculo del rango de cointegración y el supuesto de que las series sean $I(1)$ permiten obtener una estimación consistente de la matriz de cointegración a partir de un modelo formulado en Espacio de los Estados en forma de innovaciones. Los resultados ponen de manifiesto que en las situaciones estudiadas es muy probable que el uso de uno de los criterios de detección de raíces unitarias, del cálculo del rango y del estimador de la matriz de cointegración propuestos, conduzcan a estimaciones fiables de la matriz de cointegración. Todo esto sugiere la utilización del algoritmo NID, donde están implementados todos los aspectos de identificación desarrollados en este Capítulo, como un paso previo al análisis de la dinámica de un proceso multivariante.

Capítulo 4

Orden del Sistema e Índices de Kronecker

Una vez decididas por el analista las transformaciones convenientes para inducir estacionariedad en la(s) serie(s) objeto de análisis, la búsqueda del modelo más adecuado requiere determinar el índice de McMillan u orden del sistema (n), que es el mínimo orden de dinámica necesario para representar un proceso. Además, si el estudio se realiza sobre un vector de series, el cálculo de la matriz de cointegración necesita estimar un modelo en Espacio de los Estados (EE) y para ello, obtener el índice de McMillan. Por lo tanto, cualquiera que sea el propósito del análisis y/o el número de variables endógenas, la determinación de n es una decisión importante en la representación de procesos. Los trabajos que han impulsado el desarrollo de los algoritmos de subespacios no han dado demasiada importancia a este problema, probablemente debido a la utilización de muestras suficientemente largas. Sin embargo, en los últimos años la literatura presta una creciente atención a este aspecto, como muestran los trabajos de Bauer (1998), Sorelius (1999) o Bauer (2001).

En este Capítulo se derivan distintas opciones para su estimar n . Así, por un lado se presentan varios criterios y un contraste de hipótesis denominados “a priori”, ya que para su uso no es necesaria la estimación de modelos, y por otro, los llamados criterios “a posteriori”, cuya aplicación sí requiere un paso previo de estimación. En el marco de los primeros, se proponen dos criterios basados en correcciones del Criterio de Valores Singulares (*SVC*) de Bauer (2001), demostrándose su consistencia asintótica.

Por otra parte, en el ámbito de modelos multivariantes, una vez seleccionado el orden del sistema el usuario puede estar interesado en estimar un modelo en forma de innovaciones sin restricciones o, progresando en la metodología, estimar los llamados índices de observabilidad o índices de Kroneker (IK), que se corresponden

al orden de dinámica de cada una de las series del proceso. La imposición de estos IK sobre el modelo genera una estructura que reduce el número de parámetros. Casals (1997) demuestra que, a partir de la forma canónica de Luenberger (1967), un modelo de innovaciones puede escribirse como un modelo VARMAX echelon (ver Hannan y Deistler, 1988; Lütkepohl y Poskitt, 1996). El autor deriva un procedimiento que determina n y los IK simultáneamente y propone un algoritmo no iterativo que impone al modelo la estructura originada por los índices de observabilidad. En cambio, el procedimiento que se propone en este Capítulo sugiere seleccionar en un primer momento el orden del proceso y en una segunda fase, si se requiere, obtener los IK condicionados a la estimación previa de n .

El Capítulo se estructura en los siguientes apartados. En la Sección 4.1 se discuten varios métodos para obtener el índice de McMillan. Después, se trata brevemente la estimación del orden estacional en series univariantes y se analizan los resultados sobre la precisión de los diferentes procedimientos con ejercicios de Monte Carlo. En la Sección 4.2, se estudia la determinación de los IK, derivando una metodología de especificación de los mismos y realizando un análisis empírico de especificación automática. Por último, la Sección 4.3 resume las principales contribuciones del Capítulo.

4.1. Selección del Orden del Sistema

La selección de la dimensión del sistema no es un problema trivial en muestras cortas y generalmente requiere el uso de criterios o contrastes formales de hipótesis para tomar una decisión al respecto. En este apartado se revisan algunos de estos procedimientos, distinguiendo entre los métodos a priori y a posteriori.

4.1.1. Procedimientos a priori

Algunos autores, tanto en EE (Aoki, 1990), como en el contexto de modelos ARMA (Cooper y Wood, 1982; Tsay, 1989a), enfocan el problema de la identificación del orden del sistema mediante el análisis de las correlaciones canónicas. Concretamente, Tsay (1989a) propone usar el estadístico χ^2 para contrastar la significación de las $n+1, n+2, \dots, im$ menores correlaciones canónicas, donde i y m son la dimensión de los subespacios y el número de variables endógenas respectivamente, en procesos estacionarios. Casals (1997) adapta este estadístico a los métodos de subespacios formulándolo como:

$$D(n) = -(T - 2i + 1) \sum_{k=n+1}^{im} \log \left(1 - \frac{\hat{\sigma}_k^2}{\hat{d}_k} \right) \quad (4.1.1)$$

con

$$\hat{d}_k = 1 + 2 \sum_{l=1}^i \hat{\rho}_{pk}(l) \hat{\rho}_{fk}(l) \quad (4.1.2)$$

donde $\hat{\rho}_{fk}(l)$ y $\hat{\rho}_{pk}(l)$ son las autocorrelaciones de orden l de las variables canónicas \mathbf{C}_{fk} y \mathbf{C}_{pk} [ver Apéndice A.2]. Como indica Tsay (1989a), el término \hat{d}_k hace referencia a la varianza asintótica de los Coeficientes de Correlación Canónica (CCC) estimados ($\hat{\sigma}_k$) cuando existe correlación serial.

El propio Tsay (1989a) demuestra que el estadístico $D(n)$ se distribuye, bajo la hipótesis nula, como una χ^2 con $2(im - n)$ grados de libertad. Este resultado se generaliza para procesos no estacionarios en un trabajo posterior (Tiao y Tsay, 1989). Así, el estadístico (4.1.1) permite realizar contrastes para obtener el orden del sistema, tanto en procesos estacionarios como no estacionarios. El algoritmo de identificación NID introducido en el Capítulo anterior, incluye el p-valor de este contraste, que permite además, valorar la probabilidad de que el proceso estudiado pueda representarse con modelos con orden distinto.

Por otro lado, NID también comprende una familia de criterios sobre los Valores Singulares (VS). Como se vio en el Capítulo 2, introduciendo una determinada matriz de ponderación en el problema planteado por los métodos de subespacios, los VS obtenidos en la descomposición SVD coinciden con los CCC entre el subespacio de información pasada y el de información futura [ver Apéndice A.2]. Estos criterios utilizan la propiedad de que el número de VS de la descomposición de una matriz determinista es igual a su rango. En este caso, el rango de una proyección ortogonal de \mathbf{Z}_f coincide con el orden del sistema y, por tanto, con el número de valores singulares estadísticamente no nulos (Favoreel et al., 2000). A partir de esto, el criterio *SVC* de Bauer (2001) trata de obtener el menor valor singular a partir del cual el resto son despreciables o estadísticamente nulos. El autor lo define como:

$$SVC(n) = \hat{\sigma}_{n+1}^2 + C(T)d(n) \quad (4.1.3)$$

donde $\hat{\sigma}_{n+1}$ representa el valor singular $(n + 1)$ -ésimo de una proyección ortogonal de (2.4.3) y $d(n) = n(m + r) + nm + mr$ (donde r es el número de variables exógenas) denota el número de parámetros de un sistema de orden n en forma de EE (ver Hannan y Deistler, 1988, Teorema 2.5.3), excluyendo los de la matriz de covarianzas del ruido. Este criterio compara una medida de la información contenida en los VS, $\hat{\sigma}_{n+1}, \dots, \hat{\sigma}_{im}$, con una función de penalización ponderada por el número de parámetros del modelo¹. Se escoge entonces el argumento que

¹La medida utilizada por Bauer es la norma 2, con lo que $\|diag(\hat{\sigma}_{n+1}, \dots, \hat{\sigma}_{im})\| = \hat{\sigma}_{n+1}^2$, pero podrían igualmente aplicarse otras medidas.

minimiza la función (4.1.3). De nuevo, $C(T)$ es una penalización que debe cumplir ciertas características que se revisarán más adelante. Bauer (2001) propone usar la penalización habitual:

$$C(T) = \frac{\log(T)}{T} \quad (4.1.4)$$

y enumera las condiciones que tiene que cumplir una función de penalización para que el *SVC* genere estimaciones consistentes del orden del sistema. Aunque el criterio se muestra superior a otros criterios de información para tamaños muestrales relativamente grandes ($T > 100$), sigue siendo arbitraria la elección de la función de penalización. En Bauer (2001) se demuestra, además, que la matriz de ponderación \mathbf{W}_1 , empleada en la solución al problema tratado por los métodos de subespacios (2.4.4), influye notablemente en el rendimiento del *SVC* y por tanto en el resultado de la estimación de n , llegando a la conclusión de que $\mathbf{W}_1 = \mathbf{\Omega}_1^{-\frac{1}{2}} = (\mathbf{Z}_f \mathbf{\Pi}_{U_f}^\perp \mathbf{Z}'_f)^{-\frac{1}{2}}$ genera unos resultados superiores a los obtenidos con otras matrices de ponderación seleccionadas.

No obstante, cuando las muestras son más cortas, los resultados del *SVC* reflejan: 1) sensibilidad con respecto de la dimensión de los subespacios i , y 2) una importante infraparametrización, que aumenta sensiblemente en modelos multivariantes. Para adaptar el criterio a las propiedades de las series temporales económicas, donde los tamaños muestrales en muchas ocasiones son pequeños, se proponen dos modificaciones:

- (a) Sustituir la matriz de ponderación $\mathbf{\Omega}_1 = \mathbf{Z}_f \mathbf{\Pi}_{U_f}^\perp \mathbf{Z}'_f$ en (2.4.4) por la matriz $\mathbf{\Omega}_2 = \mathbf{Z}_f \mathbf{\Pi}_{M_p, U_f}^\perp \mathbf{Z}'_f$. Esta nueva matriz se corresponde con la matriz de covarianzas de los errores de predicción. Debe tenerse en cuenta que las perturbaciones del modelo (2.4.3) son errores de previsión, cuya varianza depende del horizonte temporal al que se refieren. Por ello, es adecuado ponderar la información dando más peso a los errores recientes que a los más alejados en el tiempo.
- (b) Utilizar una nueva función de penalización, denotada por $H(T, i)$, que depende del tamaño muestral (T) y de la dimensión de los subespacios de información (i), tal y como se propuso en los criterios de selección de raíces unitarias.

Así, como indica Bauer (2001), *el SVC es tan sólo uno de los criterios, sin embargo muchos otros son posibles, ya que simplemente se trata de comparar la cantidad de información que presentan los VS (medida con algún tipo de norma) con una función de penalización que depende del tamaño muestral*. A partir de esta afirmación, se propone el criterio *NIDC*, que pretende corregir los problemas

derivados del uso del *SVC* con las modificaciones (a) y (b). El criterio se formula como:

$$NIDC(n) = \hat{\sigma}_{n+1}^2 + H(T, i)d(n) \quad (4.1.5)$$

donde las diferencias con el criterio (4.1.3) radican en la matriz de ponderación (a) impuesta en (2.4.4) a partir de la cual se obtienen los valores singulares $\hat{\sigma}_{n+1}^2$ y en la función de penalización (b). Es conveniente destacar que sustituir la matriz de ponderación y la función de penalización no implican aumentar el coste computacional.

Para obtener una función de penalización que dependa del tamaño muestral y de la dimensión de los subespacios, se opera como en el Capítulo 3. Inicialmente, se parte del modelo $z_t = a_t$, donde a_t es un ruido blanco y se trabaja bajo la hipótesis de que los datos son generados por este proceso. Este supuesto tiene dos motivaciones principales: (i) las propiedades que tiene que cumplir (4.1.5) para generar estimaciones consistentes del orden del sistema con este proceso son sencillas de derivar y (ii) la función de penalización obtenida podrá utilizarse, en particular, para decidir acerca de la existencia de autocorrelación de un proceso. En este caso, para que *NIDC* seleccione $\hat{n} = 0$, en lugar de $\hat{n} = 1$ tiene que cumplirse que:

$$NIDC(0) < NIDC(1)$$

o lo que es lo mismo,

$$\hat{\sigma}_1^2 < \hat{\sigma}_2^2 + 2H(T, i)$$

ya que $d(n) = 2n$, y por tanto,

$$\nu_1 = \frac{\hat{\sigma}_1^2 - \hat{\sigma}_2^2}{2} < H(T, i)$$

De la misma forma, para que se seleccione $\hat{n} = 0$, en lugar de $\hat{n} = 2$ tiene que cumplirse,

$$\nu_2 = \frac{\hat{\sigma}_1^2 - \hat{\sigma}_3^2}{4} < H(T, i)$$

y así sucesivamente, por lo que puede definirse,

$$\nu_j = \frac{\hat{\sigma}_1^2 - \hat{\sigma}_{j+1}^2}{2j}$$

para $j = 1, \dots, i - 1$, que es una función de VS por encima de la cuál tiene que situarse la penalización para que el criterio funcione correctamente. De esta forma, a partir de la simulación de los VS de un ruido blanco podemos obtener una cota inferior para las posibles funciones de penalización, al menos para esta clase de procesos.

La Figura 4.1.1 muestra la distribución empírica de ν_1 y de ν_2 para $t = 20, 21, \dots, 500$. Estas distribuciones vienen definidas por un modelo ($z_t = a_t$) y por un percentil (95 %). Por otro lado, se muestra la función de penalización ajustada a ν_1 , $H(T, i)$ y la propuesta por la literatura, $C(T)$. La estimación de $H(T, i)$ es, para este modelo, una cota inferior por debajo de la cual se encuentran el 95 % de la distribución de ν_1 y un percentil mayor de ν_2 y en general de ν_j para $j = 3, \dots, i-1$.

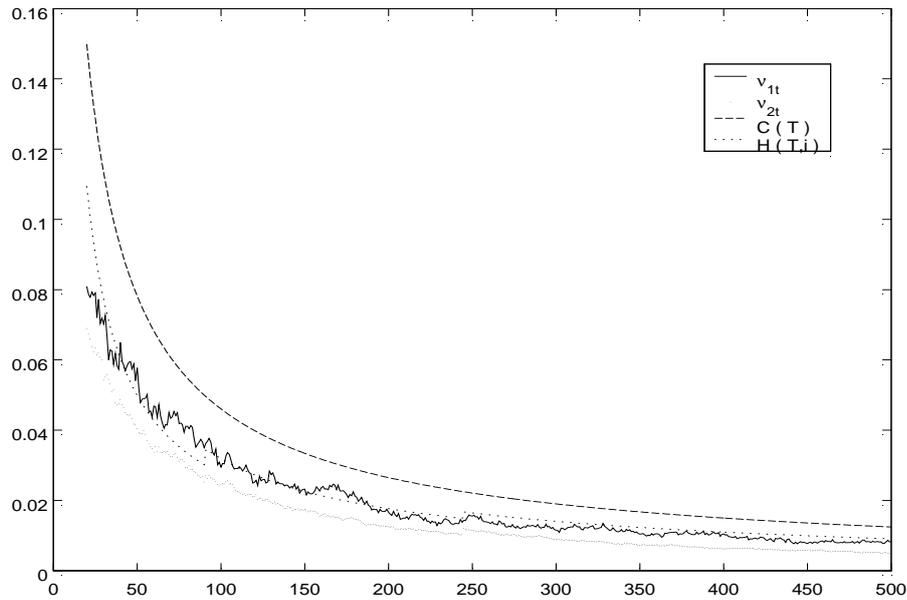


Figura 4.1.1: Función de penalización estimada para NIDC. ν_1 y ν_2 son dos funciones de valores singulares, obtenidos con un ruido blanco, que denotan el percentil 95 de las realizaciones. $H(T, i)$ es la función de penalización estimada para el criterio propuesto (NIDC) y $C(T)$ es la penalización de Bauer (2001).

La función de penalización $C(T)$ queda por encima de $H(T, i)$ fundamentalmente en muestras cortas y su distancia va disminuyendo cuando crece la muestra. Como se verá en el apartado de simulaciones, este hecho va provocar que NIDC mejore sustancialmente los resultados del SVC en la mayoría de los casos, cuando el número de observaciones es relativamente pequeño. La función de penalización $H(T, i)$ es, por tanto, válida para su utilización con el criterio NIDC cuando el proceso generador es un ruido blanco. Pero además, bajo supuestos no muy restrictivos se puede enunciar la proposición 4.1.1.

Proposición 4.1.1 *Bajo los supuestos del Teorema 3 de Bauer (2001), la función de penalización $H(T, i) = e^{-2T^{-.9}i^{1.6}}$ produce estimaciones asintóticamente*

consistentes del orden del sistema (con independencia del proceso analizado) minimizando $NIDC(n)$, ya que cumple: i) $H(T, i)T > 0$, ii) $H(T, i) \rightarrow 0$ y, iii) $\frac{H(T, i)T}{i^2 \log(\log(T))} \rightarrow \infty$, cuando $T \rightarrow \infty$.

Demostración. En la práctica se implementa $i_t = \max(4, h_t)$, siendo h_t el entero más cercano a $\log(T)$. Sin embargo, en los apartados ii) e iii) de la demostración se considera, sin pérdida de generalidad, que $i_t = \log(T)$. Durante toda la demostración se asume que $a = e^{-2}$, $b = .9$ y $c = 1.6$. Bajo estas condiciones:

i) Resulta inmediato ver que,

$$H(T, i)T = aT^{-b+1}i^c > 0 \quad \forall T > 0$$

cuando $i > 0 \quad \forall T$.

ii)

$$\lim_{T \rightarrow \infty} aT^{-b} \log(T)^c = \frac{a \log(T)^c}{T^b}$$

genera la indeterminación $\frac{\infty}{\infty}$. Aplicando la regla de l'Hôpital dos veces consecutivas se puede obtener,

$$\lim_{T \rightarrow \infty} \frac{ac(c-1) \log(T)^{c-2}}{b^2 T^b} = 0$$

y puesto que el numerador tiende a cero y el denominador a infinito,

$$\lim_{T \rightarrow \infty} H(T, i) = 0$$

iii) Por último,

$$\frac{H(T, i)T}{i^2 \log(\log(T))} = \frac{aT^{-b+1} \log(T)^{c-2}}{\log(\log(T))}$$

donde el numerador se puede escribir como,

$$\frac{a \log(T)^{c-2}}{T^{b-1}}$$

que produce una indeterminación $\frac{0}{0}$. Aplicando l'Hôpital se tiene que,

$$\lim_{T \rightarrow \infty} \frac{a(-b+1)T^{-b+3}}{(-c+2) \log(T)^{-c+1}} = \infty$$

y por tanto el límite del numerador tiende a infinito al igual que el denominador. Puede entonces aplicarse l'Hôpital sobre el límite inicial, obteniendo que:

$$\lim_{T \rightarrow \infty} \frac{aT^{-b+1} \log(T)^{c-2}}{\log(\log(T))} = \lim_{T \rightarrow \infty} aT^{1-b} [(1-b) \log(T)^{c-1} + (c-2) \log(T)^{c-2}] = \infty$$

□

Por tanto se propone un criterio para identificar el orden del sistema que, por construcción, sobreestima n con respecto del *SVC* en muestras finitas (ver Figura 4), pero que es asintóticamente consistente. Más adelante, se analizan los resultados de algunos ejercicios de simulación que comparan el funcionamiento de ambos criterios.

El hecho de preferir la sobreestimación a la infraestimación del orden del sistema se debe a que, en muchos casos, los pasos posteriores del procedimiento de análisis podrán conducir, mediante eliminación de coeficientes no significativos, a una modelización adecuada, aunque el modelo inicial esté sobreparametrizado.

4.1.2. Procedimientos a posteriori

Otro grupo de herramientas para elegir el orden del sistema implementadas en el algoritmo propuesto son los denominados criterios de información. Estos comparan el logaritmo de la función de verosimilitud de un sistema de orden n con una penalización ponderada por una medida de la complejidad del modelo, $d(n)$. De esta forma, sólo se aceptarán modelos más complejos si demuestran ser significativamente más adecuados. A esta familia pertenecen los conocidos criterios *AIC* (Akaike, 1976), *SBC* (Schwartz, 1978) y *HQ* (Hannan y Quinn, 1979). La expresión más general de los criterios de información (*CI*) es la siguiente,

$$CI(n) = \frac{-2 \log(l(n))}{T} + C(T)d(n) \quad (4.1.6)$$

donde, $l(n)$ representa el valor de la función de verosimilitud de un sistema de orden n , $C(T)$ es una función de penalización y $d(n)$ denota el número de parámetros de un sistema de orden n en forma de EE. En el algoritmo NID de identificación se usan, más concretamente, el criterio de decisión *AIC* donde $C(T) = \frac{2}{T}$, el *SBC* con $C(T) = \frac{\log(T)}{T}$ y el *HQ* con $C(T) = \frac{2 \log(\log(T))}{T}$.

Estos procedimientos a posteriori tienen la ventaja de que sus propiedades estadísticas son conocidas, pero también tienen inconvenientes. La principal desventaja de los *CI* frente a otros métodos, es la necesidad de estimar diferentes modelos.

Esto requiere, en muchos casos, una carga computacional demasiado alta. Además, por ejemplo, el criterio AIC muestra una conocida tendencia a sobreparametrizar (Gonzalo y Pitarakis, 2002). Sin embargo, estas desventajas no afectan gravemente a la metodología propuesta ya que, i) la estimación de modelos en EE mediante subespacios se realiza de forma no iterativa, por lo que la comparación entre distintos modelos es extremadamente rápida, y ii) las negativas consecuencias de la sobreparametrización (que se produce con algunos criterios) son menos graves, en principio, que las de infraparametrizar.

El algoritmo NID ofrece los valores estimados del orden del sistema con los criterios presentados. En concreto, devuelve el índice de McMillan óptimo obtenido por: 1) el criterio SVC_{Ω_2} que es el criterio de Bauer, a partir de la matriz de ponderación propuesta $\Omega_2 = \mathbf{Z}_f \mathbf{\Pi}_{M_p, U_f}^\perp \mathbf{Z}'_f$, 2) el nuevo criterio $NICD$, 3) los criterios a posteriori AIC , SBC y HQ y por último, 4) el contraste χ^2 al 95 % de Tiao y Tsay (1989). El coste computacional es muy pequeño, ya que se estima un sólo modelo en EE para cada posible orden del sistema y se realizan dos descomposiciones SVD de dos matrices distintas.

El principal objetivo del algoritmo es ofrecer al usuario una batería de herramientas para decidir el orden del proceso analizado. Así, se puede optar por uno u otro criterio dependiendo del objetivo del análisis: predicción, ajuste, etc. Además de la información aportada por los distintos criterios, el p-valor del contraste χ^2 permite valorar la distancia entre modelos con una representación de orden distinta.

Por otra parte, el cálculo de la matriz de cointegración en un conjunto de series, requiere estimar un modelo en EE y para ello, identificar la dimensión del sistema. Así, para estimar la matriz de cointegración, el algoritmo propuesto proporciona una especificación automática de n a partir de la información de los criterios y del contraste de hipótesis anteriores. En concreto, se calcula el índice de McMillan como la moda de los \hat{n} estimados por todos los criterios. Esto es, se obtiene el orden del proceso que más se repite con los distintos métodos expuestos. En caso de que exista un empate, se escoge el mayor de ellos debido a nuestra preferencia por sobreparametrizar el modelo. Dado el distinto comportamiento que muestran los procedimientos en los ejercicios de simulación según los modelos (ver Sección 4.1.3), este cálculo permite reducir el riesgo de error en la selección de la dimensión del sistema.

Selección del orden estacional:

En este apartado se expone un procedimiento para estimar el orden estacional en procesos univariantes. Casals (1997) generaliza los métodos de subespacios para la estimación de estructuras estacionales de tipo multiplicativo. Para ello, descompone el modelo original en dos subsistemas,

$$\Phi_s(B^s)z_t = \Theta_s(B^s)r_t \quad (4.1.7)$$

$$\phi(B)r_t = \theta(B)\epsilon_t \quad (4.1.8)$$

donde $\Phi_s(B^s) = 1 + \Phi_1 B^s + \Phi_2 B^{2s} + \dots + \Phi_P B^{Ps}$, $\phi(B) = 1 + \phi B + \phi_2 B^2 + \dots + \phi_p B^p$, $\Theta_s(B^s) = 1 + \Theta_1 B^s + \Theta_2 B^{2s} + \dots + \Theta_Q B^{Qs}$ y $\theta(B) = 1 + \theta B + \theta_2 B^2 + \dots + \theta_q B^q$ son los polinomios autorregresivos y media móvil, estacionales y regulares. Además, r_t es un error no observable del subsistema estacional que alberga la estructura regular de la serie z_t .

Para la estimación de la estructura estacional se define la matriz Block Hankel Estacional o de periodo s como,

$$\begin{pmatrix} \mathbf{Z}_p^s \\ \mathbf{Z}_f^s \end{pmatrix} = \begin{pmatrix} \mathbf{Z}_{0:i-1}^s \\ \mathbf{Z}_{i:2i-1}^s \end{pmatrix} = \begin{pmatrix} z_0 & z_1 & \dots & z_{T-s(2i-1)-1} \\ z_s & z_{s+1} & \dots & z_{T-s(2i-2)-1} \\ \vdots & \vdots & & \vdots \\ z_{s(i-1)} & z_{si} & \dots & z_{T-s(i-1)-1} \\ z_{si} & z_{s(i+1)} & \dots & z_{T-si-1} \\ z_{s(i+1)} & z_{s(i+2)} & \dots & z_{T-s(i+1)-1} \\ \vdots & \vdots & & \vdots \\ z_{s(2i-1)} & z_{s(2i-1)+1} & \dots & z_{T-1} \end{pmatrix} \quad (4.1.9)$$

que, al igual que la matriz Block Hankel (2.3.1), contiene toda la información muestral de z_t , dividida en la información estacional pasada, \mathbf{Z}_p^s y futura, \mathbf{Z}_f^s . Así, utilizando la descomposición (4.1.7)-(4.1.8) es posible estimar en dos etapas: i) un modelo para la parte estacional, a partir de la información ordenada en una matriz como la dada en (4.1.9) y ii) un modelo para la parte regular a partir de los residuos obtenidos en la fase anterior ordenados en la matriz (2.3.1). De esta forma, pueden utilizarse tanto los métodos a priori como a posteriori para seleccionar los órdenes estacionales y regulares de los subsistemas (4.1.7) y (4.1.8), respectivamente. En la práctica, para identificar la dimensión de los distintos subsistemas, se sigue el método siguiente:

1. Calcular el orden óptimo del subsistema estacional aplicando los procedimientos presentados en esta sección a la matriz Block Hankel Estacional (4.1.9).
2. Seleccionar un único orden estacional, \hat{n}_s , como la moda de los anteriores y, si éstos se repiten un mismo número de veces, escoger el mayor valor que más veces se repite.
3. Estimar mediante subespacios el subsistema estacional con un modelo equilibrado $\text{ARMA}(\hat{n}_s, \hat{n}_s)_s$ sin especificar estructura regular, obteniendo unos residuos \hat{r}_t .
4. Calcular \hat{n}_r , el índice de McMillan regular de \hat{r}_t , aplicando los métodos de selección.

Como indica Casals (1997), el procedimiento expuesto puede ser invertido de manera que la estructura regular se identifique en primer lugar y posteriormente la estacional. Sin embargo, el método utilizado presenta dos ventajas fundamentales: 1) en series con un periodo estacional corto (por ejemplo, trimestrales) ofrece mejores resultados debido a la mayor dificultad de los criterios en seleccionar primero el orden de la dinámica regular en la matriz (2.3.1) que el orden de dinámica estacional en (4.1.9) y, 2) la decisión sobre el orden estacional es más sencilla, en principio, que la del orden regular, lo que reduce la posibilidad de trasladar el error a la segunda decisión.

4.1.3. Ejercicios de simulación

En este apartado se realizan simulaciones de diferentes modelos univariantes y multivariantes para estudiar la capacidad de los distintos criterios y contrastes para seleccionar adecuadamente el orden del sistema. Los Cuadros 4.1.1 a 4.1.10 ofrecen la frecuencia relativa de estimar correctamente, e infraestimar o sobreestimar en una sola unidad el valor del índice de McMillan. Para cada uno de los modelos se han descartado las 50 primeras observaciones, analizando en cada caso un total de 1000 repeticiones con diferentes tamaños muestrales.

Para cada modelo y muestra, se subrayan las frecuencias relativas que ofrecen la mayor probabilidad de acierto. Los Cuadros presentan los resultados de todos los criterios expuestos en el apartado anterior que coinciden con los implementados en NID (salvo el criterio SVC_{Ω_1}). También muestran el cálculo de la moda de los valores óptimos de \hat{n} , que se utiliza en la especificación automática y la estimación de la matriz de cointegración.

El Cuadro 4.1.1 ofrece los resultados cuando el sistema es generado por un ruido blanco. El criterio *SBC* muestra resultados superiores al resto para todos los tamaños muestrales seleccionados. En este primer caso, el criterio SVC_{Ω_1} , que es el criterio original de Bauer (2001), domina a los dos propuestos ya que no existe riesgo de infraparametrización. En muestras cortas, ambos sobreestiman, generalmente en una unidad, el verdadero valor de n .

El segundo proceso generador es un autorregresivo de orden uno con poca persistencia, lo que implica un considerable deterioro de los resultados de todos los criterios. Aquí, el criterio *AIC* se muestra superior para casi todos los tamaños muestrales. Por otro lado, los criterios SVC_{Ω_2} y *NIDC* dominan ampliamente a SVC_{Ω_1} , que presenta los peores resultados del conjunto. Además, teniendo en cuenta que una leve sobreparametrización puede conducir a un modelo apropiado en el paso siguiente, el criterio *NIDC* se revela como una de las mejores opciones cuando la muestra es relativamente corta, $T \leq 100$.

El proceso generador M3 también es de orden uno. En este caso, el parámetro de la media móvil es lo suficientemente alto como para que los criterios no tengan demasiadas dificultades para estimar n correctamente, aún en muestras bastante cortas. De nuevo, el criterio *SBC* exhibe los mejores resultados. Los tres criterios basados en el estudio de los VS muestran unos resultados bastante similares.

En el Cuadro 4.1.4 se exponen los resultados de los diversos criterios para el modelo M4. Este tiene una estructura AR(2) con raíces imaginarias y parámetros poco persistentes que resulta difícil de identificar. Prueba de ello son los resultados del *SBC*, que no alcanza el 10% de estimaciones correctas de n para ningún tamaño muestral considerado. El resto de criterios y el test χ^2 ofrecen un mejor resultado al menos para muestras grandes. En muestras inferiores o iguales a 300 observaciones, el criterio propuesto *NIDC* presenta una frecuencia de acierto muy superior al resto, y especialmente, al SVC_{Ω_1} que revela un fuerte grado de infraestimación.

El modelo M5 contiene la misma estructura regular que el anterior pero añade una media móvil de orden uno en la parte estacional. El proceso emula una serie mensual, con lo que $s = 12$. En este caso, se prescinde de la simulación con 30 y 50 datos debido al escaso número de grados de libertad para estimar la parte estacional. Como era de esperar, todos los criterios estiman correctamente el orden del subsistema estacional aunque el contraste de hipótesis χ^2 encuentra una mayor dificultad que el resto, sobre todo en la muestra más corta. En la segunda etapa, los resultados en la determinación del orden regular son similares a los del Cuadro 4.1.4 lo que parece indicar que el filtrado de estructura estacional utilizado

no distorsiona significativamente los resultados de la estimación del orden regular.

El proceso M6 es igual al M4 en la parte autorregresiva añadiendo además un parámetro media móvil regular, con lo que se obtiene un ARMA(2,1) con raíces imaginarias. El modelo continúa siendo de orden 2 y la incorporación de la parte media móvil mejora la estimación de n en todos los métodos presentados. Pese a todo, salvo para los criterios AIC , HQ y $NIDC$ los resultados siguen siendo inferiores al 50% cuando $T \leq 100$ observaciones. Obsérvese cómo la utilización de la matriz de ponderación Ω_2 en el criterio SVC mejora sustancialmente los resultados.

El Cuadro 4.1.8 resume el comportamiento de los métodos de identificación del índice de McMillan en un sistema bivariante VARMA(1,1) con $n = 2$. Este presenta, en uno de los procesos univariantes, un término autorregresivo y una media móvil cercanos a la cancelación. Los resultados, que son homogéneos cuando la muestra es larga, son muy dispares en muestras cortas. El SVC propuesto por Bauer (2001) genera una alarmante tendencia a infraestimar el orden del proceso cuando $T \leq 100$ observaciones. Este hecho es corregido en los dos criterios propuestos SVC_{Ω_2} y $NIDC$, que dominan claramente al anterior en muestras cortas y presentan unos resultados muy similares con una muestra suficientemente grande. Para $T \leq 100$, $NIDC$ ofrece una mayor frecuencia de estimaciones correctas y además, un mayor grado de sobreestimación en una unidad.

Los Cuadros 4.1.9 y 4.1.10 hacen referencia a dos nuevos modelos multivariantes. En ellos, se han eliminado los resultados para $T = 30$ para aumentar los grados de libertad. El primero es un proceso trivariante de orden 3 correspondiente a una estructura VARMA(1,1) utilizada por Koreisha y Pukkila (1989). Estos autores proponen un procedimiento para la estimación lineal de modelos VARMA. Los resultados de la simulaciones están en línea con los presentados en los Cuadros anteriores. No obstante, el SVC_{Ω_1} tiende a infraestimar n para $T \leq 100$, hasta el punto de que ni siquiera es capaz de obtener $\hat{n} = n - 1$. Las correcciones realizadas en los criterios propuestos permiten evitar este efecto, mostrando unos resultados sensiblemente superiores para casi todos los tamaños muestrales estudiados. Ambos presentan una frecuencia mínima de acierto del 35% con $T = 50$ (superior al resto de métodos), siendo el comportamiento de $NIDC$ algo superior al de SVC_{Ω_2} . Además, los dos criterios exhiben un notable grado de sobreestimación cuando la muestra es muy corta (con $T = 50$, $\hat{n} = n + 1$ alrededor del 30% de las veces) lo que posibilita, en el paso siguiente, mediante contrastes de hipótesis y restricciones de exclusión, una adecuada estimación de la dimensión del sistema. Por otro lado, los mejores resultados no se corresponden a uno o dos criterios, sino que varían indistintamente dependiendo del tamaño de la muestra. El segundo proceso se corresponde también a un modelo trivariante con una es-

estructura VARMA(1,1). El orden es también igual a 3 y el sistema es usado por Reinsel et al. (1992) para proponer un procedimiento que reduce la evaluación de la función de verosimilitud condicional de un modelo VARMA a un problema de mínimos cuadrados generalizados. Los resultados de los criterios son, en general, superiores a los presentados en el cuadro anterior debido a que la estructura del proceso M8 se encuentra a menor distancia de una representación de orden inferior. Pese a ello, de nuevo el criterio SVC_{Ω_1} muestra un alto grado de infraestimación en muestras cortas, corregido en los criterios propuestos. En este caso, el contraste χ^2 ofrece los mejores resultados cuando $T \leq 100$, mientras que todos los criterios se comportan razonablemente bien con tamaños muestrales superiores.

En resumen, el ejercicio de simulación realizado revela las siguientes propiedades de los distintos criterios de selección del orden del sistema, que pueden ser muy útiles para la decisión del analista.

- Los resultados globales de las simulaciones muestran la consistencia de los métodos utilizados y en particular de los criterios propuestos, SVC_{Ω_2} y $NIDC$. Tan sólo el criterio AIC ofrece alguna evidencia de falta de consistencia en los modelos M1 y M3.
- *Criterios basados en VS.* Entre los criterios que utilizan la descomposición SVD, el SVC_{Ω_1} muestra el peor comportamiento en muestras cortas (en general el peor de todos los métodos contrastados), sobre todo en modelos multivariantes. SVC_{Ω_2} y en mayor medida $NIDC$ mejoran este hecho. Además, cuando las muestras son pequeñas, los dos criterios propuestos sobreestiman el orden del sistema. Por otro lado, $NIDC$ ofrece frecuentemente un mejor comportamiento cuando los parámetros del modelo son poco persistentes.
- *Criterios de información.* Como era de esperar, estos criterios presentan resultados diferentes entre sí. El criterio AIC sobreestima ocasionalmente, pero destaca más su comportamiento no consistente con algunos procesos estudiados. Sin embargo, funciona razonablemente bien en modelos multivariantes. El SBC es una de las mejores opciones, tanto en modelos univariantes como multivariantes, cuando la muestra es aceptablemente grande y/o los parámetros son suficientemente persistentes. No obstante, presenta peores resultados con un modelo AR(2) con raíces imaginarias. Por su parte, el criterio HQ presenta, en general, un rendimiento intermedio entre los dos anteriores. Funciona mejor que SBC (aunque peor que AIC) para autorregresivos de poca persistencia o modelos con estructuras cercanas a la cancelación, pero peor (aunque mejor que AIC) en modelos más sencillos de identificar.
- *Contraste χ^2 .* El test χ^2 al 95 % ofrece unos resultados relativamente débiles

cuando la muestra es corta, pero mejora sustancialmente cuando ésta aumenta. Aunque en procesos univariantes no suele ser la mejor opción, esto cambia en sistemas multivariantes donde su rendimiento mejora significativamente.

En conclusión, los ejercicios de simulación presentados muestran la capacidad de las correcciones propuestas en SVC_{Ω_2} y $NIDC$, para estimar el índice de McMillan en un amplio espectro de procesos. Ambos tienen, además, un coste computacional muy pequeño. Se demuestra analíticamente la consistencia de estos criterios y los resultados de los experimentos de Monte Carlo confirman este hecho.

La comparación entre los diferentes métodos de selección del orden del sistema muestra que ninguno de ellos es sistemáticamente dominado por otro. Los resultados parecen indicar que el rendimiento de cada uno de ellos depende de: i) el proceso generador de los datos, ii) la persistencia de los parámetros y iii) el tamaño de la muestra. Así, la implementación de todos los criterios presentados y el contraste de hipótesis ofrece al analista un abanico de posibilidades para determinar n . Además, para su utilización automática, el estimador \hat{n} obtenido como la moda de los órdenes óptimos de cada uno de los métodos, reduce el riesgo de error en la decisión sobre n respecto del uso continuado de uno de ellos.

Finalmente, la utilización de los métodos de subespacios permite usar todos los procedimientos presentados con un mínimo coste. Además, la formulación en EE posibilita la búsqueda del índice de McMillan adecuado avanzando de uno en uno (añadiendo cada vez un estado más), mientras que, por ejemplo, las representaciones VAR o VARMA lo hacen de m en m (añadiendo en cada paso de la búsqueda m estados más), siendo m el número de variables endógenas. Esta manera de proceder proporciona, lógicamente, ganancias sustanciales en términos de austeridad paramétrica.

Cuadro 4.1.1: Resultados de la estimación del orden del sistema para $M1^*$.

$M1(n = 0): z_t = a_t, \quad a_t \sim iidN(0, 1).$

T	SVC_{Ω_1}	SVC_{Ω_2}	$NIDC$	AIC	SBC	HQ	χ^2	$Moda$
$\hat{n} = n$								
30	0.945	0.812	0.703	0.828	<u>0.965</u>	0.846	0.948	0.899
50	0.948	0.850	0.726	0.816	<u>0.979</u>	0.889	0.950	0.919
100	0.918	0.862	0.797	0.768	<u>0.975</u>	0.908	0.955	0.927
300	0.905	0.889	0.867	0.722	<u>0.990</u>	0.944	0.939	0.948
500	0.934	0.919	0.892	0.711	<u>0.993</u>	0.947	0.950	0.953
$\hat{n} = n + 1$								
30	0.053	0.152	0.250	0.156	0.035	0.141	0.040	0.098
50	0.050	0.132	0.266	0.172	0.020	0.115	0.031	0.065
100	0.075	0.118	0.184	0.203	0.025	0.092	0.036	0.057
300	0.083	0.096	0.110	0.239	0.010	0.055	0.042	0.041
500	0.059	0.071	0.092	0.239	0.007	0.063	0.034	0.031

* Los procedimientos comparados son los siguientes: SVC_{Ω_1} corresponde al criterio de Bauer (2001), SVC_{Ω_2} es el mismo criterio con la matriz de ponderación propuesta $\mathbf{\Omega}_2 = \mathbf{Z}_f \mathbf{\Pi}_{M_p, U_f}^\perp \mathbf{Z}'_f$, $NIDC$ sustituye además la función de penalización $C(T) = \frac{\log(T)}{T}$ por la función de penalización estimada $H(T, i)$. AIC , SBC y HQ son los criterios de información de Akaike (1976), Schwartz (1978) y Hannan y Quinn (1979), respectivamente. χ^2 es el contraste de Tiao y Tsay (1989) al 95% y $Moda$ es el valor más repetido de los \hat{n} estimados por los métodos anteriores (excepto SVC_{Ω_1} , que no se incluye en NID) en cada realización. En caso de empate se selecciona el mayor valor que más veces se repite. Se subrayan las frecuencias relativas que representan un mayor acierto para cada tamaño muestral.

Cuadro 4.1.2: Resultados de la estimación del orden del sistema para $M2^*$.

$$M2(n = 1): (1 - 0.2B)z_t = a_t, \quad a_t \sim iidN(0, 1).$$

T	SVC_{Ω_1}	SVC_{Ω_2}	$NIDC$	AIC	SBC	HQ	χ^2	$Moda$
$\hat{n} = n - 1$								
30	0.913	0.801	0.384	0.680	0.866	0.682	0.878	0.721
50	0.861	0.728	0.445	0.551	0.818	0.672	0.841	0.657
100	0.782	0.676	0.498	0.314	0.697	0.521	0.729	0.532
300	0.486	0.415	0.373	0.037	0.273	0.130	0.246	0.212
500	0.243	0.204	0.166	0.007	0.081	0.027	0.051	0.078
$\hat{n} = n$								
30	0.086	0.185	<u>0.385</u>	0.315	0.134	0.305	0.098	0.272
50	0.133	0.244	0.395	<u>0.435</u>	0.182	0.311	0.127	0.324
100	0.204	0.290	0.408	<u>0.649</u>	0.301	0.470	0.238	0.446
300	0.490	0.548	0.575	<u>0.916</u>	0.724	0.849	0.713	0.771
500	0.724	0.757	0.778	<u>0.967</u>	0.915	0.961	0.906	0.922
$\hat{n} = n + 1$								
30	0.001	0.014	0.211	0.005	0.0	0.013	0.019	0.016
50	0.006	0.027	0.115	0.013	0.0	0.017	0.026	0.019
100	0.013	0.033	0.081	0.036	0.002	0.009	0.023	0.019
300	0.020	0.036	0.047	0.043	0.003	0.021	0.027	0.015
500	0.032	0.039	0.054	0.024	0.004	0.012	0.029	0.025

* Los procedimientos comparados son los siguientes: SVC_{Ω_1} corresponde al criterio de Bauer (2001), SVC_{Ω_2} es el mismo criterio con la matriz de ponderación propuesta $\Omega_2 = \mathbf{Z}_f \mathbf{\Pi}_{M_p, U_f}^\perp \mathbf{Z}_f'$, $NIDC$ sustituye además la función de penalización $C(T) = \frac{\log(T)}{T}$ por la función de penalización estimada $H(T, i)$. AIC , SBC y HQ son los criterios de información de Akaike (1976), Schwartz (1978) y Hannan y Quinn (1979), respectivamente. χ^2 es el contraste de Tiao y Tsay (1989) al 95% y $Moda$ es el valor más repetido de los \hat{n} estimados por los métodos anteriores (excepto SVC_{Ω_1} , que no se incluye en NID) en cada realización. En caso de empate se selecciona el mayor valor que más veces se repite. Se subrayan las frecuencias relativas que representan un mayor acierto para cada tamaño muestral.

Cuadro 4.1.3: Resultados de la estimación del orden del sistema para $M3^*$.

$$M3(n = 1): z_t = (1 - 0.8B)a_t, \quad a_t \sim iidN(0, 1).$$

T	SVC_{Ω_1}	SVC_{Ω_2}	$NIDC$	AIC	SBC	HQ	χ^2	$Moda$
$\hat{n} = n - 1$								
30	0.179	0.032	0.021	0.021	0.057	0.025	0.284	0.020
50	0.019	0.001	0.0	0.0	0.006	0.0	0.018	0.0
100	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
300	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
500	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
$\hat{n} = n$								
30	0.793	0.819	0.761	0.932	<u>0.934</u>	0.929	0.676	0.919
50	0.929	0.850	0.767	0.932	<u>0.988</u>	0.978	0.946	0.960
100	0.941	0.907	0.813	0.935	<u>0.998</u>	0.981	0.965	0.967
300	0.949	0.934	0.918	0.915	<u>1.0</u>	0.984	0.971	0.973
500	0.954	0.944	0.931	0.909	<u>1.0</u>	0.981	0.966	0.980
$\hat{n} = n + 1$								
30	0.027	0.124	0.166	0.045	0.009	0.041	0.03	0.055
50	0.048	0.120	0.171	0.067	0.006	0.022	0.017	0.029
100	0.052	0.076	0.126	0.063	0.002	0.019	0.012	0.034
300	0.044	0.053	0.062	0.082	0.0	0.016	0.013	0.019
500	0.040	0.047	0.053	0.090	0.0	0.019	0.01	0.014

* Los procedimientos comparados son los siguientes: SVC_{Ω_1} corresponde al criterio de Bauer (2001), SVC_{Ω_2} es el mismo criterio con la matriz de ponderación propuesta $\Omega_2 = \mathbf{Z}_f \mathbf{\Pi}_{M_p, U_f}^\perp \mathbf{Z}_f'$, $NIDC$ sustituye además la función de penalización $C(T) = \frac{\log(T)}{T}$ por la función de penalización estimada $H(T, i)$. AIC , SBC y HQ son los criterios de información de Akaike (1976), Schwartz (1978) y Hannan y Quinn (1979), respectivamente. χ^2 es el contraste de Tiao y Tsay (1989) al 95% y $Moda$ es el valor más repetido de los \hat{n} estimados por los métodos anteriores (excepto SVC_{Ω_1} , que no se incluye en NID) en cada realización. En caso de empate se selecciona el mayor valor que más veces se repite. Se subrayan las frecuencias relativas que representan un mayor acierto para cada tamaño muestral.

Cuadro 4.1.4: Resultados de la estimación del orden del sistema para M_4^* .

$$M_4(n = 2): (1 - 0.4B + 0.3B^2)z_t = a_t, \quad a_t \sim iidN(0, 1).$$

T	SVC_{Ω_1}	SVC_{Ω_2}	$NIDC$	AIC	SBC	HQ	χ^2	$Moda$
$\hat{n} = n - 1$								
30	0.247	0.419	0.415	0.652	0.441	0.628	0.257	0.563
50	0.439	0.555	0.480	0.800	0.673	0.803	0.520	0.664
100	0.653	0.640	0.517	0.814	0.939	0.909	0.767	0.756
300	0.487	0.445	0.401	0.460	0.958	0.796	0.464	0.527
500	0.290	0.263	0.233	0.229	0.908	0.605	0.204	0.249
$\hat{n} = n$								
30	0.026	0.147	<u>0.326</u>	0.085	0.022	0.062	0.082	0.110
50	0.079	0.179	<u>0.376</u>	0.126	0.018	0.049	0.111	0.167
100	0.199	0.283	<u>0.402</u>	0.182	0.019	0.081	0.157	0.216
300	0.500	0.532	<u>0.567</u>	0.539	0.042	0.204	0.481	0.463
500	0.689	0.710	0.728	<u>0.771</u>	0.092	0.395	0.738	0.738
$\hat{n} = n + 1$								
30	0.0	0.006	0.039	0.002	0.0	0.0	0.012	0.009
50	0.001	0.005	0.053	0.001	0.0	0.0	0.021	0.006
100	0.004	0.015	0.043	0.003	0.0	0.0	0.023	0.004
300	0.013	0.023	0.032	0.001	0.0	0.0	0.055	0.010
500	0.021	0.027	0.038	0.0	0.0	0.0	0.058	0.013

* Los procedimientos comparados son los siguientes: SVC_{Ω_1} corresponde al criterio de Bauer (2001), SVC_{Ω_2} es el mismo criterio con la matriz de ponderación propuesta $\Omega_2 = \mathbf{Z}_f \mathbf{\Pi}_{M_p, U_f}^\perp \mathbf{Z}_f'$, $NIDC$ sustituye además la función de penalización $C(T) = \frac{\log(T)}{T}$ por la función de penalización estimada $H(T, i)$. AIC , SBC y HQ son los criterios de información de Akaike (1976), Schwartz (1978) y Hannan y Quinn (1979), respectivamente. χ^2 es el contraste de Tiao y Tsay (1989) al 95% y $Moda$ es el valor más repetido de los \hat{n} estimados por los métodos anteriores (excepto SVC_{Ω_1} , que no se incluye en NID) en cada realización. En caso de empate se selecciona el mayor valor que más veces se repite. Se subrayan las frecuencias relativas que representan un mayor acierto para cada tamaño muestral.

Cuadro 4.1.5: Estimación del orden del sistema para M5 - Orden estacional*.

M5($n_r = 2, n_s = 1$): $(1 - 0.4B + 0.3B^2)z_t = (1 - 0.6B^{12})a_t, a_t \sim iidN(0, 1)$.

T	SVC_{Ω_1}	SVC_{Ω_2}	$NIDC$	AIC	SBC	HQ	χ^2	$Moda$
$\hat{n}_s = n_s - 1$								
100	0.026	0.014	0.007	0.0	0.008	0.001	0.474	.003
300	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
500	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
$\hat{n}_s = n_s$								
100	0.969	0.972	0.960	0.984	0.992	<u>0.996</u>	0.460	0.988
300	0.944	0.935	0.930	0.971	<u>1.0</u>	<u>0.996</u>	0.878	0.972
500	0.946	0.938	0.931	0.942	<u>1.0</u>	0.998	0.894	0.975
$\hat{n}_s = n_s + 1$								
100	0.005	0.014	0.033	0.016	0.0	0.003	0.066	0.009
300	0.053	0.066	0.070	0.029	0.0	0.004	0.122	0.028
500	0.048	0.062	0.069	0.058	0.0	0.002	0.106	0.025

* Los procedimientos comparados son los siguientes: SVC_{Ω_1} corresponde al criterio de Bauer (2001), SVC_{Ω_2} es el mismo criterio con la matriz de ponderación propuesta $\Omega_2 = Z_f \Pi_{M_p, U_f}^\perp Z_f'$, $NIDC$ sustituye además la función de penalización $C(T) = \frac{\log(T)}{T}$ por la función de penalización estimada $H(T, i)$. AIC , SBC y HQ son los criterios de información de Akaike (1976), Schwartz (1978) y Hannan y Quinn (1979), respectivamente. χ^2 es el contraste de Tiao y Tsay (1989) al 95% y $Moda$ es el valor más repetido de los \hat{n} estimados por los métodos anteriores (excepto SVC_{Ω_1} , que no se incluye en NID) en cada realización. En caso de empate se selecciona el mayor valor que más veces se repite. Se subrayan las frecuencias relativas que representan un mayor acierto para cada tamaño muestral.

Cuadro 4.1.6: Estimación del orden del sistema para M5 - Orden regular*.
M5($n_r = 2, n_s = 1$): $(1 - 0.4B + 0.3B^2)z_t = (1 - 0.6B^{12})a_t, a_t \sim iidN(0, 1)$.

T	SVC_{Ω_1}	SVC_{Ω_2}	$NIDC$	AIC	SBC	HQ	χ^2	$Moda$
$\hat{n}_r = n_r - 1$								
100	0.657	0.642	0.526	0.829	0.918	0.907	0.711	0.777
300	0.489	0.444	0.403	0.510	0.960	0.808	0.422	0.509
500	0.269	0.242	0.204	0.240	0.911	0.604	0.156	0.238
$\hat{n}_r = n_r$								
100	0.181	0.264	<u>0.401</u>	0.162	0.023	0.071	0.205	0.187
300	0.495	0.529	<u>0.562</u>	0.487	0.040	0.192	0.536	0.476
500	0.712	0.733	0.762	0.759	0.089	0.396	<u>0.802</u>	0.753
$\hat{n}_r = n_r + 1$								
100	0.001	0.003	0.027	0.0	0.0	0.0	0.021	0.002
300	0.015	0.026	0.034	0.003	0.0	0.0	0.026	0.014
500	0.017	0.023	0.032	0.001	0.0	0.0	0.024	0.007

* Los procedimientos comparados son los siguientes: SVC_{Ω_1} corresponde al criterio de Bauer (2001), SVC_{Ω_2} es el mismo criterio con la matriz de ponderación propuesta $\Omega_2 = Z_f \Pi_{M_p, U_f}^\perp Z_f'$, $NIDC$ sustituye además la función de penalización $C(T) = \frac{\log(T)}{T}$ por la función de penalización estimada $H(T, i)$. AIC , SBC y HQ son los criterios de información de Akaike (1976), Schwartz (1978) y Hannan y Quinn (1979), respectivamente. χ^2 es el contraste de Tiao y Tsay (1989) al 95% y $Moda$ es el valor más repetido de los \hat{n} estimados por los métodos anteriores (excepto SVC_{Ω_1} , que no se incluye en NID) en cada realización. En caso de empate se selecciona el mayor valor que más veces se repite. Se subrayan las frecuencias relativas que representan un mayor acierto para cada tamaño muestral.

Cuadro 4.1.7: Resultados de la estimación del orden del sistema para $M6^*$.

$$M6(n = 2): (1 - 0.4B + 0.3B^2)z_t = (1 - 0.7B)a_t, \quad a_t \sim iidN(0, 1).$$

T	SVC_{Ω_1}	SVC_{Ω_2}	$NIDC$	AIC	SBC	HQ	χ^2	$Moda$
$\hat{n} = n - 1$								
30	0.337	0.461	0.363	0.444	0.256	0.386	0.243	0.393
50	0.589	0.588	0.360	0.483	0.481	0.510	0.478	0.524
100	0.611	0.502	0.353	0.280	0.675	0.429	0.604	0.408
300	0.168	0.129	0.115	0.015	0.204	0.051	0.152	0.060
500	0.038	0.030	0.021	0.001	0.032	0.008	0.017	0.008
$\hat{n} = n$								
30	0.069	0.252	<u>0.449</u>	0.294	0.081	0.286	0.136	0.289
50	0.186	0.321	<u>0.516</u>	0.459	0.146	0.402	0.229	0.394
100	0.365	0.469	0.581	<u>0.712</u>	0.310	0.571	0.341	0.587
300	0.802	0.837	0.835	<u>0.973</u>	0.796	0.947	0.802	0.929
500	0.931	0.928	0.927	<u>0.997</u>	0.968	0.992	0.935	0.984
$\hat{n} = n + 1$								
30	0.0	0.007	0.066	0.004	0.0	0.013	0.026	0.011
50	0.0	0.008	0.092	0.003	0.0	0.014	0.023	0.009
100	0.008	0.027	0.069	0.008	0.0	0.012	0.026	0.008
300	0.028	0.031	0.044	0.011	0.0	0.015	0.021	0.011
500	0.029	0.040	0.047	0.002	0.0	0.004	0.022	0.008

* Los procedimientos comparados son los siguientes: SVC_{Ω_1} corresponde al criterio de Bauer (2001), SVC_{Ω_2} es el mismo criterio con la matriz de ponderación propuesta $\Omega_2 = \mathbf{Z}_f \mathbf{\Pi}_{M_p, U_f}^\perp \mathbf{Z}_f'$, $NIDC$ sustituye además la función de penalización $C(T) = \frac{\log(T)}{T}$ por la función de penalización estimada $H(T, i)$. AIC , SBC y HQ son los criterios de información de Akaike (1976), Schwartz (1978) y Hannan y Quinn (1979), respectivamente. χ^2 es el contraste de Tiao y Tsay (1989) al 95% y $Moda$ es el valor más repetido de los \hat{n} estimados por los métodos anteriores (excepto SVC_{Ω_1} , que no se incluye en NID) en cada realización. En caso de empate se selecciona el mayor valor que más veces se repite. Se subrayan las frecuencias relativas que representan un mayor acierto para cada tamaño muestral.

Cuadro 4.1.8: Resultados de la estimación del orden del sistema para M7*.

$$M7(n = 2): (\mathbf{I} + \Phi B)\mathbf{Z}_t = (\mathbf{I} + \Theta B)\mathbf{a}_t, \quad \mathbf{a}_t \sim iidN(\mathbf{0}, \Sigma_a)$$

$$\Phi = \begin{pmatrix} -0.4 & 0 \\ 0 & -0.8 \end{pmatrix}; \Theta = \begin{pmatrix} -0.7 & 0.5 \\ -0.3 & -0.7 \end{pmatrix}; \Sigma_a = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$$

T	SVC_{Ω_1}	SVC_{Ω_2}	$NIDC$	AIC	SBC	HQ	χ^2	$Moda$
$\hat{n} = n - 1$								
30	0.002	0.355	0.287	0.519	0.536	0.581	0.540	0.614
50	0.068	0.491	0.358	0.517	0.702	0.632	0.699	0.665
100	0.432	0.323	0.196	0.302	0.701	0.501	0.554	0.410
300	0.007	0.002	0.002	0.005	0.249	0.078	0.054	0.012
500	0.0	0.0	0.0	0.0	0.035	0.0	0.005	0.001
$\hat{n} = n$								
30	0.0	0.311	<u>0.329</u>	0.270	0.080	0.203	0.097	0.201
50	0.001	0.354	<u>0.429</u>	0.379	0.137	0.275	0.171	0.275
100	0.199	0.611	<u>0.655</u>	0.617	0.290	0.470	0.414	0.555
300	<u>0.975</u>	0.957	0.943	0.885	0.751	0.909	0.825	0.970
500	<u>0.982</u>	0.972	0.958	0.896	0.965	0.992	0.851	0.984
$\hat{n} = n + 1$								
30	0.0	0.183	0.224	0.081	0.008	0.031	0.007	0.033
50	0.0	0.072	0.143	0.084	0.003	0.026	0.011	0.021
100	0.001	0.047	0.112	0.071	0.001	0.027	0.024	0.035
300	0.018	0.040	0.053	0.106	0.0	0.014	0.119	0.018
500	0.018	0.028	0.040	0.098	0.0	0.002	0.148	0.015

* Los procedimientos comparados son los siguientes: SVC_{Ω_1} corresponde al criterio de Bauer (2001), SVC_{Ω_2} es el mismo criterio con la matriz de ponderación propuesta $\Omega_2 = \mathbf{Z}_f \mathbf{\Pi}_{M_p, U_f}^\perp \mathbf{Z}_f'$, $NIDC$ sustituye además la función de penalización $C(T) = \frac{\log(T)}{T}$ por la función de penalización estimada $H(T, i)$. AIC , SBC y HQ son los criterios de información de Akaike (1976), Schwartz (1978) y Hannan y Quinn (1979), respectivamente. χ^2 es el contraste de Tiao y Tsay (1989) al 95% y $Moda$ es el valor más repetido de los \hat{n} estimados por los métodos anteriores (excepto SVC_{Ω_1} , que no se incluye en NID) en cada realización. En caso de empate se selecciona el mayor valor que más veces se repite. Se subrayan las frecuencias relativas que representan un mayor acierto para cada tamaño muestral.

Cuadro 4.1.9: Resultados de la estimación del orden del sistema para M8*.

$$M8(n = 3): (\mathbf{I} + \Phi B)\mathbf{Z}_t = (\mathbf{I} + \Theta B)\mathbf{a}_t, \quad \mathbf{a}_t \sim iidN(\mathbf{0}, \Sigma_a)$$

$$\Phi = \begin{pmatrix} -0.7 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & -0.4 & 0 \end{pmatrix}; \Theta = \begin{pmatrix} 0 & 1.1 & 0 \\ 0 & -0.6 & 0 \\ 0 & 0 & 0.5 \end{pmatrix}; \Sigma_a = \begin{pmatrix} 1 & -0.7 & 0.4 \\ -0.7 & 1 & 0 \\ 0.4 & 0 & 1 \end{pmatrix}$$

T	SVC_{Ω_1}	SVC_{Ω_2}	$NIDC$	AIC	SBC	HQ	χ^2	$Moda$
$\hat{n} = n - 1$								
50	0.0	0.232	0.274	0.362	0.644	0.604	0.344	0.535
100	0.0	0.485	0.484	0.215	0.678	0.399	0.298	0.382
300	0.164	0.026	0.020	0.0	0.043	0.004	0.003	0.006
500	0.001	0.0	0.0	0.0	0.0	0.0	0.0	0.0
$\hat{n} = n$								
50	0.0	0.353	<u>0.429</u>	0.345	0.149	0.276	0.330	0.334
100	0.0	0.448	0.449	0.638	0.317	0.590	<u>0.641</u>	0.594
300	0.836	0.965	<u>0.966</u>	0.954	0.957	0.962	0.943	0.991
500	0.997	0.995	0.992	0.968	<u>1.0</u>	0.999	0.945	1.0
$\hat{n} = n + 1$								
50	0.0	0.270	0.246	0.178	0.013	0.052	0.023	0.099
100	0.0	0.060	0.060	0.125	0.0	0.010	0.031	0.024
300	0.0	0.009	0.014	0.044	0.0	0.004	0.047	0.003
500	0.002	0.005	0.008	0.031	0.0	0.001	0.051	0.0

* Los procedimientos comparados son los siguientes: SVC_{Ω_1} corresponde al criterio de Bauer (2001), SVC_{Ω_2} es el mismo criterio con la matriz de ponderación propuesta $\Omega_2 = \mathbf{Z}_f \mathbf{\Pi}_{M_p, U_f}^\perp \mathbf{Z}'_f$, $NIDC$ sustituye además la función de penalización $C(T) = \frac{\log(T)}{T}$ por la función de penalización estimada $H(T, i)$. AIC , SBC y HQ son los criterios de información de Akaike (1976), Schwartz (1978) y Hannan y Quinn (1979), respectivamente. χ^2 es el contraste de Tiao y Tsay (1989) al 95% y $Moda$ es el valor más repetido de los \hat{n} estimados por los métodos anteriores (excepto SVC_{Ω_1} , que no se incluye en NID) en cada realización. En caso de empate se selecciona el mayor valor que más veces se repite. Se subrayan las frecuencias relativas que representan un mayor acierto para cada tamaño muestral.

Cuadro 4.1.10: Resultados de la estimación del orden del sistema para M9*.

$$M9(n = 3): (\mathbf{I} + \Phi B)\mathbf{Z}_t = (\mathbf{I} + \Theta B)\mathbf{a}_t, \quad \mathbf{a}_t \sim iidN(\mathbf{0}, \Sigma_a)$$

$$\Phi = \begin{pmatrix} -0.4 & -0.3 & 0.6 \\ 0 & -0.8 & -0.4 \\ -0.3 & 0 & 0 \end{pmatrix}; \Theta = \begin{pmatrix} -0.7 & 0 & 0 \\ -0.1 & -0.2 & 0 \\ 0.4 & -0.5 & 0.1 \end{pmatrix}; \Sigma_a = \begin{pmatrix} 1 & 0.5 & 0.4 \\ 0.5 & 1 & 0.7 \\ 0.4 & 0.7 & 1 \end{pmatrix}$$

T	SVC_{Ω_1}	SVC_{Ω_2}	$NIDC$	AIC	SBC	HQ	χ^2	$Moda$
$\hat{n} = n - 1$								
50	0.0	0.199	0.228	0.257	0.658	0.440	0.284	0.345
100	0.002	0.314	0.311	0.055	0.295	0.110	0.057	0.116
300	0.025	0.002	0.001	0.0	0.0	0.0	0.0	0.0
500	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
$\hat{n} = n$								
50	0.0	0.380	0.429	0.428	0.264	0.490	<u>0.602</u>	0.547
100	0.0	0.607	0.608	0.789	0.700	0.869	<u>0.908</u>	0.861
300	0.971	0.978	0.974	0.979	<u>0.999</u>	0.998	0.966	0.992
500	0.998	0.987	0.983	0.987	<u>1.0</u>	0.999	0.961	0.997
$\hat{n} = n + 1$								
50	0.0	0.286	0.270	0.226	0.026	0.068	0.047	0.101
100	0.0	0.071	0.073	0.123	0.005	0.021	0.033	0.023
300	0.004	0.020	0.025	0.021	0.001	0.002	0.034	0.008
500	0.002	0.013	0.017	0.013	0.0	0.001	0.038	0.003

* Los procedimientos comparados son los siguientes: SVC_{Ω_1} corresponde al criterio de Bauer (2001), SVC_{Ω_2} es el mismo criterio con la matriz de ponderación propuesta $\Omega_2 = \mathbf{Z}_f \mathbf{\Pi}_{M_p, U_f}^\perp \mathbf{Z}_f'$, $NIDC$ sustituye además la función de penalización $C(T) = \frac{\log(T)}{T}$ por la función de penalización estimada $H(T, i)$. AIC , SBC y HQ son los criterios de información de Akaike (1976), Schwartz (1978) y Hannan y Quinn (1979), respectivamente. χ^2 es el contraste de Tiao y Tsay (1989) al 95% y $Moda$ es el valor más repetido de los \hat{n} estimados por los métodos anteriores (excepto SVC_{Ω_1} , que no se incluye en NID) en cada realización. En caso de empate se selecciona el mayor valor que más veces se repite. Se subrayan las frecuencias relativas que representan un mayor acierto para cada tamaño muestral.

4.2. Índices de Kronecker

En la sección anterior se ha tratado el problema de identificar el orden del sistema, lo que permite estimar un modelo en EE con dimensión del vector de estado igual a \hat{n} . Este modelo, estadísticamente adecuado, es útil para: i) estimar la matriz de cointegración (ver Capítulo 3), ii) realizar una descomposición estructural (ver Casals et al., 2002), o iii) prever los valores futuros de las variables endógenas del sistema. Sin embargo, es posible que el analista desee avanzar en la modelización del proceso buscando una representación más escueta. Centrándose en los modelos ARMAX, cuando el estudio se realiza sobre una única serie, esto es tan sólo posible mediante restricciones de exclusión vía contrastes de hipótesis, ya que esta representación de un modelo univariante se corresponde con una estructura canónica caracterizada por un mínimo número de parámetros. En cambio, cuando el análisis se realiza sobre un proceso formado por un conjunto de series, en general esto no se cumple. Por lo tanto, es posible reducir el número de parámetros de un modelo VARMAX general, aplicando algunas transformaciones que conducen a una forma canónica. Así, para obtener una representación VARMAX única y parsimoniosa, algunos autores como Nsiri y Roy (1992), Lütkepohl (1993) o Li y Tsay (1998) usan un modelo VARMAX en forma echelon. Esta formulación canónica proporciona una representación única del proceso, con todos los parámetros identificados y que está condicionada a los llamados índices de Kronecker (IK). Estos índices se corresponden con el máximo orden de dinámica de cada ecuación del modelo. Una definición más detallada de una forma echelon y de los IK o de observabilidad puede encontrarse en Hannan y Deistler (1988).

4.2.1. Especificación de los índices de Kronecker

En este apartado se presenta, en un primer paso hacia la parsimonia, un método de búsqueda sobre los IK más adecuados, condicionados a \hat{n} . Su estimación proporcionará una representación canónica para el vector de series analizado. Esto se desarrolla en el algoritmo SUBEST1, que devuelve el juego de índices de observabilidad óptimo y la estimación del modelo VARMAX echelon asociado.

Para la estimación de modelos sin restricciones es obvio que el resultado de la identificación sólo depende de los datos, n e i . Sin embargo, para especificar un modelo VARMAX echelon es preciso, además, identificar los IK. Pese a la idea común de la búsqueda de un modelo adecuado y parsimonioso y la extensa utilización de los modelos VARMA, la especificación de los índices de observabilidad no ha tenido un gran desarrollo en la literatura econométrica. No obstante, algunos autores han propuesto distintos métodos para su estimación. En general, los procedimientos siguen los pasos siguientes (Lütkepohl y Poskitt, 1996),

1. Ajustar un modelo VAR(h) a los datos, donde h toma un valor elevado, generalmente $\log(T)$ o $T^{\frac{1}{2}}$,

$$\mathbf{z}_t = \sum_{i=1}^h \mathbf{\Pi}_{i,h} \mathbf{z}_{t-i} + \boldsymbol{\epsilon}_t$$

donde $\mathbf{\Pi}_{i,h}$ con $i = 1, 2, \dots, h$ son matrices autorregresivas. Obtener los residuos $\hat{\boldsymbol{\epsilon}}_t$ resultantes de la estimación de este modelo.

2. Estimar diferentes modelos VARMA echelon asociados a distintos IK por mínimos cuadrados ordinarios, utilizando los datos y los residuos obtenidos en la etapa anterior:

$$\mathbf{z}_t = \mathbf{A}_1 \mathbf{z}_{t-1} + \dots + \mathbf{A}_p \mathbf{z}_{t-p} + (\mathbf{A}_0 - \mathbf{I})(\hat{\boldsymbol{\epsilon}}_t - \mathbf{z}_t) + \mathbf{M}_1 \hat{\boldsymbol{\epsilon}}_{t-1} + \dots + \mathbf{M}_p \hat{\boldsymbol{\epsilon}}_{t-p} + \boldsymbol{\nu}_t$$

donde \mathbf{A}_j y \mathbf{M}_j con $j = 1, 2, \dots, p$ son matrices autorregresivas y media móvil, respectivamente y $(\mathbf{A}_0 - \mathbf{I})$ es una matriz triangular inferior con ceros en la diagonal principal. Determinar un juego de índices de observabilidad en base a un criterio de selección de modelos.

3. Estimar por máxima verosimilitud el modelo óptimo elegido en el paso anterior.
4. Realizar contrastes de hipótesis e imponer, si procede, restricciones de exclusión sobre los parámetros.
5. Llevar a cabo un análisis de residuos y volver, en caso necesario, al punto 2.

Este Capítulo se centra en los puntos 1 y 2 del esquema anterior. A los pasos siguientes se hará especial referencia en el Capítulo 5. En general, las diferencias en los métodos propuestos en la literatura radican en el punto 2. Si bien suele proponerse el uso de los criterios de selección AIC , HQ o SBC , la búsqueda del juego de índices de observabilidad denotado por $(\rho_1, \rho_2, \dots, \rho_m)$ donde m es la dimensión del vector de variables endógenas, suele ser distinta. Los procedimientos más destacados de la literatura son los siguientes:

1) *Método de búsqueda completa.* Este procedimiento se basa en la comparación de todos los modelos posibles imponiendo una cota superior (ρ_{max} o n_{max}), para restringir así el espacio paramétrico. Esta cota puede venir expresada por el orden máximo de cualquiera de los IK o por el máximo orden del sistema. De esta forma, la exploración se realiza sobre $0 \leq \rho_k \leq \rho_{max}$ en el primer caso, o bien sobre $0 \leq \sum_{k=1}^m \rho_k \leq n_{max}$ en el segundo, siendo ρ_k el índice de observabilidad asociado a

la serie k y n_{max} el máximo orden del proceso.

2) *Método de Hannan-Kavalieris*. Hannan y Kavalieris (1984) proponen un procedimiento de búsqueda dividido en dos etapas. En la primera, comparan los modelos generados a partir de juegos de índices donde $\rho_k = \rho_j \forall k, j$, seleccionando uno de ellos. Esto equivale a obtener el orden del sistema adecuado para un modelo VARMA lleno. En la segunda etapa, se determina individualmente el último IK mediante un criterio de información, fijando el resto en el valor obtenido en la etapa previa. Después, la búsqueda se realiza sobre el penúltimo IK, fijando el resto al valor de la primera etapa salvo el último, que se fija en el valor óptimo obtenido en la segunda etapa. Así, se consigue secuencialmente el juego de índices de observabilidad óptimo.

3) *Método de ecuación en ecuación*. Poskitt (1992) desarrolla un método basándose en la propiedad de que las restricciones de la j -ésima ecuación de una forma echelon vienen determinadas por los índices ρ_k cuando $\rho_k \leq \rho_j$, siendo k y j ecuaciones del sistema con $k \neq j$. Por tanto, el autor propone realizar la especificación de ecuación en ecuación.

4) *Método secuencial de contrastes χ^2* . En este procedimiento se utiliza el estadístico (4.1.1) presentado en la sección 4.1 que además de permitir la realización de contrastes sobre la dimensión del sistema, pueden utilizarse de forma secuencial para determinar los índices de observabilidad de un proceso multivariante (ver Tsay, 1989b; Casals, 1997). En este caso, para implementar el procedimiento no son necesarios los pasos 1 y 2 descritos anteriormente.

Los tres primeros métodos estiman consistentemente los IK bajo condiciones no muy restrictivas sobre los sistemas, los criterios de selección de modelos y el orden h del VAR del paso 1. No obstante, presentan rendimientos distintos en muestras finitas y su coste computacional también es diferente. El cuarto método también es consistente y su carga computacional es reducida, pero su interpretación y mecanización es más compleja.

Más recientemente, Li y Tsay (1998) proponen un procedimiento Bayesiano para identificar y estimar conjuntamente los IK y los parámetros de un sistema. Sin embargo, aunque los resultados teóricos son positivos, los propios autores reconocen la necesidad de avances en la investigación para su uso en la práctica, ya que su carga computacional es, en muchos casos, excesiva.

En el marco de los métodos de subespacios, Casals (1997) desarrolla un nuevo algoritmo que relaciona la forma canónica de Luenberger (1967) con una repre-

sentación echelon de un modelo VARMAX. Partiendo de esta relación, aquí se propone una variación del método de búsqueda completa que aprovecha, por un lado el bajo coste computacional que conlleva la estimación mediante métodos de subespacios y por otro, la relación existente entre el orden del proceso y los IK de forma que $\sum_{k=1}^m \rho_k = n$. Efectivamente, en la sección anterior se resolvió el problema de la determinación del índice de McMillan, con lo que se abrían dos nuevas posibilidades: i) obtener una representación en EE, o una forma VARMAX equivalente, con una dimensión del vector de estado igual a \hat{n} o, ii) progresar en el análisis buscando un modelo más escueto. Si se toma esta segunda opción, la búsqueda de los índices de observabilidad debe restringirse a todas aquellas combinaciones que sumen el orden del proceso estimado previamente.

Por lo tanto, la especificación de los IK se reduce a resolver el siguiente problema de optimización con una restricción de igualdad:

$$\begin{aligned} \min_{\{\rho_1, \rho_2, \dots, \rho_m\}} \quad & \log |\hat{\Sigma}(\boldsymbol{\rho}_1, \boldsymbol{\rho}_2, \dots, \boldsymbol{\rho}_m)| + C(T)d(\rho_1, \rho_2, \dots, \rho_m) \\ \text{s.a.} \quad & \sum_{k=1}^m \rho_k = \hat{n} \end{aligned} \quad (4.2.1)$$

donde, $\hat{\Sigma}(\boldsymbol{\rho}_1, \boldsymbol{\rho}_2, \dots, \boldsymbol{\rho}_m)$ es el estimador de la matriz de covarianzas del error y $d(\rho_1, \rho_2, \dots, \rho_m)$ es el número de parámetros de la forma echelon implicada por la estructura de índices de Kronecker $(\rho_1, \rho_2, \dots, \rho_m)$. De nuevo, $C(T)$ es una función de penalización seleccionada por el usuario entre, $\frac{2}{T}$, $\frac{\log(T)}{T}$ y $\frac{2\log(\log(T))}{T}$, dependiendo del uso del criterio de selección AIC , SBC o HQ respectivamente.

Aunque este método no siempre presenta ventajas computacionales frente a procedimientos alternativos (esto dependerá principalmente del orden del proceso estimado y de la dimensión del sistema), sí reduce sensiblemente el coste respecto del método de búsqueda completa. Por otra parte, la metodología presentada pretende, separando la determinación del índice de McMillan y de los índices de observabilidad: 1) ofrecer la posibilidad al usuario de reducir los costes del análisis en función de sus objetivos y 2) disminuir el riesgo de error en la modelización. Puesto que, respecto al segundo punto, parece sensato considerar la mayor sencillez de la decisión sobre n implica una ganancia en la especificación de los IK condicionada a \hat{n} , respecto de la no condicionada.

En resumen, la especificación de los IK conduce a una disminución considerable del espacio paramétrico debida a la aplicación de transformaciones canónicas sobre el modelo identificado. Dada la velocidad de cálculo y la estabilidad de los métodos

de subespacios, la posibilidad de estimar modelos VARMAX echelon con restricciones de exclusión impuestas por el juego de índices de observabilidad, sugiere su utilización en el proceso de especificación de un modelo econométrico parsimonioso.

Una vez que el modelo está representado con una parametrización mínima, puede ser deseable realizar inferencia sobre los parámetros estimados. Por ello, en el algoritmo SUBEST1 que especifica los IK y devuelve la estimación de la forma VARMAX echelon correspondiente, se ofrecen también las desviaciones típicas de los parámetros estimados (ver Casals, 1997).

4.2.2. Análisis empírico de especificación automática

En este apartado se evalúa el método de estimación del orden del sistema y de los índices de observabilidad conjuntamente con el procedimiento de estimación mediante subespacios propuesto en Casals (1997). Todo el proceso está implementado en el algoritmo SUBEST1. El ejercicio contrasta la capacidad de esta metodología para la especificación automática de modelos en forma VARMA echelon. Para ello, se realizan 1000 repeticiones del proceso M10, correspondiente a un modelo VARMA(2,2) restringido:

$$(\mathbf{I} + \Phi_1 B + \Phi_2 B^2) \mathbf{Z}_t = (\mathbf{I} + \Theta_1 B + \Theta_2 B^2) \mathbf{a}_t, \quad \mathbf{a}_t \sim iidN(\mathbf{0}, \Sigma_a) \quad (4.2.2)$$

con,

$$\Phi_1 = \begin{pmatrix} -0.70 & 0 & 0 \\ 0.48 & -0.50 & -0.90 \\ -0.02 & 0.30 & -0.20 \end{pmatrix}; \Phi_2 = \begin{pmatrix} 0.30 & -0.20 & 0.50 \\ -0.12 & 0.08 & -0.20 \\ 0.18 & -0.12 & 0.30 \end{pmatrix};$$

$$\Theta_1 = \begin{pmatrix} -0.20 & 0.40 & 0.70 \\ 0.68 & -0.46 & -0.68 \\ 0.18 & 1.24 & -0.38 \end{pmatrix}; \Theta_2 = \begin{pmatrix} 0.30 & 0.50 & -0.80 \\ -0.12 & -0.20 & 0.32 \\ 0.18 & 0.30 & -0.48 \end{pmatrix};$$

$$\Sigma_a = \begin{pmatrix} 1 & - & - \\ 0.5 & 1 & - \\ 0.4 & 0.7 & 1 \end{pmatrix}$$

donde las restricciones implican que los coeficientes correspondientes a las posiciones (1,2) y (1,3) de la matriz Φ_1 son nulos. Este modelo puede escribirse en forma VARMA echelon equivalente como:

$$(\mathbf{F}_0 + \mathbf{F}_1 B + \mathbf{F}_2 B^2) \mathbf{Z}_t = (\mathbf{Q}_0 + \mathbf{Q}_1 B + \mathbf{Q}_2 B^2) \mathbf{a}_t \quad (4.2.3)$$

con,

$$\mathbf{F}_0 = \mathbf{Q}_0 = \begin{pmatrix} 1.0 & 0 & 0 \\ 0.4 & 1.0 & 0 \\ -0.6 & 0 & 1.0 \end{pmatrix}; \mathbf{F}_1 = \begin{pmatrix} -0.7 & 0 & 0 \\ 0.2 & -0.5 & -0.9 \\ 0.4 & 0.3 & -0.2 \end{pmatrix};$$

$$\mathbf{F}_2 = \begin{pmatrix} 0.3 & -0.2 & 0.5 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}; \mathbf{Q}_1 = \begin{pmatrix} -0.2 & 0.4 & 0.7 \\ 0.6 & -0.3 & -0.4 \\ 0.3 & 1.0 & -0.8 \end{pmatrix}; \mathbf{Q}_2 = \begin{pmatrix} 0.3 & 0.5 & -0.8 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix};$$

donde puede observarse que premultiplicando por \mathbf{F}_0^{-1} en ambos lados de (4.2.3), se obtiene directamente (4.2.2). Los índices de observabilidad de esta estructura son (2,1,1) que corresponden al máximo orden de dinámica de cada uno de los procesos univariantes. Obsérvese cómo la representación del modelo en forma VARMA echelon reduce el número de parámetros, pasando de 34 coeficientes en la especificación (4.2.2) a 24 en la formulación (4.2.3), excluyendo los de la matriz de covarianzas del error.

Los tamaños muestrales estudiados son $T = 50, 100$, y 300 , después de eliminar las 50 primeras observaciones en cada repetición. En el experimento se procede de la forma siguiente: 1) se usa el algoritmo NID, como se explicó en la sección anterior, que selecciona automáticamente el orden del proceso y, 2) se utiliza el algoritmo SUBEST1, que restringe la búsqueda de los IK a aquéllos que suman el \hat{n} estimado en la etapa anterior. El Cuadro 4.2.1 ofrece la frecuencia relativa de estimación de los IK con el criterio *SBC* (por defecto en SUBEST1) en función del índice de McMillan estimado por NID.

Los resultados muestran una alta capacidad del algoritmo para identificar la estructura de IK en el modelo estudiado. En la primera fase, se estima con una probabilidad cercana a la unidad que el orden del sistema es igual a 4. En algunos casos excepcionales se encuentra $\hat{n} > 4$, pero esto no supera el 1% de las veces en ninguna de las muestras estudiadas. En este caso, cuando se estima $\hat{n} > 4$, los IK especificados no permiten llegar exactamente al modelo (4.2.3) únicamente mediante restricciones de parámetros ya que implican estructuras de ceros distintas en las matrices \mathbf{F}_1 para el juego de índices (2,2,1) y en $\mathbf{F}_1, \mathbf{F}_2$ y \mathbf{Q}_1 para el juego de índices (3,2,0). Sin embargo, la estructura de IK (2,2,1) puede, mediante restricciones, llevar a un modelo muy similar con tan sólo un parámetro menos. Por otro lado, la falta de significación de ciertos parámetros, la existencia de cancelaciones

entre estructuras AR y MA y la elevada magnitud de algunos coeficientes estimados son claros indicios de sobreparametrización. En este tipo de situaciones, se recomienda una nueva especificación de los índices de observabilidad restringiendo la búsqueda a los que suman $\hat{n} - 1$, junto con un especial cuidado en el análisis de residuos.

El Cuadro 4.2.2 presenta el valor medio de los parámetros estimados de cada una de las matrices del sistema cuando el juego de índices se ha seleccionado correctamente. La tabla muestra unos resultados aceptablemente buenos, aún con muestras cortas y revela la consistencia de los estimadores.

Cuadro 4.2.1: Estimación de los IK para el proceso M10 con $n = 4^$.*

\hat{n}	4						5	
$\hat{\rho}_i$	(1,1,2)	(1,2,1)	(1,3,0)	(2,1,1)	(2,2,0)	(3,1,0)	(2,2,1)	(3,2,0)
50	0.008	0.061	0.005	<u>0.761</u>	0.132	0.023	0.008	0.002
T 100	0.002	0.005	0.0	<u>0.935</u>	0.042	0.008	0.008	0.0
300	0.0	0.0	0.0	<u>0.990</u>	0.002	0.0	0.008	0.0

* $i = 1, 2, 3$. \hat{n} es la estimación del orden del proceso obtenida como la moda de los criterios empleados en NID y $\hat{\rho}_i$ es la especificación del juego de índices de Kronecker a partir de \hat{n} y del criterio de información de Schwartz (1978), *SBC*. Se subrayan las frecuencias relativas que representan un mayor acierto para cada tamaño muestral.

Cuadro 4.2.2: Valores medios de las estimaciones obtenidas a partir de la especificación de los IK^*

	$F_0 = Q_0$	F_1	F_2	Q_1	Q_2
	$\begin{pmatrix} 1.0 & 0 & 0 \\ .4 & 1.0 & 0 \\ -.6 & 0 & 1.0 \end{pmatrix}$	$\begin{pmatrix} -.7 & 0 & 0 \\ .2 & -.5 & -.9 \\ .4 & .3 & -.2 \end{pmatrix}$	$\begin{pmatrix} .3 & -.2 & .5 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$	$\begin{pmatrix} -.2 & .4 & .7 \\ .6 & -.3 & -.4 \\ .3 & 1.0 & -.8 \end{pmatrix}$	$\begin{pmatrix} .3 & .5 & -.8 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$
T	$\hat{F}_0 = \hat{Q}_0$	\hat{F}_1	\hat{F}_2	\hat{Q}_1	\hat{Q}_2
50	$\begin{pmatrix} 1.0 & 0 & 0 \\ .48 & 1.0 & 0 \\ -.66 & 0 & 1.0 \end{pmatrix}$	$\begin{pmatrix} -.71 & 0 & 0 \\ .17 & -.41 & -.91 \\ .28 & .38 & -.02 \end{pmatrix}$	$\begin{pmatrix} .16 & -.04 & .65 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$	$\begin{pmatrix} -.28 & .40 & .68 \\ .54 & -.20 & -.32 \\ .17 & .84 & -.44 \end{pmatrix}$	$\begin{pmatrix} .16 & .43 & -.42 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$
100	$\begin{pmatrix} 1.0 & 0 & 0 \\ .46 & 1.0 & 0 \\ -.58 & 0 & 1.0 \end{pmatrix}$	$\begin{pmatrix} -.65 & 0 & 0 \\ .15 & -.41 & -.87 \\ .29 & .39 & -.06 \end{pmatrix}$	$\begin{pmatrix} .21 & -.08 & .56 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$	$\begin{pmatrix} -.22 & .39 & .70 \\ .50 & -.25 & -.28 \\ .22 & .81 & -.45 \end{pmatrix}$	$\begin{pmatrix} .24 & .41 & -.45 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$
300	$\begin{pmatrix} 1.0 & 0 & 0 \\ .42 & 1.0 & 0 \\ -.62 & 0 & 1.0 \end{pmatrix}$	$\begin{pmatrix} -.68 & 0 & 0 \\ .17 & -.41 & -.87 \\ .39 & .32 & -.19 \end{pmatrix}$	$\begin{pmatrix} .28 & -.16 & .50 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$	$\begin{pmatrix} -.22 & .40 & .70 \\ .51 & -.25 & -.32 \\ .29 & .89 & -.56 \end{pmatrix}$	$\begin{pmatrix} .30 & .41 & -.61 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$

* F_0, F_1, F_2, Q_0, Q_1 y Q_2 son las matrices autorregresivas y media móvil del modelo (4.2.3). Se presentan los valores medios de las estimaciones cuyas realizaciones han sido identificadas con un juego de índices de Kronecker (2,1,1) mediante el algoritmo SUBEST1.

4.3. Conclusiones del Capítulo

El algoritmo de identificación derivado, denotado por NID, además del análisis de raíces unitarias y cointegración, ofrece dos nuevas posibilidades: 1) presenta los resultados de todos los criterios, a priori y a posteriori, expuestos en la sección 4.1 para que el usuario escoja, en función de su uso, el orden del proceso que crea más conveniente y 2) especifica automáticamente el tamaño del sistema o índice de McMillan (\hat{n}) como la moda de los criterios anteriores, reduciendo el riesgo de error cometido al utilizar un único criterio. Llegado este punto, NID devuelve la estimación de un modelo en Espacio de los Estados en forma de innovaciones con un vector de estado de dimensión \hat{n} , o su representación VARMAX equivalente.

Si se desea buscar un modelo más parsimonioso, el algoritmo denominado SUBEST1 permite estimar los índices de observabilidad y las matrices del sistema, a partir de los datos y del índice de McMillan establecido anteriormente. El cálculo de los índices de Kronecker (IK) mediante el método propuesto ofrece buenos resultados con el modelo VARMA trivariante analizado, aunque puede generar un sensible coste computacional en sistemas de dimensión grande y con un orden de dinámica también elevado (ya que ambos aumentan el número de combinaciones de índices). Los resultados muestran que la sobreestimación del orden del sistema arrastra el error al cálculo de los IK. No obstante, esto genera claros indicios de sobreparametrización en la estimación del modelo que son relativamente fáciles de detectar. A su vez, la estimación de las matrices de coeficientes forzando los índices de observabilidad produce muy buenas preestimaciones de los parámetros para el usuario que desee usar un procedimiento de máxima verosimilitud.

Para depurar el modelo con un menor coste computacional que el que conlleva la estimación máximo-verosímil, el paso siguiente en la metodología propuesta ofrece un algoritmo de estimación iterativo mediante subespacios, denotado por SUBEST2, que proporciona: i) mayor precisión y eficiencia en el cálculo de los parámetros del modelo que SUBEST1 y, ii) la posibilidad de incorporar restricciones lineales individuales sobre los coeficientes.

Capítulo 5

Estimación Restringida Basada en Métodos de Subespacios

En el Capítulo anterior se ha mostrado la utilidad de los métodos de subespacios tanto para identificar la dimensión de un sistema y sus índices de observabilidad, como para estimar un modelo estadísticamente adecuado aunque no siempre escueto. No obstante, estos procedimientos presentan ciertos inconvenientes: 1) se produce un pequeño sesgo negativo en la estimación de los parámetros media móvil, que se hace más importante en modelos estacionales (ver Casals, 1997) y, 2) aún después de la selección de los índices de Kronecker, que conducen a un modelo canónico, es posible que los datos requieran imponer restricciones sobre ciertos coeficientes.

Llegados a este punto, es posible usar un algoritmo de Máxima Verosimilitud (MV) a partir de las condiciones iniciales obtenidas en el paso previo. Sin embargo, hay que tener en cuenta que la ganancia de precisión y eficiencia estadística que esto conlleva, vendrá acompañada de un aumento de la inestabilidad, la complejidad y el coste computacional. En algunos casos, como en los modelos multivariantes complejos o en la modelización de datos de alta frecuencia generados por los mercados financieros, los costes que implica la estimación por MV pueden ser excesivos.

Por ello, en este Capítulo se propone un algoritmo iterativo de estimación basado en métodos de subespacios para estimar modelos expresados en Espacio de los Estados (EE). Este algoritmo es útil para depurar el modelo obtenido en la fase anterior y generar condiciones iniciales adecuadas para la estimación máximo-verosímil o, alternativamente, para obtener estimaciones finales cuando la MV es considerada inadecuada o demasiado costosa. Además, el uso de este método implica un alto grado de independencia en cuanto a formulaciones, ya que puede aplicarse a cualquier modelo lineal con coeficientes constantes que pueda represen-

tarse en EE.

El Capítulo se estructura en tres apartados. En la Sección 5.1 se deriva el nuevo algoritmo partiendo de la notación y los resultados introducidos en el Capítulo 2. La Sección 5.2 analiza su comportamiento en muestras finitas mediante ejercicios de simulación y la Sección 5.3 resume las aportaciones del Capítulo.

5.1. Algoritmo de Estimación

En este apartado se presenta un procedimiento iterativo de estimación mediante subespacios que resulta útil para conseguir buenas condiciones iniciales para la estimación por MV o para obtener estimaciones finales en muestras con determinadas propiedades. SUBEST2 es un método de estimación que utiliza herramientas eficientes del algebra lineal comúnmente usadas en la literatura de subespacios. El algoritmo funciona únicamente con modelos en EE en forma de innovaciones. No obstante esto no implica ninguna restricción, ya que cuando el modelo tiene una representación en EE más general, se transforma la estructura del ruido a una forma de innovaciones equivalente en cada iteración, utilizando el proceso propuesto en Casals et al. (1999). Este hecho hace que SUBEST2, además de permitir restricciones adicionales sobre los coeficientes, ofrezca en muestras finitas estimaciones de los parámetros: i) más eficientes que las generadas en el paso metodológico previo, ii) generalmente más precisas, y iii) sensiblemente más rápidas que las obtenidas por MV.

Como se vio en el Capítulo 2, un modelo en EE en forma de innovaciones puede formularse mediante subespacios como:

$$\mathbf{Z}_f = \mathbf{O}_i \mathbf{X}_f + \mathbf{H}_i^d \mathbf{U}_f + \mathbf{H}_i^s \boldsymbol{\epsilon}_f \quad (5.1.1)$$

donde \mathbf{O}_i es la matriz de observabilidad y tanto \mathbf{H}_i^d como \mathbf{H}_i^s son matrices Toeplitz que albergan parámetros del modelo original. Proyectando esta ecuación en el espacio de filas de la matriz \mathbf{U} (información pasada y futura) y el espacio de filas de la matriz \mathbf{Z}_p (información pasada), se obtiene:

$$\mathbf{Z}_f \boldsymbol{\Pi}_{\mathbf{U}, \mathbf{Z}_p} = \mathbf{O}_i \mathbf{X}_f \boldsymbol{\Pi}_{\mathbf{U}, \mathbf{Z}_p} + \mathbf{H}_i^d \mathbf{U}_f \boldsymbol{\Pi}_{\mathbf{U}, \mathbf{Z}_p} \quad (5.1.2)$$

donde $\boldsymbol{\Pi}_{\mathbf{U}, \mathbf{Z}_p}$ es el proyector ortogonal de \mathbf{U} y \mathbf{Z}_p , y donde se asume que las perturbaciones futuras no tienen relación ni con el pasado del output ni con el input observable, es decir:

$$\boldsymbol{\epsilon}_f \boldsymbol{\Pi}_{\mathbf{U}, \mathbf{Z}_p} = \mathbf{0} \quad (5.1.3)$$

Así, la ecuación (5.1.2), permite calcular los estados en función de la información del output y del input observable, que está relacionada con el pasado del output y todo el input:

$$\mathbf{X}_f \mathbf{\Pi}_{U, Z_p} = \mathbf{O}_i^+ (\mathbf{Z}_f \mathbf{\Pi}_{U, Z_p} - \mathbf{H}_i^d \mathbf{U}_f \mathbf{\Pi}_{U, Z_p}) \quad (5.1.4)$$

donde \mathbf{O}_i^+ es la pseudo inversa de la matriz de observabilidad. Restando la expresión (5.1.2) a (5.1.1) puede definirse $\tilde{\mathbf{Z}}_f$ como la matriz residual resultante de ajustar la información contenida en los estados y en el futuro del input observable a la información futura del output:

$$\tilde{\mathbf{Z}}_f = \mathbf{Z}_f - \hat{\mathbf{O}}_i \hat{\mathbf{X}}_f \mathbf{\Pi}_{U, Z_p} - \hat{\mathbf{H}}_i^d \mathbf{U}_f \mathbf{\Pi}_{U, Z_p} \quad (5.1.5)$$

La matriz de residuos $\tilde{\mathbf{Z}}_f$ tiene una estructura particular con errores de previsión (EP) un periodo hacia adelante en la primera fila, EP dos periodos hacia adelante en la segunda fila y así sucesivamente. Con esto, la función de verosimilitud gaussiana, generada a partir de los *im* modelos de regresión que supone la ecuación (5.1.5), puede escribirse como:

$$\log l(\theta) = -\frac{im}{2} \log(2\pi) - \frac{i}{2} \log(|\mathbf{\Sigma}|) - \frac{1}{2} \text{tr}(\tilde{\mathbf{Z}}_f' \mathbf{\Sigma}^{-1} \tilde{\mathbf{Z}}_f) \quad (5.1.6)$$

donde $\text{tr}(\mathbf{A})$ representa la traza de la matriz \mathbf{A} , $\tilde{\mathbf{Z}}_f$ es la matriz de EP y $\mathbf{\Sigma}$ es su matriz de covarianzas. En este caso, la matriz de EP presenta autocorrelación y $\mathbf{\Sigma}$ puede formularse analíticamente como:

$$\mathbf{\Sigma} = \mathbf{O}_i \mathbf{P}_i \mathbf{O}_i' + \mathbf{H}_i^s (\mathbf{I}_i \otimes \mathbf{Q}) \mathbf{H}_i^{s'} \quad (5.1.7)$$

donde \otimes es el producto de Kronecker, \mathbf{P}_i es la matriz de covarianzas de los EP de la estimación de los estados, e \mathbf{I}_i es una matriz identidad de dimensión i . Pueden observarse dos componentes en la expresión (5.1.7): el primer sumando de la parte derecha de la ecuación, hace referencia a la varianza de los EP cometidos en la estimación de los estados, mientras que el segundo sumando se corresponde a la varianza de los EP cometidos en la estimación del output futuro, condicionada a que los estados están correctamente calculados.

Dada la estructura de $\tilde{\mathbf{Z}}_f$, la matriz de covarianzas de los EP puede definirse como:

$$\mathbf{\Sigma} = \begin{cases} \Sigma_{jk} \equiv \text{matriz de covarianzas de } EP_j & \text{si } j = k \\ \Sigma_{jk} \equiv \text{cov}(EP_j, EP_k) & \text{si } j \neq k \end{cases} \quad j, k = 1, 2, \dots, i \quad (5.1.8)$$

donde EP_l son errores de previsión l periodos hacia adelante. En efecto, la diagonal principal de $\mathbf{\Sigma}$ ofrece las matrices de covarianzas de los EP desde uno hasta i periodos hacia adelante, mientras que las matrices fuera de la diagonal principal hacen

referencia a las covarianzas entre los errores de previsión a diferentes horizontes temporales.

Por último, para definir completamente el algoritmo es necesaria, en cada iteración, una expresión para la matriz de covarianzas de los EP de los estados, denotada por \mathbf{P}_i . Para ello, se utilizan las ecuaciones de las covarianzas del Filtro de Kalman. En efecto, el Filtro de Kalman, en forma subespacial, permite obtener la estimación lineal de mínima varianza de los estados y su matriz de covarianzas de los EP recursivamente (Anderson y Moore, 1979), propagando i veces las ecuaciones:

$$\hat{\mathbf{Z}}_{i:i} = \mathbf{H}\hat{\mathbf{X}}_{i|i-1} + \mathbf{D}\mathbf{U}_{i:i} \quad (5.1.9)$$

$$\hat{\mathbf{X}}_{i+1|i} = \Phi\hat{\mathbf{X}}_{i|i-1} + \Gamma\mathbf{U}_{i:i} + \mathbf{K}_i(\mathbf{Z}_{i:i} - \hat{\mathbf{Z}}_{i:i}) \quad (5.1.10)$$

$$\mathbf{K}_i = (\Phi\mathbf{P}_{i|i-1}\mathbf{H}' + \mathbf{E}\mathbf{Q})\mathbf{B}_i' \quad (5.1.11)$$

$$\mathbf{P}_{i+1|i} = \Phi\mathbf{P}_{i|i-1}\Phi' + \mathbf{E}\mathbf{Q}\mathbf{E}' - \mathbf{K}_i\mathbf{B}_i\mathbf{K}_i' \quad (5.1.12)$$

$$\mathbf{B}_i = \mathbf{H}\mathbf{P}_{i|i-1}\mathbf{H}' + \mathbf{Q} \quad (5.1.13)$$

donde $\hat{\mathbf{X}}_{i|i-1}$ es una secuencia de vectores de estado estimados (estados Kalman), \mathbf{B}_i se corresponde con la varianza del error de predicción de $\mathbf{Z}_{i:i}$ a partir de los valores pasados del output y de la información del input, y la matriz \mathbf{K}_i es la ganancia del filtro de Kalman. En el procedimiento propuesto, tan sólo es necesario propagar las ecuaciones (5.1.11-5.1.13), que son las que generan una estimación consistente de \mathbf{P}_i . La Figura 5.1.1 representa gráficamente el algoritmo propuesto.

Estacionalidad del tipo multiplicativo:

En los modelos que presentan estacionalidad, en general, el vector de estado debe aumentar sensiblemente para captar la información estacional, lo que afecta al coste computacional de los algoritmos que usan esta representación. Por otro lado, para estimar correctamente la estructura estacional es conveniente ampliar significativamente la matriz Block Hankel donde están organizados los datos y esto produce un aumento de la carga computacional.

Para evitar este inconveniente, en estos casos se decide llevar a cabo una estimación en dos etapas. Se procede como en el apartado 4.1.2 de selección del orden estacional salvo que, en este caso, se estima utilizando el nuevo algoritmo SUBEST2. Esto es, en caso de estacionalidad de tipo multiplicativo el procedimiento opera de la manera siguiente:

1. Se estima la estructura estacional con SUBEST2 a partir de la matriz Block Hankel Estacional (4.1.9).

2. Se generan los residuos resultantes de ajustar el modelo estacional anterior a la serie analizada.
3. Se estima la estructura regular y la varianza con SUBEST2 a partir de la matriz Block Hankel (2.3.1) de los residuos obtenidos en el punto 2.

Este tratamiento por partes de la estacionalidad permite mejorar la velocidad del procedimiento y la precisión de las estimaciones tanto en muestras mensuales como trimestrales.

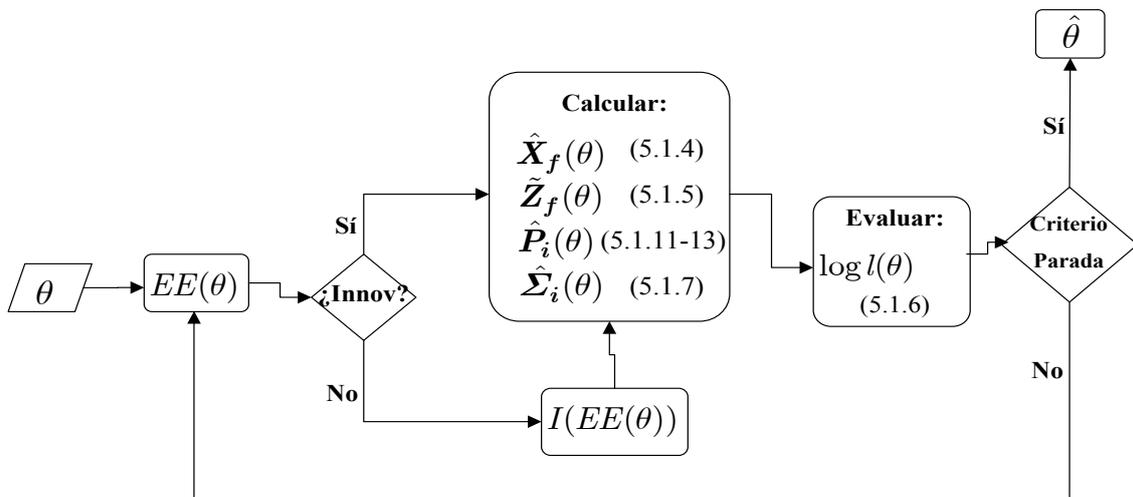


Figura 5.1.1: Algoritmo SUBEST2. θ son los parámetros del modelo original, $EE(\theta)$ es la representación del modelo original en EE, e $I(EE(\theta))$ es la transformación a un modelo en forma de innovaciones de la representación en EE más general. Cuando el modelo puede directamente representarse como una forma de innovaciones, se calculan las matrices de la función objetivo (que dependen de los parámetros del modelo original) y se evalúa. En caso contrario, en cada iteración debe realizarse una transformación del modelo general a una forma de innovaciones sobre la que se siguen los mismos pasos que en el caso anterior.

5.2. Ejercicios de Simulación

En esta Sección se analiza el comportamiento en muestras finitas del algoritmo previamente descrito, utilizando datos simulados a partir de una amplia variedad de modelos. Los resultados se estudian en términos de precisión y coste computacional, comparándolos con los obtenidos con un algoritmo de Máxima Verosimilitud

Exacta (MVE) sobre una representación en EE. La estimación máximo-verosímil se inicializa en cada caso con los verdaderos valores de los parámetros del modelo.

Los Cuadros 5.2.1-5.2.11 resumen los principales resultados. En todos ellos, se han realizado 1000 repeticiones de tamaños muestrales $T = 50$ y $T = 300$, después de descartar las 50 primeras realizaciones. Las simulaciones de los Cuadros 5.2.1-5.2.9 hacen referencia a modelos homoscedásticos. Los Cuadros 5.2.10 y 5.2.11 muestran los resultados obtenidos para dos modelos de heteroscedasticidad condicional habituales en la literatura financiera. En estos casos, se incrementa el tamaño muestral hasta $T = 500$ y $T = 3000$, con el fin de representar series temporales de alta frecuencia que típicamente usan este tipo de modelos.

Los Cuadros 5.2.1-5.2.5 muestran los resultados obtenidos por tres modelos univariantes no estacionales: AR(2) gaussiano, ARMA(2,1) gaussiano y ARMA(2,1) con errores t -Student, y un modelo bivalente VARMA(2,1), respectivamente. Los parámetros de la estructura autorregresiva están seleccionados, en todos los casos, de tal forma que sus raíces sean complejas y se mantengan fuera del círculo unidad. Esto elude situaciones de mal condicionamiento de la función objetivo evitando la cancelación de raíces autorregresivas y media móvil. Como cabía esperar:

1. Las estimaciones por MVE ofrecen una clara ventaja en precisión para modelos gaussianos en los componentes de medias móviles cuando $T = 50$. Sin embargo, cuando el modelo es puramente autorregresivo o tiene errores no gaussianos, la raíz cuadrada del error cuadrático medio (RMSE) de las estimaciones mediante SUBEST2 es comparable a la máximo-verosímil.
2. El coste computacional de MVE es 1.5/2 y 3/4 veces el coste de SUBEST2 para $T = 50$ y $T = 300$, respectivamente.

Los Cuadros 5.2.6 y 5.2.7 presentan los resultados correspondientes a un proceso gaussiano $\text{ARMA}(1, 1) \times (0, 1)_s$ para $s = 4$ y $s = 12$, respectivamente. Obsérvese que mientras que MVE mantiene la ventaja en precisión sobre los componentes MA cuando la muestra es relativamente larga o $s = 12$, esto no ocurre cuando $s = 4$ y $T = 50$. Además, cuando $T = 300$ la precisión de ambos algoritmos es muy similar en términos de RMSE. Por otro lado, la ventaja computacional de SUBEST2 se mantiene entre 2/2.5 veces para modelos trimestrales y entre 2/4 veces para modelos mensuales. Esto es debido a la mayor carga computacional que representa la dimensión del vector de estado cuando $s = 12$ frente a $s = 4$ en el algoritmo de MVE.

Los Cuadros 5.2.8 y 5.2.9 muestran los resultados obtenidos con dos formulaciones específicas de la literatura de EE: un modelo estructural con ratio señal-ruido

bajo y un modelo AR(2) con errores de observación.

SUBEST2, al igual que MVE, permite la estimación de este tipo de modelos ya que su función objetivo depende de la matriz de covarianzas de los errores, es decir, no está concentrada. Ambos algoritmos tienen en cuenta la restricción de independencia entre los errores en los estados y los de la ecuación de observación. En este caso, el coste computacional es muy similar, levemente superior con $T = 50$ y apreciablemente inferior con $T = 300$, en SUBEST2 que en MVE. Esto es debido, en parte, al escaso número de parámetros estimados. Los resultados del Cuadro 5.2.9 son muy próximos a los del caso anterior en coste computacional, siendo más lento SUBEST2 en la muestra corta pero ocurre lo contrario cuando la muestra es más larga. Obsérvese que en los modelos que no tienen una representación directa en forma de innovaciones, SUBEST2 pierde cierta ventaja computacional, debido a la transformación que hace a esta estructura en cada iteración. En cuanto a la precisión de las estimaciones, es semejante en ambos métodos.

Finalmente, los Cuadros 5.2.10 y 5.2.11 presentan los resultados obtenidos con dos modelos financieros habituales para datos de alta frecuencia: el modelo autorregresivo de volatilidad estocástica, ARSV (Harvey et al., 1994) y el modelo autorregresivo generalizado de heteroscedasticidad condicional, GARCH (Bollerslev, 1986).

Como era de esperar, los resultados del Cuadro 5.2.10 son similares a los del modelo AR con errores de observación, debido a la proximidad de ambas formulaciones en EE. Sin embargo, dado el incremento del tamaño muestral, el coste computacional de MVE se vuelve 3 y 12 veces el coste de SUBEST2, para $T = 500$ y $T = 3000$, respectivamente.

El Cuadro 5.2.11 revela que el coste computacional crece en el modelo GARCH más que en el anterior. Específicamente, aplicando MVE, éste es más de 16 veces superior al de SUBEST2, para $T = 500$. Cuando $T = 3000$, que puede ser un tamaño habitual en las series financieras, SUBEST2 converge 90 veces más rápido que MVE.

Cuadro 5.2.1: Modelo AR*: $(1 - .4B + .3B^2)Z_t = a_t; a_t \sim iidN(0, 1.0)$

Procedimiento		SUBEST2			MVE		
T	Valores reales	-.4	.3	1.0	-.4	.3	1.0
50	Media	-.367	.284	.943	-.386	.304	.946
	Desv. Típica	.136	.127	.198	.137	.135	.196
	RMSE	<u>.136</u>	<u>.127</u>	.198	.138	.135	<u>.196</u>
	Tiempo	100 %			193 %		
300	Media	-.391	.296	.992	-.395	.300	.994
	Desv. Típica	.053	.055	.082	.053	.056	.082
	RMSE	<u>.054</u>	<u>.055</u>	<u>.083</u>	<u>.054</u>	<u>.056</u>	<u>.083</u>
	Tiempo	100 %			475 %		

Cuadro 5.2.2: Modelo ARMA*: $(1 - .4B + .3B^2)Z_t = (1 - .8B)a_t; a_t \sim iidN(0, 1.0)$

Procedimiento		SUBEST2				MVE			
T	Valores reales	-.4	.3	-.8	1.0	-.4	.3	-.8	1.0
50	Media	-.307	.297	-.680	1.000	-.354	.313	-.800	.927
	Desv. Típica	.222	.141	.254	.206	.196	.141	.197	.194
	RMSE	.240	<u>.141</u>	.281	<u>.206</u>	<u>.201</u>	<u>.141</u>	<u>.197</u>	.207
	Tiempo	100 %				147 %			
300	Media	-.389	.304	-.805	1.001	-.395	.304	-.797	.993
	Desv. Típica	.066	.061	.074	.083	.066	.061	.050	.081
	RMSE	.067	<u>.062</u>	.074	.083	<u>.066</u>	<u>.062</u>	<u>.050</u>	<u>.081</u>
	Tiempo	100 %				383 %			

Cuadro 5.2.3: Modelo ARMA*: $(1 - .4B + .3B^2)Z_t = (1 - .8B)a_t; a_t \sim t$ con 8 g.d.

Procedimiento		SUBEST2				MVE			
T	Valores reales	-.4	.3	-.8	1.3	-.4	.3	-.8	1.3
50	Media	-.282	.294	-.678	.991	-.354	.308	-.802	.919
	Desv. Típica	.211	.161	.255	.247	.193	.155	.207	.238
	RMSE	.242	.161	.283	<u>.412</u>	<u>.199</u>	<u>.155</u>	<u>.207</u>	.478
	Tiempo	100 %				150 %			
300	Media	-.385	.302	-.808	.994	-.399	.300	-.803	.986
	Desv. Típica	.066	.061	.076	.057	.068	.061	.053	.055
	RMSE	<u>.068</u>	<u>.061</u>	.077	<u>.344</u>	<u>.068</u>	<u>.061</u>	<u>.053</u>	.352
	Tiempo	100 %				375 %			

* RMSE es la raíz cuadrada del error cuadrático medio. La estimación mediante MVE se inicializa con el verdadero valor de los parámetros. Se subraya el menor RMSE para cada tamaño muestral.

Cuadro 5.2.4: Modelo VARMA*: Resultados para $T = 50$

$$\begin{pmatrix} 1 - .7B + .6B^2 & 0 \\ 0 & 1 - 1.3B + .5B^2 \end{pmatrix} \begin{pmatrix} z_{1t} \\ z_{2t} \end{pmatrix} = \begin{pmatrix} 1 - .3B & -.9B \\ .6B & 1 - .8B \end{pmatrix} \begin{pmatrix} a_{1t} \\ a_{2t} \end{pmatrix}$$

$$\begin{pmatrix} a_{1t} \\ a_{2t} \end{pmatrix} \sim iidN \left[\begin{pmatrix} .0 \\ .0 \end{pmatrix}, \begin{pmatrix} .07 & .02 \\ - & .05 \end{pmatrix} \right]$$

Valores reales	-0.70	-1.30	0.60	0.50	-0.30	0.60	-0.90	-0.80	0.07	0.02	0.05
Procedimiento	SUBEST2										
Media	-.620	-1.297	.519	.500	-.331	.453	-.832	-.776	.073	.017	.056
Desv. Típica	.133	.202	.100	.151	.208	.121	.195	.251	.016	.010	.014
RMSE	.155	.202	.129	<u>.151</u>	.210	.190	<u>.207</u>	.252	<u>.017</u>	<u>.010</u>	.015
Tiempo	100 %										
Procedimiento	MVE										
Media	.686	-1.279	.589	.492	-.344	.654	-1.024	-.872	.060	.017	.044
Desv. Típica	.087	.188	.072	.155	.175	.122	.179	.220	.014	.010	.010
RMSE	<u>.088</u>	<u>.189</u>	<u>.073</u>	.155	<u>.181</u>	<u>.134</u>	.218	<u>.231</u>	<u>.017</u>	<u>.010</u>	<u>.012</u>
Tiempo	139 %										

Cuadro 5.2.5: Modelo VARMA*: Resultados para $T = 300$

Valores reales	-.70	-1.30	.60	.50	-.30	.60	-.90	-.80	.07	.02	.05
Procedimiento	SUBEST2										
Media	-.686	-1.311	.584	.506	-.289	.585	-.902	-.831	.072	.019	.053
Desv. Típica	.039	.076	.034	.062	.070	.043	.072	.089	.006	.004	.005
RMSE	.042	.076	.037	.062	.070	.046	.072	.095	<u>.006</u>	<u>.004</u>	.006
Tiempo	100 %										
Procedimiento	MVE										
Media	-.699	-1.296	.599	.500	-.305	.605	-.913	-.804	.069	.019	.049
Desv. Típica	.030	.055	.023	.047	.062	.035	.052	.066	.006	.004	.004
RMSE	<u>.030</u>	<u>.055</u>	<u>.023</u>	<u>.047</u>	<u>.062</u>	<u>.035</u>	<u>.054</u>	<u>.066</u>	<u>.006</u>	<u>.004</u>	<u>.004</u>
Tiempo	253 %										

* RMSE es la raíz cuadrada del error cuadrático medio. La estimación mediante MVE se inicializa con el verdadero valor de los parámetros. Se subraya el menor RMSE para cada tamaño muestral.

Cuadro 5.2.6: Modelo ARMA estacional*:

$$(1 - .4B)Z_t = (1 - .7B)(1 - .8B^4)a_t; a_t \sim iidN(0, 1.0)$$

Procedimiento		SUBEST2				MVE			
n	True Valores reales	-.4	-.7	-.8	1.0	-.4	-.7	-.8	1.0
	Media	-.294	-.632	-.742	.924	-.325	-.683	-.835	.886
50	Desv. Típica	.486	.482	.153	.210	.325	.329	.169	.204
	RMSE	.497	.486	<u>.163</u>	<u>.223</u>	<u>.333</u>	<u>.330</u>	.173	<u>.233</u>
	Tiempo	100 %				203 %			
	Media	-.365	-.665	-.826	.998	-.388	-.695	-.800	.990
300	Desv. Típica	.147	.132	.064	.084	.146	.122	.042	.080
	RMSE	.151	.136	.069	.084	<u>.146</u>	<u>.122</u>	<u>.042</u>	<u>.081</u>
	Tiempo	100 %				254 %			

Cuadro 5.2.7: Modelo ARMA estacional*:

$$(1 - .4B)Z_t = (1 - .7B)(1 - .8B^{12})a_t; a_t \sim iidN(0, 1.0)$$

Procedimiento		SUBEST2				MVE			
n	Valores reales	-.4	-.7	-.8	1.0	-.4	-.7	-.8	1.0
	Media	-.244	-.592	-.494	1.073	-.324	-.680	-.807	.902
50	Desv. Típica	.438	.454	.042	.237	.331	.323	.224	.225
	RMSE	.465	.467	.309	.248	<u>.340</u>	<u>.323</u>	<u>.224</u>	<u>.245</u>
	Tiempo	100 %				208 %			
	Media	-.379	-.691	-.812	.970	-.386	-.694	-.805	.984
300	Desv. Típica	.143	.128	.073	.084	.135	.112	.048	.083
	RMSE	.145	.128	.074	.089	<u>.136</u>	<u>.112</u>	<u>.048</u>	<u>.085</u>
	Tiempo	100 %				381 %			

* RMSE es la raíz cuadrada del error cuadrático medio. La estimación mediante MVE se inicializa con el verdadero valor de los parámetros. Se subraya el menor RMSE para cada tamaño muestral.

Cuadro 5.2.8: Modelo estructural*:

$$\begin{pmatrix} T_{t+1} \\ \Delta_{t+1} \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} T_t \\ \Delta_t \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \end{pmatrix} \eta_t$$

$$y_t = (1 \quad 0) \begin{pmatrix} T_t \\ \Delta_t \end{pmatrix} + \epsilon_t; \quad \begin{pmatrix} \eta_t \\ \epsilon_t \end{pmatrix} \sim iidN \left[\begin{pmatrix} .0 \\ .0 \end{pmatrix}, \begin{pmatrix} .01 & 0 \\ 0 & 10 \end{pmatrix} \right]$$

Procedimiento		SUBEST2		MVE	
T	Valores reales	.01	10	.01	10
	Media	.022	9.726	.035	9.911
50	Desv. Típica	.040	1.406	.069	1.005
	RMSE	<u>.042</u>	1.432	.073	<u>1.009</u>
	Tiempo	125 %		100 %	
	Media	.019	9.977	.008	10.005
300	Desv. Típica	.014	.443	.006	.409
	RMSE	.017	.444	<u>.007</u>	<u>.409</u>
	Tiempo	100 %		171 %	

Cuadro 5.2.9: Modelo AR con errores de observación*:

$$(1 - 1.5B + .8B^2)z_t^* = a_t; \quad z_t = z_t^* + v_t;$$

$$\begin{pmatrix} a_t \\ v_t \end{pmatrix} \sim iidN \left[\begin{pmatrix} .0 \\ .0 \end{pmatrix}, \begin{pmatrix} 1.0 & 0 \\ 0 & .5 \end{pmatrix} \right]$$

Procedimiento		SUBEST2				MVE			
T	Valores reales	-1.5	.8	1.0	.5	-1.5	.8	1.0	.5
	Media	-1.429	.737	1.045	.558	-1.480	.781	.954	.469
50	Desv. Típica	.133	.121	.195	.177	.128	.119	.185	.169
	RMSE	.150	.136	.200	.186	<u>.129</u>	<u>.121</u>	<u>.190</u>	<u>.172</u>
	Tiempo	144 %				100 %			
	Media	-1.487	.788	1.010	.513	-1.492	.793	.997	.493
300	Desv. Típica	.047	.044	.075	.057	.044	.043	.072	.055
	RMSE	.048	.046	.076	.059	<u>.045</u>	<u>.043</u>	<u>.072</u>	<u>.056</u>
	Tiempo	100 %				131 %			

* RMSE es la raíz cuadrada del error cuadrático medio. La estimación mediante MVE se inicializa con el verdadero valor de los parámetros. Se subraya el menor RMSE para cada tamaño muestral.

Cuadro 5.2.10: Modelo ARSV en forma de EE*:

$$x_{t+1} = .95x_t + w_t$$

$$z_t = x_t + 1.5u_t + v_t$$

$$E(w_t) = 0, \quad E(v_t) = 0, \quad E(w_t w_t') = .5, \quad E(v_t v_t') = 2, \quad E(w_t v_t') = 0$$

Procedimiento		SUBEST2				MVE			
T	Valores reales	.95	1.50	.50	2.0	.95	1.50	.50	2.0
500	Media	.945	1.453	.462	2.010	.933	1.471	.517	1.990
	Desv. Típica	.041	1.027	.146	.085	.034	.460	.104	.077
	RMSE	.041	1.028	.151	.085	<u>.038</u>	<u>.461</u>	<u>.106</u>	<u>.078</u>
	Tiempo	100 %				283 %			
3000	Media	.949	1.484	.494	2.004	.948	1.490	.501	2.000
	Desv. Típica	.009	.198	.044	.033	.009	.183	.038	.031
	RMSE	<u>.009</u>	.198	.044	.033	<u>.009</u>	<u>.184</u>	<u>.039</u>	<u>.031</u>
	Tiempo	100 %				1240 %			

Cuadro 5.2.11: Modelo GARCH(1,1) en representación ARMA*:

$$y_t = \epsilon_t \text{ con } \epsilon_t \sim iid(0, 1.0) \text{ y } \epsilon_t | \Omega_{t-1} \sim iidN(0, \sigma_t^2) \text{ donde}$$

$$\epsilon_t^2 = 1.0 + \frac{1-.80B}{1-.97B} v_t \quad \text{con } v_t = \epsilon_t^2 - \sigma_t^2$$

Procedimiento		SUBEST2			MVE		
T	Valores reales	1.0	-.97	-.80	1.0	-.97	-.80
500	Media	1.002	-.969	-.814	1.209	-.952	-.784
	Desv. Típica	.760	.031	.106	1.420	.037	.056
	RMSE	<u>.760</u>	<u>.031</u>	.107	1.436	.041	<u>.058</u>
	Tiempo	100 %			1650 %		
3000	Media	.998	-.972	-.809	1.095	-.967	-.797
	Desv. Típica	.313	.019	.068	.880	.011	.018
	RMSE	<u>.313</u>	.019	.069	.885	<u>.011</u>	<u>.018</u>
	Tiempo	100 %			9047 %		

* RMSE es la raíz cuadrada del error cuadrático medio. La estimación mediante MVE se inicializa con el verdadero valor de los parámetros. Se subraya el menor RMSE para cada tamaño muestral.

5.3. Conclusiones del Capítulo

En este Capítulo se deriva un algoritmo iterativo de estimación consistente basado en métodos de subespacios, denominado SUBEST2. Además, se ofrece una comparación directa para un amplio espectro de modelos de series temporales entre SUBEST2 y Máxima Verosimilitud Exacta (MVE) en términos de precisión, eficiencia y coste computacional.

La conclusión fundamental es que el procedimiento propuesto y MVE son métodos complementarios para modelos gaussianos con muestras cortas y con componentes de medias móviles, ya que el primero permite obtener buenas y rápidas condiciones iniciales, mientras que una segunda optimización de los parámetros (si se requiere) puede realizarse con MVE a partir de las anteriores.

Cuando los modelos son puramente autorregresivos, se tienen dudas acerca de la normalidad de los errores o la muestra es relativamente grande (aún en modelos estacionales), SUBEST2 puede usarse como un algoritmo de estimación alternativo con una gran ventaja en coste computacional que crece a medida que aumenta el tamaño muestral.

Por último, los ejercicios de simulación sugieren que este nuevo método es muy útil cuando se trabaja con datos financieros de alta frecuencia. En concreto, este es el caso en donde SUBEST2 ofrece mayores ventajas frente a MVE. Además, teniendo en cuenta los problemas de estimación existentes en modelos de heteroscedasticidad condicional multivariantes mediante máxima verosimilitud con muestras reales como, por ejemplo, problemas de convergencia, mal condicionamiento (ver Jerez et al., 1999) o estructuras que no cumplen las condiciones de “positividad”, la alternativa propuesta parece ser especialmente valiosa.

Capítulo 6

Resultados Empíricos

Este Capítulo está dedicado a la realización de diversos ejercicios empíricos, con datos reales y series simuladas. En él, se aplica paso a paso la metodología de identificación, especificación y estimación presentada en los Capítulos anteriores, tanto en el caso univariante regular y/o estacional como en el multivariante, con o sin cointegración.

Para la programación de los algoritmos se ha utilizado MatLab¹, un sistema de cálculo interactivo que cubre buena parte de los problemas tratados por el álgebra lineal numérica. Además de los programas, donde se instrumentan los procedimientos explicados con anterioridad, se ha empleado: 1) el paquete E4² (Terceiro et al., 2000), para la simulación de modelos, estimación mediante Máxima Verosimilitud Exacta (MVE) y el cálculo de la matriz de información exacta, y 2) el conjunto de funciones SID (Casals, 1997) para la estimación de modelos por métodos de subespacios, ambos también desarrollados en MatLab.

El Capítulo se divide en dos secciones. La Sección 6.1 usa la metodología completa de identificación y estimación para modelos univariantes. Para ello, se ilustra en 6.1.1 el procedimiento más general, abordando el análisis de una serie temporal no estacionaria con estructura estacional multiplicativa. En el apartado 6.1.2, se exponen los resultados del análisis a partir de la metodología propuesta de diversas series temporales, reales y simuladas, empleadas en la literatura. La Sección 6.2 está dedicada al análisis multivariante, centrándose en tres conjuntos de datos con propiedades dispares. El primero no muestra existencia de cointegración a pesar de la gran similitud en el movimiento de las series. El segundo y el tercero presentan relaciones de cointegración con uno y dos factores comunes respectivamente.

¹Marca registrada de The Mathworks

²<http://www.ucm.es/info/icae/e4/>

6.1. Modelos Univariantes

Los dos apartados siguientes ilustran la utilización de los algoritmos propuestos, así como los resultados de especificación y estimación obtenidos en modelos univariantes. El primero aplica paso a paso la metodología desarrollada tomando como ejemplo la serie de Tráfico de Pasajeros de Líneas Aéreas, usada por Box y Jenkins (1976). Este es un caso general, donde la serie analizada presenta no estacionariedad y estacionalidad de tipo multiplicativo. En la segunda parte, se ofrecen los resultados de especificación y estimación para otras nueve series siguiendo, los mismos pasos del ejemplo anterior.

6.1.1. Tráfico de Pasajeros de Líneas Aéreas (Box y Jenkins, 1976)

La serie temporal de Tráfico de Pasajeros de Líneas Aéreas (y_t) ha sido empleada por Box y Jenkins (1976) para ilustrar la modelización de la estacionalidad. En su análisis, éstos y otros autores identifican una estructura IMA doble.

Partiendo de la transformación Box-Cox (Box y Cox, 1964) de la serie original, $z_t = \log(y_t)$, el ejercicio se presenta de la siguiente manera: 1) se decide la transformación estacionaria de la serie, 2) se identifican los órdenes de los subsistemas regular y estacional y 3) se estima y depura el modelo propuesto.

Al tratarse de una serie mensual y teniendo en cuenta su aspecto gráfico (donde puede observarse una estacionalidad evidente), se lleva a cabo la transformación de la serie z_t que se propone en el apartado 3.1.1 del trabajo. Es decir:

$$S(B)z_t = w_t \quad (6.1.1)$$

en donde $S(B) = 1 + B + \dots + B^{11}$. De esta forma, es posible usar el algoritmo de detección de raíces unitarias e identificación del orden regular sobre la serie w_t . Al aplicar NID, se detecta al menos una raíz unitaria en la serie transformada, por lo que se repite la operación sobre su primera diferencia, ∇w_t . Los resultados se presentan en el Cuadro 6.1.1.

El criterio de detección de raíces unitarias encuentra también indicios de no estacionariedad en ∇w_t , por lo que se introduce en NID, $\nabla^2 w_t$. Esta vez, el algoritmo no detecta señales de no estacionariedad. La salida se expone en el Cuadro 6.1.2 que además ofrece la identificación del orden regular.

Con la ayuda del Cuadro 6.1.2, se establece: i) la transformación estacionaria de la serie, $\nabla^2 S(B)z_t = \nabla \nabla_{12} z_t$, y ii) la dimensión del subsistema regular, $\hat{n}_r = 1$.

Cuadro 6.1.1: Salida del algoritmo NID aplicado a ∇w_t

Raíces unitarias						
1						
Resultado de los diferentes criterios						
n	AIC	SBC	HQ	SVC_{Ω_2}	NIDC	$\chi^2(95\%)$
0	-6.13	-6.10	-6.12	10.74	10.74	-
1	-8.44	-8.38	-8.42	0.15	0.13	1.00
2	-8.41	-8.30	-8.37	0.22	0.17	0.85
3	-8.38	-8.23	-8.32	0.23	0.16	0.94
Orden regular estimado						
1	1	1	1	1	1	1

Cuadro 6.1.2: Salida del algoritmo NID aplicado a $\nabla^2 w_t$

Raíces unitarias						
0						
Resultado de los diferentes criterios						
n	AIC	SBC	HQ	SVC_{Ω_2}	NIDC	$\chi^2(95\%)$
0	-8.29	-8.27	-8.28	0.24	0.24	-
1	-8.43	-8.36	-8.40	0.16	0.14	1.00
2	-8.40	-8.29	-8.36	0.21	0.17	0.93
3	-8.38	-8.22	-8.31	0.24	0.17	0.80
Orden regular estimado						
1	1	1	1	1	1	1

Para estimar el orden del subsistema estacional, introducimos la transformación estacionaria previamente elegida, en el algoritmo de identificación. Sin embargo, ahora se especifica también en NID la periodicidad de la serie, $s = 12$. El algoritmo, que trabaja entonces con las matrices Block-Hankel Estacionales, devuelve el Cuadro 6.1.3. Con la información ofrecida, se decide iii) la dimensión del subsistema estacional, $\hat{n}_s = 1$.

A continuación, se aplica el algoritmo de estimación mediante subespacios a la serie transformada especificando los órdenes de los subsistemas regular y estacional

Cuadro 6.1.3: Salida del algoritmo NID ($s = 12$) aplicado a $\nabla\nabla_{12}z_t$

Resultado de los diferentes criterios						
n	AIC	SBC	HQ	SVC_{Ω_2}	NIDC	$\chi^2(95\%)$
0	-3.32	-3.30	-3.32	0.39	0.3	-
1	-3.57	-3.50	-3.58	0.11	0.09	0.99
2	-3.54	-3.43	-3.56	0.17	0.13	0.66
3	-3.51	-3.36	-3.53	0.24	0.17	0.66
Orden estacional estimado						
	1	1	1	1	1	1

previamente estimados. Como resultado, se obtiene el siguiente modelo equilibrado:

$$(1 - \underset{(.21)}{.06}B)(1 - \underset{(.17)}{.08}B^{12})\nabla\nabla_{12}z_t = (1 - \underset{(.21)}{.43}B)(1 - \underset{(.16)}{.58}B^{12})\hat{a}_{1t}, \quad \hat{\sigma}_{a_1} = 3.68\% \quad (6.1.2)$$

donde entre paréntesis se ofrecen las desviaciones típicas de los parámetros y donde $\hat{\sigma}_{a_1}$ denota la desviación típica porcentual del error.

Si se está interesado en depurar el modelo de parámetros no significativos y obtener, sin un alto coste computacional, una estimación final más precisa o buenas condiciones iniciales para una estimación máximo-verosímil, es posible incorporar las restricciones sobre coeficientes en el algoritmo SUBEST2. El resultado de esta estimación es el siguiente:

$$\nabla\nabla_{12}z_t = (1 - .39B)(1 - .52B^{12})\hat{a}_{2t}, \quad \hat{\sigma}_{a_2} = 3.68\% \quad (6.1.3)$$

En cualquier caso, se precise o no esta segunda estimación, es oportuno realizar una diagnosis de los residuos. En este ejemplo, los coeficientes estimados son muy próximos y los resultados apenas varían entre los modelos (6.1.2) y (6.1.3). Como ilustración, se muestra la diagnosis de los errores de la segunda estimación (6.1.3). Para ello, se aplica NID tanto para $s = 1$ como para $s = 12$, comprobando que no existe estructura dinámica regular ni estacional en los residuos \hat{a}_{2t} . Los Cuadros 6.1.4 y 6.1.5 reproducen el resultado ofrecido.

A la luz de los resultados, los residuos \hat{a}_{2t} no presentan indicios de una mala especificación ni en su parte regular, ni en su parte estacional. El criterio propuesto para la estimación del orden del sistema, *NIDC*, es el único que sugiere la existencia de estructura en la parte regular. Esto no es muy sorprendente ya que sobreestima ligeramente en muestras relativamente cortas (ver Capítulo 4). Como puede observarse, el modelo estimado corresponde con la estructura identificada

Cuadro 6.1.4: Salida del algoritmo NID aplicado a \hat{a}_{2t}

Raíces unitarias						
0						
Resultado de los diferentes criterios						
n	AIC	SBC	HQ	$SV C_{\Omega_2}$	NIDC	$\chi^2(95\%)$
0	-3.77	-3.74	-3.76	0.12	0.12	-
1	-3.76	-3.69	-3.73	0.14	0.12	0.64
2	-3.73	-3.62	-3.68	0.18	0.13	0.77
3	-3.71	-3.56	-3.65	0.23	0.17	0.57
Orden regular estimado						
	0	0	0	0	1	0

Cuadro 6.1.5: Salida del algoritmo NID ($s = 12$) aplicado a \hat{a}_{2t}

Resultado de los diferentes criterios						
n	AIC	SBC	HQ	$SV C_{\Omega_2}$	NIDC	$\chi^2(95\%)$
0	-3.77	-3.74	-3.76	0.04	0.04	-
1	-3.74	-3.67	-3.71	0.11	0.09	0.71
2	-3.72	-3.61	-3.68	0.16	0.12	0.34
3	-3.69	-3.54	-3.63	0.23	0.17	0.24
Orden estacional estimado						
	0	0	0	0	0	0

por Box y Jenkins (1976). La estimación por MVE, a partir de las condiciones iniciales del segundo modelo, resulta

$$\nabla \nabla_{12} z_t = (1 - \underset{(.08)}{.40} B)(1 - \underset{(.08)}{.56} B^{12}) \hat{a}_t, \quad \hat{\sigma}_a = 3.67\% \quad (6.1.4)$$

Consecuentemente, del análisis de este caso puede concluirse que el criterio de discriminación de raíces unitarias ha conducido a una decisión correcta sobre la transformación estacionaria de la serie. Además, el proceso de identificación de la serie revela la existencia de una representación ARMA multiplicativa acorde con la literatura. Resulta relevante comentar la proximidad entre los coeficientes estimados por los dos algoritmos de subespacios y los máximo-verosímiles. Aunque en sistemas univariantes la metodología propuesta no ofrece tantas ventajas como en el análisis de series múltiples, cabe resaltar la simplicidad para automatizar los pasos desarrollados en este apartado. Esto permite llevar a cabo el seguimiento

masivo de procesos univariantes con un bajo coste computacional y de esfuerzo del analista.

6.1.2. Identificación semi-automática de modelos para series temporales reales y simuladas

En este segundo apartado se muestran los resultados intermedios y finales de la utilización de la metodología propuesta, usando nueve series temporales habitualmente modelizadas en la literatura. En el Cuadro 6.1.6, se describen brevemente las series analizadas, mientras que en los Cuadros 6.1.7, 6.1.8 y 6.1.9 se presentan los resultados de estimación obtenidos.

En primer lugar, en todos los procesos estudiados, ya sean o no estacionales, se detecta correctamente la transformación estacionaria con el procedimiento propuesto. En cuanto al orden del sistema, en la mayoría de las series sin estacionalidad, su estimación coincide en los seis criterios usados, con el resultante de los modelos publicados en la literatura. Esto no ocurre en la serie D, donde el criterio *SBC* de Schwartz (1978) sugiere un orden del proceso igual a 1, mientras que el resto encuentra $\hat{n} = 2$ lo que coincide con el orden de dinámica de la serie simulada. Este resultado es consistente con el comportamiento mostrado por el criterio *SBC* en las simulaciones del Capítulo 4, donde el criterio infraestima claramente el orden de un modelo AR(2) con raíces imaginarias. En las series que presentan estacionalidad (series B, G y H) se siguen los pasos especificados en el apartado 4.1.2. En B, todos los criterios y el contraste indican $\hat{n}_s = 1$ y $\hat{n}_r = 1$ al igual que en la literatura. En cambio, al aplicar NID a la serie G no se encuentra, en un primer momento, estructura estacional (tan sólo el criterio *AIC* de Akaike (1976) y el test χ^2 de Tiao y Tsay (1989) muestran $\hat{n}_s = 1$). Sin embargo, todos coinciden en que la dinámica regular es $\hat{n}_r = 1$. Una vez introducida la parte regular del modelo, el análisis de residuos señala la existencia de dinámica estacional: todos los criterios indican $\hat{n}_s = 1$ sobre los residuos. Con lo que se identifica el mismo modelo que propone la literatura. Finalmente, el análisis de la serie H genera $\hat{n}_s = 1$ para todos los criterios utilizados y $\hat{n}_r = 0$ para *AIC*, *SBC*, *HQ* (Hannan y Quinn, 1979) y el test χ^2 (tanto *SVC* _{Ω_2} como *NIDC* sobreestiman, mostrando $\hat{n}_r = 2$) que coincide con el análisis manual realizado por el autor.

La única diferencia significativa entre la estimación final a partir de la especificación semi-automática y la de la literatura se presenta en la serie H. Obsérvese que el modelo estimado por SUBEST2 y el que propone Wei (1990) son iguales, con ciertas diferencias en el valor de los parámetros. Sin embargo, al estimar este modelo por MVE, el componente MA estacional converge a la unidad. Esto

ocurre porque la serie presenta una estacionalidad determinista que podría también modelizarse en niveles con un parámetro estacional para cada trimestre. Los residuos del modelo no presentan indicios de mala especificación, al ser los valores del estadístico Q de autocorrelación residual, $Q_{(10)} = 9.6$ en el modelo estimado por MVE y $Q_{(10)} = 9.9$ el valor proporcionado por el autor. Además, el valor de la función de verosimilitud en el óptimo es inferior al del modelo estimado que propone Wei (1990) ($l_{MVE} = 47.4$ por $l_{Wei} = 48.0$), por lo que se concluye que el modelo estimado en este trabajo representa adecuadamente los datos.

Los resultados expuestos en las Cuadros 6.1.7, 6.1.8 y 6.1.9 detallan los pasos realizados con la metodología de identificación de series temporales propuesta. Esta es extremadamente flexible, pues permite finalizar el análisis con modelos más o menos precisos dependiendo del uso de los mismos. Obsérvese que añadiendo un bajo coste de esfuerzo del analista y avanzando en la metodología, se obtienen modelos muy similares a los propuestos por los autores. Sin embargo, los modelos de los Cuadros 6.1.7 o 6.1.8 pueden ser adecuados, por ejemplo, para prever.

Cuadro 6.1.6: Series temporales utilizadas en los Cuadros 6.1.7, 6.1.8 y 6.1.9

Serie	T	Descripción
A	45	Burr (1976): Promedio Diario de Defectos en la Fabricación de Camiones
B*	150	Wei (1990): Serie Trimestral Simulada
C	35	Wei (1990): Desempleadas entre 16-19 años por Mes, en EEUU; Enero 1961-Diciembre 1985
D*	200	Wold (1964): Serie Simulada n°2;
E	226	Box y Jenkins (1976): Serie C-Temperatura de un Proceso Químico, medida cada minuto
F	197	Box y Jenkins (1976): Serie A-Concentración de un Proceso Químico, medida cada 2 horas
G	178	O'Donovan (1983): Número Mensual de Empleados en la Producción de Alimentos en el Estado de Wisconsin, Enero 1961-October 1975
H	32	Wei (1990): Producción Trimestral de Cerveza en EEUU; I Trim. 1975-IV Trim. 1982
I*	36	Makridakis et al. (1983): Serie Simulada

*Las tres series son simuladas. Sin embargo, en los Cuadros 6.1.7, 6.1.8 y 6.1.9, las series B y C presentan los procesos generadores de los datos, mientras que la serie I ofrece la especificación y estimación del autor.

Cuadro 6.1.7: Modelos Univariantes. Resumen de la identificación semi-automática de modelos para series reales y simuladas (1). Entre paréntesis se presentan las desviaciones típicas de los parámetros y $\hat{\sigma}_a^2$ es la varianza residual estimada. NID identifica la transformación estacionaria y los órdenes de dinámica regular y estacional. SUBESTI estima mediante subespacios el modelo identificado por NID.

Serie	T	Especificación y estimación en la literatura	NID + SUBESTI
A	45	$(1 - .43B)(Z_t - 1.79) = a_t$ (.13) (.07) $\hat{\sigma}_a^2 = .21$	$(1 - .61B)(Z_t - 3.94) = (1 - .24B)a_t$ (.30) (.31) $\hat{\sigma}_a^2 = .21$
B	150	$\nabla\nabla_4 Z_t = (1 - .80B)(1 - .60B^4)a_t$ $\hat{\sigma}_a^2 = 1.00$	$(1 + .22B)(1 + .21B^4)\nabla\nabla_4 Z_t = (1 - .56B)(1 - .49B^4)a_t$ (.13) (.13) (.12) $\hat{\sigma}_a^2 = .97$
C	35	$\nabla Z_t = (1 - .51B)a_t$ (.05) $\hat{\sigma}_a^2 = 1397.27$	$(1 - .00B)\nabla Z_t = (1 - .51B)a_t$ (.11) $\hat{\sigma}_a^2 = 1380.35$
D	200	$(1 - .70B + .49B^2)Z_t = a_t$ $\hat{\sigma}_a^2 = .60$	$(1 - .84B + .51B^2)Z_t = (1 - .16B + .05B^2)a_t$ (.11) (.10) (.11) (.11) $\hat{\sigma}_a^2 = .66$
E	226	$(1 - .82B)\nabla Z_t = a_t$ (.04) $\hat{\sigma}_a^2 = .02$	$(1 - .82B)\nabla Z_t = (1 - .02B)a_t$ (.06) (.08) $\hat{\sigma}_a^2 = .02$
F	197	$(1 - .92B)Z_t = 1.45 + (1 - .58B)a_t$ (.04) (.08) $\hat{\sigma}_a^2 = .10$	$(1 - .88B)Z_t = 2.08 + (1 - .50B)a_t$ (.09) (.02) (.10) $\hat{\sigma}_a^2 = .10$
G	178	$(1 - .79B)\nabla_{12} Z_t = (1 - .76B^{12})a_t$ (.05) (.05) $\hat{\sigma}_a^2 = 1.24$	$(1 - .56B)(1 - .05B^{12})\nabla_{12} Z_t = -2.11 + (1 - .18B)(1 - .35B^{12})a_t$ (.10) (.18) (.10) (.11) $\hat{\sigma}_a^2 = 1.54$
H	32	$\nabla_4 Z_t = 1.49 + (1 - .87B^4)a_t$ (.09) (.16) $\hat{\sigma}_a^2 = 2.39$	$(1 - .14B^4)\nabla_4 Z_t = 1.21 + (1 - .71B^4)a_t$ (.26) (.30) (.23) $\hat{\sigma}_a^2 = 2.03$
I	36	$Z_t = 51.03 + a_t$ (4.84) $\hat{\sigma}_a^2 = 818.81$	$Z_t = 47.31 + a_t$ (4.77) $\hat{\sigma}_a^2 = 790.54$

Cuadro 6.1.8: Modelos Univariantes. Resumen de la identificación semi-automática de modelos para series reales y simuladas (2). $\hat{\sigma}_a^2$ es la varianza residual estimada. Se muestran los resultados de la estimación mediante SUBEST2 eliminando los parámetros no significativos de los modelos del Cuadro 6.1.7. No se ofrecen desviaciones típicas de las estimaciones ya que, en este caso, se utiliza SUBEST2 únicamente para obtener condiciones iniciales y no como estimación final (aunque esto también es posible).

Serie	T	Especificación y estimación en la literatura	NID + SUBEST1 + SUBEST2
A	45	$(1 - .43B)(Z_t - 1.79) = a_t$ (.13) $\hat{\sigma}_a^2 = .21$	$(1 - .35B)(Z_t - 1.87) = a_t$ $\hat{\sigma}_a^2 = .21$
B	150	$\nabla\nabla_4 Z_t = (1 - .80B)(1 - .60B^4)a_t$ $\hat{\sigma}_a^2 = 1.00$	$\nabla\nabla_4 Z_t = (1 - .74B)(1 - .66B^4)a_t$ $\hat{\sigma}_a^2 = .88$
C	35	$\nabla Z_t = (1 - .51B) + a_t$ (.05) $\hat{\sigma}_a^2 = 1397.27$	$\nabla Z_t = (1 - .50B) + a_t$ $\hat{\sigma}_a^2 = 1393.54$
D	200	$(1 - .70B + .49B^2)Z_t = a_t$ $\hat{\sigma}_a^2 = .60$	$(1 - .71B + .42B^2)Z_t = a_t$ $\hat{\sigma}_a^2 = .66$
E	226	$(1 - .82B)\nabla Z_t = a_t$ (.04) $\hat{\sigma}_a^2 = .02$	$(1 - .81B)\nabla Z_t = a_t$ $\hat{\sigma}_a^2 = .02$
F	197	$(1 - .92B)Z_t = 1.45 + (1 - .58B)a_t$ (.04) $\hat{\sigma}_a^2 = .10$	$(1 - .88B)Z_t = 2.07 + (1 - .51B)a_t$ $\hat{\sigma}_a^2 = .10$
G	178	$(1 - .79B)\nabla_{12} Z_t = (1 - .76B^{12})a_t$ (.05) $\hat{\sigma}_a^2 = 1.24$	$(1 - .68B)\nabla_{12} Z_t = .07 + (1 - .56B^{12})a_t$ $\hat{\sigma}_a^2 = 1.37$
H	32	$\nabla_4 Z_t = 1.49 + (1 - .87B^4)a_t$ (.09) $\hat{\sigma}_a^2 = 2.39$	$\nabla_4 Z_t = 1.36 + (1 - .59B^4)a_t$ $\hat{\sigma}_a^2 = 2.15$
I	36	$Z_t = 51.03 + a_t$ (4.84) $\hat{\sigma}_a^2 = 818.81$	$Z_t = 51.03 + a_t$ $\hat{\sigma}_a^2 = 796.68$

Cuadro 6.1.9: Modelos Univariantes. Resumen de la identificación semi-automática de modelos para series reales y simuladas (3). Entre paréntesis se presentan las desviaciones típicas de los parámetros y $\hat{\sigma}_a^2$ es la varianza residual estimada. Se ofrece la estimación final mediante Máxima Verosimilitud Exacta (MVE) a partir de la modelización propuesta.

Serie	T	Especificación y estimación en la literatura	NID + SUBEST1 + SUBEST2 + MVE
A	45	$(1 - .43B)(Z_t - 1.79) = a_t$ (.13) $\hat{\sigma}_a^2 = .21$	$(1 - .42B)(Z_t - 1.81) = a_t$ (.07) $\hat{\sigma}_a^2 = .21$
B	150	$\nabla \nabla_4 Z_t = (1 - .80B)(1 - .60B^4)a_t$ $\hat{\sigma}_a^2 = 1.00$	$\nabla \nabla_4 Z_t = (1 - .75B)(1 - .63B^4)a_t$ (.06) $\hat{\sigma}_a^2 = .91$
C	35	$\nabla Z_t = (1 - .51B)a_t$ (.05) $\hat{\sigma}_a^2 = 1397.27$	$\nabla Z_t = (1 - .51B)a_t$ (.05) $\hat{\sigma}_a^2 = 1393.12$
D	200	$(1 - .70B + .49B^2)Z_t = a_t$ $\hat{\sigma}_a^2 = .60$	$(1 - .72B + .41B^2)Z_t = a_t$ (.06) $\hat{\sigma}_a^2 = .66$
E	226	$(1 - .82B)\nabla Z_t = a_t$ (.04) $\hat{\sigma}_a^2 = .02$	$(1 - .82B)\nabla Z_t = a_t$ (.04) $\hat{\sigma}_a^2 = .02$
F	197	$(1 - .92B)Z_t = 1.45 + (1 - .58B)a_t$ (.04) $\hat{\sigma}_a^2 = .10$	$(1 - .92B)Z_t = 1.45 + (1 - .60B)a_t$ (.04) $\hat{\sigma}_a^2 = .10$
G	178	$(1 - .79B)\nabla_{12} Z_t = (1 - .76B^{12})a_t$ (.05) $\hat{\sigma}_a^2 = 1.24$	$(1 - .77B)\nabla_{12} Z_t = (1 - .61B^{12})a_t$ (.05) $\hat{\sigma}_a^2 = 1.31$
H	32	$\nabla_4 Z_t = 1.49 + (1 - .87B^4)a_t$ (.09) $\hat{\sigma}_a^2 = 2.39$	$\nabla_4 Z_t = 1.55 + \nabla_4 a_t$ (.13) $\hat{\sigma}_a^2 = 1.64$
I	36	$Z_t = 51.03 + a_t$ (4.84) $\hat{\sigma}_a^2 = 818.81$	$Z_t = 51.03 + a_t$ (4.84) $\hat{\sigma}_a^2 = 818.86$

6.2. Modelos Multivariantes

En la sección anterior se ha visto que la metodología derivada proporciona buenos resultados en modelos univariantes. Sin embargo, es en los procesos multivariantes donde la ganancia es mayor: la incorporación de un rápido análisis de cointegración y la especificación de los Índices de Kronecker (IK) reducen sustancialmente el número de parámetros y con ello, el coste computacional y los problemas de condicionamiento de la función de verosimilitud. Este apartado se dedica a la especificación y estimación de modelos multivariantes con la metodología expuesta en este trabajo. Los sistemas de series utilizados poseen características interesantes que pueden ayudar a evaluar el comportamiento de los diferentes algoritmos presentados en los Capítulos anteriores.

En primer lugar, se analizan las series de los Índices de Precios de la Harina usadas por Tiao y Tsay (1989) y otros autores, para ilustrar sus metodologías de especificación. En el segundo subapartado, se lleva a cabo un estudio sobre los tipos de interés de la economía de Estados Unidos, analizados entre otros por Martín Manjón y Treadway (1997) y Martín Manjón (1998), que presentan distintas relaciones de cointegración.

6.2.1. Series de Precios de la Harina (Tiao y Tsay, 1989)

Las series analizadas (\mathbf{y}_t) representan los Índices mensuales de Precios de la Harina en las ciudades de Buffalo, Minneapolis y Kansas City entre los meses de Agosto de 1972 y Noviembre de 1980. Fueron presentadas por Tiao y Tsay (1989) para ilustrar su metodología de especificación y más tarde analizadas por Grubb (1992), Lütkepohl y Poskitt (1996) o Casals (1997).

Las tres variables muestran una gran similitud en los movimientos de los precios, sin embargo, tanto Tiao y Tsay (1989) como Lütkepohl y Poskitt (1996) rechazan la hipótesis de cointegración de su transformación logarítmica. La Figura 6.2.1 representa el logaritmo de los tres índices de precios.

Al procesar mediante el algoritmo NID el logaritmo de cada una de las series, se establece que todas ellas son integradas de orden $I(1)$. Por tanto, es posible contrastar la existencia de cointegración aplicando de nuevo el procedimiento al sistema conjunto. El resultado se ofrece en el Cuadro 6.2.1. La existencia de tres Coeficientes de Correlación Canónica (CCC) suficientemente cercanos a la unidad, permiten desechar la idea de cointegración. Los criterios de selección del orden del sistema pierden tamaño debido a la no estacionariedad del proceso y tan sólo son capaces de reconocer las tres raíces unitarias como el valor del índice de McMillan.

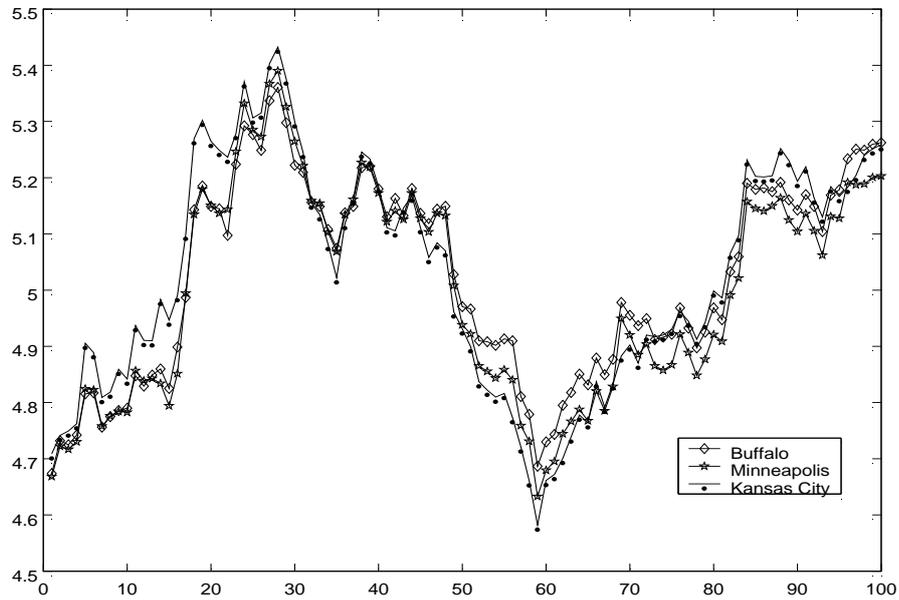


Figura 6.2.1: Logaritmo de los Precios de la Harina en las diferentes ciudades

Cuadro 6.2.1: Salida del algoritmo NID aplicado a $\log(\mathbf{y}_t)$

Raíces unitarias						
3						
Relaciones de Cointegración						
0						
Resultado de los diferentes criterios						
n	AIC	SBC	HQ	SVC_{Ω_2}	NIDC	$\chi^2(95\%)$
0	-1.18	-1.03	-1.12	56.21	56.21	-
1	-5.63	-5.32	-5.50	8.15	8.08	1.00
2	-7.60	-7.13	-7.41	4.12	3.97	1.00
3	-9.02	-8.39	-8.77	1.18	0.97	1.00
4	-8.90	-8.12	-8.59	1.32	1.04	0.01
Orden del sistema estimado						
3 3 3 3 3 3						

Para encontrar el orden del sistema se toman diferencias sobre el logaritmo de las series originales y se aplica el algoritmo de identificación, NID, sobre la transformación estacionaria de las series $\mathbf{z}_t = \nabla \log(\mathbf{y}_t)$. Los resultados aparecen en el Cuadro 6.2.2.

Cuadro 6.2.2: Salida del algoritmo NID aplicado a \mathbf{z}_t^*

Raíces unitarias						
0						
Resultado de los diferentes criterios						
n	AIC	SBC	HQ	SVC_{Ω_2}	NIDC	$\chi^2(95\%)$
0	-13.61	-13.45	-13.55	0.67	0.67	-
1	-13.83	-13.52	-13.71	0.80	0.73	0.91
2	-13.75	-13.28	-13.56	0.98	0.84	0.61
3	-13.68	-13.05	-13.43	1.13	0.92	0.08
4	-13.59	-12.81	-13.28	1.34	1.05	0.01
Orden del sistema estimado						
	1	1	1	0	0	0

* $\mathbf{z}_t = \nabla \log(\mathbf{y}_t)$

Como puede observarse el índice de McMillan estimado no se repite en todos los criterios. Como se sugiere en el planteamiento teórico, se decide el orden del sistema como la moda de los \hat{n} óptimos que devuelve NID. En este caso, al existir un empate entre 0 y 1 se toma el mayor de ellos. La decisión de probar con un modelo de orden 1 viene motivada por dos razones: 1) en el caso en que se produzca una sobreparametrización, ésta podrá resolverse en la fase de depuración de parámetros no significativamente distintos de cero y, 2) la masa de probabilidad asociada al contraste χ^2 correspondiente a $\hat{n} = 1$ es bastante elevada (91 %) con respecto del resto de valores de la columna.

Una vez decidido el orden del sistema ($\hat{n} = 1$), el algoritmo de estimación SUBEST1 genera, por un lado los índices de observabilidad estimados, y por otro, el modelo estimado en forma VARMA echelon. Los criterios *AIC*, *SBC* y *HQ* coinciden en que el juego de IK es (1,0,0). El modelo estimado es:

$$\hat{\Phi}(B)\mathbf{z}_t = \hat{\Theta}(B)\hat{\mathbf{a}}_t \quad (6.2.1)$$

con

$$\hat{\Phi}(B) = \begin{pmatrix} 1 + .28B & 0 & 0 \\ (.11) & (-) & (-) \\ -.55 & 1 & 0 \\ (.13) & (-) & (-) \\ -.11 & 0 & 1 \\ (.29) & (-) & (-) \end{pmatrix} \quad \hat{\Theta}(B) = \begin{pmatrix} 1 + .02B & .53B & -.02B \\ (.12) & (.11) & (.05) \\ -.55 & 1 & 0 \\ (.13) & (-) & (-) \\ -.11 & 0 & 1 \\ (.29) & (-) & (-) \end{pmatrix}$$

$$\hat{\Sigma}_{\alpha}(\%) = \begin{pmatrix} .23 & - & - \\ .24 & .26 & - \\ .23 & .25 & .29 \end{pmatrix} \quad (6.2.2)$$

donde entre paréntesis se ofrecen las desviaciones típicas de los parámetros y la estimación de la matriz de varianzas y covarianzas del ruido se presenta en términos porcentuales.

Como indica Casals (1997) cabe destacar que, aunque el modelo se corresponde con una estructura VARMA(1,1), los IK seleccionados imponen una estructura que reduce a seis los coeficientes estimados al margen de la matriz de varianzas y covarianzas $\hat{\Sigma}_{\alpha}$, lo que supone una disminución sustancial del número de parámetros estimados con respecto, por ejemplo, a una estructura VAR(1) como la que proponen Tiao y Tsay (1989).

La estimación del modelo anterior, mediante el algoritmo SUBEST2 eliminando el parámetro no significativamente distinto de cero correspondiente al elemento (3,1) de la matriz $\hat{\Phi}$, evidencia la existencia de estructura dinámica adicional en los residuos de la tercera serie (Precios en Kansas City). Se estima entonces el modelo obtenido en la identificación de los índices de observabilidad sin restricciones adicionales, cuyos residuos no revelan indicios de mala especificación, con el resultado siguiente:

$$\hat{\Phi}(B) = \begin{pmatrix} 1 + .49B & 0 & 0 \\ -.76 & 1 & 0 \\ -.54 & 0 & 1 \end{pmatrix} \quad \hat{\Theta}(B) = \begin{pmatrix} 1 - .94B & 1.32B & .09B \\ -.76 & 1 & 0 \\ -.54 & 0 & 1 \end{pmatrix}$$

$$\hat{\Sigma}_{\alpha}(\%) = \begin{pmatrix} .21 & - & - \\ .22 & .25 & - \\ .21 & .23 & .28 \end{pmatrix} \quad (6.2.3)$$

Con respecto de la estimación inicial, se encuentran ciertas diferencias en los valores de los parámetros estimados, fundamentalmente en la matriz media móvil. A partir de estas condiciones iniciales se lleva a cabo la estimación por MVE. La primera estimación devuelve unos valores de los parámetros muy próximos a los ofrecidos en (6.2.3) con el elemento (1,3) de la matriz $\hat{\Theta}$ no significativo. La estimación restringida conduce al modelo (6.2.1) con las matrices:

$$\hat{\Phi}(B) = \begin{pmatrix} 1 + .41B & 0 & 0 \\ (.17) & (-) & (-) \\ -.73 & 1 & 0 \\ (.11) & (-) & (-) \\ -.49 & 0 & 1 \\ (.20) & (-) & (-) \end{pmatrix} \quad \hat{\Theta}(B) = \begin{pmatrix} 1 - .94B & 1.32B & 0 \\ (.46) & (.37) & (-) \\ -.73 & 1 & 0 \\ (.11) & (-) & (-) \\ -.49 & 0 & 1 \\ (.20) & (-) & (-) \end{pmatrix}$$

$$\hat{\Sigma}_a(\%) = \begin{pmatrix} .21 & - & - \\ (.03) & & \\ .22 & .25 & - \\ (.03) & (.03) & \\ .21 & .24 & .28 \\ (.03) & (.04) & (.04) \end{pmatrix} \quad (6.2.4)$$

El modelo final muestra que las tres series comparten por un lado, una estructura autorregresiva de orden uno y por otro, los shocks producidos en los precios de las ciudades de Buffalo y Minneapolis, en el periodo anterior. En cuanto al valor de los parámetros, obsérvese que las estimaciones máximo-verosímiles son prácticamente iguales a las obtenidas por el algoritmo SUBEST2 (ver matrices (6.2.3) y (6.2.4)). Los resultados obtenidos al aplicar NID a los residuos proporcionados por la estimación (6.2.4) quedan resumidos en el Cuadro 6.2.3. Este, al igual que la matriz $\mathbf{Q}_{(7)}$ de estadísticos de Box-Ljung, no presentan ninguna señal de mala especificación en el modelo propuesto.

$$\mathbf{Q}_{(7)}(\%) = \begin{pmatrix} 3.5 & 3.3 & 3.2 \\ 4.6 & 4.3 & 4.9 \\ 6.9 & 5.6 & 4.0 \end{pmatrix} \quad (6.2.5)$$

Cuadro 6.2.3: Salida del algoritmo NID aplicado a $\hat{\mathbf{a}}_t^$*

Raíces unitarias						
0						
Resultado de los diferentes criterios						
n	AIC	SBC	HQ	SVC_{Ω_2}	$NIDC$	$\chi^2(95\%)$
0	-13.96	-13.80	-13.90	0.57	0.57	-
1	-13.87	-13.56	-13.75	0.75	0.68	0.09
2	-13.76	-13.29	-13.57	0.87	0.72	0.00
3	-13.68	-13.05	-13.42	1.09	0.88	0.01
4	-13.60	-12.82	-13.29	1.35	1.07	0.01
Orden del sistema estimado						
	0	0	0	0	0	0

* $\hat{\mathbf{a}}_t$ son los residuos del modelo (6.2.4)

Este ejemplo ilustra con sencillez la eficacia que el procedimiento propuesto ofrece para: i) identificar las raíces unitarias de un sistema, ii) identificar el orden del proceso o índice de McMillan, iii) especificar un juego de IK y estimar un modelo (con la consiguiente reducción de la complejidad), iv) depurar el modelo introduciendo restricciones de exclusión, y finalmente, v) estimar por MVE un modelo adecuado a partir de condiciones iniciales óptimas.

6.2.2. Series de tipos de interés de EEUU (Martín Manjón y Treadway, 1997)

En este apartado se analizan diversas series de tipos de interés de la economía de Estados Unidos entre los meses de agosto de 1985 y enero de 2003, un total de 210 observaciones, con el propósito de ilustrar la metodología. Martín Manjón y Treadway (1997) estudian la relación existente entre estas series con una muestra ligeramente más corta, encontrando que los tipos con un vencimiento inferior o igual a dos años, presentan una tendencia común a largo plazo con el tipo objetivo de la Reserva Federal (Fed). Por otro lado, los tipos con un vencimiento superior, muestran una tendencia común a largo plazo relacionada con una combinación lineal entre el tipo objetivo de la Fed y un tipo de interés con vencimiento a largo plazo (Martín Manjón, 1998). El ejemplo se divide en dos partes. La primera analiza, siguiendo los pasos de la metodología, el tipo objetivo de la Fed conjuntamente con el tipo de interés efectivo y los tipos a tres y seis meses del mercado secundario. En la segunda parte se estudia, siguiendo el mismo proceso, las relaciones existentes entre el tipo objetivo, el tipo de interés de los bonos a cinco años y el tipo de los bonos a treinta años.

Tipos a corto plazo:

Antes de estudiar conjuntamente los tipos de interés seleccionados, representados en la Figura 6.2.2, éstos se introducen individualmente en el algoritmo de identificación. El criterio de detección de raíces unitarias indica que todas las series son I(1). Efectivamente, los CCC para los tipos objetivo (TO_t), efectivo (TE_t), a 3 meses ($T3m_t$) y a 6 meses ($T6m_t$) son iguales a .995 en todos los casos. Como se demuestra en el Capítulo 3, cuando las series analizadas presentan un mismo orden de integración, es posible estudiar la existencia de cointegración a partir de la metodología presentada. Los resultados de aplicar el algoritmo NID al sistema multivariante se resumen en el Cuadro 6.2.4.

NID también devuelve una estimación de la matriz de cointegración Δ :

$$\hat{\Delta} = \begin{pmatrix} 1 & 0 & 0 & -1.01 \\ 0 & 1 & 0 & -0.92 \\ 0 & 0 & 1 & -0.93 \end{pmatrix} \quad (6.2.6)$$

de tal forma que $\hat{\Delta}\mathbf{Z}_{1t}$ son variables I(0), donde $\mathbf{Z}_{1t} = (TE_t \ T3m_t \ T6m_t \ TO_t)'$. Obsérvese que estas series generadas a partir de la relación de cointegración son aproximadamente el “spread” y representan una medida de la eficacia del control de tipos de interés a corto plazo.

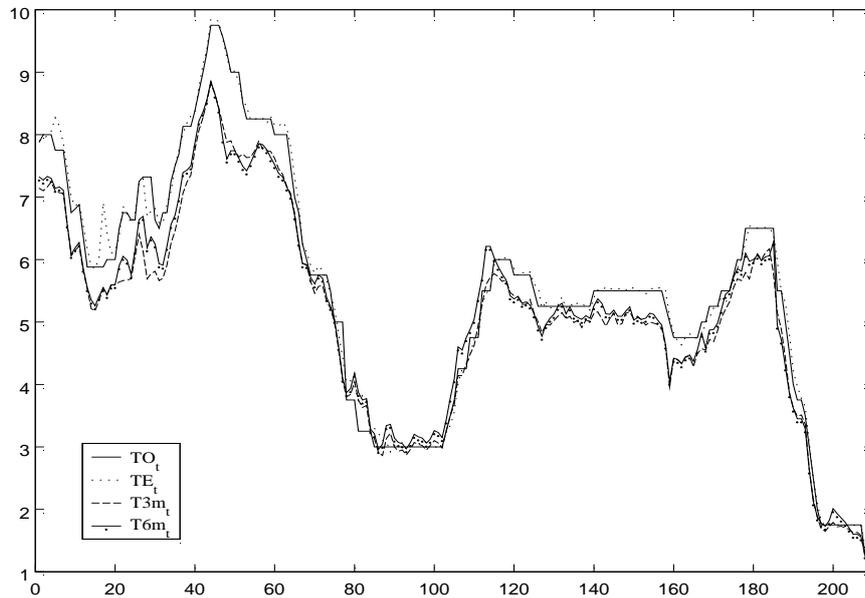


Figura 6.2.2: Tipos de interés a corto plazo. TO_t es el tipo objetivo, TE_t es el tipo efectivo y, $T3m_t$ y $T6m_t$ son los tipos a 3 y 6 meses respectivamente. La muestra analizada se extiende desde agosto de 1985 hasta enero de 2003.

Cuadro 6.2.4: Salida del algoritmo NID aplicado a \mathbf{Z}_{1t}^*

Raíces unitarias						
1						
Relaciones de Cointegración						
3						
Resultado de los diferentes criterios						
n	AIC	SBC	HQ	SVC_{Ω_2}	$NIDC$	$\chi^2(95\%)$
0	4.82	4.98	4.88	112.71	112.71	-
1	0.23	0.52	0.35	4.12	4.06	1.00
2	-1.22	-0.80	-1.05	2.35	2.23	1.00
3	-2.00	-1.46	-1.78	1.21	1.02	1.00
4	-2.18	-1.51	-1.90	1.26	1.01	1.00
Orden del sistema estimado						
	4	4	4	3	4	4

$$*\mathbf{Z}_{1t} = (TE_t \quad T3m_t \quad T6m_t \quad TO_t)'$$

El criterio sobre raíces unitarias detecta, en el sistema, la existencia de un CCC

suficientemente cercano a uno (el primero es igual a .995) y el resto diferentes de uno (los siguientes CCC son inferiores a .877). Esto conduce al estudio de la variable \mathbf{Y}_{1t} , definida como:

$$\mathbf{Y}_{1t} = \begin{pmatrix} \nabla & 0 & 0 & 0 \\ -1.01 & 1 & 0 & 0 \\ -0.92 & 0 & 1 & 0 \\ -0.93 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} TO_t \\ TE_t \\ T3m_t \\ T6m_t \end{pmatrix} \quad (6.2.7)$$

Los gráficos de las series transformadas y sus funciones de autocorrelación simples no presentan indicios de no estacionariedad. Además, al introducir \mathbf{Y}_{1t} de nuevo en el algoritmo NID, no se detectan raíces unitarias en el sistema y sí, en cambio, la existencia de estructura dinámica, como muestra el Cuadro 6.2.5.

*Cuadro 6.2.5: Salida del algoritmo NID aplicado a \mathbf{Y}_{1t} **

Raíces unitarias						
0						
Resultado de los diferentes criterios						
n	AIC	SBC	HQ	SVC_{Ω_2}	$NIDC$	$\chi^2(95\%)$
2	-4.04	-3.63	-3.87	2.41	2.39	1.00
3	-4.64	-4.10	-4.42	1.63	1.60	1.00
4	-4.88	-4.21	-4.61	1.52	1.47	1.00
5	-4.81	-4.01	-4.49	1.58	1.52	0.69
Orden del sistema estimado						
	4	4	4	4	4	4
* $\mathbf{Y}_{1t} \simeq (\nabla TO_t \quad TE_t - TO_t \quad T3m_t - TO_t \quad T6m_t - TO_t)'$						

A la vista de los resultados presentados en el Cuadro 6.2.5, se decide estimar un modelo con dinámica $\hat{n} = 4$. El algoritmo de especificación de los IK y la estimación mediante métodos de subespacios (SUBEST1), devuelve un juego de índices de observabilidad igual a (1,1,1,1), independientemente del criterio utilizado. Esto supone una estructura VARMA (1,1) “llena”, donde el juego de IK no implica ninguna restricción sobre los parámetros. En esta primera iteración, se encuentran doce elementos no significativamente distintos de cero que se eliminan en el siguiente paso con la estimación por el método SUBEST2. El resultado producido por ambos algoritmos queda resumido en el Cuadro 6.2.6. El valor de los parámetros cambia levemente acercándose, como se observará más adelante, a los valores máximo-verosímiles. Este segundo paso genera buenas condiciones iniciales para la estimación por MVE. En la tercera iteración, el cálculo de la matriz de

Cuadro 6.2.6: Secuencia de modelos identificados a partir de la metodología propuesta para el sistema \mathbf{Y}_{1t} . Entre paréntesis se presentan las desviaciones típicas de los parámetros. $\mathbf{Y}_{1t} \simeq (\nabla TO_t \quad TE_t - TO_t \quad T3m_t - TO_t \quad T6m_t - TO_t)'$.

Algoritmo	$\hat{\Phi}(B)$	$\hat{\Theta}(B)$
SUBEST1	$\begin{pmatrix} 1 - .49B & .19B & .45B & - .35B \\ (.08) & (.18) & (.15) & (.12) \\ .20B & 1 - .29B & .18B & .11B \\ (.04) & (.10) & (.08) & (.06) \\ .32B & .13B & 1 - .80B & - .10B \\ (.06) & (.14) & (.11) & (.09) \\ .23B & .0B & - .02B & 1 - .94B \\ (.08) & (.18) & (.15) & (.12) \end{pmatrix}$	$\begin{pmatrix} 1 + .04B & .36B & 1.36B & - .53B \\ (.08) & (.15) & (.16) & (.13) \\ .20B & 1 + .09B & - .32B & .20B \\ (.06) & (.10) & (.10) & (.09) \\ .40B & - .06B & 1 - .29B & - .06B \\ (.07) & (.12) & (.12) & (.11) \\ .35B & - .30B & - .02B & 1 - .15B \\ (.09) & (.16) & (.17) & (.14) \end{pmatrix}$
SUBEST2	$\begin{pmatrix} 1 - .50B & 0 & .40B & - .30B \\ (.12) & 1 - .54B & - .03B & 0 \\ .21B & 0 & 1 - .85B & 0 \\ (.18) & 0 & 0 & 1 - .86B \end{pmatrix}$	$\begin{pmatrix} 1 & .20B & 1.10B & - .28B \\ .21B & 1 & - .15B & - .01B \\ .34B & 0 & 1 - .18B & 0 \\ .38B & 0 & 0 & 1 \end{pmatrix}$
MVE	$\begin{pmatrix} 1 - .46B - .04B^2 - .19B^3 & 0 & .19B & - .10B \\ (.06) & (.05) & (.06) & (.05) \\ .16B & 1 - .66B & 0 & 0 \\ (.05) & (.08) & (-) & (-) \\ .19B & 0 & 1 - .87B & 0 \\ (.05) & (-) & (.03) & (-) \\ .23B & 0 & 0 & 1 - .89B \\ (.06) & (-) & (-) & (.03) \end{pmatrix}$	$\begin{pmatrix} 1 & .30B & .67B & 0 \\ (.08) & (.09) & (-) & (-) \\ .28B & 1 - .23B & 0 & 0 \\ (.08) & (.11) & (-) & (-) \\ .28B & 0 & 1 - .26B & 0 \\ (.07) & (-) & (.06) & (-) \\ .43B & 0 & 0 & 1 - .19B \\ (.08) & (-) & (-) & (.06) \end{pmatrix}$

información exacta junto con las estimaciones máximo-verosímiles, permiten realizar contrastes individuales con mayor precisión y eliminar ciertos parámetros no significativamente diferentes de cero que en la etapa anterior no se detectaron.

En la fase de diagnosis se encuentran e introducen en el modelo tres nuevos parámetros que resultan ser estadísticamente distintos de cero. El modelo final resulta un VARMA (3,1) restringido. La secuencia de estimaciones empleada en el análisis se expone en el Cuadro 6.2.6. La matriz de varianzas y covarianzas de los términos de error y la matriz de estadísticos de Box-Ljung son las siguientes,

$$\hat{\Sigma}_{a_1}(\%) = \begin{pmatrix} 3.33 & - & - & - \\ (.33) & & & \\ -1.59 & 2.30 & - & - \\ (.22) & (.23) & & \\ -1.33 & 1.35 & 2.84 & - \\ (.23) & (.20) & (.28) & \\ -.83 & 0.98 & 2.43 & 2.87 \\ (.22) & (.19) & (.26) & (.28) \end{pmatrix} \quad \mathbf{Q}(\tau)(\%) = \begin{pmatrix} 5.8 & 10.3 & 5.4 & 8.9 \\ 8.6 & 7.7 & 4.3 & 9.9 \\ 10.6 & 17.0 & 11.1 & 12.6 \\ 6.5 & 15.5 & 8.3 & 10.1 \end{pmatrix} \quad (6.2.8)$$

donde $\hat{\mathbf{a}}_{1t}$ son los residuos obtenidos al ajustar a los datos el modelo estimado por MVE dado en el Cuadro 6.2.6.

La matriz de estadísticos \mathbf{Q} , muestra algunos valores relativamente elevados. Como medida de diagnosis adicional, se aplica NID a los residuos del modelo final estimado por MVE. El Cuadro 6.2.7 muestra la salida del algoritmo.

Cuadro 6.2.7: Salida del algoritmo NID aplicado a $\hat{\mathbf{a}}_{1t}^*$

Raíces unitarias						
0						
Resultado de los diferentes criterios						
n	AIC	SBC	HQ	SVC_{Ω_2}	$NIDC$	$\chi^2(95\%)$
0	-4.95	-4.79	-4.89	0.81	0.81	-
1	-4.94	-4.65	-4.88	0.93	0.92	0.91
2	-5.00	-4.58	-4.82	0.93	0.91	0.68
3	-4.94	-4.40	-4.77	1.10	1.06	0.64
4	-4.95	-4.28	-4.65	1.19	1.15	0.12
Orden del sistema estimado						
	2	0	2	0	0	0

* $\hat{\mathbf{a}}_{1t}$, residuos del modelo estimado por MVE en (6.2.6)

Como cabía esperar, los criterios AIC y HQ (también el test χ^2 al 91%) del

algoritmo de identificación señalan la posibilidad de que exista aún estructura dinámica. En cambio, el resto de criterios no la reflejan. Cabe destacar que en este ejercicio no se realiza análisis de intervención, puesto que no es éste el objetivo del estudio y como es habitual en las series financieras algunos anómalos no corregidos pueden alterar el valor del estadístico y de los criterios de selección. En cualquier caso, la incorporación de nuevos parámetros al modelo final no consiguen captar estructura adicional.

Como resumen, este segundo ejemplo multivariante ilustra la capacidad del algoritmo de identificación para detectar y calcular la matriz de cointegración existente en un conjunto de series I(1). El número de relaciones de cointegración y sus estimaciones coinciden con los resultados de Martín Manjón y Treadway (1997), si bien hay que tener en cuenta que la muestra no es exactamente comparable. La identificación inicial del modelo cointegrado discrepa levemente de la del modelo final, probablemente debido a la poca persistencia de los autorregresivos de orden dos y tres del elemento (1,1) de la matriz $\hat{\Phi}$. Sin embargo, las dos primeras iteraciones con los algoritmos SUBEST1 y SUBEST2 permiten: i) una especificación del modelo en términos de IK, ii) una notable reducción del número de parámetros, y iii) unas óptimas condiciones iniciales para la estimación por MVE. En suma, una buena aproximación al modelo final, a un muy reducido coste computacional y de esfuerzo del analista.

Tipos a largo plazo:

Al igual que les ocurre a Martín Manjón y Treadway (1997), al aplicar NID al sistema formado por las series de tipos de bonos a cinco años ($T5a_t$) y el objetivo de la reserva federal, no se encuentra ninguna relación de cointegración y sí en cambio, dos raíces unitarias. En efecto, el análisis de correlaciones canónicas del sistema nos ofrece dos coeficientes superiores a .965, con lo que ambos son considerados suficientemente cercanos a la unidad por el criterio propuesto. Por esto, la segunda parte del ejemplo aborda el estudio de $T5a_t$ y TO_t junto a la serie de tipos de interés de los bonos a treinta años $T30a_t$ (vector de series denotado por Z_{2t}) tal y como realizan los autores mencionados.

Al introducir el sistema Z_{2t} en el algoritmo de identificación, el proceso encuentra dos raíces unitarias y una relación de cointegración, como muestra el Cuadro 6.2.8. Además, calcula el vector de cointegración como:

$$\hat{\Delta} = (1 \quad -.30 \quad -.68) \tag{6.2.9}$$

El proceso identifica de esta forma, que la serie $T5a_t$ opera en una relación de cointegración CI(1,1) con una combinación lineal entre las series TO_t y $T30a_t$. El

valor de los tres CCC del sistema \mathbf{Z}_{2t} es .998, .995 y .854, por lo que tan sólo los dos primeros corresponden a raíces unitarias. El análisis se centra entonces en la variable transformada:

$$\mathbf{Y}_{2t} = \begin{pmatrix} \nabla & 0 & 0 \\ 0 & \nabla & 0 \\ -0.30 & -0.68 & 1 \end{pmatrix} \begin{pmatrix} TO_t \\ T30a_t \\ T5a_t \end{pmatrix} \quad (6.2.10)$$

*Cuadro 6.2.8: Salida del algoritmo NID aplicado a \mathbf{Z}_{2t}^**

Raíces unitarias						
2						
Relaciones de Cointegración						
1						
Resultado de los diferentes criterios						
n	AIC	SBC	HQ	SVC_{Ω_2}	$NIDC$	$\chi^2(95\%)$
0	11.38	11.47	11.42	106.27	106.27	-
1	6.80	7.00	6.88	25.15	25.10	1.00
2	2.87	3.16	2.99	3.74	3.65	1.00
3	1.46	1.85	1.62	0.84	0.71	1.00
4	1.51	1.99	1.70	0.77	0.58	0.03
Orden del sistema estimado						
	3	3	3	4	4	3

$$*\mathbf{Z}_{2t} = (T5a_t \quad TO_t \quad T30a_t)'$$

Como puede observarse en la Figura 6.2.2, los gráficos de los niveles del conjunto de series \mathbf{Y}_{2t} no reflejan síntomas de no estacionariedad, además NID tampoco detecta la existencia de raíces unitarias en el sistema (ver Cuadro 6.2.9).

La identificación del modelo revela un orden del sistema entre 2 y 3. Se decide estimar un modelo con el orden del proceso que más veces se repite, $\hat{n} = 2$, y obtener un modelo canónico tal y como se propone en la metodología. Al aplicar SUBEST1, se estima el juego de IK (1,0,1), con lo que el algoritmo devuelve un modelo VARMA (1,1) echelon. Las matrices estimadas en cada etapa se presentan en el Cuadro 6.2.10.

En la primera fase, puede observarse cómo la estructura de índices de observabilidad restringe a cero ciertos parámetros, mientras que otros resultan no

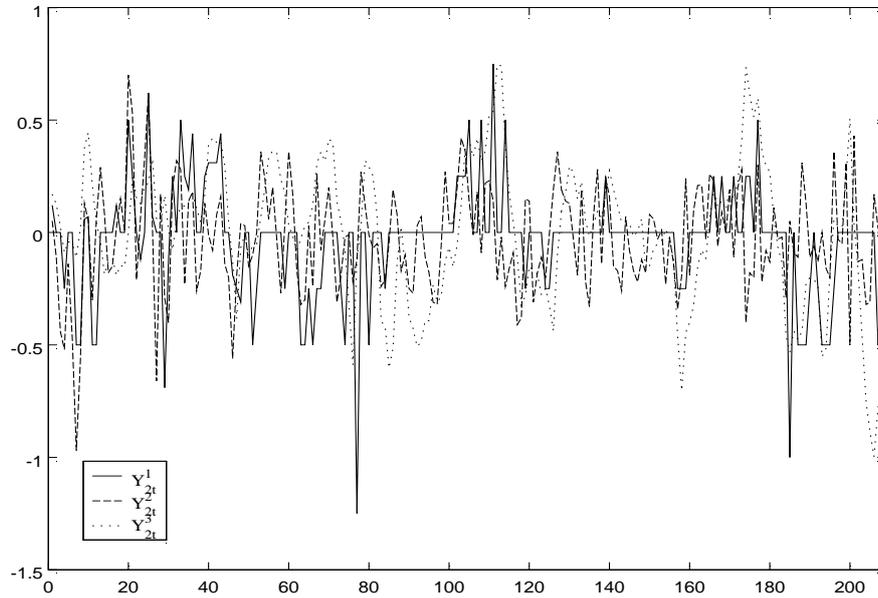


Figura 6.2.3: Transformación estacionaria de los datos. $Y_{2t}^1 = \nabla TO_t$, $Y_{2t}^2 = \nabla T30a_t$ e $Y_{2t}^3 = (T5a_t - .30TO_t - .68T30a_t)$, donde TO_t es el tipo objetivo y $T5a_t$ y $T30a_t$ son los tipos a 5 y 30 años respectivamente.

Cuadro 6.2.9: Salida del algoritmo NID aplicado a \mathbf{Y}_{2t}^*

Raíces unitarias						
0						
Resultado de los diferentes criterios						
n	AIC	SBC	HQ	SVC_{Ω_2}	NIDC	$\chi^2(95\%)$
0	0.25	0.34	0.28	6.21	6.21	-
1	-1.61	-1.42	-1.54	0.85	0.80	1.00
2	-1.75	-1.46	-1.64	0.71	0.62	1.00
3	-1.78	-1.40	-1.63	0.77	0.63	1.00
4	-1.76	-1.28	-1.57	0.89	0.71	0.87
Orden del sistema estimado						
	3	2	2	2	2	3

* $\mathbf{Y}_{2t} = (\nabla TO_t \quad \nabla T30a_t \quad T5a_t - .30TO_t - .68T30a_t)'$

estadísticamente distintos de cero, con lo que disminuye notablemente la parametrización respecto de, por ejemplo, una representación VARMA(1,1). Así en la segunda iteración, el modelo estimado mediante el algoritmo SUBEST2 presenta

Cuadro 6.2.10: Secuencia de modelos identificados a partir de la metodología propuesta para el sistema \mathbf{Y}_{2t} . $\mathbf{Y}_{2t} = (\nabla TO_t \quad \nabla T30a_t \quad T5a_t - .30TO_t - .68T30a_t)'$. Entre paréntesis se presentan las desviaciones típicas de los coeficientes estimados.

Algoritmo	$\hat{\Phi}(B)$	$\hat{\Theta}(B)$
SUBEST1	$\begin{pmatrix} 1 - .80B & 0 & -.01B \\ (.11) & (-) & (.05) \\ -.02 & 1 & 0 \\ (.14) & (-) & (-) \\ .08B & 0 & 1 - .82B \\ (.15) & (-) & (.08) \end{pmatrix}$	$\begin{pmatrix} 1 - .69B & .13B & .39B \\ (.10) & (.04) & (.08) \\ -.02 & 1 & 0 \\ (.14) & (-) & (-) \\ -.08B & .10B & 1 + .33B \\ (.04) & (.04) & (.08) \end{pmatrix}$
SUBEST2	$\begin{pmatrix} 1 - .82 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 - .76B \end{pmatrix}$	$\begin{pmatrix} 1 - .69B & .08B & .35B \\ 0 & 1 & 0 \\ 0 & .12B & 1 + .28B \end{pmatrix}$
MVE	$\begin{pmatrix} 1 - .26B - .02B^2 - .38B^3 & 0 & 0 \\ (.06) & (-) & (-) \\ 0 & 1 - .28B + .15B^2 - .14B^3 & 0 \\ (-) & (.06) & (-) \\ 0 & 0 & 1 - .81B \\ (-) & (-) & (.04) \end{pmatrix}$	$\begin{pmatrix} 1 & .16B & .48B \\ (-) & (.07) & (.10) \\ 0 & 1 & 0 \\ (-) & (-) & (-) \\ 0 & .14B & 1 + .27B \\ (-) & (.04) & (.07) \end{pmatrix}$

tan sólo siete parámetros, exceptuando los contenidos en la matriz de varianzas y covarianzas de los términos de error. En el último paso, la estimación por MVE, se elimina un parámetro más y se introducen cinco, de los que tres corresponden al elemento (2,2) de la matriz autorregresiva $\hat{\Phi}$. La razón por la cual ni el algoritmo de identificación, ni SUBEST1, detectan estos últimos parámetros incluidos en el modelo se debe principalmente a la débil persistencia que presentan. Tanto es así, que los residuos del modelo final excluyendo los parámetros ϕ_2 y ϕ_3 del elemento (2,2) de $\hat{\Phi}$, no evidencian una mala especificación. Estos parámetros tan sólo han sido detectados y añadidos tras un proceso de sobreajuste.

El modelo final genera la matriz de covarianzas $\hat{\Sigma}_{\alpha_2}$, y los residuos la matriz de estadísticos Q :

$$\hat{\Sigma}_{\alpha_2}(\%) = \begin{pmatrix} 3.90 & - & - \\ (.38) & & \\ 1.67 & 5.08 & - \\ (.31) & (.49) & \\ 0 & 1.13 & 1.88 \\ (-) & (.21) & (.18) \end{pmatrix} \quad Q_{(7)}(\%) = \begin{pmatrix} 10.2 & 12.3 & 6.6 \\ 4.9 & 4.6 & 5.9 \\ 7.6 & 14.2 & 12.3 \end{pmatrix} \quad (6.2.11)$$

Cuadro 6.2.11: Salida del algoritmo NID aplicado a $\hat{\alpha}_{2t}$ *

Raíces unitarias						
0						
Resultado de los diferentes criterios						
n	AIC	SBC	HQ	SVC_{Ω_2}	$NIDC$	$\chi^2(95\%)$
0	-1.96	-1.86	-1.92	0.42	0.42	-
1	-1.96	-1.77	-1.88	0.46	0.41	0.37
2	-1.95	-1.67	-1.84	0.53	0.44	0.18
3	-1.91	-1.53	-1.76	0.64	0.50	0.06
4	-1.87	-1.39	-1.67	0.73	0.54	0.11
Orden del sistema estimado						
	1	0	0	0	1	0

* $\hat{\alpha}_{2t}$, residuos del modelo estimado por MVE del Cuadro (6.2.10).

Como ocurre en el caso anterior, tanto la Q de Box-Ljung como el algoritmo NID aplicado a $\hat{\alpha}_{2t}$ pueden sugerir la existencia de estructura dinámica adicional. Sin embargo, un análisis detallado de los residuos y de sus funciones de correlación cruzadas no refleja evidencias de este hecho. Por tanto, se considera el modelo estimado por MVE del Cuadro 6.2.10, como adecuado para representar al sistema Y_{2t} .

La relevancia de este ejemplo es mostrar el comportamiento del algoritmo NID ante relaciones de cointegración con dos factores comunes no estacionarios. El criterio de discriminación de raíces unitarias y el cálculo de la matriz de cointegración ofrecen un resultado positivo frente a este tipo de modelos. Asimismo, la identificación de las series estacionarias transformadas también genera buenos resultados, aunque la debilidad de algunos parámetros puede infraidentificar los modelos en la primera etapa del proceso. Si este es el caso, en la fase de estimación mediante MVE y de diagnosis pueden detectarse nuevos parámetros, disminuyendo el riesgo de infraparametrización.

Por otro lado, las restricciones impuestas por la estructura de los IK, la depuración de parámetros en la primera iteración junto con la estimación mediante SUBEST2 permite, además de obtener buenas condiciones iniciales para MVE, una reducción de la parametrización del modelo y, por tanto, un aumento de los grados de libertad que ayudan a evitar los frecuentes problemas de convergencia de la estimación máximo-verosímil en modelos multivariantes complejos.

Capítulo 7

Conclusiones

En este trabajo se desarrolla una metodología completa de construcción de modelos dinámicos lineales de coeficientes fijos, mediante métodos de subespacios. Estos métodos estiman de manera no iterativa las matrices de un proceso en forma de Espacio de los Estados (EE). El procedimiento propuesto distingue entre el análisis [ver Figura 7.1]: 1) de procesos univariantes, con o sin estacionalidad y, 2) de series múltiples no estacionales. Aunque el método de análisis de series multivariantes no requiere ninguna información preliminar, a menudo resulta útil analizar previamente cada serie por separado. Esto conduce en la mayoría de los casos a una interpretación adecuada de los resultados obtenidos sobre raíces unitarias y cointegración del sistema conjunto.

El método de análisis univariante se divide en las siguientes etapas: (i) determinar el orden de integración (tanto regular como estacional), (ii) determinar los órdenes regular y estacional o dinámica de la variable y (iii) añadir, si se desea, restricciones de exclusión sobre la dinámica encontrada en la etapa previa [ver Figuras 7.4 y 7.5].

A su vez, el análisis de series múltiples consta de las siguientes fases: (i) determinar el orden de integración del proceso y, con ello, el rango y la matriz de cointegración (si existieran), (ii) determinar el orden de dinámica del sistema, (iii) especificar los índices de Kronecker, que definen cómo se distribuye la dinámica del sistema entre sus distintas series temporales, y (iv) incluir, si es necesario, restricciones de exclusión en el modelo especificado y estimado en la fase anterior [ver Figuras 7.2 y 7.3].

Las principales ventajas de la metodología propuesta frente a otras alternativas se resumen en que:

1. Se ofrece un procedimiento completo de análisis de series temporales en el que, partiendo exclusivamente de los datos, es posible obtener una representación final de la transformación estacionaria del proceso.
2. La estructura de la metodología y las propiedades de los algoritmos de subespacios facilitan su uso para especificar modelos econométricos dinámicos de forma automática o semi-automática. Esto permite su aplicación al análisis masivo de series.
3. El método proporciona una secuencia de modelos, estadísticamente adecuados y cada vez más simples, permitiendo que el analista decida en qué modelo o especificación desea parar. Recorrer más o menos fases proporciona una ganancia en términos de parsimonia, pero aumenta el coste del análisis y el riesgo de tomar decisiones erróneas. Este enfoque permite buscar, en función de los objetivos del estudio, un equilibrio idóneo entre esfuerzo, parsimonia y riesgo de error.
4. Las técnicas de subespacios utilizadas (ver Van Overschee y De Moor, 1996; Bauer, 1998; Favoreel et al., 2000) son computacionalmente eficientes y utilizan herramientas del álgebra lineal numéricamente robustas.

La contribución global del trabajo consiste, por tanto, en unificar distintos métodos de identificación y estimación mediante subespacios, algunos ya existentes (Casals, 1997; Bauer, 2001) y otros desarrollados en la Tesis, para formular una metodología completa y graduable para el análisis empírico de series temporales. Instrumentar la metodología para su uso práctico ha requerido derivar en este trabajo:

1. Un procedimiento de identificación de raíces unitarias y del rango de cointegración a partir del estudio empírico de las propiedades de los coeficientes de correlación canónica y la estimación de funciones de penalización. Los resultados de los ejercicios de simulación reflejan la capacidad de los criterios propuestos para estimar consistentemente el rango de cointegración en diferentes situaciones. La metodología, al contrario que otras en la literatura (Johansen, 1988, 1991), no requiere estimar ningún modelo a priori.
2. Un estimador consistente de la matriz de cointegración. Este se obtiene a partir de un modelo formulado en espacio de los estados en forma de innovaciones con la única información de los datos y del orden del sistema. Los resultados de las simulaciones muestran el buen comportamiento de los criterios de detección de raíces unitarias, del cálculo del rango de cointegración y del estimador de la matriz de cointegración propuestos.

3. Dos nuevos criterios de identificación del índice de McMillan basados en el Criterio de Valores Singulares SVC (Bauer, 2001) y, por tanto, en los coeficientes de correlación canónica. Se demuestra analíticamente la consistencia de estos criterios y los resultados de los experimentos de Monte Carlo confirman esta propiedad. Además, ambos procedimientos presentan un coste computacional muy pequeño.
4. Un método rápido de especificación de los índices de Kronecker, condicionado a la estimación previa del orden del sistema. El hecho de restringir la elección de los índices de observabilidad a la estimación del orden del proceso disminuye el espacio de búsqueda pero también el riesgo de error, ya que la primera decisión es, en principio, menos compleja que la segunda.
5. Un algoritmo de estimación consistente, mediante métodos de subespacios, que permite introducir restricciones de exclusión de parámetros. Los resultados de los ejercicios de simulación muestran que las estimaciones resultantes podrían usarse, dependiendo de los modelos y de la longitud de las muestras, como estimaciones finales o preestimaciones consistentes para una estimación posterior por máxima verosimilitud. En términos relativos, el procedimiento presenta una importante ventaja computacional sobre un algoritmo de máxima verosimilitud exacta.
6. Un análisis empírico de los distintos procedimientos y criterios propuestos usando ejercicios de Monte Carlo y modelizando series reales de uso habitual en la literatura, como las utilizadas en Burr (1976), Box y Jenkins (1976), Tiao y Tsay (1989), Wei (1990) y otros.

Asimismo, las extensiones que prevemos realizar en sucesivos trabajos consisten en:

- Estudiar la cointegración estacional (Hyllberg et al., 1990; Engle et al., 1993). A partir del trabajo realizado en el Capítulo 3, surge la posibilidad de estudiar las propiedades de los coeficientes de correlación canónica entre la información pasada y futura existente en la matriz Block Hankel Estacional y extender así la investigación a las relaciones de cointegración estacional.
- Refinar el tratamiento de vectores de series estacionales. Utilizando la descomposición de un proceso en un subsistema multivariante estacional y otro multivariante regular es posible desarrollar un método similar al expuesto en el Capítulo 4, para especificar los índices de Kronecker del subsistema estacional.

- Adaptar la metodología para especificar modelos canónicos de heteroscedasticidad condicional (Engle, 1982; Bollerslev, 1986). La posibilidad de aplicar estos métodos a modelos multivariantes de heteroscedasticidad condicional puede mejorar el condicionamiento de la función de verosimilitud (un problema conocido en este tipo de modelos (Jerez et al., 1999)) debido a la significativa reducción de parámetros. Además, la adaptación de los algoritmos propuestos al análisis de, por ejemplo, modelos GARCH multivariantes, reducirá sensiblemente el coste computacional en la especificación y estimación de este tipo de modelos.

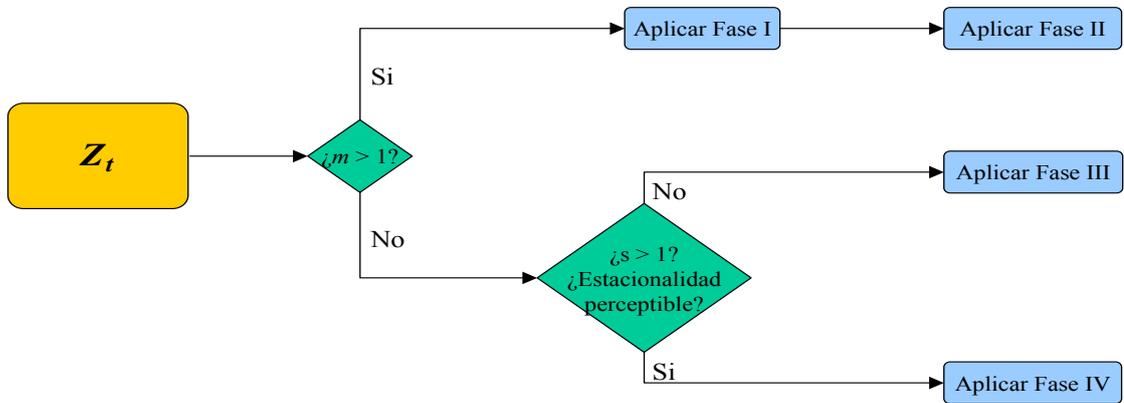


Figura 7.1: Fase Principal. Z_t es un conjunto de series temporales, m es la dimensión de este conjunto, y s es su frecuencia anual. Cada Fase hace referencia a los pasos de la metodología, esquematizados en las Figuras 7.2-7.5.

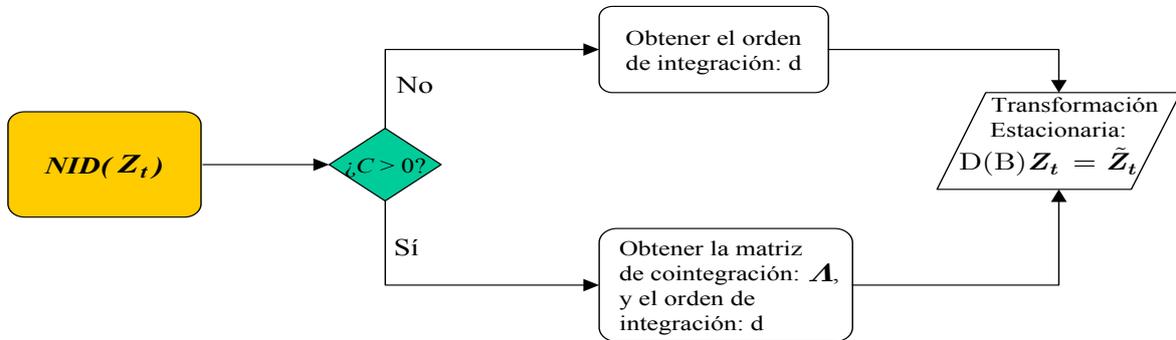


Figura 7.2: Fase I, transformaciones estacionarias. C es el rango de cointegración, \tilde{Z}_t es un conjunto de transformaciones estacionarias y $D(B)$ es una matriz que depende de d y de Λ .

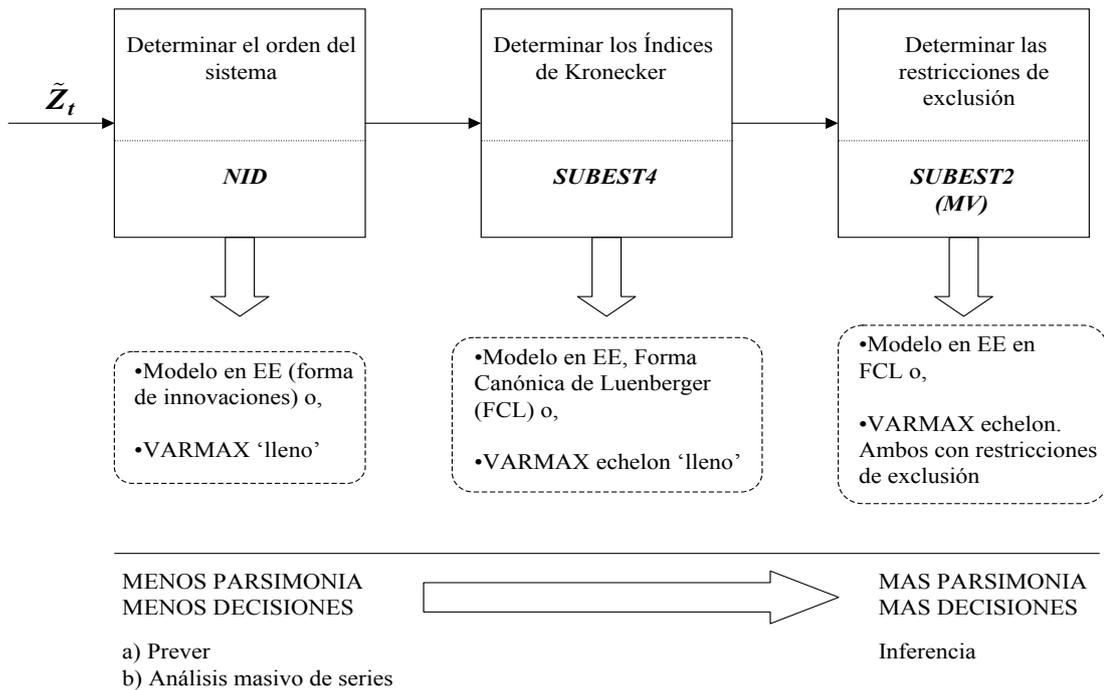


Figura 7.3: Fase II, procesos multivariantes.

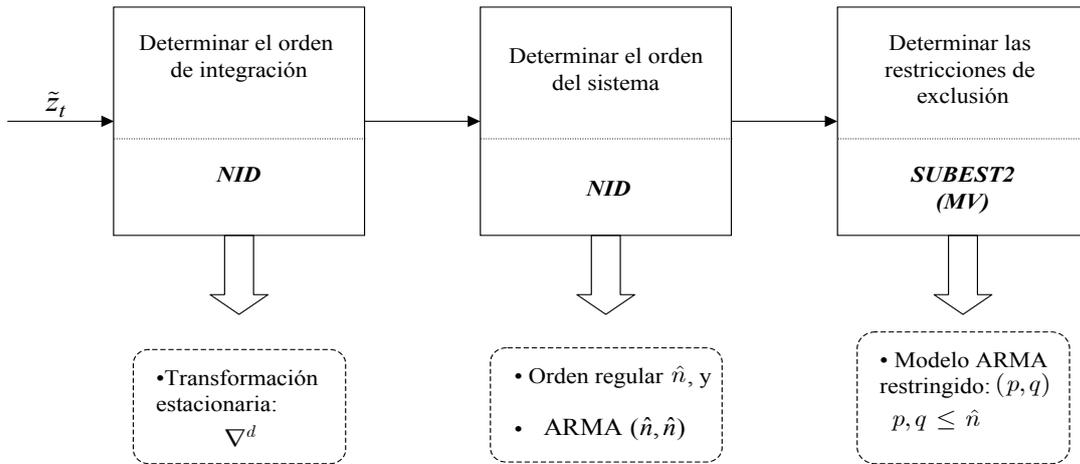


Figura 7.4: Fase III, procesos univariantes no estacionales. En este caso, $\hat{n} = \hat{n}_r$.

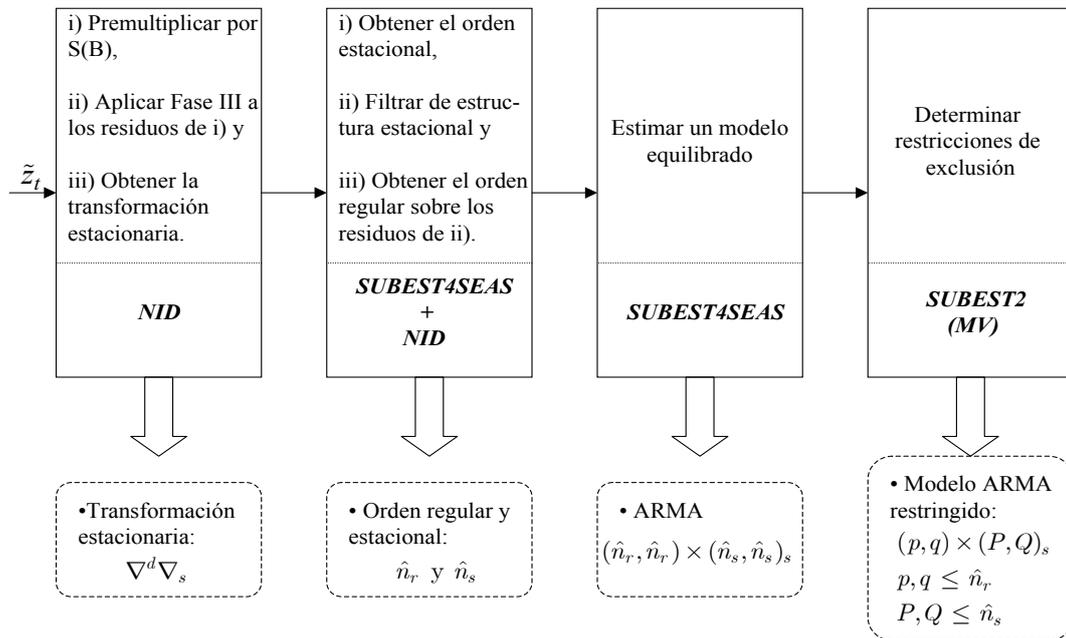


Figura 7.5: Fase IV, procesos univariantes estacionales. La suma estacional se define como $S(B) = 1 + B + \dots + B^{s-1}$.

Apéndice A

Apéndice Matemático

Este apéndice consta de tres apartados. En el primero se introducen definiciones de subespacios y algunas matrices no usuales en el campo econométrico. En el segundo y el tercero se establecen dos conceptos que se utilizan de forma continuada en el trabajo: la descomposición en valores singulares y los coeficientes de correlación canónica, poniendo de manifiesto la relación existente entre ambos.

A.1. Definiciones de Subespacios y Matrices Especiales

Sea $\mathbf{A} \in \mathbb{R}^{m \times n}$ una matriz real, entonces \mathbf{A} es un operador lineal que proyecta el espacio \mathbb{R}^n en \mathbb{R}^m , tal que $\forall \mathbf{x} \in \mathbb{R}^n$, $\mathbf{Ax} \in \mathbb{R}^m$. De la misma forma, se tiene que $\forall \mathbf{y} \in \mathbb{R}^m$, $\mathbf{A}'\mathbf{y} \in \mathbb{R}^n$. El conjunto de vectores \mathbf{Ax} , $\forall \mathbf{x} \in \mathbb{R}^n$ forman un subespacio en \mathbb{R}^m igual al número de columnas linealmente independientes de la matriz \mathbf{A} . Este subespacio se denomina “espacio de columnas” de \mathbf{A} y se denota, $im(\mathbf{A})$. Otros subespacios de interés asociados a la matriz \mathbf{A} son:

1. Espacio nulo de \mathbf{A} ($nul(\mathbf{A})$): Subespacio $S \subset \mathbb{R}^n$ tal que $\forall \mathbf{x} \in S$, $\mathbf{Ax} = \mathbf{0}$
2. Espacio de filas de \mathbf{A} ($im(\mathbf{A}')$): Subespacio en \mathbb{R}^n que se corresponde con el espacio de columnas de \mathbf{A}'
3. Espacio ortogonal al espacio de columnas de \mathbf{A} ($im(\mathbf{A})^\perp$): Subespacio en \mathbb{R}^m tal que $\forall \mathbf{y} \in im(\mathbf{A})$, $\mathbf{z} \in im(\mathbf{A})^\perp$, $\mathbf{y}'\mathbf{z} = \mathbf{0}$
4. Espacio ortogonal al espacio nulo de \mathbf{A} ($nul(\mathbf{A}^\perp)$): Subespacio en \mathbb{R}^n tal que $\forall \mathbf{x} \in nul(\mathbf{A})$, $\mathbf{y} \in nul(\mathbf{A}^\perp)$, $\mathbf{x}'\mathbf{y} = \mathbf{0}$

A continuación se definen algunas matrices usadas en este trabajo:

- Matriz ortogonal: Una matriz $\mathbf{A} \in \mathbb{R}^{m \times m}$ es ortogonal si $\mathbf{A}'\mathbf{A} = \mathbf{A}\mathbf{A}' = \mathbf{C}$, donde \mathbf{C} es una matriz diagonal.
- Matriz Hankel: Una matriz Hankel de dimensión $m \times n$ se construye a partir de $m + n - 1$ elementos $(y_1, y_2, \dots, y_{m+n-1})$, ordenados de la forma siguiente:

$$\mathbf{H} = \begin{pmatrix} y_1 & y_2 & \cdots & y_n \\ y_2 & y_3 & \cdots & y_{n+1} \\ \vdots & \vdots & & \vdots \\ y_m & y_{m+1} & \cdots & y_{m+n-1} \end{pmatrix} \quad (\text{A.1.1})$$

- Matriz Toeplitz: Una matriz Toeplitz de dimensión $m \times n$ se construye a partir de $m + n - 1$ elementos ordenados de la forma siguiente:

$$\mathbf{T} = \begin{pmatrix} y_n & y_{n-1} & \cdots & y_1 \\ y_{n+1} & y_n & \cdots & y_2 \\ \vdots & \vdots & & \vdots \\ y_{n+m-1} & y_{n+m-2} & \cdots & y_m \end{pmatrix} \quad (\text{A.1.2})$$

A.2. Descomposición en Valores Singulares (SVD)

Teorema: Si $\mathbf{A} \in \mathbb{R}^{m \times n}$ es una matriz real, existen dos matrices ortogonales,

$$\mathbf{U} = [\mathbf{u}_1, \dots, \mathbf{u}_m] \in \mathbb{R}^{m \times m} \quad \text{y} \quad \mathbf{V} = [\mathbf{v}_1, \dots, \mathbf{v}_n] \in \mathbb{R}^{n \times n}$$

tal que

$$\mathbf{U}'\mathbf{A}\mathbf{V} = \mathbf{S} = \text{diag}(\sigma_1, \dots, \sigma_p) \in \mathbb{R}^{m \times n}, \quad \text{con} \quad p = \min\{m, n\}$$

donde $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_p \geq 0$.

Demostración: Ver Golub y Van Loan (1996, Teorema 2.5.2)

Se denomina σ_i con $i = 1, \dots, p$ a los valores singulares de \mathbf{A} mientras que los vectores \mathbf{u}_i y \mathbf{v}_i son los i -ésimos vectores singulares izquierdo y derecho de \mathbf{A} , respectivamente. Algunas de las propiedades más interesantes de la descomposición SVD son las siguientes:

1. Rango (rg) y valores singulares: Si el $rg(\mathbf{A}) = k \leq p$ entonces se cumple que $rg(\mathbf{A}) = rg(\mathbf{U}\mathbf{S}\mathbf{V}') = rg(\mathbf{S}_k) = k$ donde $\mathbf{S}_k = \text{diag}(\sigma_1, \dots, \sigma_k)$ y $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_k > 0$. Así, el rango de \mathbf{A} coincide con su número de valores singulares no nulos.
2. Matriz pseudo-inversa de Moore-Penrose: La matriz $\mathbf{A}^+ = \mathbf{V}_k \mathbf{S}_k^{-1} \mathbf{U}_k'$ es la pseudo inversa de Moore Penrose de \mathbf{A} con $rg(\mathbf{A}) = k \leq p$. \mathbf{A}^+ satisface las siguientes propiedades:
 - a) $(\mathbf{A}^+)^+ = \mathbf{A}$
 - b) $\mathbf{A}\mathbf{A}^+\mathbf{A} = \mathbf{A}$
 - c) $\mathbf{A}^+\mathbf{A}\mathbf{A}^+ = \mathbf{A}^+$
 - d) $(\mathbf{A}^+)' = (\mathbf{A}')^+$
 - e) $\mathbf{A}^+ = (\mathbf{A}'\mathbf{A})^+\mathbf{A}'$
 - f) Si $rg(\mathbf{A}) = n \leq m$, entonces $\mathbf{A}^+ = (\mathbf{A}'\mathbf{A})^{-1}$ y $\mathbf{A}^+\mathbf{A} = \mathbf{I}_n$

A.3. Coeficientes de Correlación Canónica (CCC)

Sean \mathbf{x} e \mathbf{y} dos vectores de variables aleatorias de dimensiones m y n , respectivamente, con $m \geq n$, $\text{var}(\mathbf{x}) = \Sigma_{\mathbf{xx}} > \mathbf{0}$, $\text{var}(\mathbf{y}) = \Sigma_{\mathbf{yy}} > \mathbf{0}$, y $\text{cov}(\mathbf{x}, \mathbf{y}) = \Sigma_{\mathbf{xy}}$. Las n correlaciones canónicas entre \mathbf{x} e \mathbf{y} , denotadas por ρ_i , $i = 1, \dots, n$, se definen como:

$$\begin{aligned} \rho_k(\mathbf{x}, \mathbf{y}) &= \max_{\substack{\alpha_k \in \mathbb{R}^m \\ \beta_k \in \mathbb{R}^n}} \text{cov}(\alpha_k' \mathbf{x}, \beta_k' \mathbf{y}) & \text{(A.3.1)} \\ \text{s.a.} \quad & \text{var}(\alpha_k' \mathbf{x}) = 1, \quad \text{var}(\beta_k' \mathbf{y}) = 1 \\ & \text{cov}(\alpha_k' \mathbf{x}, \alpha_i' \mathbf{x}) = 0, \quad \text{cov}(\beta_k' \mathbf{y}, \beta_i' \mathbf{y}) = 0, \quad \forall i = 1, \dots, k-1 \end{aligned}$$

Como es lógico, se cumple que $\rho_1 \geq \rho_2 \geq \dots \geq \rho_n \geq 0$. La primera correlación canónica ρ_1 , es la correlación máxima entre cualquier combinación lineal de \mathbf{x} e \mathbf{y} con varianza unitaria. Los vectores $\alpha_1' \mathbf{x}$ y $\beta_1' \mathbf{y}$ se denominan “variables canónicas correspondientes a ρ_1 ”. La segunda correlación canónica ρ_2 , representa la correlación máxima entre combinaciones lineales de \mathbf{x} e \mathbf{y} incorreladas con las variables canónicas anteriores, y así sucesivamente.

Proposición: Los coeficientes de correlación canónica entre \mathbf{x} e \mathbf{y} corresponden a los valores singulares de la matriz,

$$\Sigma_{\mathbf{xx}}^{-\frac{1}{2}} \Sigma_{\mathbf{xy}} \Sigma_{\mathbf{yy}}^{-\frac{1}{2}} = \mathbf{U}_n \mathbf{S}_n \mathbf{V}_n' \quad \text{(A.3.2)}$$

con $(\alpha_1 \ \alpha_2 \ \dots \ \alpha_n) = \Sigma_{\mathbf{xx}}^{-\frac{1}{2}} \mathbf{U}_n$ y $(\beta_1 \ \beta_2 \ \dots \ \beta_n) = \Sigma_{\mathbf{yy}}^{-\frac{1}{2}} \mathbf{V}_n$.

Demostración: Ver Casals (1997, Proposición A.9.2)

En la práctica se utiliza la información muestral de $\mathbf{X} \in \mathbb{R}^{N \times m}$ e $\mathbf{Y} \in \mathbb{R}^{N \times n}$ para estimar las siguientes matrices de momentos muestrales,

$$\hat{\Sigma}_{\mathbf{xx}} = \frac{\mathbf{X}'\mathbf{X}}{N} > 0, \quad \hat{\Sigma}_{\mathbf{yy}} = \frac{\mathbf{Y}'\mathbf{Y}}{N} > 0, \quad \hat{\Sigma}_{\mathbf{xy}} = \frac{\mathbf{X}'\mathbf{Y}}{N}$$

y, por tanto, los coeficientes de correlación canónica entre \mathbf{X} e \mathbf{Y} se corresponden con los valores singulares de la matriz $(\mathbf{X}'\mathbf{X})^{-\frac{1}{2}} \mathbf{X}'\mathbf{Y} (\mathbf{Y}'\mathbf{Y})^{-\frac{1}{2}} = \mathbf{U}_n \mathbf{S}_n \mathbf{V}_n'$.

Asimismo, pueden definirse las variables canónicas como las columnas de las matrices,

$$\mathbf{C}_x = \mathbf{X} \hat{\Sigma}_{\mathbf{xx}}^{-\frac{1}{2}} \mathbf{U}_n \quad \text{y} \quad \mathbf{C}_y = \mathbf{Y} \hat{\Sigma}_{\mathbf{yy}}^{-\frac{1}{2}} \mathbf{V}_n$$

Apéndice B

Apéndice de Algoritmos

En esta sección se realiza una descripción detallada de los algoritmos implementados para la realización de la Tesis. También se incluyen, a modo de ejemplo, unas líneas de código para simular y visualizar las salidas que generan.

B.1. Algoritmo NID

Propósito

Calcula una rápida estimación del orden del sistema o índice de McMillan, las matrices estimadas de un modelo en EE con este orden de dinámica, el número de raíces unitarias y, si existen, el rango y la matriz de cointegración de un vector de series.

Synopsis

$[tab, nopts, ur, coinvec, theta, din] = nid(y, nvec, s, q, p)$

Descripción

Esta función devuelve estimaciones consistentes del orden del proceso, los parámetros estimados de un modelo en EE a partir del orden del sistema que más se repite en los diferentes métodos empleados, el número de raíces unitarias y el rango y la matriz de cointegración de un vector de series. Primero calcula los coeficientes de correlación canónica entre el pasado y el futuro de la serie. Después se analizan sus características con diversos criterios y se estiman distintos modelos en forma de innovaciones, mediante subespacios, que se comparan con criterios de información.

Los argumentos de entrada son: una matriz de datos (\mathbf{y}) y, opcionalmente, un vector de posibles índices de McMillan ($nvec$), la frecuencia anual s (en caso de buscar el orden estacional) y dos “logical flag” q y p . Si $q = 1$ (por defecto) se usan las funciones de penalización $C(G_a)$, cuando $q = 0$ se utilizan $C(G_b)$ [ver Capítulo 3]. Si $p = 1$ (por defecto) se imprimen en pantalla los resultados. Si no se especifica el intervalo de búsqueda del orden del sistema, el algoritmo usa $nvec = [0 : 2m]$, donde m denota el número de variables endógenas. Los argumentos de salida son:

1. **theta,din,lab**, estimaciones y etiquetas del modelo en EE, en formato **theta-din** correspondientes al orden del sistema óptimo, que es la moda de los nopts. En caso de que exista un empate entre varios valores de nopts se escoge el mayor de ellos.
2. **nopts**, índice de McMillan óptimo estimado por cada criterio o contraste de hipótesis.
3. **coinvec**, matriz de cointegración estimada.
4. **ur**, vector fila con dos elementos. El primero, $ur(1)$, contiene el número de raíces unitarias encontradas en el sistema y $ur(2)$ es el rango de cointegración.
5. **tab**, tabla que ofrece los resultados de los criterios de información y el contraste para cada uno de los valores de $nvec$.

Ejemplo

Las líneas de código siguientes,

```
[theta1,din1,lab1]=arma2thd([-1],[],[],[],[1],1);
[theta2,din2,lab2]=arma2thd([- .8],[],[],[],[4],1);
z1=simmod(theta1,din1,150); z2=simmod(theta2,din2,150);
u=z1+z2; u1=[z1(51:150,1) u(51:150,1)];
[theta, din, lab, nopts, coinvec, urc, tab] = nid(u1);
```

simulan el modelo, los datos y producen la correspondiente salida de NID:

*****Results from order estimation*****

n	AIC	SBC	HQ	SVC	SVC2	PVCHI2
0	11.0601	11.1383	11.0459	19.7269	19.7269	NaN
1.000	7.8123	7.9947	7.7792	1.8383	1.8309	1.0000
2.000	7.0144	7.3009	6.9624	0.8110	0.7825	1.0000
3.000	7.0823	7.4730	7.0113	0.8433	0.7817	0.5142
4.000	7.1301	7.6251	7.0403	0.8803	0.7847	0.0513

nopts

AIC	SBC	HQ	SVC	SVC2	CHI2(95%)
2	2	2	2	3	2

Unit roots

1

Cointegrating rank

1

Cointegrating matrix

1.0000 -0.9960

La tabla de resultados de la estimación del orden del sistema presenta el valor de los distintos criterios y el p-valor del contraste χ^2 para el intervalo (0:4). El valor de los nopts corresponde al índice de McMillan óptimo estimado por cada método. La tabla también ofrece el número de raíces unitarias estimadas (\hat{d}) con el criterio propuesto $C(G_a)$ y el rango de cointegración, obtenido como $m - \hat{d}$. Por último, la tabla presenta el vector de cointegración entre las dos series. Para su cálculo se usa un modelo en EE con orden de dinámica igual a la moda de los nopts obtenidos.

B.2. Algoritmo SUBEST1

Propósito

Calcula el juego de índices de Kronecker óptimo y devuelve la estimación del modelo VARMAX echelon asociado (sólo para series no estacionales).

Synopsis

[theta,din,lab,ikopt,std] = subest1(y,u,n,ct)

Descripción

La función compara los residuos de todas las combinaciones de modelos cuyos índices de Kronecker suman el orden del sistema, n . Para ello, utiliza el algoritmo de subespacios de Casals (1997) y distintos criterios de información. Una vez determinado el modelo óptimo, se devuelven el juego de índices y los parámetros estimados de la representación VARMAX echelon asociada.

Los argumentos de entrada del algoritmo son: una matriz de datos (\mathbf{y}) y el índice de McMillan del proceso (n). Opcionalmente, puede introducirse un vector de variables exógenas (\mathbf{u}) y un “logical flag” que hace referencia al criterio utilizado para discriminar los índices de Kronecker (ct). Si ct = 0 se usa el criterio *AIC*, cuando ct = 1 (por defecto) el *SBC* y cuando ct = 2 el *HQ*. Los argumentos de salida son:

1. **theta,din,lab**, estimaciones y etiquetas del modelo VARMAX echelon, en formato **theta-din**.
2. **ikopt**, índices de Kronecker óptimos del proceso.
3. **std**, desviaciones típicas estimadas de los parámetros en **theta**.

Ejemplo

Se considera el modelo VARMA restringido:

$$\left[\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} -.8 & 0 \\ -.4 & 0 \end{pmatrix} B \right] \begin{pmatrix} z_{1t} \\ z_{2t} \end{pmatrix} = \left[\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} -.4 & -1.4 \\ -.2 & -.7 \end{pmatrix} B \right] \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix};$$
$$\Sigma_{\varepsilon} = \begin{pmatrix} 1 & -.6 \\ -.6 & 1.5 \end{pmatrix}$$

Las líneas de código siguientes, 1) definen el valor de los parámetros de las matrices y simulan el modelo, y 2) estiman la representación VARMA echelon óptima correspondiente. La salida, que no se muestra aquí ya que se considera una única realización y por tanto no presenta ningún interés, devolverá los coeficientes estimados del modelo VARMA.

```
Phi=[-.8 0;-.4 0]; Theta=[-.4 -1.4;-.2 -.7]; Sigma=[1 -.6;-.6 1.5];  
[theta,din,lab]=arma2thd([Phi],[],[Theta],[Sigma],1);  
z=simmod(theta,din,350); z1=z(51:350,:);  
[theta,din,lab,ikopt,std] = subest1(z1,[],1);  
prtmod(theta,din,lab);
```

B.3. Algoritmo SUBEST1SEAS

Propósito

Calcula una estimación rápida en forma ARMA de un modelo univariante a partir de sus órdenes de dinámica regular y estacional.

Synopsis

```
[theta,din,lab,std] = subest1seas(y,u,[n ns],s)
```

Descripción

Esta función estima mediante subespacios, en dos etapas, los parámetros del modelo. En la primera, ajusta los parámetros estacionales, y en la segunda utiliza los residuos de la anterior para estimar la estructura regular. El algoritmo devuelve tanto las estimaciones de los parámetros del modelo como sus desviaciones típicas.

Los argumentos de entrada son: un vector de datos (\mathbf{y}), los órdenes de dinámica regular y estacional de la serie (n_r y n_s) y la frecuencia anual de la misma (s). Opcionalmente, puede introducirse un vector de variables exógenas (\mathbf{u}). Los argumentos que devuelve el algoritmo son los siguientes:

1. **theta,din,lab**, estimaciones y etiquetas del modelo ARMA, en formato **theta-din**.
2. **std**, desviaciones típicas estimadas de los parámetros en **theta**.

Ejemplo

Se considera el modelo univariante estacional:

$$(1 - .4B + .3B^2)z_t = -5.0 + (1 - .7B^{12})\varepsilon_t$$

El código siguiente define el modelo, simula los datos y estima un modelo ARMA(2,2) \times (1,1)₁₂ con media:

```
u = ones(350,1);  
[theta,din,lab]=arma2thd([- .4 .3], [], [], [-.7], [1], 12, -5, 1);  
z=simmod(theta,din,350,u);  
z1=z(51:350,1); u1=u(51:350,1);  
[theta,din,lab,std] = subest1seas(z1,u1,[2 1],12);  
prtmod(theta,din,lab);
```

B.4. Algoritmo SUBEST2

Propósito

Devuelve una estimación consistente de los parámetros para un modelo en estructura **theta-din**.

Synopsis

```
[topt,std] = subest2(theta,din,y)
```

Descripción

La función genera rápidas estimaciones consistentes de los parámetros **theta**. Estas estimaciones pueden tomarse como buenas condiciones iniciales para un procedimiento de máxima verosimilitud o bien, como estimación final.

El algoritmo trabaja con la representación subespacial del sistema, donde el futuro del output se expresa como una función lineal de su pasado y de la información del input. Las estimaciones se obtienen maximizando una función de los parámetros con forma de verosimilitud. La estimación, al ser iterativa, permite introducir restricciones lineales sobre cualquiera de los coeficientes.

Los argumentos de entrada son los siguientes: **theta-din**, que definen el modelo e **y**, una matriz de datos. Los argumentos de salida son:

1. **topt**, parámetros estimados.
2. **std**, desviaciones típicas estimadas de los parámetros en **theta**.

Ejemplo

Se considera el modelo univariante estacional:

$$(1 - .4B + .3B^2)z_t = -5.0 + (1 - .7B^{12})\varepsilon_t$$

Las líneas siguientes definen el modelo, simulan los datos y estiman el valor de los parámetros,

```
u = ones(350,1);  
[theta,din,lab]=arma2thd([- .4 .3], [], [], [- .7], [1], 12, -5, 1);  
z=simmod(theta,din,350,u);  
z1=z(51:350,1); u1=u(51:350,1);  
[topt,std] = subest2(theta,din,[z1 u1]);
```

Tabla de Símbolos

\mathbb{R}^n	espacio vectorial real de dimensión n
$\mathbb{R}^{m \times n}$	conjunto de las matrices reales de dimensión $m \times n$
$\mathbf{0}$	matriz de ceros
\mathbf{I}	matriz identidad
\mathbf{I}_n	matriz identidad de dimensión n
\mathbf{A}'	matriz traspuesta de \mathbf{A}
\mathbf{A}^{-1}	matriz inversa de \mathbf{A}
$\mathbf{A}^{\frac{1}{2}}$	raíz cuadrada de \mathbf{A}
\mathbf{A}^+	matriz pseudoinversa de Moore-Penrose de \mathbf{A}
$\Pi_{\mathbf{A}}$	matriz de proyección en el espacio de filas de \mathbf{A}
$\Pi_{\mathbf{A}}^{\perp}$	matriz de proyección en el espacio nulo de \mathbf{A}
σ_j	valor singular j -ésimo
$ \mathbf{A} $	valor absoluto de \mathbf{A}
$\ \mathbf{A}\ _F^2$	cuadrado de la norma Frobenius de \mathbf{A}
$diag(a_1, \dots, a_j)$	matriz diagonal
$diag(\mathbf{A}_1, \dots, \mathbf{A}_j)$	matriz diagonal por bloques
$rg(\mathbf{A})$	rango de \mathbf{A}
$tr(\mathbf{A})$	traza de \mathbf{A}
\otimes	producto de Kronecker
$\min_x f(x)$	minimizar en x la función $f(x)$
s.a.	sujeto a
$\log(x)$	logaritmo neperiano de x
$\lim_{T \rightarrow \infty}$	límite cuando T tiende a infinito
$cov(\mathbf{x}, \mathbf{y})$	covarianza entre las variables \mathbf{x} e \mathbf{y}
$var(\mathbf{x})$	varianza de \mathbf{x}
$N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$	distribución normal con media $\boldsymbol{\mu}$ y varianza $\boldsymbol{\Sigma}$
χ_n^2	distribución chi-cuadrado con n grados de libertad

B	operador retardo
∇	operador diferencia, $1 - B$
∇_s	operador diferencia estacional: $1 - B^s$ s : periodo estacional
$S(B)$	suma estacional $(1 + B + \dots + B^{s-1})$
CCC	coeficiente de correlación canónica
EE	espacio de los estados
EP	errores de previsión
l	función de verosimilitud gaussiana
MCG	mínimos cuadrados generalizados
I(d)	serie integrada de orden d
IK	índices de Kronecker
RMSE	raíz cuadrada del error cuadrático medio
VS	valor singular
AIC	criterio de información de Akaike
SBC	criterio de información de Schwartz
HQ	criterio de información de Hannan y Quinn
SVC	criterio de valores singulares de Bauer
NIDC	criterio de identificación del orden del sistema
SVD	descomposición en valores singulares
CVA	algoritmo <i>canonical variate analysis</i>
N4SID	algoritmo <i>numerical algorithm for subspace state space system identification</i>
MOESP	algoritmo <i>multivariate output error state space</i>
SID	conjunto de algoritmos <i>subspace identification</i>
MV	máxima verosimilitud
MVE	máxima verosimilitud exacta
NID	algoritmo de identificación de raíces unitarias, relaciones de cointegración, estimación de la matriz de cointegración y determinación del orden del sistema
SUBEST1	algoritmo de especificación de IK y estimación de formas VARMAX echelon
SUBEST1SEAS	algoritmo de estimación para modelos univariantes estacionales basado en subespacios
SUBEST2	algoritmo de estimación iterativo basado en subespacios

Bibliografía

- Akaike, H. (1974). A New Look at the Statistical Model Identification. *IEEE Transactions on Automatic Control*, vol. AC-19, núm. 6, 716–723.
- Akaike, H. (1975). Markovian Representation of Stochastic Processes by Canonical Variables. *SIAM Journal of Control*, vol. 13, núm. 1, 162–173.
- Akaike, H. (1976). *Canonical Correlation Analysis of Time Series and the Use of an Information Criterion*. Academic Press.
- Anderson, B. D. O. y Moore, J. B. (1979). *Optimal Filtering*. Englewood Cliffs (N.J.): Prentice Hall.
- Aoki, M. (1990). *State Space Modelling of Time Series*. Springer Verlag, New York.
- Banerjee, A., Dolado, J., Galbraith, J. W., y Hendry, D. F. (1993). *Cointegration, Error Correction and the Econometric Analysis of Non-Stationary Data*. Oxford University Press.
- Bauer, D. (1998). *Some Asymptotic Theory for the Estimation of Linear Systems Using Maximum Likelihood Methods or Subspace Algorithms*. Tesis Doctoral, TU Wien.
- Bauer, D. (2001). Order Estimation for Subspace Methods. *Automatica*, vol. 37, 1561–1573.
- Bauer, D. y Wagner, M. (2002a). Canonical Form for Unit Root Processes in the State Space Framework. *Mimeo*.
- Bauer, D. y Wagner, M. (2002b). Estimating Cointegrated Systems Using Subspace Algorithms. *Journal of Econometrics*, vol. 111, 47–84.
- Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics*, vol. 31, 307–327.

- Box, G. E. P. y Cox, D. R. (1964). An Analysis of Transformations. *Journal of the Royal Statistical Society, B Series*, vol. 26, 211–243.
- Box, G. E. P. y Jenkins, G. M. (1976). *Time Series Analysis, Forecasting and Control*. Holden-Day, San Francisco, 2ª ed.
- Box, G. E. P., Jenkins, G. M., y Reinsel, G. (1994). *Time Series Analysis, Forecasting and Control*. Prentice Hall, Englewood Cliffs, 3ª ed.
- Box, G. E. P. y Tiao, G. C. (1977). A Canonical Analysis of Multiple Time Series. *Biometrika*, vol. 64, núm. 2, 355–365.
- Burr, I. W. (1976). *Statistical Quality Control Methods*. Marcel Dekker, New York.
- Casals, J. (1997). *Métodos de Subespacios en Econometría*. Tesis Doctoral, Universidad Complutense de Madrid.
- Casals, J., Jerez, M., y Sotoca, S. (2002). An Exact Multivariate Model-Based Structural Decomposition. *Journal of the American Statistical Association*, vol. 97, núm. 458, 553–564.
- Casals, J., Sotoca, S., y Jerez, M. (1999). A Fast Stable Method to Compute the Likelihood of Time Invariant State Space Models. *Economics Letters*, vol. 65, núm. 3, 329–337.
- Chui, N. L. C. (1997). *Subspace Methods and Informative Experiments for System Identification*. Tesis Doctoral, Pembroke College Cambridge.
- Cooper, D. M. y Wood, E. F. (1982). Identifying Multivariate Time Series Models. *Journal of Time Series Analysis*, vol. 3, 277–293.
- Dagum, E. B. (1988). *X-11 ARIMA/88 Seasonal Adjustment Method - Foundations and User's Manual*. Ottawa: Statistics Canada.
- Deistler, M., Peterzell, K., y Scherrer, W. (1995). Consistency and Relative Efficiency of Subspace Methods. *Automatica*, vol. 31(12), 1865–1875.
- Dhrymes, P. (1998). *Time Series, Unit Roots and Cointegration*. San Diego, Academic Press.
- Dickey, D. A. y Fuller, W. A. (1981). Likelihood Ratio Statistics for Autorregressive Time Series with a Unit Root. *Econometrica*, vol. 49, 1063.
- Eckart, C. y Young, G. (1936). The Approximation of One Matrix by Another of Lower Rank. *Psychometrika*, vol. 1, 211–218.

- Engle, R. F. (1982). Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica*, vol. 50, 987–1007.
- Engle, R. F., Granger, C. W. J., Hylleberg, S., y Lee, H. S. (1993). Seasonal Cointegration: The Japanese Consumption Function. *Journal of Econometrics*, vol. 55, 275–298.
- Favoreel, W., De Moor, B., y Van Overschee, P. (2000). Subspace State Space System Identification for Industrial Processes. *Journal of Process Control*, vol. 10, 149–155.
- Golub, G. H. y Van Loan, C. F. (1996). *Matrix Computations*. Baltimore, tercera ed.
- Gómez, V. y Maravall, A. (1996). *Programs TRAMO and SEATS; Instructions for the Users*. Working Paper 9628, Servicio de Estudios, Banco de España.
- Gonzalo, J. y Pitarakis, J. Y. (2002). Lag Length Estimation in Large Dimensional Systems. *Journal of Time Series Analysis*, vol. 23, 401–423.
- Granger, C. W. J. y Engle, R. F. (1987). Co-Integration and Error Correction: Representation, Estimation and Testing. *Econometrica*, , núm. 55, 251–276.
- Grubb, H. (1992). A Multivariate Time Series Analysis of Some Flour Price Data. *Applied Statistics*, vol. 41, 95–107.
- Hannan, E. J. y Deistler, M. (1988). *The Statistical Theory of Linear Systems*. John Wiley, New York.
- Hannan, E. J. y Kavalieris, L. (1984). Multivariate Linear Time Series Models. *Advances in Applied Probability*, vol. 16, 492–561.
- Hannan, E. J. y Quinn, B. G. (1979). The Determination of the Order of an Autoregression. *Journal of the Royal Statistical Association, B Series*, vol. 41, 713–723.
- Harvey, A. C., Ruiz, E., y Shepard, N. (1994). Multivariate Stochastic Variance Models. *Review of Economic Studies*, vol. 61, 247–264.
- Harvey, A. C. y Shepard, N. (1993). *Structural Time Series Models*, vol. 11 of *Handbook of Statistics*. Elsevier Science Publishers B.V.
- Hotelling, H. (1936). Relation between Two Sets of Variates. *Biometrika*, vol. 28, 321–372.

- Hyllberg, S., Engle, R. F., Granger, C. W. J. G., y Yoo, B. S. (1990). Seasonal Integration and Cointegration. *Journal of Econometrics*, vol. 44, 215–238.
- Jenkins, G. M. y Alavi, A. S. (1981). Some Aspects of Modelling and Forecasting Multivariate Time Series. *Journal of Time Series Analysis*, vol. 2, núm. 1, 1–47.
- Jerez, M., Casals, J., y Sotoca, S. (1999). The Likelihood of Multivariate GARCH Models is Ill-conditioned. *Documento de Trabajo del ICAE*, , núm. 9904.
- Johansen, S. (1988). Statistical Analysis of Cointegration Vectors. *Journal of Economic Dynamics and Control*, vol. 12, 231–254.
- Johansen, S. (1991). Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autorregresive Models. *Econometrica*, vol. 59, 1551–1580.
- Knudsen, T. (2001). Consistency Analysis of Subspace Identification Methods Based on Linear Regression Approach. *Automatica*, vol. 37, 81–89.
- Koreisha, S. G. y Pukkila, T. H. (1989). Fast Linear Estimation Methods for Vector Autorregresive Moving Average Models. *Journal of Time Series Analysis*, vol. 10, 325–339.
- Larimore, W. E. (1983). System Identification, Reduced-Order Filtering and Modeling Via Canonical Variate Analysis. *Proceedings of the American Control Conference, ACC, San Francisco*, páginas 445–451.
- Larimore, W. E. (1990). Canonical Variate Analysis in Identification, Filtering and Adaptive Control. *Proceedings of the 29th Conference on Decision and Control, Hawaii*, páginas 596–604.
- Li, H. y Tsay, R. S. (1998). A Unified Approach To Identifying Multivariate Time Series Models. *Journal of the American Statistical Association*, vol. 93, núm. 442, 770–782.
- Ljung, L. (1999). *System Identification, Theory for the User*. PTR Prentice Hall.
- Luenberger, D. G. (1967). Canonical Forms for Linear Multivariate Systems. *IEEE Transactions on Automatic Control*, vol. AC-12, 290–293.
- Lütkepohl, H. (1993). *Introduction to Multiple Time Series Analysis, Second Edition*. Springer Verlag, Berlin.
- Lütkepohl, H. y Poskitt, D. S. (1996). Specification of Echelon Form VARMA Models. *Journal of Business and Economic Statistics*, vol. 14, núm. 1, 69–79.

- Lütkepohl, H. y Poskitt, D. S. (1998). *Econometrics in Theory and Practice: Festschrift for Hans Schneeweis*, cap. Consistent Estimation of the Number of Cointegration Relations in a Vector Autorregresive Model, páginas 87–100. Springer Verlag.
- Maddala, G. S. y In-Moo-Kim (1998). *Unit Roots, Cointegration and Structural Change*. Cambridge University Press.
- Makridakis, S. G., Wheelwright, S. C., y McGee, V. E. (1983). *Forecasting: Methods and Applications*. John Wiley and Sons, segunda ed.
- Martín Manjón, R. (1998). *La Reserva Federal Controla Sólo Uno de los Tipos de Interés de la Economía de Estados Unidos*. Tesis Doctoral, Universidad Complutense de Madrid.
- Martín Manjón, R. y Treadway, A. (1997). The Fed Controls Only One of the two Interest Rates in the U.S. Economy. *Documento de Trabajo del ICAE*, , núm. 9716.
- Nsiri, N. y Roy, R. (1992). On the Identification of ARMA Echelon-Form Models. *Canadian Journal of Statistics*, vol. 20, 369–386.
- O'Donovan, T. M. (1983). *Short Term Forecasting*. John Wiley and Sons Ltd.
- Peternell, K., Sherrer, W., y Deistler, M. (1996). Statistical Analysis of Novel Subspace Identification Methods. *Signal Processing*, vol. 52, 161–177.
- Poskitt, D. S. (1992). Identification of Echelon Canonical Forms for Vector Linear Processes Using Least Squares. *The Annals of Statistics*, vol. 20, 195–215.
- Poskitt, D. S. (2000). Strongly Consistent Determination of Cointegrating Rank Via Canonical Correlations. *Journal of Business and Economic Statistics*, vol. 18, 77–90.
- Reinsel, C. G. (1997). *Elements of Multivariate Time Series Analysis*. Springer-Verlag, New York, segunda ed.
- Reinsel, G. C., Basu, S., y Yap, S. F. (1992). Maximum Likelihood Estimators in the Multivariate Autorregresive Moving Average Model from a Generalized Least Squares Viewpoint. *Journal of Time Series Analysis*, vol. 13, núm. 2, 133–145.
- Riemers, H. E. (1992). Comparison of Tests for Multivariate Cointegration. *Statistical Papers*, vol. 33, 335–359.

- Rosenbrock, H. H. (1970). *State Space and Multivariable Theory*. John Wiley, New York.
- Schwartz, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics*, vol. 6, 1–48.
- Sims, C. A. (1980a). Comparison of Interwar and Postwar Cycles: Monetarism Reconsidered. *American Economic Review*, vol. 70, 250–257.
- Sims, C. A. (1980b). Macroeconomics and Reality. *Econometrica*, vol. 48, 1–47.
- Sorelius, J. (1999). *Subspace-Based Parameter Estimation Problems in Signal Processing*. Tesis Doctoral, Uppsala University.
- Terceiro, J. (1990). *Estimation of Dynamics Econometric Models with Errors in Variables*. Springer-Verlag, Berlin.
- Terceiro, J., Casals, J., Jerez, M., Serrano, G., y Sotoca, S. (2000). *A MATLAB Toolbox for Reliable Time Series Modeling and Forecasting in State Space*.
- Tiao, G. C. y Box, G. E. P. (1981). Modelling Multiple Time Series with Applications. *Journal of the American Statistical Association*, vol. 76, núm. 376, 802–816.
- Tiao, G. C. y Tsay, R. S. (1989). Model Specification in Multivariate Time Series. *Journal of the Royal Statistical Society, B Series*, vol. 51, núm. 2, 157–213.
- Toda, H. Y. (1995). Finite Sample Performance of Likelihood Ratio Tests for Cointegrating Rank in Vector Autoregressions. *Econometric Theory*, vol. 11, 1015–1032.
- Tsay, R. S. (1989a). Identifying Multivariate Time Series Models. *Journal of Time Series Analysis*, vol. 10, núm. 4, 357–372.
- Tsay, R. S. (1989b). Parsimonious Parametrization of Vector Autorregressive Moving Average Models. *Journal of Business and Economic Statistics*, vol. 7, núm. 3, 327–341.
- Tsay, R. S. y Tiao, G. C. (1985). Use of Canonical Analysis in Time Series Model Identification. *Biometrika*, vol. 72, núm. 2, 299–315.
- Van Der Veen, A., Deprettere, E. F., y Swindlehurst, A. L. (1993). Subspace-Based Signal Analysis Using Singular Value Decompositions. *Proceedings of the IEEE*, vol. 81, núm. 9, 1277–1308.

- Van Overschee, P. y De Moor, B. (1994). N4SID: Subspace Algorithms for the Identification of Combined Deterministic-Stochastic Systems. *Automatica*, vol. 30, núm. 1, 75–93.
- Van Overschee, P. y De Moor, B. (1996). *Subspace Identification for Linear Systems: Theory, Implementation, Applications*. Kluwer Academic, Dordrecht, The Netherlands.
- Vargas, M. (1999). *Modelización de Series Temporales Múltiples en Espacio de los Estados. Análisis de Procesos No Estacionarios y Cointegración*. Tesis Doctoral, Universidad de Castilla La Mancha.
- Verhaegen, M. (1994). Identification of the Deterministic Part of MIMO State Space Models Given in Innovations Form from Input-Output Data. *Automatica*, vol. 30, núm. 1, 61–74.
- Viberg, M. (1995). Subspace-Based Methods for the Identification of the Linear Time-Invariant Systems. *Automatica*, vol. 31, núm. 12, 1835–1852.
- Wei, W. W. S. (1990). *Time Series Analysis*. Addison Wesley.
- Wold, H. O. A. (1964). *A Study in the Analysis of Stationary Time Series*. Almqvist and Wiksell, Uppsala, 2^a ed.