

TESIS DOCTORAL

Programa de Doctorado en Ingeniería Industrial

ANÁLISIS DE SISTEMAS COINTEGRADOS MEDIANTE MÉTODOS DE SUBESPACIOS



**ESCUELA TÉCNICA SUPERIOR
DE INGENIEROS INDUSTRIALES**

**Dpto. de Organización de Empresas y
Comercialización e Investigación de Mercados
Universidad de Valladolid**

Autor: D. Segismundo Samuel Izquierdo Millán

Directores: D. Cesáreo Hernández Iglesias

D. Juan del Hoyo Bernat

Enero de 2005

“En general puede afirmarse que no hay cuestiones agotadas, sino hombres agotados en las cuestiones”

S. Ramón y Cajal

Agradecimientos

En primer lugar, agradezco la dirección de tesis realizada por los catedráticos D. Cesáreo Hernández Iglesias y D. Juan del Hoyo Bernat. Poder contar con su tiempo ha sido un honor y un privilegio. Sus opiniones, críticas e indicaciones han supuesto una continua fuente de estímulo, avance y enriquecimiento para mi trabajo.

Agradezco también al profesor D. Javier Pajares Gutiérrez, a quien en gran parte debo mi admiración por la profesión docente, su labor de tutor en mi formación doctoral.

Por último, agradezco al Dr. Dietmar Bauer la ayuda proporcionada con el código Matlab del algoritmo de correlaciones canónicas adaptado (ACCA) y el haberme proporcionado copia de los resultados de sus últimas investigaciones, a los doctores W. E. Larimore y M. Aoki el envío algunos de sus artículos de difícil obtención, y a L. R. Izquierdo su sugerencia de utilización del contraste binomial en la comparación de métodos

INTRODUCCIÓN	9
1 IDENTIFICACIÓN DE SISTEMAS EN EL ESPACIO DE ESTADOS	23
1.1 Introducción	23
1.2 Terminología: identificación, especificación, identificabilidad	23
1.3 Especificación estructural frente a basada en observaciones	26
1.4 El espacio de estados	31
1.5 Identificación e identificabilidad en el espacio de estados	33
1.6 Modelado económico en el espacio de estados. Interpretación	36
1.7 Estimación de parámetros en el espacio de estados	41
1.8 Motivos para una identificación en el espacio de estados	42
1.9 Equivalencia entre las representaciones ARMA y espacio de estados ...	44
2 COINTEGRACIÓN	47
2.1 Introducción	47
2.2 Series estacionarias, procesos no estacionarios y raíces unitarias	48
2.3 Contrastes de raíces unitarias	55
2.4 Cointegración de variables I(1)	58
2.5 Formulaciones de modelos cointegrados	62
2.6 Análisis estadístico de sistemas cointegrados. Método de Johansen	67
3 MÉTODO DE AOKI PARA IDENTIFICACIÓN DE PROCESOS ESTACIONARIOS	73
3.1 Introducción	73
3.2 La matriz de Hankel	74
3.3 Especificación del modelo: selección del orden. Teorema de Kronecker	75
3.4 Estimación de los parámetros del modelo	77
3.5 Propiedades del método de Aoki	82
4 MÉTODOS DE SUBESPACIOS PARA IDENTIFICACIÓN DE PROCESOS ESTACIONARIOS	85

4.1	Introducción	85
4.2	Planteamiento, notación e hipótesis básicas	89
4.3	Cálculo del orden y de una secuencia de estados.....	91
4.4	Estimación de las matrices del sistema	92
4.5	Algoritmo completo	94
4.6	Algoritmos particulares	97
4.7	El método de Aoki como método de subespacios.....	99
5	<i>IDENTIFICACIÓN DE PROCESOS NO ESTACIONARIOS EN EL ESPACIO DE ESTADOS.....</i>	<i>101</i>
5.1	Introducción	101
5.2	Descomposiciones analíticas ciclo-tendencia.....	102
5.3	Método empírico de Aoki para el modelo de cointegración	106
5.4	El algoritmo ACCA de Bauer y Wagner	109
5.5	Un algoritmo propuesto: CCA bietápico.....	110
5.6	Contrastes del número de tendencias comunes mediante métodos de subespacios	111
5.7	Propiedades de la estimación de parámetros mediante métodos de subespacios en modelos cointegrados	116
6	<i>DISEÑO DE LOS EXPERIMENTOS.....</i>	<i>119</i>
6.1	Introducción	119
6.2	Separación del problema de estimación del problema de especificación del número de tendencias comunes.....	121
6.3	Medida de la calidad de estimación del subespacio de cointegración (relación a largo plazo).....	122
6.4	Medida de la calidad de estimación de la dinámica a corto plazo	124
6.5	Términos constantes en las relaciones de cointegración	129
6.6	Ajuste de los parámetros de los métodos de estimación	131
6.7	Proceso general de las simulaciones. Programación y herramientas de análisis	133
6.8	Pruebas iniciales	137
7	<i>RESULTADOS.....</i>	<i>143</i>
7.1	Proceso generador de datos 1 (PGD1).....	144

Índice

7.2	Proceso generador de datos 2 (PGD2)	153
7.3	Proceso generador de datos 3 (PGD3)	177
7.4	Proceso generador de datos 4 (PGD4)	197
8	CONCLUSIONES	219
	<i>Anexo I . Estimación por mínimos cuadrados. Filtro de Kalman</i>	227
	I.1 Introducción	227
	I.2 Mínimos cuadrados ordinarios (MCO).....	228
	I.3 Mínimos cuadrados generalizados (MCG). Estimador de Gauss-Markov .	229
	I.4 Proyección lineal. Caso general	232
	I.5 Filtro de Kalman	233
	I.6 Resumen de los estimadores de mínimos cuadrados	237
	I.7 Aplicaciones del filtro de Kalman en econometría	239
	<i>Anexo II Realizaciones mínimas</i>	243
	<i>Anexo III . Formulaciones alternativas del espacio de estados</i>	247
	<i>Anexo IV . Identificación de sistemas estocásticos por métodos de subespacios</i>	249
	IV.1 Enfoque de Van Overschee y De Moor	249
	IV.2 Un enfoque alternativo de los métodos de subespacios.....	252
	<i>Referencias</i>	257

INTRODUCCIÓN

Presentación y objetivos

El método de análisis de sistemas cointegrados más extendido en la actualidad es el método de Johansen (Johansen 1988), que utiliza una formulación vectorial autorregresiva (VAR) de orden p del proceso generador subyacente y una técnica de estimación máximo verosímil. Por otra parte, los denominados *métodos de subespacios*, que utilizan la formulación matemática conocida como *espacio de estados* y técnicas de estimación de mínimos cuadrados, pueden ser una alternativa para el análisis de sistemas cointegrados. En particular, recientes estudios sobre las propiedades asintóticas de algunos de estos estimadores de subespacios (Bauer y Wagner 2002) ofrecen resultados muy prometedores (apartado 5.7).

Especialmente interesante es el caso de sistemas cointegrados cuyo proceso generador es de tipo VARMA, puesto que, en este caso, una formulación VAR de orden finito sólo constituirá una aproximación al verdadero proceso subyacente, mientras que la formulación en el espacio de estados no presenta esta limitación. La formulación VAR de orden finito sólo puede aproximar procesos que en el espacio de estados (con orden finito) pueden formularse de forma exacta y parsimoniosa.

En este contexto, los objetivos perseguidos en esta tesis doctoral son:

- 1- Con carácter general, analizar la utilidad de la formulación matemática conocida como *espacio de estados* en el estudio de sistemas cointegrados.
- 2- De forma específica, valorar la utilidad de los *métodos de subespacios* para obtener modelos matemáticos de sistemas cointegrados. Comparar la estimación de modelos cointegrados mediante algoritmos de subespacios con la estimación máximo verosímil por el método de Johansen, para el caso de series temporales generadas mediante procesos de tipo VARMA.

Este segundo objetivo puede formularse como el análisis (demostración o refutación) de las siguientes hipótesis:

1.- Los sistemas cointegrados admiten una formulación práctica e intuitiva en el espacio de estados, como combinación de ciclos estacionarios y tendencias comunes no estacionarias.

2.- Los algoritmos de subespacios constituyen una alternativa que puede competir con el método de Johansen para el análisis y la obtención de modelos matemáticos de sistemas cointegrados.

Motivación

La **identificación de sistemas** es una cuestión que presenta gran interés en los campos de la ingeniería y la econometría. Podemos definir la identificación de sistemas como el “arte” de encontrar modelos matemáticos simples capaces de representar un sistema a partir de observaciones ruidosas del mismo (Ljung 1999).

Conviene indicar en este punto que en econometría existe cierta ambigüedad en el uso del término *identificación*, porque con frecuencia se utilizan (indistintamente) los términos *identificación* e *identificabilidad* para referirse a la cuestión de si, dado un modelo paramétrico, existen múltiples combinaciones de valores de los parámetros que maximizan la función de verosimilitud de una muestra¹. Si el modelo no es identificable, para cualquier conjunto de observaciones dado existen múltiples combinaciones de valores de los parámetros igualmente compatibles con los datos. Nosotros nos referiremos a esta cuestión con el término *identificabilidad* exclusivamente, manteniendo el término *identificación* en su sentido de “proceso de obtención de modelos matemáticos a partir de observaciones”.

Muchos problemas del mundo real son abordados y conceptualizados como un “sistema” que ante unos determinados estímulos exteriores produce unas determinadas salidas en el tiempo. De este modo, la identificación de sistemas se convierte en una herramienta metodológica de la ciencia moderna, además de constituir una disciplina

¹ Existe aún más confusión porque algunos autores utilizan el término *identificación* para referirse a la etapa de selección de los órdenes p y q de una formulación ARMA (p,q) . Para nosotros esa será la fase de especificación del modelo, como más adelante precisaremos.

Introducción

por sí misma (Ljung 1999, p.1). Si el valor de las salidas depende no sólo del valor actual de los estímulos, sino también de valores anteriores, se dice que el sistema es dinámico. A menudo, las salidas de sistemas dinámicos cuyos estímulos no son controlados, y posiblemente no son observados, se denominan series temporales. La cotización de una acción en bolsa, el nivel del agua cada día en un embalse o la longitud de las piezas cortadas por una máquina son ejemplos de series temporales.

La clase más importante de modelos dinámicos son los modelos lineales, tanto en la práctica como en la teoría. A menudo, estos modelos lineales permiten aproximaciones suficientemente precisas incluso cuando los procesos subyacentes no son lineales, utilizando transformaciones apropiadas de las variables en un rango de funcionamiento. En esta tesis nos centramos en modelos paramétricos lineales discretos de procesos estocásticos (series temporales). La forma clásica de abordar el problema de identificación del sistema en este contexto es encontrar un filtro lineal tal que, al ser excitado por una entrada tipo ruido blanco, proporcione una salida con las mismas características estadísticas que las de la serie observada.

Otra cuestión relevante en el estudio de un sistema es la de la estacionariedad de las series observadas. La identificación de sistemas se ha desarrollado principalmente para procesos estacionarios, cuyas propiedades estadísticas son constantes en el tiempo. En un proceso estacionario, los efectos de un estímulo o perturbación transitoria (puntual) en la entrada del sistema sobre el nivel de salida del mismo se van reduciendo en el tiempo hasta desaparecer. Sin embargo, existen muchos procesos que pueden modelarse mejor mediante procesos no estacionarios, dado que las perturbaciones parecen tener efectos permanentes sobre el nivel de la serie. En el campo de la economía, Nelson y Plosser (1982) argumentaron que muchas series económicas se caracterizan mejor mediante procesos estocásticos no estacionarios que mediante funciones deterministas del tiempo (combinadas con series estacionarias). En el campo de la física o la ingeniería, aspectos como la posición de un móvil sujeto a una serie de colisiones aleatorias, o cualquier proceso en el que las desviaciones (perturbaciones) tienen efectos acumulativos permanentes (tipo “stock”), son ejemplos de procesos no estacionarios.

Las series no estacionarias exigen ciertas precauciones estadísticas y matemáticas. Así por ejemplo, los contrastes habituales de regresión no son válidos al trabajar con series no estacionarias. Un caso de especial interés en este tipo de series es el de

cointegración, que se produce si varias series no estacionarias comparten un número reducido de componentes comunes o tendencias comunes. La cointegración implica la existencia de relaciones a largo plazo entre las series no estacionarias observadas, y sería interesante poder contrastar y captar estas relaciones al identificar un modelo. En el campo de la economía, ejemplos típicos de cointegración teórica vienen dados por la teoría de la paridad de poder adquisitivo (Fuertes 1997), la relación renta-consumo o los movimientos de algunos índices bursátiles agregados (Peña 1990). En el campo de la ingeniería, la cointegración suele estar asociada a la observación de numerosas series temporales relacionadas por un proceso subyacente común no estacionario. Larimore (2000) pone como ejemplos las múltiples series de medidas obtenidas en los procesos de formación de láminas, como pueden ser láminas de aluminio o máquinas de papel, y el análisis de vibraciones con múltiples instrumentos de medida. También en el control de procesos químicos es frecuente que se presente el fenómeno de la cointegración (Harris et al. 1999).

Para llevar a cabo la identificación de un sistema estocástico lineal, el aparato matemático y estadístico actual ofrece distintas formulaciones matemáticas y distintos procedimientos asociados. La formulación matemática paramétrica más habitual en el campo de la identificación de series temporales estocásticas es la formulación VARMA, y el método de análisis de sistemas cointegrados de mayor éxito y aceptación es el método de Johansen (Johansen 1988), que utiliza una formulación de las series como vector autorregresivo (VAR) y cubre la especificación del modelo (selección del orden p de la representación VAR y contraste del número de tendencias comunes) y la estimación de parámetros. Sin embargo, existen otras formulaciones y métodos alternativos que podrían resultar más interesantes en su interpretación o en sus propiedades. Una de estas formulaciones paramétricas competidoras es la conocida como “**espacio de estados**”.

Los motivos que nos han llevado a plantearnos y a estudiar la utilidad del espacio de estados para la identificación de sistemas estocásticos cointegrados se resumen en las siguientes consideraciones:

- El espacio de estados constituye una formulación algebraica general de sistemas paramétricos dinámicos lineales. En este sentido, es equivalente a las funciones de transferencia racionales o a las formulaciones VARMA. Existen algoritmos

Introducción

para pasar un determinado modelo de un tipo de formulación a otro. Así por ejemplo, Terceiro (1990) indica un procedimiento para formular en el espacio de estados un modelo econométrico estructural general en forma reducida.

- Determinados modelos econométricos conducen de forma natural e intuitiva a una formulación en el espacio de estados. En particular, los modelos ciclo-tendencia, tan característicos del enfoque econométrico (García-Ferrer y Del Hoyo 1992; García-Ferrer y Poncela 2002; Del Hoyo y Cendejas 2003), se adaptan muy bien a esta formulación. También en ingeniería es frecuente concebir la existencia de un “estado” interno del sistema, no directamente observable pero estimable a partir de observaciones, que marca la evolución dinámica del sistema.
- La cointegración es interpretable como un modelo ciclo-tendencia. Corresponde a una descomposición de las series no estacionarias observadas en un conjunto de ciclos estacionarios y tendencias no estacionarias, donde el número de tendencias comunes no estacionarias es inferior al número de series observadas.
- Existe una batería de potentes herramientas desarrolladas sobre el espacio de estados -provenientes en su mayor parte de los campos de la ingeniería y la identificación de sistemas, y muy útiles en econometría (Hernández 1975)- que permiten la estimación de parámetros y la predicción de valores. Entre estas herramientas destacan especialmente el **filtro de Kalman** y los **métodos de subespacios**. En particular, podemos considerar que el enfoque “filtro de Kalman – función de verosimilitud” permite construir un “estimador universal” pseudo máximo verosímil de parámetros en modelos econométricos lineales.
- Se cuenta, además, con software comercial de fiabilidad probada, uso extendido y relativamente fácil acceso que implementa o facilita la implementación de muchas de estas herramientas de estimación en el espacio de estados. También

se cuenta con software académico avanzado para econometría basada en el espacio de estados¹.

- Particularmente en el caso multivariante, si se pretende realizar una especificación no estructural, sino basada en observaciones, el espacio de estados, frente a otros tipos de formulación de modelos paramétricos lineales, ofrece ventajas para la estimación del rango paramétrico u orden del sistema (ya que existe un único hiperparámetro a estimar).

Dentro de los procedimientos de identificación desarrollados sobre la formulación del espacio de estados hemos destacado especialmente la estimación máximo verosímil vía filtro de Kalman y los métodos de subespacios. En esta tesis nos planteamos la utilización de estos dos procedimientos para la identificación de sistemas estocásticos cointegrados.

Aunque el enfoque de estimación de parámetros basado en el filtro de Kalman constituye una opción para la estimación (pseudo) máximo verosímil de modelos cointegrados y para la realización de contrastes de hipótesis, se abandonó esta línea de investigación al compararla con el método de Johansen. Este último método (basado en una formulación VAR) calcula los estimadores máximo verosímiles en un solo paso, tras realizar unas regresiones auxiliares, sin necesidad de realizar una complicada y problemática búsqueda del máximo por métodos numéricos. Consideramos que, bajo este enfoque de máxima verosimilitud, sólo queda abierta la posibilidad de competir con el método de Johansen allí donde no se cumplen bien las hipótesis para su aplicación, si bien estas hipótesis no son muy restrictivas (Gonzalo, 1994). Esta reflexión nos llevó a la búsqueda de herramientas de estimación en el espacio de estados alternativas al procedimiento de estimación Kalman-máxima-verosimilitud.

Uno de los procedimientos basados en el espacio de estados que más ha evolucionado en los últimos años es la familia de algoritmos conocida bajo el nombre genérico de **métodos de subespacios**. Los métodos de subespacios han sido desarrollados

¹ Destacamos en particular el paquete de software E⁴, (Estimación Econométrica en el Espacio de Estados), desarrollado sobre Matlab, y ofrecido libremente a la comunidad académica con fines de investigación (Terceiro, Casals, Jerez, Serrano y Sotoca, 2000).

Introducción

principalmente en el ámbito de la identificación de sistemas, y proporcionan representaciones mínimas de un sistema en el espacio de estados a partir de un conjunto de observaciones del mismo. El profesor Ljung indicaba en 1993 que “El desarrollo de los métodos de subespacios es el aspecto más apasionante que ha sucedido en el campo de la identificación de sistemas en los últimos cinco años...”¹.

Los algoritmos de subespacios no son iterativos ni requieren optimización numérica, sino que utilizan operaciones robustas de álgebra matricial. Además de proporcionar una rápida estimación de parámetros del sistema², permiten estimar en un solo paso el orden del sistema (esto es, especificar el modelo basándose en observaciones).

Los métodos de subespacios son una familia de algoritmos distintos con unas características comunes. Cada tipo de algoritmo de subespacios elige de forma característica unas determinadas matrices de pesos. Dentro de la familia de métodos de subespacios, destaca especialmente el algoritmo de correlaciones canónicas (CCA). Bauer y Wagner (2002) indican que, para procesos estocásticos estacionarios, el algoritmo CCA es asintóticamente equivalente a pseudo máxima verosimilitud, es decir, que los estimadores tienen la misma distribución asintótica. Bauer y Ljung (2002) demuestran que, para procesos estocásticos estacionarios, el algoritmo CCA elige las matrices de pesos de forma óptima.

Desarrollados inicialmente para su aplicación a procesos estacionarios, la utilidad de los métodos de subespacios para identificación de procesos no estacionarios apenas ha sido abordada aún. Recientemente, Bauer y Wagner (2002) proponen un algoritmo de subespacios para procesos cointegrados que presenta además ventajas teóricas de estimación consistente sobre el método de Johansen (ver apartado 5.7).

¹ La cita aparece en Van Overschee y De Moor (1996). Traducción propia.

² Esta estimación puede usarse como alternativa a otros procedimientos de estimación recursivos (Kalman-máxima-verosimilitud, por ejemplo), más lentos y problemáticos, o para proporcionar valores iniciales adecuados para los mismos.

Llegados a este punto tenemos en contexto el objetivo principal de esta tesis: valorar en la práctica la utilidad de los métodos de subespacios para proporcionar modelos lineales paramétricos de series temporales estocásticas cointegradas.

Aplicación de métodos de subespacios a procesos no estacionarios

Hasta donde nuestra investigación alcanza, existen dos algoritmos de subespacios para procesos no estacionarios propuestos en la literatura: el algoritmo State Space Aoki Time Series (SSATS)¹ en dos etapas (Aoki 1997) y el método de correlaciones canónicas adaptado (ACCA) de Bauer y Wagner (2002). El análisis de ambos algoritmos nos llevará a plantear y estudiar un nuevo método de subespacios aplicado a procesos cointegrados: el “algoritmo de correlaciones canónicas bietápico”, o CCA2.

El algoritmo SSATS, desde su aparición inicial en Aoki (1983), ha dado lugar a diversas aplicaciones prácticas, revisiones teóricas, artículos y tesis doctorales (Vargas 2000). Sin embargo, ha sido criticado por falta de solidez estadística.

El método ACCA es una modificación del algoritmo Canonical Correlations Analysis (CCA) simple, el cual, dentro de la familia de algoritmos de subespacios, presenta características especialmente apropiadas para su aplicación a sistemas estocásticos (Bauer y Ljung 2002). ACCA cuenta con una sólida base estadística y matemática, y se publicó durante el desarrollo de esta tesis doctoral, por lo que apenas cuenta aún con estudios comparativos² o aplicados. El método ACCA modifica el algoritmo CCA simple para conseguir (utilizando una determinada forma canónica de representación en el espacio de estados) una estimación consistente de todos los parámetros de un sistema cointegrado generado por un proceso tipo VARMA, lo que supone una importante ventaja teórica sobre el método de Johansen. En este sentido, los principales resultados relativos a la consistencia de los estimadores son (apartado 5.7):

¹ El algoritmo SSATS no fue concebido inicialmente como método de subespacios, pero, como mostraremos posteriormente, podemos clasificarlo como tal.

² Con la excepción del trabajo (no publicado hasta la fecha) de Bauer y Wagner (2003b).

Introducción

- El procedimiento de Johansen aplicado sobre un proceso generador VARMA cointegrado y considerando un VAR de orden p (fijo) proporciona estimadores consistentes del subespacio de cointegración, pero los parámetros asociados a la dinámica a corto plazo no se estiman de forma consistente (Wagner 1999 y Wagner 2000).
- El procedimiento CCA simple aplicado sobre procesos generadores VARMA cointegrados proporciona estimadores consistentes del subespacio de cointegración (Bauer y Wagner 2002).
- El procedimiento ACCA aplicado sobre procesos generadores VARMA cointegrados proporciona estimadores consistentes de todos los parámetros del modelo (Bauer y Wagner 2002).

El método original CCA2 que también estudiaremos proporciona la misma estimación del subespacio de cointegración que el método CCA, por lo que la consistencia de la estimación del subespacio de cointegración está asegurada. Este método recupera la idea intuitiva de Aoki de obtener el modelo global en dos etapas: una primera etapa que obtenga el modelo para los movimientos de baja frecuencia, o tendencias no estacionarias, y una segunda etapa para obtener un modelo para los movimientos de mayor frecuencia, o ciclos estacionarios. Al proceder en dos etapas, en la primera etapa esperamos captar de forma precisa la tendencia, porque de los movimientos de alta frecuencia estacionarios esperamos un efecto muy limitado sobre el nivel medio de la serie. Del mismo modo, esperamos encontrar un modelo preciso para el ciclo porque en la segunda etapa utilizamos observaciones en las que se ha eliminado el efecto distorsionador de la tendencia.

Por otra parte, con el método CCA2 esperamos superar algunas de las limitaciones del método de Aoki. Si bien el método de Aoki utiliza en cada etapa un algoritmo de componentes principales (Van Overschee y De Moor 1996, p. 79), CCA2 utiliza en cada una de las etapas el algoritmo CCA simple, el cual, como ya hemos indicado, presenta, dentro de la familia de subespacios, características óptimas para la identificación de procesos estocásticos (Bauer y Ljung 2002).

Método de estudio

Nuestro objetivo principal es valorar la utilidad práctica de los métodos de subespacios para analizar y proporcionar modelos lineales paramétricos de series temporales estocásticas cointegradas. Para ello realizaremos un estudio mediante simulación en el que compararemos los resultados proporcionados por distintos métodos de subespacios entre sí y con los resultados proporcionados por el método de Johansen, que tomaremos como método de referencia. Los distintos métodos de subespacios estudiados son el algoritmo SSATS de Aoki (1997), el algoritmo CCA simple de Larimore (1996), el algoritmo CCA adaptado (ACCA) de Bauer y Wagner (2002) y el algoritmo CCA bietápico (CCA2) original de esta tesis.

Aunque, para el caso de procesos generadores VARMA, el recientemente propuesto método ACCA es el único que garantiza la estimación consistente de todos los parámetros del sistema, no se asegura que su comportamiento sea superior al de las otras alternativas al identificar sistemas cointegrados a partir de muestras finitas de distintos tamaños. Puesto que la distribución de los estimadores en muestras finitas es, en general, desconocida, la manera más directa de comparar distintos procedimientos de identificación es la simulación de Montecarlo.

Hacemos notar que, dada la escasez de estudios prácticos comparativos y el desconocimiento de las distribuciones de los estimadores en muestras finitas, el solo planteamiento y estudio de nuevos procesos generadores presenta ya un notable interés, ante la sospecha de que el mejor o peor comportamiento de unos métodos de identificación frente a otros pueda depender de las características del proceso generador.

Para cada uno de los procedimientos de identificación de sistemas cointegrados analizados hemos estudiado dos aspectos: la calidad de estimación del subespacio de cointegración y la capacidad de predicción a corto plazo de los modelos obtenidos. El primer aspecto mide la capacidad para captar la relación a largo plazo entre las variables. El segundo aspecto mide la capacidad para captar la dinámica a corto plazo del sistema.

Entre los antecedentes destacan los trabajos de Gonzalo (1994) y Bauer y Wagner (2003b). Gonzalo (1994) compara distintos métodos de estimación del vector de cointegración (mínimos cuadrados simple, mínimos cuadrados no lineales, Johansen,

Introducción

componentes principales, correlaciones canónicas), para el caso de dos variables cointegradas, tanto en sus distribuciones asintóticas como en sus resultados prácticos (mediante simulación). Aunque todos los estimadores que utiliza son superconsistentes, las simulaciones demuestran que las estimaciones pueden variar de forma significativa dependiendo del método utilizado. En su estudio, el método de máxima verosimilitud de Johansen es el que mejores resultados obtiene.

Tomaremos el trabajo de Gonzalo (1994) como base para nuestros experimentos, si bien nos distinguimos respecto a los antecedentes en los siguientes aspectos:

- Los métodos de estimación utilizados. Reproduciremos en uno de los experimentos el proceso generador de Gonzalo (1994) y varios de los estimadores de su trabajo, pero nos centraremos en los estimadores proporcionados por los métodos de subespacios, considerando también el nuevo algoritmo CCA2.
- Frente a Gonzalo (1994), no sólo compararemos la calidad de estimación del subespacio de cointegración (relación a largo plazo), sino que valoraremos también la capacidad de predicción a corto plazo de los modelos proporcionados por los distintos procedimientos de identificación de sistemas.
- Trabajaremos con conjuntos de 2 y 3 series temporales, y con subespacios de cointegración de dimensiones 1 y 2. Consideraremos procesos generadores en los que no se cumplen las hipótesis del método de Johansen: esto es, procesos generadores VARMA y perturbaciones no gaussianas.
- Gonzalo (1994) utiliza una medida de calidad de estimación del vector de cointegración que resulta difícilmente extensible al caso de más de dos series temporales cointegradas. Bauer y Wagner (2003b) utilizan una medida de distancia entre subespacios que puede resultar poco intuitiva (la distancia de Hausdorff). Nosotros propondremos como medida de calidad de estimación del subespacio de cointegración el ángulo formado entre los subespacios real y estimado. Esta forma de medir propuesta, además de resultar intuitiva, es extensible a varias dimensiones, y no requiere ni se ve afectada por una normalización arbitraria de los vectores de cointegración.

- Tanto los experimentos de Gonzalo (1994) como los de Bauer y Wagner (2003b) se limitan al caso en que los residuos u_t de las relaciones de cointegración son series estacionarias de esperanza cero, y se considera esta restricción en el modelo. Puesto que en series reales no existen motivos generales para imponer dicha condición, en este trabajo abordamos el caso (más general) en que los residuos de las relaciones de cointegración no tienen necesariamente esperanza cero (esto obligará a modificar los estimadores del vector de cointegración para mantener la consistencia).

Se seleccionó Matlab para la programación y ejecución de las simulaciones. Matlab ofrece gran capacidad de cálculo matricial y soporte gráfico, cuenta con una larga trayectoria de aplicación en investigación científica y dispone de un módulo de identificación de sistemas desarrollado por el profesor Ljung (Ljung 2002), lo que lo convierte en un programa especialmente fiable y apropiado para la identificación de modelos de series temporales¹. La mayor parte de los métodos han sido programados por el autor, si bien en las simulaciones de los métodos CCA y ACCA se utilizó finalmente el código amablemente proporcionado por el Dr. Bauer, coautor del algoritmo ACCA.

Los diferentes métodos de identificación utilizados requieren, a lo largo de cada algoritmo, una serie de decisiones que se detallan en el capítulo 6, junto con las condiciones de realización de los experimentos. Para complementar las capacidades de análisis ofrecidas por Matlab, los resultados de las simulaciones fueron exportados a ficheros de texto e importados en Excel para su tratamiento mediante tablas y gráficos dinámicos. Los resultados de estos análisis se muestran en el capítulo 7 y se resumen en las conclusiones.

¹ El módulo de identificación de sistemas llega incluso a implementar algunos métodos de subespacios (como el algoritmo CCA estándar, ó CVA).

Estructura del documento

En el capítulo 1 se centra el problema de identificación de sistemas estocásticos y se concretan los principales términos y conceptos utilizados. Se discute la relevancia de la especificación basada en observaciones, particularmente en el contexto de la econometría. Se tratan también las características más relevantes de la formulación en el espacio de estados: la interpretación de los modelos, la equivalencia con otras formulaciones matemáticas de modelos lineales, y los métodos de especificación y estimación de sistemas desarrollados sobre este tipo de formulación matemática.

El capítulo 2 se dedica al concepto de cointegración y a las distintas formulaciones matemáticas de modelos cointegrados. Se sugiere una posible forma de representación de sistemas cointegrados en el espacio de estados que resulta estar ligada a la forma canónica de Bauer y Wagner (2003), si bien se llega a ella desde una perspectiva diferente. De forma general, se analiza en qué consiste el estudio de sistemas cointegrados y se indican las características de algunos de los métodos más utilizados (el método de Johansen en particular).

En el capítulo 3 se expone una versión actualizada del método de Aoki (Aoki y Havenner 1991) para identificación de procesos estacionarios, incorporando los avances, las innovaciones y las variaciones propuestas sobre dicho algoritmo en los últimos años (Vargas 2000).

En el capítulo 4 se discuten los métodos de subespacios, entendidos como algoritmos de identificación de sistemas estacionarios en el espacio de estados. Se analizan las características principales de esta familia de algoritmos y se plantean distintas interpretaciones de los mismos. Se aporta también una interpretación del método de Aoki como algoritmo de subespacios.

En el capítulo 5 abordamos la identificación de sistemas cointegrados en el espacio de estados mediante métodos de subespacios. El método de Aoki aplicado a sistemas cointegrados y el algoritmo ACCA de Bauer y Wagner (2002) aparecen como los dos únicos antecedentes existentes en la literatura, surgidos por adaptación de los métodos para procesos estacionarios discutidos en los dos capítulos previos.

También en este capítulo se discuten las propiedades teóricas de los métodos de Johansen y ACCA cuando el proceso generador subyacente es un proceso VARMA. Finalmente, combinando las ideas y propiedades de los métodos de Aoki y CCA, se propone, para su inclusión en el campo de pruebas, un nuevo algoritmo de subespacios para identificación de sistemas cointegrados: el algoritmo CCA bietápico ó CCA2.

En el capítulo 6 se detalla el diseño y las condiciones de realización de las simulaciones planteadas para comparar los distintos métodos de subespacios, tomando como referencia el método de máxima verosimilitud de Johansen. Se discuten los distintos aspectos a comparar (calidad de estimación del subespacio de cointegración, calidad de predicción) y se seleccionan estimadores. La aplicación de los algoritmos exigirá una serie de decisiones y ajuste de parámetros que se describen con detalle en este capítulo, así como otros aspectos de la metodología de trabajo y análisis.

En el capítulo 7 se presentan los resultados del análisis comparativo de métodos de identificación aplicados sobre distintos procesos generadores de datos cointegrados. Para ello se utilizará un conjunto seleccionado de tablas, gráficos y contrastes representativos.

Por último, las conclusiones resumen los resultados obtenidos en las comparaciones de los diferentes métodos, así como su relevancia práctica. Se indican también posibles extensiones de la investigación a partir de este trabajo.

1 IDENTIFICACIÓN DE SISTEMAS EN EL ESPACIO DE ESTADOS

1.1 Introducción

En este capítulo pretendemos centrar el ámbito específico de esta tesis (identificación de sistemas estocásticos discretos lineales mediante modelos paramétricos en el espacio de estados) en el marco general de la identificación de sistemas y la econometría.

En primer lugar, tras concretar algunos aspectos de la terminología utilizada, ofrecemos nuestra interpretación de las diferencias entre el enfoque de especificación basada en observaciones y el enfoque de especificación estructural. Posteriormente, discutiremos las características principales del espacio de estados, los motivos que nos han llevado a seleccionar esta formulación, su interpretación y sus aplicaciones más habituales en econometría.

1.2 Terminología: identificación, especificación, identificabilidad

Siguiendo al profesor Lennart Ljung (1999), podemos definir la identificación de sistemas como el arte de crear modelos matemáticos sencillos y precisos para sistemas complejos a partir de series temporales de observaciones “ruidosas”. Bennet (1979, p. 20) indica las siguientes etapas del proceso general de análisis de sistemas¹:

¹ Traducción propia. Bennet se refiere a la etapa 2 como “identificación y especificación del sistema”, en consonancia con la metodología de Box-Jenkins. Como indicamos más adelante, nosotros utilizaremos el término “identificación” en un sentido más amplio, conforme a su uso en teoría de sistemas.

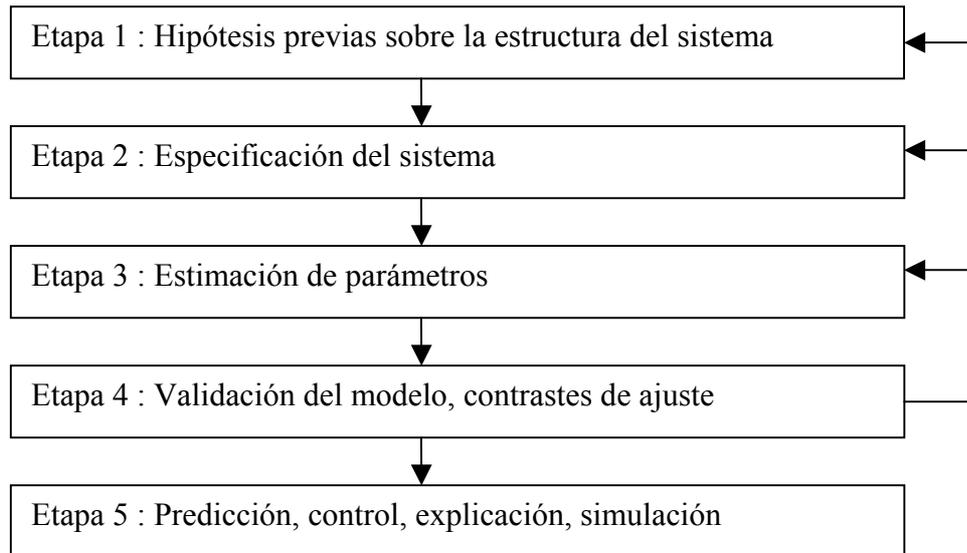


Figura 1-1. *Etapas del análisis de sistemas. Basado en Bennet (1979).*

Centrándonos en la etapa 2, especificar un modelo paramétrico para un sistema es encontrar una expresión matemática paramétrica, con un número finito y determinado de parámetros desconocidos, que consideramos capaz de representar o captar las relaciones que existen entre las variables en estudio.

A modo de ejemplo, una posible especificación de un modelo para una serie temporal es la especificación conocida como AR(1):

$$(y_t - \mu) = \phi_1 (y_{t-1} - \mu) + e_t$$

donde e_t es ruido blanco i.i.d. $(0, \sigma^2)$, $\mu = E(y_t)$ y ϕ_1 es un parámetro del modelo (que podría ser desconocido a priori y a estimar, al igual que σ^2 y μ).

Para especificar un modelo se debe elegir primero, de forma explícita o implícita, un tipo de formulación matemática. Por ejemplo, para sistemas lineales podemos considerar, entre las distintas opciones, una formulación ARMA, o una formulación en el espacio de estados. En su formulación genérica, estos modelos cuentan con un número indeterminado de parámetros: por ejemplo, si elegimos un modelo ARMA (p, q) , los órdenes p y q , de la parte autorregresiva y de medias móviles, son a priori desconocidos y tan elevados como sea necesario. Sin embargo, una vez fijados, los valores p y q limitan el número de parámetros del modelo y concretan el “aspecto” y

posibles propiedades del modelo, por lo que a estos valores se les conoce como “hiperparámetros”. Del mismo modo, para una formulación en el espacio de estados, el orden (n) de un modelo es el (único) hiperparámetro.

Especificar un modelo consiste en elegir un tipo de formulación matemática del modelo y fijar un valor para los hiperparámetros, además de (posiblemente) fijar un valor (o unas restricciones) para algunos de los parámetros.

En el campo de la teoría de sistemas, el término “identificación” abarca todo el proceso de búsqueda de un modelo para el sistema en estudio, incluidas las etapas de especificación del sistema (selección del tipo de modelo y del orden del mismo) y de estimación de los parámetros del modelo. Sin embargo, en el análisis de series temporales, diversos autores (en particular, la metodología Box-Jenkins) utilizan el término “identificación” para referirse en particular a la etapa de selección del orden del modelo (p.e., la determinación del número de retardos significativos en un modelo ARMA).

En este trabajo utilizaremos el término “identificación” en el sentido amplio correspondiente a la teoría de sistemas, reservaremos el término “especificación” para la etapa de selección del tipo de modelo y su orden, y utilizaremos el término “estimación” para referirnos únicamente a la etapa de estimación de los parámetros de un modelo ya especificado.

En cuanto a la propiedad de identificabilidad, un modelo paramétrico es identificable a partir de observaciones del sistema si la función de verosimilitud de una muestra (expresada en función de los parámetros del modelo) presenta un máximo en un único punto (vector) correspondiente a unos valores concretos de los parámetros, de forma que esos valores de los parámetros serían identificables por estimación máximo-verosímil. Si para cualquier muestra observada del sistema existen distintas combinaciones de valores de los parámetros que generarían esa muestra con igual probabilidad, el modelo no es identificable, y podemos elegir entre distintas representaciones equivalentes (observacionalmente). Si trabajamos con formulaciones no identificables, como es el caso de una formulación en el espacio de estados, podemos, sin perder generalidad, imponer condiciones adicionales sobre los parámetros del modelo para obtener una representación particular, de entre las representaciones

equivalentes. Estas representaciones particulares identificables se denominan formas canónicas.

1.3 Especificación estructural frente a basada en observaciones

El enfoque de obtención de modelos paramétricos en identificación de sistemas es distinto al enfoque que utiliza la econometría basada en modelos o “estructural”. La identificación de sistemas parte de formulaciones o tipos de modelos matemáticos flexibles y generales (modelos ARMA, espacio de estados, redes neuronales, etc.) y obtiene una representación particular lo más sencilla posible que “capte” las características o relaciones dinámicas entre las variables del sistema observado. El enfoque estructural, en cambio, parte de teorías o hipótesis sobre el “proceso generador” de las observaciones, que se traducen en unas ecuaciones matemáticas que relacionan entre sí las variables del sistema.

Para modelos paramétricos, tanto el enfoque de identificación de sistemas como el enfoque estructural proporcionan en un primer paso un conjunto de ecuaciones con un número limitado de parámetros a estimar, o modelo especificado. Valores apropiados para esos parámetros desconocidos son entonces calculados mediante algún procedimiento de estimación.

Especificación estructural

El enfoque estructural determina la “estructura” paramétrica de las ecuaciones del modelo basándose en alguna teoría económica o hipótesis sobre la relación existente entre las variables en estudio. En definitiva, este enfoque introduce en el modelo el conocimiento “a priori” que pueda existir sobre el proceso generador de las observaciones. El “tipo de modelo” o formulación matemática elegida bajo un enfoque estructural tiene, normalmente, una interpretación sencilla en relación con la teoría económica subyacente (regresión, ecuaciones en diferencias,...). Además de determinar la estructura general de las ecuaciones del modelo (hiperparámetros), el enfoque estructural suele establecer a priori los valores de algunos de los parámetros.

El enfoque estructural es el más característico de la econometría, que algunos autores definen como “la parte de la economía que se ocupa de la estimación empírica de

relaciones económicas”¹. La Figura 1-2 es ilustrativa del enfoque habitual en econometría, donde el modelo econométrico se alcanza basándose en una teoría económica, que es plasmada mediante algún tipo de formulación matemática. La formulación del modelo se trata en realidad de un proceso iterativo: si el modelo no es validado se procede a modificarlo o reformularlo.

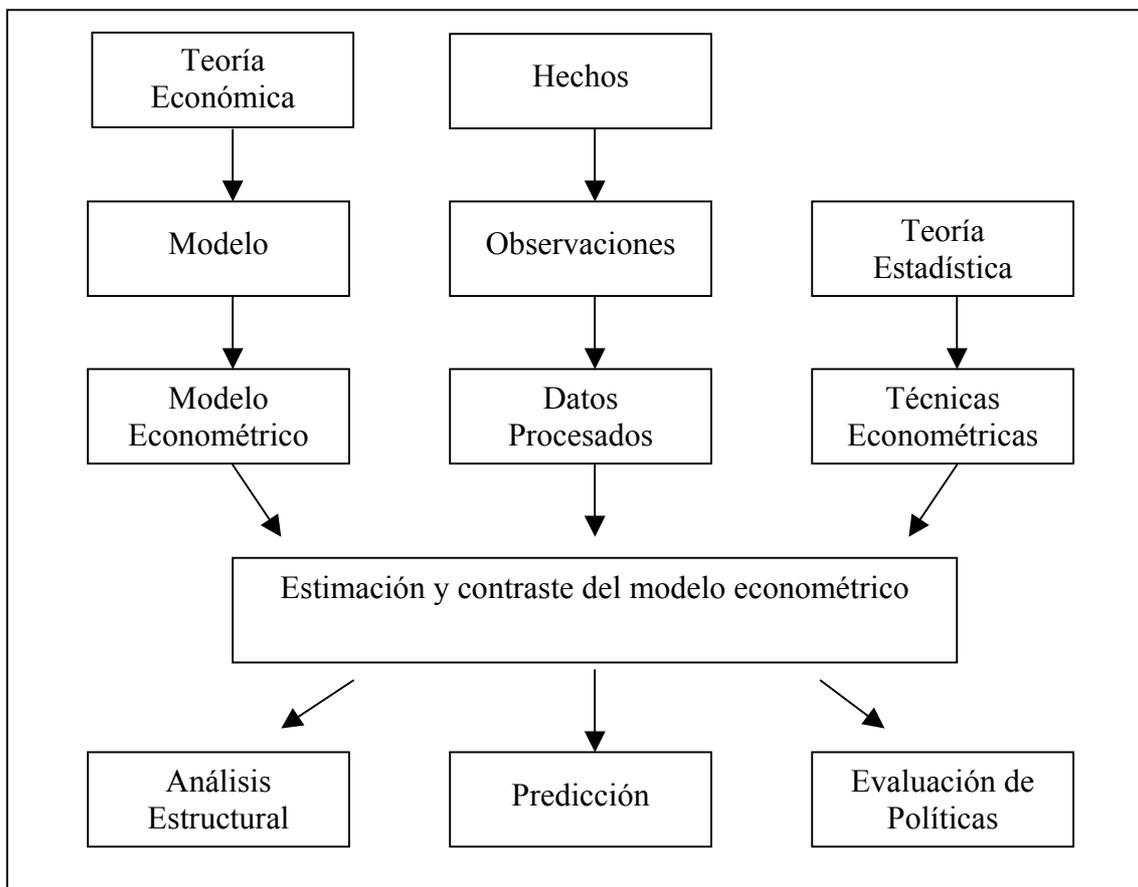


Figura 1-2. El enfoque econométrico. Basado en Intriligator, Bodkin y Hsiao (1996).

Sin embargo, la econometría también ha abordado la cuestión de especificación de modelos basándose en observaciones, propia de la identificación de sistemas. El ejemplo más conocido es posiblemente la metodología Box-Jenkins de identificación de modelos ARMA.

¹ Intriligator, Bodkin y Hsiao (1996).

Especificación basada en observaciones

El enfoque de especificación basada en observaciones parte explícitamente de algún tipo de formulación matemática genérica flexible, con capacidad de representar una amplia variedad de procesos diferentes. Para el caso de sistemas estocásticos lineales, por ejemplo, podemos partir de formulaciones ARMA o espacio de estados. Dependiendo de las características estadísticas existentes en una muestra observada del sistema a modelar, se elige un determinado valor mínimo para los hiperparámetros del modelo, de forma que el modelo sea capaz de ajustar las características estadísticas de la muestra observada. En definitiva, a partir de los datos, se determina un conjunto de restricciones sobre los parámetros del modelo genérico. En general, y buscando la parsimonia, estas restricciones establecen que todos los parámetros del modelo genérico son nulos, salvo un conjunto reducido de ellos que resulta suficiente para captar (reproducir) las características del sistema en estudio.

Comparación entre los enfoques de especificación

Como hemos indicado, al buscar un modelo apropiado para un sistema en estudio podemos distinguir dos procedimientos alternativos para especificar un modelo: el enfoque “estructural”, de acuerdo a la teoría (carácter deductivo), y el enfoque basado en observaciones, de carácter inductivo.

Ninguno de los dos enfoques de especificación está exento de riesgos. La especificación estructural impone una serie de restricciones que, si son ciertas, permiten estimar mejor los parámetros del modelo, ya que los parámetros fijados a priori no se ven sujetos a variabilidad muestral, y el resto puede ser estimado con mayor precisión. Además, estos modelos pueden venir respaldados por principios cuya validez empírica está contrastada históricamente más allá de lo que resulta posible mediante el rango de observaciones del sistema en estudio. Sin embargo, conviene no olvidar que la forma estructural de las ecuaciones que relacionan las variables se ha impuesto desde un punto de vista teórico, en forma de hipótesis del modelo. Si los datos prácticos observados no se corresponden con las características del modelo teórico especificado, quizá podamos rechazar (estadísticamente) las hipótesis del modelo. Por el contrario, si las observaciones sí se

ajustan al modelo teórico, todo cuanto podemos aceptar es que la realidad observada no ofrece suficiente evidencia en contra de nuestras hipótesis¹.

La especificación basada en observaciones asegura, en cierta medida², que el modelo especificado es capaz de captar la estructura subyacente a las observaciones en las que se ha basado la especificación. Sin embargo, el modelo sólo será válido si en el futuro las series observadas mantienen las mismas características estadísticas presentes en el conjunto de observaciones utilizado para la especificación.

Deistler (1996) realiza un análisis histórico de la inclinación de los econométricos hacia uno u otro tipo de especificación. Entre otras consideraciones interesantes, cita la decepción provocada por los grandes modelos estructurales en los años 70, al comprobar que sus predicciones eran superadas por las de sencillos modelos univariantes identificados mediante la metodología Box-Jenkins. Según Deistler, no existe actualmente una opinión clara sobre qué aproximación es preferible, si bien existe una tendencia a la especificación estructural cuando los objetivos son el análisis económico, la simulación de políticas o la predicción a medio plazo, y una tendencia a la especificación basada en observaciones, tipo VAR o con modelos “caja negra”, cuando se busca predicción a corto plazo.

El enfoque estructural es el enfoque habitual si, al abordar el estudio de unos datos, contamos a priori con información sobre el modelo generador de esos datos, o contamos con unas hipótesis que consideramos fiables sobre cómo es ese modelo generador. Si contamos con información suficiente como para poder establecer de forma precisa la estructura de las ecuaciones del modelo y se cumple que los datos estudiados corresponden realmente al modelo considerado, entonces la imposición a priori de la

¹ No olvidemos que, en el mejor de los casos, estaremos contrastando la probabilidad de que unas observaciones se correspondan con un determinado modelo matemático, y no la correspondencia con unas determinadas hipótesis económicas. Normalmente existirán múltiples teorías económicas distintas y compatibles con un mismo modelo matemático, además de múltiples modelos matemáticos compatibles con unos datos. Uno de los casos más habituales (donde a menudo se olvida este punto) es el de la teoría de la eficiencia de los mercados, junto con los contrastes de camino aleatorio para los precios.

² También en el enfoque “caja negra” existe la hipótesis previa de que el sistema es lineal de parámetros constantes.

estructura de las ecuaciones nos permite, en general, mejorar las estimaciones de los parámetros. Sin embargo, si los datos no corresponden al modelo considerado, tendremos simplemente un modelo mal especificado, que posiblemente no cumplirá bien sus objetivos y será incapaz de captar correctamente la estructura del proceso generador. Pese a ser el enfoque más frecuente en econometría, no está exento de riesgos. Como indican Aoki y Havenner (1991), en ocasiones las restricciones sobre las ecuaciones del modelo son discrecionales, y teorías alternativas conducen a restricciones contradictorias. Imponer restricciones erróneas conduce a modelos mal especificados.

El enfoque de “caja negra” apenas impone restricciones a priori sobre el modelo generador: se parte de un modelo general con un amplio número de parámetros y es el análisis de los datos el que va permitiendo imponer restricciones sobre esos parámetros, de forma que podamos encontrar un modelo generador lo más sencillo posible sin dejar de “explicar” los datos en estudio. Este es, por ejemplo, el enfoque de Box-Jenkins, que considera modelos ARMA generales y los particulariza a unos datos determinados: establece primero, a partir del análisis de los datos, qué parámetros es probable que sean significativos (es decir, especifica el modelo, lo deja expresado en función de unos pocos parámetros, imponiendo que el resto son nulos) y, una vez especificado el modelo, proporciona estimaciones de los parámetros del modelo.

Independientemente de las preferencias particulares o prevención que podamos tener hacia modelos estructurales o hacia modelos tipo “caja negra”, existen aspectos, como la predicción, en los que los modelos pueden ser fácilmente contrastados. Como indican Aoki y Havenner (1991), refiriéndose a los métodos univariantes de series temporales, “es notable que funcionen tan bien como lo hacen en tantas ocasiones, puesto que difícilmente se puede sostener que los bajos órdenes obtenidos en el análisis de series temporales univariantes reflejen el orden probable del modelo estructural subyacente, con independencia de la teoría estructural particular que se adopte”¹.

¹ Traducción propia

1.4 El espacio de estados

Una forma general de representación de un modelo lineal invariante en el tiempo con perturbación o ruido añadido es la función de transferencia (Ljung 1999, p. 45):

$$\mathbf{y}_t = \mathbf{G}(L) \mathbf{u}_t + \mathbf{H}(L) \mathbf{e}_t$$

Donde L es el operador retardo,

$$\mathbf{G}(L) = \sum_{k=1}^{\infty} \mathbf{G}_k L^k, \quad \mathbf{H}(L) = \mathbf{I} + \sum_{k=1}^{\infty} \mathbf{H}_k L^k,$$

y \mathbf{e}_t es una secuencia de variables aleatorias independientes con esperanza cero y matriz de covarianzas \mathbf{R} .

En la práctica trabajamos con modelos en los que realizamos una simplificación: suponemos que $\mathbf{G}(L)$ y $\mathbf{H}(L)$ son expresables mediante un número finito de parámetros. Las funciones de transferencia racionales y las formulaciones en el espacio de estados con dimensión finita con coeficientes constantes son formulaciones generales que permiten representar este tipo de modelos lineales invariantes en el tiempo.

Otras formulaciones paramétricas generales de sistemas lineales son:

- El modelo estructural dinámico general:

$$\mathbf{F}(L) \mathbf{y}_t = \mathbf{G}(L) \mathbf{u}_t + \mathbf{H}(L) \mathbf{e}_t$$

$$\text{con } \mathbf{F}(L) = \sum_{k=0}^{\infty} \mathbf{F}_k L^k, \quad \mathbf{G}(L) = \sum_{k=0}^{\infty} \mathbf{G}_k L^k, \quad \mathbf{H}(L) = \sum_{k=0}^{\infty} \mathbf{H}_k L^k$$

- La forma reducida, en la que $\mathbf{F}_0 = \mathbf{I}$, y a la que se puede llegar a partir del modelo general si existe $\mathbf{F}_0^{-1}(L)$.
- La formulación VARMAX, que es una representación en forma reducida en la que además se impone $\mathbf{H}_0 = \mathbf{I}$:

$$(\mathbf{I} + \mathbf{F}_1 L + \mathbf{F}_2 L^2 + \dots) \mathbf{y}_t = \mathbf{G}(L) \mathbf{u}_t + (\mathbf{I} + \mathbf{H}_1 L + \mathbf{H}_2 L^2 + \dots) \mathbf{e}_t$$

Para sistemas estocásticos, tendríamos la formulación VARMA:

$$(\mathbf{I} + \mathbf{F}_1 L + \mathbf{F}_2 L^2 + \dots) \mathbf{y}_t = (\mathbf{I} + \mathbf{H}_1 L + \mathbf{H}_2 L^2 + \dots) \mathbf{e}_t$$

Por su parte, la formulación general del espacio de estados admite distintas expresiones, como se detalla en el Anexo III; tras examinar las distintas posibilidades, en este trabajo hemos optado por tomar como referencia la representación y notación de Ljung (1999) para modelos en el espacio de estados con coeficientes constantes. La única salvedad respecto a su notación es que utilizamos el símbolo \mathbf{z} en vez de \mathbf{x} para el vector de estados. El motivo es que el símbolo \mathbf{x} se ha utilizado históricamente en distintos modelos econométricos, como el de regresión o el de datos de panel, para representar las variables exógenas; para no crear confusión, reservaremos los símbolos \mathbf{x} (habitual en econometría) y \mathbf{u} (habitual en teoría de sistemas) para representar variables exógenas, y representaremos el vector de estados mediante el símbolo \mathbf{z} .

Así pues, la representación y notación adoptada (en forma de innovaciones y con parámetros constantes, que será el caso en que nos centraremos) es

$$\mathbf{z}_{t+1} = \mathbf{A} \mathbf{z}_t + \mathbf{B} \mathbf{u}_t + \mathbf{K} \mathbf{e}_t \quad \text{Ecuación de transición de estados}$$

$$\mathbf{y}_t = \mathbf{C} \mathbf{z}_t + \mathbf{e}_t \quad \text{Ecuación de observaciones}$$

donde \mathbf{e}_t es ruido blanco de esperanza $E(\mathbf{e}_t) = \mathbf{0}$ y matriz de covarianzas $E(\mathbf{e}_t \mathbf{e}'_t) = \mathbf{R}$.

Para el caso de procesos estocásticos, en que no se consideran variables exógenas \mathbf{u}_t , tendríamos

$$\mathbf{z}_{t+1} = \mathbf{A} \mathbf{z}_t + \mathbf{K} \mathbf{e}_t \quad \text{Ecuación de transición de estados}$$

$$\mathbf{y}_t = \mathbf{C} \mathbf{z}_t + \mathbf{e}_t \quad \text{Ecuación de observaciones}$$

donde \mathbf{e}_t es ruido blanco de esperanza $\mathbf{0}$ y matriz de covarianzas $E(\mathbf{e}_t \mathbf{e}'_t) = \mathbf{R}$.

Así pues, una representación en el espacio de estados de un sistema estocástico lineal de series temporales \mathbf{y}_t quedaría definida por el conjunto de matrices $\{\mathbf{A}, \mathbf{K}, \mathbf{C}, \mathbf{R}\}$.

Los motivos para seleccionar la representación de Ljung como base son:

- Es la notación más utilizada, prácticamente estándar, en el campo de la Teoría de Sistemas (Teoría de Control), que es el campo en el que más se ha desarrollado el estudio y aplicación del filtro de Kalman. Trabajos avanzados

en el estudio de series temporales en el espacio de estados – Aoki (1990), Aoki y Havenner (1991), Bauer y Wagner (2002)- han adoptado también una notación similar.

- Las variables exógenas aparecen de forma explícita, con su propia matriz de parámetros, lo cual supone claridad en la representación e interpretación.
- Es la representación con la que trabaja el módulo de Identificación de Sistemas de MATLAB.

1.5 Identificación e identificabilidad en el espacio de estados

Identificación estocástica en el espacio de estados

Esta tesis está centrada en la identificación de sistemas lineales estocásticos, en los que no se consideran variables exógenas de entrada al sistema, discretos y de parámetros constantes.

La metodología de análisis de procesos estocásticos se ha desarrollado particularmente para el caso de los procesos estacionarios, cuyas propiedades estadísticas son constantes en el tiempo. La forma típica de abordar el estudio de estos sistemas es la de encontrar un filtro lineal tal que, al ser excitado por una entrada tipo ruido blanco, produce una salida cuyas características estadísticas coinciden con las del proceso estudiado.

Precisando un poco más, como indican Vaccaro y Vukina (1993), un modelo matemático para un proceso estocástico estacionario de dimensión finita es un sistema lineal invariante en el tiempo tal que, al ser excitado por ruido blanco, produce una salida que es equivalente (en sus propiedades estadísticas) al proceso estocástico considerado. Los parámetros del modelo matemático pueden ser, entre otras opciones, coeficientes de una ecuación en diferencias (modelo ARMA) o parámetros de un modelo en el espacio de estados.

Van Oberschee y De Moor, (1996, p. 58) formulan el problema general de identificación estocástica en el espacio de estados con coeficientes constantes al que nos enfrentamos (ver anexo IV.1). Si adaptamos este problema a la formulación en

innovaciones que hemos adoptado, podemos formular nuestro objetivo de identificación como:

Dado un conjunto de T observaciones \mathbf{y}_t , de dimensión m (m series temporales), correspondientes a un sistema de parámetros desconocidos que evoluciona conforme a las ecuaciones

$$\mathbf{z}_{t+1} = \mathbf{A} \mathbf{z}_t + \mathbf{K} \mathbf{e}_t \quad \text{Ecuación de transición de estados}$$

$$\mathbf{y}_t = \mathbf{C} \mathbf{z}_t + \mathbf{e}_t \quad \text{Ecuación de observaciones}$$

donde \mathbf{e}_t es un vector ruido blanco de esperanza cero que cumple

$$E(\mathbf{e}_k \mathbf{e}'_j) = \mathbf{R} \delta_{kj}$$

Encontrar:

- El orden n del sistema (especificación del modelo).
- Unas matrices \mathbf{A} , \mathbf{K} y \mathbf{C} (susceptibles de transformación correspondiente a un cambio de base de los estados) y una matriz \mathbf{R} , tales que los estadísticos de segundo orden de la salida del modelo y de las series observadas sean iguales.

El espacio de estados ofrece características propias respecto a la interpretación del modelo, la especificación del mismo y la estimación de parámetros. Aunque la formulación en el espacio de estados es equivalente a otras formulaciones alternativas generales y paramétricas de un sistema lineal, como la función de transferencia o las formulaciones VARMA, existen algoritmos y métodos de identificación desarrollados específicamente para trabajar con cada tipo de formulación. Cuando el modelo es desconocido y nuestro objetivo es identificar un modelo apropiado a partir de observaciones (inferencia sin teoría económica de partida), los distintos métodos de identificación conducirán (posiblemente) a la obtención de distintos modelos.

Identificabilidad de un modelo en el espacio de estados.

Como hemos indicado, en muchas ocasiones los parámetros del modelo subyacente a las observaciones serán desconocidos, y nuestro objetivo será encontrar una representación de ese modelo (Harvey 1989, p. 205). La formulación general en el espacio de estados subraya una particularidad en este caso: una representación en el espacio de estados no es identificable, al menos hasta que impongamos algunas condiciones, que posiblemente serán arbitrarias¹. Una forma canónica en el espacio de estados es una formulación que establece suficientes condiciones adicionales sobre los parámetros del sistema para que la representación sea única, identificable.

Supongamos que un sistema es representable en el espacio de estados por las ecuaciones

$$\mathbf{z}_{t+1} = \mathbf{A} \mathbf{z}_t + \mathbf{K} \mathbf{e}_t \quad \text{Ecuación de transición de estados}$$

$$\mathbf{y}_t = \mathbf{C} \mathbf{z}_t + \mathbf{e}_t \quad \text{Ecuación de observaciones}$$

donde \mathbf{e}_t es ruido blanco de esperanza $\mathbf{0}$ y matriz de covarianzas $E(\mathbf{e}_t \mathbf{e}'_t) = \mathbf{R}$. Consideremos ahora unos nuevos estados definidos por $\mathbf{z}^*_t = \mathbf{T} \mathbf{z}_t$, donde \mathbf{T} es una matriz no singular cualquiera de dimensiones coherentes con el vector de estados. Entonces, las primeras igualdades de las ecuaciones

$$\mathbf{z}^*_{t+1} = \mathbf{A}^* \mathbf{z}^*_t + \mathbf{K}^* \mathbf{e}_t = \mathbf{T} \mathbf{A} \mathbf{T}^{-1} \mathbf{T} \mathbf{z}_t + \mathbf{T} \mathbf{K} \mathbf{e}_t$$

$$\mathbf{y}_t = \mathbf{C}^* \mathbf{z}^*_t + \mathbf{e}_t = \mathbf{C} \mathbf{T}^{-1} \mathbf{T} \mathbf{z}_t + \mathbf{e}_t$$

donde las matrices \mathbf{A}^* , \mathbf{K}^* y \mathbf{C}^* están definidas de forma que se cumplan las segundas igualdades (esto es, $\mathbf{A}^* = \mathbf{T} \mathbf{A} \mathbf{T}^{-1}$, $\mathbf{K}^* = \mathbf{T} \mathbf{K}$ y $\mathbf{C}^* = \mathbf{C} \mathbf{T}^{-1}$), representan también el mismo sistema estudiado. Ambas representaciones generan la misma función de distribución conjunta para las variables, por lo que la probabilidad de generar un conjunto de observaciones dado es la misma para las dos. Se dice que ambas representaciones son indistinguibles observacionalmente: dada una secuencia de

¹ Otros tipos de formulaciones matemáticas, como los modelos ARMA, también precisan imponer, de forma implícita o explícita, una serie de restricciones para que la representación sea identificable (Bauer y Wagner 2003, p.7).

variables observadas y_t , ésta puede haber sido generada con igual probabilidad por múltiples posibles representaciones distintas en el espacio de estados. Se pueden encontrar infinitos vectores de variables de estado que, junto con unas matrices del sistema adecuadas, generarían con igual probabilidad la secuencia observada. Por este motivo decimos que la representación de un sistema en el espacio de estados no es única. Dentro de todas las distintas representaciones posibles, preferiremos, en general, aquellas que minimicen el número de variables de estado y el número de parámetros.

Consecuencia de lo anterior es que, incluso con una representación que utilice el mínimo número de estados necesario (ver Anexo II), para poder identificar los parámetros de una representación en el espacio de estados es necesario imponer unas condiciones sobre las variables. En particular, dado que los estados pueden sufrir rotación y cambio de escala sin que por ello dejen de poder representar el sistema (cambiarían las matrices \mathbf{A} , \mathbf{K} y \mathbf{C} según lo indicado anteriormente, absorbiendo los cambios en los estados), para la identificación es necesario fijar un sistema coordinado para los estados, lo que normalmente se traduce en imponer condiciones adicionales sobre las matrices \mathbf{A} , \mathbf{K} y \mathbf{C} .

1.6 Modelado económico en el espacio de estados. Interpretación

Componentes no observados

La formulación en el espacio de estados trabaja con un conjunto de componentes no directamente observados pero de dinámica conocida (estados), subyacentes a un conjunto de variables observadas.

Podemos considerar que una serie observada ($y_{1,t}$) está compuesta por un conjunto de n componentes no observados $\mathbf{u}_t = [u_{1,t} \ u_{2,t} \ \dots \ u_{n,t}]'$, es decir,

$$y_t = \sum_{i=1}^n c_i u_{i,t} = [c_1, c_2, \dots, c_n] \mathbf{u}_t.$$

Si postulamos una descomposición de este tipo, y postulamos también la dinámica de evolución de los componentes $u_{i,t}$, obtenemos una formulación en el espacio de estados.

Revisaremos brevemente algunos modelos habituales en econometría que se adaptan en su interpretación a una formulación en el espacio de estados.

Modelos de series temporales estructurales

Un modelo de series temporales estructurales (Durbin y Koopman 2001, p. 39; Harvey 1989) es un modelo que postula que la serie observada está formada por un componente tendencial, un componente estacional o cíclico y un término de error, y donde dichos componentes son modelados de forma explícita. Consideremos por ejemplo el modelo

$$y_t = \mu_t + e_{1t} \quad e_{1t} \sim N(0, \sigma_1^2)$$

$$\mu_{t+1} = \mu_t + \theta_t + e_{2t} \quad e_{2t} \sim N(0, \sigma_2^2)$$

$$\theta_{t+1} = \theta_t + e_{3t} \quad e_{3t} \sim N(0, \sigma_3^2)$$

En este modelo la variable μ se interpreta como un “nivel aleatorio” o tendencia estocástica para la serie observada, y la variable θ , que sigue un camino aleatorio, se interpreta como una “pendiente aleatoria”.

Aunque este enfoque de análisis de series temporales utiliza el término “estructural”, lo cierto es que a menudo las ecuaciones del modelo no vienen marcadas por ninguna teoría económica, sino que responden a una descomposición “intuitiva” de una serie temporal en distintos componentes no directamente observados. A uno de los componentes se le denomina tendencia porque se le impone un modelo que implique el seguimiento de la serie observada con una cierta suavidad. A la diferencia entre la tendencia y la serie observada se le denomina ciclo, y se le impone un modelo estacionario¹.

Clark (1987), siguiendo este enfoque, postula un modelo para la actividad económica (y_t) de la forma

$$y_t = T_t + C_t$$

¹ Se entiende el ciclo como la diferencia entre la tendencia y la serie observada, para el que se postula un modelo estacionario, pero no necesariamente se le exige que presente raíces complejas, es decir, no se le exige que su respuesta a impulsos sea sinusoidal.

$$T_t = T_{t-1} + d_t + e_{1t}$$

$$d_t = d_{t-1} + e_{2t}$$

$$\Phi(L) C_t = e_{3t}$$

donde e_{1t} , e_{2t} , e_{3t} son procesos de ruido blanco independientes y de varianzas σ_1^2 , σ_2^2 , σ_3^2 , respectivamente, y $\Phi(L)$ es un polinomio finito en el operador retardo L . Clark estudia distintas especificaciones para $\Phi(L)$.

Modelos de componentes comunes

Estos modelos postulan la existencia de factores no observados comunes a un grupo de series observadas. En cualquier caso, se requiere especificar la dinámica de los componentes no observados.

Así por ejemplo, Stock y Watson (1991) postularon la existencia de un componente no directamente observado C_t que representa el estado del ciclo económico, de forma que un conjunto de n variables macroeconómicas observadas (y_{1t} , y_{2t}, \dots , y_{nt}) estarían influenciadas por el estado del ciclo además de contar con un componente individual propio (s_{1t} , s_{2t}, \dots , s_{nt}) que se supone incorrelacionado con el de las demás. Estos modelos exigen especificar tanto la dinámica del componente C_t como la dinámica de los componentes individuales de cada serie. Si, por ejemplo, se supone que tanto el ciclo como los componentes individuales siguen una estructura AR(1), un modelo en el espacio de estados sería

$$\begin{pmatrix} C_{t+1} \\ s_{1,t+1} \\ s_{2,t+1} \\ \vdots \\ s_{n,t+1} \end{pmatrix} = \begin{pmatrix} \phi_c & 0 & 0 & \dots & 0 \\ 0 & \phi_1 & 0 & \dots & 0 \\ 0 & 0 & \phi_2 & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & 0 & \dots & \phi_n \end{pmatrix} \begin{pmatrix} C_t \\ s_{1,t} \\ s_{2,t} \\ \vdots \\ s_{n,t} \end{pmatrix} + \begin{pmatrix} e_{C,t} \\ e_{1,t} \\ e_{2,t} \\ \vdots \\ e_{n,t} \end{pmatrix}$$

$$\begin{pmatrix} y_{1,t} \\ y_{2,t} \\ \vdots \\ y_{n,t} \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{pmatrix} + \begin{pmatrix} \gamma_1 & 1 & 0 & \dots & 0 \\ \gamma_2 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ \gamma_n & 0 & 0 & \dots & 1 \end{pmatrix} \begin{pmatrix} C_t \\ s_{1,t} \\ s_{2,t} \\ \vdots \\ s_{n,t} \end{pmatrix}$$

donde el parámetro γ_i representa la sensibilidad de la variable i respecto al ciclo económico.

También bajo este enfoque de componentes comunes, Peña y Poncela (2005) abordan el análisis de factores comunes no observables dinámicos y no estacionarios. El modelo de cointegración responde en particular a estas características, ya que la cointegración equivale a la existencia de tendencias estocásticas comunes a un conjunto de series temporales no estacionarias.

Modelos de coeficientes variables

Podemos considerar un modelo de estructura similar al de regresión ($y_t = \mathbf{\beta}' \mathbf{x}_t + e_t$) en el que los coeficientes $\mathbf{\beta}$ no sean estáticos, sino que varíen en el tiempo conforme a una dinámica determinada, por ejemplo, que sean paseos aleatorios $\mathbf{\beta}_{t+1} = \mathbf{\beta}_t + \mathbf{v}_t$.

Wells (1996) estudia y sugiere una amplia serie de fenómenos económicos (principalmente financieros) planteables en términos de coeficientes dinámicos, siendo el ejemplo más típico el de los coeficientes beta dinámicos para las acciones en finanzas, en un modelo tipo CAPM (del Olmo y Hernández 1991). Este modelo puede formularse en el espacio de estados mediante las ecuaciones

$$\mathbf{\beta}_{t+1} = \mathbf{\beta}_t + \mathbf{v}_t$$

$$r_t = r_f + (r_{m,t} - r_f) \mathbf{\beta}_t + e_t$$

donde (r_t) es la rentabilidad¹ de un activo en el periodo t , ($r_{m,t}$) es la rentabilidad del mercado en ese periodo, (r_f) es la rentabilidad libre de riesgo y (\mathbf{v}_t , e_t) son perturbaciones tipo ruido blanco de esperanza cero y matriz de covarianzas a estimar.

¹ En realidad, el CAPM está formulado en expectativas, por lo que se debería trabajar con las rentabilidades esperadas. Lamentablemente, las rentabilidades esperadas no son observables, lo cual, por otro lado, convierte al CAPM en un modelo imposible de contrastar (si no se añade un modelo de formación de expectativas).

Podríamos aumentar la generalidad contemplando para los coeficientes variables una dinámica de la forma $\beta_{t+1} = A \beta_t + v_t$, o podríamos incluir entre las variables explicativas retardos de las variables observadas.

En cualquier caso, en estos modelos de coeficientes variables, los parámetros variables son estados del sistema, para los que es necesario especificar una ecuación dinámica, y su valor en cada instante es estimable mediante el filtro de Kalman. Entre otros aspectos, esta estimación de un modelo de coeficientes variables permite realizar contrastes sobre la estabilidad de los mismos. Wells (1996) ofrece una revisión muy completa de diversos contrastes de estabilidad de coeficientes, desde el conocido *cusumsq* (Cumulative Sum of Squared Recursive Residuals, basado en los residuos de una estimación recursiva por mínimos cuadrados) hasta el que Wells denomina *flexible least squares*, basado en el filtro de Kalman.

Pese al atractivo teórico de estos modelos de coeficientes variables, se presentan varias dificultades teóricas al abordarlos. La primera de ellas está asociada al propio concepto de parámetro. Al estimar un parámetro de un modelo tratamos de captar una característica estable en los datos observados. Si esa característica no es estable, se enturbia el concepto de parámetro. La segunda dificultad teórica es que para estimar esos parámetros cambiantes debemos especificar a priori su estructura dinámica, su ecuación de transición en el tiempo. Los resultados que obtengamos dependerán en gran medida de la especificación de esa estructura dinámica, que en la mayoría de los casos se impone de forma arbitraria, sin sujeción a una teoría económica. Lamentablemente, este enfoque de parámetros variables conduce en muchos casos a modelos que, aunque resultan compatibles con los datos, no aumentan la capacidad predictiva y, a menudo, enturbian su capacidad explicativa.

Por los motivos anteriores, en esta tesis nos hemos limitado a los modelos de coeficientes constantes, reconociendo la utilidad que estos modelos de coeficientes variables pueden ofrecer para estudiar modelos no lineales y modelos con cambio de régimen.

Citando algunos otros problemas económicos que de forma natural se adaptan a una formulación en el espacio de estados, o que han sido abordados desde la misma, Gourieroux y Monfort (1997) destacan los modelos de “expectativas racionales” y los

modelos de aprendizaje (estos últimos caen dentro del campo de los parámetros variables).

1.7 Estimación de parámetros en el espacio de estados

Una vez especificado el modelo en el espacio de estados, quizá el enfoque más obvio para la estimación de parámetros sea el de máxima verosimilitud, supuestas perturbaciones gaussianas. Este enfoque (Harvey 1989, pp. 125-127; Hamilton 1994, pp. 385-387; Gouriéroux y Monfort 1997, pp. 592-597), consiste en un cálculo recursivo de la función de verosimilitud mediante el filtro de Kalman y una búsqueda del máximo por métodos numéricos, tal como se detalla en el anexo I.7.

Desde un punto de vista teórico (sin considerar los problemas numéricos de este procedimiento e imponiendo condiciones apropiadas y suficientes para solucionar el problema de la identificabilidad del modelo), este enfoque resuelve el problema de la estimación en el espacio de estados. Vargas (1999) indica que éste puede ser el motivo por el que en el campo de la Economía Cuantitativa no se ha prestado gran atención a algoritmos alternativos de estimación como el propuesto por Aoki. Sin embargo, los resultados de los trabajos realizados sobre este tema (ver, por ejemplo, Harvey y Peters 1984) indican que el enfoque de máxima verosimilitud puede ser un método singularmente complejo, incluso en los modelos más simples.

Por otro lado, el enfoque de máxima verosimilitud precisa contar con buenas estimaciones iniciales para comenzar los métodos numéricos de búsqueda. Esto justificaría también la utilización de métodos alternativos de estimación, si resultan suficientemente rápidos.

En los últimos años se han desarrollado particularmente los denominados métodos de subespacios, que permiten tanto la especificación del modelo en el espacio de estados como la estimación de los parámetros del mismo, a partir de observaciones del sistema. Como veremos más adelante, el método de Aoki comparte muchos puntos en común con estos algoritmos, y puede ser considerado como un caso particular de los mismos.

1.8 Motivos para una identificación en el espacio de estados

El espacio de estados permite partir de un modelo estructural (especificado), en el que las variables y sus relaciones tengan una interpretación física o económica, y llevar a cabo la estimación de los parámetros del modelo. Del mismo modo, es posible plantear a priori modelos en los que las series observadas se ven formadas por componentes “interpretables intuitivamente” y cuya dinámica es postulada a priori, y los parámetros de estos modelos pueden entonces ser estimados. Además de esta importante ventaja de permitir especificaciones “intuitivas” e interpretables, podemos plantearnos qué justifica la utilización del espacio de estados de cara a la identificación de un modelo para un sistema, qué ventajas puede presentar frente a otras formulaciones de modelos lineales como la formulación VARMA.

Ljung (1999, p. 524), refiriéndose a los aspectos prácticos de la identificación de sistemas multivariable, indica:

“Los sistemas multivariable son a menudo más complejos de modelar. En particular, los sistemas con varias salidas pueden resultar difíciles. Una de las razones de esta dificultad es que el acoplamiento entre varias entradas y salidas conduce a modelos complejos: las estructuras son más ricas y se requieren más parámetros para conseguir un buen ajuste.

Con carácter general, es preferible trabajar con modelos en el espacio de estados para el caso multivariable, ya que la complejidad en la estructura del modelo es más sencilla de tratar aquí. Básicamente es una cuestión de elegir el orden del modelo.”¹

Volviendo entonces a la cuestión de qué ventajas ofrece la representación en el espacio de estados para la identificación del sistema, resaltamos en particular:

1.- Facilidad comparativa en la especificación del modelo, especialmente en el caso de series multivariable. Nos referimos aquí a una especificación basada en la información contenida en los datos u observaciones del sistema, y no a una especificación basada en

¹ Traducción propia.

intuiciones o teorías previas. Es importante resaltar también que el espacio de estados permite además trabajar con modelos tipo “caja gris” (Ljung 1999), que combinan una especificación basada en observaciones con la consideración de determinadas características o restricciones estructurales a priori.

2.- Como se indicó en el apartado 1.7, el enfoque Kalman-máxima-verosimilitud para estimación de parámetros resuelve, desde el punto de vista teórico, la estimación máximo verosímil de parámetros en modelos lineales, una vez éstos han sido formulados en el espacio de estados.

3.- Existen métodos de especificación y estimación de parámetros del modelo desarrollados específicamente para la representación en el espacio de estados. En particular, pueden aplicarse los denominados métodos de subespacios, que analizaremos posteriormente. La aplicación de estos métodos puede realizarse como alternativa o como complemento a otros procedimientos de estimación, tales como la optimización numérica de la función de verosimilitud.

Curiosamente, la utilidad del espacio de estados para realizar una especificación de sistemas multivariante basada en observaciones pasa desapercibida en gran parte de la literatura econométrica actual sobre espacio de estados y series temporales, más centrada en la estimación de parámetros dado el modelo que en la especificación del propio modelo. A modo de ejemplo, indicaremos que ni el libro de Harvey (1989) sobre filtro de Kalman y series temporales estructurales (compuestas por tendencia, ciclo y error, que son modelados por separado), ni el manual de Hamilton (1994) dedicado al análisis de series temporales, ni el libro de Durbin y Koopman (2001), sobre análisis de series temporales en el espacio de estados, refieren esta utilidad del espacio de estados. En estos textos, el espacio de estados se plantea básicamente como una manera de formular un modelo previa a la aplicación del filtro de Kalman, sin referencia a la utilidad del espacio de estados para la especificación del modelo. La especificación en el espacio de estados basada en observaciones ha sido desarrollada particularmente en el campo de la identificación de sistemas o teoría de control. Es en este campo en el que los algoritmos de subespacios han evolucionado, relacionados sólo con algunas aportaciones aisladas, aunque notables, desde el campo de la econometría, como el algoritmo de Aoki - Aoki y Havenner (1991)-.

1.9 Equivalencia entre las representaciones ARMA y espacio de estados

La representación VARMA de un sistema y la representación en el espacio de estados son dos formas alternativas de representar un mismo objeto. Considerando procesos con función de transferencia racional (ver Hannan y Deistler 1988 o Gilbert 1993 para una discusión detallada), una determinada representación en el espacio de estados corresponde a una determinada función de transferencia, al igual que una determinada representación VARMA también corresponde a una determinada función de transferencia. Sin embargo, existen múltiples representaciones VARMA y espacio de estados para una misma función de transferencia dada.

Existe una equivalencia algebraica entre ambos tipos de representaciones de forma que, dada una representación de una función de transferencia en el espacio de estados (/VARMA), es posible encontrar una representación VARMA (/espacio de estados) equivalente, que corresponde a la misma función de transferencia. Algoritmos para pasar de una formulación a otra se discuten en Aoki y Havenner (1991).

Hamilton (1994, p.375) indica un procedimiento para representar un modelo ARMA(p,q) en el espacio de estados. Definiendo $r = \max(p, q+1)$, un modelo ARMA(p,q) se puede representar como

$$(y_t - \mu) = \phi_1(y_{t-1} - \mu) + \phi_2(y_{t-2} - \mu) + \dots + \phi_r(y_{t-r} - \mu) + u_t + \theta_1 u_{t-1} + \dots + \theta_{r-1} u_{t-r+1}$$

donde $\phi_j = 0$ para $j > p$ y $\theta_j = 0$ para $j > q$.

Este modelo es formulable de forma equivalente en el espacio de estados mediante las ecuaciones

$$y_t = \mu + [1 \ \theta_1 \ \theta_1 \ \dots \ \theta_{r-1}] z_t$$

Identificación de Sistemas en el Espacio de Estados

$$\begin{pmatrix} \mathbf{z}_{t+1} \end{pmatrix} = \begin{pmatrix} \phi_1 & \phi_2 & \dots & \phi_{r-1} & \phi_r \\ 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \cdot & \cdot & \dots & \cdot & \cdot \\ 0 & 0 & \dots & 1 & 0 \end{pmatrix} \begin{pmatrix} \mathbf{z}_t \end{pmatrix} + \begin{pmatrix} \mathbf{u}_{t+1} \\ 0 \\ \cdot \\ 0 \end{pmatrix}$$

Pero esta transformación no es única, existiendo distintas opciones para pasar de una representación ARMA a una representación en el espacio de estados.

Como hemos indicado, dentro de las posibles representaciones en el espacio de estados, son de interés especialmente las denominadas representaciones mínimas (ver Anexo II), que utilizan la mínima dimensión posible del vector de estados.

Aoki (1976, 1990) y Terceiro (1990) indican un procedimiento para formular en el espacio de estados y mediante una representación mínima las ecuaciones de un modelo estructural lineal en forma reducida, es decir, un modelo de la forma

$$\mathbf{F}(\mathbf{L}) \mathbf{y}_t = \mathbf{G}(\mathbf{L}) \mathbf{x}_t + \mathbf{H}(\mathbf{L}) \mathbf{e}_t$$

donde \mathbf{y}_t es un vector de variables endógenas; \mathbf{x}_t es un vector de variables exógenas; \mathbf{e}_t es un vector de ruido blanco con $E(\mathbf{e}_t) = \mathbf{0}$ y $E(\mathbf{e}_t \mathbf{e}_t') = \mathbf{R}$; \mathbf{L} es el operador retardo; y $\mathbf{F}_0 = \mathbf{I}$, lo que caracteriza la forma reducida del modelo. Como vimos, los modelos VARMAX (y dentro de éstos los ARMA) son casos particulares de este modelo econométrico estructural lineal en forma reducida, que además cumplen $\mathbf{H}_0 = \mathbf{I}$.

Aunque, dado un modelo especificado, podemos encontrar representaciones VARMA y espacio de estados equivalentes, la especificación del modelo puede ser desconocida y constituir entonces uno de los objetivos de la identificación del sistema. Si la especificación del modelo es desconocida, los procedimientos de identificación basados en formulaciones distintas conducirán probablemente a modelos que no son exactamente equivalentes. Esto es debido a que los procedimientos de identificación asociados a cada formulación son distintos, con diferente complejidad, diferente capacidad de ajustar con parsimonia las restricciones de los datos, y, en consecuencia, distintos resultados.

Análisis de Sistemas Cointegrados mediante Métodos de Subespacios

Como hemos visto, la identificación de sistemas basada en la formulación en el espacio de estados presenta determinadas características (tanto en la especificación como en la estimación) que la hacen especialmente interesante, particularmente para el caso multivariable.

2 COINTEGRACIÓN

2.1 Introducción

En este capítulo desarrollamos los principales conceptos y métodos asociados al análisis de la cointegración de series no estacionarias. Una serie es estacionaria si su distribución de probabilidad no depende del tiempo, lo que implica una media y varianza uniformes en el tiempo. Muchas de las series del mundo real, tanto en economía como en ingeniería, resultan o parecen ser no estacionarias (una vez eliminados los posibles componentes deterministas de la serie). El tipo de series no estacionarias más habituales son las series integradas de orden 1 ó series $I(1)$, cuyas primeras diferencias son estacionarias. El tratamiento estadístico de estas series exige precauciones, porque los métodos de inferencia habituales pueden no ser válidos. Así por ejemplo, los estadísticos de contraste de regresión no siguen las distribuciones estándar correspondientes a procesos estacionarios, lo que da lugar a las denominadas regresiones espurias (con frecuencia se identifican como significativas relaciones que no existen realmente).

Una forma de abordar las series $I(1)$ es tomar diferencias, para trabajar con series estacionarias. Sin embargo, si existe una relación a largo plazo entre los niveles de las series estudiadas (cointegración), tomar diferencias hace perder esa relación. La cointegración corresponde a una relación a largo plazo entre variables $I(1)$, relación que no se detecta bien con los contrastes habituales de regresión.

Aunque el análisis de cointegración ha sido estudiado principalmente en el campo de la econometría, presenta también interés en ingeniería: el análisis de cointegración proporciona información sobre el proceso subyacente y permite construir modelos con menos parámetros a estimar y más precisos (suponiendo que realmente se produzca cointegración en el proceso generador subyacente). Larimore (2000) indica aplicaciones industriales en los procesos de fabricación de láminas y análisis de vibraciones. En econometría, el análisis de cointegración suele emplearse para contrastar la evidencia empírica de relaciones lineales a largo plazo establecidas por la teoría económica, y/o para estimar los coeficientes de esas relaciones lineales.

Después de presentar las ideas principales, discutiremos distintas representaciones habituales de un sistema cointegrado. Nosotros buscamos una representación en el

espacio de estados, y propondremos una a partir de la representación en forma de “tendencias comunes” (representación que presenta una clara similitud con la formulación en el espacio de estados). La representación que proponemos resulta estar relacionada con la forma canónica de Bauer y Wagner (2003).

Textos generales introductorios al concepto de cointegración pueden encontrarse en Greene (1999, cap. 18), y en Intriligator, Bodkin y Hsiao (1996, cap.11). Escribano (1990) realiza una revisión general de la literatura hasta el momento. Un análisis detallado se presenta en Hamilton (1994, cap. 19 y 20), y en Banerjee, Dolado, Galbraith y Hendry (1993). El enfoque de estimación máximo verosímil de sistemas cointegrados se desarrolla en Johansen (1988).

2.2 Series estacionarias, procesos no estacionarios y raíces unitarias

El concepto de cointegración está íntimamente ligado con el de estacionariedad. Un proceso aleatorio es estrictamente estacionario si su distribución de probabilidad no depende del tiempo. Un proceso aleatorio es débilmente estacionario (o estacionario en covarianza) si los dos primeros momentos de su distribución de probabilidad (media y covarianzas) no dependen del tiempo.

Con carácter general utilizaremos el término estacionario para los procesos débilmente estacionarios. Tenemos entonces que una serie y_t es estacionaria (débilmente o de covarianza estacionaria) si

$$E(y_t) = \mu$$

$$Var(y_t) = \sigma_y^2$$

$$Cov(y_t, y_s) = f(t-s)$$

donde μ y σ_y^2 son constantes para todo t , y donde $f(t-s)$ es una función que depende exclusivamente de la diferencia $(t-s)$ y no del periodo t .

Cointegración

Intuitivamente, una serie temporal estocástica es estacionaria si los valores se mueven alrededor de su valor medio, con un comportamiento (aspecto de las desviaciones respecto a la media, suavidad de cambio) similar a lo largo del tiempo.

Un ejemplo extremo de serie estacionaria es la serie denominada ruido blanco, esto es, una serie y_t que cumple

$$E(y_t) = 0,$$

$$\text{Var}(y_t) = \sigma^2$$

$$\text{Cov}(y_t, y_s) = 0.$$

Si una serie no cumple las condiciones de estacionariedad indicadas, entonces es no estacionaria.

Ejemplos típicos de series no estacionarias son:

Paseo aleatorio puro $y_t = y_{t-1} + \varepsilon_t$

Paseo aleatorio con deriva $y_t = \delta + y_{t-1} + \varepsilon_t$

Proceso estacionario con tendencia $y_t = a + \delta t + \varepsilon_t$

donde ε_t es ruido blanco. En el primer caso se tiene que la varianza de y_t no está acotada, aunque la esperanza sea estable; se trata de un camino aleatorio o paseo aleatorio. La segunda serie contiene, además, una tendencia causada por el término δ , al que se denomina “deriva”; se puede interpretar como una combinación de una tendencia determinista (causada por el término δ) y un paseo aleatorio. La tercera de las series contiene una tendencia determinista (δt) y es estacionaria en torno a dicha tendencia.

Fenómenos causados por la no estacionariedad

Durante mucho tiempo, la econometría ha limitado su análisis a los procesos estacionarios. Gouriéroux y Monfort (1997)¹ indican que la razón práctica de esta

¹ Traducción al inglés de la edición en francés de 1990

limitación es, probablemente, que los métodos econométricos clásicos (desarrollados bajo la hipótesis de estacionariedad) presentan propiedades complejas e incluso sorprendentes al ser aplicados a series no estacionarias. Así por ejemplo, los estimadores de regresión no siguen sus distribuciones habituales al aplicarse sobre series no estacionarias. Entre estos fenómenos asociados a la falta de estacionariedad destacan:

- La aparición de regresiones espurias. Este fenómeno consiste en la aparición de altos coeficientes de correlación R^2 entre procesos independientes (por ejemplo, entre dos caminos aleatorios independientes), de forma que su relación lineal sería juzgada como significativa bajo los contrastes habituales.

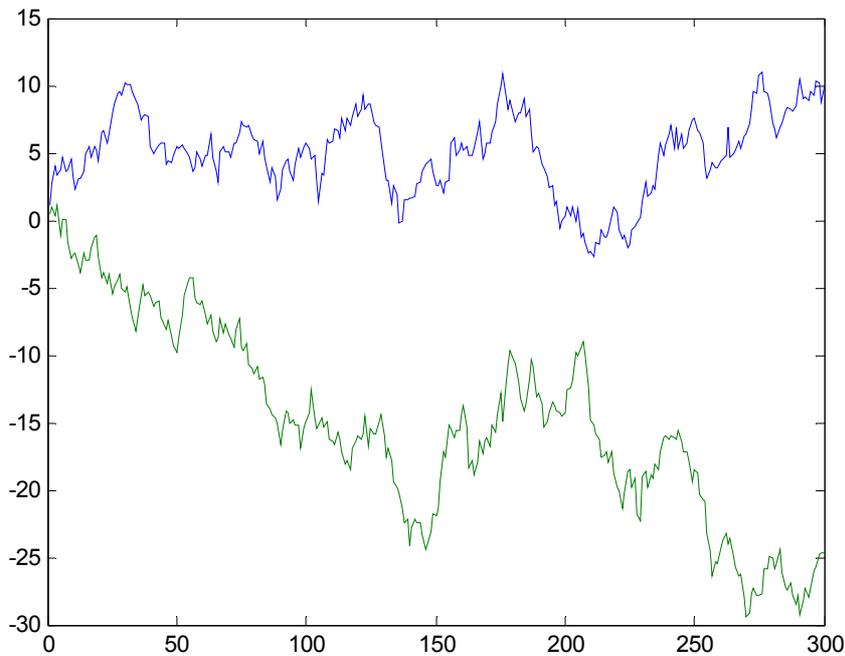


Figura 2.1. Una realización de dos caminos aleatorios independientes, con origen cero

Cointegración

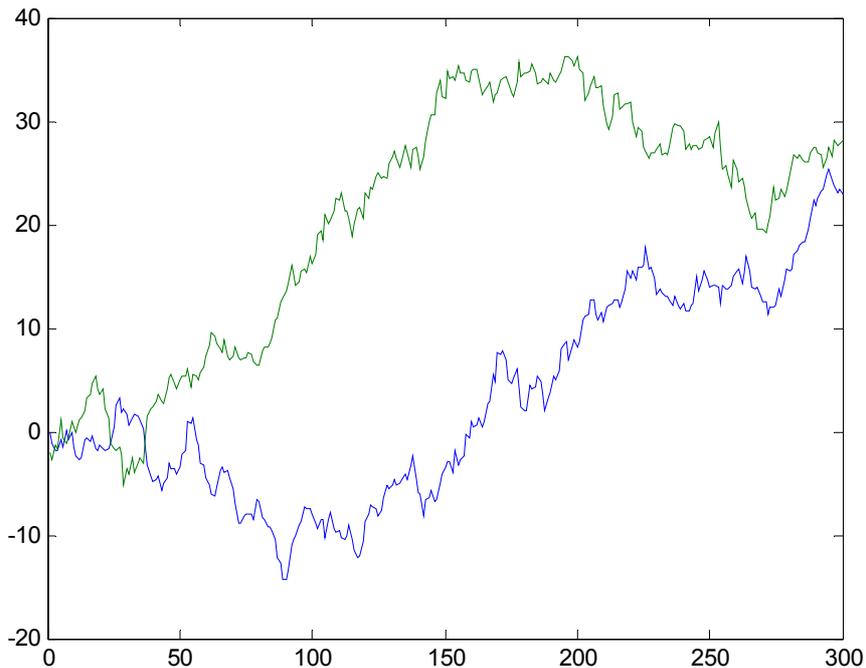


Figura 2.2. Ejemplo de realización de dos caminos aleatorios independientes. Se simuló con origen cero para ambos. Aunque la casualidad ha provocado que las series volvieran a juntarse, no existe ninguna fuerza que las obligue a converger o a mantener una relación lineal determinada.

- Una velocidad de convergencia mayor de los estimadores de mínimos cuadrados ordinarios (MCO) de coeficientes autorregresivos. Así por ejemplo, en el modelo autorregresivo $y_t = \rho y_{t-1} + e_t$ con $|\rho| < 1$ (caso estacionario), el estimador MCO de ρ converge al verdadero valor del parámetro con velocidad de convergencia $1/\sqrt{T}$; en cambio, si el valor de ρ es 1 (caso no estacionario), la velocidad de convergencia es $1/T$. Este fenómeno se conoce como superconsistencia. Además, el sesgo que en el caso estacionario se produce en estos estimadores debido a perturbaciones autocorrelacionadas, tiende a cero rápidamente en el caso no estacionario, por lo que los estimadores MCO mantienen la consistencia.
- Los estadísticos t-Student asociados al contraste $\rho = 1$ no convergen a una distribución normal $N(0,1)$, lo que exige modificar los procedimientos de contraste habituales. Las distribuciones asintóticas de estos estadísticos son distribuciones inusuales.

Así pues, el tratamiento de series no estacionarias exige algunas precauciones.

Tendencias deterministas y tendencias estocásticas

Como indica Hamilton (1994), pese al amplio desarrollo de la teoría de series temporales estacionarias, muchas de las series económicas y financieras que se encuentran en la práctica presentan tendencias (piénsese, p.e., en el PIB nominal), las cuales no son representables mediante procesos estacionarios¹.

Hay dos aproximaciones popularizadas para describir estas tendencias. La primera es considerar una tendencia temporal determinista, tal como $(\alpha + \delta t)$, más un término estacionario $\Psi(L)e_t$:

$$y_t = \alpha + \delta t + \Psi(L)e_t$$

En este caso los efectos de una perturbación concreta puntual e_t sobre el nivel de la serie desaparecen con el tiempo, dado el carácter estacionario de $\Psi(L)e_t$.

La segunda especificación es un proceso con raíz unitaria:

$$(1-L) y_t = \delta + \Psi(L)e_t$$

donde $\Psi(L)e_t$ es estacionario y $\Psi(1)$ es distinto de cero². El ejemplo más típico es el del camino aleatorio con deriva: $y_t = y_{t-1} + \delta + e_t$. En este caso, la desviación de la serie y_t respecto a la tendencia lineal $(y_0 + \delta t)$ no es estacionaria, ya que es una desviación de varianza creciente (la propia desviación es un camino aleatorio). Además, el efecto de una perturbación e_t sobre el nivel esperado de los valores futuros de la serie es permanente.

Nótese que el modelo de tendencia determinista $y_t = \alpha + \delta t + e_t$ y el modelo de camino aleatorio con deriva $y_t = y_{t-1} + \delta + e_t$ son ambos representables de la forma

$$y_t = \alpha + \delta t + a_t$$

¹ Un proceso estacionario implica media constante en el tiempo.

² Esta condición asegura que la serie original no es estacionaria. Hamilton (1994, p.436).

Cointegración

pero en el caso de tendencia determinista, el factor a_t es estacionario, mientras que en el caso de camino aleatorio con deriva, el factor a_t (desviación respecto a la tendencia determinista δt) es a su vez un camino aleatorio.

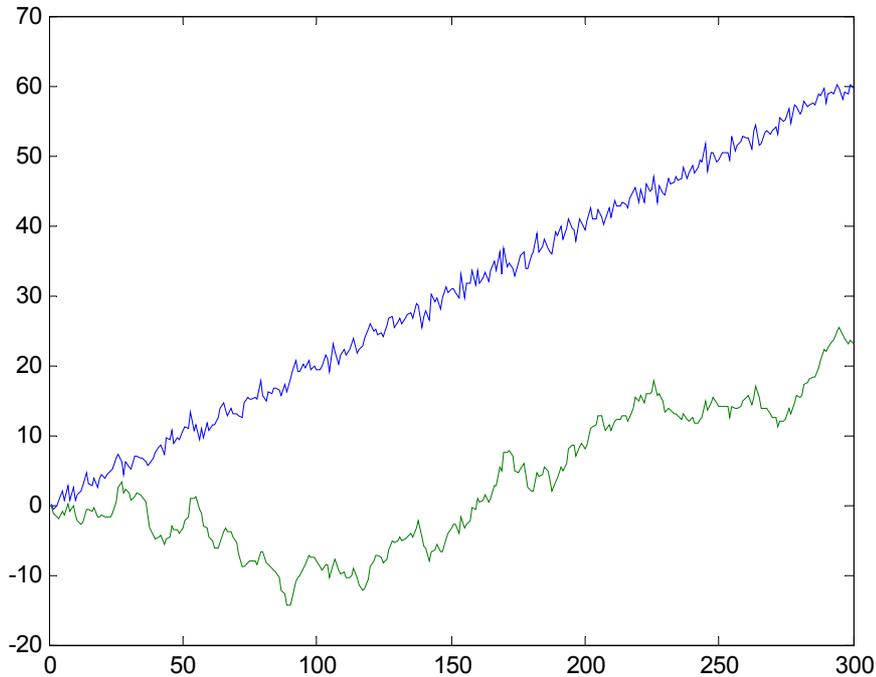


Figura 2.3. Representación de una serie estacionaria con tendencia lineal determinista ($y_t = 0,2 t + e_t$, serie superior) y una serie no estacionaria con deriva y tendencia estocástica ($y_t = 0,2 + y_{t-1} + e_t$, serie inferior), para los mismos valores de la perturbación e_t .

Hamilton (1994, cap. 15) desarrolla una comparación detallada de ambos tipos de tendencias. Un resultado interesante es que la transformación adecuada para conseguir estacionariedad es diferente en cada caso: eliminación de la tendencia lineal en el caso determinista y diferenciación en el caso de raíz unitaria (la diferenciación de una serie con tendencia lineal determinista introduce un término de media móvil con raíz unitaria, no invertible, que genera dificultades en su tratamiento).

Estas dos tendencias de carácter lineal ofrecen bastante flexibilidad para representar series no estacionarias. Aunque muchas de las series reales en economía presentan un

crecimiento de tipo exponencial¹ ($y_t = e^{\delta t}$), en vez de lineal, una transformación logarítmica daría como resultado una tendencia lineal: $\log(y_t) = \delta t$.

Separación entre tendencias deterministas y estocásticas

El que una serie económica sea modelada con una tendencia estocástica o determinista tiene implicaciones económicas importantes. De esto depende, por ejemplo, que el efecto de una recesión sobre el PIB tenga consecuencias permanentes (en el modelo) o que su efecto a largo plazo sea despreciable.

Nelson y Plosser (1982) establecieron la diferenciación entre ambos tipos de tendencias y desarrollaron contrastes estadísticos para juzgar qué modelo resulta más apropiado para unos datos observados. Argumentaban que muchas series económicas se caracterizan mejor mediante raíces unitarias (tendencias estocásticas) que mediante tendencias deterministas. Hamilton (1994) presenta un resumen de varios trabajos realizados en los 90 sobre este tema: aunque sería muy interesante saber si una serie tiene o no una raíz unitaria, no es posible dar respuesta a esta pregunta a partir de un número finito de observaciones muestrales. El motivo es que, dado un proceso con raíz unitaria, siempre existe un proceso estacionario lo suficientemente próximo a él² como para que no sea discriminable en base a un número de datos dado. El recíproco también es cierto.

Dicho esto, en vez de la cuestión de si modelar con tendencia estocástica o determinista, quizá tiene más sentido plantearse cuestiones tales como si las perturbaciones tienen un efecto significativo sobre el nivel de la serie durante un horizonte temporal definido.

Series integradas I(d)

En general, a partir de una serie no estacionaria es posible obtener una serie estacionaria, mediante una adecuada transformación de los datos. Una de las transformaciones más habituales para buscar la estacionariedad es la toma de diferencias entre un término y el siguiente. Siendo L el operador retardo, tal que $Ly_t =$

¹ Una tasa de crecimiento porcentual constante supone un crecimiento exponencial para una serie.

² Con alguna raíz lo suficientemente próxima a uno.

Cointegración

y_{t-1} , y siendo Δ el operador diferencia, tal que $\Delta y_t = y_t - y_{t-1} = (1-L) y_t$, la serie diferenciada de y_t es

$$z_t = y_t - y_{t-1} = \Delta y_t = (1-L) y_t$$

De forma general, si una serie y_t sin componentes deterministas puede representarse, después de diferenciar d veces, como un proceso ARMA estacionario e invertible, la serie es integrada de orden d , representado como $I(d)$.

Así pues, una serie estacionaria es $I(0)$, y una serie no estacionaria sin componentes deterministas es $I(1)$ si su serie de primeras diferencias $\Delta y_t = (1-L) y_t$ es estacionaria.

Ejemplos típicos de series integradas $I(1)$ son el paseo aleatorio puro ($y_t = y_{t-1} + \varepsilon_t$) y el paseo aleatorio con deriva ($y_t = \alpha + y_{t-1} + \varepsilon_t$). Ambas cumplen que Δy_t es estacionaria:

$$\Delta y_t = \mu + \varepsilon_t$$

La metodología de análisis de series temporales está centrada en las series estacionarias, por lo que el tratamiento de series no estacionarias pasa típicamente por su transformación a estacionarias y su análisis mediante los métodos VAR o Box-Jenkins.

Sin embargo, la transformación adecuada no siempre es evidente, y una transformación para conseguir estacionariedad puede implicar pérdida de información. Ya hemos indicado que la diferenciación de la serie no es adecuada si la tendencia es determinista. Otra situación en la que la diferenciación no es adecuada, porque supone perder la relación a largo plazo existente entre los niveles de dos (o más) series temporales, es el caso de cointegración.

2.3 Contrastes de raíces unitarias

El primer paso en el análisis de cointegración de unas variables es comprobar que todas ellas están integradas con el mismo orden, es decir, que todas son $I(d)$. En economía no son frecuentes las series con orden de integración superior a uno, por lo que nos limitaremos al caso en que las series presentan tendencias estocásticas $I(1)$.

Análisis de Sistemas Cointegrados mediante Métodos de Subespacios

Para comprobar que las series en estudio son $I(1)$ suelen utilizarse los contrastes de tipo Dickey-Fuller. Estos contrastes parten de las hipótesis de que la serie y_t tiene una raíz unitaria y sigue un proceso autorregresivo de orden p :

$$y_t = a_1 y_{t-1} + \dots + a_p y_{t-p} + \varepsilon_t$$

donde ε_t es i.i.d. $(0, \sigma^2)$. Expresando (para $i = 2, \dots, p$) cada término y_{t-i} , en función de y_{t-1} y de los incrementos $\Delta y_{t-1}, \dots, \Delta y_{t-i+1}$, podemos reformular la expresión anterior como:

$$y_t = (a_1 + \dots + a_p) y_{t-1} + b_1 \Delta y_{t-1} + \dots + b_{p-1} \Delta y_{t-p+1} + \varepsilon_t$$

y se tiene

$$\Delta y_t = \gamma^* y_{t-1} + b_1 \Delta y_{t-1} + \dots + b_{p-1} \Delta y_{t-p+1} + \varepsilon_t$$

donde $\gamma^* = a_1 + \dots + a_p - 1$.

Ahora bien, si y_t tiene una raíz unitaria, entonces $(a_1 + \dots + a_p) = 1$, y por tanto $\gamma^* = 0$. Se estima γ^* en la regresión anterior y se compara el estadístico normalizado en las tablas correspondientes, bajo la hipótesis nula de que la serie es $I(1)$, es decir, bajo la hipótesis nula $H_0: \gamma^* = 0$, frente a la hipótesis alternativa $H_1: \gamma^* < 0$. El caso $\gamma^* > 0$ no se contempla porque corresponde a procesos no estables de comportamiento explosivo, en general fáciles de distinguir (incluso en muestras pequeñas) de procesos estacionarios en diferencias, por lo que de partida se asume $\gamma^* \leq 0$.

Si el γ^* estimado es suficientemente negativo, rechazamos la hipótesis nula y admitimos que la serie es estacionaria. Lamentablemente, el estadístico normalizado del contraste de la t (dividiendo el estimador de regresión de γ^* entre el estimador de la desviación típica) no sigue, bajo la hipótesis nula, la distribución t estándar, ni ninguna otra distribución estándar. Tampoco la distribución asintótica es normal. A modo de ejemplo, en el modelo $\Delta y_t = \gamma^* y_{t-1} + \varepsilon_t$, la distribución asintótica del estimador de regresión del contraste de la t bajo la hipótesis $H_0: \gamma^* = 0$, es (Hamilton 1994, p. 489)

$$\frac{(1/2) \{ [W(1)]^2 - 1 \}}{\left\{ \int_0^1 [W(r)]^2 dr \right\}^{1/2}}$$

Cointegración

donde $W(r)$ representa un movimiento browniano estándar (Hamilton 1994, p. 478).

Los valores críticos para estos estadísticos se calculan por simulación y pueden encontrarse tabulados para distintos órdenes p y tamaños muestrales. Los contrastes pueden ampliarse para considerar términos deterministas, como derivas y tendencias.

Los valores críticos son substancialmente menores (más negativos) que los valores de la distribución t , como puede verse en la **Tabla 2-1**, que muestra valores críticos del estadístico t normalizado en el contraste de Dickey-Fuller para el modelo

$$\Delta y_t = \alpha + \gamma^* y_{t-1} + \varepsilon_t$$

Nivel de significación	Tamaño muestral					t_∞ ó $N(0,1)$
	25	50	100	500	∞	
0,01	-3,75	-3,58	-3,51	-3,44	-3,43	-2,33
0,05	-3,00	-2,93	-2,89	-2,87	-2,86	-1,65
0,1	-2,63	-2,60	-2,58	-2,57	-2,57	-1,28

Tabla 2-1. Valores críticos para el contraste de Dickey-Fuller basado en el estadístico t , modelo $\Delta y_t = \alpha + \gamma^* y_{t-1} + \varepsilon_t$

Existen diversas modificaciones y variantes propuestas de contrastes tipo Dickey-Fuller, entre las que citamos en particular el contraste DF-GLS (Elliott, Rothenberg y Stock, 1996), que contrasta la hipótesis nula de que la serie es un camino aleatorio con deriva frente a la hipótesis alternativa de que la serie es estacionaria en torno a una tendencia lineal determinista.

2.4 Cointegración de variables I(1)

Sean y_t, z_t series no estacionarias I(1). Sea la diferencia $\varepsilon_t = y_t - \beta z_t$. En general, esta diferencia ε_t será también no estacionaria I(1), pero puede que exista un β tal que ε_t sea I(0). Si es así, la diferencia $y_t - \beta z_t$ es estacionaria, “estable” alrededor de una media fija. Se dice entonces que las dos series están cointegradas, con vector de cointegración $[1, -\beta]'$.

Esto nos permite estudiar una relación a largo plazo definida por el vector de cointegración y una relación a corto plazo marcada por la serie estacionaria de residuos. Diferenciar las dos series haría perder la relación a largo plazo. Además, si hay cointegración, la representación VAR de las series diferenciadas contiene infinitos términos (si no incluimos niveles, sólo diferencias. Hamilton 1994, p. 573), por lo que no es práctico diferenciar.

Un ejemplo de sistema cointegrado es

$$y_{1t} = \gamma y_{2t} + \varepsilon_{1t}$$

$$y_{2t} = y_{2, t-1} + \varepsilon_{2t}$$

donde $\varepsilon_{1t}, \varepsilon_{2t}$ son ruidos blancos independientes y γ es una constante. Obsérvese que las dos series son I(1), pero la relación $y_{1t} - \gamma y_{2t}$ es estacionaria.

Intuitivamente, que dos series cointegran quiere decir que podemos conseguir que se muevan “casi” juntas si a una de ellas le cambiamos la escala y le sumamos un valor constante.

A modo de ejemplo se han representado (figuras 2.4, 2.5 y 2.6) distintas realizaciones del sistema cointegrado

$$y_{1t} = b + \gamma y_{2t} + \varepsilon_{1t}$$

$$y_{2t} = \pi + y_{2, t-1} + \varepsilon_{2t}$$

que incluye una deriva π y un desplazamiento b (también llamado “término constante de la relación de cointegración”).

Cointegración

En los tres casos se ha tomado una deriva de valor $\pi = 1$. En cada caso se ha modificado el vector de cointegración $[1 - \gamma]'$, el desplazamiento b y la matriz de covarianzas de las perturbaciones.

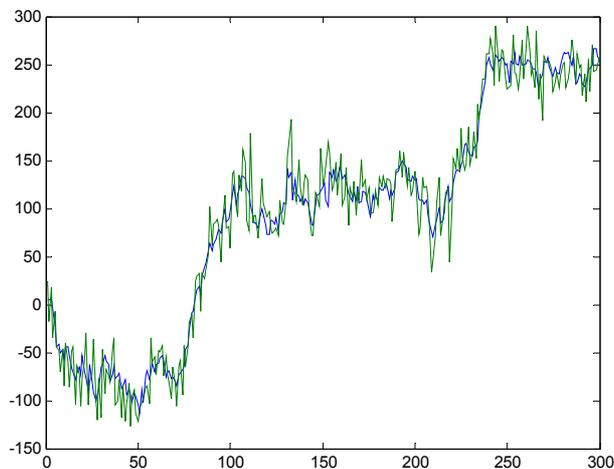


Figura 2.4. Sistema cointegrado con vector de cointegración $[1 - 1]'$ y desplazamiento nulo.

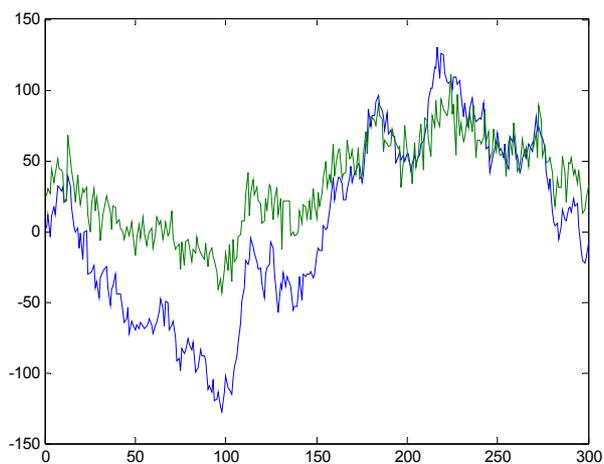


Figura 2.5. Sistema cointegrado con vector de cointegración $[1 - 0,5]'$ y desplazamiento 30.

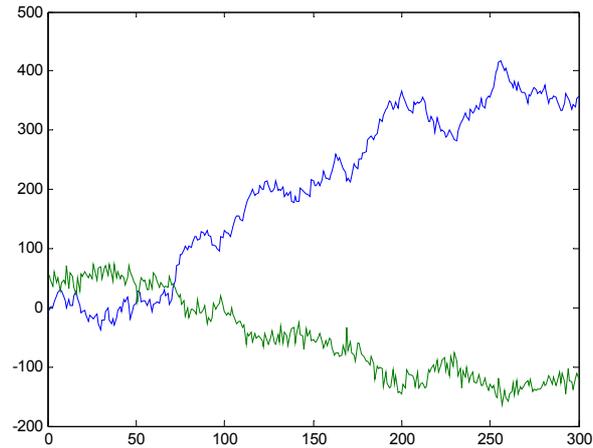


Figura 2.6. Sistema cointegrado con vector de cointegración $[1 \ 0,5']$ y desplazamiento 50.

Stock y Watson (2003) indican tres métodos no excluyentes para decidir si dos variables pueden ser modeladas de forma plausible como un sistema cointegrado:

- El conocimiento experto y la teoría económica.
- Representar gráficamente las series y observar si parece existir una tendencia estocástica común.
- Realizar contrastes estadísticos de cointegración. Estos suelen partir de dos series $I(1)$ y la hipótesis nula de que la relación $(y_{1t} - \gamma y_{2t})$ es también $I(1)$, es decir, no estacionaria. Si se rechaza la hipótesis nula¹, se admite que el sistema está cointegrado.

Ejemplos de cointegración teórica

Ejemplos típicos de aplicación de la teoría de cointegración son los estudios de la paridad del poder adquisitivo entre distintos países (Fuertes 1997) y el estudio de la relación entre los ingresos y el consumo. En ambos casos se trata de series no estacionarias para las que la teoría económica establece una relación a largo plazo.

¹ El contraste es diferente dependiendo de que el vector de cointegración sea conocido a priori o se estime a partir de las observaciones.

Cointegración

La teoría de la paridad del poder adquisitivo (purchasing power parity, o PPP) establece que el precio efectivo de los bienes tiende a ser el mismo en distintos países, de modo que el tipo de cambio ajusta las diferencias que se produzcan en los niveles de precios. Tendríamos un modelo de la forma

$$P_t \approx S_t P^*_t$$

donde P es el precio de unos bienes en €, P^* es el precio en \$ y S el tipo de cambio (precio del \$ en €). Según este modelo, y trabajando con los logaritmos, se tiene que

$$\log P_t = \log S_t + \log P^*_t + z_t$$

donde la variable z_t es estacionaria, pese a que cada uno de los otros términos de la expresión son $I(1)$.

Otro caso típico de cointegración teórica es el de la relación entre el consumo y los ingresos: el consumo C_t tiende a ser una proporción constante de los ingresos I_t , y ambas series son no estacionarias. Conforme a esta teoría, la diferencia $\log C_t - \log I_t$ sería una variable estacionaria y las dos series estarían cointegradas.

Cointegración multivariable. Vectores y rango de cointegración

Con carácter más general que el caso de dos variables, considerando un vector y_t de m variables no estacionarias $I(d)$, si existe un vector $\beta \neq \mathbf{0}$ tal que $\beta'y_t$ es $I(d-b)$ para algún $b > 0$, entonces se dice que las variables y_t están cointegradas $CI(d,b)$. A β se le denomina vector de cointegración. Con más de dos variables no estacionarias, puede ser que existan varios vectores linealmente independientes de cointegración.

El caso más habitual es el de series $I(1)$. Siendo y_t el vector de m variables $I(1)$, podemos tener r vectores independientes de cointegración ($0 \leq r < m$) agrupables en filas en la matriz Γ de dimensiones $r \times m$, de modo que $\Gamma y_t = u_t$, donde u_t es $I(0)$. A r se le denomina rango de cointegración, y es el rango de la matriz Γ .

Los vectores de cointegración forman un subespacio vectorial. Para expresar una base de este subespacio podemos tomar unas condiciones de normalización.

Obsérvese que si Γy_t es estacionario, entonces $C\Gamma y_t$ es también estacionario, para cualquier matriz C constante $r \times r$. Si Γ es una base del subespacio de cointegración y C es no singular, entonces $\Gamma^* = C\Gamma$ es otra base del subespacio. Para normalizar, se puede exigir que la matriz de vectores independientes de cointegración $\Gamma^* = C\Gamma$ sea de la forma (I_r, Γ_2^*) . Podemos llegar a una base de esta forma partiendo de una descomposición de una base cualquiera $\Gamma = (\Gamma_1, \Gamma_2)$ de modo que Γ_1 sea $r \times r$ y no singular (esto puede exigir un determinado orden de las variables). Tomando $C = \Gamma_1^{-1}$ tenemos

$$\Gamma^* = \Gamma_1^{-1}\Gamma = (I_r, \Gamma_1^{-1}\Gamma_2)$$

y obtendríamos una base de vectores de cointegración de la forma buscada. Esta base corresponde a r ecuaciones de cointegración que relacionan cada una de las r primeras variables con las $m-r$ variables finales, de forma que los residuos son estacionarios.

2.5 Formulaciones de modelos cointegrados

Entre las representaciones alternativas habituales de un sistema cointegrado destacan la representación en forma de tendencias comunes y la representación denominada de “corrección del error”. En este apartado introducimos una posible forma de representación de sistemas cointegrados en el espacio de estados partiendo de la representación en forma de tendencias comunes.

Representación en forma de tendencias comunes

Stock y Watson (1988) demuestran que si las m variables $I(1)$ del vector y_t cointegran con rango de cointegración r , entonces podemos encontrar $m-r$ tendencias comunes τ_t (camino aleatorios con tendencia) tales que

$$\tau_t = \pi + \tau_{t-1} + H \varepsilon_t$$

$$y_t = y_0 + C \tau_t + \phi(L) \varepsilon_t$$

donde ε_t es ruido blanco iid $(0, \Omega)$ y $\phi(L) \varepsilon_t$ es estacionario.

Cointegración

Intuitivamente, esto quiere decir que todas las variables pueden expresarse en función de $m-r$ paseos aleatorios con tendencia, o tendencias comunes, o “factores comunes”¹ ocultos, más un residuo estacionario. Ya anteriormente habíamos hecho notar que en un sistema cointegrado con rango r , r variables pueden expresarse como combinación lineal de las otras $(m-r)$ variables, salvo por un residuo estacionario.

En cuanto a la recuperación de los vectores de cointegración a partir de un modelo de tendencias comunes, nótese que, si conocemos la matriz C de la ecuación de observaciones, las relaciones de cointegración vienen definidas por una base del subespacio ortogonal a C' , ya que, por definición, los vectores de cointegración son vectores b , de dimensión m , tales que $b' y_t$ es $I(0)$. Para que $b' y_t$ sea $I(0)$, teniendo en cuenta que τ_t es $I(1)$ y que

$$y_t = y_0 + C \tau_t + \phi(L) \varepsilon_t$$

debe cumplirse que $b'C = 0$. Por otro lado, si $b'C = 0$, entonces $b' y_t$ es $I(0)$. Así pues, los vectores de cointegración son soluciones de $b'C = 0$.

Representación de un sistema cointegrado en el espacio de estados

Obsérvese que la representación en forma de tendencias comunes de un vector y_t de m series temporales $I(1)$:

$$\tau_t = \pi + \tau_{t-1} + v_t$$

$$y_t = y_0 + C \tau_t + u_t$$

donde v_t es ruido blanco y u_t es $I(0)$, es una representación muy próxima al espacio de estados, en la que la matriz de transición de estados sería la matriz identidad, y en la que tendríamos tantos estados como tendencias comunes. Esta representación indica que la información no estacionaria de las m series y_t cointegradas puede condensarse en $(m-r)$ estados no estacionarios (tendencias comunes).

¹ El enfoque de “factores comunes”, que puede encontrarse en Escribano y Peña (1994), está muy relacionado con el de tendencias comunes de Stock y Watson, si bien es más general (Peña, 1990).

La representación en forma de tendencias comunes no es exactamente una representación en el espacio de estados porque los residuos estacionarios \mathbf{u}_t no son ruido blanco, sino que pueden presentar autocorrelación temporal. Para representarlo en el espacio de estados podríamos modelar las perturbaciones estacionarias \mathbf{u}_t en el espacio de estados:

$$\boldsymbol{\eta}_t = \mathbf{A}_\eta \boldsymbol{\eta}_{t-1} + \mathbf{K}_\tau \mathbf{e}_t$$

$$\mathbf{u}_t = \mathbf{C}_\eta \boldsymbol{\eta}_t + \mathbf{e}_t$$

y anidar este modelo en el de tendencias comunes, lo cual supone combinar (m-r) estados no estacionarios (tendencias comunes) con otra serie de estados estacionarios utilizados para modelar los residuos estacionarios \mathbf{u}_t .

Prescindiendo de los componentes deterministas, un modelo completo en el espacio de estados, en forma de innovaciones, para m series \mathbf{y}_t no estacionarias I(1) cointegradas con r relaciones de cointegración, sería de la forma

$$\begin{bmatrix} \boldsymbol{\tau}_{t+1} \\ \boldsymbol{\eta}_{t+1} \end{bmatrix} = \begin{bmatrix} \mathbf{I}_{n-r} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_\eta \end{bmatrix} \begin{bmatrix} \boldsymbol{\tau}_t \\ \boldsymbol{\eta}_t \end{bmatrix} + \begin{bmatrix} \mathbf{K}_\tau \\ \mathbf{K}_\eta \end{bmatrix} \mathbf{e}_t$$

$$\mathbf{y}_t = \begin{bmatrix} \mathbf{C}_\tau & \mathbf{C}_\eta \end{bmatrix} \begin{bmatrix} \boldsymbol{\tau}_t \\ \boldsymbol{\eta}_t \end{bmatrix} + \mathbf{e}_t$$

donde \mathbf{e}_t es i.i.d. $(\mathbf{0}, \mathbf{R})$ y existen n-r tendencias comunes $\boldsymbol{\tau}_t$ que son caminos aleatorios cuya dinámica es independiente del valor de los estados estacionarios.

Bauer y Wagner (2003), partiendo de un enfoque distinto al aquí expuesto, proponen esta misma forma de representación de sistemas cointegrados en el espacio de estados, y profundizan en las condiciones que se deben cumplir para que esta forma de representación constituya una forma canónica (identificable).

Nótese que un cambio de base de los estados daría lugar a una representación equivalente en el espacio de estados en la que las matrices del sistema han sufrido la

Cointegración

transformación correspondiente, de modo que no tiene por qué conservarse la estructura diagonal por bloques de la matriz de transición de estados **A** (aunque sí se mantendrían sus autovalores), y tendríamos representaciones totalmente equivalentes a la anterior en las que los componentes tendencial y cíclico (estacionario) no evolucionan por separado. Volveremos sobre este punto al estudiar el método de Aoki para procesos no estacionarios.

Teniendo en cuenta que los autovalores de la matriz de transición de estados **A** son invariantes a la elección de sistema coordenado, se tiene que, como indica Gilbert (1993, p.17), un modelo de n series temporales cointegra con rango r ($0 < r < n$) si la matriz dinámica del sistema (la matriz **A** de transición de estados) contiene $n-r$ autovalores unitarios y el resto de autovalores tiene módulo menor que uno, y cada uno de los n modelos individuales para las observaciones contiene un autovalor unitario.

Recordamos en este punto que, teóricamente, supuestas perturbaciones gaussianas, la estimación de los parámetros se podría llevar a cabo calculando mediante el filtro de Kalman la función de verosimilitud y maximizándola a continuación. Como indicamos en la introducción, esta alternativa no se ha desarrollado por las desventajas que de partida ofrece respecto al método de Johansen, que también realiza una estimación máximo-verosímil, pero proporciona los estimadores máximo-verosímiles en un solo paso (tras realizar unas regresiones auxiliares), sin necesidad de realizar una búsqueda numérica del máximo de una función complicada. Aunque bajo el enfoque de máxima verosimilitud quedaría abierta la posibilidad de competir con el método de Johansen allí donde no se cumplen bien las hipótesis para su aplicación (en especial, la hipótesis de que la serie sigue un VAR en niveles), los experimentos prácticos— ver Gonzalo (1994) o Wagner (2000)— indican que el método resulta robusto ante desviaciones en las hipótesis .

Modelo de corrección del error

La cointegración es equivalente a la existencia de un proceso de ajuste que impide que los errores respecto a la relación a largo plazo crezcan indefinidamente. Este aspecto es visible en el modelo llamado de “corrección del error”, en el que se relacionan los

incrementos de las variables cointegradas con la desviación respecto al equilibrio a largo plazo.

Para el caso general de n series temporales, Hamilton (1994, p. 580) demuestra que si el vector de n series no estacionarias \mathbf{y}_t puede representarse como un vector autorregresivo de orden p :

$$\mathbf{y}_t = \boldsymbol{\alpha} + \boldsymbol{\Phi}_1 \mathbf{y}_{t-1} + \boldsymbol{\Phi}_2 \mathbf{y}_{t-2} + \dots + \boldsymbol{\Phi}_p \mathbf{y}_{t-p} + \boldsymbol{\varepsilon}_t$$

$$\text{ó, de forma compacta, } \boldsymbol{\Phi}(L) \mathbf{y}_t = \boldsymbol{\alpha} + \boldsymbol{\varepsilon}_t$$

y si las series \mathbf{y}_t cointegran con r relaciones de cointegración, entonces se tiene que $\boldsymbol{\Phi}(z)$ contiene $(n-r)$ raíces unitarias y el sistema se puede representar de la forma

$$\Delta \mathbf{y}_t = \boldsymbol{\zeta}_1 \Delta \mathbf{y}_{t-1} + \boldsymbol{\zeta}_2 \Delta \mathbf{y}_{t-2} + \dots + \boldsymbol{\zeta}_{p-1} \Delta \mathbf{y}_{t-p+1} + \boldsymbol{\alpha} + \boldsymbol{\zeta}_0 \mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t$$

con $\boldsymbol{\zeta}_0 = -\mathbf{B}\mathbf{A}'$ y donde la matriz \mathbf{A}' , de dimensiones $r \times n$, contiene por filas los vectores de cointegración. Esta es la representación vectorial en forma de corrección del error (VEC, vector error correction) de un sistema cointegrado.

Observando que el término $\Delta \mathbf{y}_t$ es estacionario, este modelo sólo es internamente consistente si las variables \mathbf{y}_t están cointegradas con vectores de cointegración contenidos en $\boldsymbol{\zeta}_0$.

Es fácil comprobar que la relación entre los coeficientes $\boldsymbol{\Phi}_i$ de una representación VAR y los coeficientes $\boldsymbol{\zeta}_i$ de una representación VEC es:

$$\boldsymbol{\zeta}_0 = -(\mathbf{I}_n - \boldsymbol{\Phi}_1 - \boldsymbol{\Phi}_2 - \dots - \boldsymbol{\Phi}_p) = -\boldsymbol{\Phi}(1)$$

$$\boldsymbol{\zeta}_s = -(\boldsymbol{\Phi}_{s+1} + \boldsymbol{\Phi}_{s+2} + \dots + \boldsymbol{\Phi}_p) \text{ para } s = 1, 2, \dots, p-1$$

Utilizaremos la representación vectorial en forma de corrección del error (VEC) en el apartado 2.6, al estudiar el método de Johansen.

2.6 Análisis estadístico de sistemas cointegrados. Método de Johansen

El análisis de cointegración en series I(1) consiste en

- 1 Comprobar que las series son I(1): contraste de raíces unitarias.
- 2 Estimar el rango de cointegración.
- 3 Estimar los vectores de cointegración.

Como ya indicamos (apartado 2.3), para comprobar que las series en estudio son I(1) suelen utilizarse los contrastes de tipo Dickey-Fuller.

Para estimar el rango y los vectores de cointegración, el método que mayor popularidad ha alcanzado es el método de máxima verosimilitud de Johansen (1988), aunque se han ido proponiendo diversas técnicas en la literatura (Engel y Granger 1987; Stock y Watson 1988). Gonzalo (1994) realiza un estudio comparativo de cinco métodos alternativos de estimación de las relaciones de cointegración: mínimos cuadrados ordinarios (Engel y Granger), mínimos cuadrados no lineales, componentes principales (Stock y Watson), correlaciones canónicas, y máxima verosimilitud (Johansen). De su estudio se desprende que el método de Johansen es el que presenta las mejores propiedades para estimar los vectores de cointegración.

El método de Johansen parte de las siguientes hipótesis:

- El vector \mathbf{y}_t de n series temporales I(1), sigue un VAR(p), y admite por tanto una formulación en forma de corrección del error (VEC¹):

$$\Delta \mathbf{y}_t = \zeta_1 \Delta \mathbf{y}_{t-1} + \zeta_2 \Delta \mathbf{y}_{t-2} + \dots + \zeta_{p-1} \Delta \mathbf{y}_{t-p+1} + \boldsymbol{\alpha} + \zeta_0 \mathbf{y}_{t-1} + \mathbf{e}_t$$

- Las perturbaciones son ruido blanco gaussiano $N(\mathbf{0}, \boldsymbol{\Omega})$.

¹ A partir de la cual puede obtenerse, si se prefiere, una formulación VAR en niveles (apartado 2.5).

- El número de tendencias comunes es h . Esto implica que el número de relaciones de cointegración es $r = n - h$, y que ζ_0 puede escribirse como $\zeta_0 = -\mathbf{BA}'$, donde \mathbf{A}' es de dimensiones $r \times n$.

Consideremos una muestra de $T + p$ observaciones de \mathbf{y} : $(\mathbf{y}_{-p+1}, \mathbf{y}_{-p+2}, \dots, \mathbf{y}_T)$. Si las perturbaciones son gaussianas, entonces el logaritmo de la función de verosimilitud de $(\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_T)$ condicionados a $(\mathbf{y}_{-p+1}, \mathbf{y}_{-p+2}, \dots, \mathbf{y}_0)$ viene dado por

$$\begin{aligned}
 L(\boldsymbol{\Omega}, \zeta_1, \zeta_2, \dots, \zeta_{p-1}, \boldsymbol{\alpha}, \zeta_0) \\
 = - (Tn/2) \log(2\pi) - (T/2) \log |\boldsymbol{\Omega}| \\
 - (1/2) \sum_{t=1}^T [(\Delta \mathbf{y}_t - \zeta_1 \Delta \mathbf{y}_{t-1} - \zeta_2 \Delta \mathbf{y}_{t-2} - \dots - \zeta_{p-1} \Delta \mathbf{y}_{t-p+1} - \boldsymbol{\alpha} - \zeta_0 \mathbf{y}_{t-1})' \boldsymbol{\Omega}^{-1} \\
 \times (\Delta \mathbf{y}_t - \zeta_1 \Delta \mathbf{y}_{t-1} - \zeta_2 \Delta \mathbf{y}_{t-2} - \dots - \zeta_{p-1} \Delta \mathbf{y}_{t-p+1} - \boldsymbol{\alpha} - \zeta_0 \mathbf{y}_{t-1})]
 \end{aligned}$$

El objetivo es encontrar los valores $(\boldsymbol{\Omega}, \zeta_1, \zeta_2, \dots, \zeta_{p-1}, \boldsymbol{\alpha}, \zeta_0)$ que maximizan la función de verosimilitud sujetos a la restricción $\zeta_0 = -\mathbf{BA}'$.

Johansen demostró que, bajo las hipótesis anteriores, los estimadores máximo verosímiles de las relaciones de cointegración pueden encontrarse calculando los autovectores de la matriz de correlaciones de los residuos de las regresiones (vectoriales) de $\Delta \mathbf{y}_t$ y de \mathbf{y}_{t-1} sobre $\Delta \mathbf{y}_{t-1}, \Delta \mathbf{y}_{t-2}, \dots, \Delta \mathbf{y}_{t-p+1}$.

El método proporciona el valor de la función de verosimilitud de la muestra y estimadores máximo verosímiles de los parámetros de una formulación VEC.

El algoritmo procede en tres etapas: cálculo de las regresiones auxiliares, cálculo de las correlaciones canónicas de los residuos y cálculo de los estimadores máximo verosímiles.

Método de Johansen con términos constantes en las relaciones de cointegración

Las relaciones de cointegración son combinaciones lineales de series no estacionarias que resultan ser estacionarias. En el caso general no hay una razón a priori por la que las relaciones de cointegración deban tener esperanza cero, por lo que debemos considerar

Cointegración

la presencia de términos constantes en las relaciones de cointegración, asociados a las esperanzas de las relaciones de cointegración.

En este apartado se detalla el algoritmo de Johansen para la estimación máximo verosímil de sistemas cointegrados con términos constantes en las relaciones de cointegración y en ausencia de tendencias deterministas (Hamilton 1994, p. 643). Este algoritmo será utilizado posteriormente en los experimentos prácticos realizados.

Sea \mathbf{y}_t un vector de dimensión n que sigue un VAR(p) en niveles, correspondiente a un sistema cointegrado con h relaciones de cointegración. Consideremos su representación en forma de “corrección del error”:

$$\Delta \mathbf{y}_t = \zeta_1 \Delta \mathbf{y}_{t-1} + \zeta_2 \Delta \mathbf{y}_{t-2} + \dots + \zeta_{p-1} \Delta \mathbf{y}_{t-p+1} + \boldsymbol{\alpha} + \zeta_0 \mathbf{y}_{t-1} + \mathbf{e}_t$$

donde \mathbf{e}_t es ruido blanco gaussiano $(\mathbf{0}, \boldsymbol{\Omega})$.

La cointegración implica una restricción sobre ζ_0 , que debe poder expresarse como

$$\zeta_0 = -\mathbf{B}\mathbf{A}'$$

donde \mathbf{B} es una matriz $(n \times h)$ y \mathbf{A}' es una matriz $(h \times n)$.

La ausencia de tendencias deterministas implica la restricción (Hamilton 1994, p. 581)

$$\boldsymbol{\alpha} = \mathbf{B} \boldsymbol{\mu}$$

siendo $\boldsymbol{\mu} = E(\mathbf{A}'\mathbf{y}_t)$.

El método de Johansen en estas condiciones procede de la siguiente forma:

Paso 1: Cálculo de tres regresiones auxiliares:

- 1) Una regresión de $\Delta \mathbf{y}_t$ sobre $(\Delta \mathbf{y}_{t-1}, \Delta \mathbf{y}_{t-2}, \dots, \Delta \mathbf{y}_{t-p+1})$ sin término constante,

$$\Delta \mathbf{y}_t = \boldsymbol{\Pi}_1 \Delta \mathbf{y}_{t-1} + \boldsymbol{\Pi}_2 \Delta \mathbf{y}_{t-2} + \dots + \boldsymbol{\Pi}_{p-1} \Delta \mathbf{y}_{t-p+1} + \mathbf{u}_t$$

- 2) Una regresión de un término constante sobre $(\Delta \mathbf{y}_{t-1}, \Delta \mathbf{y}_{t-2}, \dots, \Delta \mathbf{y}_{t-p+1})$,

$$1 = \omega'_1 \Delta y_{t-1} + \omega'_2 \Delta y_{t-2} + \dots + \omega'_{p-1} \Delta y_{t-p+1} + w_t$$

3) Una regresión de y_{t-1} sobre $(\Delta y_{t-1}, \Delta y_{t-2}, \dots, \Delta y_{t-p+1})$ sin término constante,

$$y_{t-1} = \aleph_1 \Delta y_{t-1} + \aleph_2 \Delta y_{t-2} + \dots + \aleph_{p-1} \Delta y_{t-p+1} + v_t$$

Paso 2: Cálculo de correlaciones canónicas

Agrupar los residuos w_t y v_t en el vector $\mathbf{w}_t = [w_t, v_t]'$ y calcular las matrices de covarianzas muestrales

$$\Sigma_{ww} = (1/T) \sum_{t=1}^T \mathbf{w}_t \mathbf{w}'_t$$

$$\Sigma_{uu} = (1/T) \sum_{t=1}^T \mathbf{u}_t \mathbf{u}'_t$$

$$\Sigma_{uw} = (1/T) \sum_{t=1}^T \mathbf{u}_t \mathbf{w}'_t$$

Formar la matriz $\Sigma_{ww}^{-1} \Sigma_{wu} \Sigma_{uu}^{-1} \Sigma_{uw}$, de dimensiones $(n+1) \times (n+1)$, y encontrar sus autovalores: $\lambda_1 > \lambda_2 > \dots > \lambda_{n+1}$. El máximo valor obtenido por el logaritmo de la función de verosimilitud de la muestra, sujeto a la restricción de que existen h relaciones de cointegración y no existen tendencias deterministas, es

$$L_h = -(Tn/2) \log(2\pi) - (Tn/2) - (T/2) \log|\Sigma_{uu}| - (T/2) \sum_{i=1}^h \log(1 - \lambda_i)$$

Paso 3. Cálculo de los estimadores máximo verosímiles de los parámetros

Sean $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_{n+1}$ los autovectores de la matriz $\Sigma_{ww}^{-1} \Sigma_{wu} \Sigma_{uu}^{-1} \Sigma_{uw}$ correspondientes a los autovalores ordenados $\lambda_1 > \lambda_2 > \dots > \lambda_{n+1}$, y normalizados por la condición $\mathbf{a}'_i \Sigma_{ww} \mathbf{a}_i = 1$. Sea $\tilde{\mathbf{A}}$ la matriz $[\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_h]$. Entonces, los estimadores máximo verosímiles de α y ζ_0 son

Cointegración

$$[\hat{\alpha} \quad \hat{\zeta}_0] = \Sigma_{uw} \tilde{\mathbf{A}} \tilde{\mathbf{A}}'$$

El estimador máximo verosímil de cada ζ_i es

$$\hat{\zeta}_i = \Pi_i - \hat{\alpha} \omega'_i - \hat{\zeta}_0 \mathbf{x}_i \quad \text{para } i = 1, 2, \dots, p-1,$$

y el estimador máximo verosímil de Ω es

$$\hat{\Omega} = (1/T) \sum_{t=1}^T [(\mathbf{u}_t - \hat{\alpha} \mathbf{w}_t - \hat{\zeta}_0 \mathbf{v}_t) (\mathbf{u}_t - \hat{\alpha} \mathbf{w}_t - \hat{\zeta}_0 \mathbf{v}_t)']$$

3 MÉTODO DE AOKI PARA IDENTIFICACIÓN DE PROCESOS ESTACIONARIOS

3.1 Introducción

Aoki (1983), basándose en avances en el campo de la teoría de sistemas, desarrolla un método para la identificación de procesos estacionarios multivariable en el espacio de estados que, desde su versión inicial, ha sido posteriormente revisado y completado en trabajos como los de Aoki (1990), Aoki y Havenner (1991), Östermark y Aoki (1992) o Vargas (1999). Este último presenta una excelente revisión de los trabajos realizados desde el algoritmo inicial, algunas de las variantes propuestas y una versión revisada y actualizada del algoritmo. En este apartado expondremos las características principales de este método, al que nos referimos como método de Aoki o algoritmo SSATS¹.

Recordemos que el problema de identificación de un sistema estocástico en el espacio de estados es el de encontrar un orden (una dimensión del vector de estados) y unas matrices $\{A, C, K, R\}$ tales que la representación

$$z_{t+1} = A z_t + K e_t \quad \text{Ecuación de transición de estados}$$

$$y_t = C z_t + e_t \quad \text{Ecuación de observaciones}$$

donde e_t es ruido blanco de matriz de covarianzas R , corresponda a una serie dinámica cuyas propiedades estadísticas (hasta segundo orden) coincidan con las de la serie observada.

El método de Aoki se basa en la información contenida en una matriz de bloques (tipo Hankel) formada por las matrices de autocorrelaciones (o de autocovarianzas) muestrales, que capta la correlación existente entre observaciones con distintos retardos (correlación entre las observaciones pasadas y las futuras). El análisis de esta matriz de Hankel permitirá tanto la especificación del modelo (estimación de la dimensión del

¹ Aoki (1997) se refiere al algoritmo con las siglas SSATS: State Space Aoki Time Series. Otros autores se refieren al “método de Aoki de modelización de series temporales en el espacio de estados”.

vector de estados) como la estimación de unas matrices $\{\mathbf{A}, \mathbf{C}, \mathbf{K}, \mathbf{R}\}$ que proporcionen una adecuada representación del sistema en el espacio de estados.

3.2 La matriz de Hankel

En el contexto de predicción lineal de los valores futuros de una serie a partir de los valores pasados de la misma, aparece una matriz de tipo Hankel (es decir, una matriz formada por bloques dispuestos de forma simétrica respecto a la diagonal), cuyos bloques constituyentes son matrices de autocorrelación de la serie observada correspondientes a distintos retardos.

Con carácter general, siendo $\boldsymbol{\beta}$ e \mathbf{y} vectores aleatorios cualesquiera, se cumple que el estimador lineal \mathbf{b} de $\boldsymbol{\beta}$ ($\mathbf{b} = \mathbf{K} \mathbf{y}$, siendo \mathbf{K} una matriz a definir) que minimiza el error cuadrático medio es (Luenberger, 1969):

$$\mathbf{b}_{\text{MECM}} = E(\boldsymbol{\beta} \mathbf{y}') [E(\mathbf{y} \mathbf{y}')]^{-1} \mathbf{y}$$

El estimador de menor error cuadrático medio coincide además (Hamilton 1994, p. 73) con la esperanza condicionada de $\boldsymbol{\beta}$ dado \mathbf{y} : $E(\boldsymbol{\beta}|\mathbf{y})$.

Supongamos que deseamos estimar los valores futuros del vector \mathbf{y}_t de dimensión m (m series observadas) basándonos en sus valores pasados hasta el periodo $t-1$. Si queremos predecir N_f periodos, estamos buscando el vector \mathbf{y}_t^+ de dimensión $(m \times N_f)$, definido como

$$\mathbf{y}_t^+ = [\mathbf{y}_t', \mathbf{y}_{t+1}', \dots, \mathbf{y}_{t+N_f-1}']'$$

para lo cual contamos con el vector de observaciones pasadas, definido como

$$\mathbf{y}_{t-1}^- = [\mathbf{y}_{t-1}', \mathbf{y}_{t-2}', \dots, \mathbf{y}_{t-N_p}']'$$

En estas condiciones, tenemos la siguiente predicción lineal del vector de observaciones futuras (\mathbf{y}_t^+) condicionada a los valores pasados de la serie observada (\mathbf{y}_{t-1}^-):

$$\hat{\mathbf{y}}_t^+ | \mathbf{y}_{t-1}^- = E(\mathbf{y}_t^+ | \mathbf{y}_{t-1}^-) [E(\mathbf{y}_{t-1}^- \mathbf{y}_{t-1}^{-'})]^{-1} \mathbf{y}_{t-1}^- = \mathbf{H} \mathbf{L}^{-1} \mathbf{y}_{t-1}^-$$

Donde \mathbf{H} es la matriz $E(\mathbf{y}_t^+ \mathbf{y}_{t-1}^-)$, de dimensiones $(mN_f \times mN_p)$, que resulta ser una matriz de tipo Hankel y \mathbf{L} es la matriz de covarianzas $E(\mathbf{y}_{t-1}^- \mathbf{y}_{t-1}^-)$, de dimensiones $(mN_p \times mN_p)$. Si denotamos por $\mathbf{\Gamma}_s$ la matriz $(m \times m)$ de autocorrelación con retardo s : $\mathbf{\Gamma}_s = E(\mathbf{y}_{t+s} \mathbf{y}_t')$, entonces

$$\mathbf{H} = \begin{pmatrix} \mathbf{\Gamma}_1 & \mathbf{\Gamma}_2 & \mathbf{\Gamma}_3 & \text{-----} & \mathbf{\Gamma}_{N_p} \\ \mathbf{\Gamma}_2 & \mathbf{\Gamma}_3 & \mathbf{\Gamma}_4 & \text{-----} & \mathbf{\Gamma}_{N_p+1} \\ \mathbf{\Gamma}_3 & \mathbf{\Gamma}_4 & \mathbf{\Gamma}_5 & \text{-----} & \mathbf{\Gamma}_{N_p+2} \\ \vdots & \vdots & \vdots & & \vdots \\ \mathbf{\Gamma}_{N_f} & \mathbf{\Gamma}_{N_f+1} & \mathbf{\Gamma}_{N_f+2} & \text{-----} & \mathbf{\Gamma}_{N_f+N_p-1} \end{pmatrix}$$

3.3 Especificación del modelo: selección del orden. Teorema de Kronecker

El teorema de Kronecker (Gantmacher, 1960) establece que, para valores suficientemente altos de N_p y N_f , el rango de la matriz \mathbf{H} es el número de estados requeridos para condensar la información histórica del sistema que se propaga hacia el futuro.

En la práctica, deben elegirse unos valores determinados finitos para N_p y N_f . La elección del valor N_p supone un equilibrio entre la adopción de modelos de mayor generalidad y el modelado de relaciones de autocorrelación para excesivos retardos (lo que supondrá mayor error muestral en los estimadores de los parámetros). Si el valor de N_f excede el número desconocido de estados n , entonces el rango de la matriz \mathbf{H} no está restringido por el número de filas (Aoki y Havenner 1991, p.17). Para valores suficientemente altos de N_p y N_f , el rango de la matriz de Hankel no aumenta ya al aumentar N_p o N_f .

Elegidos unos valores para N_p y N_f , pasamos al problema de determinar el rango de la matriz \mathbf{H} . El procedimiento será, en primer lugar, encontrar un estimador $\hat{\mathbf{H}}$ para \mathbf{H} , y, a continuación, calcular el rango de $\hat{\mathbf{H}}$ mediante una descomposición en valores singulares (DVS). El rango de una matriz es el número de valores singulares no nulos.

Estimaremos los bloques $\mathbf{\Gamma}_s$ de \mathbf{H} mediante las matrices de autocovarianza muestral

$$\hat{\Gamma}_s = \frac{1}{T} \sum_{t=1}^{T-s} \mathbf{y}_{t+s} \mathbf{y}_t', \quad s=0,1,2,\dots, N_p + N_f$$

La composición por bloques de estas matrices $\hat{\Gamma}_s$ nos proporciona la estimación $\hat{\mathbf{H}}$ buscada. Ahora, la descomposición en valores singulares proporciona un método computacionalmente robusto de determinación del orden de \mathbf{H} :

$$\mathbf{H} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}'$$

de modo que la matriz $\mathbf{\Sigma}$ de dimensión $(mN_f \times mN_p)$ está formada por ceros salvo en su diagonal, donde contiene los valores singulares ordenados en orden decreciente $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_{mN_f}$ (supuesto $N_f \leq N_p$), y las matrices \mathbf{U} y \mathbf{V} son ortonormales: $\mathbf{U}'\mathbf{U} = \mathbf{I}_{mN_f}$ y $\mathbf{V}'\mathbf{V} = \mathbf{I}_{mN_p}$.

Puesto que $\hat{\mathbf{H}}$ es una estimación muestral (ruidosa) de \mathbf{H} , no se cumplirá que sus valores singulares a partir del correspondiente al número mínimo de estados (n) son exactamente nulos. Para decidir cuáles son significativos debe tomarse un criterio.

Con los valores singulares ordenados de mayor a menor, Aoki y Havenner (1991) eligen un número de estados \hat{n} tal que el cociente de valores singulares $\sigma_{\hat{n}+1}/\sigma_1$ sea de orden $1/\sqrt{T}$. También con este orden decreciente, Havenner (1997) recomienda considerar como último valor singular significativo el primero que, sumado a los anteriores, consigue una suma acumulada igual al 80% de la suma total de los valores singulares¹. Vargas (1999) estudia una serie de contrastes estadísticos que han sido propuestos en la literatura en relación a este problema de determinación del orden (rango) de la matriz \mathbf{H} , basados en aproximaciones a la distribución estadística de los valores singulares de $\hat{\mathbf{H}}$ o en criterios de información.

Una vez seleccionado el orden, el modelo está especificado, y podemos pasar a la etapa de estimación de parámetros del modelo.

¹ Havenner (1997) encuentra que esta regla “ad hoc” resulta sorprendentemente consistente en la práctica con la decisión que se adoptaría mediante un estudio exhaustivo.

Dado que elegir un orden \hat{n} equivale a suponer que los valores singulares de \mathbf{H} a partir del \hat{n} son nulos, modificaremos nuestra estimación de la matriz \mathbf{H} y tomaremos el estimador

$$\hat{\mathbf{H}}_{\hat{n}} = \mathbf{U}_{\hat{n}} \boldsymbol{\Sigma}_{\hat{n}} \mathbf{V}_{\hat{n}}'$$

Donde, partiendo de la descomposición $\hat{\mathbf{H}} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}'$, el subíndice \hat{n} indica que en $\boldsymbol{\Sigma}_{\hat{n}}$ sólo se han considerado los \hat{n} mayores valores singulares de $\boldsymbol{\Sigma}$ y en $\mathbf{U}_{\hat{n}}$ y $\mathbf{V}_{\hat{n}}$ sólo se han considerado los vectores singulares asociados a esos \hat{n} mayores valores singulares. De este modo, nuestro estimador $\hat{\mathbf{H}}_{\hat{n}}$ de \mathbf{H} es una matriz de rango exactamente \hat{n} extraída a partir de la matriz de covarianzas muestral $\hat{\mathbf{H}}$.

3.4 Estimación de los parámetros del modelo

Sea un sistema representable en el espacio de estados en forma de innovaciones mediante una representación mínima de orden n :

$$\mathbf{z}_{t+1} = \mathbf{A} \mathbf{z}_t + \mathbf{K} \mathbf{e}_t \quad \text{Ecuación de transición de estados}$$

$$\mathbf{y}_t = \mathbf{C} \mathbf{z}_t + \mathbf{e}_t \quad \text{Ecuación de observaciones}$$

donde \mathbf{e}_t es ruido blanco de esperanza cero y $E(\mathbf{e}_t \mathbf{e}_t') = \mathbf{R}$. Recordemos que, una vez determinado el orden del sistema, nuestro objetivo es encontrar esas matrices $\{\mathbf{A}, \mathbf{C}, \mathbf{K}, \mathbf{R}\}$ a partir de una secuencia de observaciones \mathbf{y}_t .

Selección de un sistema coordinado para los estados: selección de una matriz de observabilidad

Si el sistema es representable mediante la representación mínima $\{\mathbf{A}, \mathbf{C}, \mathbf{K}, \mathbf{R}\}$, recordemos que cualquier otro conjunto de matrices $\{\mathbf{TAT}^{-1}, \mathbf{CT}^{-1}, \mathbf{TK}, \mathbf{R}\}$, para cualquier matriz invertible \mathbf{T} , constituiría una representación equivalente del mismo sistema (indistinguible observacionalmente), relacionada con la anterior mediante un cambio de base para los estados de la forma $\mathbf{z}^*_t = \mathbf{T} \mathbf{z}_t$. Es decir, que lo que vamos a buscar es una representación mínima $\{\mathbf{A}, \mathbf{C}, \mathbf{K}, \mathbf{R}\}$ determinada dentro de la clase de

representaciones mínimas equivalentes. Elegir una representación determinada dentro de las representaciones mínimas equivalentes corresponde a fijar un sistema coordinado determinado para los estados, y viceversa. Elegiremos en particular una representación que cumpla determinadas condiciones que resultarán de interés.

Para identificar unas matrices $\{\mathbf{A}, \mathbf{C}, \mathbf{K}, \mathbf{R}\}$ apropiadas, trabajaremos de nuevo con la matriz de Hankel \mathbf{H} formada por las matrices de autocorrelación $\mathbf{\Gamma}_s = E(\mathbf{y}_{t+s} \mathbf{y}_t')$ correspondientes a distintos retardos. Si el sistema está representado por $\{\mathbf{A}, \mathbf{C}, \mathbf{K}, \mathbf{R}\}$, entonces se tiene que las matrices de autocorrelación son

$$\mathbf{\Gamma}_s = E(\mathbf{y}_{t+s} \mathbf{y}_t') = \mathbf{C}\mathbf{A}^{s-1}E(\mathbf{z}_{t+1|t} \mathbf{y}_t') = \mathbf{C}\mathbf{A}^{s-1}\mathbf{\Omega}$$

donde se ha definido $\mathbf{\Omega} = E(\mathbf{z}_{t+1|t} \mathbf{y}_t')$. Se tiene entonces la descomposición:

$$\mathbf{H} = \begin{pmatrix} \mathbf{\Gamma}_1 & \mathbf{\Gamma}_2 & \mathbf{\Gamma}_3 & \text{-----} & \mathbf{\Gamma}_{N_p} \\ \mathbf{\Gamma}_2 & \mathbf{\Gamma}_3 & \mathbf{\Gamma}_4 & \text{-----} & \mathbf{\Gamma}_{N_p+1} \\ \mathbf{\Gamma}_3 & \mathbf{\Gamma}_4 & \mathbf{\Gamma}_5 & \text{-----} & \mathbf{\Gamma}_{N_p+2} \\ \vdots & \vdots & \vdots & & \vdots \\ \mathbf{\Gamma}_{N_f} & \mathbf{\Gamma}_{N_f+1} & \mathbf{\Gamma}_{N_f+2} & \text{-----} & \mathbf{\Gamma}_{N_f+N_p-1} \end{pmatrix} = \underbrace{\begin{pmatrix} \mathbf{C} \\ \mathbf{C}\mathbf{A} \\ \mathbf{C}\mathbf{A}^2 \\ \vdots \\ \mathbf{C}\mathbf{A}^{N_f-1} \end{pmatrix}}_{\mathbf{O}} \underbrace{\begin{pmatrix} \mathbf{\Omega} & \mathbf{A}\mathbf{\Omega} & \mathbf{A}^2\mathbf{\Omega} & \text{-----} & \mathbf{A}^{N_p-1}\mathbf{\Omega} \end{pmatrix}}_{\mathbf{\Pi}}$$

De modo que hemos descompuesto \mathbf{H} en el producto de una matriz de observabilidad \mathbf{O} y otra matriz que denominamos $\mathbf{\Pi}$.

Por otro lado, podemos considerar la descomposición en valores singulares $\mathbf{H} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}'$. Vamos a tratar de relacionar ambas descomposiciones.

Nótese que la matriz \mathbf{H} es común (única) para toda representación mínima equivalente $\{\mathbf{A}, \mathbf{C}, \mathbf{K}, \mathbf{R}\}$ del sistema. Es una propiedad del proceso, no de la representación. Sin embargo, las matrices \mathbf{O} y $\mathbf{\Pi}$ dependen de la representación concreta del sistema, de modo que para una representación mínima equivalente $\{\mathbf{T}\mathbf{A}\mathbf{T}^{-1}, \mathbf{C}\mathbf{T}^{-1}, \mathbf{T}\mathbf{K}, \mathbf{R}\}$ correspondiente a un cambio de base en los estados $\mathbf{z}_t^* = \mathbf{T} \mathbf{z}_t$ tendríamos unas nuevas matrices $\mathbf{O}^* = \mathbf{O} \mathbf{T}^{-1}$ y $\mathbf{\Pi}^* = \mathbf{T} \mathbf{\Pi}$, y se seguiría cumpliendo

$$\mathbf{H} = \mathbf{O} \mathbf{\Pi} = \mathbf{O} \mathbf{T}^{-1} \mathbf{T} \mathbf{\Pi} = \mathbf{O}^* \mathbf{\Pi}^*$$

Así pues, y dado que la matriz \mathbf{T}^{-1} es arbitraria, existe una amplia libertad para descomponer \mathbf{H} en forma del producto de dos matrices \mathbf{O} y $\mathbf{\Pi}$ con las dimensiones y rangos apropiados. Tomar una descomposición determinada equivale a seleccionar una determinada representación mínima $\{\mathbf{A}, \mathbf{C}, \mathbf{K}, \mathbf{R}\}$ del sistema.

El procedimiento de Aoki se basa en elegir una descomposición $\mathbf{H} = \mathbf{O} \mathbf{\Pi}$ a partir de la descomposición en valores singulares $\mathbf{H} = \mathbf{U} \mathbf{\Sigma} \mathbf{V}' = (\mathbf{U} \mathbf{\Sigma}^{1/2}) (\mathbf{\Sigma}^{1/2} \mathbf{V}')$, de forma que tendríamos una selección de matrices \mathbf{O} y $\mathbf{\Pi}$ dada por $\mathbf{O} = \mathbf{U} \mathbf{\Sigma}^{1/2}$ y $\mathbf{\Pi} = \mathbf{\Sigma}^{1/2} \mathbf{V}'$. Las matrices $\{\mathbf{A}, \mathbf{C}, \mathbf{K}, \mathbf{R}\}$ quedan determinadas y son recuperables a partir de esta elección de \mathbf{O} y $\mathbf{\Pi}$.

Recuperación de las matrices del modelo

Para recuperar las matrices \mathbf{C} y \mathbf{A} a partir de la estimación de \mathbf{H} y de su descomposición en \mathbf{O} y $\mathbf{\Pi}$, Aoki y Havenner (1991) proponen considerar las siguientes relaciones:

$$\mathbf{C} \mathbf{\Pi} = [\mathbf{\Gamma}_1 \quad \mathbf{\Gamma}_2 \quad \dots \quad \mathbf{\Gamma}_{N_p}]$$

$$\mathbf{O} \mathbf{A} \mathbf{\Pi} = \begin{pmatrix} \mathbf{C} \\ \mathbf{C} \mathbf{A} \\ \mathbf{C} \mathbf{A}^2 \\ \vdots \\ \mathbf{C} \mathbf{A}^{N_f-1} \end{pmatrix} \begin{bmatrix} \mathbf{A} \mathbf{\Omega} & \mathbf{A}^2 \mathbf{\Omega} & \dots & \mathbf{A}^{N_p} \mathbf{\Omega} \end{bmatrix} = \begin{pmatrix} \mathbf{\Gamma}_2 & \mathbf{\Gamma}_3 & \dots & \mathbf{\Gamma}_{N_p+1} \\ \mathbf{\Gamma}_3 & \mathbf{\Gamma}_4 & \dots & \mathbf{\Gamma}_{N_p+2} \\ \mathbf{\Gamma}_4 & \mathbf{\Gamma}_5 & \dots & \mathbf{\Gamma}_{N_p+3} \\ \vdots & \vdots & & \vdots \\ \mathbf{\Gamma}_{N_f+1} & \mathbf{\Gamma}_{N_f+2} & \dots & \mathbf{\Gamma}_{N_f+N_p} \end{pmatrix}$$

La matriz de la derecha de esta última expresión es $E(\mathbf{y}_{t+1}^+ \mathbf{y}_{t-1}^-)$, que denotamos por \mathbf{H}_{+1} .

En la práctica, dispondremos de las autocovarianzas muestrales $\hat{\mathbf{\Gamma}}_i$ para formar los estimadores $\hat{\mathbf{H}}_{\hat{n}}$ (de \mathbf{H}) y $\hat{\mathbf{H}}_{+1}$ (de \mathbf{H}_{+1}). Descompondremos $\hat{\mathbf{H}}_{\hat{n}}$ en el producto de dos matrices en la forma

$$\hat{\mathbf{H}}_{\hat{n}} = \mathbf{U}_{\hat{n}} \mathbf{\Sigma}_{\hat{n}} \mathbf{V}_{\hat{n}}' = (\mathbf{U}_{\hat{n}} \mathbf{\Sigma}_{\hat{n}}^{1/2}) (\mathbf{\Sigma}_{\hat{n}}^{1/2} \mathbf{V}_{\hat{n}}') = \hat{\mathbf{O}} \hat{\mathbf{\Pi}}$$

Consideraremos como inversas generalizadas

$$\hat{\mathbf{O}}^{-1} = \boldsymbol{\Sigma}^{-1/2}_{\hat{n}} \mathbf{U}'_{\hat{n}}$$

$$\hat{\Pi}^{-1} = \mathbf{V}_{\hat{n}} \boldsymbol{\Sigma}^{-1/2}_{\hat{n}}$$

y tomaremos los estimadores de \mathbf{C} y \mathbf{A} :

$$\hat{\mathbf{C}} = [\hat{\Gamma}_1 \ \hat{\Gamma}_2 \ \dots \ \hat{\Gamma}_{Np}] \hat{\Pi}^{-1} = [\hat{\Gamma}_1 \ \hat{\Gamma}_2 \ \dots \ \hat{\Gamma}_{Np}] \mathbf{V}_{\hat{n}} \boldsymbol{\Sigma}^{-1/2}_{\hat{n}}$$

$$\hat{\mathbf{A}} = \hat{\mathbf{O}}^{-1} \hat{\mathbf{H}}_{+1} \hat{\Pi}^{-1} = \boldsymbol{\Sigma}^{-1/2}_{\hat{n}} \mathbf{U}'_{\hat{n}} \hat{\mathbf{H}}_{+1} \mathbf{V}_{\hat{n}} \boldsymbol{\Sigma}^{-1/2}_{\hat{n}}$$

Sólo nos queda entonces estimar la matriz \mathbf{K} . Para ello, el algoritmo original estima en primer lugar la matriz $\boldsymbol{\Omega}$ y requiere a continuación solucionar una ecuación tipo Riccati (Aoki y Havenner, 1991). Vargas (1999) revisa los principales problemas y limitaciones que puede presentar este algoritmo, así como algunas de las variantes propuestas en la literatura para solucionarlos.

Estimación de \mathbf{K} basada en la secuencia de estados

Un enfoque alternativo para estimar la matriz \mathbf{K} que no requiere solucionar la ecuación de Riccati se encuentra en los trabajos de Östermark y Aoki (1992), Östermark (1997) y Vargas (1999). Este enfoque consiste en estimar primero la secuencia de los estados y a partir de ésta utilizar las ecuaciones del modelo

$$\mathbf{z}_{t+1} = \mathbf{A} \mathbf{z}_t + \mathbf{K} \mathbf{e}_t \quad \text{Ecuación de transición de estados}$$

$$\mathbf{y}_t = \mathbf{C} \mathbf{z}_t + \mathbf{e}_t \quad \text{Ecuación de observaciones}$$

para estimar la matriz \mathbf{K} . Nótese que si conociéramos la secuencia de los estados y las matrices \mathbf{A} y \mathbf{C} , podríamos usar las ecuaciones

$$\mathbf{e}_t = \mathbf{y}_t - \mathbf{C} \mathbf{z}_t$$

$$\mathbf{K} \mathbf{e}_t = \mathbf{z}_{t+1} - \mathbf{A} \mathbf{z}_t$$

para estimar los primeros términos de cada ecuación (innovaciones \mathbf{e}_t y términos $\mathbf{K} \mathbf{e}_t$), y a continuación usar la ecuación

$$E[(\mathbf{K} \mathbf{e}_t) \mathbf{e}_t'] = \mathbf{K} E[\mathbf{e}_t \mathbf{e}_t']$$

para estimar la matriz \mathbf{K} ¹.

Sólo falta saber cómo estimar la secuencia de estados. Los estados en una representación mínima son estadísticos mínimo suficientes para condensar la historia pasada del sistema que resulta relevante para su evolución futura. Si consideramos las ecuaciones del modelo en el espacio de estados y sustituimos recursivamente para expresar \mathbf{y}_t^+ en función de \mathbf{z}_t , tenemos que

$$E(\mathbf{y}_t^+ | \mathbf{y}_{t-1}^-) = \mathbf{O} \mathbf{z}_t$$

Por otro lado, realizando la proyección lineal de \mathbf{y}_t^+ sobre \mathbf{y}_{t-1}^- , tenemos que

$$E(\mathbf{y}_t^+ | \mathbf{y}_{t-1}^-) = E(\mathbf{y}_t^+ \mathbf{y}_{t-1}^{-'}) [E(\mathbf{y}_{t-1}^- \mathbf{y}_{t-1}^{-'})]^{-1} \mathbf{y}_{t-1}^- = \mathbf{H} \mathbf{L}^{-1} \mathbf{y}_{t-1}^- = \mathbf{O} \mathbf{\Pi} \mathbf{L}^{-1} \mathbf{y}_{t-1}^-$$

Igualando estas dos expresiones encontramos que podemos tomar como vector de estados:

$$\mathbf{z}_t = \mathbf{\Pi} \mathbf{L}^{-1} \mathbf{y}_{t-1}^-$$

Podemos considerar entonces un estimador $\mathbf{z}_{t|t-1}^e$ del valor de \mathbf{z}_t condicionado a los valores anteriores, definido por (trabajando ya con las matrices estimadas):

$$\mathbf{z}_{t|t-1}^e = \hat{\mathbf{O}}^\dagger \hat{\mathbf{H}}_n \hat{\mathbf{L}}^{-1} \mathbf{y}_{t-1}^-$$

siendo $\hat{\mathbf{O}}^\dagger$ una inversa generalizada de $\hat{\mathbf{O}}$ (el resultado sería el mismo para cualquier $\hat{\mathbf{O}}^\dagger$).

A partir del conocimiento de la secuencia de estados, junto con las matrices \mathbf{A} y \mathbf{C} ¹, podemos estimar, como hemos indicado, la secuencia de innovaciones (\mathbf{e}_t) y la matriz \mathbf{K} .

Vargas (1999) discute también modificaciones del método de Aoki propuestas por otros autores que han tratado de facilitar la labor de estimación.

¹ De hecho, también las matrices \mathbf{A} y \mathbf{C} podrían ser estimadas a partir de la secuencia de estados \mathbf{z}_t , aplicando mínimos cuadrados. Desarrollaremos este punto al estudiar los métodos de subespacios.

3.5 Propiedades del método de Aoki

Los procedimientos de análisis de series temporales parten de formulaciones generales, como puede ser una formulación VARMA, e imponen condiciones sobre los parámetros para que el modelo sea más fácilmente tratable y para evitar un excesivo error muestral en los estimadores de los parámetros, lo que conduciría posiblemente a malas predicciones. Las restricciones impuestas sobre los modelos son a veces arbitrarias (incluso aunque esas restricciones no supongan pérdida de generalidad, pueden dar lugar a especificaciones más complicadas. Por ejemplo, limitar el modelo a términos autorregresivos). El método de Aoki impone las restricciones sobre los coeficientes del modelo basándose en la matriz Hankel de autocovarianzas muestrales. Sustituir esta matriz de Hankel por otra en la que se han despreciado algunos de los valores singulares equivale a imponer unas condiciones sobre los parámetros del modelo.

La selección de sistema coordinado o base de representación para los estados tiene también importancia para a las propiedades del algoritmo. Recordemos que el sistema coordinado se fija al seleccionar como matriz de observabilidad del sistema la matriz $\mathbf{O} = \mathbf{U} \boldsymbol{\Sigma}^{1/2}$, correspondiente a una descomposición de la matriz de Hankel de la forma $\mathbf{H} = \mathbf{U} \boldsymbol{\Sigma} \mathbf{V}' = (\mathbf{U} \boldsymbol{\Sigma}^{1/2}) (\boldsymbol{\Sigma}^{1/2} \mathbf{V}') = \mathbf{O} \boldsymbol{\Pi}$. Esta elección del sistema coordinado da lugar a una representación del sistema que en teoría de sistemas se conoce como “representación equilibrada”, ya que cumple que $\mathbf{O}' \mathbf{O} = \boldsymbol{\Pi} \boldsymbol{\Pi}' = \boldsymbol{\Sigma}$. Aoki y Havenner (1991) indican varias de las propiedades del algoritmo asociadas a la elección de este sistema coordinado equilibrado. En particular, indicamos:

i) Inclusión estricta y consistencia frente a especificación errónea.

Los valores estimados para los parámetros en un modelo con n estados son un subconjunto de los valores estimados en un modelo con $n+1$ estados. Así por ejemplo, la esquina noroeste $n \times n$ de la matriz $\hat{\mathbf{A}}$ es la misma en el modelo con n y con $n+1$ estados. Aoki y Havenner (1991) denominan a esta propiedad “strict nesting”, que aquí traducimos como “inclusión estricta”. Así pues, al estimar con un modelo de orden n , los valores estimados de los parámetros para modelos de orden inferior son conocidos sin necesidad de recalcularlos. Además, puesto que los estados van entrando en el modelo por orden decreciente de los valores singulares asociados, los efectos más

importantes son captados primero, sin que la estimación de sus parámetros asociados cambie al incluir nuevos estados.

ii) Estabilidad de los submodelos

Hemos visto que, partiendo de una representación equilibrada en el espacio de estados, cualquier aproximación de menor orden puede obtenerse sin necesidad de reestimar el modelo, sin más que eliminar los elementos que corresponda de las matrices del sistema. Se tiene además que, si el modelo original es estable, también lo es cualquiera de los modelos así estimados de menor orden.

iii) Especificación eficiente del modelo

Aoki y Havenner (1991) inciden en que el principal determinante del éxito del modelado de series temporales es realizar una apropiada especificación del modelo, pese a que la mayoría de econométricos y estadísticos se viene centrando históricamente en la cuestión de realizar una estimación eficiente *dado el modelo*.

Por otra parte, los modelos de series temporales tratan de aproximar, explícita o implícitamente, la secuencia de autocovarianzas muestral de los datos.

Siendo $\hat{\mathbf{H}}$ la matrix de Hankel formada por las matrices de autocovarianza muestrales, el método de Aoki encuentra una representación cuya matriz de autocovarianzas es $\hat{\mathbf{H}}_n$, donde $\hat{\mathbf{H}}_n$ se construye a partir de $\hat{\mathbf{H}}$ basándose en la descomposición en valores singulares en orden decreciente, de forma que el primer valor singular excluido es σ_{n+1} . La norma de Hankel (Aoki 1990, p. 68) de esta aproximación a la secuencia de autocovarianzas es el primer valor singular excluido σ_{n+1} . El método de Aoki proporciona una representación en el espacio de estados tal que ninguna otra representación de igual o menor orden puede aproximar mejor (en el sentido de norma de Hankel) la secuencia de autocovarianzas (o autocorrelaciones) observada.

En cuanto a las posibles dificultades o inconvenientes que pueden surgir al aplicar el método, indicamos en particular el del cumplimiento de la condición de secuencias de covarianzas reales positivas. El método de Aoki en su versión de 1990 estima las matrices $\{\mathbf{A}, \mathbf{C}, \mathbf{\Omega}, \mathbf{\Gamma}_0\}$ como paso previo antes de llegar a las matrices buscadas $\{\mathbf{A}, \mathbf{C}, \mathbf{K}\}$. El problema es que las matrices $\{\mathbf{A}, \mathbf{C}, \mathbf{\Omega}, \mathbf{\Gamma}_0\}$ deben cumplir una determinada condición para poder corresponder a un proceso estocástico, y esta condición no viene

garantizada por el algoritmo de Aoki. Existen combinaciones de matrices $\{\mathbf{A}, \mathbf{C}, \mathbf{\Omega}, \mathbf{\Gamma}_0\}$ que generarían una secuencia de autocovarianzas $\mathbf{\Gamma}_i$ que no es una secuencia real positiva, por lo que no pueden corresponder a un proceso estocástico real. Havenner (1997) indica que, cuando el procedimiento de Aoki estima una de estas combinaciones, existen dos opciones: o cambiar los estimadores, o cambiar la especificación del modelo. Aunque se han desarrollado estimadores alternativos para evitar este problema, como los de Vaccaro y Vukina (1993), Havenner recomienda probar con otra especificación (otro orden) del modelo.

Si, en vez de utilizar las matrices $\{\mathbf{A}, \mathbf{C}, \mathbf{\Omega}, \mathbf{\Gamma}_0\}$ y la ecuación de Riccati, se utiliza la estimación de \mathbf{K} basada en la secuencia de estados, entonces no se presenta el problema de la falta de positividad de la secuencia de covarianzas.

4 MÉTODOS DE SUBESPACIOS PARA IDENTIFICACIÓN DE PROCESOS ESTACIONARIOS

4.1 Introducción

Los métodos de subespacios son una familia de algoritmos de identificación de sistemas en el espacio de estados basados en técnicas de proyección lineal. Una buena presentación de estos métodos puede encontrarse en De Cock y De Moor (2003) y en Viberg (1994). Ljung (1999, cap. 7 y cap. 10) ofrece también un tratamiento general, y mayor detalle puede encontrarse en Van Overschee y De Moor (1996).

Si bien los orígenes de estos algoritmos se encuentran en la teoría de realización clásica – suele citarse el trabajo de Ho y Kalman (1966) como uno de los primeros antecedentes- , es principalmente desde finales de los años 80 cuando los trabajos y el interés por estos métodos se multiplican. El profesor Ljung indicaba en 1993 que “El desarrollo de los métodos de subespacios es el aspecto más apasionante que ha sucedido en el campo de la identificación de sistemas en los últimos cinco años...”¹

Los algoritmos de subespacios se han desarrollado para la identificación de sistemas deterministas, de sistemas estocásticos, y de sistemas mixtos deterministas-estocásticos. Bauer y Ljung (2002) distinguen también entre dos tipos de algoritmos de subespacios: los de tipo Larimore (o tipo CCA, por el algoritmo conocido como “análisis de correlaciones canónicas, de Larimore 1983) y los de tipo MOESP². En esta tesis nos centramos en sistemas estocásticos y algoritmos tipo Larimore, en particular, en el algoritmo CCA, que se considera especialmente apropiado para la identificación de sistemas estocásticos (Bauer y Wagner 2002).

La principal característica de los algoritmos de subespacios tipo Larimore es que una secuencia de estados es calculada (implícita o explícitamente) a partir de los datos entrada-salida del sistema, sin necesidad de conocer las matrices del modelo. Esta

¹ La cita aparece en Van Overschee y De Moor (1996). Traducción propia.

² Los algoritmos de tipo MOESP se centran en las características de una matriz de observabilidad estimada para, a partir de ésta, obtener estimadores de las matrices del sistema.

Análisis de Sistemas Cointegrados mediante Métodos de Subespacios

secuencia de estados es utilizada posteriormente para estimar las matrices del modelo. Nótese que, frente a este enfoque, la aproximación clásica a la estimación de los estados es la utilización del filtro de Kalman una vez conocidas las matrices del modelo.

Si se conoce la secuencia de estados, la estimación de las matrices del sistema $\{\mathbf{A}, \mathbf{K}, \mathbf{C}, \mathbf{R}\}$ puede realizarse por mínimos cuadrados. Para ver esto, consideremos el modelo general

$$\mathbf{z}_{t+1} = \mathbf{A} \mathbf{z}_t + \mathbf{K} \mathbf{e}_t \quad \text{Ecuación de transición de estados}$$

$$\mathbf{y}_t = \mathbf{C} \mathbf{z}_t + \mathbf{e}_t \quad \text{Ecuación de observaciones}$$

donde \mathbf{e}_t es un vector ruido blanco de esperanza cero y matriz de covarianzas \mathbf{R} . Si conocemos la secuencia de estados, tenemos las ecuaciones

$$\begin{pmatrix} \mathbf{z}_{t+1} \\ \mathbf{y}_t \end{pmatrix} = \begin{pmatrix} \mathbf{A} \\ \mathbf{C} \end{pmatrix} \mathbf{z}_t + \begin{pmatrix} \mathbf{K} \mathbf{e}_t \\ \mathbf{e}_t \end{pmatrix}$$

conocido

que permiten estimar las matrices del modelo por mínimos cuadrados.

Como sucedía en el método de Aoki, los algoritmos de subespacios parten de una matriz de proyección lineal de las observaciones “futuras” en las observaciones “pasadas”. Tomando

$$\mathbf{y}_t^+ = [\mathbf{y}_t', \mathbf{y}_{t+1}', \dots, \mathbf{y}_{t+f-1}']'$$

$$\mathbf{y}_{t-1}^- = [\mathbf{y}_{t-1}', \mathbf{y}_{t-2}', \dots, \mathbf{y}_{t-p}']'$$

la matriz de proyección ortogonal de las observaciones futuras (\mathbf{y}_t^+) en las pasadas (\mathbf{y}_{t-1}^-) es

$$E(\mathbf{y}_t^+ \mathbf{y}_{t-1}^{-'}) [E(\mathbf{y}_{t-1}^- \mathbf{y}_{t-1}^{-'})]^{-1}$$

Y la proyección ortogonal de \mathbf{y}_t^+ sobre \mathbf{y}_{t-1}^- es

$$E(y_t^+ | y_{t-1}^-) [E(y_{t-1}^- | y_{t-1}^-)]^{-1} y_{t-1}^-$$

Básicamente, los métodos de subespacios encuentran una combinación [matriz de observabilidad + secuencia de estados] que proporcione las mismas predicciones que proporciona una estimación mediante proyección lineal.

Una de las características más favorables de los métodos de subespacios es que las operaciones requeridas pueden implementarse mediante algoritmos de álgebra lineal numérica robustos y de propiedades bien conocidas, como son la descomposición en valores singulares y la descomposición matricial **QR** (donde **R** es una matriz triangular superior y **Q** es una matriz ortogonal), lo que permite expresar de forma compacta los estimadores finales en función de las matrices de datos observados -ver, p.e., las fórmulas utilizadas en De Cock y De Moor (2003)-.

Los algoritmos de subespacios no son iterativos, lo que favorece su rapidez y evita problemas de convergencia (Van Oberschee y De Moor 1996, pp. 11-12). La rápida estimación de los parámetros del sistema que proporcionan puede usarse como alternativa a otros procedimientos de estimación recursivos (Kalman-máxima-verosimilitud, por ejemplo), más lentos y problemáticos, o para proporcionar valores iniciales adecuados para los mismos.

Identificación por métodos de subespacios frente al enfoque Kalman-máxima verosimilitud.

Posiblemente, la mayor aportación de los métodos de subespacios consiste en haber establecido y combinado las siguientes dos ideas:

- 1 Que es posible estimar una secuencia de estados (tipo Kalman) a partir de observaciones del sistema y en un solo paso, sin necesidad de conocer las matrices del sistema (sin conocer ni aplicar el filtro de Kalman).
- 2 Que, una vez conocida una secuencia de estados, todas las matrices del sistema son estimables por mínimos cuadrados.

La diferencia en el enfoque de estimación de parámetros entre los algoritmos de subespacios tipo Larimore y el método filtro de Kalman + máxima verosimilitud puede observarse en la figura 4.1.

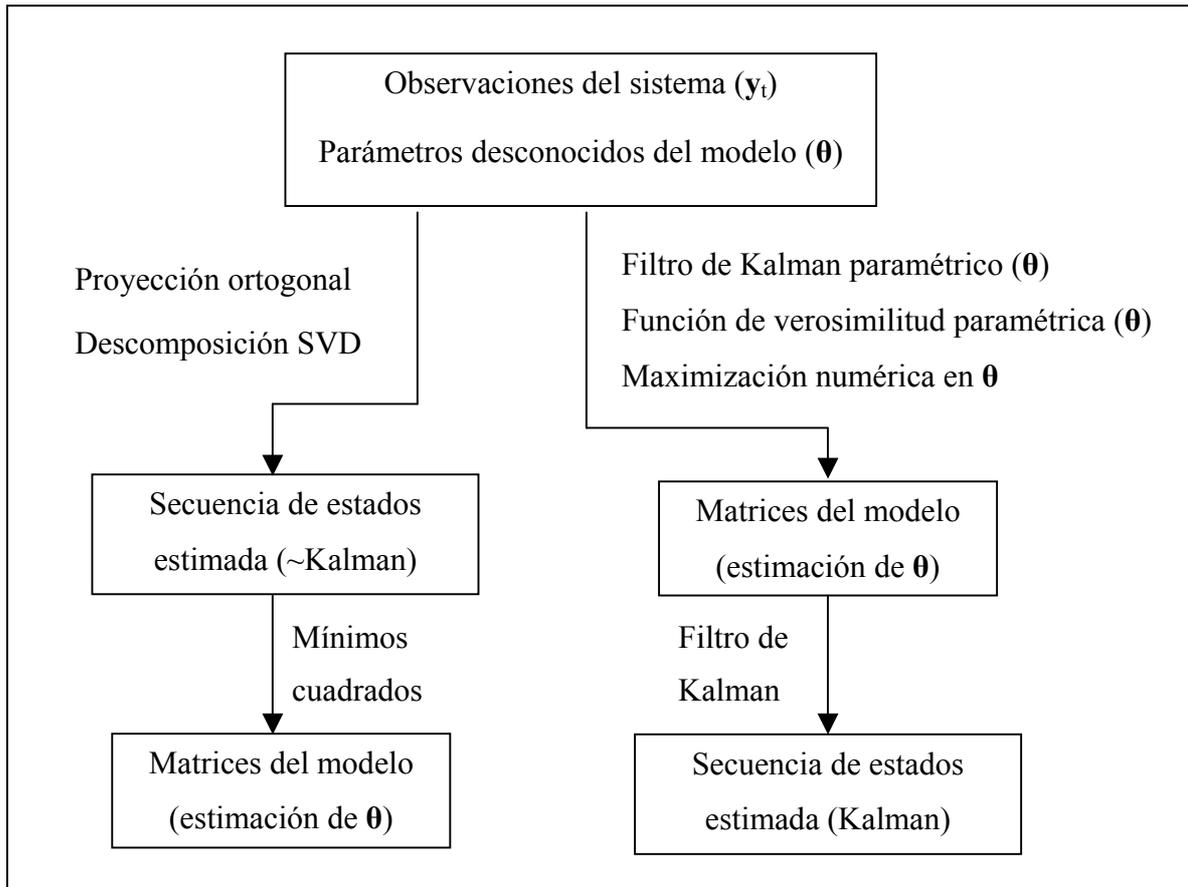


Figura 4-1. Diferencia entre la estimación por métodos de subespacios y la estimación basada en el filtro de Kalman + máxima verosimilitud. La parte izquierda resume el enfoque de los métodos de subespacios, en los que una secuencia de estados (tipo Kalman) es calculada en primer lugar y a partir de observaciones del sistema, sin necesidad de conocer o estimar previamente las matrices del modelo.

En comparación al enfoque Kalman-máxima-verosimilitud destacamos las principales diferencias en la aplicación de los métodos de subespacios:

- 1.- La estimación no requiere una optimización iterativa de una función no lineal. Se emplean, en cambio, métodos robustos de álgebra numérica lineal, como la descomposición en valores singulares y la descomposición **QR**.
- 2.-Frente al enfoque Kalman-máxima-verosimilitud, la consideración o imposición de restricciones a priori sobre algunos de los parámetros del modelo no resulta sencilla de considerar bajo los métodos de subespacios, que parten de un enfoque “caja negra”.

3.- La fase de especificación de los algoritmos se encarga de buscar el orden de una representación mínima. La identificabilidad se consigue fijando un sistema coordinado para los estados mediante la elección de una matriz invertible arbitraria (cada tipo de algoritmo de subespacios elige esta matriz de una forma determinada)¹.

4.2 Planteamiento, notación e hipótesis básicas

El planteamiento del problema que buscamos resolver es el correspondiente al problema de identificación estocástica en el espacio de estados, tal como se enunció en el apartado 1.5.

Suponemos que el proceso a identificar es estacionario y ergódico, y que el modelo es controlable y observable (la representación que buscamos es una representación mínima).

A partir de la secuencia de observaciones $\mathbf{y}_0, \dots, \mathbf{y}_T$ definimos la matriz de tipo Hankel $\mathbf{Y}_{0|2i-1}$ y las (sub)matrices $\mathbf{Y}_p, \mathbf{Y}_f, \mathbf{Y}_{p+}, \mathbf{Y}_{f-}$ como se indica a continuación:

$$\mathbf{Y}_{0|2i-1} = \begin{array}{c} \begin{array}{c} \uparrow \\ \textit{i} \\ \downarrow \end{array} \left(\begin{array}{cccc} \mathbf{y}_0 & \mathbf{y}_1 & \dots & \mathbf{y}_{j-1} \\ \dots & \dots & \dots & \dots \\ \mathbf{y}_{i-2} & \mathbf{y}_{i-1} & \dots & \mathbf{y}_{i+j-3} \\ \mathbf{y}_{i-1} & \mathbf{y}_i & \dots & \mathbf{y}_{i+j-2} \\ \hline \mathbf{y}_i & \mathbf{y}_{i+1} & \dots & \mathbf{y}_{i+j-1} \\ \mathbf{y}_{i+1} & \mathbf{y}_{i+2} & \dots & \mathbf{y}_{i+j} \\ \dots & \dots & \dots & \dots \\ \mathbf{y}_{2i-1} & \mathbf{y}_{2i} & \dots & \mathbf{y}_{2i+j-2} \end{array} \right) \begin{array}{c} \uparrow \\ \textit{"pasado"} \\ \downarrow \\ \uparrow \\ \textit{"futuro"} \\ \downarrow \end{array} \end{array} = \frac{\mathbf{Y}_p}{\mathbf{Y}_f} =$$

¹ Esto es en cierto sentido equivalente a la utilización de parametrizaciones canónicas del enfoque Kalman-máxima-verosimilitud para conseguir un modelo identificable.

$$\begin{array}{c}
 \begin{array}{c} \longleftarrow j \longrightarrow \\ \uparrow i+1 \\ \downarrow i-1 \end{array} \left(\begin{array}{cccc} \mathbf{y}_0 & \mathbf{y}_1 & \dots & \mathbf{y}_{j-1} \\ \dots & \dots & \dots & \dots \\ \mathbf{y}_{i-2} & \mathbf{y}_{i-1} & \dots & \mathbf{y}_{i+j-3} \\ \mathbf{y}_{i-1} & \mathbf{y}_i & \dots & \mathbf{y}_{i+j-2} \\ \mathbf{y}_i & \mathbf{y}_{i+1} & \dots & \mathbf{y}_{i+j-1} \\ \hline \mathbf{y}_{i+1} & \mathbf{y}_{i+2} & \dots & \mathbf{y}_{i+j} \\ \dots & \dots & \dots & \dots \\ \mathbf{y}_{2i-1} & \mathbf{y}_{2i} & \dots & \mathbf{y}_{2i+j-2} \end{array} \right) \begin{array}{c} \uparrow \text{“pasado”} \\ \downarrow \text{“futuro”} \end{array} = \frac{\mathbf{Y}_{p+}}{\mathbf{Y}_f}
 \end{array}$$

El valor de i es un índice fijado por el usuario que debe ser superior al orden n del sistema. Puesto que ese orden es desconocido a priori, se tomará un i que sea al menos mayor que el máximo orden que pretendamos identificar. Se asumirá en adelante $i > n$.

El valor de j suele elegirse de forma que se aprovechen las $T+1$ observaciones disponibles, lo que supone tomar $2i + j - 2 = T$, luego $j = T - 2i + 2$.

Definimos una secuencia de estados \mathbf{Z}_i como la matriz de dimensiones $n \times j$

$$\mathbf{Z}_i = [\mathbf{z}_i \quad \mathbf{z}_{i+1} \quad \dots \quad \mathbf{z}_{i+j-2} \quad \mathbf{z}_{i+j-1}]$$

Definimos una¹ matriz de observabilidad extendida \mathbf{O}_i ($i > n$) del sistema como una matriz de la forma

$$\mathbf{O}_i = \begin{pmatrix} \mathbf{C} \\ \mathbf{CA} \\ \mathbf{CA}^2 \\ \vdots \\ \mathbf{CA}^{i-1} \end{pmatrix}$$

¹ Si nos referimos al sistema, y fijado el índice i , deberíamos hablar de “una” matriz de observabilidad extendida porque cada matriz de observabilidad está asociada a una representación particular del modelo, que, como sabemos, no es única. Dos matrices de observabilidad extendida \mathbf{O}_i y \mathbf{O}_i^* correspondientes a sendas representaciones mínimas de un modelo de orden n están relacionadas mediante la ecuación $\mathbf{O}_i^* = \mathbf{O}_i \mathbf{T}^{-1}$, para alguna matriz de cambio de base \mathbf{T}^{-1} de dimensiones $n \times n$.

donde \mathbf{A} y \mathbf{C} corresponden a una representación mínima del sistema.

4.3 Cálculo del orden y de una secuencia de estados

Los teoremas principales de la identificación por métodos de subespacios se preporcionan en el Anexo IV. En este apartado realizamos una interpretación intuitiva de estos métodos.

La proyección de las observaciones futuras sobre las pasadas $\mathbf{Y}_f / \mathbf{Y}_p$ es estimable de forma consistente para procesos estacionarios mediante el teorema de la proyección lineal:

$$\mathbf{Y}_f / \mathbf{Y}_p \equiv \Phi_i \approx \mathbf{Y}_f \mathbf{Y}'_p (\mathbf{Y}_p \mathbf{Y}'_p)^{\dagger} \mathbf{Y}_p$$

Por otra parte, para modelos en el espacio de estados, se tiene que

- El rango de la matriz Φ_i es el orden del sistema.
- El estado \mathbf{z}_t condensa toda la información histórica de la serie relevante para el futuro. Una estimación de \mathbf{y}_{t+k} ($k \geq 0$) basada en la información hasta el momento $t-1$ sería $\hat{\mathbf{y}}_{t+k|t-1} = \mathbf{C}\mathbf{A}^k \mathbf{z}_{|t-1}$, lo que, matricialmente, conduce a la estimación $\hat{\mathbf{Y}}_f = \mathbf{O}_i \mathbf{Z}_i$

Considerando la relación $\hat{\mathbf{Y}}_f = \mathbf{O}_i \mathbf{Z}_i \approx \mathbf{Y}_f / \mathbf{Y}_p \equiv \Phi_i$, los algoritmos de subespacios estiman la matriz Φ_i y llevan a cabo una descomposición de la misma en valores singulares para:

- Calcular el orden del sistema (teóricamente, el orden es el número de autovalores de la matriz Φ_i distintos de cero).
- Descomponer la matriz Φ_i estimada en el producto de una matriz de observabilidad extendida \mathbf{O}_i con una secuencia de estados \mathbf{Z}_i .

Para calcular la descomposición $\mathbf{O}_i \mathbf{Z}_i$ existe una amplia libertad. Nótese que el producto $\mathbf{O}_i \mathbf{Z}_i$ es el mismo para cualquier representación mínima del sistema, de modo que, si consideramos la descomposición en valores singulares de Φ_i (que nos permite estimar el orden del sistema), tenemos:

$$\Phi_i = U \Sigma V' = O_i Z_i = O_i T^{-1} T Z_i = O_i^* Z_i^*$$

donde T es una matriz invertible $n \times n$ cualquiera y O_i^* es, por tanto, una matriz de observabilidad correspondiente a otra representación mínima del sistema. Esto implica que cualquier matriz de observabilidad de una representación mínima del sistema debe ser de la forma

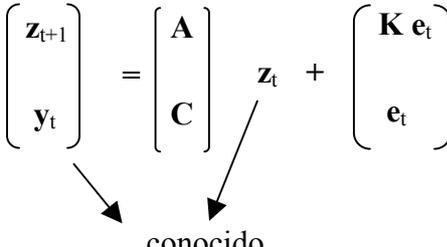
$$O_i = U T^{-1}$$

para alguna matriz T^{-1} de dimensión $n \times n$ que, al ser elegida libremente entre las matrices invertibles de rango n , determina el sistema coordinado para los estados.

4.4 Estimación de las matrices del sistema

Como ya hemos indicado, una vez conocida la secuencia de estados, las matrices del sistema $\{A, K, C, R\}$ pueden estimarse por mínimos cuadrados: si además de las observaciones (y_t) disponemos de z_{t+1} y de z_t , podemos considerar las ecuaciones

$$\begin{pmatrix} z_{t+1} \\ y_t \end{pmatrix} = \begin{pmatrix} A \\ C \end{pmatrix} z_t + \begin{pmatrix} K e_t \\ e_t \end{pmatrix}$$



Dado que las perturbaciones e_t están incorrelacionadas con los estados, estas ecuaciones pueden resolverse por mínimos cuadrados para encontrar estimadores de A y de C . Los residuos de esas mismas regresiones, calculados mediante las relaciones

$$\rho_{wt} = z_{t+1} - A z_t$$

$$\rho_{vt} = y_t - C z_t$$

donde las matrices $\{\mathbf{A}, \mathbf{C}\}$ y los estados deben ser sustituidos por sus valores estimados, proporcionan estimadores de las perturbaciones que podemos utilizar para calcular las matrices \mathbf{K} y \mathbf{R} .

En las ecuaciones anteriores hemos utilizado la secuencia de estados desplazados un periodo \mathbf{z}_{t+1} . La matriz que contiene esta secuencia de estados $\mathbf{Z}_{i+1} = [\mathbf{z}_{i+1} \ \mathbf{z}_{i+2} \ \dots \ \mathbf{z}_{i+j}]$ puede calcularse a partir de los datos teniendo en cuenta la relación

$$\mathbf{Y}_f / \mathbf{Y}_{p+} = \mathbf{O}_{i-1} \mathbf{Z}_{i+1}$$

que nos proporciona la fórmula

$$\mathbf{Z}_{i+1} = (\mathbf{O}_{i-1})^\dagger (\mathbf{Y}_f / \mathbf{Y}_{p+}).$$

Además, la matriz \mathbf{O}_{i-1} es, por definición, la propia matriz \mathbf{O}_i suprimiendo las últimas m filas.

Alternativamente a la estimación por mínimos cuadrados, la propia matriz de observabilidad (que depende de \mathbf{A} y de \mathbf{C}) puede proporcionar directamente estimadores para las matrices \mathbf{A} y \mathbf{C} , como vimos al estudiar el método de Aoki. Recordemos que la dimensión de \mathbf{y}_t es m y la de \mathbf{z}_t es n , de modo que \mathbf{C} es de dimensiones $m \times n$, \mathbf{A} es de dimensiones $n \times n$ y \mathbf{O}_i es de dimensiones $m_i \times n$. A partir de la definición de \mathbf{O}_i observamos que

$$\mathbf{C} = \mathbf{O}_i (1:m, 1:n)$$

$$\mathbf{O}_i (1:m(i-1), 1:n) \mathbf{A} = \mathbf{O}_i (m:m_i, 1:n)$$

donde estamos utilizando la notación de MATLAB, de modo que $\mathbf{M} (i:j, k:l)$ representa la matriz formada extrayendo de \mathbf{M} la intersección de las filas i hasta j con las columnas k hasta l . Estas dos ecuaciones nos permiten estimar \mathbf{C} y \mathbf{A} . Dado que estaremos trabajando con un estimador de \mathbf{O}_i , la última ecuación, que contiene más condiciones que incógnitas, se resolvería en el sentido de mínimos cuadrados.

4.5 Algoritmo completo

Van Overschee y De Moor (1996) proponen tres variantes alternativas para la solución del problema de identificación de sistemas estocásticos mediante métodos de subespacios.

La primera variante estima las secuencias de estados \mathbf{Z}_i , \mathbf{Z}_{i+1} y, a partir de éstas, las matrices \mathbf{A} y \mathbf{C} por mínimos cuadrados. La segunda variante corresponde al procedimiento alternativo descrito en el apartado anterior: primero se estima una matriz de observabilidad extendida \mathbf{y} , a partir de ésta, las matrices \mathbf{A} y \mathbf{C} . A partir de este punto, las dos variantes prosiguen la identificación del sistema en función de las matrices $\mathbf{\Omega} = E(\mathbf{z}_{t+1} \mathbf{y}'_t)$ y $\mathbf{\Gamma} = E(\mathbf{y}_t \mathbf{y}'_t)$. Con esto se consiguen estimadores insesgados, pero se presenta el problema de la “secuencia real positiva”: las matrices \mathbf{A} , \mathbf{C} , $\mathbf{\Omega}$ y $\mathbf{\Gamma}$ así encontradas pueden no corresponder a ningún proceso estocástico real.

La tercera variante propuesta por Van Overschee y De Moor (1996) calcula una secuencia de estados, estima las matrices \mathbf{A} y \mathbf{C} por regresión sobre los estados¹, y estima después las matrices \mathbf{Q} , \mathbf{R} y \mathbf{S} ⁽²⁾ a partir de los residuos de la regresión. De este modo, se asegura que el modelo identificado genera una secuencia real positiva de covarianzas (corresponde a un proceso estocástico), si bien los estimadores de \mathbf{Q} , \mathbf{R} y \mathbf{S} resultan sesgados³. Si se busca una representación en innovaciones $\{\mathbf{A}, \mathbf{C}, \mathbf{K}, \mathbf{R}\}$, se propone calcular las matrices auxiliares $\mathbf{\Omega}$ y $\mathbf{\Gamma}$ a partir de \mathbf{Q} , \mathbf{R} y \mathbf{S} , y resolver a continuación una ecuación de Riccati para encontrar la matriz \mathbf{K} .

En este trabajo optamos por utilizar la secuencia auxiliar de estados \mathbf{Z}_i , \mathbf{Z}_{i+1} y la regresión posterior para el cálculo de las matrices \mathbf{A} y \mathbf{C} . Este es el enfoque de la tercera variante propuesta anteriormente. Sin embargo, y frente a dicha variante, para el cálculo de la matriz \mathbf{K} del modelo en innovaciones utilizaremos las fórmulas alternativas que vimos anteriormente al estudiar el método de Aoki -Östermark y Aoki (1992), Vargas

¹ Hasta aquí procede igual que el primer algoritmo.

² El algoritmo de Van Overschee y De Moore no trabaja con la representación en innovaciones. Ver Anexo IV, apartado IV.1, para una definición de las matrices \mathbf{Q} , \mathbf{R} y \mathbf{S} .

³ Con un sesgo que tiende a cero cuando i tiende a infinito.

(1999)- y que evitan tener que resolver una ecuación de Riccati. Además, al estimar \mathbf{K} por este procedimiento se evita también el problema de la secuencia real positiva.

Las fórmulas a utilizar para estimar \mathbf{K} son:

$$\mathbf{e}_t = \mathbf{y}_t - \mathbf{C} \mathbf{z}_t$$

$$E [(\mathbf{z}_{t+1} - \mathbf{A} \mathbf{z}_t) \mathbf{e}'_t] = \mathbf{K} E [\mathbf{e}_t \mathbf{e}'_t]$$

Lo que se traduce en

$$\mathbf{K} = E [(\mathbf{z}_{t+1} - \mathbf{A} \mathbf{z}_t) \mathbf{e}'_t] E [\mathbf{e}_t \mathbf{e}'_t]^{-1}$$

Nótese que los estimadores de los términos $(\mathbf{z}_{t+1} - \mathbf{A} \mathbf{z}_t)$ y \mathbf{e}_t son precisamente los residuos de la regresión con la que se estimaron \mathbf{A} y \mathbf{C} .

Otra opción para estimar las matrices por mínimos cuadrados, una vez estimada la secuencia de estados, es utilizar la ecuación de observaciones

$$\mathbf{y}_t = \mathbf{C} \mathbf{z}_t + \mathbf{e}_t$$

para estimar la matriz \mathbf{C} y los residuos \mathbf{e}_t (y la matriz \mathbf{R} a partir de éstos) y a continuación utilizar la ecuación de transición de estados

$$\mathbf{z}_{t+1} = \mathbf{A} \mathbf{z}_t + \mathbf{K} \mathbf{e}_t$$

para estimar las matrices \mathbf{A} y \mathbf{K} .

Con estas consideraciones, indicamos a continuación el algoritmo completo que proponemos y que utilizaremos como algoritmo de subespacios genérico.

1 Calcular las proyecciones:

$$\Phi_i = Y_f / Y_p \approx Y_f Y'_p (Y_p Y'_p)^{\dagger} Y_p$$

$$\Phi_{i-1} = Y_{f-} / Y_{p+} \approx Y_{f-} Y'_{p+} (Y_{p+} Y'_{p+})^{\dagger} Y_{p+}$$

2 Calcular la descomposición en valores singulares de la matriz de proyección estimada Φ_i acondicionada por las matrices de pesos W_1 y W_2

$$W_1 \Phi_i W_2 = U \Sigma V'$$

3 Determinar el orden n del sistema por inspección de los valores singulares de Σ . Seleccionar las n primeras columnas de U y la esquina superior izquierda $n \times n$ de Σ para obtener las matrices U_1 y Σ_1 .

4 Estimar una matriz de observabilidad O_i mediante la fórmula:

$$O_i = W_1^{-1} U_1 \Sigma_1^{1/2}$$

y obtener O_{i-1} por eliminación de las últimas m filas de O_i

5 Calcular Z_i y Z_{i+1} de la forma:

$$Z_i = (O_i)^{\dagger} \Phi_i \quad Z_{i+1} = (O_{i-1})^{\dagger} \Phi_{i-1}$$

6 Estimar C por regresión de $[y_{i+1}, y_{i+2}, \dots, y_{i+j}]$ sobre Z_i y tomar los residuos de regresión E :

$$[y_{i+1}, y_{i+2}, \dots, y_{i+j}] = C Z_i + E$$

$$\text{Calcular } R = E E'$$

7 Estimar A y K por regresión de Z_{i+1} sobre Z_i y sobre E :

$$Z_{i+1} = A Z_i + K E$$

Cuadro 4.1. Algoritmo genérico de subespacios. Variante propuesta.

4.6 Algoritmos particulares

La elección de las matrices \mathbf{W}_1 (invertible, arbitraria) y \mathbf{W}_2 (de rango completo, arbitraria) permite incluir como casos particulares del método general de subespacios una amplia serie de algoritmos propuestos en la literatura. Este es el caso de los denominados algoritmos de componentes principales (PC) y algoritmos de análisis de variables canónicas (CVA ó CCA).

Algoritmo de componentes principales

El algoritmo de componentes principales determina las componentes principales de la matriz $\mathbf{C}_i = E(\mathbf{Y}_f \mathbf{Y}_p')$. Esta matriz está constituida por matrices de autocovarianza ($\mathbf{\Gamma}$) en la forma:

$$\mathbf{C}_i = \begin{pmatrix} \mathbf{\Gamma}_i & \mathbf{\Gamma}_{i-1} & \dots & \mathbf{\Gamma}_2 & \mathbf{\Gamma}_1 \\ \mathbf{\Gamma}_{i+1} & \mathbf{\Gamma}_i & \dots & \mathbf{\Gamma}_3 & \mathbf{\Gamma}_2 \\ \mathbf{\Gamma}_{i+2} & \mathbf{\Gamma}_{i+1} & \dots & \mathbf{\Gamma}_4 & \mathbf{\Gamma}_3 \\ \vdots & \vdots & & \vdots & \vdots \\ \mathbf{\Gamma}_{2i-1} & \mathbf{\Gamma}_{2i-2} & \dots & \mathbf{\Gamma}_{i+1} & \mathbf{\Gamma}_i \end{pmatrix}$$

El método de Aoki pertenece a este tipo de algoritmos. Recordemos que el método de Aoki se basa en la matriz $\mathbf{H} = E(\mathbf{y}_t^+ \mathbf{y}_{t-1}')$

$$\mathbf{H} = \begin{pmatrix} \mathbf{\Gamma}_1 & \mathbf{\Gamma}_2 & \mathbf{\Gamma}_3 & \dots & \mathbf{\Gamma}_{N_p} \\ \mathbf{\Gamma}_2 & \mathbf{\Gamma}_3 & \mathbf{\Gamma}_4 & \dots & \mathbf{\Gamma}_{N_p+1} \\ \mathbf{\Gamma}_3 & \mathbf{\Gamma}_4 & \mathbf{\Gamma}_5 & \dots & \mathbf{\Gamma}_{N_p+2} \\ \vdots & \vdots & \vdots & & \vdots \\ \mathbf{\Gamma}_{N_f} & \mathbf{\Gamma}_{N_f+1} & \mathbf{\Gamma}_{N_f+2} & \dots & \mathbf{\Gamma}_{N_f+N_p-1} \end{pmatrix}$$

Obsérvese que, tomando unos valores $N_f = N_p = i$, las matrices de covarianzas que componen la matriz \mathbf{C}_i son las mismas que componen la matriz \mathbf{H} del método de Aoki. El motivo de la diferente ordenación de las matrices de covarianzas es la distinta ordenación que los dos métodos hacen al trabajar con las observaciones.

Van Overschee y De Moor (1996, p. 79) demuestran que este algoritmo corresponde a una elección particular de las matrices de acondicionamiento \mathbf{W}_1 y \mathbf{W}_2 en el contexto del algoritmo general. En concreto, corresponde a la elección

$$\mathbf{W}_1 = \mathbf{I}_{mi}$$

$$\mathbf{W}_2 = \mathbf{Y}'_p (\mathbf{j}^{-1} \mathbf{Y}_p \mathbf{Y}'_p)^{-1/2} \mathbf{Y}_p$$

Lo que permite incluir el algoritmo de componentes principales en el marco del algoritmo general de subespacios considerado.

Algoritmo de correlaciones canónicas CCA

El algoritmo de correlaciones canónicas (CCA ó CVA, Larimore 1983) realiza un análisis de correlaciones canónicas entre las observaciones “pasadas” \mathbf{Y}_p y las “futuras” \mathbf{Y}_f . Van Overschee y De Moor (1996) demuestran que este algoritmo corresponde a una elección particular de las matrices \mathbf{W}_1 y \mathbf{W}_2 en el contexto del algoritmo general. En concreto, corresponde a la elección

$$\mathbf{W}_1 = (\mathbf{j}^{-1} \mathbf{Y}_f \mathbf{Y}'_f)^{-1/2}$$

$$\mathbf{W}_2 = \mathbf{I}_j.$$

Larimore (1996) y Bauer y Ljung (2002) defienden la optimalidad del método CCA frente a los otros algoritmos de subespacios, en particular para la identificación de procesos estocásticos. Además, CCA, frente a muchos de los otros algoritmos de subespacios, es insensible a la escala de las variables, ya que el propio método realiza una normalización. Bauer y Ljung (2002) refieren un artículo de Bauer (aún no publicado en el momento) en el que demuestra que, para procesos estocásticos estacionarios, el método CCA es asintóticamente equivalente a (pseudo) máxima verosimilitud, es decir, que los estimadores tienen la misma varianza asintótica que los obtenidos maximizando la función de verosimilitud gaussiana.

4.7 El método de Aoki como método de subespacios

Llegados a este punto podemos interpretar el método de Aoki como un método de subespacios. Aoki (1997) realiza una comparación entre ambos métodos (el suyo y los de subespacios) e indica que, aunque los procedimientos numéricos utilizados para el cálculo de las proyecciones ortogonales son distintos, los dos comparten el mismo marco conceptual, en particular para series sin señales exógenas.

Desde el punto de vista del procedimiento práctico, sabemos ya que, utilizando una matriz de acondicionamiento apropiada en el algoritmo de subespacios (la correspondiente al método de componentes principales), los dos métodos trabajarían sobre la misma matriz y con los mismos objetivos. La especificación del modelo (determinación del orden) sería común. En cuanto a la estimación, los algoritmos de subespacios ofrecen alternativas más sencillas que el método de Aoki, basadas en herramientas de álgebra numérica eficientes y robustas, como la descomposición **QR**.

Desde el punto de vista de la interpretación teórica, el método de Aoki busca una representación en el espacio de estados que proporcione una buena aproximación a la secuencia de autocovarianzas muestral contenida en la matriz de Hankel **H**, y fue desarrollado inicialmente como un método de variables instrumentales. Aunque ya se indicaba la posibilidad de calcular la secuencia de estados a partir de las observaciones, el algoritmo original no utiliza esta secuencia como herramienta para la estimación de las matrices del modelo.

El método de Aoki presenta además las ventajas de anidamiento estricto, estabilidad de los submodelos y especificación eficiente, propiedades que se mantendrían si la estimación del modelo se llevara a cabo mediante el método de subespacios equivalente (método de componentes principales).

Por otro lado, podríamos decir que las principales aportaciones de los algoritmos de subespacios respecto al método de Aoki consisten en:

- La concepción de la secuencia de estados como un instrumento directamente calculable a partir de las observaciones y que permite la estimación por mínimos cuadrados de todas las matrices del modelo.

Análisis de Sistemas Cointegrados mediante Métodos de Subespacios

- La implementación de los cálculos mediante operaciones algebraicas numéricas robustas como la descomposición **QR**.
- Proporcionar un marco muy general (algoritmo general) para este tipo de métodos de identificación, que contempla el enfoque de Aoki como caso particular. Los distintos algoritmos de subespacios corresponden a distintas selecciones de las matrices de pesos \mathbf{W}_1 y \mathbf{W}_2 . Además, este marco general permite abordar la cuestión de seleccionar las matrices de pesos \mathbf{W}_1 y \mathbf{W}_2 de forma óptima, lo que en el caso estocástico estacionario conduce al algoritmo CCA (Bauer y Ljung 2002).

5 IDENTIFICACIÓN DE PROCESOS NO ESTACIONARIOS EN EL ESPACIO DE ESTADOS

5.1 Introducción

El estudio de la cointegración supone abordar procesos no estacionarios. En un contexto general de análisis estadístico de modelos no estacionarios en el espacio de estados, citamos, en particular, a Gourieroux y Monfort (1997), que proporcionan también (cap. 16) una excelente panorámica de las aplicaciones de modelos en el espacio de estados; Peña y Poncela (2005) presentan un tratamiento teórico general de modelos con factores comunes dinámicos no estacionarios; Bauer y Wagner (2003) desarrollan una forma canónica en el espacio de estados para representación de procesos con raíces unitarias.

Aunque los algoritmos de subespacios vienen desarrollándose desde la década de los 80, su aplicación a procesos no estacionarios ha sido poco abordada aún. Una revisión de la literatura proporciona el algoritmo de Aoki (Aoki y Havenner 1991) como el primer caso de aplicación de un algoritmo de subespacios a procesos no estacionarios, y el único que hasta el momento ha conseguido una cierta continuidad y relevancia en su aplicación práctica (Havenner 1997, Östermark 1997, Vargas 2000). Como ya hemos indicado, aunque el algoritmo de Aoki no fue inicialmente concebido como un método de subespacios, es conceptualmente equivalente (Aoki 1997) y lo clasificamos como tal. Dentro de los métodos de subespacios propiamente dichos, la cuestión de aplicación a sistemas cointegrados ha sido abordada recientemente en los trabajos de Bauer y Wagner (2002) y en Larimore (2000), centrados en el algoritmo de correlaciones canónicas CCA.

En este capítulo analizaremos en primer lugar, y desde el espacio de estados, distintas posibilidades de descomposición analítica de una serie temporal en componentes cíclicos y tendenciales, siguiendo principalmente a Aoki y Havenner (1991). Tras ese análisis, revisaremos el método de dos etapas propuesto por Aoki para modelar procesos no estacionarios, elaborado a partir de su algoritmo para procesos estacionarios que ya discutimos en el capítulo 3. Vargas (2000) presenta una revisión más actualizada del algoritmo original, al que también aporta un contraste estadístico de cointegración. Continuaremos el capítulo con la revisión del método ACCA de Bauer y Wagner (2002). Una vez vistos los dos algoritmos propuestos en la literatura, la revisión del

procedimiento propuesto por Aoki para análisis de procesos no estacionarios, en combinación con ciertos resultados relativos a la optimalidad del método de subespacios conocido como CCA para la identificación de sistemas estocásticos, nos condujo a plantearnos un método de subespacios para procesos estocásticos no estacionarios que denominaremos “algoritmo de correlaciones canónicas bietápico”, o CCA2. Concluiremos el capítulo con un breve análisis sobre los contrastes de tendencias comunes (especificación del modelo de tendencias).

5.2 Descomposiciones analíticas ciclo-tendencia

Con modelos estrictamente estocásticos¹, es frecuente realizar una descomposición de la serie en componentes basándose en las diferentes frecuencias presentes en la dinámica de los datos: un modelo para capturar la tendencia (o componente de baja frecuencia) y un modelo para capturar los ciclos (o desviaciones de alta frecuencia con respecto a la tendencia)². El modelo de tendencia capturaría las posibles faltas de estacionariedad junto con los componentes de baja frecuencia, de modo que los datos que usemos para calcular el modelo de altas frecuencias sean estacionarios y no se vean afectados por los movimientos de baja frecuencia.

Dado que los efectos de tendencia y ciclo no se observan por separado, cualquier descomposición en tendencia y ciclo depende necesariamente de un conjunto particular de hipótesis o restricciones impuestas. Además, no hay consenso respecto a las restricciones que deben usarse para la identificación de estas componentes. La validez del modelo conjunto puede contrastarse, pero no es posible contrastar cada componente (tendencia y ciclo) por separado.

¹ En la práctica, no podemos saber a partir de un número finito de observaciones si la tendencia es estocástica o es una tendencia determinista de condiciones suficientemente cercanas, aunque contrastes como el anteriormente citado de Elliott, Rothenberg y Stock (1996) pueden servirnos para tomar una decisión.

² En muchos casos esta descomposición se impone para interpretar o explicar mejor el comportamiento de la serie y es una descomposición arbitraria, pero que puede resultar práctica.

Como puede verse en Vargas (2000), el comportamiento a largo plazo del sistema depende de los autovalores de la matriz dinámica \mathbf{A} . Si alguno de los autovalores tiene un módulo superior a la unidad, el sistema presenta un comportamiento explosivo. Si todos los autovalores son inferiores a la unidad en módulo, el sistema es estable, y el efecto de cualquier perturbación en el vector de estado va convergiendo a cero paulatinamente. En este último caso, la serie es estacionaria (es agregación de estados estacionarios). Si algunos de los autovalores presentan módulos unitarios, el sistema no es estable, y esta falta de estacionariedad de algunos estados se transmitirá a algunas de las series observadas.

Aoki y Havenner (1991) plantean tres descomposiciones alternativas tendencia-ciclo de una serie temporal no estacionaria basadas en la descomposición de la matriz dinámica \mathbf{A} de una representación en el espacio de estados del sistema: una utiliza la descomposición espectral de \mathbf{A} y las otras dos se basan en la descomposición de Schur de esta matriz (éstas últimas corresponden a los modelos de cointegración y de corrección del error). En los tres casos se procede del mismo modo:

- a) Partición del vector de estados en dos grupos, conforme a los autovalores de \mathbf{A} : un grupo de estados que llevan información sobre la dinámica de largo plazo (o tendencia, o movimientos de baja frecuencia), asociados a los autovalores de mayor valor de la matriz \mathbf{A} , y otro grupo de estados que llevan información sobre la dinámica de alta frecuencia (corto plazo, ciclo).
- b) Selección de una base para el espacio de estados tal que la representación del sistema tenga una matriz dinámica \mathbf{A}^* diagonal (descomposición espectral), o triangular (descomposición de Schur).

1 Descomposición espectral

Descompongamos la matriz dinámica \mathbf{A} de una representación en el espacio de estados en sus autovectores y autovalores ordenados por orden decreciente, en la forma¹ :

¹ Asumimos, por simplicidad, que el número de autovectores independientes es igual a la dimensión de \mathbf{A} .

$$\mathbf{A} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^{-1} = \begin{bmatrix} \mathbf{Q}_1 & \mathbf{Q}_2 \end{bmatrix} \begin{bmatrix} \mathbf{\Lambda}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{\Lambda}_2 \end{bmatrix} \begin{bmatrix} \mathbf{Q}^1 \\ \mathbf{Q}^2 \end{bmatrix}$$

donde $\mathbf{\Lambda}_1$ es una matriz diagonal que contiene los k mayores autovalores de \mathbf{A} , y $\mathbf{\Lambda}_2$ es una matriz diagonal que contiene los $n-k$ autovalores menores. \mathbf{Q}_1 y \mathbf{Q}_2 corresponden a las k primeras y $n-k$ últimas columnas de \mathbf{Q} respectivamente. \mathbf{Q}^1 y \mathbf{Q}^2 corresponden a las k primeras y $n-k$ últimas filas de \mathbf{Q}^{-1} respectivamente.

Buscamos trabajar con un nuevo vector de estados $[\boldsymbol{\tau}'_t \boldsymbol{\eta}'_t]'$ tal que los estados $\boldsymbol{\tau}_t$ estén asociados a los k mayores autovalores (y por tanto al modelo para la tendencia) y los $\boldsymbol{\eta}_t$ estén asociados a los $n-k$ menores autovalores (y por tanto al modelo para el ciclo). Recordemos que ante el cambio de base para los estados $\mathbf{z}^*_t = \mathbf{T} \mathbf{z}_t$ se tiene una nueva matriz dinámica $\mathbf{A}^* = \mathbf{T}\mathbf{A}\mathbf{T}^{-1}$. Teniendo esto en cuenta, junto con la descomposición $\mathbf{A} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^{-1}$, encontramos que, para que la matriz dinámica sea diagonal, debemos tomar un vector de estados¹ $[\boldsymbol{\tau}'_t \boldsymbol{\eta}'_t]' = \mathbf{Q}^{-1} \mathbf{z}_t$. La ecuación de transición de estados queda entonces:

$$\begin{bmatrix} \boldsymbol{\tau}_{t+1} \\ \boldsymbol{\eta}_{t+1} \end{bmatrix} = \mathbf{\Lambda} \begin{bmatrix} \boldsymbol{\tau}_t \\ \boldsymbol{\eta}_t \end{bmatrix} + \mathbf{Q}^{-1} \mathbf{K} \mathbf{e}_t$$

que corresponde a modelos separados para la tendencia $\boldsymbol{\tau}_t$ y para el ciclo $\boldsymbol{\eta}_t$:

$$\boldsymbol{\tau}_{t+1} = \mathbf{\Lambda}_1 \boldsymbol{\tau}_t + \mathbf{Q}^1 \mathbf{K} \mathbf{e}_t$$

$$\boldsymbol{\eta}_{t+1} = \mathbf{\Lambda}_2 \boldsymbol{\eta}_t + \mathbf{Q}^2 \mathbf{K} \mathbf{e}_t$$

2 Descomposición de Schur

Mantendremos una partición del vector de estados en dos grupos, conforme a la distinta magnitud de los autovalores de la matriz \mathbf{A} : un grupo asociado a la tendencia (k autovalores mayores) y un grupo asociado al ciclo ($n-k$ autovalores menores).

¹ Esto es, $\boldsymbol{\tau}_t = \mathbf{Q}^1 \mathbf{z}_t$ y $\boldsymbol{\eta}_t = \mathbf{Q}^2 \mathbf{z}_t$. También se tiene $\mathbf{z}_t = \mathbf{Q}_1 \boldsymbol{\tau}_t + \mathbf{Q}_2 \boldsymbol{\eta}_t$

De igual modo que en el apartado anterior buscábamos una representación del sistema en la que la matriz dinámica fuera diagonal, podemos buscar una representación en la que dicha matriz sea triangular. Para ello podemos basarnos en la descomposición de Schur (Golub y Van Loan, 1983): $\mathbf{A} = \mathbf{Q}\mathbf{U}\mathbf{Q}'$, donde \mathbf{U} es triangular superior y \mathbf{Q} cumple $\mathbf{Q}\mathbf{Q}' = \mathbf{I}_n$

Tomar unos nuevos estados \mathbf{z}^*_t de la forma $\mathbf{z}^*_t = \mathbf{Q}'\mathbf{z}_t$ nos conduce a una representación cuya matriz de transición de estados es triangular superior. Si la descomposición de Schur se realiza tomando un orden decreciente para los valores singulares y consideramos la partición del vector de estados en tendencia y ciclo $\mathbf{z}^*_t = [\boldsymbol{\tau}'_t \boldsymbol{\eta}'_t]'$, tenemos el modelo

$$\begin{bmatrix} \boldsymbol{\tau}_{t+1} \\ \boldsymbol{\eta}_{t+1} \end{bmatrix} = \mathbf{U} \begin{bmatrix} \boldsymbol{\tau}_t \\ \boldsymbol{\eta}_t \end{bmatrix} + \mathbf{Q}' \mathbf{K} \mathbf{e}_t = \begin{bmatrix} \mathbf{U}_{11} & \mathbf{U}_{12} \\ \mathbf{0} & \mathbf{U}_{22} \end{bmatrix} \begin{bmatrix} \boldsymbol{\tau}_t \\ \boldsymbol{\eta}_t \end{bmatrix} + \mathbf{Q}' \mathbf{K} \mathbf{e}_t$$

Esta representación con matriz dinámica triangular superior y estados asociados a los autovalores ordenados por orden decreciente corresponde a un modelo en el que la tendencia $\boldsymbol{\tau}_t$ (que es el componente que modela las no estacionariedades) se ve afectada por el estado del ciclo $\boldsymbol{\eta}_t$, de la forma $\boldsymbol{\tau}_{t+1} = \mathbf{U}_{11} \boldsymbol{\tau}_t + \mathbf{U}_{12} \boldsymbol{\eta}_t + \mathbf{Q}' \mathbf{K} \mathbf{e}_t$. Nótese que los dos últimos sumandos son estacionarios. Este análisis es la base del método empírico de identificación de sistemas cointegrados en dos etapas de Aoki que veremos más adelante.

Si en vez de trabajar con una matriz dinámica triangular superior consideramos una matriz dinámica triangular inferior¹, estaremos tomando una descomposición en la que el modelo para el ciclo depende del modelo para la tendencia (“equilibrio” a largo plazo). Aoki y Havenner (1991, p. 46) indican que esta representación corresponde a la noción de “corrección del error”, mientras que la representación con matriz dinámica triangular superior corresponde a la noción de cointegración.

¹ Aoki y Havenner (1991) indican que se puede llegar a esta representación utilizando la misma descomposición de Schur con matriz triangular superior, pero con autovalores ordenados en orden creciente, lo que supone cambiar en el vector de estados el orden tendencia-ciclo $[\boldsymbol{\tau}'_t \boldsymbol{\eta}'_t]'$ por el de ciclo-tendencia $[\boldsymbol{\eta}'_t \boldsymbol{\tau}'_t]'$.

Equivalencia de las descomposiciones

Como hemos visto, existe una equivalencia matemática entre las descomposiciones estudiadas tal que, si la especificación del modelo es conocida, las distintas descomposiciones corresponden a representaciones equivalentes que producirían las mismas predicciones (aunque la descomposición en ciclo y tendencia sería distinta).

Resulta muy interesante la observación de Gilbert (1993) respecto a que, dada la equivalencia de estos modelos, la alternativa “cointegración frente a corrección del error” es una cuestión para la teoría económica que no puede ser resuelta mediante el análisis estadístico.

Sin embargo, si la especificación del modelo no es conocida, cada descomposición puede tener asociado un procedimiento empírico distinto de búsqueda de la especificación. Esto puede dar lugar a modelos con distintas especificaciones dependiendo de la descomposición utilizada, y, por tanto, a distintas predicciones y distintas calidades de predicción.

5.3 Método empírico de Aoki para el modelo de cointegración

Basándose en la descomposición de Schur anterior, Aoki (1990) propone un procedimiento en dos etapas que modela primero la tendencia o componentes de baja frecuencia del sistema. Las autocovarianzas para el modelo de altas frecuencias son calculadas en una segunda etapa a partir de desviaciones medidas respecto a una media dinámica (tendencia) ajustada al sistema.

Descompondremos el vector de series observadas \mathbf{y}_t en una componente asociada a la tendencia a largo plazo o movimientos de baja frecuencia $\mathbf{C}\boldsymbol{\tau}_t$ y un residuo estacionario \mathbf{y}^*_t , probablemente con correlación temporal. Los estados $\boldsymbol{\tau}_t$ condensan la información correspondiente a los movimientos de baja frecuencia. El modelo para las observaciones es:

$$\mathbf{y}_t = \mathbf{C}\boldsymbol{\tau}_t + \mathbf{y}^*_t$$

Identificación de Procesos No Estacionarios en el Espacio de Estados

con ecuación de transición de estados¹

$$\boldsymbol{\tau}_{t+1} = \mathbf{A}_\tau \boldsymbol{\tau}_t + \mathbf{K}_\tau \mathbf{y}^*_t$$

Como hemos indicado, la idea es calcular primero, a partir de las observaciones, un modelo que capture las tendencias y movimientos amplios, y obtener un residuo estacionario \mathbf{y}^*_t que es explicado a continuación mediante otro modelo en el espacio de estados:

$$\mathbf{y}^*_t = \mathbf{C}_\eta \boldsymbol{\eta}_t + \mathbf{e}_t$$

$$\boldsymbol{\eta}_{t+1} = \mathbf{A}_\eta \boldsymbol{\eta}_t + \mathbf{K}_\eta \mathbf{e}_t$$

De forma que el modelo completo es:

$$\begin{bmatrix} \boldsymbol{\tau}_{t+1} \\ \boldsymbol{\eta}_{t+1} \end{bmatrix} = \begin{bmatrix} \mathbf{A}_\tau & \mathbf{K}_\tau \mathbf{C}_\eta \\ \mathbf{0} & \mathbf{A}_\eta \end{bmatrix} \begin{bmatrix} \boldsymbol{\tau}_t \\ \boldsymbol{\eta}_t \end{bmatrix} + \begin{bmatrix} \mathbf{K}_\tau \\ \mathbf{K}_\eta \end{bmatrix} \mathbf{e}_t$$

$$\mathbf{y}_t = \begin{bmatrix} \mathbf{C}_\tau & \mathbf{C}_\eta \end{bmatrix} \begin{bmatrix} \boldsymbol{\tau}_t \\ \boldsymbol{\eta}_t \end{bmatrix} + \mathbf{e}_t$$

Obsérvese que es un modelo en el que la evolución de las tendencias comunes se ve afectada por el estado del ciclo, pero el ciclo evoluciona independientemente del estado de la tendencia.

Con este modelo, si las series en \mathbf{y}_t son no estacionarias, cualquier vector $\boldsymbol{\alpha}$ que cumpla $\boldsymbol{\alpha}'\mathbf{C}_\tau = 0$ (es decir, cualquier vector que pertenezca al espacio nulo de \mathbf{C}'_τ), cumplirá que la combinación $\boldsymbol{\alpha}'\mathbf{y}_t$ es estacionaria, luego $\boldsymbol{\alpha}$ es, por tanto, un vector de cointegración.

¹ Considerado aisladamente, este modelo para la tendencia no cumple en realidad las condiciones de una representación en el espacio de estados, dada la correlación temporal de los residuos \mathbf{y}^*_t . Esta correlación debería ser eliminada, que es en realidad la función del segundo modelo o modelo anidado para la perturbación \mathbf{y}^*_t . El modelo conjunto sí cumple las condiciones de una representación en el espacio de estados.

Aoki (1990) propone utilizar el “algoritmo de Aoki” que vimos anteriormente para encontrar tanto un modelo para la tendencia como un modelo para el ciclo. El parámetro N_p es utilizado para establecer la separación entre dinámica a corto y a largo plazo. En Aoki y Havenner (1991) se indica que, heurísticamente, si se realiza una regresión de una serie sobre su valor anterior, se obtiene un indicador del efecto medio a largo plazo. Si la regresión se realiza sobre varios de los valores retardados, se puede obtener un cierto detalle cíclico. Teniendo esto en cuenta, al buscar un modelo para la tendencia se elige un valor de N_p muy pequeño, típicamente 1 para series multivariantes¹. Los residuos de este modelo entran en un nuevo proceso de identificación, en el que el parámetro N_p es incrementado según convenga, conforme a lo indicado al desarrollar el algoritmo de Aoki para series estacionarias.

Tanto para el modelo de tendencia como para el de ciclos, se utiliza la matriz de Hankel de autocovarianzas para fijar el número de estados necesarios en el modelo. Si para el modelo de tendencias se requiere un número de estados superior al número de series en y_t , será necesario un modelo para las tendencias superior al primer orden. Si el número de estados que condensan la información sobre las tendencias es menor que el número de series no estacionarias, entonces hay “tendencias comunes”, compartidas por varias de las series en estudio.

Determinación del orden del modelo de tendencias

La determinación del orden del modelo de tendencias se lleva a cabo en el método de Aoki mediante el análisis de la matriz de Hankel de autocovarianzas muestrales para valores pequeños de N_p (típicamente 1). Esto quiere decir que trabajaremos con la matriz $\mathbf{H}' = [\hat{\Gamma}'_1 \hat{\Gamma}'_2 \dots \hat{\Gamma}'_{N_f}]$. Si tomamos $N_f = 1$, entonces estimamos el orden a partir de la matriz de autocovarianzas muestral de primer orden $\hat{\Gamma}'_1$.

Aoki (1990) o Aoki y Havenner (1991) sugieren utilizar el mismo procedimiento que para series estacionarias, que consiste entonces en determinar el rango de $\hat{\Gamma}'_1$ mediante el análisis de sus valores singulares, adoptando un criterio para decidir si un valor singular es significativo o no. Así por ejemplo, Arndt y Foster (1997) ordenan de mayor a menor

¹ Un modelo con N_p retardos para una serie vectorial de dimensión m equivale a modelos univariantes de hasta mN_p retardos para las series individuales (Aoki y Havenner 1991).

los valores singulares (σ_i) de $\hat{\Gamma}_1$ y eligen un número de estados \tilde{n} tal que el cociente de valores singulares $\sigma_{\tilde{n}+1}/\sigma_1$ sea de orden $1/\sqrt{T}$.

5.4 El algoritmo ACCA de Bauer y Wagner

Bauer y Wagner (2002) proponen un método de subespacios para el análisis de sistemas cointegrados. Citan el método de Aoki como el único caso previo de aplicación de algoritmos de subespacios a procesos no estacionarios.

El método constituye una variante del algoritmo conocido como CCA o CVA (canonical variate analysis), por lo que se denomina CCA adaptado o ACCA. Recordamos (apartado 4.6) que CCA resulta especialmente adecuado para procesos estocásticos (sin variables exógenas) y, en el caso estacionario, es asintóticamente equivalente a la estimación (pseudo) máximo-verosímil (Bauer y Ljung 2002).

En el caso de series no estacionarias cointegradas, Bauer y Wagner parten de una forma canónica de representación de un sistema cointegrado con c tendencias comunes:

$$\begin{pmatrix} \boldsymbol{\tau}_{t+1} \\ \boldsymbol{\eta}_{t+1} \end{pmatrix} = \begin{pmatrix} \mathbf{I}_c & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_\eta \end{pmatrix} \begin{pmatrix} \boldsymbol{\tau}_t \\ \boldsymbol{\eta}_t \end{pmatrix} + \begin{pmatrix} \mathbf{K}_\tau \\ \mathbf{K}_\eta \end{pmatrix} \mathbf{e}_t$$

$$\mathbf{y}_t = \begin{pmatrix} \mathbf{C}_\tau & \mathbf{C}_\eta \end{pmatrix} \begin{pmatrix} \boldsymbol{\tau}_t \\ \boldsymbol{\eta}_t \end{pmatrix} + \mathbf{e}_t$$

donde \mathbf{C}_τ es una matriz ortonormal, $\boldsymbol{\tau}_t$ es un vector de c caminos aleatorios (tendencias comunes, estados no estacionarios) y $\boldsymbol{\eta}_t$ es un vector de $n-c$ estados estacionarios. Frente al modelo de Aoki, y por la naturaleza diagonal por bloques de la matriz de transición de estados, en este caso la descomposición ciclo-tendencia es tal que los ciclos y las tendencias evolucionan con dinámica independiente. El algoritmo CCA estándar estima de forma consistente los vectores de cointegración, pero no se ha demostrado la estimación consistente de los parámetros correspondientes a la parte estacionaria del sistema. El algoritmo ACCA modifica al CCA para conseguir una estimación consistente de todos los parámetros del sistema.

En la práctica, ACCA requiere modificar la matriz \mathbf{W}_1 del método CCA, sustituyéndola por una matriz que utiliza ya una primera estimación de los vectores de cointegración. Para conseguir esa primera estimación de los vectores de cointegración puede usarse, en una primera fase, el algoritmo CCA estándar. Los detalles prácticos del algoritmo no son inmediatos y admiten diversas variantes, por lo que no los detallamos aquí, pudiendo consultarse en Bauer y Wagner (2002).

Bauer y Wagner proponen además varios contrastes estadísticos del número de tendencias comunes y un criterio de estimación consistente del orden del sistema.

5.5 Un algoritmo propuesto: CCA bietápico

Rescatando la idea intuitiva del método de Aoki de realizar una estimación bietápica para abordar procesos no estacionarios, pero considerando que, frente a otros métodos de subespacios como el de componentes principales (equivalente al método de Aoki), el algoritmo CCA resulta especialmente adecuado para la identificación de procesos estocásticos (Bauer y Ljung 2002), proponemos una aplicación en dos etapas del algoritmo CCA para identificar procesos no estacionarios. La primera etapa debe captar las tendencias no estacionarias, y la segunda etapa modela los residuos estacionarios de la primera. La anidación de ambos modelos proporciona un único modelo cointegrado con la misma estructura que el proporcionado por el método de Aoki (sección 5.3).

Frente al algoritmo ACCA, que proporciona un modelo en el que la dinámica de tendencias y ciclos evoluciona independientemente, el método CCA2 proporciona (como el método de Aoki) un modelo en el que la dinámica de la tendencia depende del estado del ciclo. En cualquier caso, una vez obtenida una representación en el espacio de estados, puede rotarse la base de estados y pasar de un tipo de representación al otro.

Aunque la identificación de la segunda etapa (ciclos) se llevaría a cabo conforme al algoritmo estándar CCA para procesos estacionarios, para la primera etapa (tendencias) debemos modificar el criterio de selección del orden, que en este caso determina el número de tendencias comunes del modelo. Discutiremos este aspecto de estimación del número de tendencias comunes en el siguiente apartado (5.6).

En resumen, proponemos el método CCA2 porque:

- Al ser un método bietápico, mantiene la intuición subyacente al método de Aoki. La ventaja teórica esperada de un procedimiento bietápico para modelar sistemas con frecuencias muy diferentes está en que el modelo para las altas frecuencias se calcula una vez se ha eliminado el efecto de las bajas frecuencias, de modo que esperamos que la tendencia no distorsione el cálculo del modelo para el ciclo.
- Frente a las otras variantes de algoritmos de subespacios, y, en particular, frente al algoritmo de componentes principales utilizado por el método de Aoki, el algoritmo CCA resulta teóricamente óptimo para identificar procesos estocásticos (Bauer y Ljung 2002).

Por estos motivos proponemos el algoritmo CCA2 como método de subespacios para identificación de procesos estocásticos cointegrados, y lo incluiremos como un algoritmo más en los experimentos de simulación que realizaremos para valorar la utilidad práctica de los métodos de subespacios, comparando los distintos métodos entre sí y con el procedimiento de máxima verosimilitud de Johansen.

5.6 Contrastes del número de tendencias comunes mediante métodos de subespacios

Basados en el método de Aoki. Contraste de tendencias comunes de Vargas

Estimar el número de tendencias comunes es el mismo problema que estimar el orden del modelo de tendencias. Los métodos de determinación del orden del modelo de tendencias desarrollados como extensión natural del método de Aoki para procesos estacionarios se basan en el análisis de los valores singulares de la matriz de Hankel $\hat{\Gamma}_1$. Aoki y Havenner (1991) o Havenner (1997) no indicaban un criterio distinto al seguido para procesos estacionarios a la hora de seleccionar el orden del modelo de tendencias, y el procedimiento ha sido criticado por falta de solidez estadística. Östermark (1997) recurre en este punto a métodos alternativos de determinación del número de tendencias

comunes¹: utiliza el método de Johansen y, una vez cuenta con un orden estimado para el modelo de tendencias, prosigue la estimación en el espacio de estados.

Trabajar con las covarianzas de series no estacionarias del mismo modo a como estas funciones se definieron para series estacionarias puede resultar problemático, ya que la varianza muestral de una serie no estacionaria es divergente a infinito. Sin embargo, se puede demostrar (Peña y Poncela 2005) que, para un conjunto de m series integradas de orden 1 y rango de cointegración r (esto es, con $m-r$ raíces unitarias independientes), las matrices de autocovarianza muestral generalizada ($T^{-1}\hat{\Gamma}_k$), definidas como

$$T^{-1}\hat{\Gamma}_k = T^{-2} \sum_{t=k+1}^T (y_{t-k} - \bar{y})(y_t - \bar{y})'$$

convergen débilmente a una misma matriz aleatoria que tiene $m-r$ valores singulares no nulos y r valores singulares nulos. Así pues, cualquiera de estas matrices $T^{-1}\hat{\Gamma}_k$ presenta r autovalores convergentes a cero y $n-r$ convergentes a constantes no nulas. Considerando este resultado, si calculamos algunas de las matrices de covarianza muestral $T^{-1}\hat{\Gamma}_0$, $T^{-1}\hat{\Gamma}_1$, $T^{-1}\hat{\Gamma}_2$, ..., el número de valores singulares significativos (significativamente no nulos) debería ser $n-r$, igual al número de tendencias comunes. Además, los autovectores asociados a los valores singulares (asintóticamente) nulos convergen a los vectores de cointegración.

Esto proporciona una justificación teórica al uso de la matriz $T^{-1}\hat{\Gamma}_1$ para la determinación del orden del modelo de tendencias, o al uso de $\hat{\Gamma}_1$ si se utilizan cocientes de valores singulares.

Vargas (2000) aporta al método de Aoki un procedimiento de contraste estadístico secuencial del orden de cointegración del modelo, o número de tendencias comunes, que, en definitiva, consiste en contrastar el número de valores singulares asintóticamente no nulos de la matriz $T^{-1}\hat{\Gamma}_1$.

¹ Puesto que el problema en este punto es el de determinación del rango de cointegración de las series, podemos considerar cualquiera de las alternativas desarrolladas a este efecto, como las que se indicaron en el capítulo correspondiente al tema de cointegración.

Identificación de Procesos No Estacionarios en el Espacio de Estados

El contraste propuesto por Vargas (2000) comienza por ordenar de mayor a menor los valores singulares ($\sigma_1 > \sigma_2 > \dots > \sigma_m$) de la matriz de autocorrelaciones muestral $\hat{\Gamma}_1$ y elegir como estadísticos los cocientes ($\delta_2, \delta_3, \dots, \delta_m$) entre cada valor singular y el primero (y mayor) de ellos: $\delta_i = \sigma_i/\sigma_1$. Una apreciación cualitativa interesante es que, en general, se manifiesta un salto de magnitud muy apreciable en estos cocientes correspondiente al paso entre los valores singulares que convergen a cero y los que no.

Comenzando por δ_2 , se realiza un contraste secuencial bajo la hipótesis nula de que el cociente en cuestión es asintóticamente no nulo (correspondiente a un estado no estacionario). Si para algún cociente se rechaza la hipótesis nula, se acepta que ese cociente y los siguientes son no significativos (son asintóticamente nulos). El número de cocientes δ_i ($i = 2, \dots, m$) que se suponen convergentes a cero es el número estimado de relaciones de cointegración¹.

Vargas (2000) calcula y tabula, para distintos tamaños muestrales (15, 25, 50, 75, 100) y número de series (de 2 a 5), la distribución empírica de los cocientes de valores singulares bajo la hipótesis nula (cociente δ_i asintóticamente no nulo). Sin embargo, la posible deriva existente en los procesos estocásticos considerados altera la distribución de los valores singulares de la matriz de autocorrelación de primer o segundo orden. Para evitar esta influencia de la deriva sobre los valores singulares, Vargas recomienda eliminar primero su efecto realizando una regresión sobre una función lineal del tiempo (\mathbf{x}_t) que elimine los términos deterministas, y tomar los residuos \mathbf{u}_t . El modelo completo que considera es:

$$\mathbf{y}_t = \boldsymbol{\beta} \mathbf{x}_t + \mathbf{u}_t$$

$$\mathbf{z}_{t+1} = \mathbf{A} \mathbf{z}_t + \mathbf{K} \mathbf{e}_t$$

$$\mathbf{u}_t = \boldsymbol{\mu}_u + \mathbf{C} \mathbf{z}_t + \mathbf{e}_t$$

donde \mathbf{x}_t es una función lineal del tiempo, \mathbf{z}_t es el vector de estados utilizado para modelar los residuos estocásticos \mathbf{u}_t (posiblemente no estacionarios) y \mathbf{e}_t es i.i.d. $(\mathbf{0}, \mathbf{R})$.

¹ Si todas las series individuales son I(1). Si hay mezcla de series individuales I(1) y I(0), el número de valores singulares que se admiten como significativos es el número de tendencias comunes independientes que se considera en el modelo.

Al estimar el modelo, la regresión inicial de y_t sobre unas variables x_t apropiadas¹ permite considerar los residuos u_t libres de tendencias deterministas y libres de deriva, y aplicar el contraste de cointegración sobre los valores singulares de la matriz de autocorrelaciones de primer orden. Si se considera que alguna de las series y_t es tendencia estocástica con deriva, para la estimación del modelo de tendencias se volverían a tomar los valores originales de y_t , en vez de los residuos (u_t) de la regresión inicial².

Nótese que, aunque se propone el procedimiento práctico anterior para abordar la influencia de componentes deterministas, no queda determinado el efecto sobre la validez de los contrastes. Por otro lado, los contrastes de Vargas (2000) abordan la dependencia de la distribución de los valores singulares respecto del número de tendencias comunes (contraste secuencial), respecto del número de series (tablas para distintas dimensiones de y_t) y respecto del tamaño muestral. Sin embargo, no aparece considerada de forma explícita la (posible) dependencia de la distribución de los estimadores δ_i respecto a los valores de otros parámetros (que pueden ser desconocidos a priori) del sistema, ya que las distribuciones teóricas no son conocidas. Si se produjera esta dependencia, la validez de las tablas del contraste quedaría limitada.

Basados en el algoritmo CCA. Contrastes de Bauer y Wagner.

Para un sistema cointegrado de m series temporales y_t con c tendencias comunes, que sea representable en el espacio de estados con un modelo de orden n , se tiene que, bajo condiciones muy generales (Bauer y Wagner 2002), los estimadores de exactamente c valores singulares del algoritmo CCA convergen a uno (con velocidad de convergencia T), mientras que los $n-c$ restantes convergen a valores menores que uno (con velocidad

¹ Havenner (1997) indica que si los regresores son una constante y t , es decir, $x_t = (1, t)'$, entonces los coeficientes pueden ser estimados por mínimos cuadrados ordinarios, el coeficiente de t es superconsistente con velocidad de convergencia $T^{3/2}$ y la ordenada en el origen es consistente con velocidad \sqrt{T} .

² Sin embargo, dos tendencias estocásticas que, sin considerar la deriva, comparten una misma tendencia común, pueden no estar cointegradas al considerar la deriva.

de convergencia $T^{1/2}$). Este resultado proporciona la base para realizar un contraste del número de tendencias comunes a partir de los valores singulares.

Aunque es conocida la distribución asintótica de algunos estadísticos (Bauer y Wagner 2002) formados a partir de los valores singulares, y pueden utilizarse esos estadísticos para realizar contrastes de cointegración, las distribuciones dependen en general de parámetros desconocidos del sistema, lo que hace problemática su aplicación práctica.

Otra posibilidad es buscar un criterio consistente de estimación¹, en vez de un contraste estadístico. Siguiendo este enfoque (Poskitt 2000; Wagner 2000, p.60), podemos estimar c como el número de valores singulares σ_i que cumplen la condición

$$\sigma_i^2 > 1 - h_T/T$$

donde h_T es una función de T tal que $h_T \rightarrow \infty$ y $h_T/T^{1/4} \rightarrow 0$ cuando $T \rightarrow \infty$. Sin embargo, la elección de la función afecta a los resultados de la estimación e introduce un elemento de arbitrariedad.

Bauer y Wagner (2002) proponen también contrastes de cointegración basados en los autovalores de la matriz de transición de estados (**A**) estimada conforme al algoritmo ACCA. Estos contrastes ofrecen la ventaja de que la distribución asintótica de dichos autovalores no depende de parámetros desconocidos del sistema. En un estudio mediante simulación de Bauer y Wagner (2003b) se comparan los resultados proporcionados por este tipo de contrastes con los proporcionados por el método de Johansen. Los resultados no son del todo concluyentes: aunque en muestras pequeñas ($T = 100, 200, 300$) los contrastes de subespacios analizados batían² en general al método de Johansen, resultan ser a veces mejores y a veces peores que dicho método, dependiendo de los parámetros del proceso generador, sin que se aprecien unas pautas o razones claras de los motivos de esas diferencias.

¹ Como vimos, Aoki también propone un criterio para su método, aunque la consistencia de ese criterio no aparece demostrada.

² Estiman con mayor índice de aciertos el número de tendencias comunes correcto

5.7 Propiedades de la estimación de parámetros mediante métodos de subespacios en modelos cointegrados

Como ya hemos indicado, el método de análisis de sistemas cointegrados más extendido es el método de Johansen, por lo que es el método que usaremos como referencia. El estudio comparativo de Gonzalo (1994) indica además una superioridad práctica en muestras finitas del método de Johansen frente a varios otros métodos habituales de cálculo del vector de cointegración (caso de 2 variables).

Recordemos que el método de Johansen utiliza una representación del sistema mediante un VAR de orden p . Una cuestión interesante es saber en qué medida este método puede verse afectado o limitado cuando el proceso generador no es un VAR puro, sino que sigue un proceso VARMA, de la forma $\mathbf{A}(L) \mathbf{y}_t = \mathbf{B}(L) \mathbf{e}_t$. En estas condiciones, la formulación VAR de orden finito, de la forma $\mathbf{A}^*(L) \mathbf{y}_t = \mathbf{e}_t^*$, sólo podrá constituir una aproximación al verdadero proceso generador. Sin embargo, el proceso VARMA de órdenes p y q finitos sí puede ser representado de forma exacta en el espacio de estados de orden finito.

Los principales resultados teóricos relativos a consistencia de los estimadores en procesos VARMA cointegrados son:

- El procedimiento de Johansen aplicado sobre un proceso generador VARMA cointegrado y considerando un VAR de orden p (fijo) proporciona estimadores consistentes del subespacio de cointegración (relaciones estables a largo plazo), pero los parámetros asociados a la dinámica a corto plazo no se estiman de forma consistente (Wagner 1999 y Wagner 2000). En particular, la matriz de covarianzas de las perturbaciones no se estima de forma consistente.
- El procedimiento CCA simple aplicado sobre procesos generadores VARMA cointegrados proporciona estimadores consistentes del subespacio de cointegración (Bauer y Wagner 2002).
- El procedimiento ACCA aplicado sobre procesos generadores VARMA cointegrados proporciona estimadores consistentes de todos los parámetros del modelo (Bauer y Wagner 2002).

Identificación de Procesos No Estacionarios en el Espacio de Estados

En el plano teórico, el método de subespacios de correlaciones canónicas simple (CCA) proporciona, al igual que el método de Johansen, estimadores consistentes del subespacio de cointegración¹, pero no está demostrada la estimación consistente del resto de parámetros del sistema.

Como hemos indicado, el algoritmo ACCA utiliza la forma canónica de Bauer y Wagner (2003) y modifica al algoritmo CCA para conseguir una estimación consistente de todos los parámetros del sistema, tanto los correspondientes a la parte no estacionaria (subespacio de cointegración, tendencias, relaciones a largo plazo) como los correspondientes a la parte estacionaria (ciclos, perturbaciones, dinámica a corto plazo).

El método CCA2 que proponemos proporciona la misma estimación del subespacio de cointegración que el método CCA, y por tanto, queda asegurada la estimación consistente de las relaciones de cointegración. Aunque con el procedimiento de dos etapas esperamos estimar de forma consistente también el resto de parámetros, postergamos el análisis teórico a la espera de que las pruebas prácticas a las que vamos a someter al método CCA2 justifiquen su utilización.

¹ Bauer y Wagner (2002, p.57). Siendo c el número de tendencias comunes, supongamos que estimamos el modelo por el método CCA y, mediante una regresión de las observaciones y_t sobre los primeros c estados, obtenemos la matriz C_1 . El subespacio nulo de C_1 es un estimador (super)consistente del subespacio de cointegración. Se trata, en definitiva, de aplicar el algoritmo CCA estándar tomando un número de estados igual al número de tendencias comunes c .

6 DISEÑO DE LOS EXPERIMENTOS

6.1 Introducción

En este capítulo exponemos la planificación y las condiciones de realización (material y métodos) de las simulaciones llevadas a cabo con el objetivo de valorar la calidad de distintos procedimientos de subespacios para identificar sistemas estocásticos cointegrados.

Trabajaremos con diferentes procesos generadores de observaciones: de dos y tres series temporales, con una o dos relaciones de cointegración, con distintos tamaños muestrales y valores de los parámetros generadores. Todos los sistemas generadores serán formulados en el espacio de estados.

Los procedimientos estudiados serán el algoritmo de correlaciones canónicas CCA, el recientemente propuesto algoritmo ACCA, de Bauer y Wagner (2002), y el algoritmo CCA2, original de esta investigación. En una primera fase se contemplará también el método de Aoki. Todos estos métodos serán comparados entre sí y con el método que hemos adoptado como referencia: el procedimiento de Johansen – máxima verosimilitud bajo especificación VAR(p)-.

Los aspectos valorados serán la capacidad de estimación de las relaciones de cointegración (subespacio de cointegración) y la calidad de predicción a corto plazo.

Antecedentes

Entre los antecedentes destacan los trabajos de Gonzalo (1994) y Bauer y Wagner (2003b).

Gonzalo (1994) compara, para el caso de dos variables cointegradas, distintos estimadores de la relación de cointegración (mínimos cuadrados simples, mínimos cuadrados no lineales, Johansen, componentes principales, correlaciones canónicas). Gonzalo obtiene la distribución asintótica de estos estimadores y realiza también un estudio práctico comparativo de los resultados obtenidos por cada estimador en muestras finitas. La metodología de nuestro estudio está basada en este trabajo, si bien nos centraremos en los métodos de estimación de subespacios, tomaremos distintos

procesos generadores (más generales), añadiremos nuevos aspectos a comparar (capacidad de predicción) y emplearemos herramientas gráficas de comparación.

Bauer y Wagner (2003b) realizan diversas simulaciones donde comparan el algoritmo ACCA con el método de Johansen. Contemplan los aspectos de contraste de cointegración, estimación del subespacio de cointegración y capacidad de predicción del modelo estimado, para distintos procesos generadores VARMA.

Aportaciones originales respecto a los antecedentes

- Un nuevo método de estimación: el algoritmo CCA bietápico o CCA2. Este método es una propuesta original de esta tesis doctoral y por tanto no ha sido estudiado con anterioridad.
- Un tipo de relación de cointegración más general que la considerada por Gonzalo (1994) o Bauer y Wagner (2003). En estos trabajos sólo se considera el caso en que las relaciones de cointegración son series estacionarias de esperanza cero, lo que supone una restricción que no tiene por qué cumplirse en series reales. Frente a los citados trabajos, nosotros consideramos la (posible) existencia de término constante en la relación de cointegración, lo que requiere alterar los procedimientos de estimación (con modificación de las características de los estimadores).
- Una forma original de medir la separación entre subespacios real y estimado: el ángulo que forman dichos subespacios. Esta medida de separación, además de resultar muy intuitiva en el caso de 2 o tres dimensiones, es generalizable a mayor número de series observadas y mayor dimensión del subespacio de cointegración. Por su parte, Gonzalo (1994) normaliza la primera componente del vector de cointegración al valor 1 y mide la distancia en la segunda componente entre el valor real y el estimado, pero este procedimiento sólo es válido para el caso de dos series y un único vector de cointegración. Bauer y Wagner (2003b) utilizan la distancia de Hausdorff, que es extensible a varias dimensiones pero resulta menos intuitiva que el ángulo, especialmente para subespacios representables en el espacio (dos o tres series cointegradas con uno o dos vectores de cointegración).

- Nuevos procesos generadores. Una de las limitaciones del método de Johansen es que trabaja con una formulación VAR, lo que puede resultar problemático cuando el proceso generador es un VARMA. La sospecha de que el mejor o peor rendimiento de unos métodos de identificación respecto a otros (Johansen / subespacios) pueda depender de las características del proceso generador hace interesante el estudio de una amplia variedad de procesos. Nosotros trabajaremos con procesos generadores de distintas dimensiones, formulados en el espacio de estados y que admiten una formulación VARMA equivalente. Trataremos de observar cómo influyen distintos parámetros del proceso generador en la calidad de cada algoritmo de identificación.

6.2 Separación del problema de estimación del problema de especificación del número de tendencias comunes

Tanto el método de Johansen como el algoritmo ACCA necesitan calcular el número de tendencias comunes como paso intermedio. La estimación de este valor podría llevarse a cabo por cualquier criterio al respecto (incluidos los criterios asociados con estos dos métodos). Si algún criterio permite estimar de forma óptima el número de tendencias comunes, puede ser usado en este punto, con independencia de que la estimación de parámetros se lleve a cabo después por otro procedimiento.

Por este motivo, al comparar los algoritmos, puede resultar interesante separar el problema de especificación del número de tendencias comunes del problema de estimación de parámetros, y es la línea que hemos adoptado. En definitiva, en los experimentos prácticos compararemos los estimadores obtenidos por los distintos métodos suponiendo siempre que el número de tendencias comunes calculado o impuesto es el correcto. En cualquier caso, recordemos que, si queremos estimar el número de tendencias comunes mediante métodos de subespacios, tenemos las opciones que se discutieron en el apartado 5.6, algunas de las cuales han sido comparadas con el método de Johansen sin resultados concluyentes respecto a la idoneidad de uno u otro procedimiento.

Aunque imponemos el número de tendencias comunes, estimamos el orden del modelo (p para una formulación VAR o n para una formulación en el espacio de estados)

mediante distintos criterios, asociados a los distintos procedimientos de identificación. Dado que obtendremos modelos con especificaciones diferentes a la del proceso generador, la calidad de los modelos obtenidos no podemos compararla en función de la distancia de valores de parámetros reales-estimados (salvo para el subespacio de cointegración, donde sí fijamos su dimensión para todos los algoritmos). Volveremos sobre este punto en el apartado 6.4.

6.3 Medida de la calidad de estimación del subespacio de cointegración (relación a largo plazo)

Para comparar la precisión en la estimación de los vectores de cointegración debemos elegir una medida de proximidad o calidad. Dado que los vectores de cointegración no son únicos, puesto que cualquier combinación lineal de vectores de cointegración es otro vector de cointegración, se trata de medir la separación entre dos subespacios: el subespacio formado por los vectores de cointegración estimados y el subespacio formado por los vectores de cointegración auténticos (los del proceso generador). Surgen entonces distintas posibilidades:

- Una posibilidad para medir la calidad de estimación es comparar dos bases normalizadas del subespacio de cointegración. Así por ejemplo, Gonzalo (1994), para el caso de un solo vector de cointegración con dos componentes $[v_1, v_2]'$, normaliza a 1 el valor de la primera componente y mide la separación entre el vector de cointegración $[1, \beta]'$ y el vector de cointegración estimado $[1, \hat{\beta}]'$ por la diferencia existente en el valor de la segunda componente. Aunque este método es simple e intuitivo para el caso de dos variables cointegradas, su extensión a más de dos variables presenta complicaciones, y la separación entre los vectores depende de qué componente elijamos normalizar a 1.
- Wagner (2000, p. 62) y Bauer y Wagner (2003b) eligen la distancia de Hausdorff para medir la calidad en la estimación del subespacio de cointegración. Si M y N son dos subespacios en R^s , esta distancia entre ambos viene dada por

$$d_H(M,N) = \max \left[\sup_{x \in M, \|x\|=1} \|(I-Q)x\| \quad \sup_{x \in N, \|x\|=1} \|(I-P)x\| \right]$$

donde \mathbf{Q} denota la proyección ortogonal sobre N , \mathbf{P} la proyección ortogonal sobre M , y $\|x\|$ denota la norma euclídea en \mathbb{R}^s . Para el caso de subespacios de distinta dimensión, esta distancia vale 1. Este sistema permite comparar las estimaciones entre métodos en el caso en que el número estimado de tendencias comunes sea distinto. Sin embargo, la penalización de tomar una distancia de valor 1 en caso de especificación incorrecta no deja de ser arbitraria. De hecho, en Bauer y Wagner (2003b) se separan los problemas de especificación y estimación y, tal y como hacemos en este trabajo, se comparan las estimaciones bajo la hipótesis de que el número de tendencias comunes se ha especificado correctamente.

En esta tesis proponemos una medida diferente de la separación entre subespacios, que, además de resultar intuitiva, no requiere ni se ve afectada por una normalización arbitraria de los vectores de cointegración. En el caso de vectores con representación espacial, esto es, de dos o tres dimensiones, una medida muy intuitiva de la separación entre los vectores es el ángulo que forman¹ (“el ángulo que los separa”, en el lenguaje habitual). Esta es la medida de separación (calidad) que proponemos en este trabajo. La utilización del ángulo como medida de separación entre los subespacios exige sin embargo algunas consideraciones:

- En el caso de dos series temporales, el subespacio de cointegración viene definido por un único vector en el plano. El ángulo entre vectores puede medir la separación entre subespacios. Podemos además considerar un sólo semiplano y definir ángulos positivos y negativos, para detectar sesgos en uno u otro sentido del vector estimado respecto al auténtico.

¹ El menor de los posibles, esto es, tomado en $(-\pi/2, \pi/2]$ si se distingue sentido o en $[0, \pi/2]$ si no se distingue sentido.

- En el caso de tres series temporales, el subespacio de cointegración puede estar definido por uno o por dos vectores de cointegración en el espacio (tres dimensiones). Si el subespacio viene definido por un único vector, el ángulo entre vectores puede medir la separación entre subespacios. Si el subespacio de cointegración viene definido por dos vectores (un plano), podemos asociar a cada subespacio un único vector, el vector ortogonal al plano que lo define, o vector director del plano. La separación entre subespacios puede entonces medirse por el ángulo entre vectores directores.
- En el caso de más de tres series temporales, podemos utilizar una generalización matemática del ángulo como medida de separación entre subespacios, aunque la interpretación intuitiva se hace más complicada. Matlab ofrece la función “subspace” a este efecto, definida precisamente como una medida del ángulo entre subespacios. Este caso de más de tres series temporales no ha sido abordado en los experimentos prácticos.

6.4 Medida de la calidad de estimación de la dinámica a corto plazo

Sabemos que el algoritmo CCA estándar proporciona una estimación consistente de los vectores de cointegración (relación a largo plazo entre las variables), y que las modificaciones ACCA y CCA2 de ese algoritmo buscan obtener también una buena estimación de los parámetros correspondientes al ciclo, que influye en la dinámica a corto plazo. Para medir la calidad de la estimación de la dinámica a corto plazo existen varias posibilidades.

Teóricamente, una de las posibilidades es expresar todos los modelos, incluido el modelo generador, en una formulación identificable común tomada como referencia (por ejemplo, la forma canónica de Bauer y Wagner 2003, o una representación VARMA). El problema es que las especificaciones (los órdenes) de cada uno de los modelos estimados pueden ser diferentes, lo que hace complicada la comparación de parámetros bajo este enfoque, ya que, entre otras cosas, el número de parámetros de cada modelo es diferente. Es el problema que existe si, por ejemplo, queremos medir la proximidad entre un modelo AR(3) y un MA(1) observando directamente los valores de los parámetros: aunque ambos modelos podrían ser considerados un ARMA(3,1) con

valores nulos para algunos de los parámetros, es difícil saber directamente cómo de próximos están dos modelos en base a los valores de los parámetros.

Parece mejor camino comparar las propiedades estadísticas que se derivan de cada modelo (que son precisamente, además, lo que los modelos buscan captar), o bien comparar directamente el error de predicción a corto plazo, que es una medida práctica de la calidad con la que el modelo estimado ha captado la dinámica a corto plazo del proceso generador. Esta última línea es la que hemos adoptado, centrándonos en el estudio mediante simulación del error de predicción a corto plazo (a horizonte k , con $k = 1, 2, 3$) proporcionado por cada uno de los distintos modelos.

Sea $\hat{y}_{t|t-k}$ la predicción del valor y_t a horizonte k (predicción del valor y_t utilizando valores previos hasta y_{t-k}) conforme al modelo que estemos valorando. Algunos estadísticos habituales para medir la calidad de predicción en una muestra de tamaño h ($y_{T+1}, y_{T+2}, \dots, y_{T+h}$) son:

- Raíz del error cuadrático medio (RMSE):

$$\sqrt{\sum_{t=T+1}^{T+h} (y_t - \hat{y}_{t|t-k})^2 / h}$$

- Error absoluto medio (MAE):

$$\sum_{t=T+1}^{T+h} |y_t - \hat{y}_{t|t-k}| / h$$

- Error porcentual absoluto medio (MAPE)

$$100 \sum_{t=T+1}^{T+h} \left| \frac{y_t - \hat{y}_{t|t-k}}{y_t} \right| / h$$

- Coeficiente de desigualdad de Theil

$$\frac{\sqrt{\sum_{t=T+1}^{T+h} (y_t - \hat{y}_{t|t-k})^2 / h}}{\sqrt{\sum_{t=T+1}^{T+h} y_t^2 / h + \sum_{t=T+1}^{T+h} \hat{y}_{t|t-k}^2 / h}}$$

Cualquiera de estos cuatro estadísticos resulta útil para comparar las predicciones proporcionadas por distintos modelos, pero la interpretación de los valores exige algunas consideraciones. Estas consideraciones nos llevaron a no seleccionar el coeficiente de desigualdad de Theil, pese a ser nuestro primer candidato.

Obsérvese que los dos primeros estadísticos anteriores (RMSE, MAE) dependen de la escala, y los dos últimos (MAPE, Theil), por ser porcentuales, dependen del nivel (medio) de la serie. La dependencia del valor del estadístico respecto del nivel de la serie puede resultar problemática en series no estacionarias $I(1)$, en las que el nivel de la serie no es estacionario, pero la amplitud media de las perturbaciones o de los incrementos sí lo es. Para ver esto, supongamos que generamos 110 observaciones de un camino aleatorio $y_{t+1} = y_t + e_t$, donde e_t es ruido blanco $N(0, \sigma^2)$, y calculamos el coeficiente de desigualdad de Theil entre las observaciones 101 y 110 para predicciones $\hat{y}_{t+1|t} = y_t$. Obsérvese que el numerador del coeficiente de Theil es un estimador de σ^2 , mientras que el denominador depende del nivel de la serie. Si, por casualidad, al empezar la predicción a partir de la observación y_{100} la serie presenta niveles altos (en valor absoluto y en relación al valor medio de la perturbación e_t), el coeficiente de Theil será próximo a cero. El coeficiente de Theil será mayor si el nivel medio de la serie en la observación 100 resulta ser bajo. El resultado es que el coeficiente de Theil se ve afectado, no tanto por la calidad de ajuste (en términos absolutos), sino por el nivel que el camino aleatorio ha alcanzado al calcular el estimador. En múltiples simulaciones podemos esperar niveles muy distintos alcanzados por un camino aleatorio, y por tanto una alta variabilidad del coeficiente de Theil, causada no por diferencias en el error absoluto de ajuste, sino por medirlo en términos porcentuales. En modelos estocásticos con nivel dinámico en los que la amplitud de las perturbaciones es estática (no depende del nivel de la serie), no parece apropiado medir el error en términos porcentuales respecto al nivel de la serie.

Diebold y López (1995) indican que una manera sencilla y popular de comparar la precisión de predicciones se consigue mediante el estadístico U de Theil, definido, para horizonte de predicción k y una muestra de predicción de tamaño h ($y_{T+1}, y_{T+2}, \dots, y_{T+h}$) por:

$$U_k = \frac{\sum_{t=T+1}^{T+h} (y_t - \hat{y}_{t|t-k})^2}{\sum_{t=T+1}^{T+h} (y_t - y_{t-k})^2}$$

donde $\hat{y}_{t|t-k}$ es la predicción a horizonte k (predicción del valor y_t utilizando valores previos hasta y_{t-k}) conforme al modelo que estamos valorando. Este estadístico compara el error cuadrático de la predicción a horizonte k de un modelo dado, con el error cuadrático de la predicción $\hat{y}_{t|t-k} = y_{t-k}$ (la predicción que correspondería a un camino aleatorio). Valores de U_k menores a 1 indican mejores predicciones (en error cuadrático) que la predicción “trivial” $\hat{y}_{t|t-k} = y_{t-k}$.

Propiedades del estadístico U de Theil que resultan especialmente apropiadas para valorar la calidad de predicción en muestras de series no estacionarias son la independencia respecto del nivel y de la escala de la serie. Por ese motivo es el estadístico que hemos seleccionado para las comparaciones: confrontaremos las distribuciones del estadístico U de Theil obtenidas (en muestras de predicción) por cada uno de los modelos estimados conforme a los distintos procedimientos de identificación estudiados. La propia distribución del estadístico U para cada modelo, además de facilitar las comparaciones, nos aporta información directa sobre la calidad de predicción del modelo.

Muestra de predicción y cálculo del estado inicial para predicción

Al generar las observaciones del sistema, hemos tomado muestras de trabajo de tamaño T y hemos generado un 10% adicional de observaciones para valorar la capacidad de predicción de los modelos (muestra de predicción). Una vez estimados los distintos modelos, la predicción de valores la calculamos en el espacio de estados mediante el filtro de Kalman.

La predicción basada en el espacio de estados (recursiva) exige contar con un estimador del estado inicial y de su matriz de covarianzas. La calidad de predicción de la muestra de validación se puede ver afectada por la estimación de ese estado inicial¹.

¹ Matlab ofrece distintas posibilidades ya programadas a este efecto.

En realidad, al calcular el estado inicial para la muestra de predicción contamos ya con toda la muestra de trabajo. La opción elegida en este trabajo ha sido utilizar para la muestra de trabajo un estado inicial suavizado, que minimiza el error de ajuste de la muestra¹. Para la muestra de predicción utilizamos como estado inicial el estado final estimado en la muestra de trabajo. En cualquier caso, se comprobó (y así mostraremos posteriormente) que, siguiendo este procedimiento de utilizar toda la muestra de trabajo para estimar el estado inicial de la muestra de predicción, el método de estimación del estado inicial de la muestra de trabajo tiene normalmente poca influencia en la muestra de predicción.

Así por ejemplo, la Figura 6-1 muestra las predicciones a horizonte uno de una serie generada con el proceso generador PGD2 (que veremos posteriormente), modelo estimado por el procedimiento CCA simple, y dos estimaciones distintas del estado inicial: una nula y otra suavizada.

La diferencia debida a la distinta estimación del estado inicial, notable en las primeras observaciones, va desapareciendo al contar con más observaciones. Para un número suficiente de observaciones de la muestra de trabajo, la estimación del estado final (que es el estado inicial para la muestra de predicción) será prácticamente la misma.

¹ La instrucción “predict” de Matlab permite utilizar esta opción preprogramada, por lo que no entraremos en detalle.

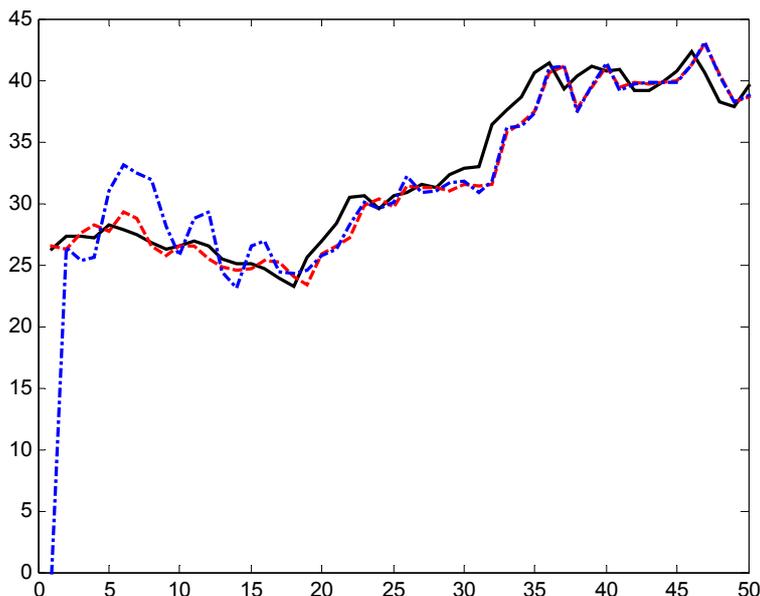


Figura 6-1. Predicción a horizonte 1 de una de las series (línea continua, negro) generadas por el proceso generador PGD2. Modelo estimado por el método CCA. En rojo (línea discontinua), la predicción con estado inicial suavizado. En azul (punto y raya), la predicción con estado inicial nulo.

6.5 Términos constantes en las relaciones de cointegración

Las relaciones de cointegración son de la forma

$$y_{1,t} = a_2 y_{2,t} + \dots + a_m y_{m,t} + u_t,$$

donde las series $y_{i,t}$ son $I(1)$ y u_t es estacionario. En general, los experimentos comparativos realizados en la literatura, tales como Gonzalo (1994) o Bauer y Wagner (2003b), se limitan al caso en que los residuos u_t de las relaciones de cointegración son series estacionarias de esperanza cero. Puesto que esta restricción no tiene por qué cumplirse en series reales, en este trabajo hemos querido abordar el caso más general en que los residuos de las relaciones de cointegración no tienen necesariamente esperanza cero. Esta circunstancia exige modificar los estimadores del vector de cointegración para mantener la consistencia.

Si los residuos u_t de las relaciones de cointegración son estacionarios pero de esperanza no nula, pueden expresarse como $u_{i,t} = \mu_i + v_{i,t}$, donde $\mu_i = E(u_{i,t})$ y $v_{i,t}$ es estacionaria de

esperanza cero. Al término μ_i lo denominamos término constante de la relación de cointegración.

La estimación máximo verosímil por el método de Johansen para el caso de que aparezcan términos constantes (pero no tendencias deterministas) en las relaciones de cointegración fue desarrollada en el apartado 2.6. El procedimiento es similar al caso general, pero requiere obtener los residuos de tres regresiones auxiliares.

Para los estimadores de subespacios, proponemos restar las medias muestrales antes de estimar las relaciones de cointegración. Esto elimina en gran medida el desplazamiento causado por el término constante de la relación de cointegración. Obsérvese que si

$$y_{1,t} = \beta y_{2,t} + u_t$$

donde u_t es estacionaria de esperanza μ_u , entonces

$$y_{1,t} - \bar{y}_1 = \beta (y_{2,t} - \bar{y}_2) + (u_t - \bar{u})$$

y, por la convergencia de la media \bar{u} a la esperanza μ_u , el último término converge a una serie estacionaria de esperanza cero.

Del mismo modo, en el caso $y_{1,t} = \beta y_{2,t} + u_t$, con esperanza de u_t igual a cero, el estimador de mínimos cuadrados consistente para β es el que resulta de hacer una regresión de $y_{1,t}$ sobre $y_{2,t}$. Pero si la esperanza de u_t es distinta de cero, el estimador de mínimos cuadrados consistente para β es el que resulta de hacer una regresión de $y_{1,t}$ sobre $y_{2,t}$ y sobre un término constante. Este estimador es el mismo que el de hacer una regresión de $(y_{1,t} - \bar{y}_1)$ sobre $(y_{2,t} - \bar{y}_2)$, es decir, el mismo estimador que para el caso en que no hay término constante en la relación de cointegración, pero trabajando en diferencias respecto a las medias.

Los experimentos prácticos han proporcionado buenos resultados para estos estimadores tomando los datos en diferencias respecto a las medias. Como medida de precaución, comparamos también las estimaciones obtenidas por el método de Johansen siguiendo las dos opciones:

- Sin restar las medias, permitiendo término constante en las relaciones de cointegración.

- Restando las medias, suponiendo que el término constante es nulo.

Si nuestro enfoque de restar las medias es adecuado, esperamos que no haya diferencia entre los dos procedimientos al estimar los vectores de cointegración.

6.6 Ajuste de los parámetros de los métodos de estimación

Orden de la aproximación autorregresiva VAR(p) en el método de Johansen

Aunque podríamos seguir el enfoque de Gonzalo (1994) y probar, para cada proceso generador, los resultados obtenidos con aproximaciones VAR de distintos órdenes, en la práctica el experimentador elegirá un determinado criterio para seleccionar, dada una muestra de observaciones del sistema, el orden p del vector autorregresivo con el que aproxima el sistema. Al comparar métodos de identificación, resulta interesante incorporar el efecto del criterio de selección (especificación) del modelo. Entre los procedimientos de selección del orden más utilizados se encuentran el criterio de información de Akaike (AIC) y el criterio de información de Schwarz (BIC). Decidimos utilizar este último, que es un criterio consistente de estimación del orden para modelos VAR (Lütkepohl 1991, p. 132).

El criterio BIC consiste en elegir aquel modelo que minimiza el valor

$$\text{BIC} = -2 (\log L)/T + k (\log T) / T$$

donde L es el máximo valor de la función de verosimilitud y k es el número de parámetros estimados. Para hallar el valor de L , en la sección 2.6 puede encontrarse la fórmula del método de Johansen que proporciona el valor de la función de verosimilitud de la muestra, utilizando una formulación VEC de un sistema cointegrado con rango de cointegración r , suponiendo perturbaciones normales, y suponiendo que no existen tendencias deterministas pero sí existen términos constantes en las relaciones de cointegración.

Se consideró un valor mínimo $p = 2$ para el orden autorregresivo de la representación VAR en niveles, que corresponde a un orden 1 para la representación VEC.

Orden del modelo en los métodos de subespacios. Parámetros p y f

La determinación del orden (n) del modelo en los métodos de subespacios está basada en los valores singulares de la matriz $\mathbf{W}_1 \Phi_1 \mathbf{W}_2$. Sólo los n primeros valores singulares de esta matriz (desconocida pero estimable) son no nulos (ver Anexo IV). En la práctica, los n primeros valores singulares estimados ($\hat{\sigma}_i$) convergen a valores positivos no nulos, y el resto converge a cero.

Existe una amplia variedad de criterios propuestos (Vargas 1999 realiza una revisión general). Revisando las últimas publicaciones, hemos seleccionado el criterio BA(n) de Bauer y Wagner (2003b), que consiste en elegir aquel valor de n que minimice la función

$$BA(n) = -\log(1 - \hat{\sigma}_{n+1}^2) + 2nm \log(T) / T$$

donde m es el número de series temporales (dimensión de \mathbf{y}_t), y donde T es el tamaño muestral. Bauer y Wagner (2003b) proponen este criterio por sus propiedades teóricas de consistencia y por sus buenos resultados en muestras finitas, en comparación con otras alternativas.

En las simulaciones se han tomado, dependiendo del proceso generador, valores fijos e iguales de los parámetros p y f de los algoritmos de subespacios. Existen distintas opciones para elegir valores para estos parámetros (Ljung 1999, p. 351). Una opción es elegir los valores p y f en base a un criterio de información¹. Para que el método funcione bien, se debe cumplir $f \geq n$ (desconocido a priori), y suele tomarse $p = f$.

En nuestros experimentos se comprobó que el valor de los parámetros p y f tenía poca influencia sobre los resultados, una vez superado el valor del orden del sistema. Tomar a priori un valor fijo de p y de f puede interpretarse como la consideración de un límite superior a priori sobre el orden del sistema subyacente.

¹ Bauer y Wagner (2003b), basándose en una demostración de consistencia, toman $p = f = 2 \times \rho(\text{AIC})$, donde $\rho(\text{AIC})$ es el orden de una aproximación VAR que minimiza el criterio de información de Akaike.

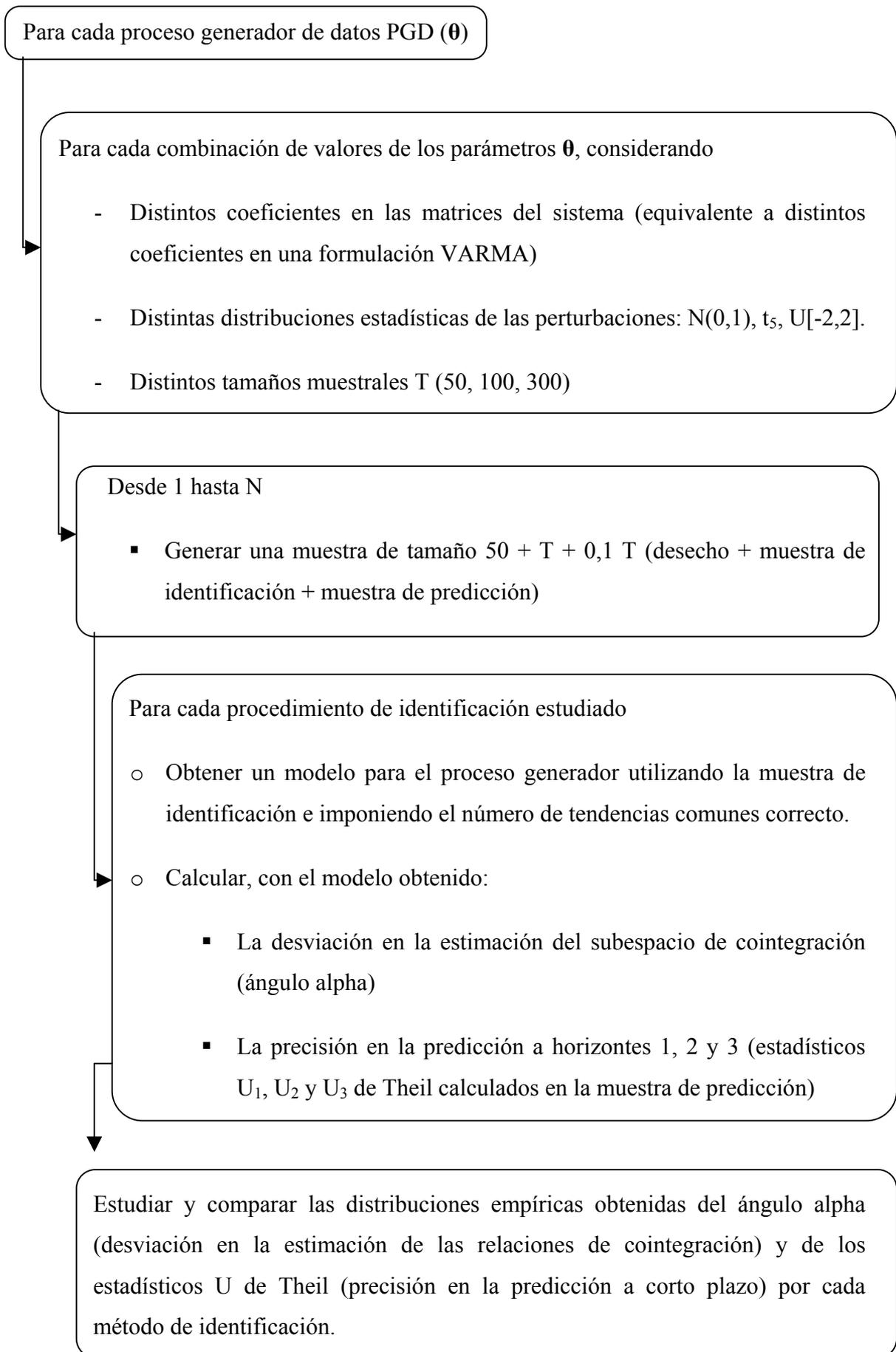
6.7 Proceso general de las simulaciones. Programación y herramientas de análisis

Se han seleccionado varios procesos generadores para las simulaciones, con distinto número de variables y de relaciones de cointegración. Cada proceso generador viene definido en función de un conjunto de parámetros que toman distintos valores en las simulaciones, buscando cubrir un rango amplio y variado de características del proceso generador.

Para cada proceso generador y combinación de valores de los parámetros, se generaron por simulación muestras de identificación y de predicción. En predicción se usaron muestras de un tamaño igual al 10% del tamaño de las muestras de identificación. Sobre la muestra de identificación se aplicaron los distintos métodos de identificación estudiados, obteniendo los respectivos modelos estimados. Para cada modelo, se calculó la separación entre los subespacios de cointegración real y estimado (ángulo α) y se calcularon los estadísticos U de Theil (calidad de predicción a corto plazo) a distintos horizontes. Repitiendo N veces la generación de muestras, se obtuvieron las distribuciones empíricas del ángulo α (separación entre los subespacios de cointegración real y estimado) y de los estadísticos U de Theil.

Los resultados han sido representados gráficamente mediante histogramas y exportados a Excel para analizar, mediante tablas y gráficos dinámicos, las medidas descriptivas (media, mediana, sesgo, dispersión,...) de las distribuciones obtenidas.

De forma más visual, el procedimiento seguido es:



Otro aspecto considerado al estudiar los diferentes procedimientos de identificación ha sido el orden estimado del sistema (esto es, el número de estados para una formulación en el espacio de estados, o el número de retardos para una formulación VAR).

Al valorar la capacidad predictiva de los modelos obtenidos por los distintos procedimientos comparados se ha incluido también un contraste no paramétrico de igualdad. También se ha estudiado la predicción mediante combinación de modelos.

Se seleccionó un número de repeticiones N buscando un equilibrio entre el tiempo de simulación y la precisión y estabilidad de los resultados. Se comprobó que, en general, con un número de repeticiones $N = 1.000$:

- Los histogramas de las distribuciones empíricas estudiadas presentan un aspecto suave, “continuo”.
- El aspecto de los histogramas prácticamente no varía al repetir series de simulaciones, de forma que las diferencias visuales existentes y detectadas entre las distribuciones de los distintos estimadores se mantienen al repetir el experimento.
- La variabilidad de los estadísticos descriptivos (media, mediana, IQR,...) de las distribuciones empíricas calculadas en series de simulaciones es reducida con relación a las diferencias existentes entre los distintos métodos comparados: las conclusiones comparativas observadas se mantienen en las repeticiones.

Para generar las perturbaciones pseudo-normales se utilizó la función “randn” de matlab. Para analizar el efecto de perturbaciones no gaussianas se escogió una distribución t con cinco grados de libertad (Figura 6-2), generada mediante la función “trnd”, y, en ocasiones, se utilizó también una distribución uniforme $[-2,2]$. En todos los casos se despreciaron las primeras 50 observaciones de los sistemas simulados.

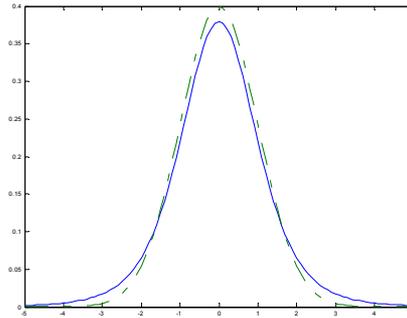


Figura 6-2. Comparación entre una distribución normal (0,1) (línea discontinua) y una t con cinco grados de libertad (línea continua).

La plataforma de trabajo utilizada ha consistido en un ordenador personal con microprocesador AMD Athlon XP 1600+ y sistema operativo Windows 2000. Todos los métodos de identificación de sistemas cointegrados analizados (CCA, ACCA, CCA2, Johansen) han sido programados en matlab por el autor. Para la estimación de la relación de cointegración en el caso de dos variables, también se programaron los métodos de mínimos cuadrados ordinarios (MCO), componentes principales (CP) y Aoki.

El método de Johansen y el método de Johansen para el caso de existir términos constantes en las relaciones de cointegración (en ausencia de tendencias deterministas) se programaron basándose en las descripciones de los algoritmos que aparecen en Hamilton (1994, c. 20). El método CCA y ACCA fueron programados basándose en una versión inicial del documento de Bauer y Wagner (2002), y fueron posteriormente depurados por comparación con el código amablemente proporcionado por el Dr. Dietmar Bauer¹. El método CCA2 se programó conforme a la descripción del mismo realizada en esta tesis doctoral. Para la estimación del vector de cointegración en caso de dos variables, los métodos de mínimos cuadrados y componentes principales se programaron basándose en Gonzalo (1994), y el método de Aoki se programó basándose en Aoki y Havenner (1991).

¹ En las simulaciones de los métodos CCA y ACCA se utilizó finalmente el código proporcionado por el Dr. Bauer, coautor del algoritmo ACCA.

6.8 Pruebas iniciales

Para comprobar el funcionamiento de los métodos programados y contrastar resultados de investigadores anteriores tomamos un proceso generador de datos (PGD) de Gonzalo (1994), quien cita a su vez a varios autores (Engel y Granger 1987, Hansen y Philips 1990) que lo han empleado con anterioridad. Llamaremos a este proceso generador de datos PGD0, y lo utilizaremos para contrastar los resultados obtenidos por Gonzalo (1994) con los obtenidos en nuestras simulaciones, por lo que en este caso nos limitaremos a estudiar la calidad de estimación del subespacio de cointegración.

El proceso generador de datos PGD0 considerado es

$$\begin{aligned} y_{1,t} - \beta y_{2,t} &= u_t & u_t &= \rho u_{t-1} + e_{1,t} \\ a_1 y_{1,t} - a_2 y_{2,t} &= w_t & w_t &= w_{t-1} + e_{2,t} \end{aligned} \quad \begin{pmatrix} e_{1,t} \\ e_{2,t} \end{pmatrix} \equiv \text{iid N} \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \theta\sigma \\ \theta\sigma & \sigma^2 \end{pmatrix} \right]$$

La variable intermedia u_t es un proceso autorregresivo AR(1) de parámetro ρ , estacionario si $|\rho| < 1$. La variable intermedia w_t es un camino aleatorio.

El proceso PGD0 es posiblemente más fácil de interpretar en una formulación (reducida) ciclo-tendencia. Si consideramos las dos primeras ecuaciones, que en forma matricial son del tipo $\mathbf{F}_0 (y_{1,t}, y_{2,t})' = (u_t, w_t)'$, y supuesto que $\det(\mathbf{F}_0) \neq 0$ (en nuestro caso, $a_1 \beta - a_2 \neq 0$), podemos despejar $(y_{1,t}, y_{2,t})' = \mathbf{F}_0^{-1} (u_t, w_t)'$. Es decir, podemos expresar las variables observadas $(y_{1,t}, y_{2,t})$ en función de las variables intermedias (u_t, w_t) , obteniendo

$$y_{1,t} = \frac{1}{a_1 \beta - a_2} (\beta w_t - a_2 u_t) = \beta \tau_t - a_2 \eta_t$$

$$y_{2,t} = \frac{1}{a_1 \beta - a_2} (w_t - a_1 u_t) = \tau_t - a_1 \eta_t$$

donde $\tau_t = \frac{1}{a_1 \beta - a_2} w_t$ es un camino aleatorio (tendencia estocástica) y $\eta_t = \frac{1}{a_1 \beta - a_2} u_t$

es un proceso autorregresivo AR(1) de parámetro ρ (estacionario, supuesto $|\rho| < 1$). Las

dos series observadas en PGD0 son entonces combinación de un camino aleatorio común y un ciclo estacionario común.

El parámetro σ^2 es la varianza de la perturbación responsable de generar el camino aleatorio. Por otro lado, la varianza de la perturbación que genera el ciclo estacionario es la unidad, por lo que σ^2 se puede interpretar como un cociente señal-ruido: cuanto mayor es σ^2 , mayor es el peso de la tendencia respecto al ciclo, por lo que la relación de cointegración debe aparecer de forma más clara.

Realizaciones del proceso PGD0 (sin término constante en la relación de cointegración)

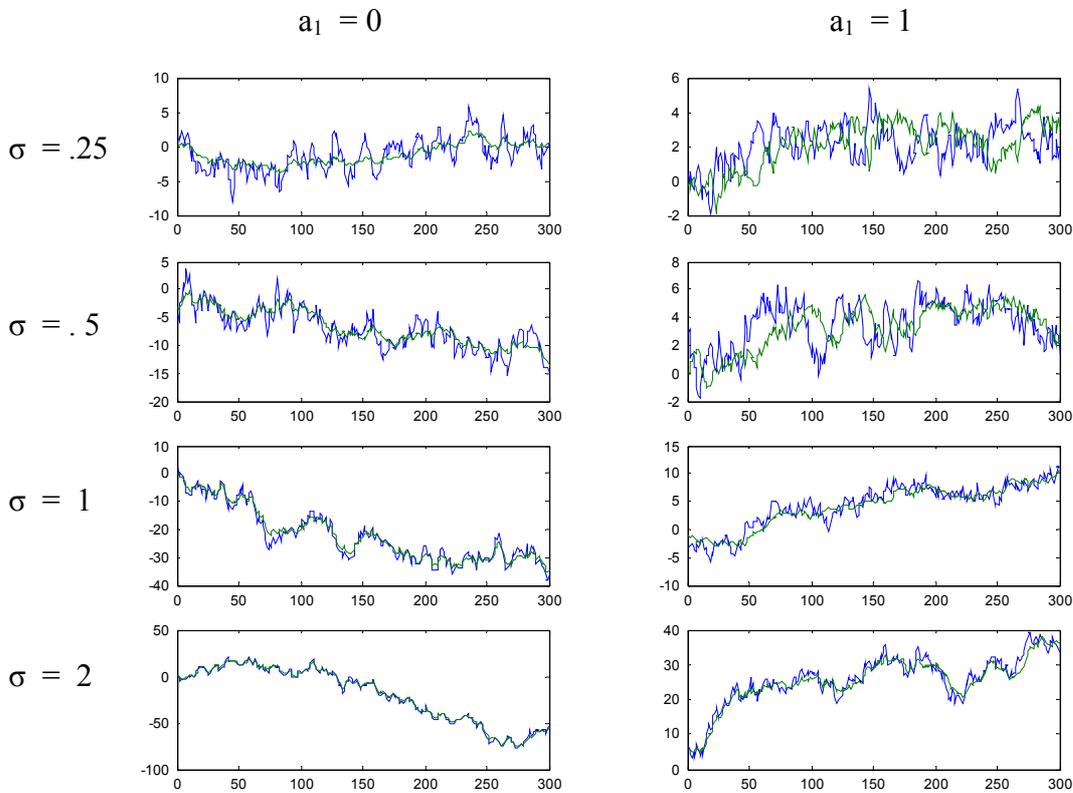


Figura 6-3. Realizaciones del proceso PGD0 sin término constante en la relación de cointegración, para distintas combinaciones de valores de los parámetros. Este es uno de los procesos estudiados por Gonzalo (1994).

La Figura 6-3 representa distintas realizaciones del proceso PGD0 correspondientes a distintas combinaciones de valores de los parámetros. En este primer caso no se ha incluido término constante en la relación de cointegración para poder comparar con los

resultados de Gonzalo (1994). El espacio paramétrico considerado en este caso es el mismo que el que utiliza Gonzalo (1994) en sus resultados: $(a_1 \times a_2 \times \beta \times \rho \times \sigma \times \theta \times T)$ con $a_1 = (0, 1)$, $a_2 = -1$, $\beta = 1$, $\rho = 0.8$, $\sigma = (0.25, 0.5, 1, 2)$, $\theta = (-0.5)$ y $T = (100, 300)$.

Estimación de la relación de cointegración

Los métodos de estimación utilizados son:

- mínimos cuadrados (MCO),
- componentes principales de la matriz de correlaciones muestral $\hat{\Gamma}_0$ (CP),
- componentes principales de la matriz de correlaciones muestral $\hat{\Gamma}_1$ (Aoki),
- correlaciones canónicas (CCA)
- máxima verosimilitud (ML).

En el caso de máxima verosimilitud se consideran distintos órdenes para la estimación de la serie observada y_t en forma de vector autorregresivo: orden 1 y orden 5 (que corresponden a autorregresivos en diferencias de órdenes cero y cuatro, por lo que nos referiremos a los estimadores como ML(0) y ML(4) respectivamente. Nótese que el proceso generador se puede formular como un VAR(1).

Los resultados obtenidos para la distribución del parámetro β (relación de cointegración) estimado por los distintos procedimientos y utilizando 5.000 repeticiones son muy similares a los proporcionados por Gonzalo (1994) utilizando 500 repeticiones.

A modo de ejemplo se presenta la tabla 6.1, correspondiente a la combinación de valores $T = 300$ y $a_1 = 1$. Las columnas que añaden al nombre del estimador el sufijo – JG corresponden a los resultados obtenidos por Gonzalo (1994). Los estadísticos seleccionados en la tabla Tabla 6-1 para comparar los distintos procedimientos de estimación de beta son el sesgo de la media, el sesgo de la mediana, el rango entre el primer y tercer cuartil (rango intercuartílico o IQR), la desviación típica (SD) del estimador y la probabilidad de que la distancia del valor estimado al real sea menor de 0,05.

Análisis de Sistemas Cointegrados mediante Métodos de Subespacios

		<i>Método</i>										
<i>Datos</i>	<i>Sigma</i>	<i>MCO-JG</i>	<i>MCO</i>	<i>CP-JG</i>	<i>CP</i>	<i>Aoki</i>	<i>CC-JG</i>	<i>CCA</i>	<i>ML(0)-JG</i>	<i>ML(0)</i>	<i>ML(4)-JG</i>	<i>ML(4)</i>
<i>Sesgo media</i>	0.25	-0.701	-0.694	-0.549	-0.520	-0.493	-0.140	0.017	0.296	0.046	-0.174	0.034
	0.5	-0.320	-0.311	-0.140	-0.129	-0.088	-0.010	-0.007	0.009	0.010	0.014	0.011
	1	-0.125	-0.121	-0.057	-0.054	-0.038	-0.008	-0.007	0.002	0.002	0.004	0.002
	2	-0.048	-0.045	-0.028	-0.026	-0.018	-0.005	-0.004	0.000	0.000	0.001	0.000
<i>Sesgo mediana</i>	0.25	-0.632	-0.650	-0.282	-0.263	-0.200	-0.058	-0.046	-0.005	-0.005	-0.003	-0.002
	0.5	-0.252	-0.250	-0.092	-0.083	-0.063	-0.029	-0.023	-0.003	0.000	-0.001	0.000
	1	-0.094	-0.088	-0.041	-0.040	-0.030	-0.014	-0.012	-0.001	0.000	-0.001	0.000
	2	-0.034	-0.032	-0.020	-0.020	-0.015	-0.007	-0.006	-0.001	0.000	0.000	0.000
<i>IQR</i>	0.25	0.600	0.585	0.890	0.830	0.549	0.268	0.249	0.275	0.252	0.281	0.266
	0.5	0.334	0.315	0.185	0.167	0.151	0.139	0.133	0.137	0.132	0.140	0.135
	1	0.139	0.126	0.078	0.078	0.071	0.070	0.069	0.069	0.065	0.070	0.067
	2	0.055	0.049	0.039	0.037	0.034	0.035	0.033	0.034	0.032	0.035	0.033
<i>SD</i>	0.25	0.378	0.369	0.757	1.766	5.102	3.076	0.459	5.908	0.399	5.006	1.013
	0.5	0.239	0.231	0.241	0.362	0.207	0.171	0.156	0.162	0.148	0.176	0.154
	1	0.114	0.109	0.080	0.076	0.071	0.075	0.070	0.072	0.068	0.078	0.072
	2	0.050	0.047	0.038	0.036	0.034	0.036	0.035	0.035	0.034	0.038	0.036
<i>P(dist. < 0.05)</i>	0.25	1%	0%	10%	10%	12%	18%	20%	19%	22%	20%	22%
	0.5	6%	6%	26%	28%	32%	38%	37%	37%	40%	37%	39%
	1	29%	29%	51%	53%	58%	62%	63%	64%	66%	65%	65%
	2	64%	67%	78%	81%	85%	88%	88%	89%	89%	87%	88%

Tabla 6-1. Resultados de la distribución del estimador $\hat{\beta}$. Con $\beta = 3$, $T = 300$ y $a_1 = 1$. Para distintos procedimientos de estimación y valores de σ . En negrita los mejores valores de cada fila.

La comparación puede visualizarse mejor en la Figura 6-4, que muestra el sesgo en la mediana que corresponde a cada uno de los estimadores, para distintas combinaciones de valores de los parámetros del proceso generador. Puede observarse que, en sesgo de la mediana, los estimadores de correlaciones canónicas y, especialmente, máxima verosimilitud, son muy superiores a las otras alternativas.

Diseño de los Experimentos

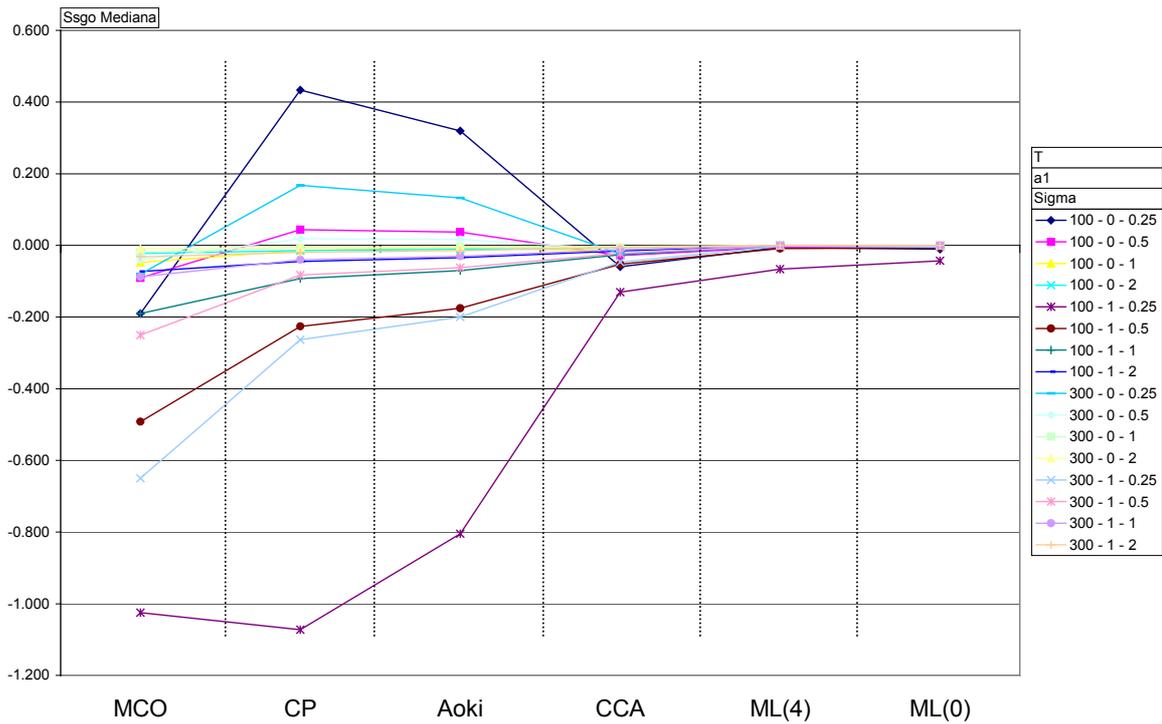


Figura 6-4. Sesgo de la mediana de las distribuciones del valor $\hat{\beta}$, estimado por los procedimientos MCO, CP, Aoki, CCA, ML(0) y ML(4), para distintas combinaciones de valores de los parámetros del proceso generador PGD0.

Si representamos el rango intercuartílico IQR (Figura 6-5), observamos que los métodos CCA y ML resultan similares (algo más concentrado el primero), y los métodos de componentes principales y Aoki presentan un problema de dispersión muy acusado para valores bajos de la señal-ruido.

Por último, si representamos la probabilidad de que el valor estimado $\hat{\beta}$ se encuentre a una distancia del verdadero β menor que una cierta cantidad de referencia (0,05), obtenemos la Figura 6-6, donde se aprecia de nuevo que, en general, los métodos CCA y ML (incluso en un modelo sobreespecificado) son los que mejores resultados proporcionan.

Análisis de Sistemas Cointegrados mediante Métodos de Subespacios

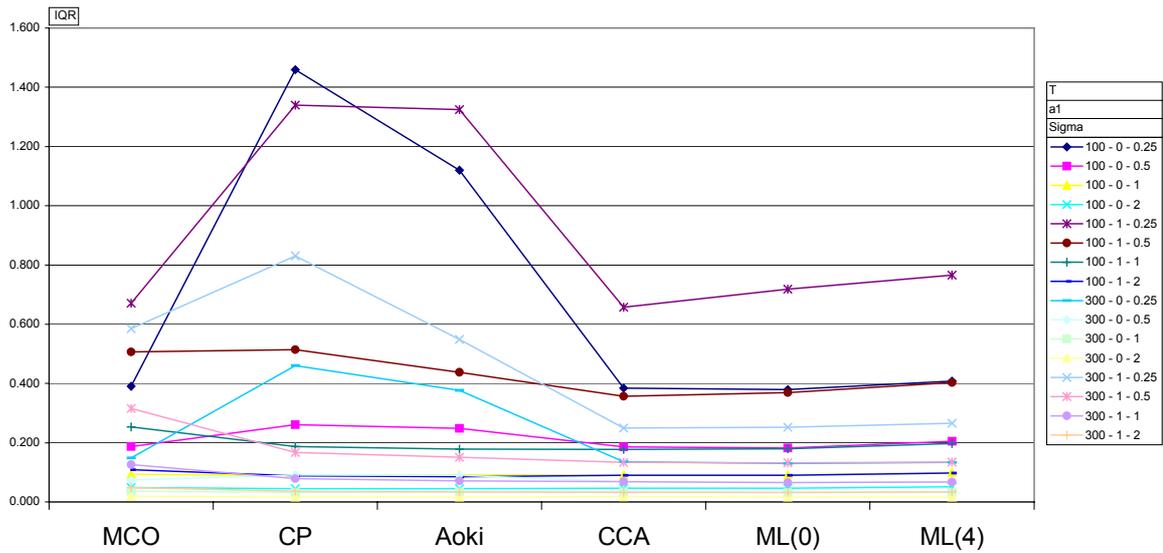


Figura 6-5. Rango intercuartílico (IQR) de las distribuciones del valor $\hat{\beta}$, estimado por los procedimientos MCO, CP, Aoki, CCA, ML(0) y ML(4), para distintas combinaciones de valores de los parámetros del proceso generador PGD0, sin término constante en la relación de cointegración.

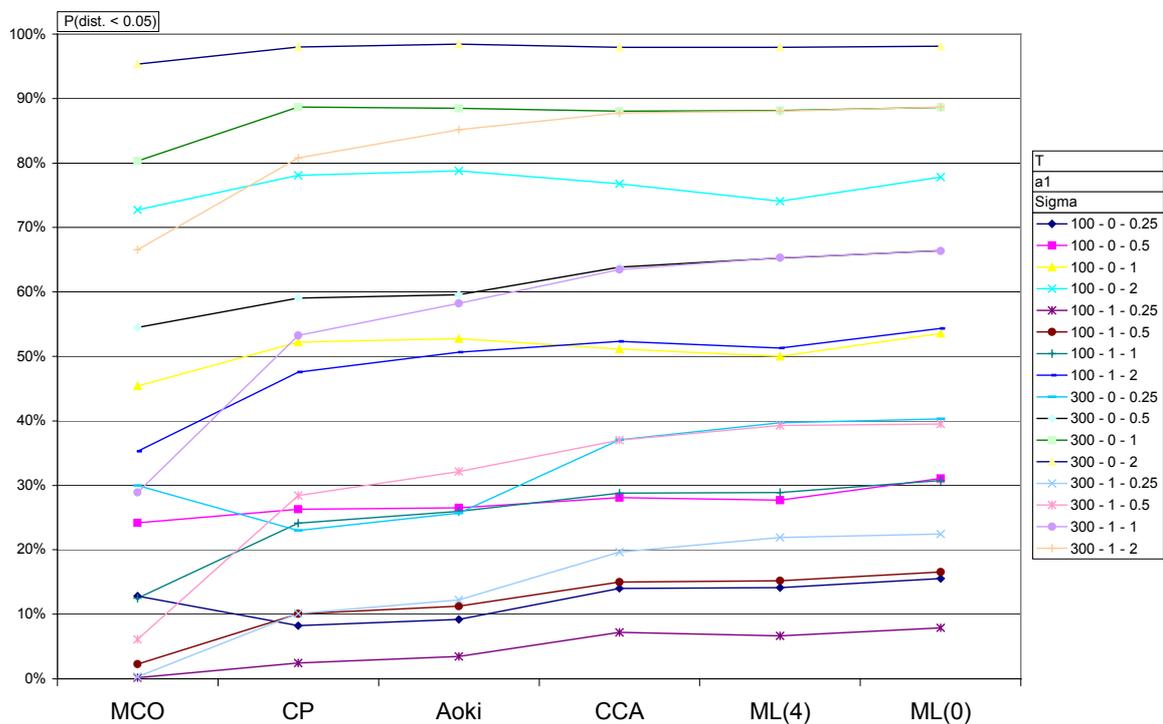


Figura 6-6. $P(|\hat{\beta} - \beta| < 0,05)$, para beta estimado por los procedimientos MCO, CP, Aoki, CCA, ML(0) y ML(4), para distintas combinaciones de valores de los parámetros del proceso generador PGD0.

7 RESULTADOS

En este capítulo se presenta una serie seleccionada de contrastes, tablas y gráficas para ilustrar los resultados más significativos de las simulaciones realizadas, conforme a las condiciones indicadas en el capítulo anterior.

Se considerarán cuatro diferentes procesos generadores de datos, todos ellos formulados en el espacio de estados como combinación de ciclos estacionarios y tendencias no estacionarias $I(1)$. El proceso PGD1 genera dos series temporales cointegradas y es similar al proceso PGD0 que vimos en el capítulo anterior (tomado de Gonzalo 1994), pero introduce término constante en la relación de cointegración. Para este primer proceso PGD1 nos limitaremos a mostrar los resultados de la estimación del vector de cointegración, lo que nos llevará a prescindir del algoritmo SSATS a favor de los algoritmos de correlaciones canónicas.

El proceso PGD2 añade al proceso PGD1 términos de media móvil, con lo que esperamos dificultar los resultados de una aproximación VAR (método de Johansen). Estudiaremos la estimación del vector de cointegración y la capacidad predictiva de los modelos identificados.

Los procesos PGD3 y PGD4 generan tres series temporales cointegradas. El PGD3 genera una única tendencia común (dos relaciones de cointegración) y el proceso PGD4 genera dos tendencias comunes (una relación de cointegración).

Dada la variedad de procesos generadores de datos considerados y la gran cantidad de combinaciones de valores de parámetros dentro de un mismo proceso generador, sólo se mostrarán algunos resultados representativos. Salvo indicación expresa, los gráficos y tablas corresponden a simulaciones realizadas utilizando perturbaciones gaussianas.

7.1 Proceso generador de datos 1 (PGD1)

El proceso generador de datos 0 (PGD0) se introdujo en el capítulo anterior para contrastar nuestros resultados con los de Gonzalo (1994). Buscando mayor generalidad en nuestro estudio, suprimimos a continuación una de las restricciones del PGD0 de Gonzalo (1994) e introducimos un término constante en la relación de cointegración, dando lugar al proceso PGD1. Esto supone un desplazamiento en vertical de todos los valores de una de las series con respecto a la otra, como puede verse en la Figura 7-1.

Realizaciones del proceso PGD1, con término constante en la relación de cointegración

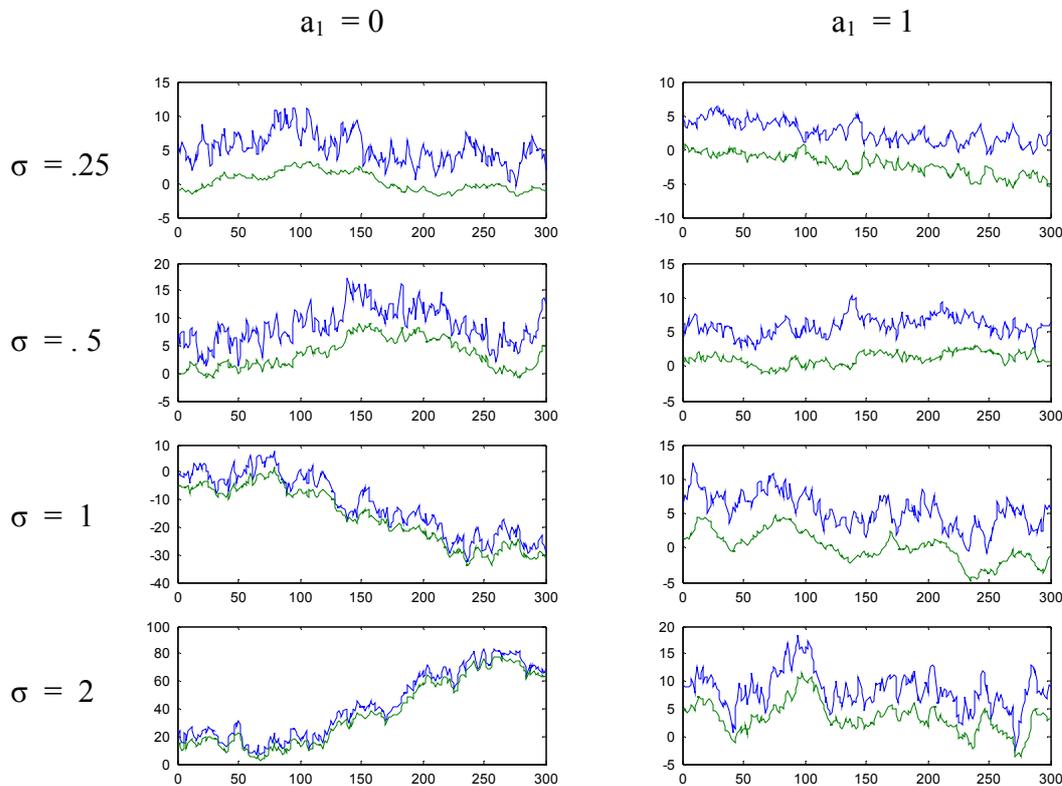


Figura 7-1. Realizaciones del proceso PGD1 con término constante en la relación de cointegración. Para distintas combinaciones de valores de los parámetros.

Otra de las diferencias introducidas es que la medida de separación utilizada ahora será el ángulo formado por el vector de cointegración estimado con el vector de cointegración real. Como ya indicamos, esta medida permite generalización a más de

Resultados

dos dimensiones y no se ve sujeta a variación por normalización arbitraria de alguna de las componentes. En cualquier caso, repitiendo el experimento de Gonzalo pero utilizando el ángulo como medida de separación entre los subespacios de cointegración real y estimado, se obtienen, como era de esperar, las mismas conclusiones respecto a la calidad de los estimadores.

La presencia de términos constantes en la relación de cointegración permite estudiar un caso de mayor generalidad que el de Gonzalo (1994), pero obliga a modificar los estimadores del vector de cointegración para mantener la consistencia.

Nuestro objetivo ahora es comparar los mismos procedimientos de estimación anteriores (ver apartado 6.8), pero en el caso en que existe término constante en la relación de cointegración. Para eliminar el efecto del término constante trabajaremos en diferencias respecto a las medias. En el caso de máxima verosimilitud se han comparado dos procedimientos de estimación para el vector de cointegración: el procedimiento que corresponde a utilizar los datos sin restar las medias (y estima también el término constante de la relación de cointegración), y el procedimiento que resta las medias y supone término constante nulo. Si al restar las medias no perdemos información sobre la relación de cointegración, esperamos que los dos estimadores sean equivalentes. Para distinguirlos, al estimador por máxima verosimilitud (método de Johansen) que no resta las medias y que estima el término constante le denominamos “ML_b”.

Nótese que, si se restan las medias muestrales, la magnitud del valor de los desplazamientos (términos constantes) no influye en los resultados.

La Tabla 7-1 es representativa del tipo de resultados obtenidos, que se complementan con las figuras Figura 7-2 a Figura 7-5. Podemos extraer las siguientes conclusiones:

- En sesgo de la mediana, los estimadores CCA y ML superan en general a MCO, CP o Aoki. Esta superioridad es muy acusada para bajos valores de la señal-ruido, pero desaparece para valores altos.
- Los estimadores de ML obtenidos tras restar las medias presentan un sesgo de la mediana algo más reducido que los estimadores de ML obtenidos sin restar las medias (estos últimos obtenidos por el método de Johansen adaptado para estimar también el término constante de la relación de cointegración). Salvo por este detalle, las distribuciones del vector de cointegración son prácticamente idénticas restando la media y sin restarla (procedimiento de Johansen adaptado).

Análisis de Sistemas Cointegrados mediante Métodos de Subespacios

- Los estimadores MCO y, especialmente, CP y Aoki, presentan un problema de dispersión acusada para bajos valores de la señal-ruido. Sólo en algunas ocasiones, para valores elevados de la señal-ruido (relación de cointegración acusada), llegan a igualar o superar la concentración de los estimadores ML en torno al valor real del vector de cointegración.

		<i>Método</i>								
<i>Estadístico</i>	<i>Sigma</i>	<i>MCO</i>	<i>CP</i>	<i>Aoki</i>	<i>CCA</i>	<i>ACCA</i>	<i>ML0</i>	<i>ML_b0</i>	<i>ML4</i>	<i>ML_b4</i>
<i>Media</i>	0.25	-0.604	0.081	0.068	0.005	0.049	-0.003	-0.004	-0.004	-0.006
	0.5	-0.186	-0.005	0.003	0.003	-0.004	-0.001	-0.001	0.000	-0.001
	1	-0.053	-0.009	-0.004	0.002	-0.009	0.000	0.000	0.000	0.000
	2	-0.018	-0.007	-0.004	0.001	-0.007	0.000	0.000	0.000	0.000
<i>Mediana</i>	0.25	-0.545	0.049	0.046	0.005	0.037	-0.001	-0.001	-0.001	-0.002
	0.5	-0.150	-0.004	0.002	0.002	-0.004	-0.001	-0.001	0.000	-0.001
	1	-0.042	-0.007	-0.003	0.001	-0.007	0.000	-0.001	0.000	-0.001
	2	-0.014	-0.005	-0.003	0.001	-0.005	0.000	0.000	0.000	0.000
<i>IQR</i>	0.25	0.561	0.132	0.123	0.094	0.115	0.092	0.091	0.093	0.094
	0.5	0.168	0.047	0.046	0.046	0.047	0.045	0.046	0.048	0.048
	1	0.049	0.024	0.023	0.024	0.024	0.023	0.023	0.024	0.024
	2	0.019	0.012	0.012	0.012	0.013	0.011	0.011	0.012	0.012
<i>SD</i>	0.25	0.379	0.156	0.133	0.084	0.108	0.083	0.083	0.090	0.091
	0.5	0.139	0.043	0.043	0.043	0.043	0.042	0.042	0.046	0.046
	1	0.041	0.020	0.020	0.022	0.020	0.021	0.021	0.023	0.024
	2	0.016	0.010	0.010	0.011	0.011	0.010	0.010	0.011	0.011
<i>P($\alpha < 0.05$)</i>	0.25	0%	41%	42%	52%	43%	53%	53%	53%	53%
	0.5	12%	80%	81%	80%	80%	82%	82%	79%	79%
	1	58%	96%	97%	97%	96%	97%	97%	96%	96%
	2	96%	100%	100%	100%	100%	100%	100%	100%	100%

Tabla 7-1. Resultados de la distribución del ángulo formado entre los vectores de cointegración estimado y real. 5.000 simulaciones, para distintos procedimientos de estimación y valores de sigma, con $\beta = 3$, $T = 300$ y $a_1 = 1$. Todos los estimadores restan las medias salvo $ML_b(0)$ y $ML_b(4)$. En *negrita* los mejores valores de cada fila.

Resultados

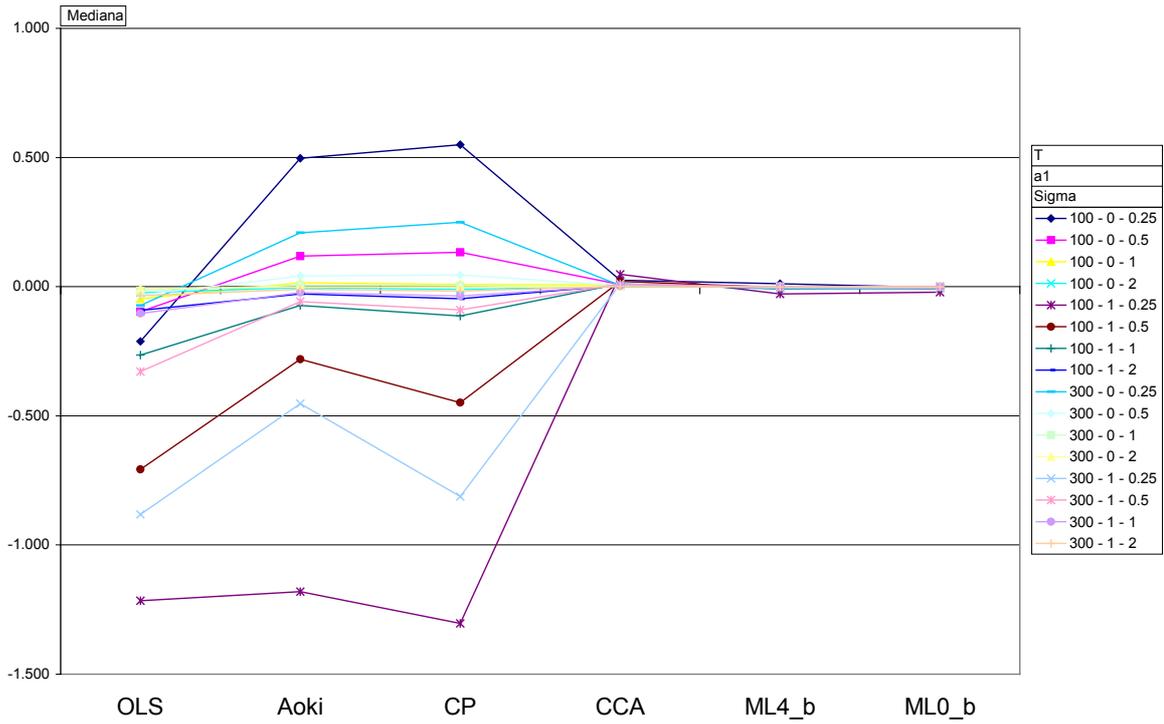


Figura 7-2. Mediana de la distribución del ángulo formado entre los vectores de cointegración estimado y real. Procedimientos de estimación MCO, CP, Aoki, CCA, ML_b(0) y ML_b(4). Para distintas combinaciones de valores de los parámetros del proceso generador PGD1, con $\beta = 1$.

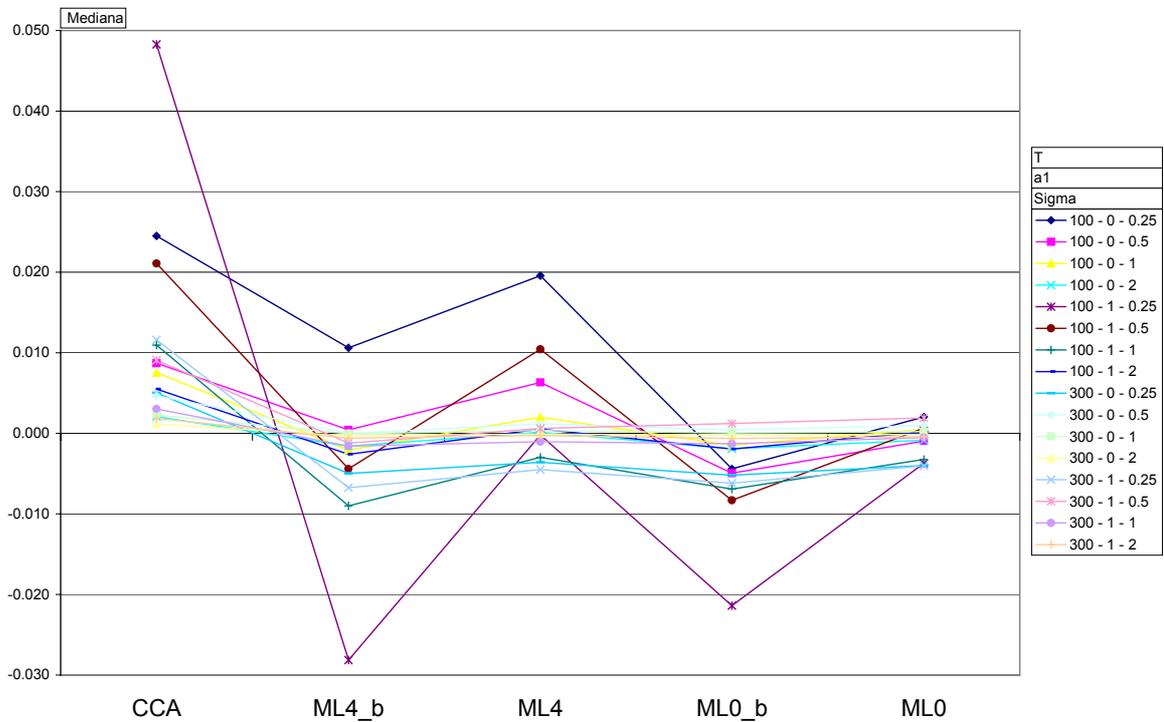


Figura 7-3. Mediana de la distribución del ángulo formado entre los vectores de cointegración estimado y real. Procedimientos de estimación CCA, ML(0), ML_b(0), ML(4) y ML_b(4). Para distintas combinaciones de valores de los parámetros del proceso generador PGD1, con $\beta = 1$.

Análisis de Sistemas Cointegrados mediante Métodos de Subespacios

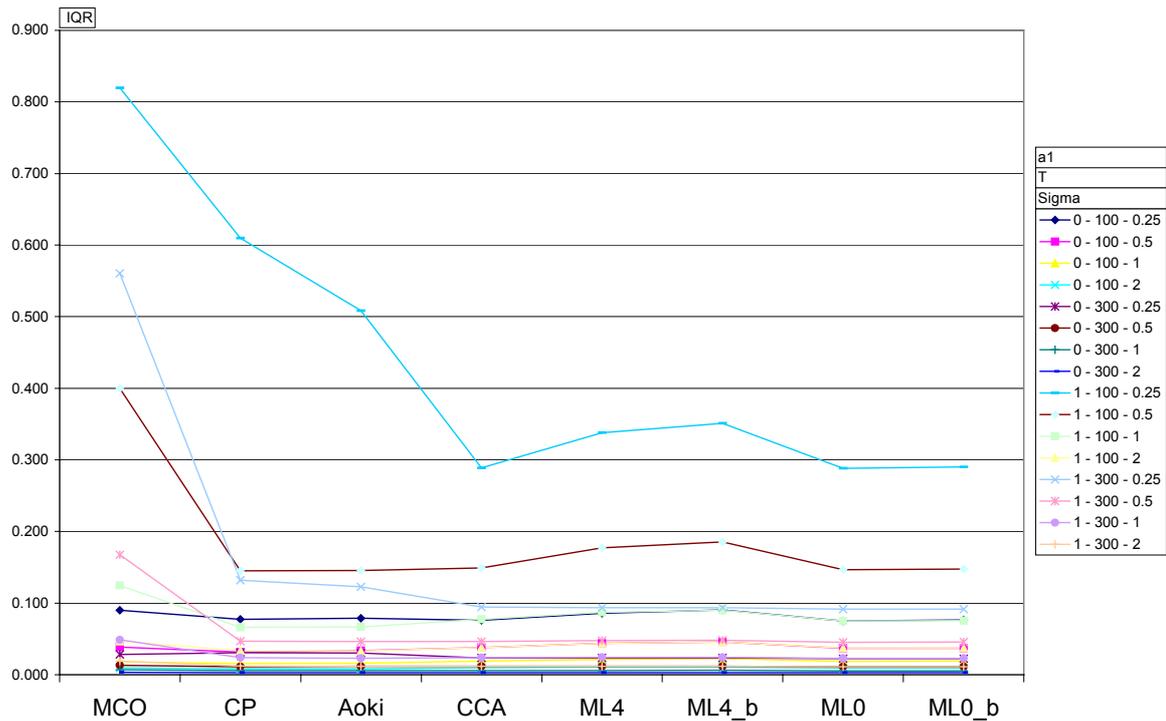


Figura 7-4. Rango intercuartílico (IQR) de la distribución del ángulo formado entre los vectores de cointegración estimado y real. Procedimientos de estimación MCO, CP, Aoki, CCA, ML_b(0) y ML_b(4). Distintas combinaciones de valores de los parámetros del proceso generador PGD1 con $\beta = 3$.

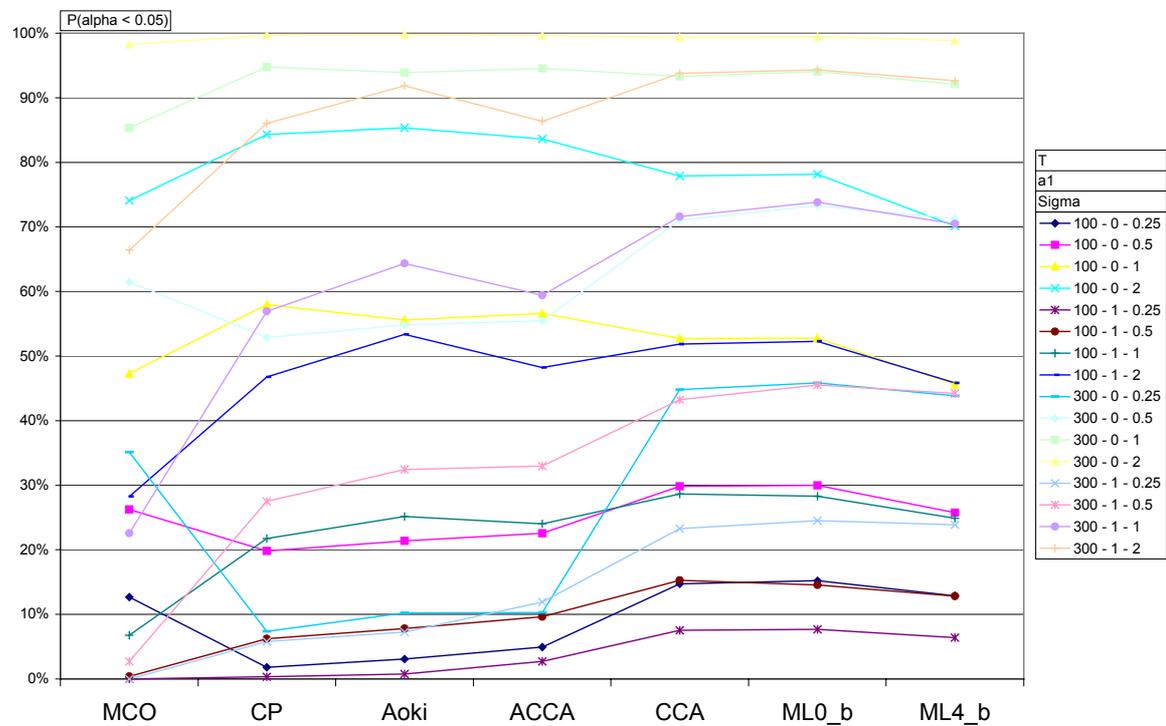


Figura 7-5. $P(|\alpha| < 0,05)$, siendo α el ángulo formado entre los vectores de cointegración estimado y real. Procedimientos de estimación MCO, CP, Aoki, CCA, ML_b(0) y ML_b(0). Para distintas combinaciones de valores de los parámetros del proceso generador PGD1, con $\beta = 1$.

Resultados

A la vista de estos resultados, prescindimos en adelante de los métodos menos precisos (MCO, CP, Aoki) y nos centramos en los estimadores ML (obtenidos por el método de Johansen) y en los estimadores de correlaciones canónicas. Introducimos los procedimientos de correlaciones canónicas adaptado (ACCA) de Bauer y Wagner (2002) y el procedimiento CCA2.

En las simulaciones de este proceso generador PGD1, los estimadores de correlaciones canónicas se han calculado tomando los valores de los parámetros $p = f = n = 2$. La diferencia entre el método CCA y el CCA2 al estimar el vector de cointegración es que el segundo trabaja con un número impuesto de estados igual al número de tendencias comunes, y el primero trabaja con un número estimado de estados total (estacionarios y no estacionarios). Para el proceso PGD1 las estimaciones por el método CCA2 coinciden prácticamente con las estimaciones obtenidas por el método CCA simple, por lo que no se han representado. Los resultados pueden observarse en la Figura 7-6 y en la Figura 7-7.

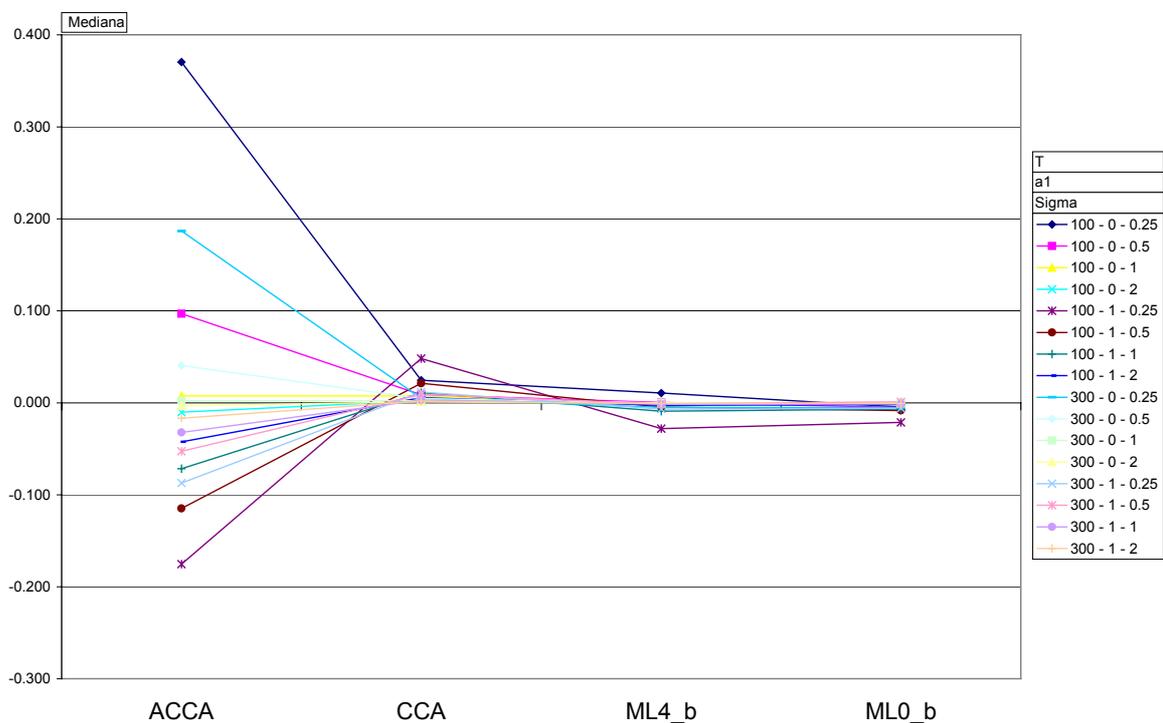


Figura 7-6. Mediana de la distribución del ángulo formado entre los vectores de cointegración estimado y real. Procedimientos de estimación ACCA, CCA, $ML_b(0)$ y $ML_b(4)$. Para distintas combinaciones de valores de los parámetros del proceso generador PGD1, con $\beta=1$ y con término constante en la relación de cointegración.

Análisis de Sistemas Cointegrados mediante Métodos de Subespacios

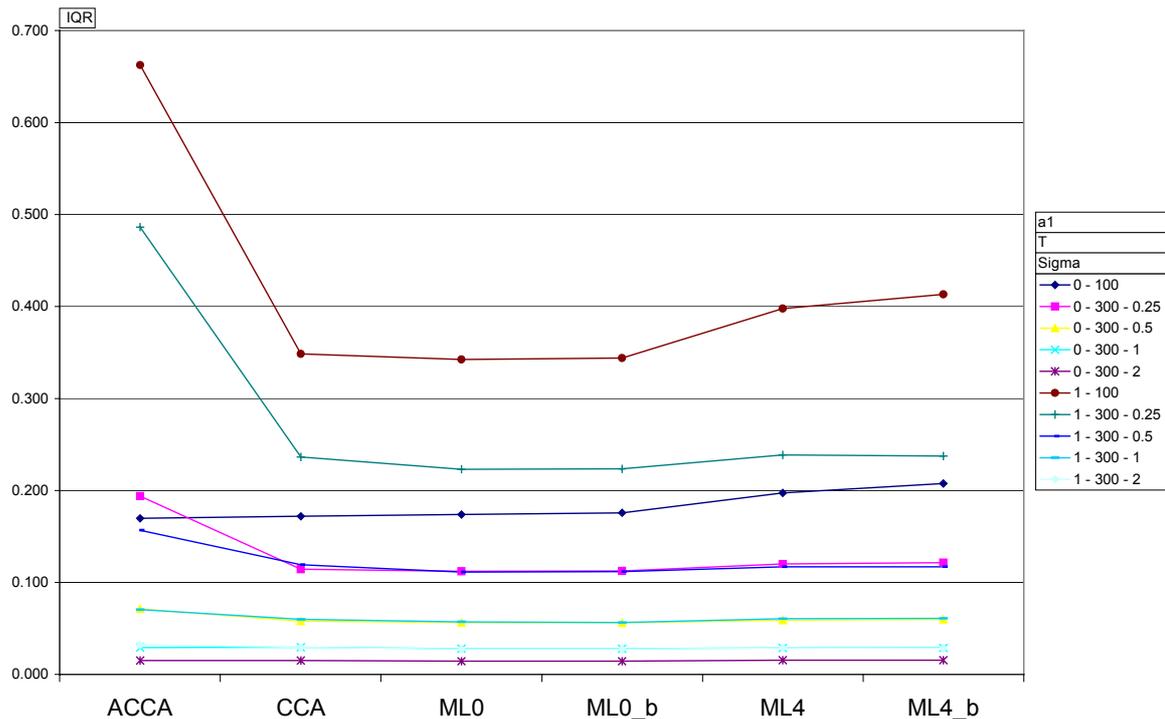


Figura 7-7. Rango intercuartílico (IQR) de la distribución del ángulo formado entre los vectores de cointegración estimado y real. Procedimientos de estimación ACCA, CCA, ML₍₀₎, ML_{b(0)}, ML₍₄₎ y ML_{b(4)}. Para distintas combinaciones de valores de los parámetros del proceso generador PGD1, con $\beta=1$ y con término constante en la relación de cointegración.

Se comprueba que:

- La distribución del ángulo que forman los vectores de cointegración real y estimado presenta medianas notablemente más sesgadas (alejadas del origen) si se usa el procedimiento de estimación ACCA que si se usa CCA simple o ML. También la dispersión del estimador es en general mayor (o mucho mayor) cuando se utiliza el procedimiento ACCA.
- La estimación del vector de cointegración por ML restando las medias y sin restarlas (ML_b) es prácticamente la misma.

Podemos concluir que, para el proceso generador PGD1 con los valores estudiados, el algoritmo ACCA produce peores estimaciones del subespacio de cointegración que el algoritmo CCA simple o el algoritmo de máxima verosimilitud de Johansen.

Los resultados anteriores han sido contrastados observando las distribuciones empíricas (histogramas) del ángulo α formado entre el vector de cointegración real y observado,

Resultados

para cada uno de los procedimientos de estimación estudiados y para las distintas combinaciones de valores de los parámetros del proceso generador.

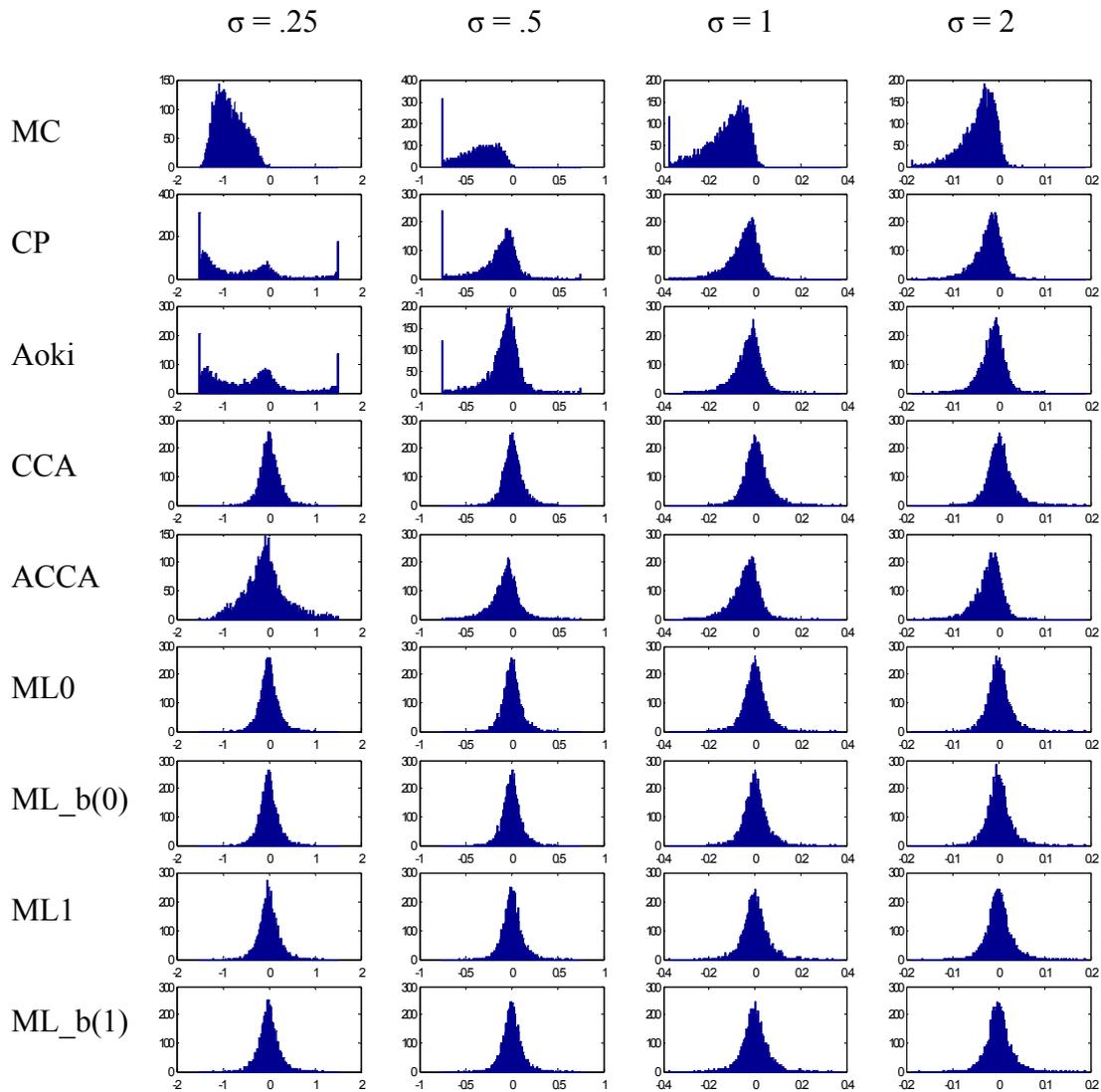


Figura 7-8. Distribuciones empíricas (histogramas) del ángulo α formado entre los vectores de cointegración estimado y real. Procedimientos de estimación MCO, CP, Aoki, CCA, ACCA, ML(0), ML_b(0), ML(4) y ML_b(4). Para el proceso generador PGDI con término constante en la relación de cointegración, $a=1$, $\beta=1$, $T=300$ y distintos valores del parámetro señal-ruido σ .

A modo de ejemplo se presenta la Figura 7-8. Obsérvese que, en esta representación, todos los histogramas de una misma columna tienen la misma escala en el eje horizontal (valor del ángulo α), y que esta escala aumenta al moverse por columnas hacia la

derecha, porque la distribución del ángulo entre el vector real y estimado se concentra en torno a cero al aumentar el valor de la señal-ruido σ .

7.2 Proceso generador de datos 2 (PGD2)

7.2.1 Características generales

El proceso generador de datos PGD2 es una extensión del proceso PGD1. Una formulación del proceso PGD2 en el espacio de estados es

$$\begin{bmatrix} \tau \\ \eta_1 \\ \eta_2 \end{bmatrix}_{t+1} = \begin{bmatrix} 1 & a & 0 \\ 0 & \varphi_1 & 0 \\ 0 & 0 & \varphi_2 \end{bmatrix} \begin{bmatrix} \tau \\ \eta_1 \\ \eta_2 \end{bmatrix}_t + \begin{bmatrix} 1 & 1 \\ \varphi_1 + \theta & 0 \\ 0 & \varphi_2 + \theta \end{bmatrix} \begin{bmatrix} e_1 \\ e_2 \end{bmatrix}_t$$

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix}_t = \begin{bmatrix} \beta & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{\sigma}{1+a} \tau \\ \eta_1 \\ \eta_2 \end{bmatrix}_t + \begin{bmatrix} e_1 \\ e_2 \end{bmatrix}_t$$

Puede comprobarse que, al igual que en el PGD1, las series observadas son generadas por una combinación tendencia-ciclos:

$$y_{1,t} = \beta \frac{\sigma}{1+a} \tau_t + \eta_{1,t} + e_{1,t}$$

$$y_{2,t} = \frac{\sigma}{1+a} \tau_t + \eta_{2,t} + e_{2,t}$$

Sin embargo, buscando modelos más generales que el PGD1 e introduciendo términos de media móvil (con lo que esperamos dificultar la capacidad de una formulación VAR para representar el sistema), hemos considerado dos ciclos independientes (asociados a $\eta_{1,t} + e_{1,t}$ y a $\eta_{2,t} + e_{2,t}$), que no serán autorregresivos AR(1), sino procesos más generales ARMA(1,1)¹ de la forma

$$(\eta_{1,t} + e_{1,t}) = \varphi_1 (\eta_{1,t-1} + e_{1,t-1}) + e_{1,t} + \theta e_{1,t-1}$$

¹ Por otra parte (y aunque en el PGD2 los ciclos entran independientemente en cada serie observada), una suma de procesos AR(1) genera, con carácter general, procesos de tipo ARMA(2,1). (Hamilton 1994, p. 108).

Para la tendencia común (τ_t) tenemos:

$$\tau_{t+1} = \tau_t + a \eta_{1,t} + e_{1,t}.$$

Dependiendo del valor del parámetro a (0 ó 1), la tendencia común será, o bien un camino aleatorio, o bien una serie cuyas primeras diferencias coinciden con uno de los ciclos estacionarios ($\eta_{1,t} + e_{1,t}$).

Para el vector de perturbaciones e_t se ha considerado ruido blanco gaussiano de esperanza cero y matriz de covarianzas \mathbf{I} . Para analizar el efecto causado por la falta de normalidad de las perturbaciones, también se ha considerado ruido blanco conforme a una distribución t de Student con cinco grados de libertad.

El parámetro σ es un parámetro de amplificación de la tendencia respecto al ciclo (parámetro señal-ruido para el subespacio de cointegración).

Obsérvese que la relación de cointegración se puede extraer de la primera columna de la matriz \mathbf{C} de la ecuación de observaciones, que es el vector $[\beta, 1]'$: el subespacio (vector) ortogonal al vector $[\beta, 1]'$ define la relación de cointegración.

Para incluir un término constante en la relación de cointegración, se ha sumado un valor fijo δ a todos los valores de una de las series (desplazamiento en vertical de todos los valores de una de las series con respecto a la otra).

El espacio paramétrico considerado es $(T \times \beta \times \sigma \times a \times \varphi_1 \times \varphi_2 \times \theta \times \delta)$. Eligiendo una serie de tamaños muestrales variada y representativa de series económicas $T = (50, 100, 300)$, y ajustando los valores de los otros parámetros tratando de cubrir un conjunto variado de características del proceso generador (Figura 7-9), hemos considerado los valores $\beta = 1$, $\sigma = (.2, .5)$, $a = (0, 1)$, $\varphi_1 = (.5, .9)$, $\varphi_2 = (.5, .9)$, $\theta = (.5, .9)$ y $\delta = 50$.

La Figura 7-9 representa un conjunto de realizaciones del proceso generador PGD2 que corresponde a distintas combinaciones de valores de los parámetros.

Resultados

Realizaciones PGD2

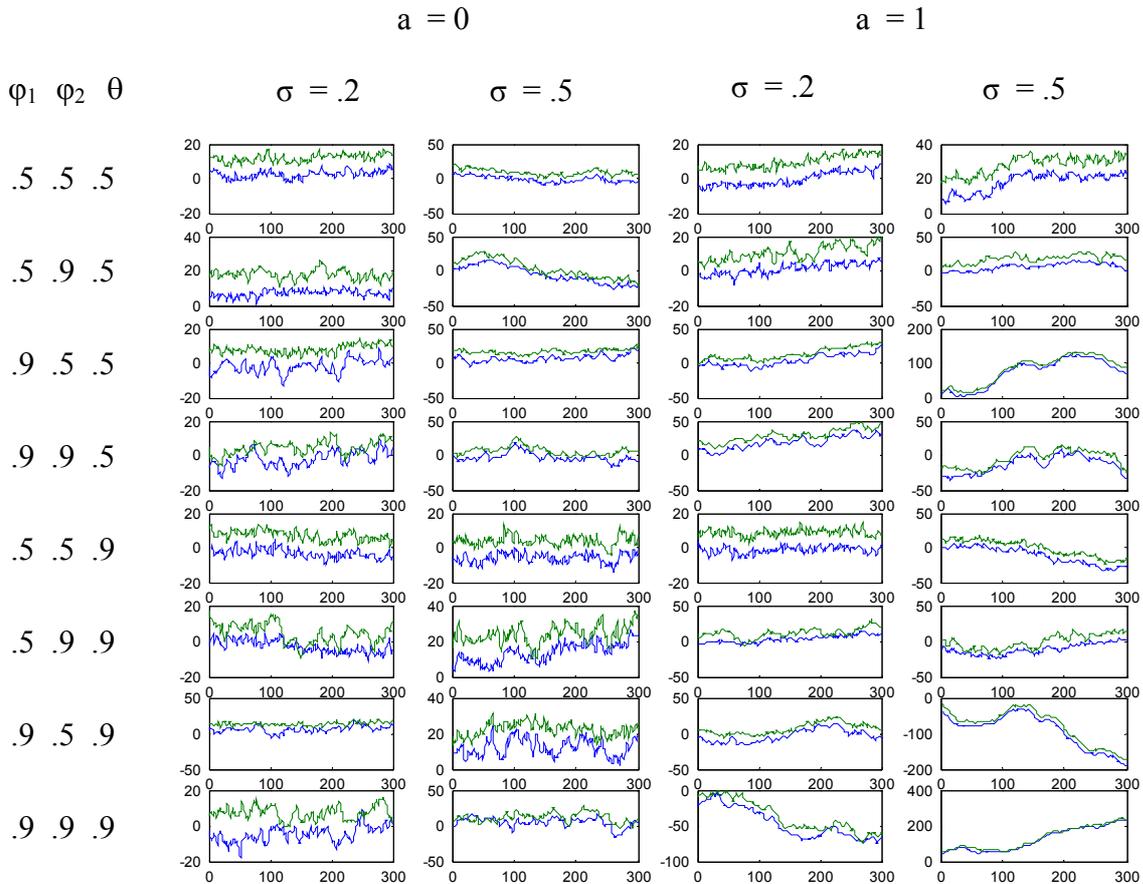


Figura 7-9. Realizaciones del proceso PGD2 con componente constante en la relación de cointegración. Para distintas combinaciones de valores de los parámetros, con desplazamiento $\delta = 10$.

Ajuste de parámetros de los algoritmos

Para todas las simulaciones que siguen, el orden p de la representación VAR del vector de series temporales \mathbf{y}_t se ha seleccionado minimizando el criterio de información de Schwarz (BIC).

Para los métodos de subespacios y para este proceso generador PGD2, se tomaron valores de los parámetros $p = f = 5$. El orden n de la representación en el espacio de estados se calculó por el criterio $BA(n)$ de Bauer y Wagner (sección 6.6).

La Figura 7-13 muestra un resultado representativo (típico) de la estimación del orden del modelo para los procedimientos CCA2 y ACCA.

Análisis de Sistemas Cointegrados mediante Métodos de Subespacios

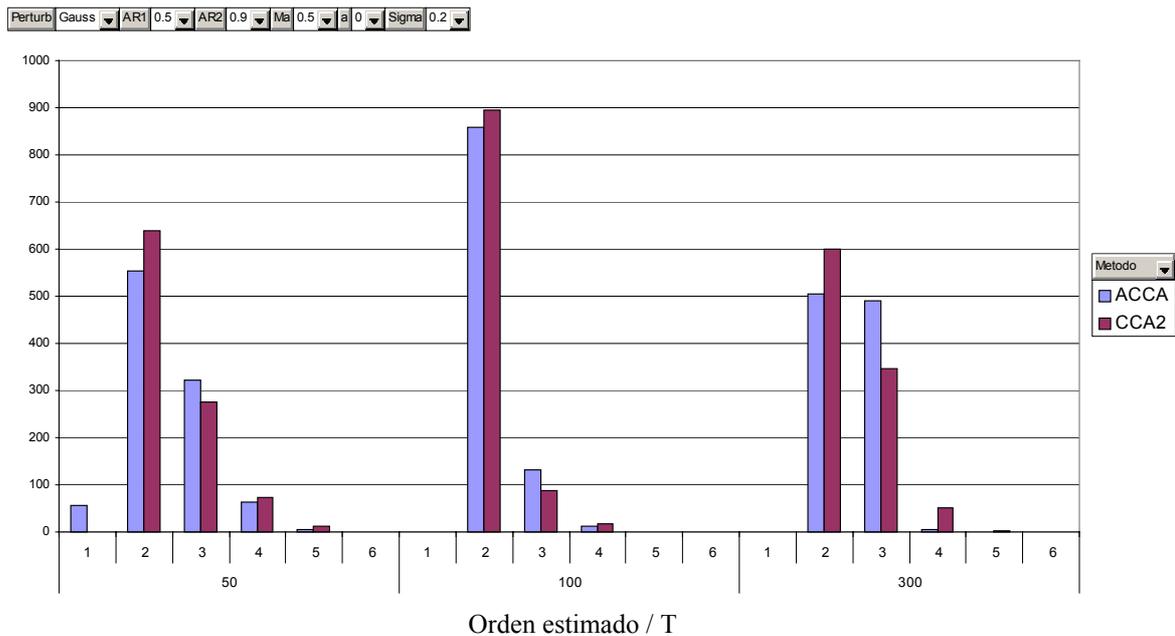


Figura 7-10. Histogramas del orden estimado (de modelos en el espacio de estados) por los procedimientos ACCA y CCA2, con el criterio $BA(n)$, para distintos tamaños muestrales T .

Se observa que para tamaños muestrales reducidos la dispersión en la estimación del orden del modelo (número de estados) es mayor. También se observa que en general se subestima el verdadero orden del proceso generador (el orden del PGD2 es 3, puesto que la formulación del PGD2 con tres estados utilizada es controlable y observable para los valores de los parámetros considerados). Al crecer el tamaño muestral el orden del modelo se estima con mayor precisión. No se observan diferencias notables entre los dos procedimientos representados.

En cuanto a la estimación del orden de una aproximación VAR por el criterio BIC, la Figura 7-11 muestra resultados representativos, correspondientes a un valor del parámetro $\theta = 0,5$. Se observa una tendencia a aumentar el orden estimado del modelo al aumentar T . También un aumento en el valor del parámetro θ provoca un aumento en el orden de la aproximación VAR (Figura 7-12).

Resultados

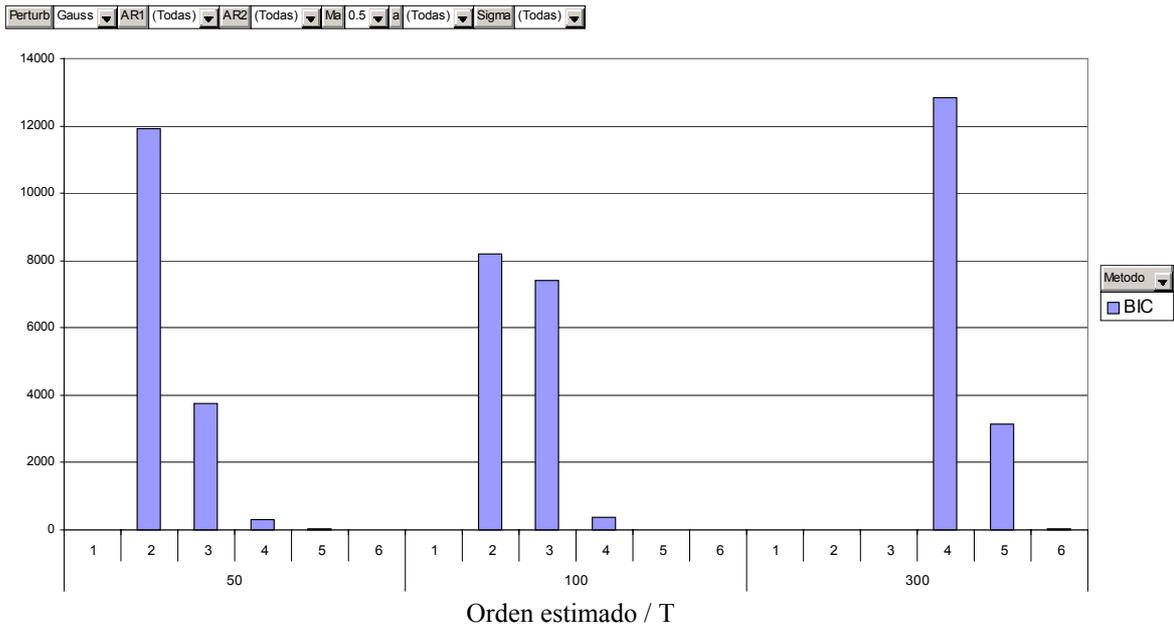


Figura 7-11. Histogramas del orden estimado (de modelos VAR) por el procedimiento de Johansen, con el criterio BIC, para distintos tamaños muestrales T , con $\theta = 0,5$.

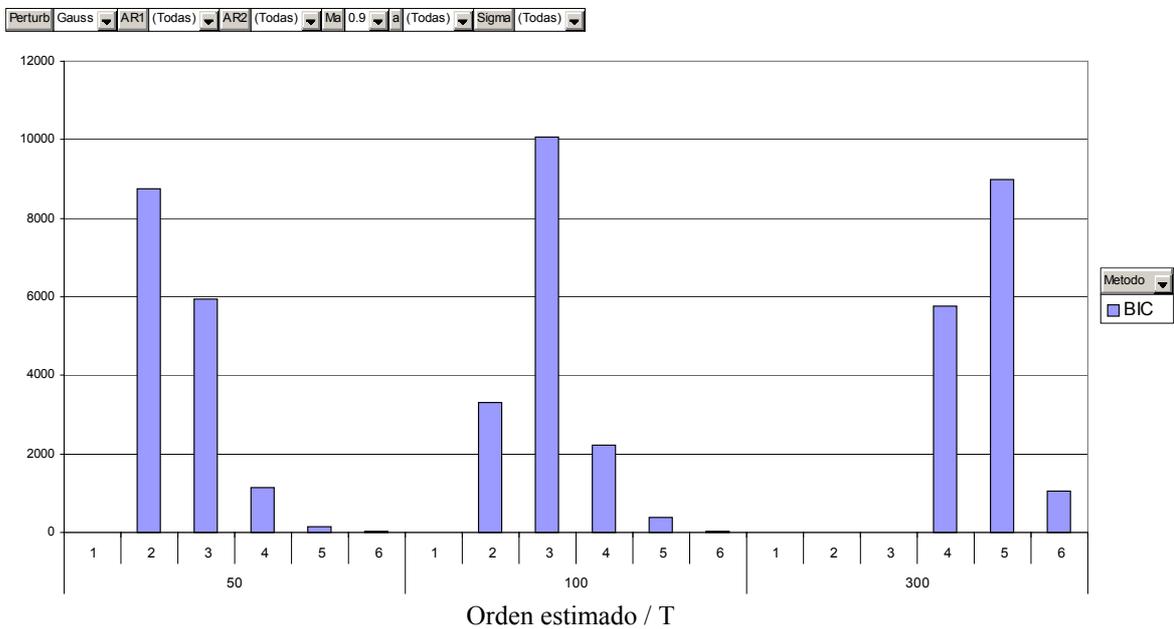


Figura 7-12. Histogramas del orden estimado (de modelos VAR) por el procedimiento de Johansen, con el criterio BIC, para distintos tamaños muestrales T , con $\theta = 0,9$.

7.2.2 Estimación del vector de Cointegración

La Tabla 7-2 presenta los valores de varios estadísticos de la distribución del ángulo que separa los vectores de cointegración real y estimado, para distintos estimadores y distintas combinaciones de valores de los parámetros del proceso generador, con perturbaciones gaussianas y manteniendo algunos de los parámetros fijos.

T	Sigma	Estadístico	Método				
			ML_b	ML	ACCA	CCA	CCA2
50	0.2	Media	-0.020	-0.018	-0.020	-0.013	-0.010
		Mediana	-0.019	-0.017	-0.013	-0.001	-0.001
		IQR	0.779	0.766	0.501	0.433	0.422
		SD	0.670	0.665	0.446	0.408	0.404
		$P(\alpha < 0.05)$	8%	8%	11%	13%	13%
	0.5	Media	0.013	0.014	-0.008	-0.008	-0.007
		Mediana	-0.006	-0.009	-0.012	-0.008	-0.010
		IQR	0.341	0.332	0.242	0.241	0.232
		SD	0.429	0.419	0.251	0.264	0.257
		$P(\alpha < 0.05)$	19%	19%	22%	25%	25%
100	0.2	Media	0.037	0.037	0.005	0.004	0.005
		Mediana	0.018	0.018	0.004	0.005	0.004
		IQR	0.383	0.383	0.280	0.234	0.234
		SD	0.430	0.428	0.238	0.208	0.207
		$P(\alpha < 0.05)$	16%	16%	21%	21%	22%
	0.5	Media	-0.001	-0.001	0.000	0.000	0.000
		Mediana	-0.002	-0.002	0.001	0.001	0.002
		IQR	0.137	0.137	0.120	0.117	0.118
		SD	0.155	0.152	0.102	0.100	0.098
		$P(\alpha < 0.05)$	37%	36%	43%	43%	44%
300	0.2	Media	-0.001	-0.001	-0.001	-0.001	-0.001
		Mediana	0.000	0.000	-0.001	0.000	-0.001
		IQR	0.113	0.113	0.097	0.090	0.091
		SD	0.131	0.131	0.085	0.075	0.074
		$P(\alpha < 0.05)$	45%	45%	51%	54%	55%
	0.5	Media	0.000	0.000	0.000	0.000	0.000
		Mediana	0.000	0.000	0.000	0.000	0.000
		IQR	0.041	0.041	0.038	0.039	0.039
		SD	0.041	0.041	0.035	0.034	0.034
		$P(\alpha < 0.05)$	83%	83%	86%	89%	89%

Tabla 7-2. Resultados de la distribución del ángulo formado entre los vectores de cointegración estimado y real. 1.000 simulaciones del proceso PGD2, para distintos procedimientos de estimación y valores de sigma y T, con $\beta = 1$, $\alpha = 0$, $\varphi_1 = .5$, $\varphi_2 = .5$, $\theta = .5$. Todos los estimadores restan las medias salvo ML_b. En negrita los mejores valores de cada fila.

Resultados

Aunque en la Tabla 7-2 el método CCA2 ofrece muy buenos resultados, no es el caso general, como se comprueba en los gráficos que van de la Figura 7-13 a la Figura 7-15.

Se observan los siguientes resultados:

- Entre sí, los estimadores ML y ML_b y los estimadores CCA y CCA2 resultan prácticamente equivalentes.
- Se observa que el sesgo de la mediana es, en general, muy inferior en la estimación ML respecto a los métodos de subespacios. Sin embargo, la dispersión de los estimadores es muy variable, y los métodos de ML ofrecen a veces una dispersión muy superior a los métodos de subespacios, dependiendo de los valores de los parámetros del proceso generador, pero sin que se detecte una relación clara con un parámetro concreto (interacciones).
- ACCA proporciona en general mejores estimaciones que CCA, aunque se producen excepciones.

Estos resultados pueden confirmarse mediante los histogramas. A modo de ejemplo se muestran algunos de estos histogramas entre la Figura 7-16 y la Figura 7-20.

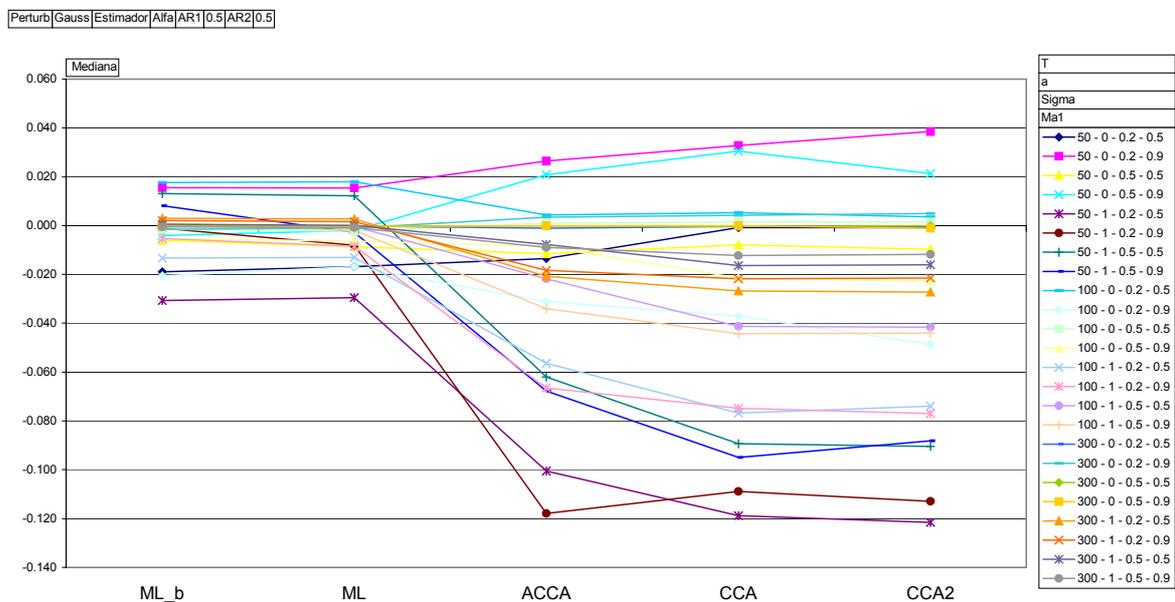


Figura 7-13. Mediana de la distribución del ángulo formado entre los vectores de cointegración estimado y real. Procedimientos de estimación ML_b, ML, ACCA, CCA y CCA2. Para distintas combinaciones de valores de los parámetros del proceso generador PGD2.

Análisis de Sistemas Cointegrados mediante Métodos de Subespacios

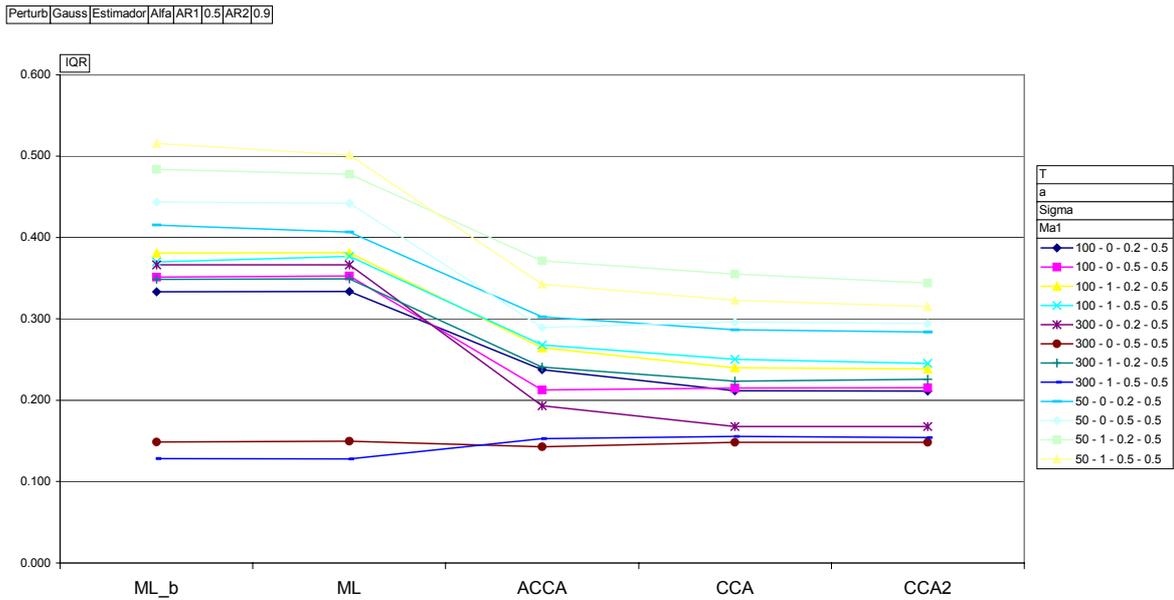


Figura 7-14. Rango intercuartílico (IQR) de la distribución del ángulo formado entre los vectores de cointegración estimado y real. Procedimientos de estimación ML_b, ML, ACCA, CCA y CCA2. Para distintas combinaciones de valores de los parámetros del proceso generador PGD2, con $\phi_1 = .5$ y $\phi_2 = .9$

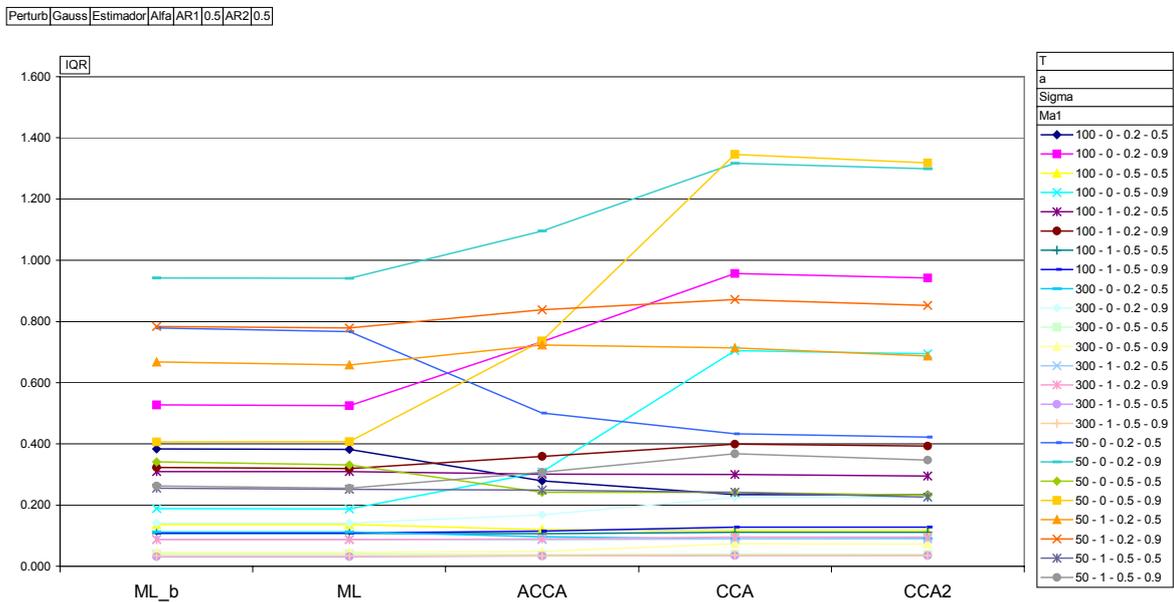


Figura 7-15. Rango intercuartílico (IQR) de la distribución del ángulo formado entre los vectores de cointegración estimado y real. Procedimientos de estimación ML, ML_b, ACCA, CCA y CCA2. Para distintas combinaciones de valores de los parámetros del proceso generador PGD2, con $\phi_1 = .5$ y $\phi_2 = .5$

Resultados

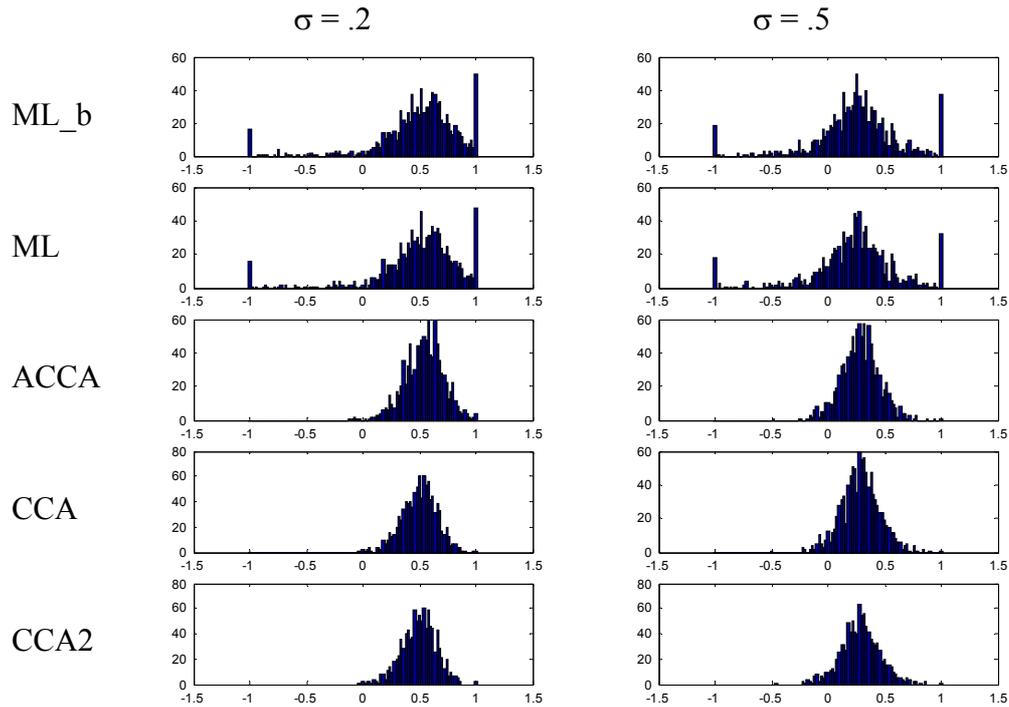


Figura 7-16. Histogramas del ángulo que separa el vector de cointegración estimado del real. 1000 realizaciones del proceso PGD2. $T = 100$, $\beta = 1$, $\sigma = (.2, .5)$, $a = 0$, $\varphi_1 = .9$, $\varphi_2 = .5$, $\theta = .5$

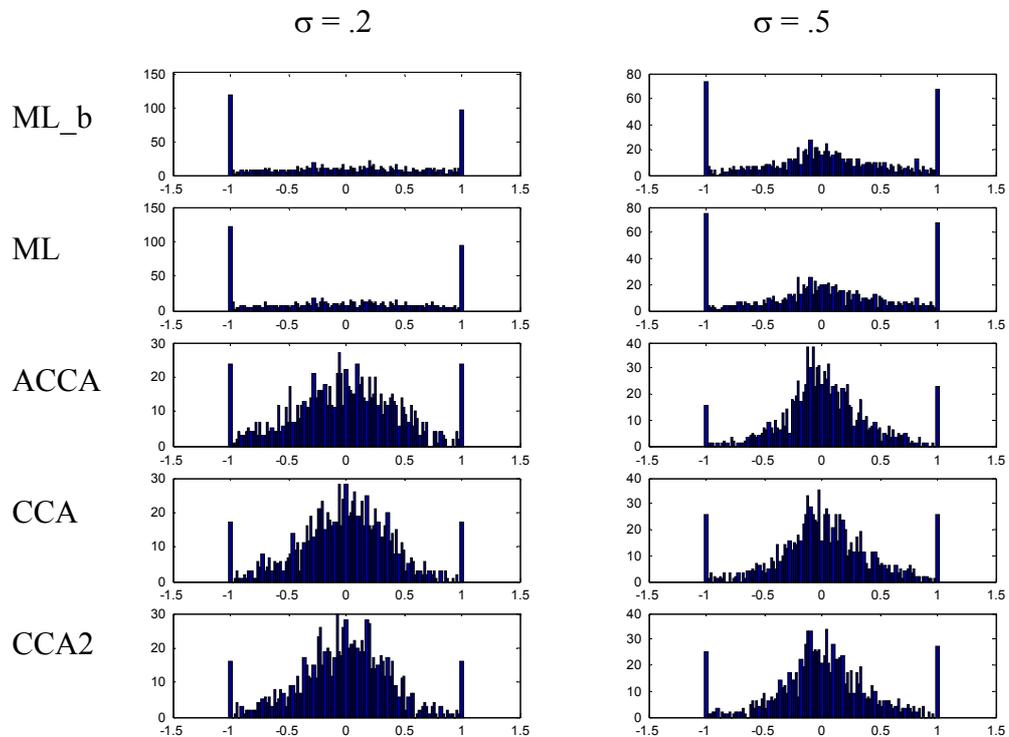


Figura 7-17. Histogramas del ángulo que separa el vector de cointegración estimado del real. 1000 realizaciones del proceso PGD2. $T = 100$, $\beta = 1$, $\sigma = (.2, .5)$, $a = 0$, $\varphi_1 = .9$, $\varphi_2 = .9$, $\theta = .5$

Análisis de Sistemas Cointegrados mediante Métodos de Subespacios

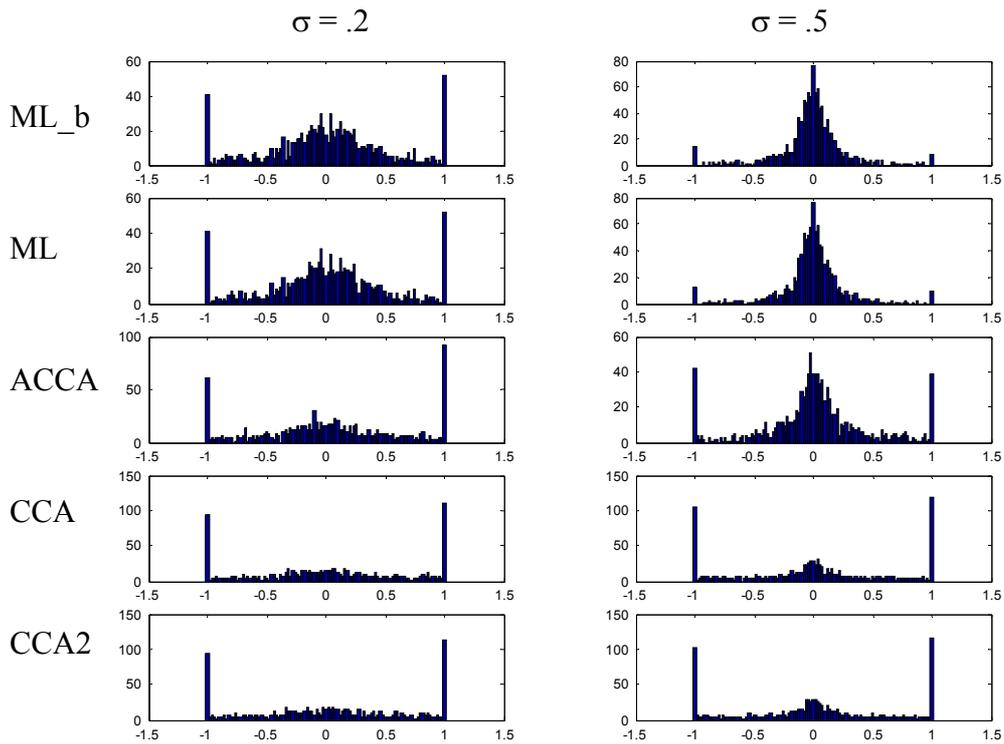


Figura 7-18. Histogramas del ángulo que separa el vector de cointegración estimado del real. 1000 realizaciones del proceso PGD2. $T = 100$, $\beta = 1$, $\sigma = (.2, .5)$, $a = 0$, $\varphi_1 = .5$, $\varphi_2 = .5$, $\theta = .9$

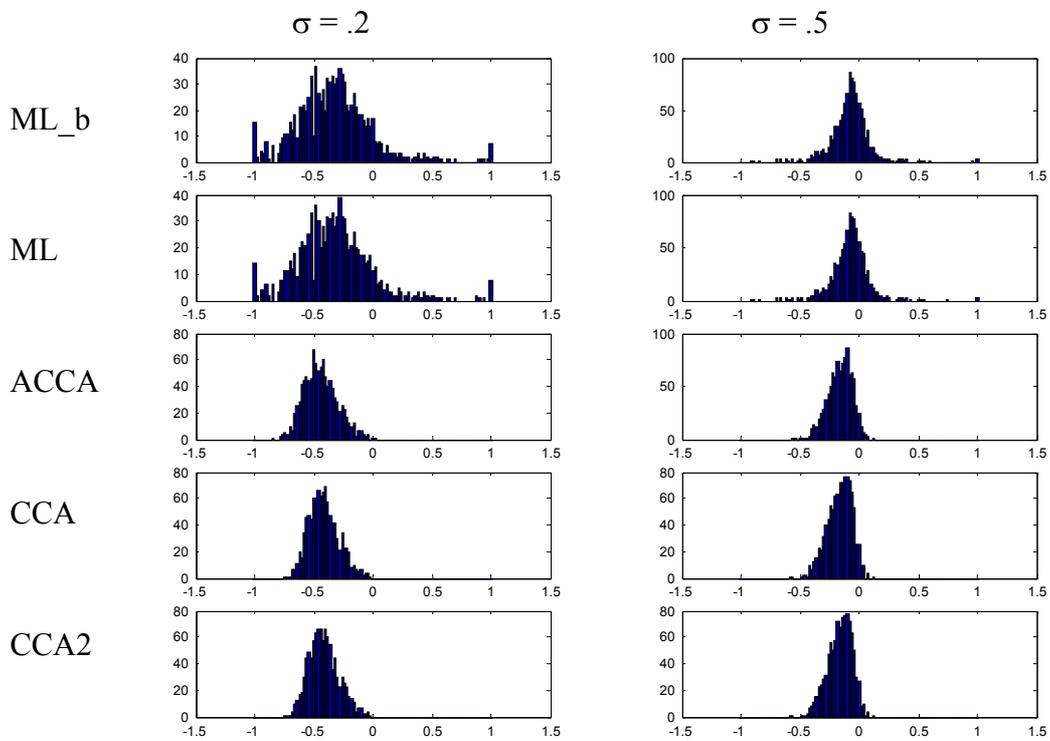


Figura 7-19. Histogramas del ángulo que separa el vector de cointegración estimado del real. 1000 realizaciones del proceso PGD2. $T = 300$, $\beta = 1$, $\sigma = (.2, .5)$, $a = 0$, $\varphi_1 = .5$, $\varphi_2 = .9$, $\theta = .5$

Resultados

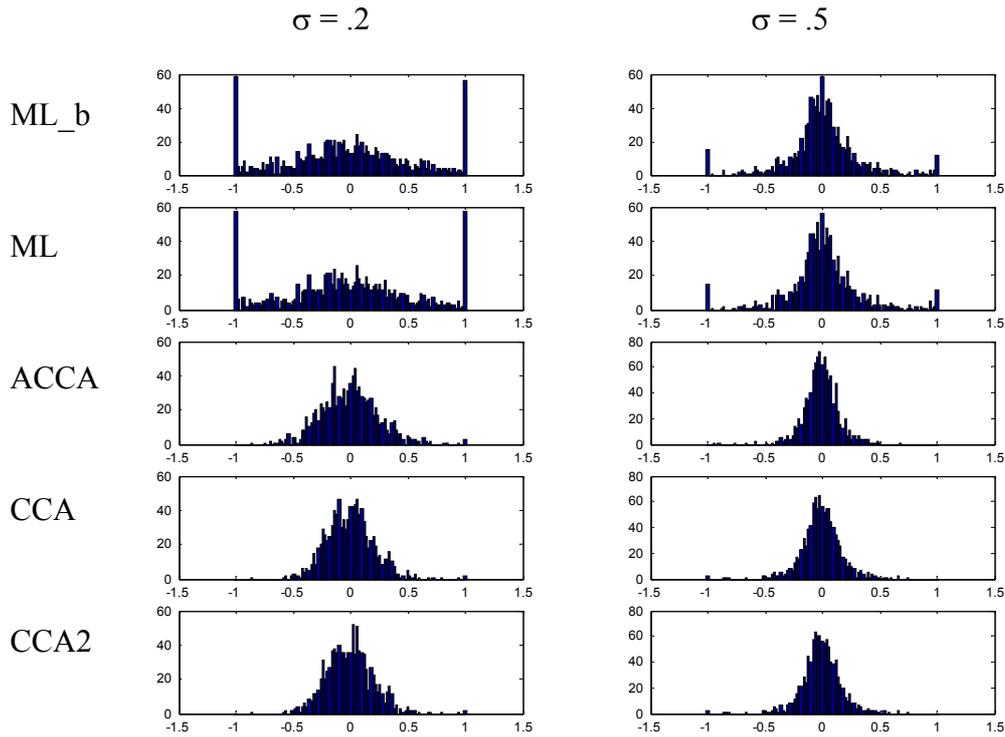


Figura 7-20. Histogramas del ángulo que separa el vector de cointegración estimado del real. 1000 realizaciones del proceso PGD2. $T = 300$, $\beta = 1$, $\sigma = (.2, .5)$, $a = 0$, $\varphi_1 = .9$, $\varphi_2 = .9$, $\theta = .5$

7.2.3 Error de predicción a corto plazo

El segundo aspecto utilizado para comparar los distintos algoritmos es la capacidad para captar la dinámica a corto plazo del proceso. Como ya hemos indicado, valoraremos esta capacidad mediante el estadístico U de Theil, que compara el error cuadrático de predicción a horizonte k de cada modelo con el de la predicción $\hat{y}_{t+k|t} = y_t$.

Para cada combinación de valores de los parámetros del proceso generador PGD2 se realizaron 1000 simulaciones. En cada repetición, se generó una muestra de trabajo y una muestra de validación (predicción); con la muestra de trabajo se estimaron los distintos modelos para el proceso generador, conforme a los distintos algoritmos de identificación comparados; con la muestra de validación y con cada uno de los modelos estimados se calculó el error cuadrático medio de predicción a horizontes 1, 2 y 3. Se estudió la distribución del estadístico U de Theil mediante tablas, gráficas e histogramas.

Para la muestra de predicción se utilizó, como estado inicial, el estado final estimado en la muestra de trabajo. Para la muestra de trabajo se utilizó un estado inicial suavizado.

Como curiosidad, se comprobó que otras formas habituales de estimar el estado inicial en la muestra de trabajo tienen, en general, poca influencia sobre el valor estimado del estado al comienzo de la muestra de predicción, como se puede observar en la Figura 7-21, donde se muestran las diferencias en predicción utilizando dos estimaciones diferentes del estado inicial: una estimación suavizada y una estimación de estado inicial nulo. La predicción con estado inicial nulo suele ser mucho peor para las primeras observaciones de la muestra de trabajo (especialmente con series no estacionarias, donde el nivel de partida puede ser alto; ver Figura 6-1), pero al cabo de un número suficiente de observaciones de la muestra de trabajo (veinte en general, en el caso representado, y recordemos que no comenzaremos a valorar la capacidad de predicción hasta llegar al final de la muestra de trabajo), la diferencia de predicción debida a diferente estimación del estado inicial es, en la mayoría de los casos, despreciable.

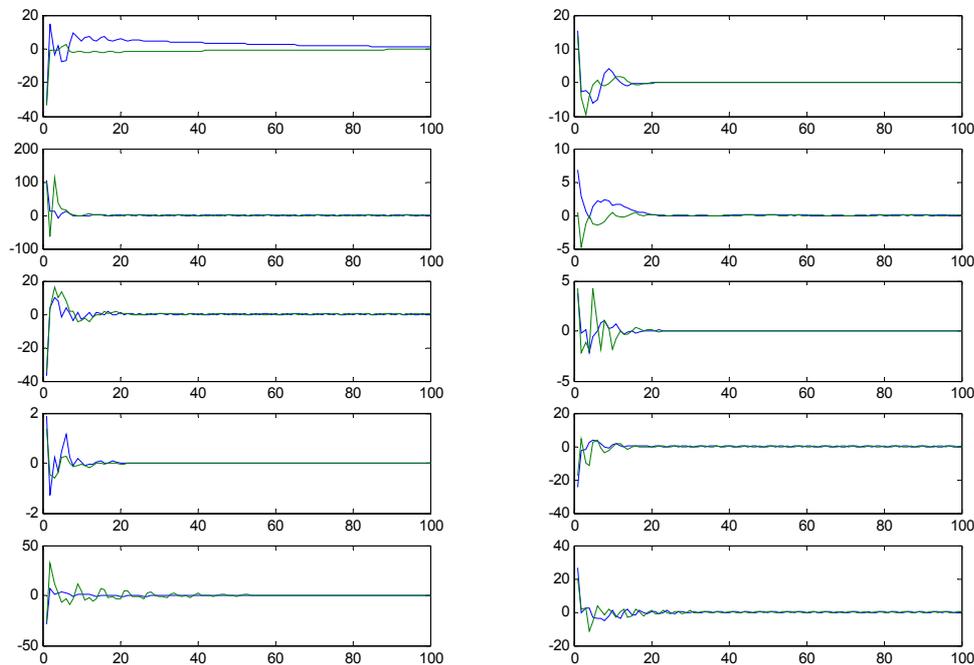


Figura 7-21 *Diferencias en las predicciones a horizonte uno con un modelo estimado por el método CCA y distintos estimadores del estado inicial. 10 simulaciones del proceso PGD2. Cada línea corresponde a cada una de las dos series observadas. Las diferencias corresponden a predicciones con un estado inicial suavizado y un estado inicial nulo*

Resultados

La Tabla 7-3 muestra los resultados de predicción correspondientes a una combinación de parámetros particular, para la que el método CCA2 proporciona los mejores resultados. Éste es en realidad un caso aislado, más representativos son los resultados que mostramos entre la Figura 7-22 y la Figura 7-27, donde se observa que:

- La mediana de la distribución del coeficiente de Theil es en general menor que uno para todos los métodos, lo que indica predicciones mejores a la del modelo “camino aleatorio”. La ventaja es más marcada para valores altos de los coeficientes autorregresivos del ciclo y para $a = 1$. Sin embargo, para $T = 50$, se obtienen con frecuencia coeficientes de Theil superiores a uno.
- El mejor funcionamiento de un método de estimación u otro (ML / subespacios) depende en gran medida de los valores de los parámetros del proceso generador, especialmente de los coeficientes autorregresivos de los ciclos. El tamaño muestral y la señal-ruido tienen gran influencia en la calidad de predicción, pero es en general una influencia uniforme sobre los distintos métodos. Por otro lado, los resultados no son iguales para las dos series estudiadas.
- En general se observa una ventaja de los métodos de subespacios frente a ML en la predicción a horizonte 1, ventaja que va disminuyendo al ir ampliando el horizonte de predicción. La ventaja también va disminuyendo al aumentar el tamaño muestral, aunque con $T = 300$ sigue siendo apreciable. Con $T = 50$ hay alta variabilidad respecto al método que mejor funciona.
- Los métodos ML (que resta las medias muestrales) y ML_b (que no resta las medias y admite término constante en la relación de cointegración) resultan prácticamente equivalentes.
- Los métodos ACCA y CCA presentan resultados de predicción muy similares, sin que se aprecie ventaja del método ACCA (CCA adaptado) frente al método CCA simple. El método CCA2 presenta resultados en general similares a CCA, pero suelen ser algo inferiores en calidad. En particular, existen algunos casos en que los modelos proporcionados por el método CCA2 resultan muy inferiores en capacidad predictiva a los proporcionados por el método CCA simple, como se comprueba por la elevada dispersión de la distribución del coeficiente de Theil,

Análisis de Sistemas Cointegrados mediante Métodos de Subespacios

particularmente para tamaños muestrales reducidos (50, 100): ver Figura 7-25 y Figura 7-26.

- No existen diferencias en las conclusiones que se extraen utilizando las perturbaciones no gaussianas (distribución t con 5 grados de libertad o uniforme de -2 a 2).

Serie	Sigma	Estadístico	Método				
			ML_b	ML	ACCA	CCA	CCA2
1	0.2	Media	0.914	0.914	0.875	0.865	0.858
		Mediana	0.867	0.862	0.796	0.782	0.772
		IQR	0.375	0.383	0.364	0.367	0.368
		SD	0.339	0.339	0.382	0.384	0.370
		$P(U < 1)$	69%	69%	75%	76%	76%
	0.5	Media	0.929	0.928	0.909	0.902	0.900
		Mediana	0.877	0.874	0.837	0.820	0.815
		IQR	0.373	0.373	0.364	0.394	0.387
		SD	0.333	0.332	0.369	0.385	0.389
		$P(U < 1)$	67%	67%	72%	73%	72%
2	0.2	Media	0.920	0.920	0.888	0.880	0.874
		Mediana	0.859	0.859	0.793	0.780	0.774
		IQR	0.380	0.383	0.361	0.375	0.395
		SD	0.351	0.352	0.402	0.406	0.402
		$P(U < 1)$	70%	69%	75%	75%	74%
	0.5	Media	0.942	0.942	0.936	0.931	0.919
		Mediana	0.884	0.885	0.830	0.815	0.819
		IQR	0.374	0.373	0.391	0.418	0.418
		SD	0.388	0.389	0.559	0.563	0.541
		$P(U < 1)$	66%	66%	71%	71%	70%

Tabla 7-3. Resultados de la distribución del estadístico U de Theil en predicción a horizonte 1 para 1.000 simulaciones del proceso PGD2, distintos procedimientos de estimación del modelo y distintos valores de sigma, con $T=100$, $\beta=1$, $\alpha=0$, $\varphi_1=.5$, $\varphi_2=.5$, $\theta=.5$. Todos los estimadores restan las medias salvo ML_b. En negrita los mejores valores de cada fila.

Resultados

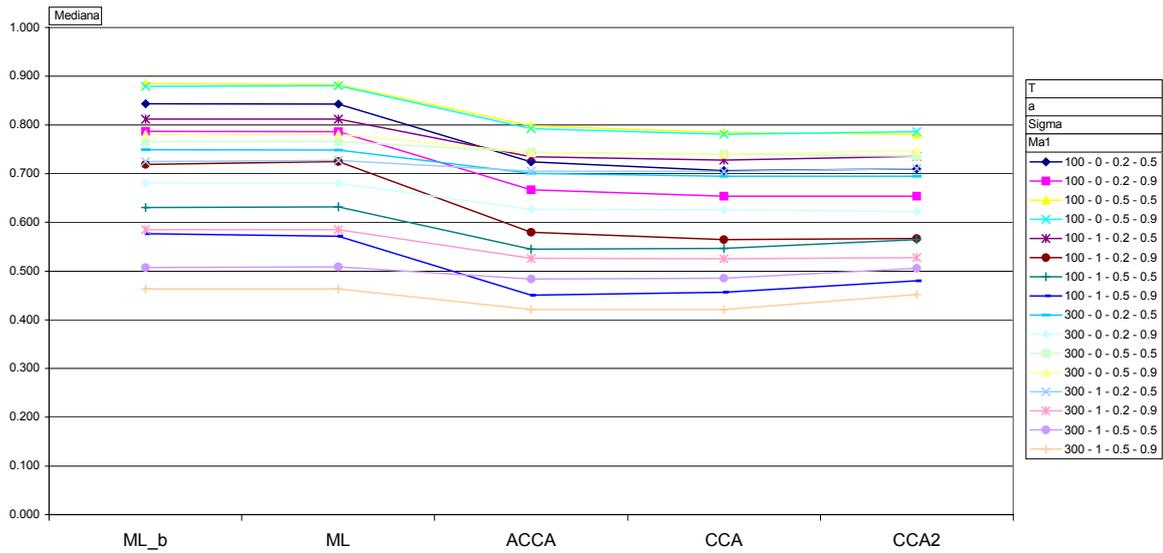


Figura 7-22. Mediana de la distribución del estadístico U de Theil en predicción de la serie y_2 a horizonte 1, para modelos estimados mediante los procedimientos ML_b, ML, ACCA, CCA, y CCA2. Para distintas combinaciones de valores de los parámetros del proceso generador PGD2.

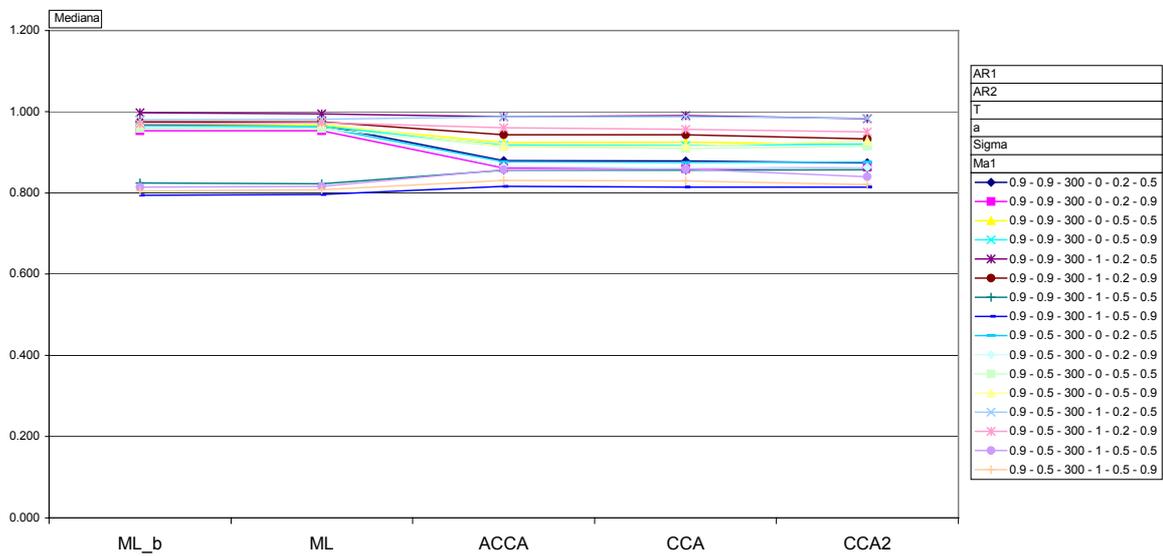


Figura 7-23. Mediana de la distribución del estadístico U de Theil en predicción de la serie y_1 a horizonte 3, para modelos estimados mediante los procedimientos ML_b, ML, ACCA, CCA, y CCA2. Para distintas combinaciones de valores de los parámetros del proceso generador PGD2.

Análisis de Sistemas Cointegrados mediante Métodos de Subespacios

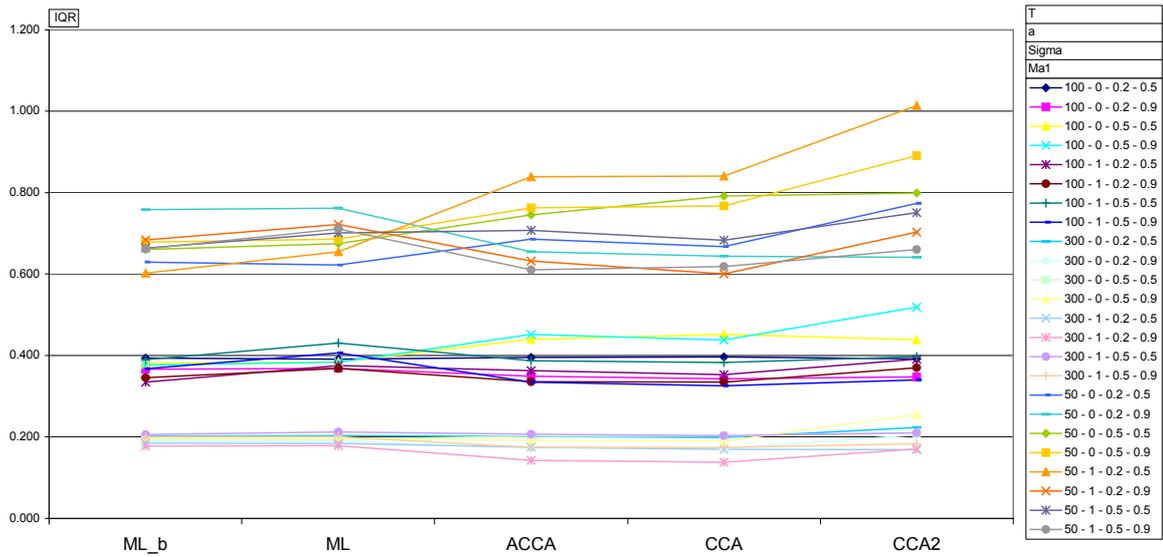


Figura 7-24. Rango intercuartílico (IQR) de la distribución del estadístico U de Theil en predicción de la serie y_2 a horizonte 1, para modelos estimados mediante los procedimientos ML_b, ML, ACCA, CCA, y CCA2. Para distintas combinaciones de valores de los parámetros del proceso generador PGD2.

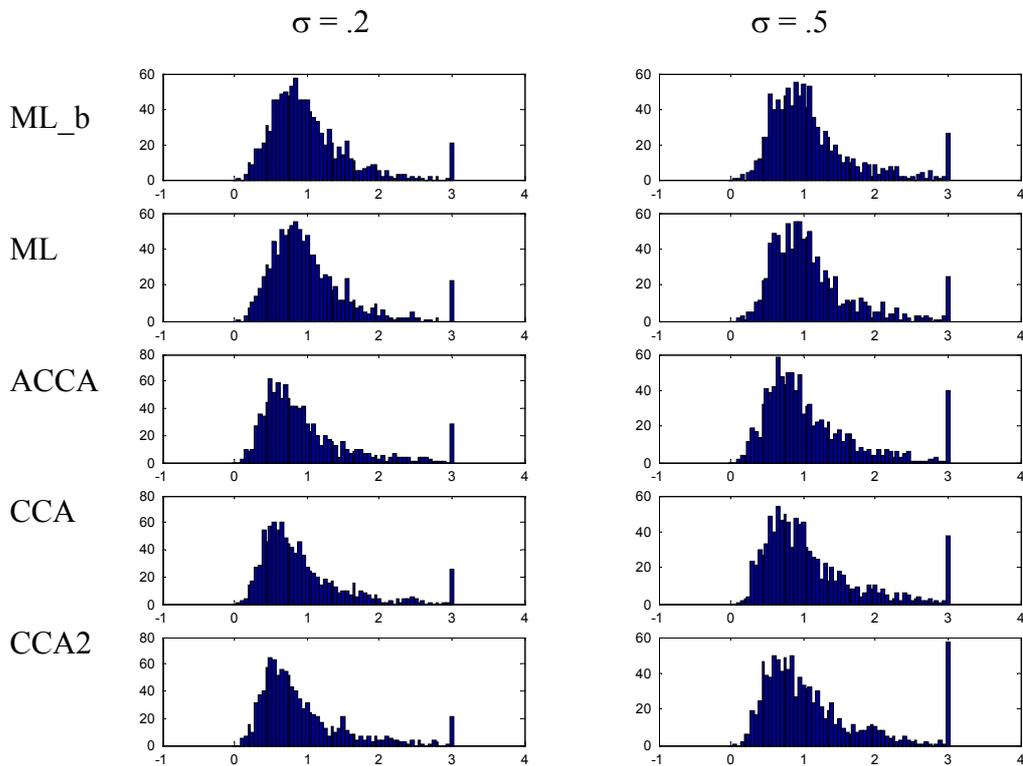


Figura 7-25. Histogramas del estadístico U de Theil para la serie y_1 a horizonte 1. 1000 realizaciones del proceso PGD2. $T = 50$, $\beta = 1$, $\sigma = (.2, .5)$, $a = 0$, $\varphi_1 = .9$, $\varphi_2 = .5$, $\theta = .9$

Resultados

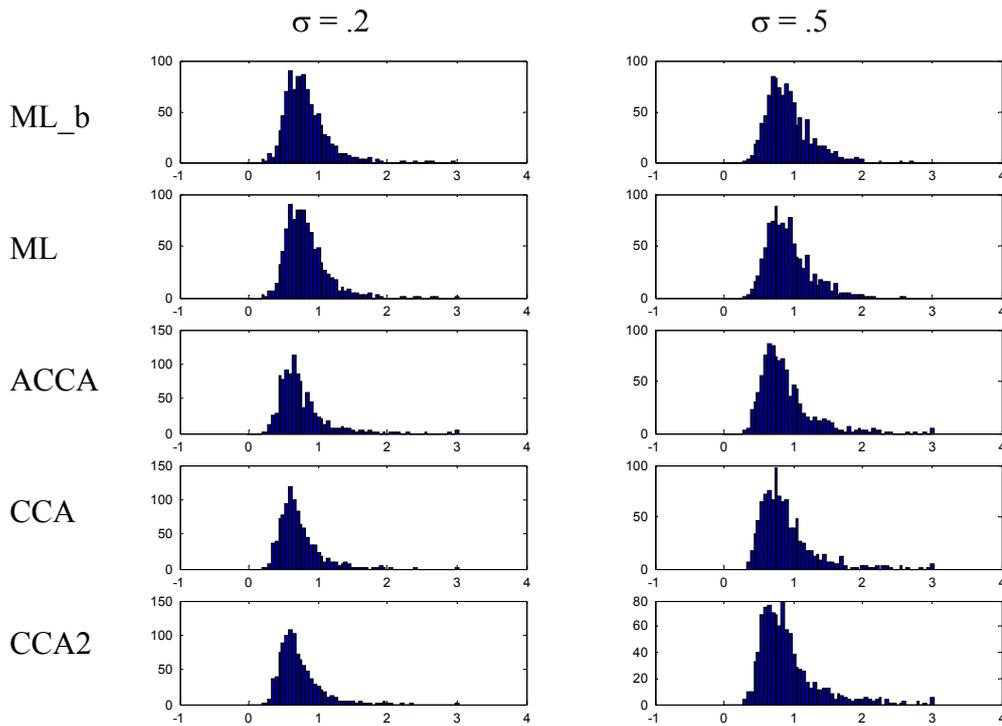


Figura 7-26. Histogramas del estadístico U de Theil para la serie y_1 a horizonte 1. 1000 realizaciones del proceso PGD2. $T = 100$, $\beta = 1$, $\sigma = (.2, .5)$, $a = 0$, $\varphi_1 = .9$, $\varphi_2 = .5$, $\theta = .9$

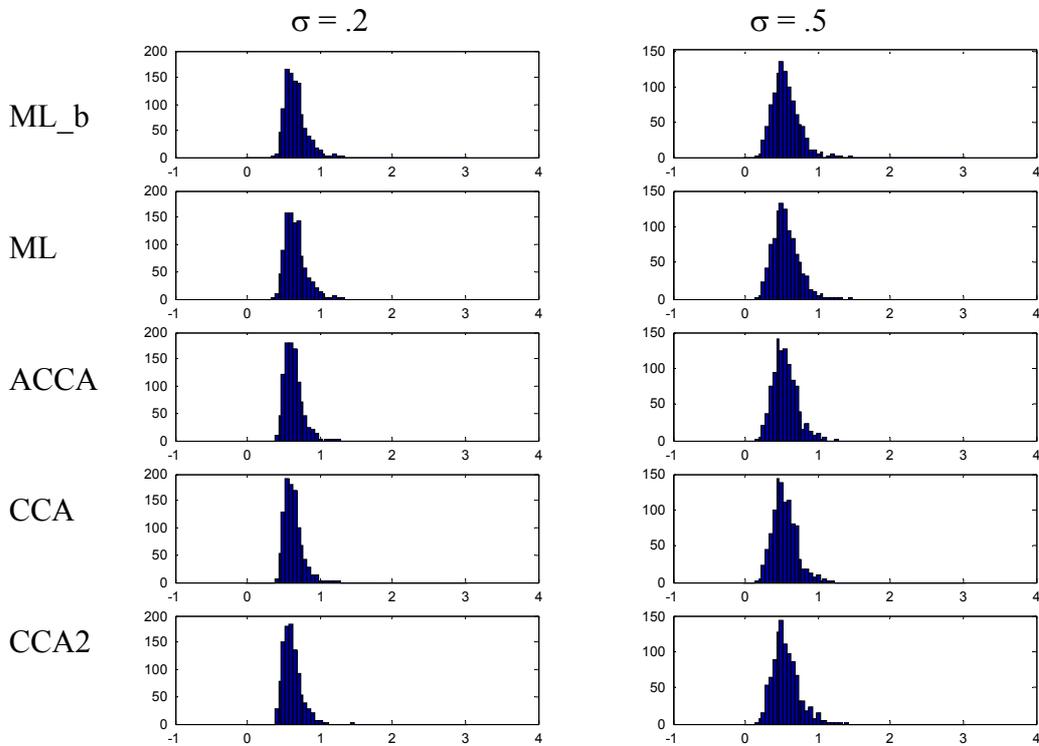


Figura 7-27. Histogramas del estadístico U de Theil para la serie y_1 a horizonte 1. 1000 realizaciones del proceso PGD2. $T = 300$, $\beta = 1$, $\sigma = (.2, .5)$, $a = 1$, $\varphi_1 = .9$, $\varphi_2 = .5$, $\theta = .9$

7.2.4 Contraste de igualdad de capacidad predictiva

Podemos plantearnos si es significativo el número de veces que el modelo identificado por un procedimiento supera en capacidad predictiva (error cuadrático) al modelo identificado por otro procedimiento. Para ello podemos utilizar el contraste binomial de Siegel (1985), cuyo aplicación exponemos a continuación.

Bajo la hipótesis de que dos procedimientos de identificación (A y B) son equivalentes (a efectos de calidad de predicción de los modelos proporcionados), la variable $n =$ "número de veces que el modelo del método A supera al modelo del método B en N repeticiones independientes" seguiría una distribución binomial de parámetro $p = 1/2$ (probabilidad de que el modelo del método A proporcione un menor error cuadrático de predicción que el modelo del método B). Tendríamos entonces que

$$P(n \leq k) = \sum_{i=0}^k \binom{N}{i} p^i (1-p)^{N-i} = \frac{1}{2^N} \sum_{i=0}^k \binom{N}{i}$$

Lo que nos permite realizar contrastes de igualdad de métodos.

Utilizando la distribución anterior, bajo la hipótesis nula de igualdad y tomando $N = 1.000$ repeticiones, la probabilidad de que uno de los métodos supere al otro (contraste de dos lados) en más de 541 ocasiones es inferior al 1%, e inferior al 1 por mil si consideramos 552 o más ocasiones. En las comparaciones posteriores tomaremos 541 como valor crítico superior (ó, equivalentemente, 459 como valor inferior, que es $1.000 - 541$).

Hemos decidido adoptar el método CCA simple como referencia en las comparaciones dos a dos, ya que tanto ACCA como CCA2 son modificaciones de este método que buscan mejorarlo. La Tabla 7-4 muestra, para distintas combinaciones de valores de parámetros del proceso generador PGD2, el número de veces (de entre 1.000 repeticiones) en que los modelos identificados por los métodos de Johansen, ACCA o CCA2 proporcionaron predicciones de menor error cuadrático medio que el modelo correspondiente estimado por el algoritmo CCA simple.

Resultados

T	φ_1	φ_2	θ	Sigma					
				0.2			0.5		
				Johansen	ACCA	CCA2	Johansen	ACCA	CCA2
50	0.5	0.5	0.5	416	471	514	471	490	533
			0.9	320	454	505	455	528	462
		0.9	0.5	404	465	513	408	446	517
			0.9	361	447	461	410	484	497
	0.9	0.5	0.5	443	497	512	453	537	513
			0.9	328	504	496	423	496	463
		0.9	0.5	404	450	521	456	492	514
			0.9	344	449	519	426	476	519
100	0.5	0.5	0.5	377	441	518	392	413	526
			0.9	307	454	513	419	443	423
		0.9	0.5	278	380	488	347	430	519
			0.9	234	395	503	375	448	449
	0.9	0.5	0.5	372	503	525	409	476	503
			0.9	275	502	529	377	478	445
		0.9	0.5	254	441	514	320	454	533
			0.9	240	437	535	314	455	510
300	0.5	0.5	0.5	370	427	311	395	415	382
			0.9	307	411	330	383	450	302
		0.9	0.5	226	433	480	301	401	458
			0.9	237	405	534	339	469	354
	0.9	0.5	0.5	399	511	366	441	440	278
			0.9	347	499	345	374	493	257
		0.9	0.5	224	411	575	352	441	542
			0.9	179	441	543	295	485	454

Tabla 7-4. Número de veces entre 1.000 en que el modelo identificado por los métodos de Johansen, ACCA o CCA2 bate en error cuadrático medio de predicción al modelo identificado por el método CCA simple. Se considera significativo a favor de CCA un valor inferior a 459 (celdas en negrita), y significativo a favor del método en cuestión un valor superior a 541 (celdas con relleno). La tabla se ha realizado con el error de predicción de la serie y_2 a horizonte 1, con $a = 0$, $\beta = 1$ y perturbaciones gaussianas.

En la Tabla 7-4 pueden apreciarse unos resultados del método CCA simple para predicción de la serie y_2 notablemente superiores a los de las otras alternativas, y, en particular, muy superiores a los del método de Johansen. Sólo el método CCA2, para tamaños muestrales $T = 50$ y $T = 100$, resulta comparable. Esta ventaja de CCA simple se mantiene, aunque no tan notable, para la serie y_1 , y no se ve afectada por violaciones de la hipótesis de normalidad de las perturbaciones. Sin embargo, otros valores del parámetro β ($\beta = -3$) provocaron que la ventaja de CCA simple respecto a Johansen o a

ACCA disminuyera o incluso desapareciera, mientras que el método CCA2 empeoró notablemente su rendimiento.

7.2.5 Combinación de predicciones

Aunque unos algoritmos de identificación proporcionen mejores modelos predictivos que otros, podemos plantearnos si la combinación de modelos identificados por los distintos procedimientos puede mejorar notablemente las predicciones de un modelo único.

Una primera idea la proporciona el estudio gráfico de los residuos de predicción correspondientes a cada modelo, como puede verse en la Figura 7-28.

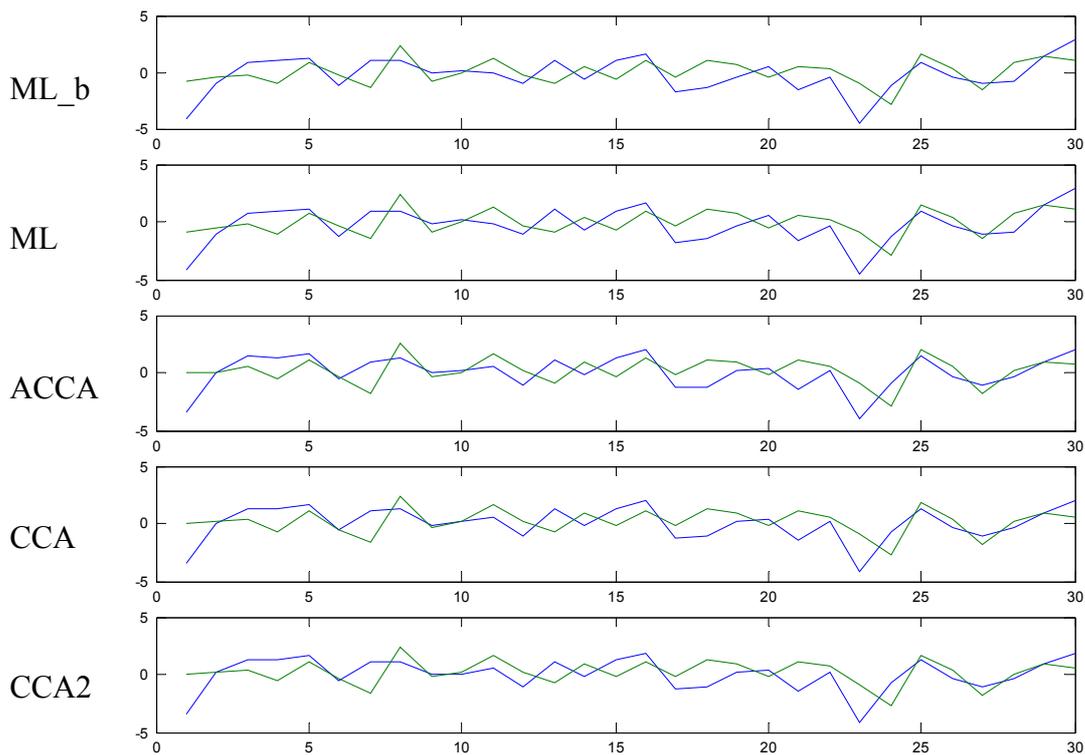


Figura 7-28. Errores de predicción a horizonte uno de dos series generadas por el proceso PGD2, con las predicciones realizadas mediante modelos estimados por los distintos procedimientos comparados.

Resultados

Los residuos de predicción de los distintos modelos se mueven en paralelo, lo cual parece indicar que no hay información complementaria para predicción entre los distintos modelos.

Para un tratamiento analítico acudimos al estudio de métodos de combinación de predicciones de Diebold y López (1995). Si consideramos las predicciones obtenidas por dos modelos diferentes $\hat{y}_{t+k|t}^{(1)}$ e $\hat{y}_{t+k|t}^{(2)}$, podemos buscar la combinación lineal

$$\hat{y}_{t+k|t}^c = \omega \hat{y}_{t+k|t}^{(1)} + (1-\omega) \hat{y}_{t+k|t}^{(2)}$$

que minimiza el error cuadrático de predicción. El peso óptimo ω^* se puede estimar mediante la fórmula

$$\hat{\omega}^* = \frac{\hat{\sigma}_{22}^2 - \hat{\sigma}_{12}}{\hat{\sigma}_{22}^2 + \hat{\sigma}_{11}^2 - 2\hat{\sigma}_{12}}$$

donde $\hat{\sigma}_{11}^2$, $\hat{\sigma}_{22}^2$, $\hat{\sigma}_{12}$ son los estimadores de las varianzas y covarianza de los errores de predicción ($e_{t+k,t}^i = \hat{y}_{t+k|t}^i - y_{t+k}$) de cada modelo, proporcionados por la fórmula

$$\hat{\sigma}_{ij} = \frac{1}{T} \sum_{t=1}^T e_{t+k,t}^i e_{t+k,t}^j$$

Un valor de ω^* igual a uno (cero) indica que el primer (segundo) modelo produce mejores predicciones que el otro, el cual no aportaría información adicional para mejorar la predicción. Valores de ω^* distintos de cero o uno indican que existe información en los dos modelos que puede combinarse para mejorar la predicción.

El método puede extenderse para comparar simultáneamente varios modelos, pero los resultados son más fáciles de visualizar y analizar comparando modelos dos a dos. En este caso tomaremos como referencia el algoritmo CCA2, por ser original de esta tesis doctoral.

En la Figura 7-29 se han representado histogramas del peso obtenido por los modelos ML al combinarlos en predicción con modelos CCA2, y en la Figura 7-30 se estudian combinaciones de modelos CCA con CCA2. La concentración en los valores cero y uno indica que normalmente uno de los dos modelos resulta mejor en predicción, y el otro no añade información útil.

En general, resultan mejor los modelos identificados por CCA2 (concentración en el valor $\omega^* = 0$) que los ML, pero esta ventaja disminuye con el horizonte de predicción (H) y, especialmente, con el tamaño muestral. El método ML llega a resultar mejor que CCA2 para determinadas combinaciones de parámetros del proceso generador.

El estudio de combinaciones de modelos CCA con CCA2 permite observar que, en general, los modelos obtenidos por ambos procedimientos tampoco se complementan. Para tamaños muestrales pequeños (50, 100), CCA2 resulta en muchas ocasiones mejor a CCA, pero suele resultar peor para tamaño muestral 300. Pese a que en muchos casos CCA2 resulta mejor que ML ó CCA, ya hemos observado que los resultados obtenidos con el método CCA2 presentan más variabilidad que las otras alternativas (mayor IQR en la distribución del estadístico U). Por otro lado, en este punto se estudió la influencia del valor de la relación de cointegración (parámetro β), y experimentos con otros valores del vector de cointegración ($\beta = -3$) condujeron a resultados muy inferiores del método CCA2 respecto a ML ó CCA, particularmente para el caso $T = 300$.

Resultados

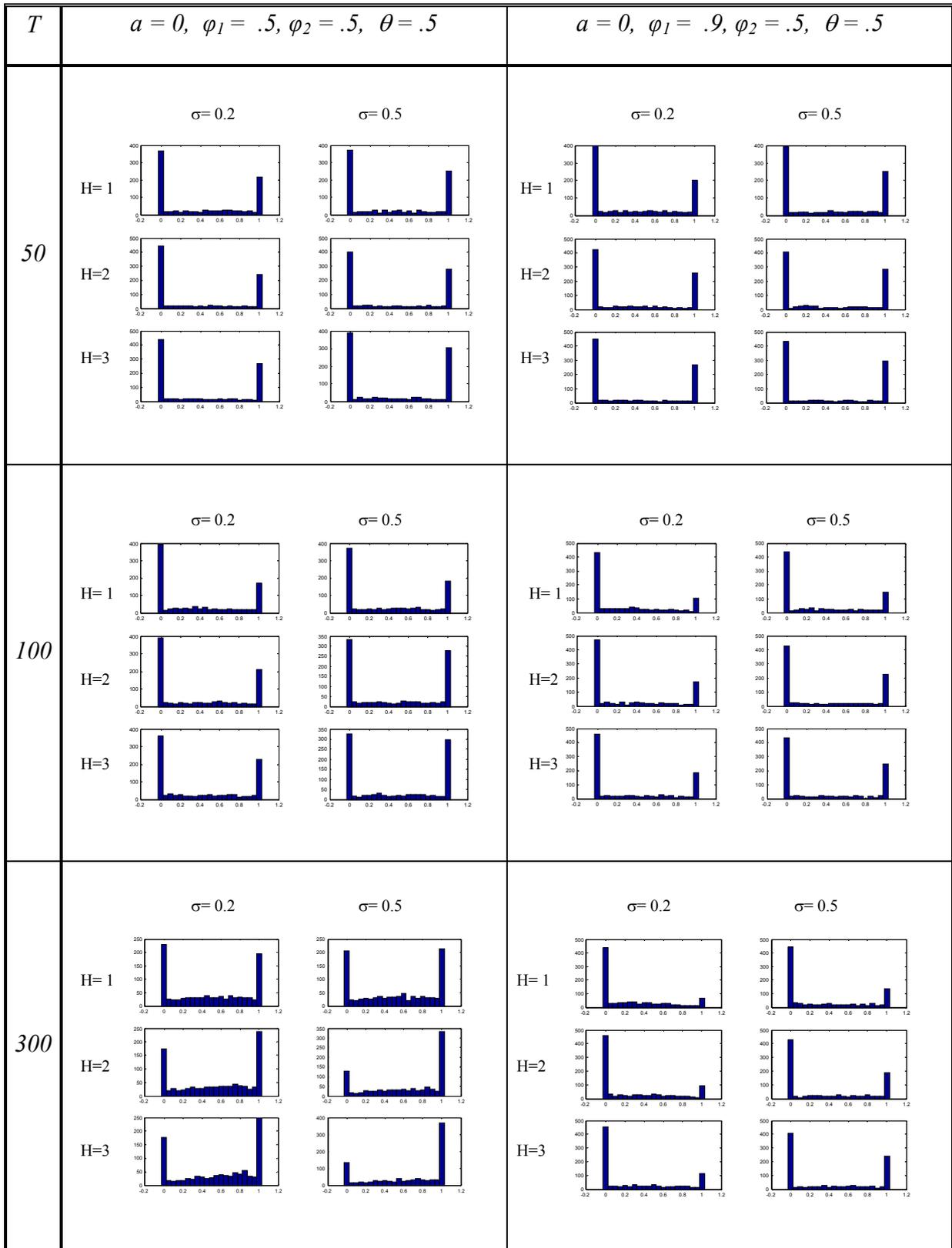


Figura 7-29. Histogramas de los pesos ω^* para modelos ML en combinaciones con modelos CCA2 para predicción de la serie 1 a horizonte H .

Análisis de Sistemas Cointegrados mediante Métodos de Subespacios

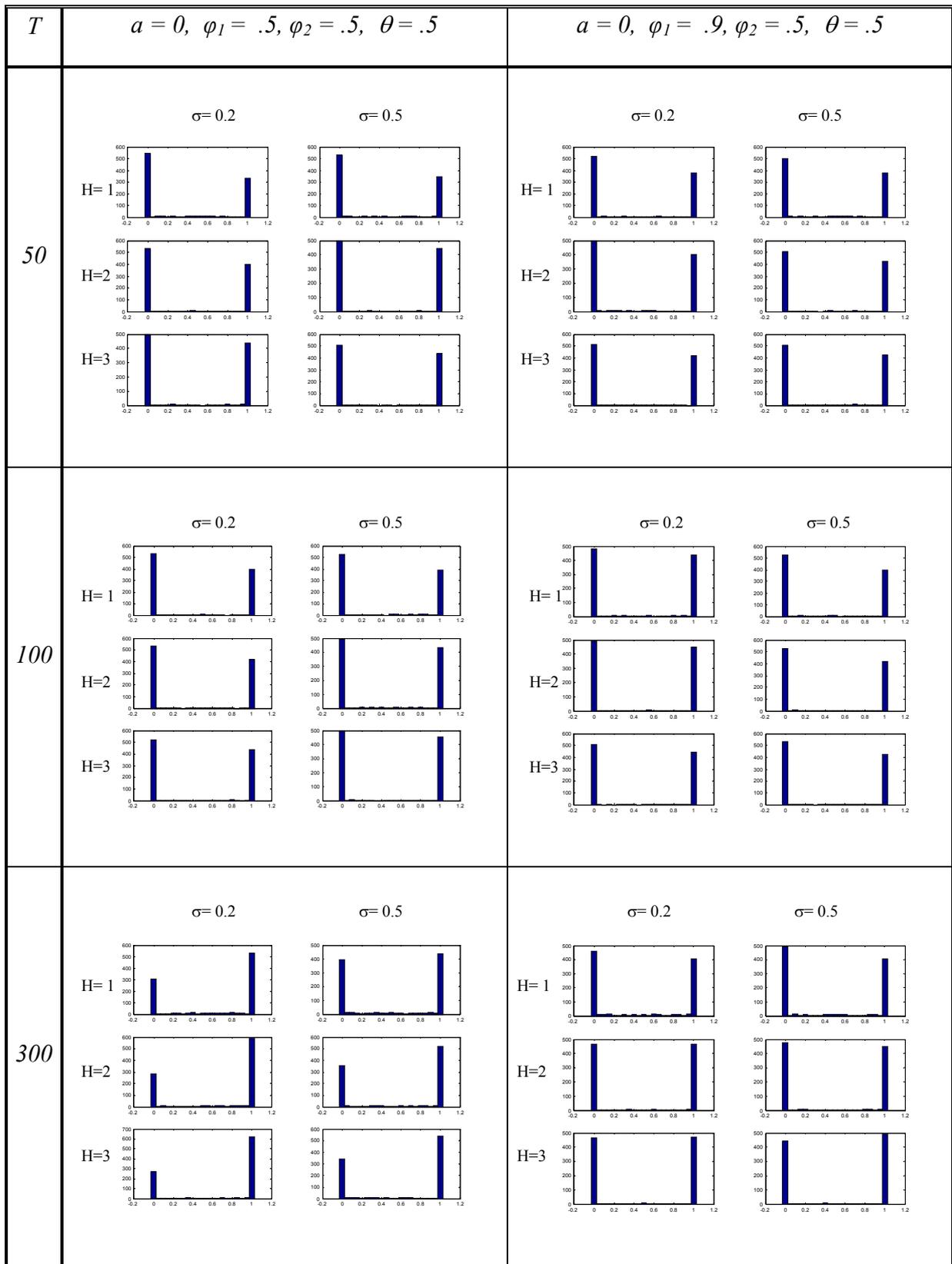


Figura 7-30. Histogramas de los pesos ω^* para modelos CCA en combinaciones con modelos CCA2 para predicción de la serie 1 a horizonte H .

7.3 Proceso generador de datos 3 (PGD3)

7.3.1 Características generales

El proceso generador de datos 3 (PGD3) genera tres series temporales por combinación de una tendencia común (lo que implica 2 vectores de cointegración) y dos ciclos autorregresivos AR(1). La tendencia común es de la forma

$$\tau_{t+1} = \tau_t + e_{1,t} + a \theta e_{1,t-1}$$

lo que corresponde a un camino aleatorio puro (si $a = 0$), o bien a una serie cuyas primeras diferencias siguen un proceso MA(1) de parámetro θ .

Una formulación en el espacio de estados para el PGD3 es

$$\begin{bmatrix} \tau \\ \eta_1 \\ \eta_2 \\ \eta_3 \end{bmatrix}_{t+1} = \begin{bmatrix} 1 & 0 & 0 & a \\ 0 & \varphi_1 & 0 & 0 \\ 0 & 0 & \varphi_2 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \tau \\ \eta_1 \\ \eta_2 \\ \eta_3 \end{bmatrix}_t + \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ \theta & 0 & 0 \end{bmatrix} \begin{bmatrix} e_1 \\ e_2 \\ e_3 \end{bmatrix}_t$$

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}_t = \begin{bmatrix} 0.8 & | & 1 & 0 & 0 \\ 1 & | & 0 & 1 & 0 \\ -1 & | & 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} \sigma \tau \\ \eta_1 \\ \eta_2 \\ \eta_3 \end{bmatrix}_t + \begin{bmatrix} e_1 \\ e_2 \\ e_3 \end{bmatrix}_t$$

donde el vector de perturbaciones e_t es ruido blanco de esperanza cero y matriz de covarianzas la identidad. Nótese que los estados estacionarios $\eta_{1,t}$ y $\eta_{2,t}$ corresponden a dos ciclos autorregresivos AR(1), y el estado $\eta_{3,t}$ guarda el valor de una perturbación hasta el periodo siguiente, lo que introduce componentes de media móvil.

Las relaciones de cointegración pertenecen al subespacio (plano) ortogonal al primer vector de la matriz C (este primer vector de C es, precisamente, el vector director del plano de cointegración).

Para incluir términos constantes en las relaciones de cointegración, se ha sumado un valor fijo δ_1 a todos los valores de una de las series y un valor fijo δ_2 a todos los valores de otra de las series (desplazamientos en vertical de todos los valores de una de las series con respecto a las otras).

Análisis de Sistemas Cointegrados mediante Métodos de Subespacios

El espacio paramétrico considerado es $(T \times \sigma \times a \times \varphi_1 \times \varphi_2 \times \theta \times \delta_1 \times \delta_2)$. Partiendo de las simulaciones anteriores y buscando procesos con características estadísticas variadas (Figura 7-31), hemos considerado los valores $T = (50, 100, 300)$, $\sigma = (.2, .5)$, $a = (0, 1)$, $\varphi_1 = (.5, .9)$, $\varphi_2 = (.5, .9)$, $\theta = (.5, .9)$, $\delta_1 = 50$ y $\delta_2 = -30$. También se han considerado perturbaciones no gaussianas, conforme a una distribución t de Student con 5 grados de libertad.

Realizaciones del proceso PGD3

Se representan a continuación un conjunto de realizaciones del proceso generador PGD3, correspondientes a distintas combinaciones de valores de los parámetros.

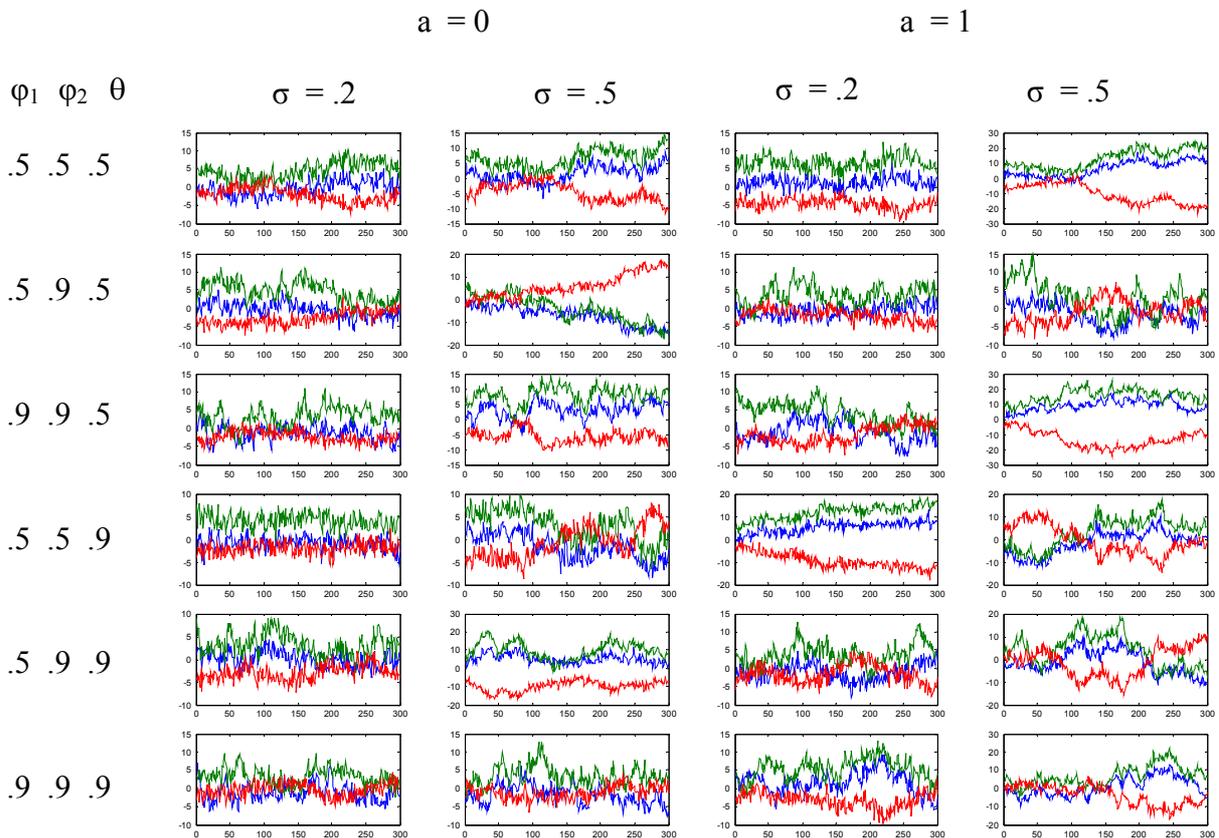


Figura 7-31. Realizaciones del proceso PGD3 con componente constante en la relación de cointegración. Para distintas combinaciones de valores de los parámetros, con desplazamientos $\delta_1 = 5$, $\delta_2 = -3$.

Resultados

Ajuste de los algoritmos

El orden p de la representación VAR del vector de series temporales y_t se ha seleccionado minimizando el criterio de información de Schwarz (BIC). Para los métodos de subespacios se tomaron valores de los parámetros $p = f = 5$. El orden n de la representación en el espacio de estados se calculó por el procedimiento BA(n) de Bauer y Wagner (sección 6.6).

La Figura 7-37 muestra un resultado representativo (típico) de la estimación del orden del modelo con los procedimientos CCA2 y ACCA.

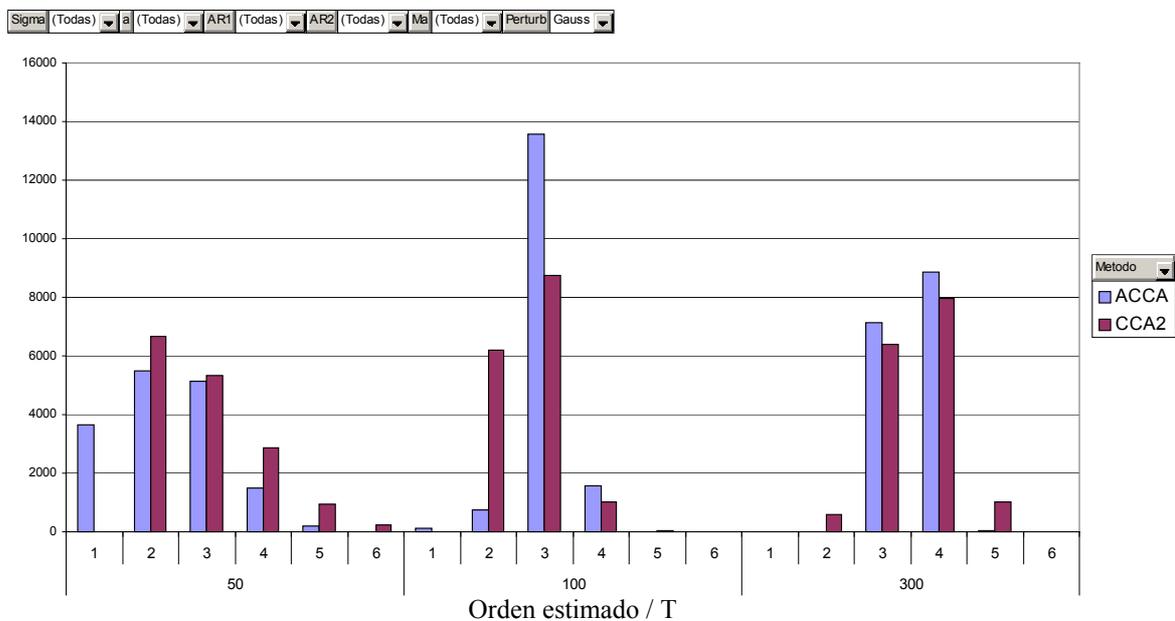


Figura 7-32. Histogramas del orden estimado (de modelos en el espacio de estados) por los procedimientos ACCA y CCA2, con el criterio BA(n), para distintos tamaños muestrales T.

Como sucedía con el PGD2, se observa que al aumentar T se reduce la variabilidad en la estimación del orden del modelo (número de estados), y el orden estimado tiende a aumentar y a concentrarse en torno al verdadero orden del proceso generador (se puede comprobar que la formulación del PGD3 que hemos considerado, con los valores de los parámetros considerados, es controlable y observable, y corresponde por tanto a una representación mínima del sistema PGD3, que es de orden 4).

Análisis de Sistemas Cointegrados mediante Métodos de Subespacios

La Figura 7-38 muestra resultados representativos la estimación del orden de una aproximación VAR por el criterio BIC, y también se observa una tendencia a aumentar el orden estimado del modelo al aumentar T.

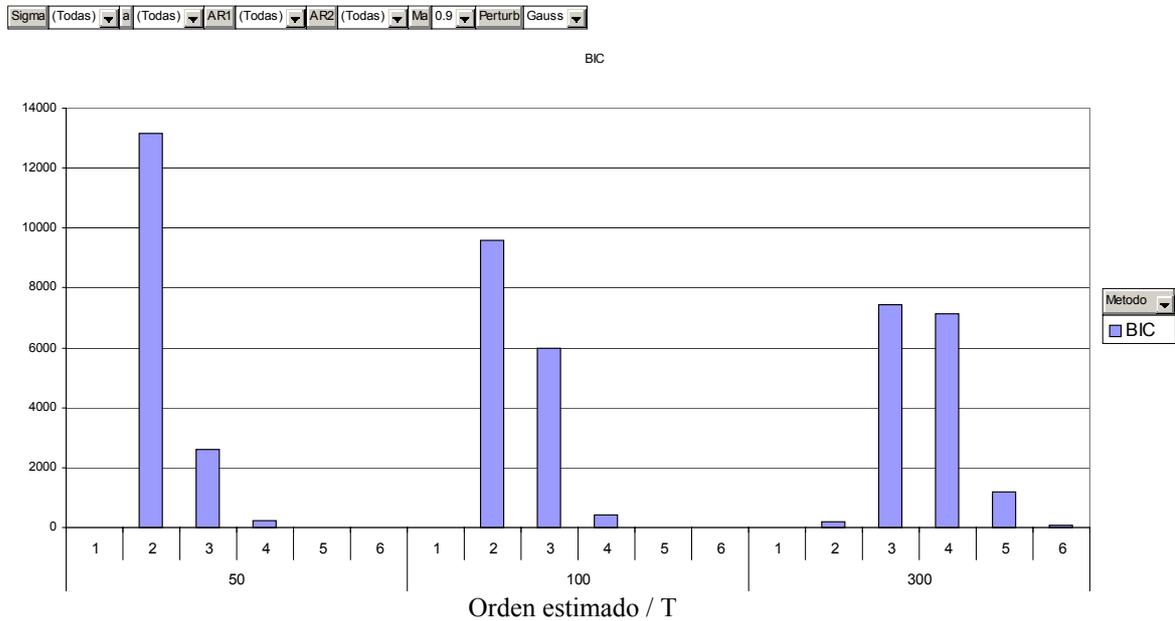


Figura 7-33. Histogramas del orden estimado (de modelos VAR) por el procedimiento de Johansen, con el criterio BIC, para distintos tamaños muestrales T, con $\theta = 0,9$.

7.3.2 Estimación del subespacio de cointegración

El subespacio de cointegración es en este caso un plano. La calidad de estimación la mediremos por el ángulo formado entre los vectores directores de los planos de cointegración real y estimado.

La Tabla 7-5 y la Tabla 7-6 y los gráficos entre la Figura 7-34 y la Figura 7-38 muestran estadísticos correspondientes a la distribución de este ángulo, para los distintos métodos de estimación y distintos parámetros del proceso generador PGD3.

Podemos extraer las siguientes conclusiones:

- Como en todos los casos anteriores de estimación de las relaciones de cointegración, los estimadores ML_b y ML son prácticamente idénticos, al igual que los estimadores CCA y CCA2.

Resultados

- En general, la estimación ACCA resulta algo peor que CCA simple. Aunque no suele haber grandes diferencias entre los métodos ML y CCA, existen varios casos en los que la estimación por CCA resulta mejor (más concentrada y con mediana más próxima al valor correcto) que la estimación ML, pero se producen algunas excepciones, sin que se identifique una relación clara entre estas excepciones y los valores de un parámetro concreto del proceso generador.
- Los resultados comparativos se mantienen al considerar perturbaciones conforme a una distribución t con 5 grados de libertad (los gráficos y tablas mostrados corresponden a perturbaciones gaussianas).

Análisis de Sistemas Cointegrados mediante Métodos de Subespacios

T	Sigma	Estadístico	Método				
			ML_b	ML	ACCA	CCA	CCA2
50	0.2	Media	0.470	0.465	0.518	0.543	0.540
		Mediana	0.374	0.371	0.438	0.446	0.444
		IQR	0.406	0.400	0.435	0.482	0.472
		SD	0.331	0.328	0.342	0.360	0.358
		$P(\alpha < 0.05)$	2%	1%	1%	1%	1%
	0.5	Media	0.173	0.167	0.171	0.197	0.195
		Mediana	0.128	0.125	0.134	0.145	0.142
		IQR	0.132	0.132	0.144	0.159	0.159
		SD	0.165	0.157	0.153	0.189	0.189
		$P(\alpha < 0.05)$	14%	14%	13%	12%	12%
100	0.2	Media	0.220	0.219	0.254	0.251	0.248
		Mediana	0.175	0.174	0.204	0.196	0.193
		IQR	0.188	0.187	0.226	0.219	0.210
		SD	0.183	0.181	0.195	0.201	0.198
		$P(\alpha < 0.05)$	9%	9%	6%	6%	6%
	0.5	Media	0.072	0.072	0.074	0.076	0.075
		Mediana	0.059	0.058	0.060	0.063	0.062
		IQR	0.057	0.058	0.064	0.064	0.064
		SD	0.057	0.055	0.054	0.055	0.055
		$P(\alpha < 0.05)$	41%	42%	41%	40%	40%
300	0.2	Media	0.060	0.060	0.092	0.083	0.083
		Mediana	0.048	0.048	0.073	0.063	0.063
		IQR	0.050	0.050	0.080	0.074	0.073
		SD	0.048	0.048	0.068	0.066	0.066
		$P(\alpha < 0.05)$	53%	53%	31%	38%	38%
	0.5	Media	0.021	0.021	0.023	0.023	0.023
		Mediana	0.018	0.018	0.019	0.019	0.019
		IQR	0.018	0.017	0.019	0.020	0.020
		SD	0.016	0.016	0.016	0.016	0.016
		$P(\alpha < 0.05)$	95%	95%	94%	94%	94%

Tabla 7-5. Resultados de la distribución del ángulo formado entre los vectores de cointegración estimado y real. 1.000 simulaciones del proceso PGD3, para distintos procedimientos de estimación y valores de sigma y T, con $a = 1$, $\varphi_1 = .5$, $\varphi_2 = .5$, $\theta = .5$. En negrita los mejores valores por fila.

Resultados

T	Sigma	Estadístico	Método				
			ML_b	ML	ACCA	CCA	CCA2
50	0.2	Media	0.828	0.823	0.713	0.697	0.695
		Mediana	0.798	0.795	0.653	0.622	0.617
		IQR	0.580	0.578	0.515	0.504	0.495
		SD	0.372	0.375	0.360	0.360	0.358
		$P(\alpha < 0.05)$	0%	0%	0%	0%	0%
	0.5	Media	0.491	0.475	0.314	0.323	0.317
		Mediana	0.376	0.366	0.258	0.262	0.254
		IQR	0.427	0.418	0.244	0.244	0.235
		SD	0.369	0.363	0.237	0.251	0.245
		$P(\alpha < 0.05)$	2%	2%	3%	3%	3%
100	0.2	Media	0.613	0.609	0.490	0.473	0.468
		Mediana	0.522	0.517	0.427	0.416	0.416
		IQR	0.474	0.473	0.340	0.308	0.315
		SD	0.364	0.363	0.293	0.279	0.278
		$P(\alpha < 0.05)$	0%	0%	1%	1%	1%
	0.5	Media	0.331	0.323	0.180	0.185	0.182
		Mediana	0.224	0.220	0.150	0.160	0.155
		IQR	0.274	0.264	0.146	0.139	0.139
		SD	0.316	0.314	0.128	0.127	0.125
		$P(\alpha < 0.05)$	4%	4%	9%	9%	9%
300	0.2	Media	0.285	0.284	0.241	0.231	0.230
		Mediana	0.211	0.211	0.201	0.197	0.196
		IQR	0.245	0.246	0.203	0.190	0.186
		SD	0.243	0.244	0.160	0.149	0.148
		$P(\alpha < 0.05)$	6%	5%	4%	4%	4%
	0.5	Media	0.110	0.107	0.072	0.076	0.076
		Mediana	0.073	0.072	0.062	0.065	0.066
		IQR	0.082	0.081	0.058	0.063	0.062
		SD	0.124	0.121	0.049	0.051	0.051
		$P(\alpha < 0.05)$	31%	33%	39%	36%	36%

Tabla 7-6. Resultados de la distribución del ángulo formado entre los vectores de cointegración estimado y real. 1.000 simulaciones del proceso PGD3, para distintos procedimientos de estimación y valores de sigma y T, con $a = 1$, $\varphi_1 = .9$, $\varphi_2 = .9$, $\theta = .9$. En negrita los mejores valores por fila.

Análisis de Sistemas Cointegrados mediante Métodos de Subespacios

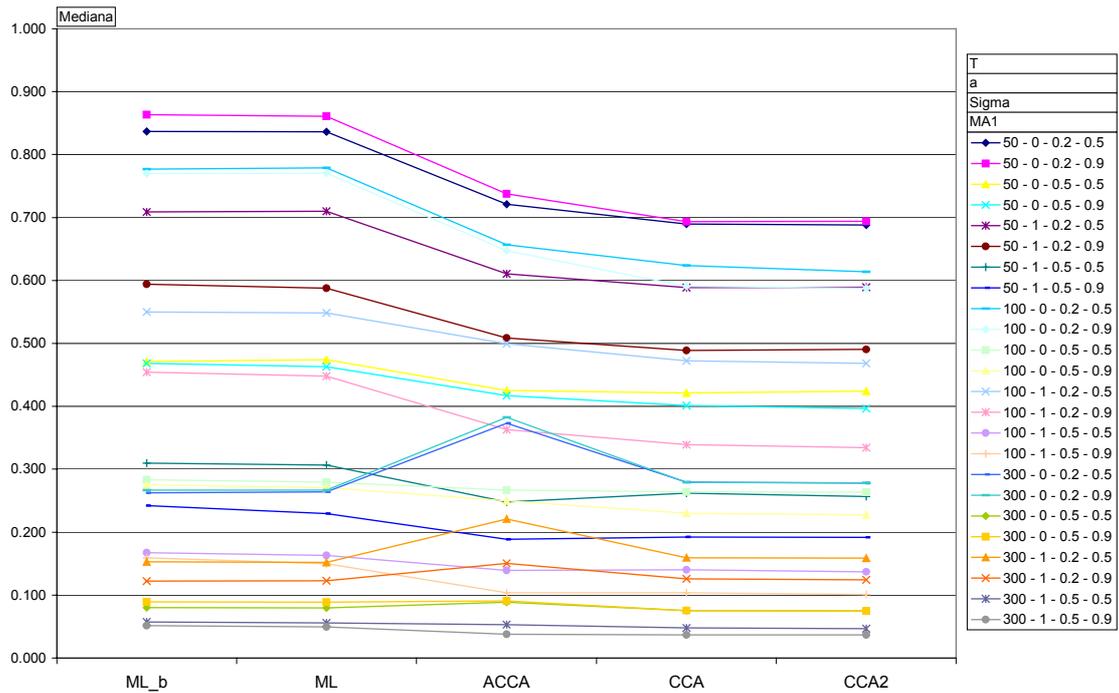


Figura 7-34. Mediana de la distribución del ángulo formado entre los subespacios de cointegración estimado y real. Procedimientos de estimación ML_b, ML, ACCA, CCA, y CCA2. Para distintas combinaciones de valores de los parámetros del proceso generador PGD3, con $\phi_1 = .5$ y $\phi_2 = .9$

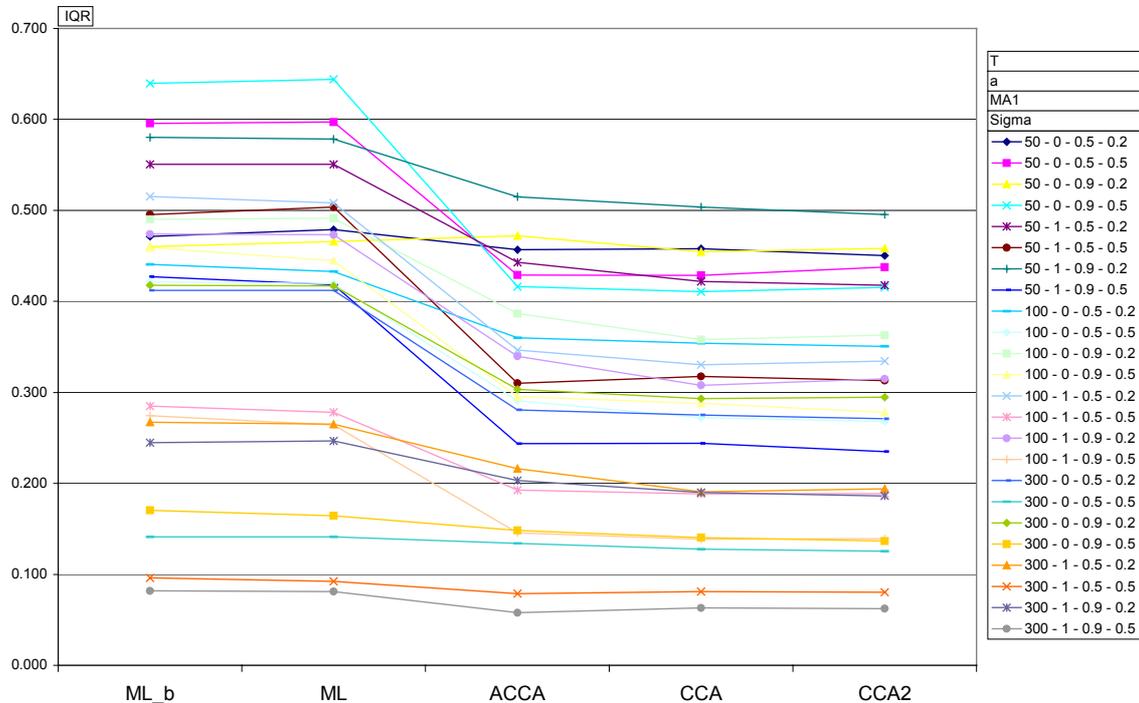


Figura 7-35. Rango intercuartílico (IQR) de la distribución del ángulo formado entre los subespacios de cointegración estimado y real. Procedimientos de estimación ML_b, ML, ACCA, CCA, y CCA2. Para distintas combinaciones de valores de los parámetros del proceso PGD3, con $\phi_1 = .9$ y $\phi_2 = .9$

Resultados

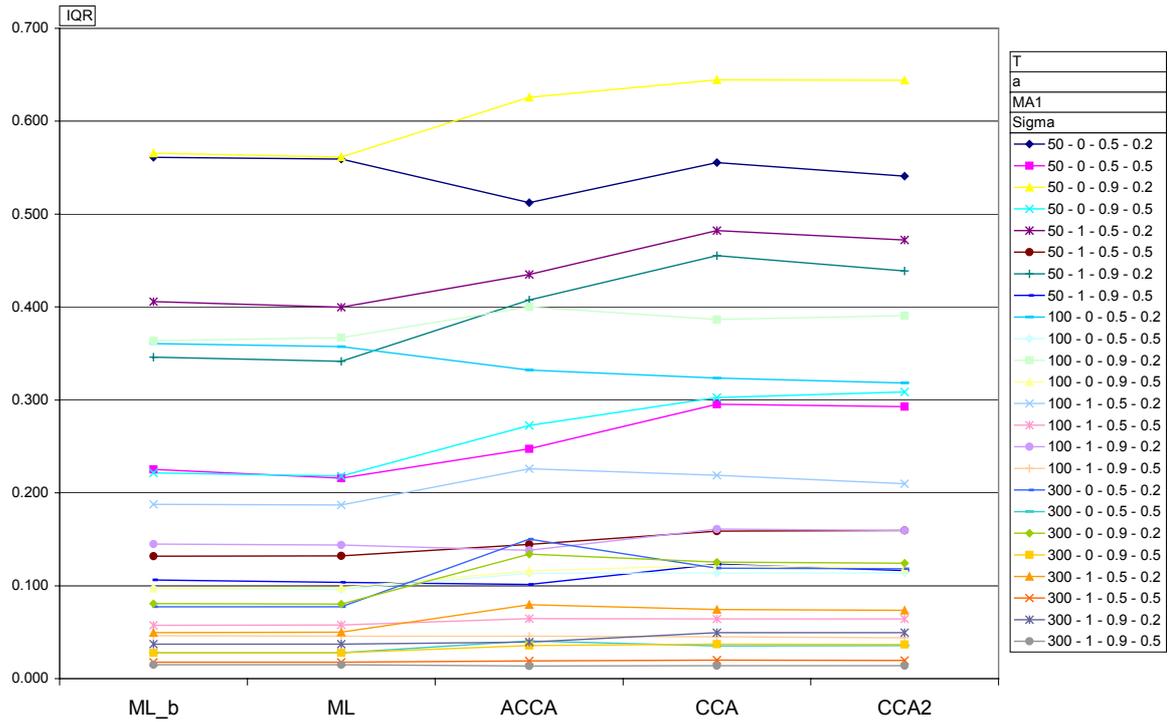


Figura 7-36. Rango intercuartílico (IQR) de la distribución del ángulo formado entre los subespacios de cointegración estimado y real. Procedimientos de estimación ML_b, ML, ACCA, CCA, y CCA2. Para distintas combinaciones de valores de los parámetros del proceso PGD3, con $\phi_1 = .5$ y $\phi_2 = .5$

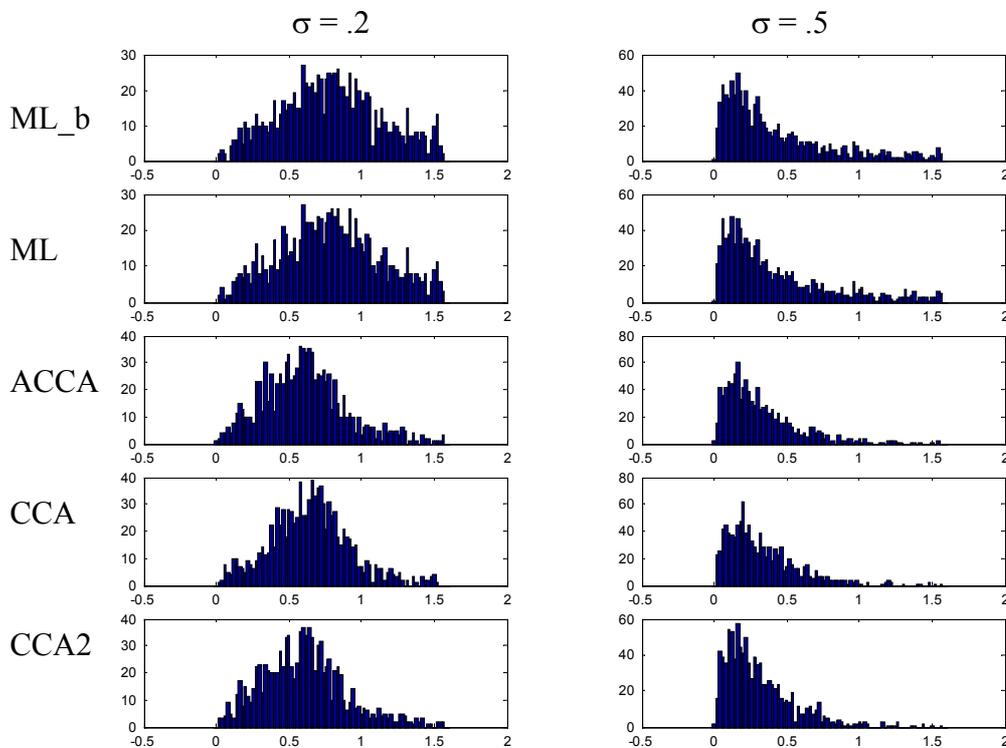


Figura 7-37. Histogramas del ángulo que separa el subespacio de cointegración estimado del real. 1000 realizaciones del proceso PGD3. $T = 100$, $\sigma = (.2, .5)$, $a = 0$, $\phi_1 = .5$, $\phi_2 = .9$, $\theta = .9$

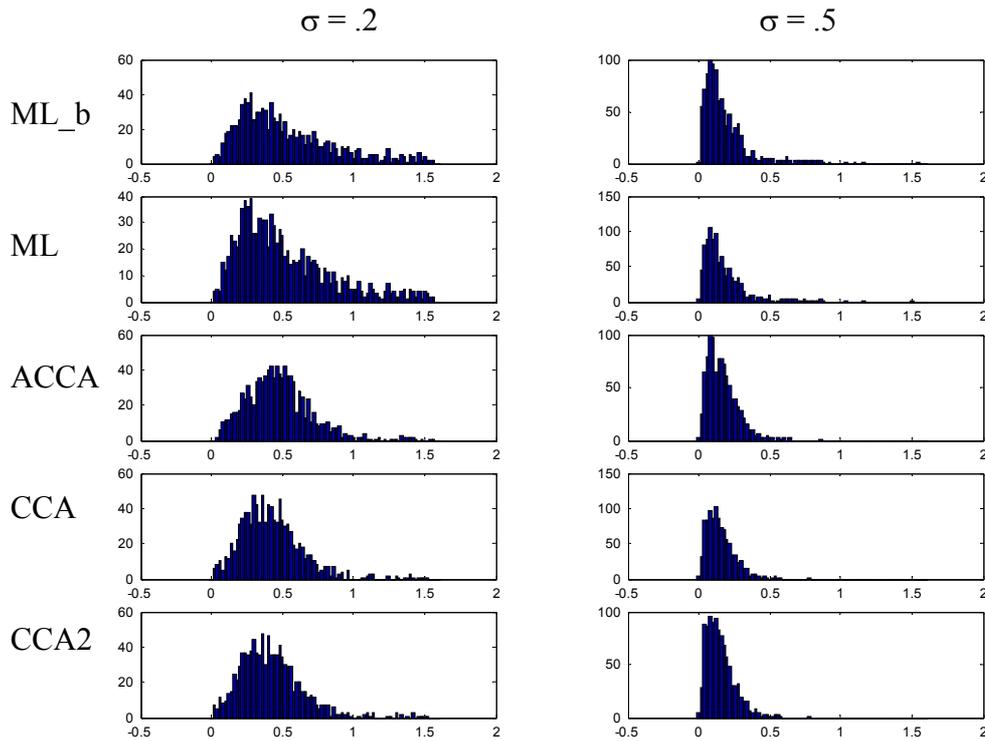


Figura 7-38. Histogramas del ángulo que separa el subespacio de cointegración estimado del real. 1000 realizaciones del proceso PGD3. $T = 300$, $\sigma = (.2, .5)$, $a = 0$, $\varphi_1 = .9$, $\varphi_2 = .9$, $\theta = .9$

7.3.3 Error de predicción a corto plazo

Las condiciones del experimento son las mismas que se indicaron para el proceso PGD2, utilizando, para cada combinación de valores de parámetros del proceso generador PGD3, una muestra de validación de tamaño igual al 10% del tamaño de la muestra de trabajo, y calculando, para cada modelo, el valor del estadístico U de Theil y su distribución.

La predicción se realiza expresando el modelo en el espacio de estados. Como estado inicial para la muestra de predicción se utiliza el estado final estimado en la muestra de trabajo (o de identificación). Para la muestra de trabajo se utiliza un estado inicial suavizado. En general, el efecto de tomar otro estimador del estado inicial iría desapareciendo a lo largo de la muestra de trabajo, como se observa en la Figura 7-39.

Resultados

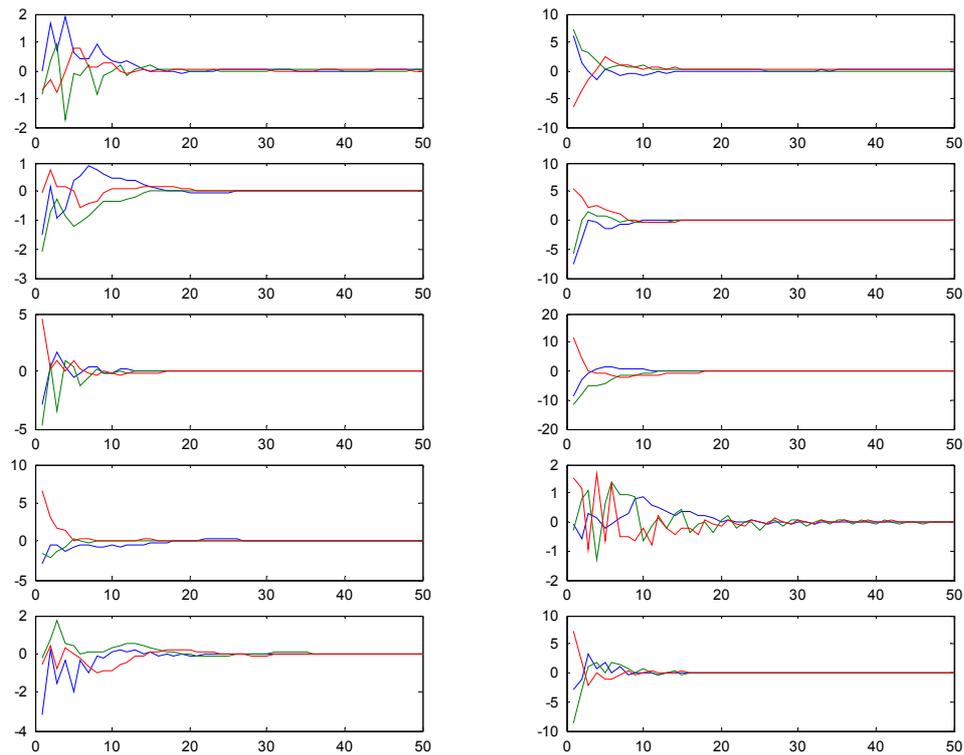


Figura 7-39. *Diferencias en la predicción a horizonte uno de tres series generadas mediante el proceso PGD3. Las diferencias corresponden a dos estimaciones distintas del estado inicial: una estimación suavizada y una estimación nula. En las 10 realizaciones el modelo usado para predicción se estimó por el procedimiento CCA.*

La Tabla 7-7 y los gráficos que van de la Figura 7-40 a la Figura 7-46 muestran varios de los resultados obtenidos. Se observa que:

- La estimación ML_b y ML resultan equivalentes. En general, también resultan equivalentes ACCA y CCA simple.
- El método CCA2 resulta, con diferencia, el peor método para la predicción. En muchas ocasiones, los modelos identificados por este procedimiento predicen peor que el modelo “camino aleatorio”. El procedimiento funciona especialmente mal para valores bajos del parámetro señal-ruido (σ).
- Si comparamos ML con los métodos de subespacios, observamos que para tamaño muestral $T = 50$, los modelos estimados por subespacios ofrecen resultados muy variables, y, en general, predicen mejor los modelos estimados

por ML. Sin embargo, para tamaños muestrales de 100 y 300, los métodos CCA y ACCA suelen proporcionar estadísticos U de Theil menores (mejores resultados de predicción a corto plazo) que ML, si bien el estadístico suele tener dispersión algo mayor. La ventaja de los métodos de subespacios disminuye al aumentar el horizonte de predicción.

- Los resultados dependen de los parámetros del proceso generador y de la serie particular que se esté prediciendo ($y_{1,t}$, $y_{2,t}$ ó $y_{3,t}$), por lo que resulta difícil encontrar pautas de validez global.
- Pequeñas desviaciones de la hipótesis de normalidad de las perturbaciones no afectan a los resultados anteriores.

Resultados

T	Sigma	Estadístico	Método				
			ML_b	ML	ACCA	CCA	CCA2
50	0.2	Media	0.894	0.880	1.057	1.011	1.878
		Mediana	0.612	0.608	0.657	0.636	0.675
		IQR	0.611	0.585	0.735	0.688	0.894
		SD	0.992	0.953	1.387	1.304	12.169
		P(U < 1)	75%	76%	71%	71%	66%
	0.5	Media	1.006	0.984	1.169	1.196	1.520
		Mediana	0.629	0.619	0.709	0.707	0.777
		IQR	0.687	0.683	0.850	0.864	0.997
		SD	1.637	1.587	1.608	1.711	2.859
		P(U < 1)	72%	72%	65%	66%	62%
100	0.2	Media	0.660	0.658	0.606	0.589	0.756
		Mediana	0.555	0.557	0.519	0.503	0.606
		IQR	0.346	0.345	0.315	0.293	0.379
		SD	0.394	0.391	0.333	0.307	0.546
		P(U < 1)	87%	87%	90%	92%	83%
	0.5	Media	0.653	0.651	0.609	0.591	0.776
		Mediana	0.554	0.553	0.516	0.498	0.622
		IQR	0.378	0.378	0.344	0.333	0.433
		SD	0.387	0.382	0.368	0.345	0.520
		P(U < 1)	88%	87%	90%	91%	80%
300	0.2	Media	0.520	0.520	0.486	0.484	0.562
		Mediana	0.502	0.502	0.469	0.469	0.527
		IQR	0.172	0.174	0.147	0.146	0.188
		SD	0.138	0.138	0.122	0.121	0.171
		P(U < 1)	99%	99%	100%	100%	98%
	0.5	Media	0.538	0.538	0.511	0.502	0.620
		Mediana	0.521	0.521	0.497	0.487	0.590
		IQR	0.187	0.191	0.168	0.164	0.204
		SD	0.151	0.151	0.137	0.136	0.178
		P(U < 1)	99%	99%	99%	99%	96%

Tabla 7-7. Resultados de la distribución del estadístico U de Theil en predicción de la serie y_1 a horizonte 1 para 1.000 simulaciones del proceso PGD3, para distintos procedimientos de estimación del modelo y distintos valores de sigma y T, con $a=1$, $\varphi_1 = .5$, $\varphi_2 = .5$, $\theta = .9$. Todos los estimadores restan las medias salvo ML_b. En negrita los mejores valores de cada fila.

Análisis de Sistemas Cointegrados mediante Métodos de Subespacios

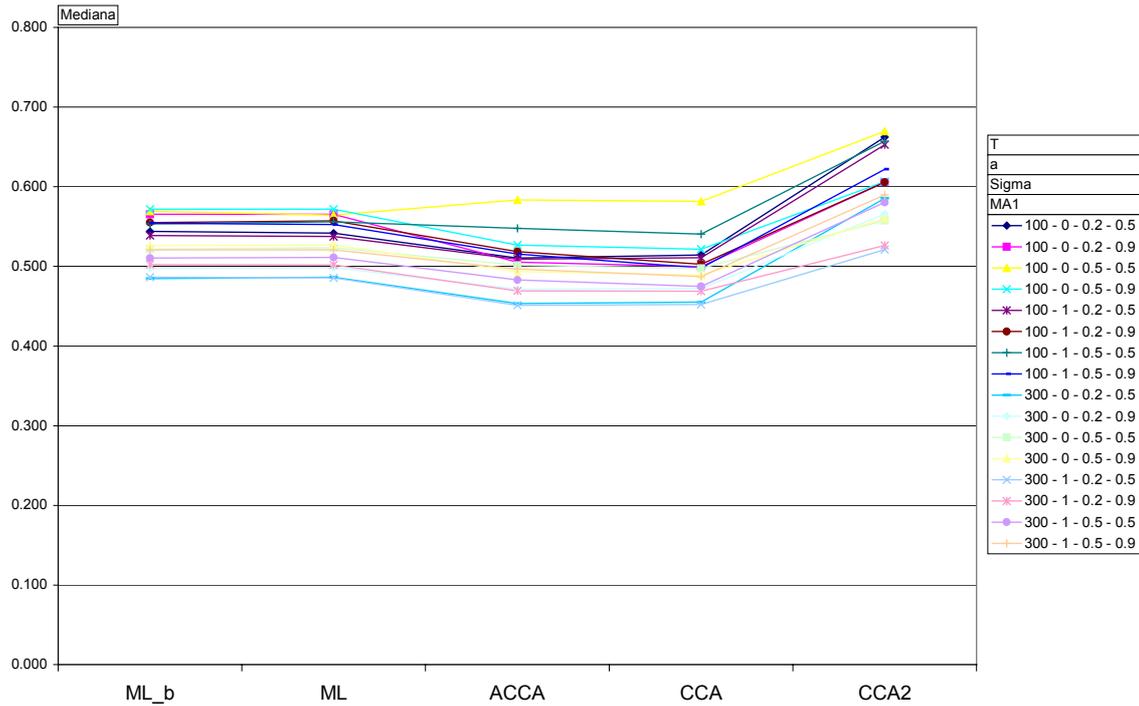


Figura 7-40. Mediana de la distribución del estadístico U de Theil en predicción de la serie y_1 a horizonte 1, para modelos identificados mediante distintos procedimientos y con distintas combinaciones de valores de los parámetros del proceso generador PGD3, con $T=100, 300$, $\phi_1 = .5$ y $\phi_2 = .5$

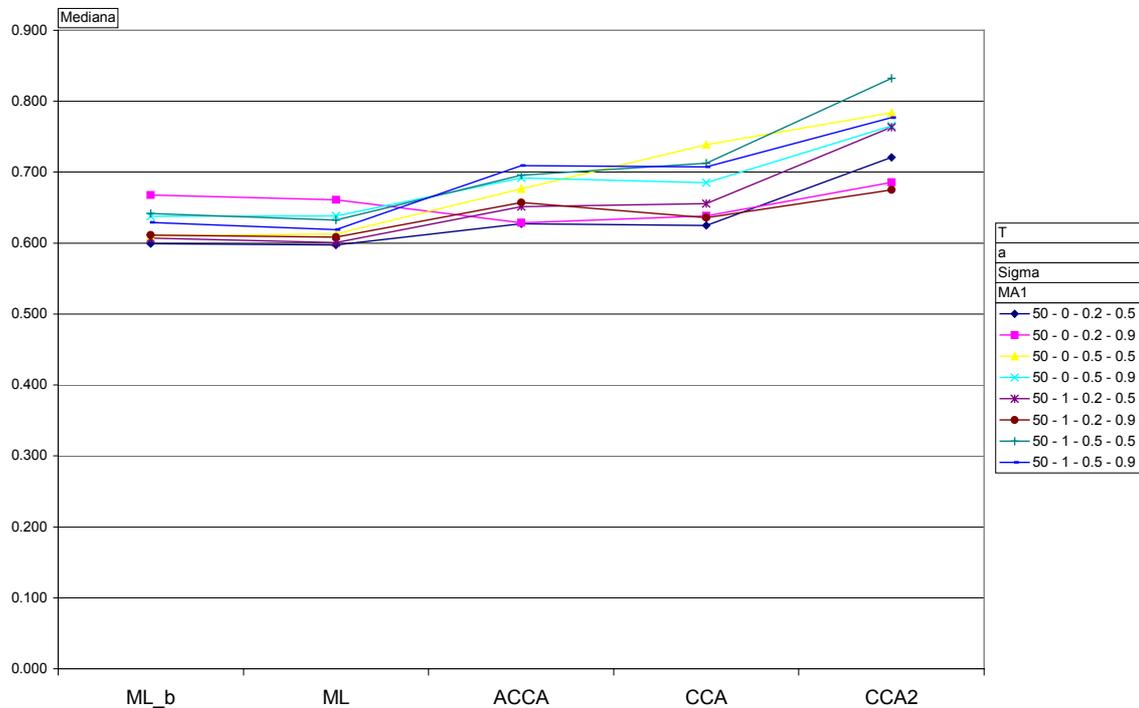


Figura 7-41. Mediana de la distribución del estadístico U de Theil en predicción de la serie y_1 a horizonte 1, para modelos identificados mediante distintos procedimientos y con distintas combinaciones de valores de los parámetros del proceso generador PGD3, con $T=50$, $\phi_1 = .5$ y $\phi_2 = .5$

Resultados

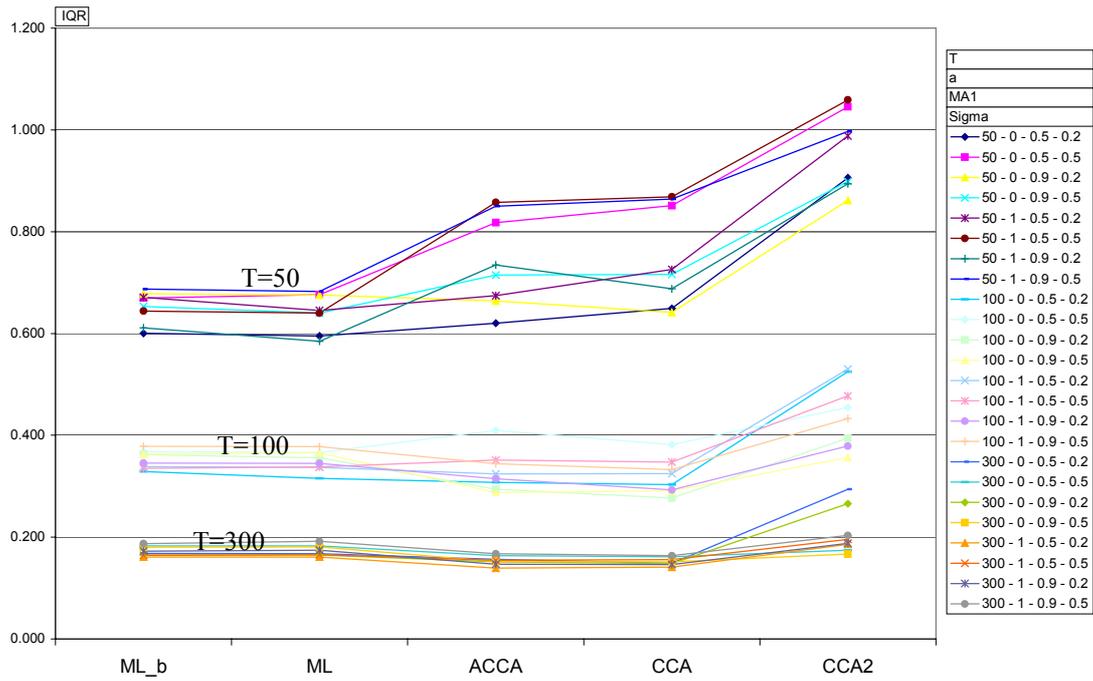


Figura 7-42. Rango intercuartílico (IQR) de la distribución estadístico U de Theil en predicción de la serie y_1 a horizonte 1, para modelos identificados mediante distintos procedimientos y para distintas combinaciones de valores de los parámetros del proceso generador PGD3, con $\varphi_1 = .5$ y $\varphi_2 = .5$

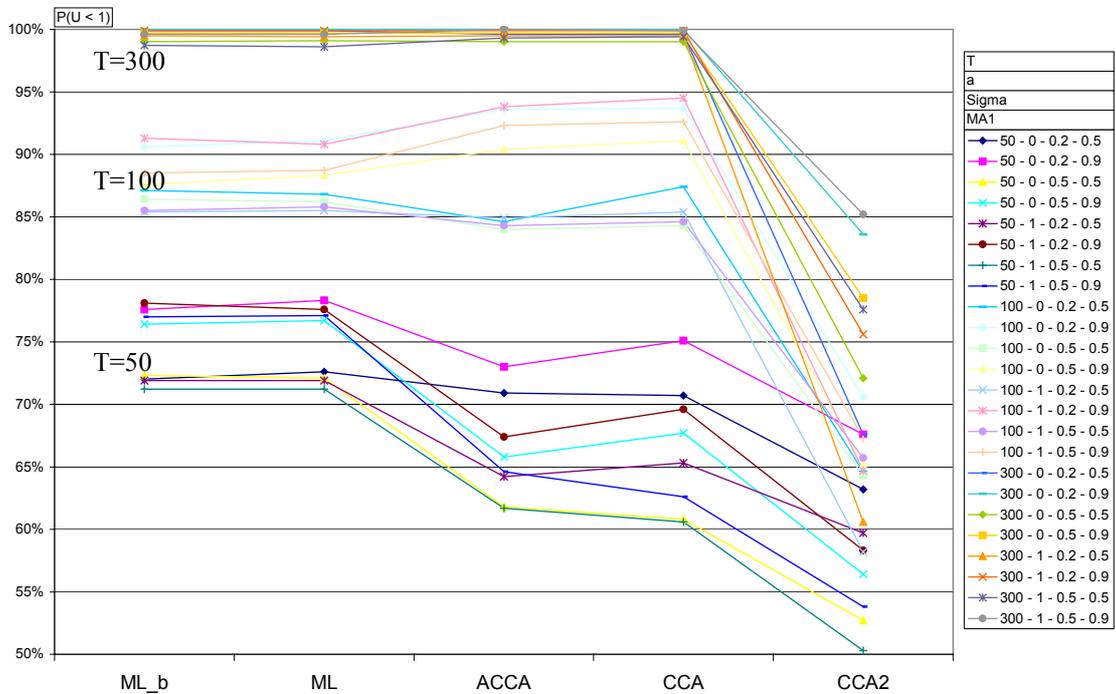


Figura 7-43. Probabilidad de que el estadístico U de Theil en predicción de la serie y_3 a horizonte 1 sea menor que uno, para modelos identificados mediante distintos procedimientos, y para distintas combinaciones de valores de los parámetros del proceso PGD3, con $\varphi_1 = .5$ y $\varphi_2 = .5$

Análisis de Sistemas Cointegrados mediante Métodos de Subespacios

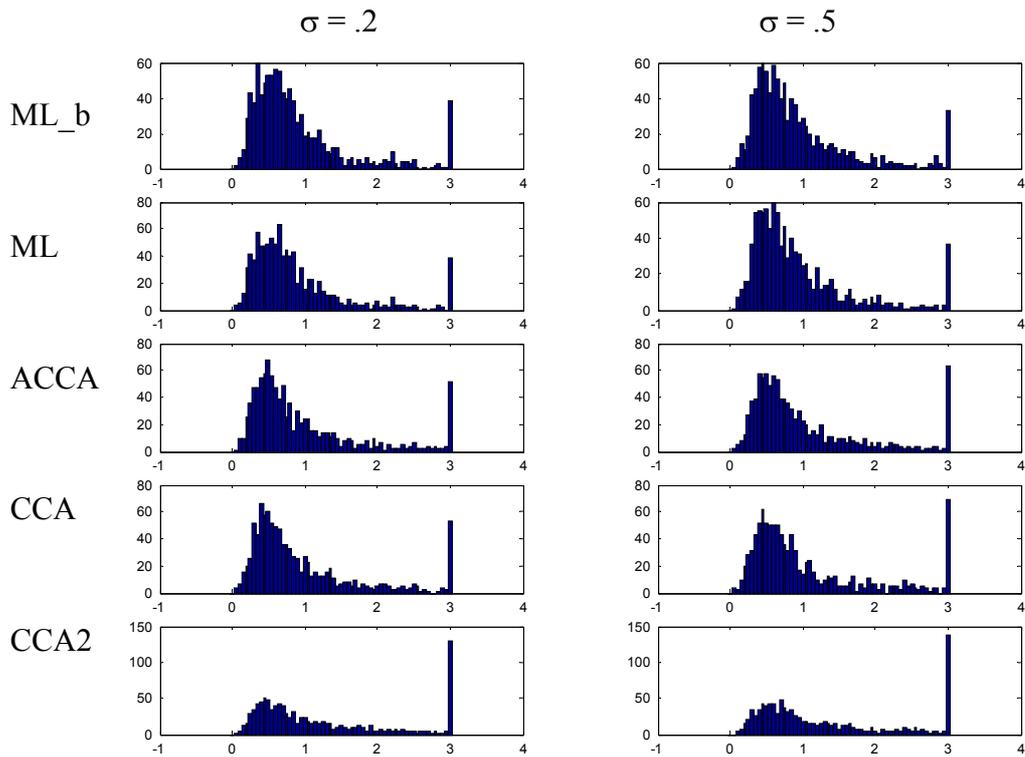


Figura 7-44. Histogramas del estadístico U de Theil en predicción de la serie y_2 a horizonte 1. Proceso PGD3. $T = 50$, $\sigma = (.2, .5)$, $a = 0$, $\varphi_1 = .5$, $\varphi_2 = .5$, $\theta = .9$

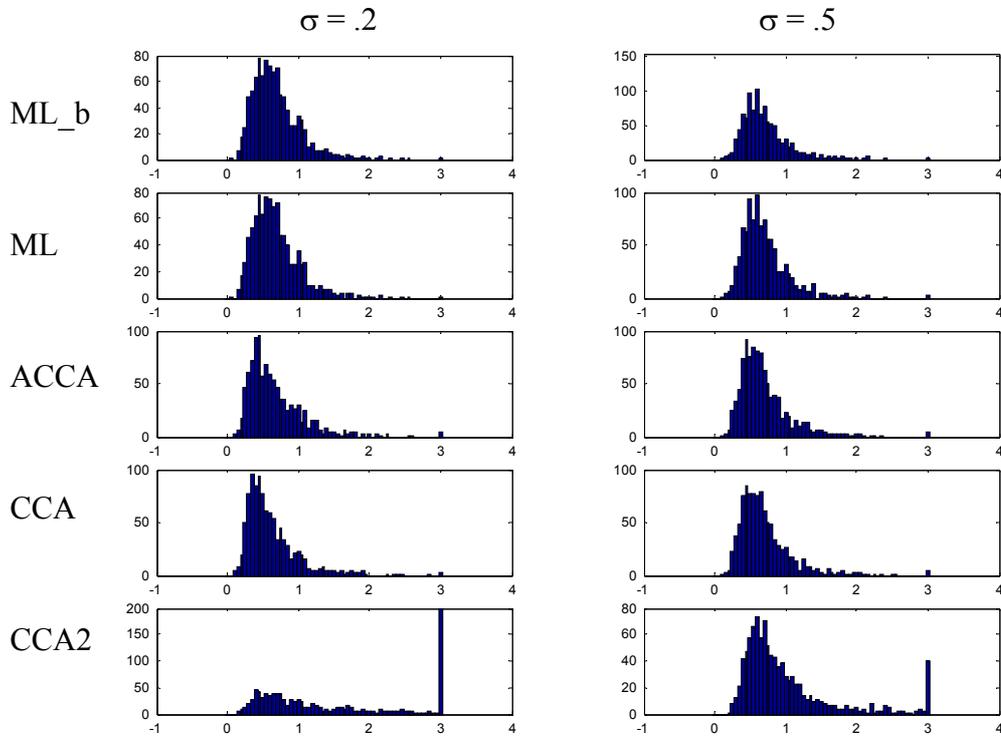


Figura 7-45. Histogramas del estadístico U de Theil en predicción de la serie y_2 a horizonte 1. Proceso PGD3. $T = 100$, $\sigma = (.2, .5)$, $a = 0$, $\varphi_1 = .5$, $\varphi_2 = .5$, $\theta = .5$

Resultados

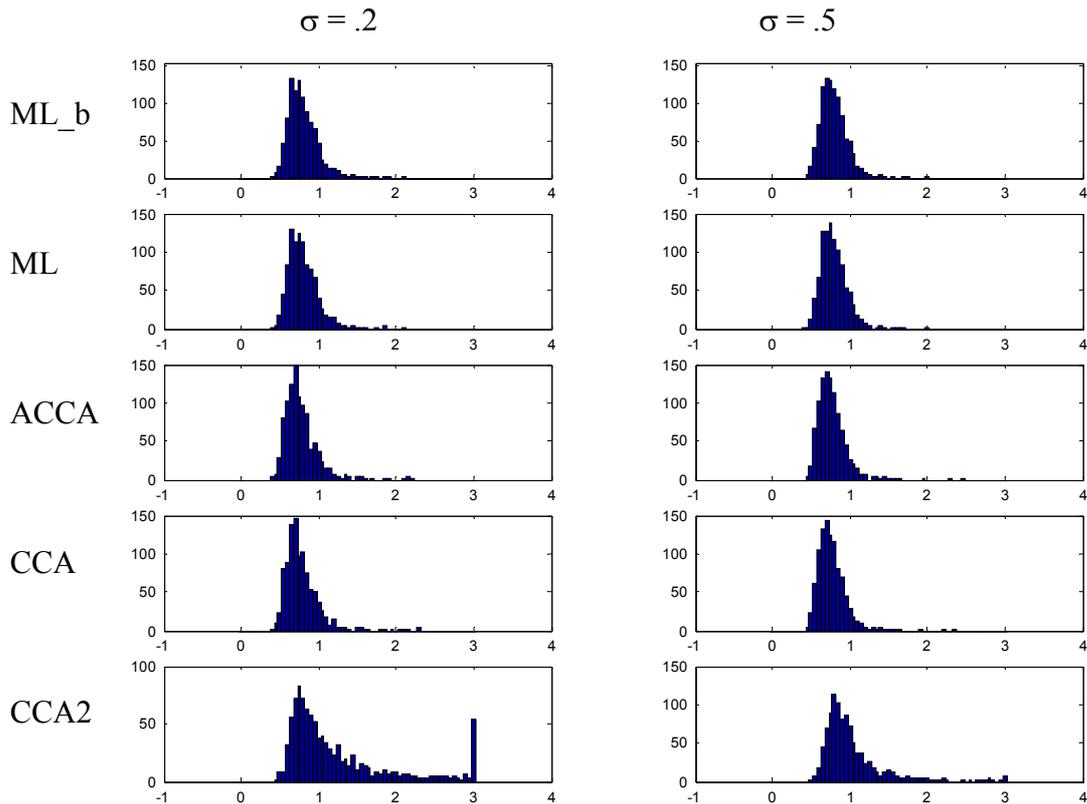


Figura 7-46. Histogramas del estadístico U de Theil en predicción de la serie y_1 a horizonte 1. Proceso PGD3. $T = 300$, $\sigma = (.2, .5)$, $a = 0$, $\varphi_1 = .9$, $\varphi_2 = .9$, $\theta = .9$

7.3.4 Contraste de igualdad de capacidad predictiva

Como hicimos con el proceso generador anterior, nos planteamos si es significativo el número de veces que el modelo identificado por un procedimiento supera en capacidad predictiva (error cuadrático) al modelo identificado por otro procedimiento.

Recordemos (apartado 7.2.4) que, bajo la hipótesis de equivalencia de los dos procedimientos de identificación (A y B), la variable $n =$ "número de veces que el modelo del método A supera al modelo del método B en N repeticiones independientes" seguiría una distribución binomial de parámetro $p = 1/2$, lo que nos permite realizar contrastes de igualdad de métodos, a efectos de capacidad predictiva de los modelos identificados. El valor crítico superior en 1.000 repeticiones elegido (al 99%) es de 541 (459 como valor inferior).

Análisis de Sistemas Cointegrados mediante Métodos de Subespacios

Tomando como referencia el método CCA simple, la Tabla 7-10 muestra, para distintas combinaciones de valores de parámetros del proceso generador PGD3, el número de veces (de entre 1.000 repeticiones) en que los modelos identificados por los métodos de Johansen, ACCA o CCA2 proporcionaron predicciones de menor error cuadrático medio que el modelo correspondiente estimado por el algoritmo CCA simple.

T	φ_1	φ_2	θ	σ					
				0,2			0,5		
				Johansen	ACCA	CCA2	Johansen	ACCA	CCA2
50	0.5	0.5	0.5	532	520	417	584	549	466
			0.9	502	490	446	517	482	447
		0.9	0.5	554	508	420	585	510	443
			0.9	504	469	438	524	469	422
	0.9	0.5	0.5	611	568	456	644	554	482
			0.9	522	549	422	563	498	436
		0.9	0.5	606	526	418	650	521	465
			0.9	567	513	412	589	538	434
100	0.5	0.5	0.5	412	516	239	525	435	328
			0.9	302	464	287	340	417	291
		0.9	0.5	475	444	197	534	421	255
			0.9	332	402	230	311	375	218
	0.9	0.5	0.5	430	573	318	566	489	454
			0.9	394	521	345	544	496	459
		0.9	0.5	488	490	271	579	494	374
			0.9	423	467	344	523	449	343
300	0.5	0.5	0.5	248	554	115	281	397	183
			0.9	279	510	197	272	394	158
		0.9	0.5	316	426	56	301	369	79
			0.9	398	407	53	294	322	107
	0.9	0.5	0.5	279	588	141	350	443	255
			0.9	253	498	143	279	430	175
		0.9	0.5	343	485	82	392	406	172
			0.9	358	531	142	300	447	142

Tabla 7-8. Número de veces entre 1.000 en que el modelo identificado por los métodos de Johansen, ACCA o CCA2 bate en error cuadrático medio de predicción al modelo identificado por el método CCA simple. Se considera significativo a favor de CCA un valor inferior a 459 (celdas en negrita), y significativo a favor del método en cuestión un valor superior a 541 (celdas con relleno). La tabla se ha realizado con el error de predicción de la serie y_1 a horizonte 1, con $a = 0$ y perturbaciones gaussianas.

En la Tabla 7-10, que resulta representativa de los resultados generales, se aprecia que el método de Johansen es el que proporciona los mejores modelos para bajo tamaño muestral ($T = 50$). Este resultado se mantiene para las otras series y otros valores de los

parámetros. En general, para tamaño muestral $T = 100$ y $T = 300$, el método CCA simple es el que proporciona los mejores modelos predictivos. El método CCA2 es notablemente inferior a CCA simple. ACCA proporciona resultados variables, en algunos casos supera a CCA simple, pero suele resultar inferior, en especial para $T = 300$.

7.3.5 Combinación de predicciones

Pese a los malos resultados de predicción obtenidos por los modelos identificados mediante el método CCA2, nos planteamos si estos modelos pueden aportar información útil para predicción, al combinarlos con modelos identificados por otros procedimientos.

Puesto que el método CCA2 es una modificación del método CCA, que trata de mejorarlo en sistemas cointegrados, parece lógico realizar, en primer lugar, una comparación entre estos dos procedimientos. La Figura 7-47 muestra la distribución de pesos óptimos para predicción en combinaciones lineales de modelos CCA con modelos CCA2 (ver apartado 7.2.4). Un valor del peso próximo a uno (/cero) indica que el modelo CCA (/CCA2) proporciona la mejor predicción y el otro no aporta información útil para predicción.

Se comprueba que:

- La mejor predicción corresponde, en general, a uno de los modelos, y no a una combinación de ambos.
- Para $T = 50$ los dos métodos (CCA y CCA2) están básicamente empatados en cuanto a probabilidad de proporcionar el mejor modelo en predicción. La ventaja del método CCA sobre el CCA2 se va haciendo notable al aumentar el tamaño muestral y es en general muy evidente (depende de los valores de los parámetros del proceso generador) para tamaño muestral $T = 300$. Para este tamaño muestral sí existen algunos casos, aunque reducidos, en que la combinación de modelos CCA2-CCA supera, en predicción, al modelo CCA simple.

Análisis de Sistemas Cointegrados mediante Métodos de Subespacios

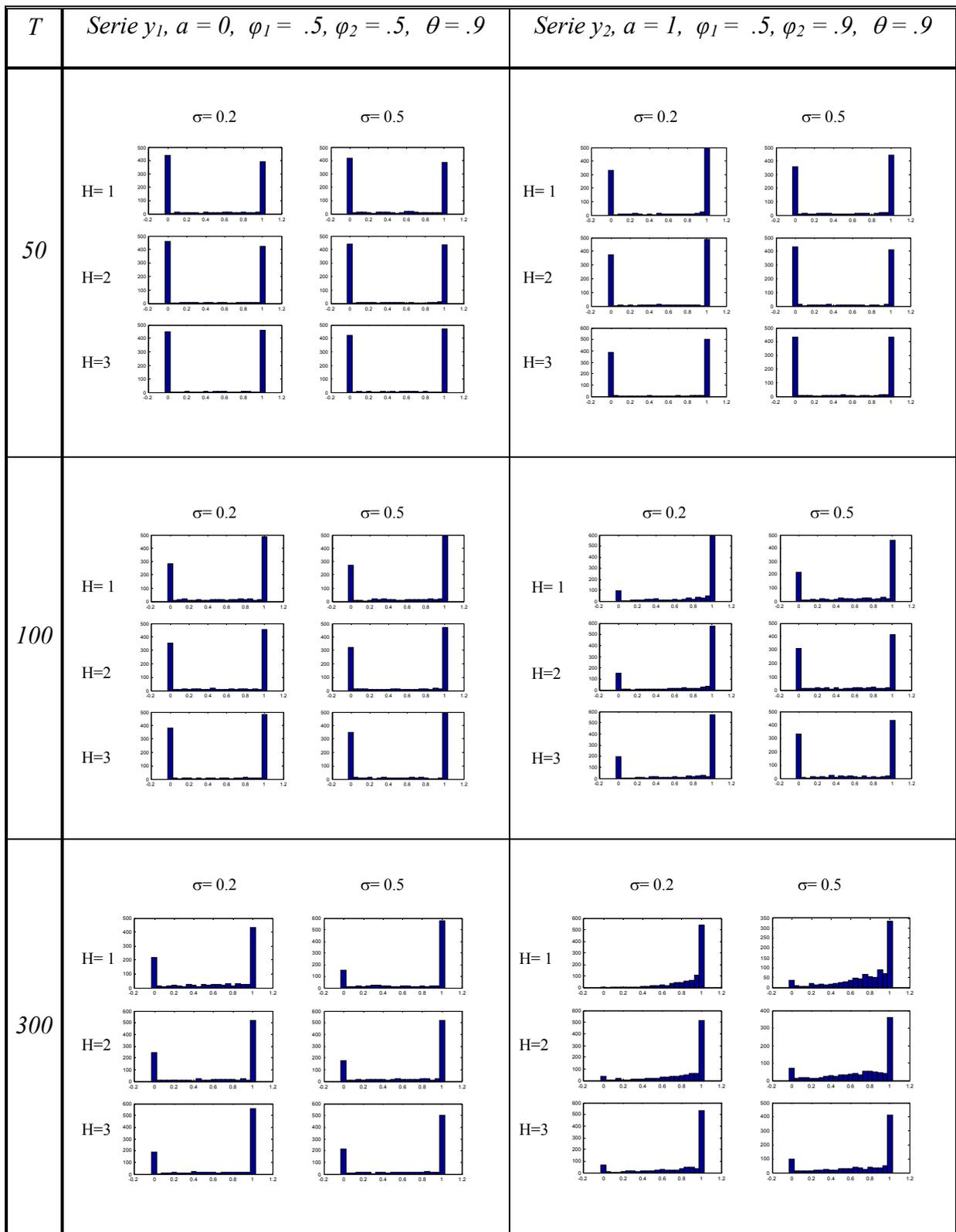


Figura 7-47. Histogramas de los pesos ω^* de modelos CCA en combinaciones con modelos CCA2 para predicción a horizonte H .

7.4 Proceso generador de datos 4 (PGD4)

7.4.1 Características generales

El proceso generador de datos 4 (PGD4) genera tres series temporales como combinación de 2 tendencias comunes (lo que implica un vector de cointegración) y tres ciclos, donde uno de los ciclos sigue un AR(1) y los otros dos recuerdan perturbaciones anteriores (componentes MA).

Una formulación en el espacio de estados para el PGD4 es

$$\begin{bmatrix} \tau_1 \\ \tau_2 \\ \eta_1 \\ \eta_2 \\ \eta_3 \end{bmatrix}_{t+1} = \begin{bmatrix} 1 & 0 & 0 & a & 0 \\ 0 & 1 & 0 & 0 & a \\ 0 & 0 & \varphi_1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \tau_1 \\ \tau_2 \\ \eta_1 \\ \eta_2 \\ \eta_3 \end{bmatrix}_t + \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ \theta_1 & 0 & 0 \\ 0 & \theta_2 & 0 \end{bmatrix} \begin{bmatrix} e_1 \\ e_2 \\ e_3 \end{bmatrix}_t$$

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}_t = \begin{bmatrix} 1 & 1 & | & 1 & 0 & 0 \\ 0 & 1 & | & 0 & 1 & 0 \\ 1 & 0 & | & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \sigma \tau_1 \\ \sigma \tau_2 \\ \eta_1 \\ \eta_2 \\ \eta_3 \end{bmatrix}_t + \begin{bmatrix} e_1 \\ e_2 \\ e_3 \end{bmatrix}_t$$

donde el vector de perturbaciones \mathbf{e}_t es ruido blanco gaussiano de esperanza cero y matriz de covarianzas la identidad.

La relación de cointegración pertenece al subespacio ortogonal a los dos primeros vectores de la matriz \mathbf{C} , por lo esta relación viene definida por el vector $[1, -1, -1]'$.

Para incluir términos constantes en las relaciones de cointegración, se ha sumado un valor fijo δ_1 a todos los valores de una de las series y un valor fijo δ_2 a todos los valores de otra de las series (desplazamientos en vertical de todos los valores de una de las series con respecto a las otras). Si se restan las medias muestrales, la magnitud del valor de los desplazamientos no influye en los resultados.

Análisis de Sistemas Cointegrados mediante Métodos de Subespacios

El espacio paramétrico considerado es $(T \times \sigma \times a \times \varphi_1 \times \theta_1 \times \theta_2 \times \delta_1 \times \delta_2)$. A partir de las consideraciones realizadas en los procesos anteriores, tomamos los valores $T = (50, 100, 300)$, $\sigma = (.2, .5)$, $a = (0, 1)$, $\varphi_1 = (.5, .9)$, $\theta_1 = (.5, .9)$, $\theta_2 = (.5, .9)$, $\delta_1 = 50$ y $\delta_2 = -30$.

La Figura 7-48 representa un conjunto de realizaciones del proceso generador PGD4 que corresponde a distintas combinaciones de valores de los parámetros.

Realizaciones PGD4

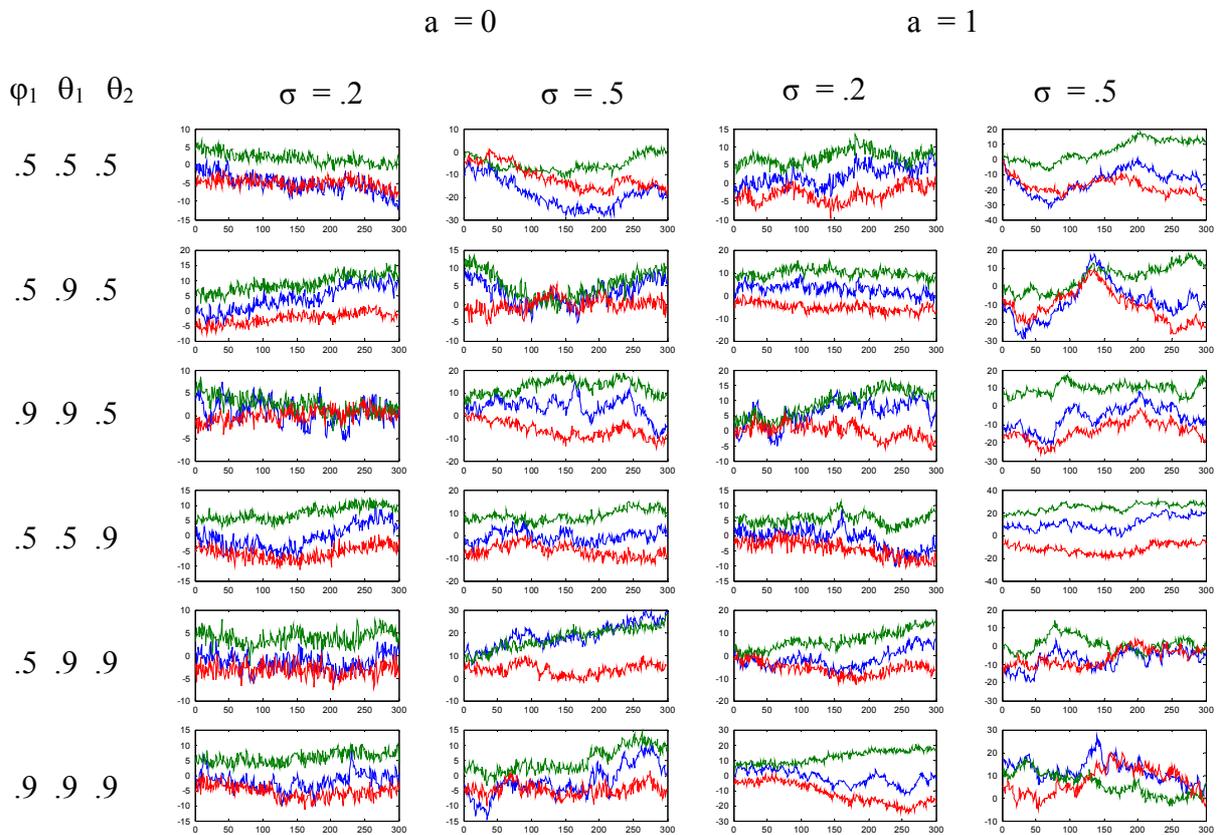


Figura 7-48. Realizaciones del proceso PGD4 con componente constante en la relación de cointegración.

Para distintas combinaciones de valores de los parámetros, con desplazamientos $\delta_1 = 50$ y $\delta_2 = -30$.

Resultados

Ajuste de los algoritmos

Como en los casos anteriores, el orden p de la representación VAR del vector de series temporales y_t se ha seleccionado minimizando el criterio de información BIC. Para los métodos de subespacios se tomaron valores de los parámetros $p = f = 5$. El orden n de la representación en el espacio de estados se calculó por el procedimiento BA(n) de Bauer y Wagner (sección 6.6).

La Figura 7-56 muestra un resultado representativo (típico) de la estimación del orden del modelo con los procedimientos CCA2 y ACCA para el PGD4.

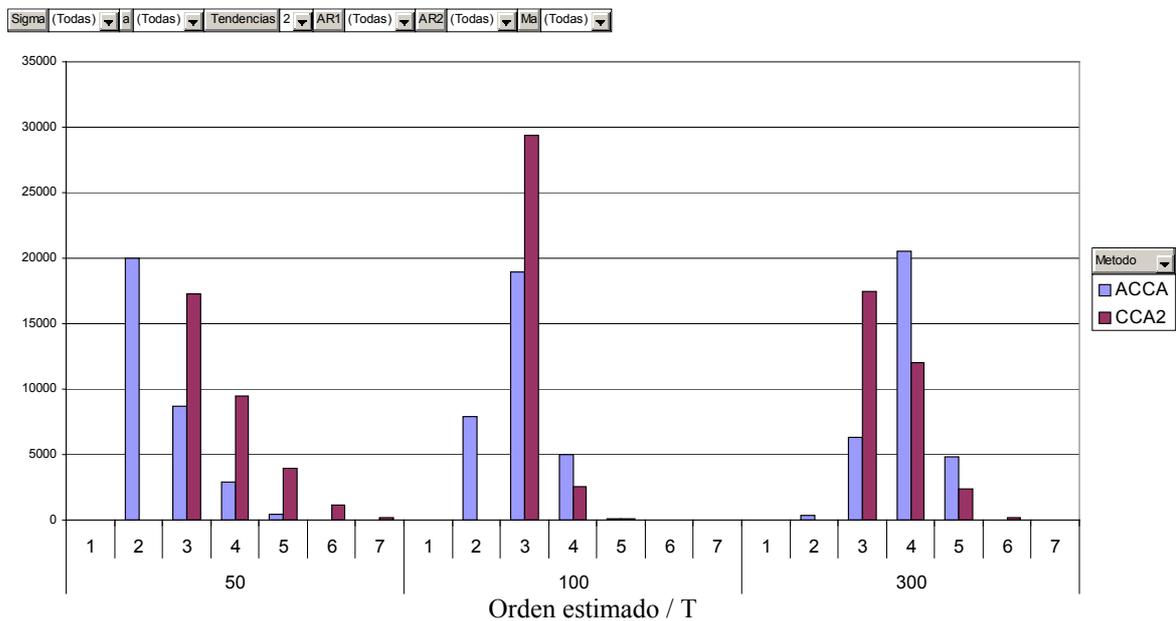


Figura 7-49. Histogramas del orden estimado (de modelos en el espacio de estados) por los procedimientos ACCA y CCA2, con el criterio BA(n), para distintos tamaños muestrales T.

Como en los casos anteriores, al aumentar T se va reduciendo la variabilidad en la estimación del orden del modelo (número de estados), y el orden estimado se va aproximando más al verdadero orden del proceso generador (orden 5), si bien en general el orden resulta subestimado, incluso para $T = 300$.

La Figura 7-57 muestra resultados representativos de la estimación del orden de una aproximación VAR por el criterio BIC. Se observa una tendencia a aumentar el orden

estimado del modelo al aumentar T. También se detectó que el aumento de valor de los parámetros θ_1 y θ_2 provoca un aumento de valor del orden estimado.

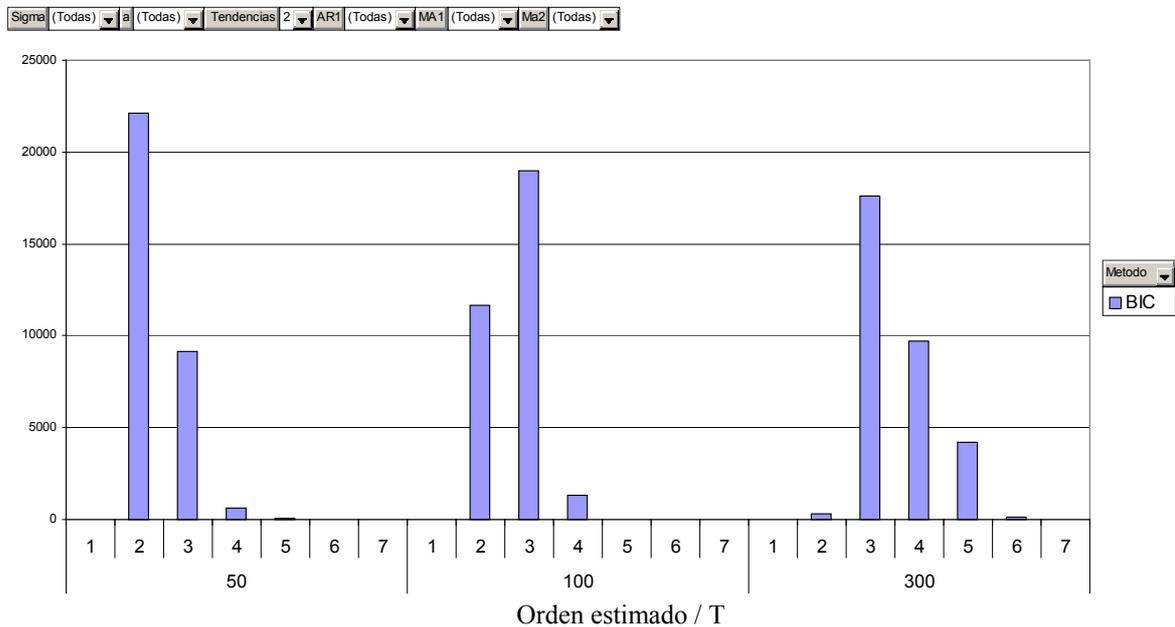


Figura 7-50. Histogramas del orden estimado (de modelos VAR) por el procedimiento de Johansen, con el criterio BIC, para distintos tamaños muestrales T.

7.4.2 Estimación del subespacio de cointegración

El subespacio de cointegración es en este caso un vector. La calidad de estimación la mediremos por el ángulo formado entre los vectores de cointegración real y estimado.

La Tabla 7-9 y la Tabla 7-10 y los gráficos que van de la Figura 7-51 a la Figura 7-57 muestran estadísticos correspondientes a la distribución de este ángulo, para los distintos métodos de estimación y distintos parámetros del proceso generador PGD4. Se observa que:

- Como en todos los casos anteriores de estimación de las relaciones de cointegración, los estimadores ML_b y ML son prácticamente idénticos, al igual que los estimadores CCA y CCA2. Para este proceso generador PGD4 se tiene además que, en general, la estimación ML resulta muy superior a la estimación por subespacios.

Resultados

- En cuanto al método ACCA, proporciona en general (con excepciones, sobre todo para valores altos de φ_1) peores resultados que la estimación ML, y resulta en ocasiones mejor y en otras peor que CCA simple, dependiendo de los parámetros del proceso generador.

Análisis de Sistemas Cointegrados mediante Métodos de Subespacios

T	Sigma	Estadístico	Método				
			ML_b	ML	ACCA	CCA	CCA2
50	0.2	Media	0.439	0.437	0.766	0.771	0.771
		Mediana	0.329	0.329	0.725	0.713	0.709
		IQR	0.414	0.413	0.692	0.691	0.690
		SD	0.352	0.352	0.407	0.410	0.409
		$P(\alpha < 0.05)$	3%	3%	1%	0%	1%
	0.5	Media	0.160	0.157	0.311	0.340	0.340
		Mediana	0.116	0.115	0.197	0.221	0.220
		IQR	0.130	0.126	0.291	0.313	0.312
		SD	0.170	0.165	0.314	0.335	0.336
		$P(\alpha < 0.05)$	18%	18%	8%	7%	7%
100	0.2	Media	0.173	0.173	0.436	0.449	0.448
		Mediana	0.138	0.137	0.341	0.357	0.357
		IQR	0.138	0.138	0.359	0.375	0.374
		SD	0.143	0.142	0.322	0.324	0.323
		$P(\alpha < 0.05)$	11%	11%	2%	2%	2%
	0.5	Media	0.052	0.052	0.093	0.096	0.096
		Mediana	0.041	0.041	0.072	0.073	0.073
		IQR	0.043	0.042	0.073	0.078	0.078
		SD	0.040	0.040	0.090	0.092	0.092
		$P(\alpha < 0.05)$	60%	60%	32%	33%	33%
300	0.2	Media	0.044	0.044	0.128	0.149	0.148
		Mediana	0.035	0.035	0.107	0.121	0.121
		IQR	0.035	0.035	0.106	0.121	0.121
		SD	0.033	0.033	0.099	0.109	0.108
		$P(\alpha < 0.05)$	70%	70%	17%	12%	13%
	0.5	Media	0.012	0.012	0.028	0.025	0.025
		Mediana	0.010	0.010	0.023	0.020	0.020
		IQR	0.010	0.010	0.024	0.023	0.022
		SD	0.009	0.009	0.020	0.019	0.019
		$P(\alpha < 0.05)$	99%	99%	87%	90%	91%

Tabla 7-9. Resultados de la distribución del ángulo formado entre los vectores de cointegración estimado y real. 1.000 simulaciones del proceso PGD4, para distintos procedimientos de estimación y valores de sigma y T, con $\alpha = 1$, $\varphi_1 = .5$, $\theta_1 = .5$ y $\theta_2 = .5$. En negrita los mejores valores de cada fila

Resultados

T	Sigma	Estadístico	Método				
			ML_b	ML	ACCA	CCA	CCA2
50	0.2	Media	0.706	0.712	0.770	0.764	0.765
		Mediana	0.636	0.639	0.723	0.722	0.718
		IQR	0.544	0.547	0.590	0.609	0.609
		SD	0.373	0.373	0.375	0.380	0.380
		$P(\alpha < 0.05)$	0%	0%	0%	0%	0%
	0.5	Media	0.429	0.418	0.366	0.397	0.396
		Mediana	0.324	0.318	0.278	0.308	0.306
		IQR	0.376	0.355	0.292	0.346	0.353
		SD	0.336	0.324	0.293	0.314	0.313
		$P(\alpha < 0.05)$	2%	2%	3%	2%	2%
100	0.2	Media	0.531	0.532	0.607	0.613	0.611
		Mediana	0.471	0.470	0.514	0.532	0.526
		IQR	0.353	0.350	0.449	0.467	0.458
		SD	0.319	0.317	0.356	0.359	0.359
		$P(\alpha < 0.05)$	0%	0%	0%	0%	0%
	0.5	Media	0.273	0.262	0.197	0.231	0.229
		Mediana	0.197	0.190	0.154	0.179	0.178
		IQR	0.208	0.198	0.144	0.174	0.174
		SD	0.259	0.248	0.182	0.204	0.201
		$P(\alpha < 0.05)$	6%	6%	8%	7%	7%
300	0.2	Media	0.248	0.247	0.314	0.338	0.336
		Mediana	0.204	0.203	0.253	0.283	0.280
		IQR	0.191	0.190	0.213	0.244	0.245
		SD	0.183	0.182	0.234	0.237	0.236
		$P(\alpha < 0.05)$	5%	5%	2%	1%	1%
	0.5	Media	0.085	0.083	0.069	0.086	0.086
		Mediana	0.065	0.063	0.059	0.072	0.071
		IQR	0.073	0.071	0.056	0.070	0.069
		SD	0.075	0.073	0.047	0.065	0.065
		$P(\alpha < 0.05)$	38%	38%	42%	34%	34%

Tabla 7-10. Resultados de la distribución del ángulo formado entre los vectores de cointegración estimado y real. 1.000 simulaciones del proceso PGD4, para distintos procedimientos de estimación y valores de sigma y T, con $a = 1$, $\varphi_1 = .9$, $\theta_1 = .9$ y $\theta_2 = .9$. En negrita los mejores valores de cada fila

Análisis de Sistemas Cointegrados mediante Métodos de Subespacios

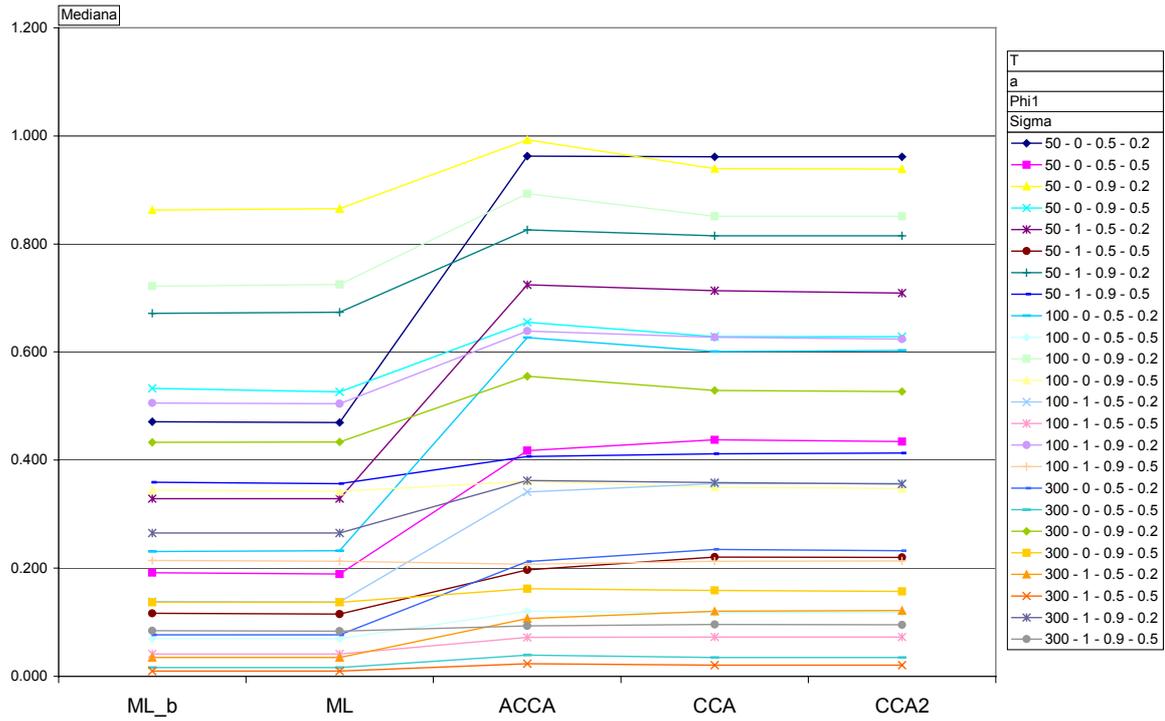


Figura 7-51. Mediana de la distribución del ángulo formado entre los subespacios de cointegración estimado y real. Procedimientos de estimación ML_b, ML, ACCA, CCA, y CCA2. Para distintas combinaciones de valores de los parámetros del proceso PGD4, con $\theta_1 = .5$ y $\theta_2 = .5$

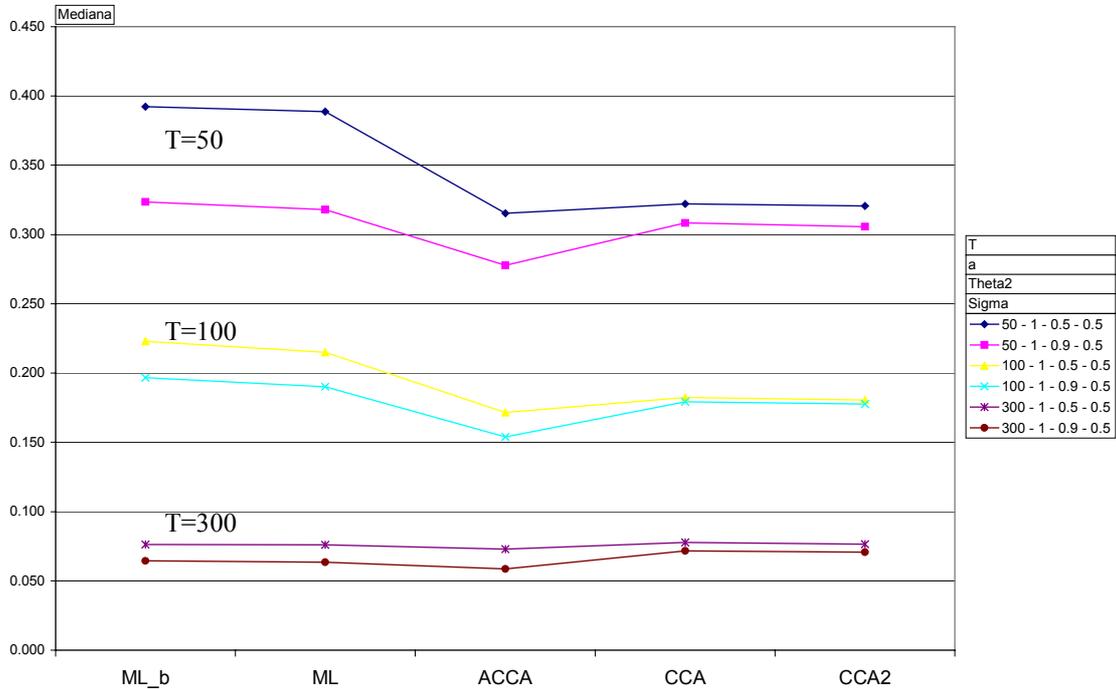


Figura 7-52. Mediana de la distribución del ángulo formado entre los subespacios de cointegración estimado y real. Procedimientos de estimación ML_b, ML, ACCA, CCA, y CCA2. Para distintas combinaciones de valores de los parámetros del proceso PGD4, con $a = 1$, $\phi_1 = .9$ y $\theta_1 = .9$

Resultados

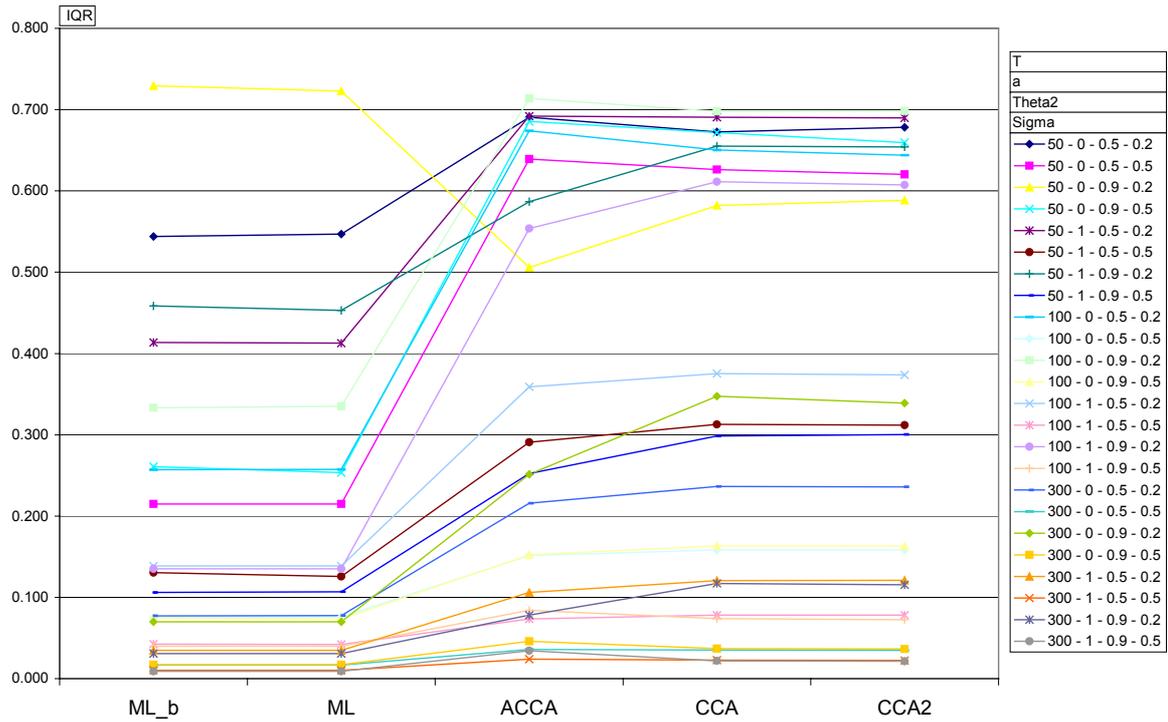


Figura 7-53. Rango intercuartílico (IQR) de la distribución del ángulo formado entre los subespacios de cointegración estimado y real. Procedimientos de estimación ML_b, ML, ACCA, CCA, y CCA2. Para distintas combinaciones de valores de los parámetros del proceso PGD4, con $\phi_1 = .5$ y $\theta_1 = .5$

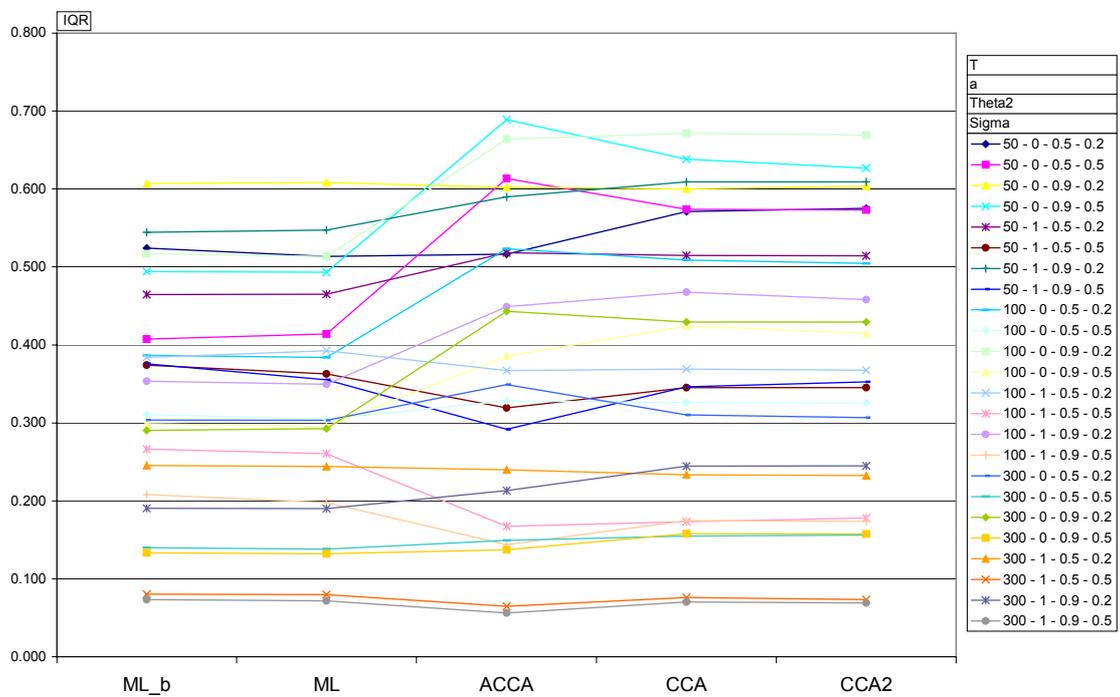


Figura 7-54. Rango intercuartílico (IQR) de la distribución del ángulo formado entre los subespacios de cointegración estimado y real. Procedimientos de estimación ML_b, ML, ACCA, CCA, y CCA2. Para distintas combinaciones de valores de los parámetros del proceso PGD4, con $\phi_1 = .9$ y $\theta_1 = .9$

Análisis de Sistemas Cointegrados mediante Métodos de Subespacios

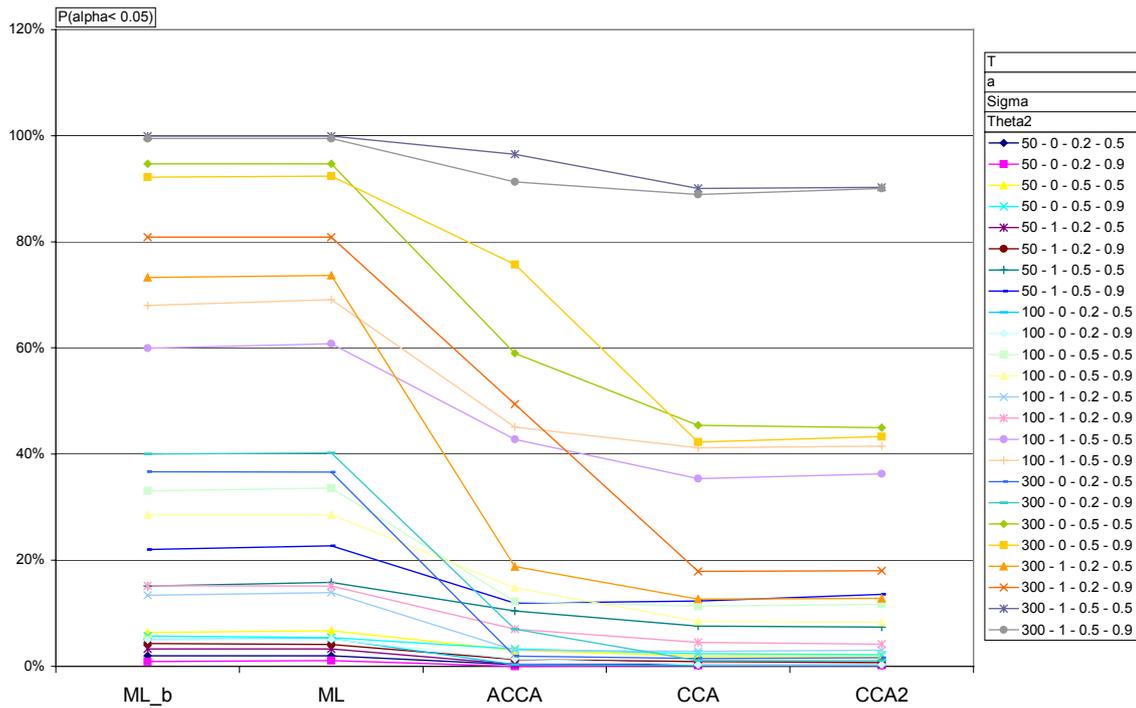


Figura 7-55. Probabilidad de que el ángulo formado entre los subespacios de cointegración estimado y real sea menor que un valor de referencia k , para cada procedimiento de estimación y para distintas combinaciones de valores de los parámetros del proceso PGD4, con $\varphi_1 = .5$ y $\theta_1 = .9$

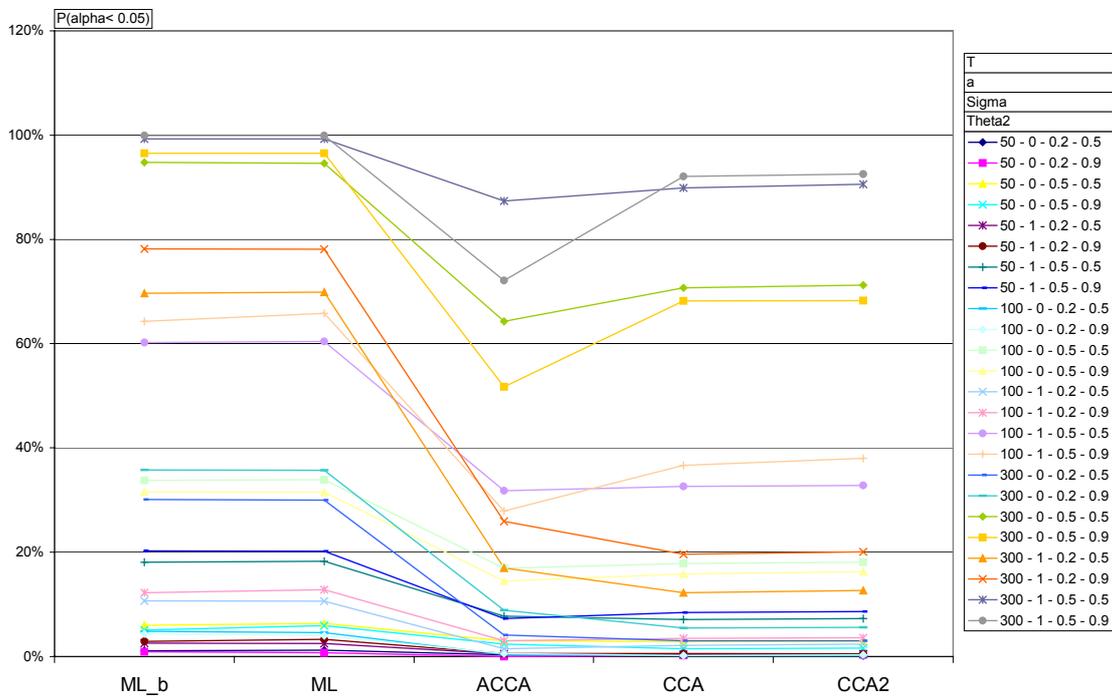


Figura 7-56. Probabilidad de que el ángulo formado entre los subespacios de cointegración estimado y real sea menor que un valor de referencia k , para cada procedimiento de estimación y para distintas combinaciones de valores de los parámetros del proceso PGD4, con $\varphi_1 = .5$ y $\theta_1 = .5$

Resultados

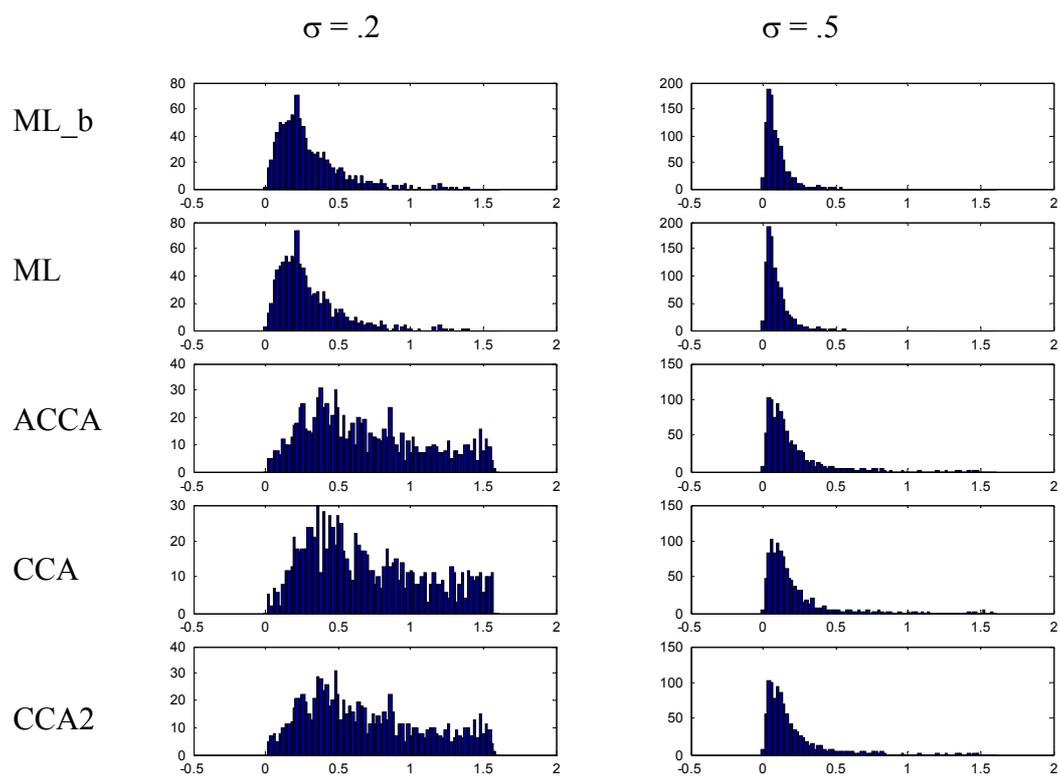


Figura 7-57. Histogramas del ángulo que separa el subespacio de cointegración estimado del real. 1000 realizaciones del proceso PGD4. $T = 100$, $\sigma = (.2, .5)$, $a = 0$, $\varphi_1 = .5$, $\theta_1 = .5$ y $\theta_2 = .5$

7.4.3 Error de predicción a corto plazo

Como en los experimentos anteriores, la capacidad de predicción se ha estudiado en muestras de predicción de tamaño igual al 10% del tamaño de la muestra de trabajo. La predicción se calcula en el espacio de estados. Como estado inicial para la muestra de predicción se utiliza el estado final estimado en la muestra de trabajo. Para la muestra de trabajo se utiliza un estado inicial suavizado. De nuevo comprobamos que, en general (Figura 7-58), otros métodos de cálculo del estado inicial de la muestra de trabajo supondrían poca o nula diferencia en la predicción obtenida al final de la muestra de trabajo (momento en que comenzamos a valorar la capacidad de predicción).

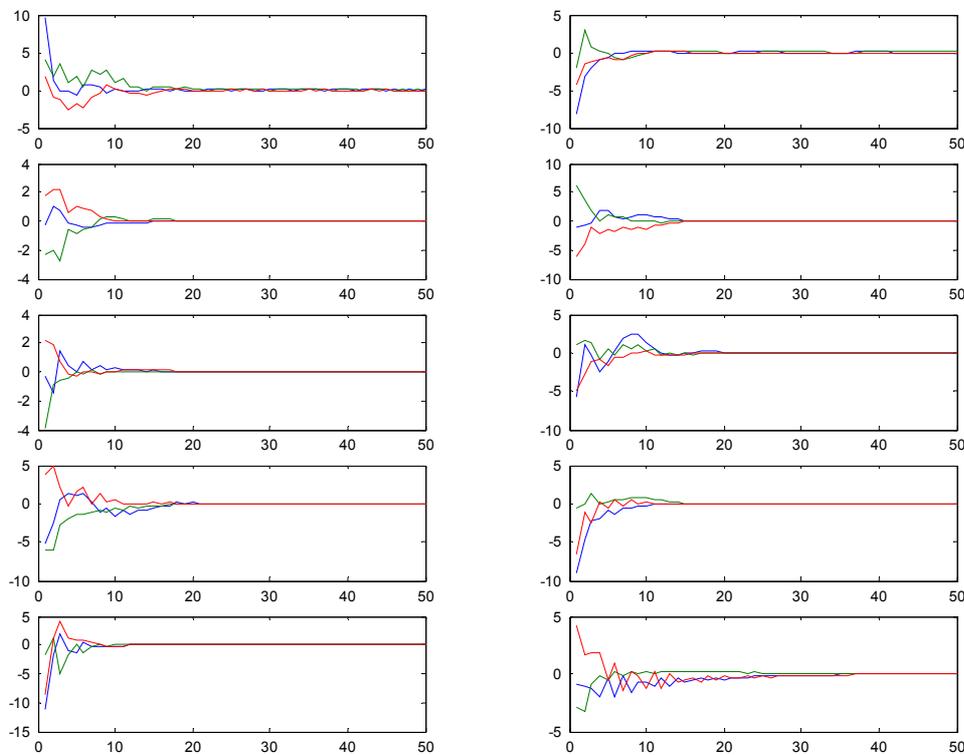


Figura 7-58. Cada línea mide la diferencia en la predicción a horizonte uno de cada una de las tres series generadas mediante el proceso PGD4. Las diferencias corresponden a dos estimaciones distintas del estado inicial: una estimación suavizada y una estimación nula. En las 10 realizaciones, el modelo usado para predicción se estimó por el procedimiento CCA.

La Tabla 7-11 y los gráficos que van de la Figura 7-59 a la Figura 7-66 se han seleccionado para ilustrar la comparación de métodos de estimación, en base a la

Resultados

distribución del estadístico U de Theil en predicción a corto plazo de los modelos proporcionados por cada método. Se comprueba que:

- En general, los modelos identificados por el método de Johansen (especificación BIC y estimación ML) resultan mejores para predicción que los modelos identificados por métodos de subespacios para bajo tamaño muestral ($T = 50, 100$) y horizontes de predicción crecientes (3, ...). Sin embargo, el método CCA simple (y en ocasiones el ACCA) supera al método de Johansen para tamaño muestral alto ($T = 300$) y bajo horizonte de predicción (1, 2).
- Pese a las ventajas asintóticas teóricas del algoritmo ACCA sobre el algoritmo CCA simple, los experimentos prácticos no nos permiten inclinarnos por su utilización: en muchas ocasiones ACCA resulta inferior a CCA simple, y tampoco se aprecia que el rendimiento de este algoritmo mejore (en términos relativos) con el tamaño muestral, dentro de los tamaños considerados (50, 100, 300).
- Los resultados obtenidos en predicción dependen de los parámetros del proceso generador y de la serie particular que se esté prediciendo (la primera, segunda o tercera: $y_{1,t}$, $y_{2,t}$ ó $y_{3,t}$), por lo que resulta difícil encontrar resultados de validez general.
- El método CCA2 es, con diferencia, el que peores resultados proporciona.

Análisis de Sistemas Cointegrados mediante Métodos de Subespacios

T	Sigma	Datos	Serie 1					Serie 2				
			Método					Método				
			ML_b	ML	ACCA	CCA	CCA2	ML_b	ML	ACCA	CCA	CCA2
50	0.2	Media	0.868	0.863	0.895	1.007	1.196	0.784	0.776	1.094	1.113	1.182
		Mediana	0.586	0.582	0.549	0.591	0.644	0.482	0.481	0.559	0.575	0.569
		IQR	0.609	0.600	0.592	0.676	0.760	0.533	0.512	0.748	0.788	0.826
		SD	1.346	1.341	1.518	1.740	2.909	1.064	1.049	1.861	1.864	2.079
		P(U < 1)	76%	77%	77%	74%	70%	82%	83%	73%	71%	70%
	0.5	Media	0.743	0.735	0.758	0.845	1.178	0.881	0.873	1.429	1.754	2.329
		Mediana	0.491	0.491	0.486	0.510	0.557	0.564	0.554	0.686	0.809	0.909
		IQR	0.527	0.533	0.549	0.612	0.729	0.711	0.693	1.060	1.258	1.457
		SD	0.953	0.938	1.101	1.105	3.345	1.136	1.145	2.330	4.631	9.736
		P(U < 1)	81%	81%	82%	78%	73%	74%	74%	63%	58%	53%
100	0.2	Media	0.562	0.559	0.556	0.569	0.630	0.512	0.509	0.649	0.579	0.712
		Mediana	0.481	0.480	0.467	0.477	0.506	0.419	0.417	0.432	0.424	0.484
		IQR	0.336	0.334	0.302	0.321	0.381	0.327	0.318	0.396	0.358	0.446
		SD	0.315	0.310	0.376	0.409	0.476	0.383	0.376	1.141	0.871	0.879
		P(U < 1)	92%	92%	92%	92%	89%	92%	93%	86%	90%	82%
	0.5	Media	0.485	0.488	0.490	0.485	0.533	0.528	0.526	0.610	0.562	1.352
		Mediana	0.407	0.408	0.415	0.408	0.432	0.449	0.450	0.485	0.452	0.761
		IQR	0.291	0.293	0.293	0.282	0.327	0.344	0.344	0.423	0.384	0.965
		SD	0.298	0.304	0.305	0.306	0.370	0.344	0.341	0.487	0.440	2.020
		P(U < 1)	94%	93%	94%	94%	92%	91%	91%	87%	89%	62%
300	0.2	Media	0.458	0.458	0.429	0.427	0.609	0.393	0.393	0.374	0.374	0.584
		Mediana	0.442	0.442	0.407	0.402	0.540	0.373	0.374	0.350	0.351	0.502
		IQR	0.156	0.155	0.142	0.145	0.249	0.164	0.163	0.156	0.158	0.273
		SD	0.123	0.123	0.127	0.126	0.286	0.134	0.133	0.137	0.137	0.339
		P(U < 1)	100%	100%	100%	100%	93%	100%	100%	100%	100%	92%
	0.5	Media	0.393	0.394	0.391	0.391	0.442	0.402	0.402	0.405	0.401	0.753
		Mediana	0.375	0.376	0.370	0.371	0.412	0.381	0.381	0.379	0.375	0.643
		IQR	0.137	0.136	0.140	0.142	0.166	0.160	0.161	0.165	0.165	0.398
		SD	0.120	0.120	0.120	0.122	0.157	0.132	0.133	0.143	0.141	0.421
		P(U < 1)	100%	100%	100%	100%	100%	100%	100%	100%	100%	81%

Tabla 7-11. Resultados de la distribución del estadístico U de Theil en predicción de las series y_1 e y_2 a horizonte 1 para 1.000 simulaciones del proceso PGD4, para distintos procedimientos de estimación del modelo y distintos valores de sigma y T, con $a=1$, $\varphi_1 = .5$, $\theta_1 = .9$ y $\theta_2 = .5$. Todos los estimadores restan las medias salvo ML_b. En negrita los mejores valores de cada fila, por cada serie.

Resultados

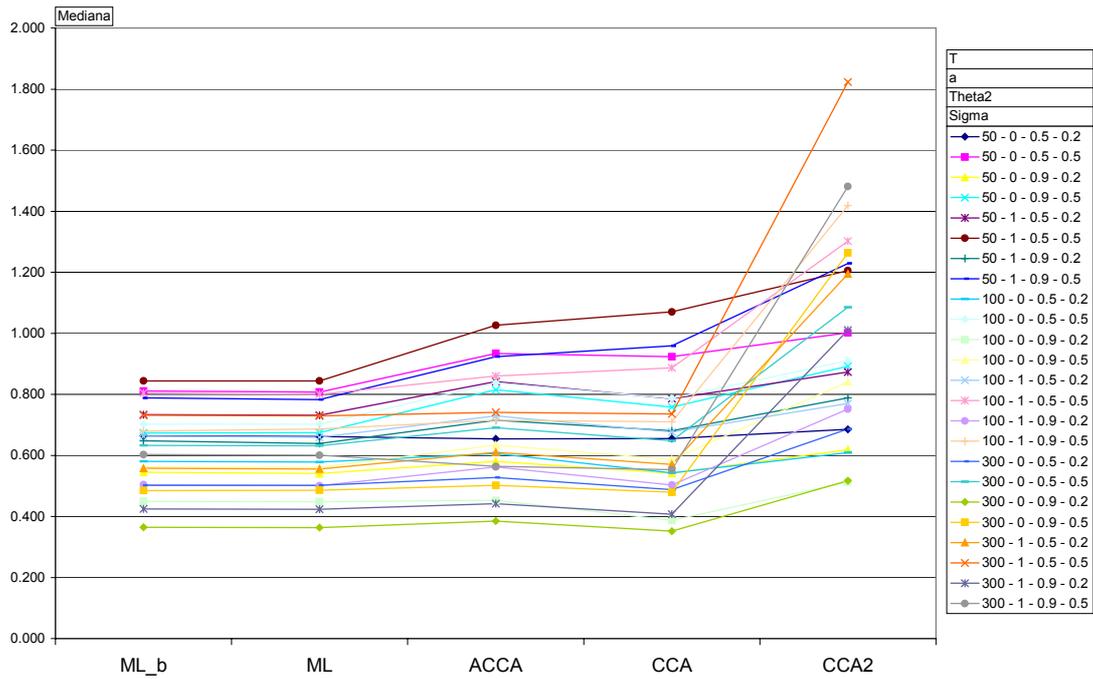


Figura 7-59. Mediana de la distribución del estadístico U de Theil en predicción de la serie y_3 a horizonte 1, para modelos identificados mediante distintos procedimientos y para distintas combinaciones de valores de los parámetros del proceso generador PGD4, con $\phi_1 = .9$ y $\theta_1 = .9$

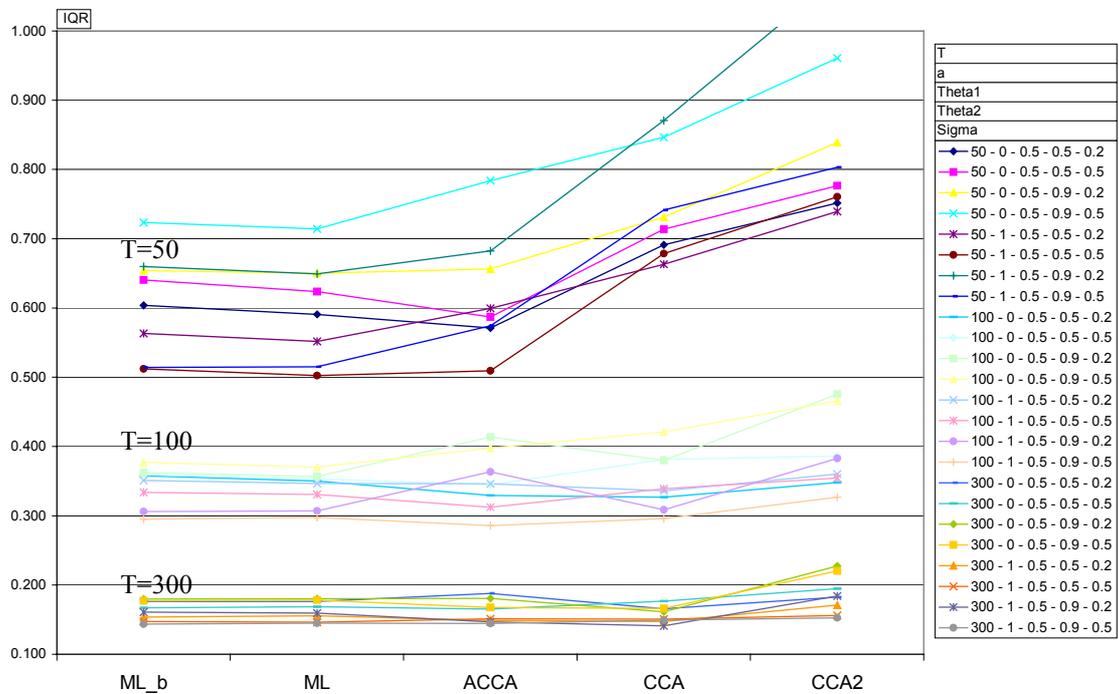


Figura 7-60. Rango intercuartílico (IQR) de la distribución del estadístico U de Theil en predicción de la serie y_1 a horizonte 1, para modelos identificados mediante distintos procedimientos y para distintas combinaciones de valores de los parámetros del proceso generador PGD4, con $\phi_1 = .5$ y $\theta_1 = .5$.

Análisis de Sistemas Cointegrados mediante Métodos de Subespacios

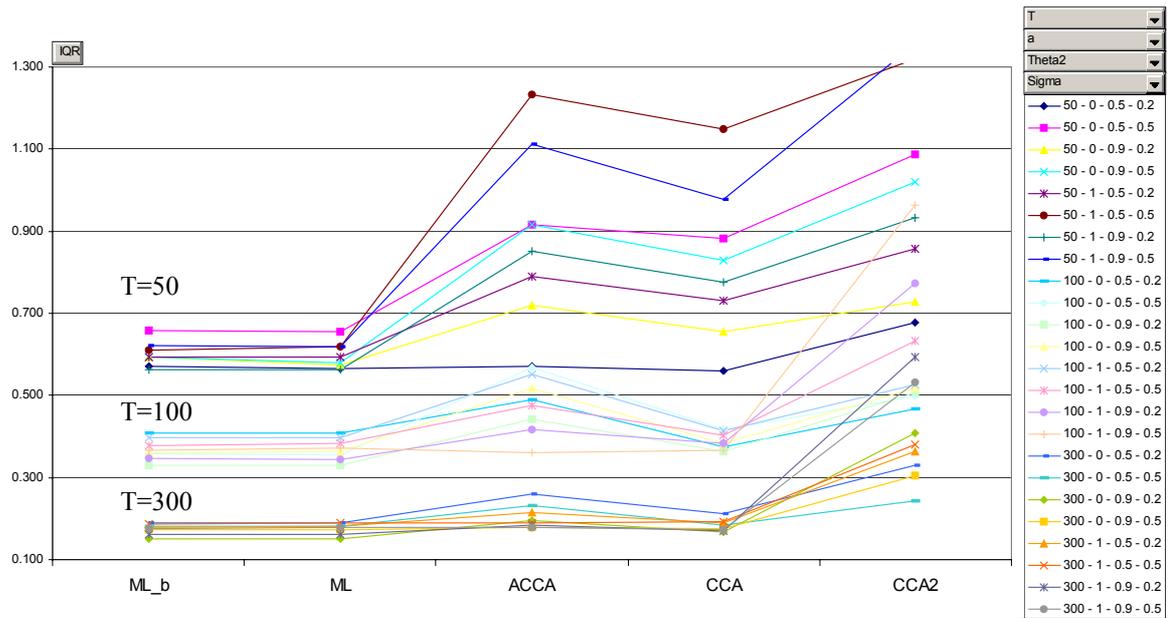


Figura 7-61. Rango intercuartílico (IQR) de la distribución del estadístico U de Theil en predicción de la serie y_3 a horizonte 1, para modelos identificados mediante distintos procedimientos y para distintas combinaciones de valores de los parámetros del proceso generador PGD4, con $\phi_1 = .5$ y $\theta_1 = .5$.

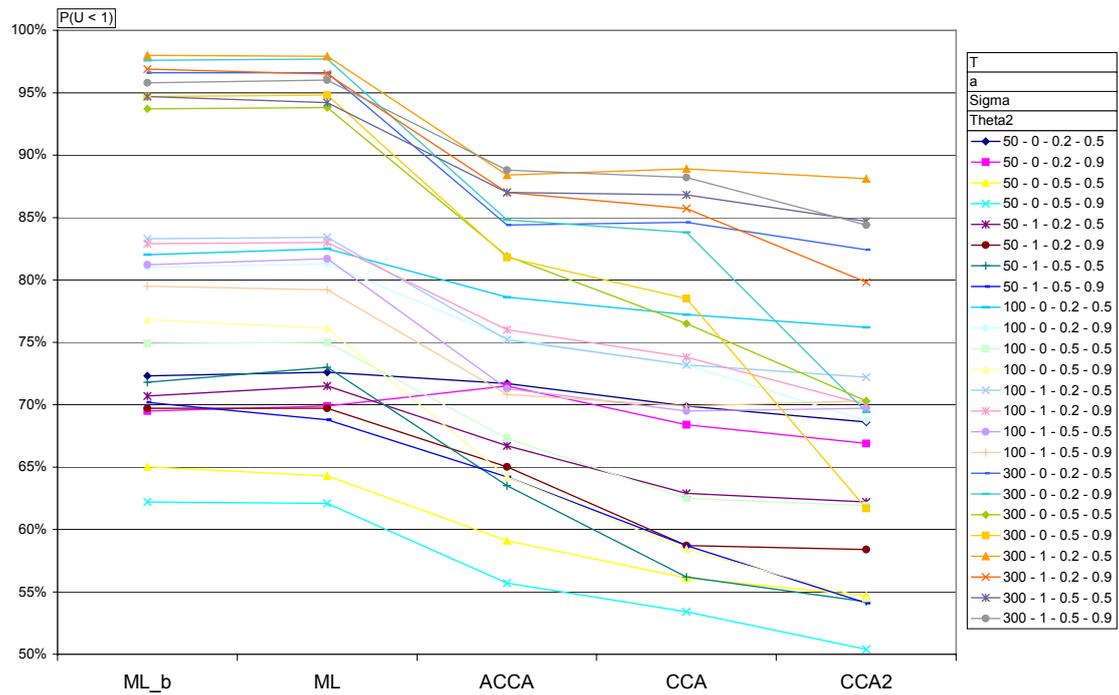


Figura 7-62. Probabilidad de que el estadístico U de Theil en predicción de la serie y_1 a horizonte 3 sea menor que uno, para modelos identificados mediante distintos procedimientos, y para distintas combinaciones de valores de los parámetros del proceso generador PGD4, con $\phi_1 = .5$ y $\theta_1 = .5$.

Resultados

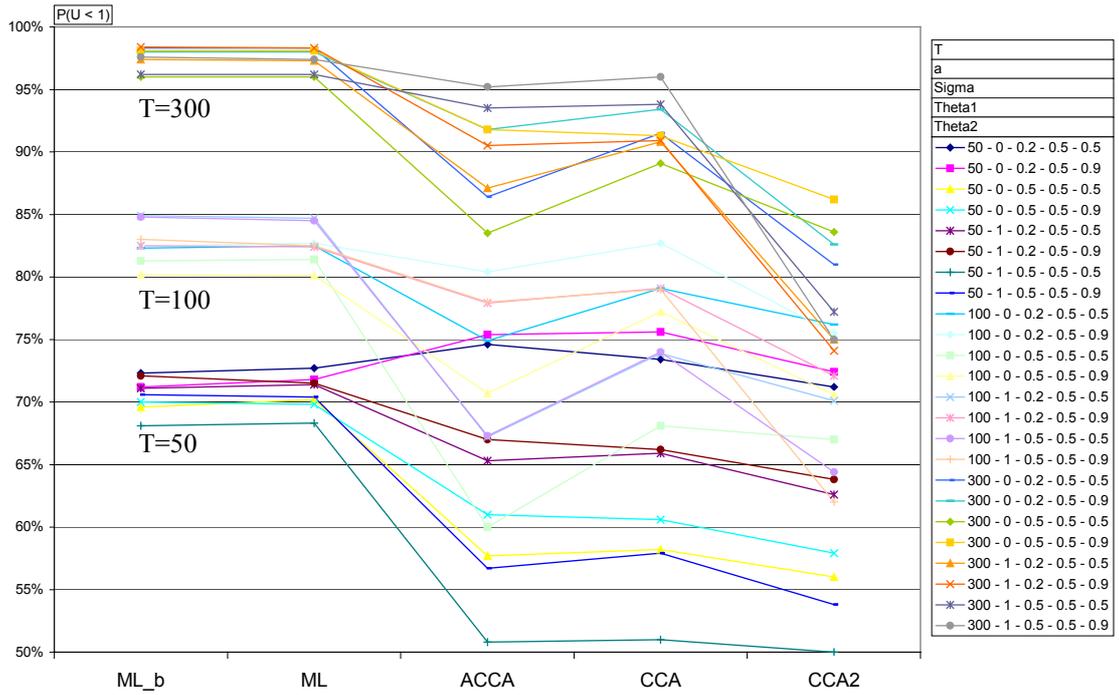


Figura 7-63. Probabilidad de que el estadístico U de Theil en predicción de la serie y_3 a horizonte 2 sea menor que uno, para modelos identificados mediante distintos procedimientos, y para distintas combinaciones de valores de los parámetros del proceso generador PGD4, con $\phi_1 = .5$ y $\theta_1 = .5$

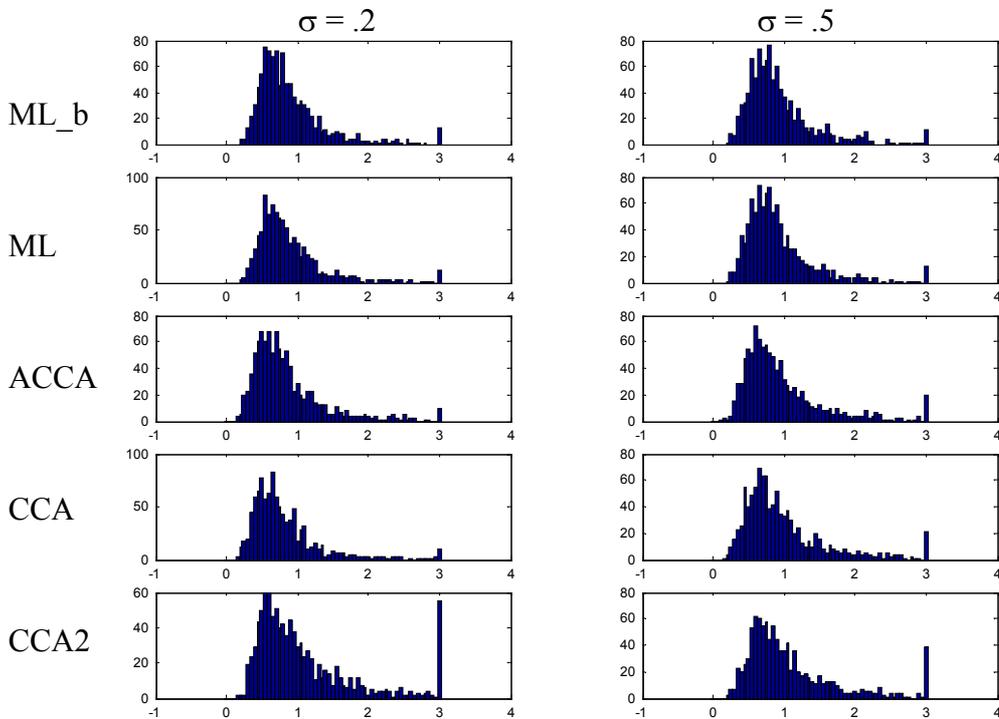


Figura 7-64. Histogramas del estadístico U de Theil en predicción de la serie y_2 a horizonte 1. 1000 realizaciones del proceso PGD4. $T = 100$, $\sigma = (.2, .5)$, $a = 0$, $\phi_1 = .9$, $\theta_1 = .9$ y $\theta_2 = .9$

Análisis de Sistemas Cointegrados mediante Métodos de Subespacios

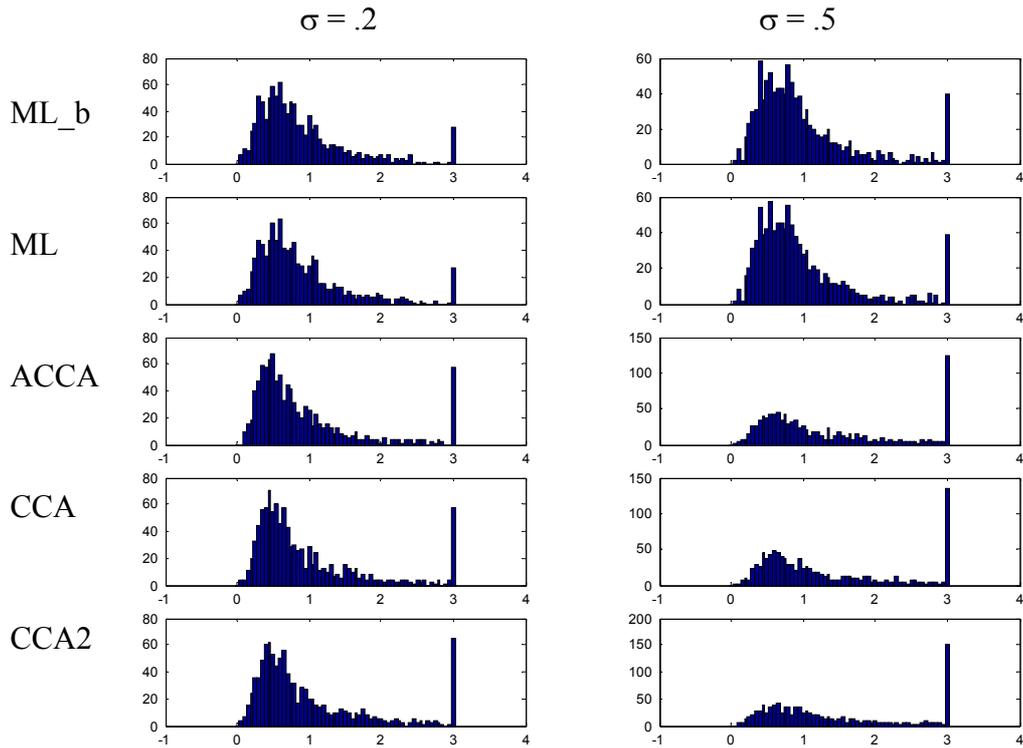


Figura 7-65. Histogramas del estadístico U de Theil en predicción de la serie y_2 a horizonte 1. 1000 realizaciones del proceso PGD4. $T = 50$, $\sigma = (.2, .5)$, $a = 0$, $\varphi_1 = .5$, $\theta_1 = .5$ y $\theta_2 = .5$

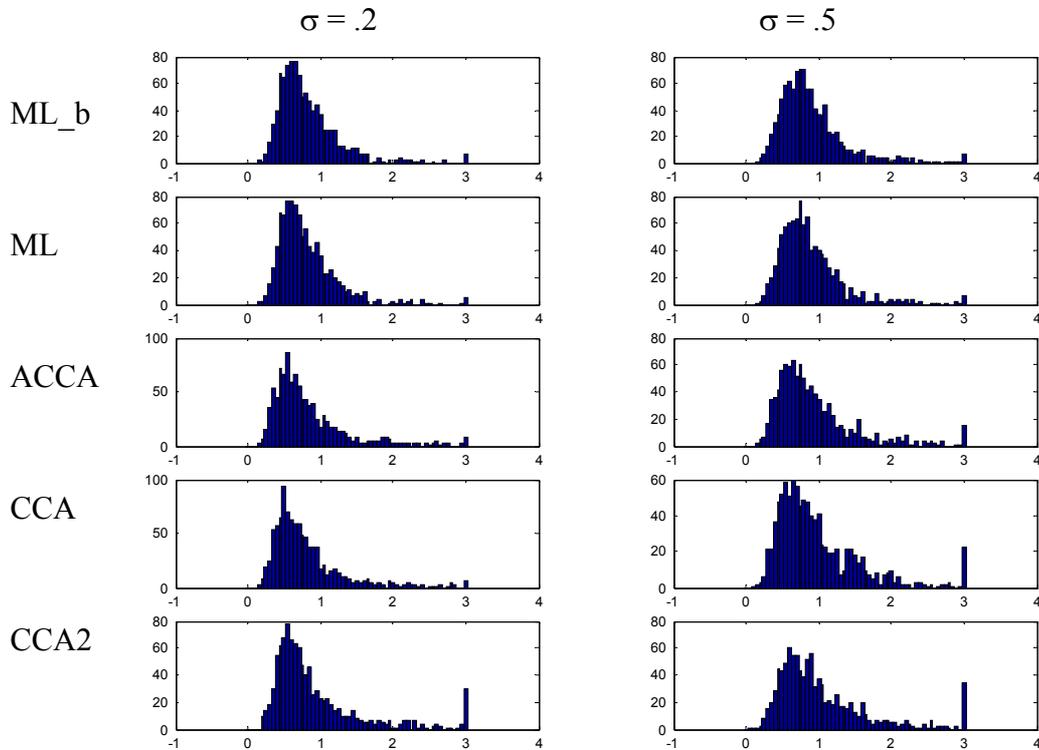


Figura 7-66. Histogramas del estadístico U de Theil en predicción de la serie y_1 a horizonte 1. 1000 realizaciones del proceso PGD4. $T = 100$, $\sigma = (.2, .5)$, $a = 0$, $\varphi_1 = .9$, $\theta_1 = .9$ y $\theta_2 = .5$

7.4.4 Contraste de igualdad de capacidad predictiva

Como en los casos anteriores, analizamos el número de veces que el modelo identificado por un procedimiento supera en capacidad predictiva (error cuadrático) al modelo identificado por otro procedimiento. El contraste binomial se detalló en el apartado 7.2.4. La Tabla 7-12 muestra, para distintas combinaciones de valores de parámetros del proceso generador PGD4, el número de veces (de entre 1.000 repeticiones) en que los modelos identificados por los métodos de Johansen, ACCA o CCA2 proporcionaron predicciones de menor error cuadrático medio que el modelo correspondiente estimado por el algoritmo CCA simple.

T	φ_1	θ_1	θ_2	σ					
				0.2			0.5		
				Johansen	ACCA	CCA2	Johansen	ACCA	CCA2
50	0.5	0.5	0.5	541	497	470	658	496	434
			0.9	582	484	446	621	456	371
		0.9	0.5	514	503	455	595	506	382
			0.9	560	479	415	580	476	337
	0.9	0.5	0.5	573	492	452	669	540	454
			0.9	566	465	450	641	505	422
		0.9	0.5	564	465	447	617	519	428
			0.9	533	436	428	614	509	380
100	0.5	0.5	0.5	567	382	458	581	411	371
			0.9	504	353	232	421	450	159
		0.9	0.5	525	404	368	530	601	190
			0.9	448	338	152	444	482	104
	0.9	0.5	0.5	544	377	430	604	464	385
			0.9	445	310	266	432	376	205
		0.9	0.5	537	364	352	622	557	276
			0.9	510	320	207	553	514	187
300	0.5	0.5	0.5	564	297	176	456	402	117
			0.9	475	321	44	351	323	29
		0.9	0.5	436	327	63	370	372	15
			0.9	343	316	18	306	305	2
	0.9	0.5	0.5	650	178	172	627	401	191
			0.9	504	187	96	422	315	63
		0.9	0.5	535	220	72	519	408	45
			0.9	381	226	49	315	339	34

Tabla 7-12. Número de veces entre 1.000 en que el modelo identificado por los métodos de Johansen, ACCA o CCA2 bate en error cuadrático medio de predicción al modelo identificado por el método CCA simple. Se considera significativo a favor de CCA un valor inferior a 459 (celdas en negrita), y significativo a favor del método en cuestión un valor superior a 541 (celdas con relleno). La tabla se ha realizado con el error de predicción de la serie y_3 a horizonte 1, con $a = 1$ y perturbaciones gaussianas.

La Tabla 7-12, que resulta representativa de los resultados generales para otras combinaciones de valores de parámetros, permite obtener unas conclusiones similares a las obtenidas para el proceso generador PGD3. Para bajo tamaño muestral ($T = 50$), se aprecia que el método de Johansen proporciona, en general, mejores modelos predictivos que el algoritmo CCA simple. Al crecer el tamaño muestral, el método CCA simple se vuelve comparable al de Johansen, y normalmente lo supera para alto tamaño muestral ($T = 300$). El método CCA2 resulta claramente inferior al CCA simple, para prácticamente todos los casos. El método ACCA supera en ocasiones al CCA simple, pero en general resulta inferior, y especialmente para alto tamaño muestral ($T = 300$).

7.4.5 Combinación de predicciones

Pese a los malos resultados de predicción obtenidos por los modelos identificados mediante el método CCA2, nos planteamos de nuevo si estos modelos pueden aportar información útil para predicción, al combinarlos con modelos identificados por otros procedimientos.

La Figura 7-67 muestra la distribución de pesos óptimos para predicción en combinaciones lineales de modelos CCA con modelos CCA2 (ver apartado 7.2.5). Un valor del peso próximo a uno (/cero) indica que el modelo CCA (/CCA2) proporciona la mejor predicción y el otro no aporta información útil.

Las conclusiones para este proceso generador coinciden con las obtenidas para el proceso generador PGD3:

- La mejor predicción corresponde, en general, a uno de los modelos, y no a una combinación de ambos.
- Para $T = 50$ los dos métodos (CCA y CCA2) están básicamente empatados en cuanto a probabilidad de proporcionar el mejor modelo en predicción. La ventaja del método CCA sobre el CCA2 se va haciendo notable al aumentar el tamaño muestral y es en general muy evidente (depende de los valores de los parámetros del proceso generador) para tamaño muestral $T = 300$.

Resultados

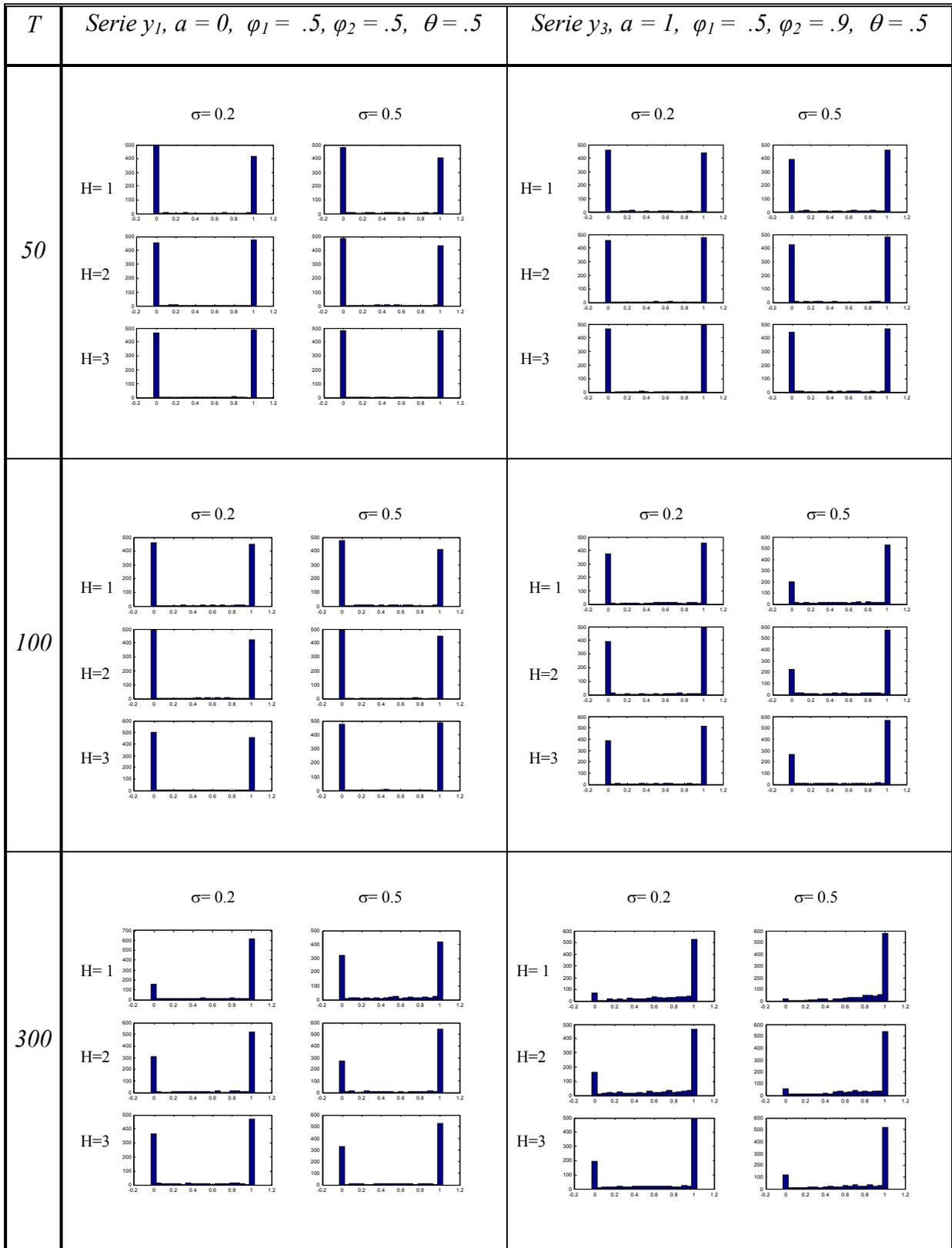


Figura 7-67. Histogramas de los pesos ω^* de modelos CCA en combinaciones con modelos CCA2 para predicción a horizonte H .

8 CONCLUSIONES

El análisis de sistemas cointegrados es una cuestión de gran interés actual en econometría (recordemos el premio Nobel de economía de 2003, concedido a Clive W. J. Granger por sus métodos de análisis de series temporales económicas con tendencias comunes). Sin embargo, los procesos no estacionarios y cointegrados no son exclusivos de los fenómenos económicos, sino que son frecuentes en la naturaleza y aparecen en múltiples procesos de ingeniería, como indicábamos en la introducción. En general, todas las disciplinas que, entre sus herramientas, utilicen la identificación de sistemas, podrán verse beneficiadas por avances en la metodología de análisis de procesos cointegrados, que es el objeto de estudio de esta tesis doctoral.

El método de análisis de procesos cointegrados más extendido y desarrollado actualmente es el método de máxima verosimilitud de Johansen, que está basado en una formulación vectorial autorregresiva (VAR) de las series estudiadas. Al comienzo de esta tesis nos planteamos estudiar la utilidad de otro tipo de formulación matemática de sistemas lineales, el espacio de estados, para el análisis de sistemas cointegrados. Los principales motivos por los que nos hicimos este planteamiento son la mayor parsimonia de la formulación del espacio de estados respecto a la formulación VAR para representar procesos VARMA, y la serie de potentes técnicas y herramientas desarrolladas sobre el espacio de estados, entre las que destacamos el filtro de Kalman y los métodos de subespacios. Además, el espacio de estados constituye una forma natural e intuitiva para formular modelos ciclo-tendencia, de los cuales los modelos cointegrados son un caso particular.

El estudio del problema y de las diferentes técnicas y métodos que hemos citado, desarrollado a lo largo de los primeros capítulos de esta tesis, nos condujo a las siguientes consideraciones:

- Para abordar el análisis de sistemas cointegrados en el espacio de estados necesitamos en primer lugar analizar cómo puede representarse uno de estos sistemas en el espacio de estados. Este análisis se llevó a cabo en el capítulo 2, en el que relacionamos la representación en forma de tendencias comunes de Stock y Watson (1988), con la formulación en el espacio de estados. Además, la propia formulación en el espacio de estados nos permitió llamar la atención

sobre el hecho de que la descomposición ciclo-tendencia, tan habitual en econometría, en general no es única, no es identificable¹.

- Una de las utilidades más potentes y más utilizadas del espacio de estados es la obtención de estimadores pseudo máximo verosímiles mediante optimización numérica de la función de verosimilitud paramétrica calculada de forma recursiva mediante el filtro de Kalman (ver anexo I.7). Tras el análisis realizado consideramos que, para la estimación de sistemas cointegrados, este enfoque difícilmente puede competir con el método de Johansen, que también es un enfoque de máxima verosimilitud, pero no precisa entrar en problemáticas optimizaciones por métodos numéricos.
- Los métodos de subespacios sí constituyen, teóricamente, una alternativa o un complemento factible e interesante para el estudio de sistemas cointegrados. Son rápidos, no iterativos, descansan sobre herramientas algebraicas robustas y los últimos avances (Bauer y Wagner 2002) indican que pueden competir e incluso superar al método de Johansen en el plano teórico (asintótico).

Tras alcanzar estas reflexiones, nos centramos en el objetivo más concreto de esta investigación doctoral: valorar la utilidad práctica de los métodos de subespacios para el análisis de sistemas cointegrados, averiguar en qué medida pueden competir en la práctica con el método de referencia (método de Johansen).

La revisión de la literatura nos proporcionó dos algoritmos de subespacios previos propuestos para ser aplicados a sistemas cointegrados: el algoritmo SSATS de Aoki (1997) y el algoritmo ACCA de Bauer y Wagner (2002). El estudio de ambos algoritmos nos sugirió un nuevo método: el algoritmo CCA bietápico, o CCA2, que combina la intuición del método de Aoki con las propiedades deseables del algoritmo CCA (Bauer y Ljung 2002). También decidimos considerar los modelos proporcionados

¹ Salvo que se impongan a priori una serie de condiciones que, en muchos casos, resultan arbitrarias o argumentables. Es frecuente que esas condiciones se impongan precisamente para conseguir la identificabilidad, y no porque las respalde una teoría económica sólida. El problema de identificabilidad llama la atención sobre el peligro de aceptar como válida una determinada descomposición estructural sólo porque ésta resulta compatible con las observaciones.

Conclusiones

por el método CCA estándar, que proporciona estimadores consistentes del subespacio de cointegración.

Se diseñó una batería de experimentos para comparar los distintos métodos, tomando como referencia el método de máxima verosimilitud de Johansen. Se seleccionaron procesos generadores cointegrados formulados en el espacio de estados (o, equivalentemente, procesos VARMA cointegrados) de mayor generalidad a los considerados en trabajos previos (Gonzalo 1994, Bauer y Wagner 2003b). Ante la sospecha de que el mejor o peor funcionamiento de uno u otro método de identificación pudiera depender de las características del proceso generador (como efectivamente sucede), se tomaron procesos generadores variados y de distintas dimensiones. También se plantearon medidas de calidad de estimación originales (tales como el ángulo de separación entre subespacios), que tratan de superar algunas de las limitaciones de los antecedentes.

En cuanto a las conclusiones obtenidas por comparación de los distintos métodos de subespacios, resaltamos:

- Ninguno de los algoritmos de subespacios considerados para el análisis de sistemas cointegrados ha demostrado ventajas prácticas claras sobre el algoritmo CCA simple o estándar.
- El método SSATS de Aoki para análisis de sistemas cointegrados se desechó por su desventaja frente al método CCA al estimar el subespacio de cointegración (sin embargo, las ideas de Aoki nos llevaron a proponer el método CCA2, que no presenta esa desventaja).
- El método CCA2, planteado y estudiado en esta tesis, resultó proporcionar, pese a su carácter intuitivo, modelos que, en general, presentan peor capacidad de predicción que los obtenidos mediante el método CCA simple, por lo que, conforme a los experimentos realizados, su utilización no se encuentra en conjunto justificada.
- El método ACCA, o CCA adaptado, de Bauer y Wagner (2002), es el único que cuenta con una prueba teórica de estimación consistente de todos los parámetros del modelo. Pese a sus ventajas teóricas asintóticas sobre el método CCA

simple, no se aprecian ventajas sobre el mismo en los experimentos realizados, ni en la estimación del subespacio de cointegración, ni en la capacidad de predicción a corto plazo de los modelos estimados.

- En los experimentos realizados estas conclusiones han resultado robustas frente a desviaciones de la hipótesis de normalidad de las perturbaciones.

En lo relativo a la comparación Johansen / subespacios, las conclusiones más relevantes desde el punto de vista práctico son:

- Para identificación de procesos generadores VARMA, ninguno de los métodos proporciona uniformemente mejores resultados que el otro: el mejor o peor funcionamiento de un método de identificación respecto ha mostrado una alta dependencia de las características del proceso generador (tanto de la estructura como de los valores de los parámetros). Este resultado obliga a considerar con precaución las conclusiones de cualquier estudio práctico comparativo sobre métodos de identificación o estimación que esté basado en un número reducido de procesos generadores.
- En general, en los experimentos realizados con una muestra de identificación de tamaño reducido ($T = 50$ ó 100), los modelos obtenidos mediante el algoritmo de Johansen han proporcionado los mejores resultados predictivos (con una ventaja más notable para $T = 50$ que para $T = 100$). Por el contrario, cuando se cuenta con tamaños muestrales mayores ($T = 300$), el método CCA simple proporcionó, en general, los mejores modelos predictivos a corto plazo (horizontes de predicción 1, 2, 3). Por otra parte, al aumentar el horizonte de predicción aumenta la precisión relativa de los modelos de Johansen respecto a los modelos CCA. En la mayoría de los casos estudiados se observaron estas pautas de forma significativa.
- El hecho de que los métodos de subespacios hayan proporcionado, en muchas ocasiones, modelos que superaban (en precisión de estimación de las relaciones de cointegración y, particularmente, en capacidad predictiva) a los modelos VAR proporcionados por el método de Johansen, justifica, cuando menos, la utilización práctica de los métodos de subespacios como complemento o alternativa al método de Johansen.

Limitaciones y extensiones

- Posiblemente la mayor aportación de esta tesis doctoral es mostrar que, para estimar modelos de sistemas cointegrados en los que el proceso generador subyacente es un proceso VARMA, los métodos de subespacios pueden competir con el método de Johansen, tanto en la estimación del subespacio de cointegración como, especialmente, en la calidad predictiva de los modelos identificados, resultando en muchos casos superiores. El aspecto de estimación del número de tendencias comunes ha sido estudiado por Bauer y Wagner (2002), que sostienen que, en este sentido, los métodos de subespacios también compiten en términos de igualdad o superioridad con el método de Johansen.
- El siguiente paso debería ser la búsqueda de cuáles son las características del proceso generador que hacen más recomendable uno u otro método de estimación. Conforme a los experimentos realizados, el tamaño muestral constituye un indicador para inclinarnos bien por el método de Johansen o bien por los métodos de subespacios, pero no parece que se pueda dar un tamaño muestral de referencia con validez general para todos los procesos, pues el tamaño muestral no es el único parámetro influyente.
- Analizar mediante simulación el efecto de los diferentes parámetros del proceso generador sobre el rendimiento de cada método de identificación presenta varias dificultades. Una de estas dificultades es que muchos de esos parámetros, tal como aparecen en el modelo generador, pueden no resultar recuperables a partir de las observaciones¹. Incluso si fueran recuperables, es de esperar que su efecto no venga marcado por funciones lineales sencillas, y que existan interacciones en los efectos causados por distintos parámetros. Desde nuestro punto de vista, para abordar este problema y proporcionar una guía práctica de cuándo es más recomendable uno u otro método, debemos acudir a características recuperables

¹ Por dos motivos. Primero, porque hemos trabajado con procesos generadores “estructurales”, independientemente de que correspondan o no a formas canónicas en el espacio de estados. Segundo, porque aunque trabajáramos con procesos generadores en formas canónicas, tampoco aseguramos la recuperación de los parámetros, puesto que la fase de especificación puede conducir a modelos con especificaciones distintas a la del proceso generador.

(estimables) del sistema independientes de la formulación paramétrica elegida, tales como los invariantes de Markov, si bien se mantiene el problema asociado a la gran variedad de parámetros existente en un sistema multivariante y a las interacciones de los efectos.

- Existe una diferencia entre el método de Johansen y los métodos de subespacios que puede explicar la ventaja predictiva relativa ofrecida por los modelos identificados mediante el método de Johansen al ir aumentando el horizonte de predicción en las simulaciones: el método de Johansen proporciona modelos con raíces unitarias para las tendencias, mientras que los algoritmos de subespacios estiman el valor de esas raíces (autovalores de la matriz de transición de estados \mathbf{A}). Al haber simulado modelos con raíces unitarias, es de esperar una cierta ventaja en el procedimiento que incorpora esa información, ventaja que se hace especialmente notable a largo plazo (el efecto predictivo de una raíz próxima a uno, pero algo mayor o menor, se va haciendo más notable al aumentar el horizonte de predicción). Sin embargo, utilizando técnicas de regresión de rango reducido es posible modificar un algoritmo de subespacios para imponer que las raíces no estacionarias sean exactamente unitarias (Larimore 2000, Bauer y Wagner 2002), lo que supone una línea de ampliación de esta investigación comparativa.
- Otras fuentes de investigación futura son la influencia y el tratamiento de tendencias deterministas en las series estudiadas, así como el efecto de perturbaciones heterocedásticas

ANEXOS

Anexo I. Estimación por mínimos cuadrados. Filtro de Kalman

Anexo II. Realizaciones mínimas

Anexo III. Formulaciones alternativas del espacio de estados

Anexo IV. Identificación de sistemas estocásticos por métodos de subespacios

Anexo I . Estimación por mínimos cuadrados. Filtro de Kalman

I.1 Introducción

El filtro de Kalman es una técnica recursiva de estimación que se engloba dentro de los métodos de estimación por mínimos cuadrados –Sorenson (1970), Luenberger (1969)-. Este anexo desarrolla los principales conceptos de esta técnica de estimación, partiendo de los casos más simples de estimación por mínimos cuadrados ordinarios y complicando progresivamente los modelos, a la vez que aumentando su generalidad, hasta llegar al modelo de espacio de estados y el filtro de Kalman.

Partiremos de un procedimiento de ajuste de observaciones por mínimos cuadrados en un contexto independiente del modelo “generador” de los datos, donde el objetivo será minimizar la suma cuadrática de las distancias euclídeas entre observación y estimación (error cuadrático), lo que da lugar al método de mínimos cuadrados ordinarios (MCO).

Al considerar un determinado modelo generador de los datos, lineal pero estocástico, el problema de estimación por mínimos cuadrados pasará del ámbito de ajustar una muestra de datos particular, al ámbito probabilístico, y la distancia a minimizar mediante el ajuste será la esperanza de la norma del vector de residuos, es decir, buscaremos un ajuste que minimice el error cuadrático medio. Esto conducirá al estimador de mínimos cuadrados generalizados (MCG).

Por último, se considerará un modelo aún más general para el sistema generador de datos: el modelo lineal de coeficientes estocásticos. Como caso particular, aunque con un grado de generalidad muy elevado, de este modelo de coeficientes estocásticos, se presentará el espacio de estados, y se introducirá el filtro de Kalman como método de estimación recursivo en este contexto.

Varios de los resultados de este capítulo se presentan sin demostración. El lector interesado puede acudir a Luenberger (1969). También Ljung (1999, Appendix II, p. 543) resume los principales aspectos de la estimación por mínimos cuadrados con claridad y brevedad.

I.2 Mínimos cuadrados ordinarios (MCO)

El estimador de mínimos cuadrados ordinarios proporciona el ajuste lineal (en base a unos regresores \mathbf{w}_i) de menor error cuadrático para una muestra de datos observados cualesquiera.

Supongamos que queremos aproximar un vector \mathbf{y} de dimensión n , conocido, mediante otro vector $\hat{\mathbf{y}}$ que sea combinación lineal de k vectores conocidos \mathbf{w}_i . Es decir, buscamos una proyección lineal del vector \mathbf{y} sobre el espacio generado por los \mathbf{w}_i . Podemos agrupar los vectores \mathbf{w}_i en la matriz \mathbf{W} , y estamos buscando un vector $\hat{\mathbf{y}}$ de la forma

$$\hat{\mathbf{y}} = \mathbf{W}\mathbf{b}$$

Esta aproximación $\hat{\mathbf{y}}$ presentará unas diferencias \mathbf{e} con el vector \mathbf{y} , definidas por

$$\mathbf{e} = \mathbf{y} - \mathbf{W}\mathbf{b}$$

donde \mathbf{y} , \mathbf{e} son vectores de dimensión n ; \mathbf{b} es un vector de dimensión k ; \mathbf{W} es una matriz ($n \times k$) cuyas columnas son los vectores \mathbf{w}_i .

Si buscamos la combinación $\hat{\mathbf{y}} = \mathbf{W}\mathbf{b}$ que mejor aproxima \mathbf{y} en el sentido de menor suma de errores cuadráticos (distancia euclídea), la solución es la denominada solución de mínimos cuadrados ordinarios (MCO). Se trata de proyectar \mathbf{y} sobre las columnas de \mathbf{W} . La solución de mínima distancia buscada, que denotamos $\hat{\mathbf{y}}_{\text{MCO}} = \mathbf{W}\mathbf{b}_{\text{MCO}}$, cumple que su proyección sobre las columnas de \mathbf{W} es la misma que la proyección de \mathbf{y} sobre las columnas de \mathbf{W} , es decir:

$$\mathbf{W}'\mathbf{W}\mathbf{b}_{\text{MCO}} = \mathbf{W}'\mathbf{y}$$

de donde

$$\mathbf{b}_{\text{MCO}} = (\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}'\mathbf{y}$$

supuesto que $(\mathbf{W}'\mathbf{W})^{-1}$ exista.

Obsérvese que, por construcción del estimador, los residuos \mathbf{e} son ortogonales a los vectores explicativos \mathbf{w}_i : $\mathbf{W}'\mathbf{e} = \mathbf{0}$, por lo que nos referimos a esta proyección como proyección ortogonal. Estas k condiciones de ortogonalidad (euclídea) del vector de residuos \mathbf{e} con cada vector \mathbf{w}_i (regresores) caracterizan al estimador MCO.

Nótese que no hemos hecho ninguna hipótesis sobre el modelo que genera los datos, ninguna hipótesis sobre la estructura o la distribución de probabilidad de los residuos \mathbf{e} . Simplemente hemos minimizado la suma cuadrática de residuos muestrales $(\mathbf{y}-\mathbf{Wb})$. En el caso de dos variables, hemos encontrado la recta que mejor ajusta los datos (en el sentido de menor error cuadrático), independientemente de que el modelo generador de los datos sea o no sea lineal.

Demostración de la fórmula de cálculo de \mathbf{b}_{MCO}

El vector buscado \mathbf{b}_{MCO} es el \mathbf{b} que minimiza el error cuadrático $(\mathbf{y}-\mathbf{Wb})'(\mathbf{y}-\mathbf{Wb})$. Por derivación con respecto a \mathbf{b} , la condición para \mathbf{b}_{MCO} es

$$\mathbf{W}'(\mathbf{y}-\mathbf{Wb}_{MCO}) = \mathbf{0} ,$$

de donde se sigue que

$$\mathbf{b}_{MCO} = (\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}'\mathbf{y} .$$

I.3 Mínimos cuadrados generalizados (MCG). Estimador de Gauss-Markov

Al considerar modelos lineales generadores de las observaciones, $\boldsymbol{\beta}$ representará el vector de coeficientes de un modelo lineal $\mathbf{y} = \mathbf{W}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, y representaremos mediante \mathbf{b} un estimador de $\boldsymbol{\beta}$

Para datos provenientes de un modelo lineal estocástico $\mathbf{y} = \mathbf{W}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, el estimador de mínimos cuadrados generalizados proporciona el estimador lineal de menor error cuadrático medio del parámetro $\boldsymbol{\beta}$.

Observemos que, frente al caso anterior de mínimos cuadrados ordinarios, ahora admitimos una hipótesis sobre el modelo que genera los datos. Quizá podamos aprovechar esa información para realizar una estimación lineal de $\boldsymbol{\beta}$ que minimice la esperanza del error cuadrático (error cuadrático **medio**), en vez de minimizar la suma cuadrática de residuos (error cuadrático) particular de la muestra de datos con la que trabajamos.

Así pues, supongamos que los datos son generados por el modelo

$$y = \mathbf{W}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad \text{con } E(\boldsymbol{\varepsilon}) = \mathbf{0} \text{ y } E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = \boldsymbol{\Sigma},$$

y busquemos el estimador lineal \mathbf{b} de $\boldsymbol{\beta}$ que minimiza el error cuadrático medio.

Para el caso particular en que suponemos (a priori) que las perturbaciones ($\boldsymbol{\varepsilon}$) siguen una distribución iid $(0, \sigma^2)$, es decir que la matriz $E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = \sigma^2\mathbf{I}$, el estimador lineal ($\mathbf{b}=\mathbf{K}\mathbf{y}$) de $\boldsymbol{\beta}$ que minimiza el error cuadrático esperado coincide con el estimador \mathbf{b}_{MCO} calculado anteriormente

$$\mathbf{b}_{\text{MCO}} = (\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}'\mathbf{y}$$

Con carácter general, siendo $E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = \boldsymbol{\Sigma}$, tenemos que el estimador \mathbf{b} de $\boldsymbol{\beta}$ que minimiza el error cuadrático esperado es el estimador de mínimos cuadrados generalizados (MCG):

$$\mathbf{b}_{\text{MCG}} = (\mathbf{W}'\boldsymbol{\Sigma}^{-1}\mathbf{W})^{-1}\mathbf{W}'\boldsymbol{\Sigma}^{-1}\mathbf{y}$$

cuya matriz de covarianzas es

$$E[(\boldsymbol{\beta} - \mathbf{b}_{\text{MCG}})(\boldsymbol{\beta} - \mathbf{b}_{\text{MCG}})'] = (\mathbf{W}'\boldsymbol{\Sigma}^{-1}\mathbf{W})^{-1}$$

Al estimador MCG del modelo considerado también se le conoce como estimador lineal insesgado de mínima varianza o estimador de Gauss-Markov. El estimador MCG pondera el peso de cada observación (y_i) conforme a las características de su perturbación (ε_i) en relación al resto de perturbaciones (por las varianzas y covarianzas de las perturbaciones): penaliza a las observaciones con alta variabilidad y adapta los pesos de cada observación teniendo en cuenta las covarianzas.

Hemos pasado de un estimador que minimiza el error cuadrático para una muestra dada a un estimador que minimiza el error cuadrático esperado, siempre que los datos se ajusten a (sean generados conforme a) el modelo lineal estocástico heterocedástico considerado.

Anexo I

La interpretación en términos de ortogonalidad es más complicada al trabajar con variables aleatorias. Buscábamos un estimador lineal $\mathbf{b}=\mathbf{K}\mathbf{y}$. Las condiciones de ortogonalidad del estimador $E[(\mathbf{K}\mathbf{y}-\boldsymbol{\beta})\mathbf{y}'] = \mathbf{0}$ son $k \times n$ condiciones de ortogonalidad de cada residuo e_i a cada variable y_i , o condición de ortogonalidad de cada residuo al subespacio generado por las y_i . Estas condiciones caracterizan al estimador MCG. En realidad, caracterizan al estimador $\mathbf{b} = \mathbf{K} \mathbf{y}$ que minimiza la norma

$$\| \mathbf{K}\mathbf{y}-\boldsymbol{\beta} \|^2 = E [(\mathbf{K}\mathbf{y}-\boldsymbol{\beta})'(\mathbf{K}\mathbf{y}-\boldsymbol{\beta})]$$

Es decir, caracterizan al estimador de menor error cuadrático medio. Una posible interpretación es ésta: la correlación es una medida de la relación lineal entre unas variables; si ajustamos un modelo lineal de forma que la correlación entre los residuos y las variables del modelo sea cero, no hay más relación lineal que extraer en el modelo.

Se puede dar una interpretación intuitiva al estimador MCG. Supongamos que disponemos de tres observaciones que sabemos son una muestra particular de un modelo

$$y_i = a + \beta w_i + \varepsilon_i \quad i = 1, 2, 3.$$

Donde sabemos además que la perturbación ε_i es despreciable para las observaciones 1 y 2, pero es muy elevada para la observación 3. Queremos estimar los parámetros a y β del modelo a partir de las observaciones. Con la información disponible sobre variabilidad de las perturbaciones, es de esperar que la recta que estamos tratando de estimar pase muy cerca de las dos primeras observaciones, si bien la tercera no es tan fiable. Eso es precisamente lo que tiene en cuenta el estimador MCG, que penaliza la “atracción” que ejerce la observación 3 sobre la recta de regresión debido a su mayor variabilidad.

En la figura I.1 se han representado tres observaciones obtenidas del modelo $y_i = a + w_i\beta + \varepsilon_i$, con varianzas despreciables para las dos observaciones de la izquierda y varianza elevada para la observación de la derecha. Una estimación por mínimos cuadrados ordinarios proporcionaría la recta marcada como MCO, que es la que mejor aproxima los datos de la muestra. Una estimación por mínimos cuadrados generalizados tendría en cuenta que la tercera observación es menos fiable y proporcionaría la recta MCG, de parámetros más próximos (en general) a los de la recta objetivo.

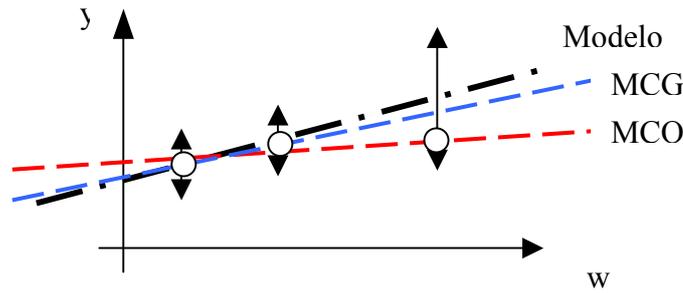


Figura I.1. Una comparación intuitiva de los estimadores MCO y MCG

I.4 Proyección lineal. Caso general

Con carácter general, siendo β e y vectores aleatorios cualesquiera, se cumple que el estimador lineal b de β ($b = K y$) que minimiza la esperanza del error cuadrático

$$E[(\beta - b)'(\beta - b)]$$

es

$$b_{MECM} = E(\beta y') [E(y y')]^{-1} y$$

lo cual puede demostrarse a partir de las ecuaciones normales

$$E[(\beta - K y) y'] = 0$$

Son las mismas condiciones que para el estimador de MCG, pero allí β se suponía constante.

Se tiene además que la matriz de covarianzas del error del estimador es

$$E[(\beta - b)(\beta - b)'] = E(\beta\beta') - E(\beta y') [E(y y')]^{-1} E(y \beta')$$

Al estimador $b_{MECM} = E(\beta y') [E(y y')]^{-1} y$ se le denomina también proyección lineal de β sobre y .

Proyección lineal en un modelo lineal estocástico con coeficientes estocásticos

Basándonos en los resultados anteriores, podemos generalizar aún más el modelo lineal $\mathbf{y} = \mathbf{W}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ admitiendo la posibilidad de que $\boldsymbol{\beta}$ sea también un vector aleatorio del que conocemos cierta información. Utilizaremos esa información que tengamos sobre $\boldsymbol{\beta}$ para conseguir un estimador con el menor error cuadrático medio posible.

Supongamos el caso particular

$$\mathbf{y} = \mathbf{W}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad \text{con} \quad E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = \boldsymbol{\Sigma}, \quad E(\boldsymbol{\beta}\boldsymbol{\beta}') = \mathbf{C} \quad \text{y} \quad E(\boldsymbol{\varepsilon}\boldsymbol{\beta}') = \mathbf{0}$$

donde \mathbf{y} es un vector conocido de dimensión n , \mathbf{W} es una matriz conocida constante ($n \times k$), $\boldsymbol{\beta}$ es un vector aleatorio desconocido de dimensión k y $\boldsymbol{\varepsilon}$ es un vector aleatorio desconocido de dimensión n , y donde asumimos que $\mathbf{W}\mathbf{C}\mathbf{W}' + \boldsymbol{\Sigma}$ es no singular.

Para este modelo, particularizando la fórmula general, se tiene que el estimador lineal \mathbf{b} de $\boldsymbol{\beta}$ ($\mathbf{b} = \mathbf{K}\mathbf{y}$) que minimiza el error cuadrático medio $E[(\boldsymbol{\beta} - \mathbf{b})'(\boldsymbol{\beta} - \mathbf{b})]$ es

$$\mathbf{b}_{\text{MECM}} = \mathbf{C}\mathbf{W}'(\mathbf{W}\mathbf{C}\mathbf{W}' + \boldsymbol{\Sigma})^{-1} \mathbf{y}$$

y la matriz de covarianzas del error es

$$E[(\boldsymbol{\beta} - \mathbf{b})(\boldsymbol{\beta} - \mathbf{b})'] = \mathbf{C} - \mathbf{C}\mathbf{W}'(\mathbf{W}\mathbf{C}\mathbf{W}' + \boldsymbol{\Sigma})^{-1}\mathbf{W}\mathbf{C}$$

Expresado de forma alternativa tenemos

$$\mathbf{b}_{\text{MECM}} = (\mathbf{W}'\boldsymbol{\Sigma}^{-1}\mathbf{W} + \mathbf{C}^{-1})^{-1} \mathbf{W}'\boldsymbol{\Sigma}^{-1} \mathbf{y}$$

La forma alternativa permite comparar el estimador directamente con el de MCG. En definitiva, la información de covarianzas de $\boldsymbol{\beta}$ contenida en la matriz \mathbf{C} nos permite encontrar un estimador de menor $E(\|\mathbf{b} - \boldsymbol{\beta}\|^2)$ que el estimador de MCG.

1.5 Filtro de Kalman

El filtro de Kalman es un algoritmo de cálculo recursivo del estimador de menor error cuadrático medio del denominado “vector de estados”. Es aplicable sobre sistemas cuya

dinámica viene representada mediante unas ecuaciones conocidas como “modelo del espacio de estados”, o simplemente “espacio de estados”.

El espacio de estados es un modelo particular dentro de los modelos lineales estocásticos con coeficientes estocásticos. Viene caracterizado por una ecuación vectorial en diferencias para los coeficientes (llamados estados):

$$\beta_{t+1} = \phi_t \beta_t + u_t$$

que se conoce como ecuación de transición de estados, y una ecuación lineal para las observaciones:

$$y_t = W_t \beta_t + e_t$$

que se conoce como ecuación de observaciones. Debe cumplirse además que las perturbaciones e_t y u_t , de esperanza cero, no presenten autocorrelación temporal, siendo sus matrices de covarianzas $E(u_t u_t') = \Sigma_t$ y $E(e_t e_t') = C_t$, definida positiva.

El filtro de Kalman, aplicable sobre estos modelos, es un algoritmo de cálculo recursivo del estimador de menor error cuadrático medio (MECM) de los estados, y proporciona también la matriz de covarianzas del error de estimación.

Para no asociar excesivamente la interpretación del vector de estados con la interpretación de coeficientes variables de regresión, lo cual puede resultar incómodo o limitante en algunos casos, prescindiremos del símbolo β_t y adoptaremos la notación y_t para el vector de observaciones y z_t para el vector de estados.

Dado su carácter recursivo, el filtro de Kalman precisa partir de una estimación (\hat{z}_0) del estado inicial (z_0) y de una estimación de la matriz de covarianzas del error de la estimación inicial, es decir, una estimación (P_0) de $E[(\hat{z}_0 - z_0)(\hat{z}_0 - z_0)']$.

Considerando el modelo

$$z_{t+1} = A_t z_t + u_t \quad \text{Ecuación de transición de estados}$$

$$y_t = C_t z_t + e_t \quad \text{Ecuación de observaciones}$$

Anexo I

con las condiciones indicadas anteriormente y siendo las matrices de covarianzas $E(\mathbf{u}_t \mathbf{u}_t') = \Sigma_t$ y $E(\mathbf{e}_t \mathbf{e}_t') = \mathbf{R}_t$, definida positiva, las ecuaciones del filtro de Kalman son (Luenberger, 1969):

$$\mathbf{e}^-_t = \mathbf{y}_t - \mathbf{C}_t \mathbf{z}^e_{t|t-1}$$

$$\mathbf{K}_t = \mathbf{P}_t \mathbf{C}'_t [\mathbf{C}_t \mathbf{P}_t \mathbf{C}'_t + \mathbf{R}_t]^{-1}$$

$$\mathbf{z}^e_{t+1|t} = \mathbf{A}_t (\mathbf{z}^e_{t|t-1} + \mathbf{K}_t \mathbf{e}^-_t)$$

$$\mathbf{P}_{t+1} = \mathbf{A}_t (\mathbf{P}_t - \mathbf{K}_t \mathbf{C}_t \mathbf{P}_t) \mathbf{A}'_t + \Sigma_t$$

Donde $\mathbf{z}^e_{t+1|t}$ es el valor esperado del vector de estados para el periodo $t + 1$ cuando se cuenta con observaciones hasta el periodo t , y \mathbf{P}_{t+1} es la matriz de covarianzas de \mathbf{z}_{t+1} condicionada a las observaciones anteriores hasta el periodo t .

El estimador de \mathbf{y}_t cuando se cuenta con observaciones hasta \mathbf{y}_{t-1} proporcionado por el filtro de Kalman es

$$\hat{\mathbf{y}}_{t|t-1} = \mathbf{C}_t \mathbf{z}^e_{t|t-1}$$

Este es el estimador lineal de menor error cuadrático medio de \mathbf{y}_t . Si el estimador lineal es óptimo, coincide (Hamilton 1994, pp. 73 y 77) con la esperanza condicionada $E(\mathbf{y}_t | \mathbf{y}_0, \dots, \mathbf{y}_{t-1})$, y la matriz de covarianzas condicionada viene dada por

$$E[(\mathbf{y}_t - \hat{\mathbf{y}}_{t|t-1})(\mathbf{y}_t - \hat{\mathbf{y}}_{t|t-1})'] = \mathbf{C}_t \mathbf{P}_t \mathbf{C}'_t + \mathbf{R}_t$$

En el algoritmo de cálculo, \mathbf{e}^-_t y \mathbf{K}_t son variables auxiliares. La variable \mathbf{e}^-_t es el error de predicción, diferencia entre el valor esperado y el valor observado de \mathbf{y}_t . El vector $\mathbf{K}_t \mathbf{e}^-_t$ es el vector de ganancias de Kalman, y corrige (mejora) la estimación del vector de estados en el instante t al obtener la nueva observación \mathbf{y}_t . Nótese que la estimación del vector de estados en el instante t cuando se cuenta con la observación \mathbf{y}_t es precisamente $\mathbf{z}^e_{t|t-1} + \mathbf{K}_t \mathbf{e}^-_t$. Esto quiere decir que, si la observación \mathbf{y}_t de la salida del sistema no coincide con la que se había previsto (la previsión es $\mathbf{C}_t \mathbf{z}^e_{t|t-1}$), se tiene un error de predicción $\mathbf{e}^-_t = \mathbf{y}_t - \mathbf{C}_t \mathbf{z}^e_{t|t-1}$ no nulo y se modifica en consecuencia la estimación del estado del sistema.

Análisis de Sistemas Cointegrados mediante Métodos de Subespacios

La denominación de “vector de estados” se debe a que este vector contiene la información suficiente sobre el estado del sistema en un instante determinado como para poder determinar la evolución y la respuesta en el instante siguiente, si conociéramos los valores de las entradas al sistema y de las perturbaciones. En este sentido, el vector de estados condensa toda la información histórica del sistema relevante para su evolución futura. Por evolucionar conforme a un proceso de Markov de primer orden, el camino seguido hasta llegar a un estado determinado no es relevante para su evolución futura: toda la información relevante está contenida en el vector de estados.

Maybeck (1979) presenta una interesante introducción al filtro de Kalman, con ejemplos físicos y de ingeniería. Uno de los ejemplos que utiliza para ilustrar la utilidad del filtro es el cálculo de la posición (z_t) de un barco que se mueve por una trayectoria unidimensional; cada cierto tiempo podemos contar con medidas (y_t) de su distancia al origen, y cada una de estas medidas con precisión distinta (variabilidades de e_t distintas y conocidas):

$$y_t = z_t + e_t$$

Sabemos además que el estado del barco (su posición o distancia al origen, z_t) evoluciona conforme a la ecuación

$$z_{t+1} = z_t + v \Delta t + w_t$$

donde v es la velocidad nominal del barco y w_t es una variable normal de esperanza cero y varianza $\sigma_w^2 \Delta t$, asociada a factores aleatorios que pueden influir en la velocidad, como pueden ser las ráfagas de viento. En estas condiciones, el filtro de Kalman permite obtener estimaciones óptimas (menor error cuadrático medio) de la posición del barco, combinando de manera adecuada el valor de la posición esperada conforme a la ecuación de estados, con el valor de cada nueva medida que se vaya obteniendo, ponderando ambas conforme a su variabilidad respectiva (variabilidad del viento y precisión de los instrumentos de medida de la posición utilizados).

El algoritmo no se restringe a procesos estacionarios, sino que incluye también los no estacionarios, siempre que el proceso sea estable. El carácter recursivo del filtro permite además una rápida actualización de los estimadores al ir recibiendo nuevas observaciones. Esta característica de actualización rápida recursiva presenta especial

interés en procesos donde se necesite una pronta respuesta ante nuevas observaciones del sistema (De la Fuente, Hernández y Del Olmo 1989).

El filtro de Kalman proporciona un estimador lineal óptimo. Si en el modelo de espacio de estados las perturbaciones son gaussianas, el estimador óptimo es lineal, y es entonces proporcionado por el filtro de Kalman. Si las perturbaciones no son gaussianas, el filtro óptimo no es lineal, y el filtro de Kalman sólo proporciona el estimador óptimo dentro de los lineales.

Desde el punto de vista de la estimación de parámetros del modelo, el filtro de Kalman suele utilizarse en econometría como una herramienta que permite el cálculo recursivo de la función de verosimilitud de una muestra. Como hemos indicado, si el modelo es lineal y las perturbaciones son gaussianas, el filtro de Kalman proporciona de forma recursiva la esperanza y la matriz de covarianzas condicionadas del vector de observaciones. Además, con funciones gaussianas, la esperanza y la varianza son suficientes para definir la función de probabilidad, por lo que, en ese caso, el filtro de Kalman proporciona de forma recursiva la función de probabilidad condicionada. Esto, a su vez, permite calcular la función de verosimilitud de la muestra, lo que posibilita el cálculo de estimadores máximo-verosímiles de parámetros del modelo. Si el procedimiento indicado se aplica con perturbaciones no gaussianas, se obtienen estimadores denominados cuasi-máximo-verosímiles.

I.6 Resumen de los estimadores de mínimos cuadrados

Proyección lineal. Caso general.

Siendo β e y vectores aleatorios cualesquiera, el estimador lineal \mathbf{b} de β ($\mathbf{b} = \mathbf{K} y$) que minimiza la esperanza del error cuadrático $E \|\beta - \mathbf{b}\|^2$ es

$$\mathbf{b}_{\text{MECM}} = E(\beta y') [E(y y')]^{-1} y$$

Mínimos Cuadrados Ordinarios (MCO)

Independientemente del modelo generador de los datos, siendo \mathbf{y} el vector de observaciones y \mathbf{W} la matriz de regresores, el vector \mathbf{b} que minimiza el error cuadrático $\|\mathbf{y} - \mathbf{Wb}\|^2$ de ajuste lineal de la muestra es

$$\mathbf{b}_{\text{MCO}} = (\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}'\mathbf{y}$$

Mínimos Cuadrados Generalizados (MCG)

Si se tiene el modelo $\mathbf{y} = \mathbf{W}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, donde $\boldsymbol{\beta}$ es constante y $\boldsymbol{\varepsilon}$ es aleatorio con matriz de covarianzas $\boldsymbol{\Sigma}$, entonces el vector \mathbf{b} de la forma $\mathbf{b} = \mathbf{K}\mathbf{y}$ (lineal) que minimiza el error cuadrático medio $E\|\boldsymbol{\beta} - \mathbf{b}\|^2$ es

$$\mathbf{b}_{\text{MCG}} = (\mathbf{W}'\boldsymbol{\Sigma}^{-1}\mathbf{W})^{-1}\mathbf{W}'\boldsymbol{\Sigma}^{-1}\mathbf{y}$$

Si se tiene el caso particular $\varepsilon_i \sim \text{iid}(0, \sigma^2)$, entonces \mathbf{b}_{MCG} coincide con \mathbf{b}_{MCO}

Estimador de menor error cuadrático medio (MECM) en un modelo lineal de coeficientes estocásticos

Si se tiene el modelo $\mathbf{y} = \mathbf{W}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, donde $\boldsymbol{\beta}$ y $\boldsymbol{\varepsilon}$ son vectores aleatorios con $E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = \boldsymbol{\Sigma}$, $E(\boldsymbol{\beta}\boldsymbol{\beta}') = \mathbf{C}$, y $E(\boldsymbol{\varepsilon}\boldsymbol{\beta}') = \mathbf{0}$, entonces el estimador lineal ($\mathbf{b} = \mathbf{K}\mathbf{y}$) que minimiza el error cuadrático medio $E\|\boldsymbol{\beta} - \mathbf{b}\|^2$ es

$$\mathbf{b}_{\text{MECM}} = \mathbf{C}\mathbf{W}'(\mathbf{W}\mathbf{C}\mathbf{W}' + \boldsymbol{\Sigma})^{-1}\mathbf{y}$$

Si $E(\boldsymbol{\beta}\boldsymbol{\beta}')^{-1} = \mathbf{0}$, lo que puede interpretarse como que no hay información en el modelo (información a priori) sobre $\boldsymbol{\beta}$, entonces \mathbf{b}_{MECM} coincide con \mathbf{b}_{MCG} .

El filtro de Kalman es un algoritmo para calcular de forma recursiva este estimador \mathbf{b}_{MECM} , supuesto un modelo lineal dinámico de primer orden para la evolución de los estados $\boldsymbol{\beta}_{t+1} = \mathbf{A}_t\boldsymbol{\beta}_t + \mathbf{w}_t$ (que incluye la perturbación adicional \mathbf{w}_t), característico del espacio de estados.

I.7 Aplicaciones del filtro de Kalman en econometría

Los primeros desarrollos de la metodología de espacio de estados tuvieron lugar en el campo de la ingeniería en vez de la estadística, comenzando por el famoso artículo de Kalman (1960). En ese artículo, Kalman mostró dos aspectos cruciales. En primer lugar, mostró que una amplia variedad de problemas podían ser formulados mediante un modelo lineal simple: el modelo de espacio de estados. En segundo lugar, mostró cómo, debido a la naturaleza Markoviana del modelo, los cálculos necesarios para la aplicación práctica del modelo podían realizarse de forma recursiva, de modo particularmente conveniente para un computador. Hasta comienzos de los años 80, las contribuciones de estadistas y econométricos a la metodología del espacio de estados fueron aisladas y esporádicas. Sin embargo, en los últimos años el interés de la estadística y la econometría por el espacio de estados ha sido creciente.

Predicción, suavizado y filtrado de los estados, conocido el modelo

Dentro de la teoría de estimación, podemos distinguir entre filtrado, predicción y suavizado (o alisado). Entendemos por predicción la estimación del valor que tomará una variable en el periodo $(t + p)$ utilizando la información disponible hasta el periodo t . Filtrado es la estimación del valor que toma en el periodo t una variable no directamente observable utilizando la información hasta el periodo t . Suavizado es la estimación del valor que tomó en el periodo $(t - p)$ una variable no directamente observable, utilizando la información disponible hasta el periodo t .

Conocido el modelo en el espacio de estados, el filtro de Kalman proporciona recursivamente el estimador lineal óptimo de la secuencia de estados en base a observaciones del sistema. Las fórmulas pueden adaptarse sin gran dificultad para proporcionar los estimadores suavizados (Durbin y Koopman 2001) y la predicción es sencilla a partir del valor estimado del estado del sistema.

Estimación Kalman-máxima-verosimilitud de parámetros del modelo

El filtro de Kalman proporciona de forma recursiva el estimador de mínima varianza de los estados, así como la matriz de covarianzas de ese estimador. Aparte de la predicción de valores utilizando el modelo, su aplicación principal en econometría es la estimación de parámetros de un modelo. Esto es posible porque el filtro de Kalman permite calcular

recursivamente la función de verosimilitud de una muestra¹, expresada en función de los parámetros desconocidos del modelo (Durbin y Koopman 2001). La maximización por métodos numéricos de la función de verosimilitud proporciona estimaciones máximo verosímiles (MLE) de los parámetros. Este enfoque se muestra en la figura I.2.

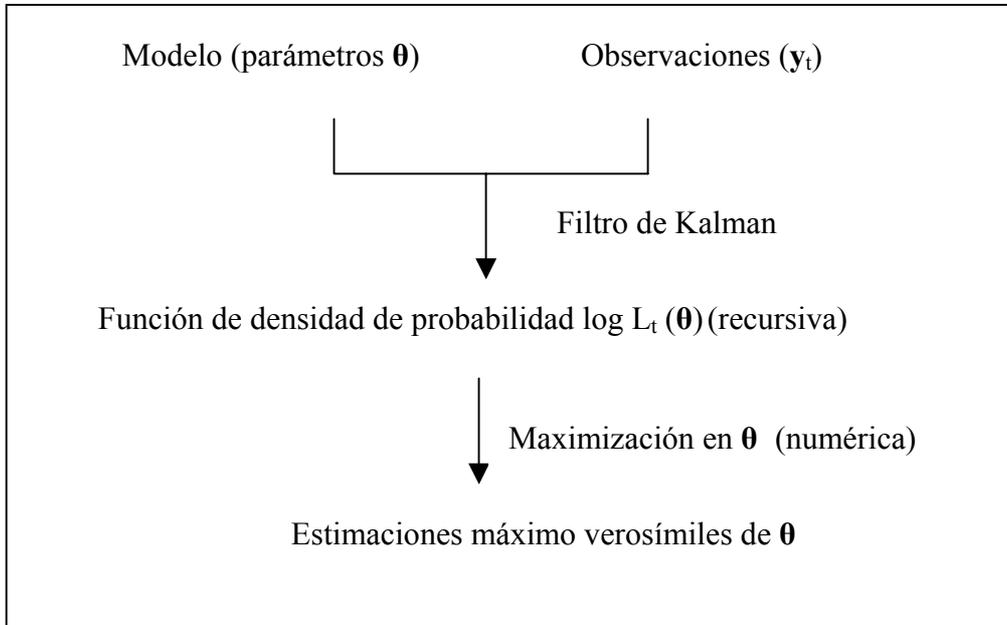


Figura I.2. Utilización del filtro de Kalman para la obtención de estimadores máximo verosímiles de los parámetros de un modelo

La idea es la siguiente: podemos expresar la función de densidad de probabilidad conjunta (L) de una serie temporal en función de las densidades de probabilidad condicionadas de cada observación dadas las anteriores, de la forma

$$L = \prod_{t=1}^T p(\mathbf{y}_t | \mathbf{y}_0, \dots, \mathbf{y}_{t-1})$$

Donde $p(\mathbf{y}_t | \mathbf{y}_0, \dots, \mathbf{y}_{t-1})$ es la función de densidad de probabilidad de la observación \mathbf{y}_t condicionada a todas las anteriores. Si las perturbaciones y el estado inicial siguen distribuciones normales (multivariante), entonces cada distribución condicionada es

¹ Suponiendo que las perturbaciones sean gaussianas, de forma que la media y varianza sean suficientes para definir la función de probabilidad, y el estimador lineal de menor error cuadrático medio proporcione la esperanza condicionada.

Anexo I

también normal. Además, una distribución normal (condicionada) viene completamente caracterizada por su media (condicionada) y matriz de covarianzas (condicionada), y estas dos variables son proporcionadas por el filtro de Kalman (I.5):

$$p(\mathbf{y}_t | \mathbf{y}_0, \dots, \mathbf{y}_{t-1}) = (2\pi)^{-n/2} | \mathbf{C} \mathbf{P}_t \mathbf{C}' + \mathbf{R} |^{-1/2} \\ \times \exp \left\{ -\frac{1}{2} [\mathbf{y}_t - \mathbf{C} \mathbf{z}_{t|t-1}]' [\mathbf{C} \mathbf{P}_t \mathbf{C}' + \mathbf{R}]^{-1} [\mathbf{y}_t - \mathbf{C} \mathbf{z}_{t|t-1}] \right\}$$

Podemos construir entonces el logaritmo de la función de verosimilitud de la muestra como

$$\text{Log } L = \sum_{t=1}^T \log p(\mathbf{y}_t | \mathbf{y}_0, \dots, \mathbf{y}_{t-1})$$

Expresando las matrices del sistema en función de un conjunto de parámetros θ desconocidos del modelo, la fórmula anterior proporciona la función de verosimilitud de la muestra en función de θ , y podemos maximizar numéricamente esta función con respecto a los parámetros desconocidos θ .

Si las perturbaciones no son gaussianas, la función de verosimilitud obtenida será sólo una aproximación a la real, y las estimaciones de los parámetros obtenidas por este procedimiento (tratando el modelo como si fuera gaussiano) se denominan cuasi-máximo verosímiles (QMLE) o pseudo-máximo verosímiles.

El filtro de Kalman se puede aplicar también sobre el modelo que considera coeficientes dinámicos (por ejemplo, para estimación de vectores autorregresivos con coeficientes dinámicos).

Hamilton (1994, cap. 13), indica también la utilidad del filtro para factorización de las funciones matriciales de generación de autocovarianzas o densidades espectrales.

Anexo II Realizaciones mínimas

Consideremos un sistema que evoluciona conforme al siguiente modelo representado en el espacio de estados

$$\mathbf{z}_{t+1} = \mathbf{A} \mathbf{z}_t + \mathbf{K} \mathbf{e}_t \quad \text{Ecuación de transición de estados}$$

$$\mathbf{y}_t = \mathbf{C} \mathbf{z}_t + \mathbf{e}_t \quad \text{Ecuación de observaciones}$$

con $E(\mathbf{e}_k \mathbf{e}'_j) = \mathbf{R} \delta_{kj}$. Como hemos indicado, el problema de identificación (también llamado problema de representación o de realización) es el de encontrar, a partir de observaciones del sistema, un conjunto de matrices $\{\mathbf{A}^*, \mathbf{K}^*, \mathbf{C}^*, \mathbf{R}\}$ tales que el modelo en el espacio de estados construido con esas matrices genere, al ser excitado por un ruido blanco de esperanza cero y matriz de covarianzas \mathbf{R} , una salida con las mismas propiedades estadísticas¹ que \mathbf{y}_t .

Sin embargo, existen infinitas combinaciones de matrices $\{\mathbf{A}^*, \mathbf{K}^*, \mathbf{C}^*\}$ que ante la entrada \mathbf{e}_t generarían una salida \mathbf{y}_t con las mismas propiedades estadísticas². Dentro de las posibles representaciones, una representación mínima es aquella en la que la dimensión del vector de estados es mínima.

Tampoco una representación mínima es única. Dada una representación mínima $\{\mathbf{A}, \mathbf{K}, \mathbf{C}, \mathbf{R}\}$, el sistema de matrices $\{\mathbf{A}^*, \mathbf{K}^*, \mathbf{C}^*, \mathbf{R}\} = \{\mathbf{TAT}^{-1}, \mathbf{TK}, \mathbf{CT}^{-1}, \mathbf{R}\}$ constituye una representación mínima equivalente, correspondiente al cambio de base de los estados $\mathbf{z}^* = \mathbf{T} \mathbf{z}$, donde \mathbf{T} es una matriz invertible cualquiera de dimensiones $n \times n$.

Se tiene que una representación es mínima si y sólo si es observable y controlable (Kalman, 1966). Introducimos brevemente a continuación estos dos conceptos.

¹ Hasta segundo orden, lo cual sería suficiente para que las propiedades estadísticas fueran las mismas si consideramos procesos gaussianos.

² Distintas combinaciones de las matrices del modelo que corresponden a la misma relación o función de transferencia entrada-salida.

Observabilidad

La observabilidad de un modelo de un sistema en el espacio de estados es la capacidad que presenta el modelo para reconstruir los estados a partir del conocimiento de las entradas y las salidas. Así, un modelo de un sistema dinámico lineal es observable si y sólo si su estado es determinable de forma única a partir del conocimiento de las ecuaciones del modelo, de los valores de las entradas y de los valores de las salidas (Grewal y Andrews, 2001, p. 43).

Para los sistemas lineales discretos de parámetros constantes que estamos considerando, la matriz de observabilidad del modelo se define como

$$\mathbf{O} = \begin{pmatrix} \mathbf{C} \\ \mathbf{CA} \\ \mathbf{CA}^2 \\ \vdots \\ \mathbf{CA}^{n-1} \end{pmatrix}$$

siendo n la dimensión del vector de estados. Se demuestra que el modelo del sistema es observable si la matriz de observabilidad es de rango completo.

Nótese que un cambio de base de los estados $\mathbf{z}^* = \mathbf{T} \mathbf{z}$ para \mathbf{T} invertible supone un cambio en las matrices que representan el sistema y una nueva matriz de observabilidad de la forma $\mathbf{O}^* = \mathbf{O}\mathbf{T}^{-1}$, pero este cambio de base no afecta al rango de la matriz de observabilidad.

Controlabilidad

La controlabilidad de un modelo lineal de un sistema discreto es la capacidad que presenta el modelo para poder alcanzar cualquier estado final deseado en un número de pasos finito, mediante una secuencia apropiada de valores de las entradas.

Anexo II

Para los sistemas lineales discretos estocásticos de parámetros constantes que estamos considerando, la matriz de controlabilidad del modelo es¹ (siendo n la dimensión del vector de estados):

$$\Psi = [\mathbf{K} \ \mathbf{AK} \ \mathbf{A}^2\mathbf{K} \ \dots \ \mathbf{A}^{n-1}\mathbf{K}]$$

Se demuestra que el modelo del sistema es controlable si la matriz de controlabilidad es de rango completo.

Nótese que un cambio de base de los estados $\mathbf{z}^* = \mathbf{T} \mathbf{z}$ para \mathbf{T} invertible supone un cambio en las matrices que representan el sistema y una nueva matriz de controlabilidad de la forma $\Psi^* = \mathbf{T} \Psi$, pero este cambio de base no afecta al rango de la matriz de controlabilidad.

¹ En los algoritmos de identificación estocástica suele jugar un papel importante la matriz $\mathbf{\Pi}$, que algunos autores denominan matriz de controlabilidad estocástica, o, simplemente, matriz de controlabilidad, en el contexto estocástico (Aoki y Havenner 1991, Van Oberschee y De Moor 1996), definida como

$$\mathbf{\Pi} = [\mathbf{\Omega} \ \mathbf{A}\mathbf{\Omega} \ \mathbf{A}^2\mathbf{\Omega} \ \dots \ \mathbf{A}^{n-1}\mathbf{\Omega}]$$

donde $\mathbf{\Omega}$ es la matriz $E(\mathbf{z}_{t+1|t} \mathbf{y}'_t)$ y n es la dimensión del vector de estados. Aoki (1990), en cambio, hace la distinción semántica entre matriz de controlabilidad Ψ y matriz de controlabilidad estocástica $\mathbf{\Pi}$.

Anexo III . Formulaciones alternativas del espacio de estados

Al abordar el estudio del espacio de estados encontraremos que no existe unicidad entre los distintos autores, ni en la forma utilizada de las ecuaciones, ni en los símbolos utilizados para referirse a los vectores y matrices del modelo. Tomaremos los símbolos y_t para representar el vector de observaciones y z_t para el vector de estados.

Para sistemas discretos, el modelo más general que podemos encontrar - Kalman (1960), Luenberger (1969), Anderson y Moore (1979)-, es de la forma

$$z_{t+1} = A_t z_t + v_t \quad \text{Ecuación de transición de estados}$$

$$y_t = C_t z_t + e_t \quad \text{Ecuación de observaciones}$$

donde z_t es un vector de n estados no directamente observables; y_t es un vector de m observaciones; A_t y C_t son matrices que, de momento, se suponen conocidas en cada periodo, pudiendo variar entre periodos; v_t , e_t son vectores aleatorios de esperanza cero y tales que

$$E(v_t v'_s) = Q_t \delta_{ts},$$

$$E(e_t e'_s) = R_t \delta_{ts} \quad \text{donde } R_t \text{ es definida positiva}$$

$$E(v_t e'_s) = S_t \delta_{ts}$$

siendo δ_{ts} la función delta de Kronecker.

Algunas de las hipótesis, como la ausencia de correlación temporal de las perturbaciones, pueden relajarse – ver p.e. Grewal y Andrews (2001, pp. 84, 85, 129)-.

En el espacio de estados, la dinámica del sistema está contenida en la ecuación de transición de estados, que es una ecuación vectorial en diferencias de primer orden (recordemos que una ecuación en diferencias de orden n es expresable como una ecuación vectorial en diferencias de primer orden). De este modo, los estados guardan, de forma más o menos condensada, la información histórica del sistema que influye en su evolución futura.

Vale la pena considerar otras formulaciones del espacio de estados frecuentemente utilizadas en la literatura. Todas ellas son expresables conforme a la formulación general indicada anteriormente, y constituyen casos particulares de la misma.

Autores como Harvey (1989) o Hamilton (1994) utilizan una formulación con las perturbaciones desplazadas un periodo en la ecuación de transición de estados, de la forma

$$\mathbf{z}_{t+1} = \mathbf{\Phi} \mathbf{z}_t + \mathbf{v}_{t+1} \quad \text{Ecuación de transición de estados}$$

$$\mathbf{y}_t = \mathbf{H} \mathbf{z}_t + \mathbf{D} \mathbf{x}_t + \mathbf{e}_t \quad \text{Ecuación de observaciones}$$

Harvey (1989, pp. 103, 112) discute las implicaciones de esta representación desplazada.

Para modelos de coeficientes (matrices) constantes, Terceiro (1990) utiliza una representación de la forma

$$\mathbf{z}_{t+1} = \mathbf{\Phi} \mathbf{z}_t + \mathbf{\Gamma} \mathbf{u}_t + \mathbf{E} \mathbf{v}_t \quad \text{Ecuación de transición de estados}$$

$$\mathbf{y}_t = \mathbf{H} \mathbf{z}_t + \mathbf{D} \mathbf{u}_t + \mathbf{C} \mathbf{w}_t \quad \text{Ecuación de observaciones}$$

donde \mathbf{y}_t son variables observadas, \mathbf{z}_t son los estados del sistema (en principio no observables), \mathbf{u}_t es de forma explícita un vector de variables exógenas (o variables de control o entradas al sistema), las matrices \mathbf{H} , \mathbf{D} , \mathbf{C} , $\mathbf{\Phi}$, $\mathbf{\Gamma}$, \mathbf{E} se suponen conocidas, \mathbf{v}_t y \mathbf{w}_t son perturbaciones ruido blanco de esperanza $\mathbf{0}$, sin correlación entre periodos y con matrices de varianzas y covarianzas $E(\mathbf{v}_t \mathbf{v}'_t)$, $E(\mathbf{w}_t \mathbf{w}'_t)$ y $E(\mathbf{v}_t \mathbf{w}'_t)$ conocidas.

Una de las formulaciones más habituales es la denominada “representación en forma de innovaciones”, en la que el mismo vector de perturbaciones de las observaciones (innovaciones) aparece en la ecuación de estados. Por ejemplo, para coeficientes constantes y modelo en innovaciones, Aoki utiliza la representación

$$\mathbf{z}_{t+1} = \mathbf{A} \mathbf{z}_t + \mathbf{B} \mathbf{e}_t \quad \text{Ecuación de transición de estados}$$

$$\mathbf{y}_t = \mathbf{C} \mathbf{z}_t + \mathbf{e}_t \quad \text{Ecuación de observaciones}$$

donde \mathbf{e}_t es ruido blanco de esperanza $E(\mathbf{e}_t) = \mathbf{0}$ y matriz de covarianzas $E(\mathbf{e}_t \mathbf{e}'_t) = \mathbf{R}$. Se puede comprobar que utilizar esta representación en forma de innovaciones no supone pérdida de generalidad (Hannan y Deistler 1988, Aoki 1990).

Anexo IV . Identificación de sistemas estocásticos por métodos de subespacios

IV.1 Enfoque de Van Overschee y De Moor

Resumimos a continuación los principales aspectos y teoremas relativos a la identificación de sistemas estacionarios estocásticos por métodos de subespacios, conforme al enfoque de Van Overschee y De Moor (1996).

Van Overschee y De Moor, (1996, p. 58) formulan el problema general de identificación estocástica en el espacio de estados con coeficientes constantes de la siguiente forma:

Dado un conjunto de T observaciones del vector \mathbf{y}_t , de dimensión m (m series temporales), correspondiente a un sistema de parámetros desconocidos que evoluciona conforme a las ecuaciones

$$\mathbf{z}_{t+1} = \mathbf{A} \mathbf{z}_t + \mathbf{w}_t \quad \text{Ecuación de transición de estados}$$

$$\mathbf{y}_t = \mathbf{C} \mathbf{z}_t + \mathbf{v}_t \quad \text{Ecuación de observaciones}$$

donde \mathbf{w}_t y \mathbf{v}_t son vectores ruido blanco de esperanza cero que cumplen

$$E(\mathbf{w}_k \mathbf{w}'_j) = \mathbf{Q} \delta_{kj}, \quad E(\mathbf{v}_k \mathbf{v}'_j) = \mathbf{R} \delta_{kj} \quad \text{y} \quad E(\mathbf{w}_k \mathbf{v}'_j) = \mathbf{S} \delta_{kj}$$

Encontrar:

- El orden n del sistema (especificación del modelo).
- Unas matrices \mathbf{A} y \mathbf{C} (susceptibles de transformación correspondiente a un cambio de base de los estados) y unas matrices \mathbf{Q} , \mathbf{R} y \mathbf{S} , tales que los estadísticos de segundo orden de la salida del modelo y de la salida observada sean iguales.

El teorema principal de la identificación estocástica de Van Overschee y De Moor (1996, p. 74), que a continuación enunciamos, es la base para el cálculo de una secuencia de estados a partir de observaciones del sistema:

- *Bajo las hipótesis:*

1 El ruido de proceso \mathbf{w}_t y el ruido de medida \mathbf{v}_t no son idénticamente cero.

2 El número de observaciones tiende a infinito.

3 Las matrices definidas por el usuario \mathbf{W}_1 de dimensiones $(m_i \times m_i)$ y \mathbf{W}_2 de dimensiones $(j \times j)$ son tales que \mathbf{W}_1 es de rango completo y \mathbf{W}_2 cumple que $\text{rango}(\mathbf{Y}_p) = \text{rango}(\mathbf{Y}_p, \mathbf{W}_2)$, donde \mathbf{Y}_p es la matriz tipo Hankel por bloques que contiene las observaciones “pasadas”, tal como se definió en la sección 4.2 .

- Siendo Φ_i la matriz de proyección lineal de las observaciones futuras en las pasadas¹:

$$\Phi_i = \mathbf{Y}_f / \mathbf{Y}_p \approx \mathbf{Y}_f \mathbf{Y}_p' (\mathbf{Y}_p \mathbf{Y}_p')^\dagger \mathbf{Y}_p$$

Y considerando la descomposición en valores singulares

$$\mathbf{W}_1 \Phi_i \mathbf{W}_2 = \begin{bmatrix} \mathbf{U}_1 & \mathbf{U}_2 \end{bmatrix} \begin{bmatrix} \Sigma_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{V}_1' \\ \mathbf{V}_2' \end{bmatrix} = \mathbf{U}_1 \Sigma_1 \mathbf{V}_1'$$

- *Se tiene:*

1 La matriz Φ_i es igual al producto de la matriz de observabilidad extendida \mathbf{O}_i y la secuencia de estados \mathbf{Z}_i que proporcionaría un filtro de Kalman:

$$\Phi_i = \mathbf{O}_i \mathbf{Z}_i$$

2 El orden del sistema a identificar es igual al número de valores singulares distintos de cero en la descomposición anterior.

¹ Utilizamos la notación desarrollada en el apartado 4.2.

Anexo IV

3 Una matriz de observabilidad extendida \mathbf{O}_i y la matriz de controlabilidad extendida asociada $\mathbf{\Pi}_i$ pueden calcularse como

$$\mathbf{O}_i = \mathbf{W}_1^{-1} \mathbf{U}_1 \boldsymbol{\Sigma}_1^{1/2}$$

$$\mathbf{\Pi}_i = \mathbf{O}_i^\dagger \mathbf{E}(\mathbf{Y}_f \mathbf{Y}_p')$$

4 La secuencia de estados asociada a la matriz de observabilidad \mathbf{O}_i es:

$$\mathbf{Z}_i = \mathbf{O}_i^\dagger \boldsymbol{\Phi}_i$$

Este teorema puede resumirse algebraicamente mediante las expresiones:

- Rango $(\mathbf{Y}_f / \mathbf{Y}_p) = n$
- Espacio generado por las filas de $\mathbf{Y}_f / \mathbf{Y}_p =$ espacio generado por las filas de \mathbf{Z}_i
- Espacio generado por columnas de $\mathbf{Y}_f / \mathbf{Y}_p =$ espacio generado por columnas de \mathbf{O}_i

Los resultados de este teorema nos permiten estimar, a partir de las observaciones, el orden del sistema y la secuencia de estados, sin necesidad de conocer o estimar previamente las matrices del modelo. Una vez conocida la secuencia de estados, la estimación de estas matrices es un problema sencillo de mínimos cuadrados.

Nótese que, en vez de trabajar directamente sobre la matriz $\boldsymbol{\Phi}_i$, el algoritmo trabaja sobre la matriz $\mathbf{W}_1 \boldsymbol{\Phi}_i \mathbf{W}_2$. El efecto de la matriz \mathbf{W}_1 es acondicionar la matriz $\boldsymbol{\Phi}_i \mathbf{W}_2$ antes de llevar a cabo la descomposición en valores singulares. Aunque en teoría su efecto desaparece al estimar la matriz de observabilidad como $\mathbf{O}_i = \mathbf{W}_1^{-1} \mathbf{U}_1 \boldsymbol{\Sigma}_1^{1/2}$, puede tener implicaciones prácticas sobre el resultado de la descomposición en valores singulares y permite dotar de mayor generalidad al algoritmo presentado, de forma que distintos algoritmos existentes en la literatura pueden englobarse dentro de este algoritmo general. Por su parte, el efecto de fijar una matriz \mathbf{W}_2 determinada es el de fijar una base (sistema coordenado) para la secuencia de estados.

IV.2 Un enfoque alternativo de los métodos de subespacios

Un enfoque de los algoritmos de subespacios distinto al de Van Overschee y De Moor (1996) es el que proporcionan Ljung¹ (1999) o Verhaegen (1991), centrado en la estimación consistente de una matriz de observabilidad del sistema. Indicaremos las principales diferencias de este enfoque.

Por aplicación recursiva de las ecuaciones del espacio de estados se tiene

$$\mathbf{y}_{t+f-1} = \mathbf{CA}^{f-1} \mathbf{z}_t + \mathbf{CA}^{f-2} \mathbf{e}_t + \mathbf{CA}^{f-3} \mathbf{e}_{t+1} + \dots + \mathbf{C} \mathbf{e}_{t+f-2} + \mathbf{e}_{t+f-1}$$

de donde

$$\mathbf{y}_t^f = \mathbf{O}_f \mathbf{z}_t + \mathbf{H} \mathbf{e}_t^f$$

siendo $\mathbf{e}_t^f \equiv [\mathbf{e}'_t, \mathbf{e}'_{t+1}, \dots, \mathbf{e}'_{t+f-1}]'$ y siendo \mathbf{H} una matriz triangular inferior compuesta por bloques identidad (en la diagonal) y bloques de productos \mathbf{CA}^j .

Definiendo las matrices

$$\mathbf{Y}_f \equiv [\mathbf{y}'_p, \mathbf{y}'_{p+1}, \dots, \mathbf{y}'_T]$$

$$\mathbf{Z} \equiv [\mathbf{z}'_p, \mathbf{z}'_{p+1}, \dots, \mathbf{z}'_T]$$

$$\mathbf{E}_f \equiv [\mathbf{e}'_p, \mathbf{e}'_{p+1}, \dots, \mathbf{e}'_T]$$

podemos expresar matricialmente la relación anterior como

$$\mathbf{Y}_f = \mathbf{O}_f \mathbf{Z} + \mathbf{H} \mathbf{E}_f$$

Si ahora encontramos una matriz de instrumentos Ψ que elimine asintóticamente el término de las perturbaciones y tal que el producto $\mathbf{Z} \Psi$ sea (asintóticamente) una matriz de rango n , tendremos la igualdad (asintótica) $\mathbf{Y}_f \Psi = \mathbf{O}_f \mathbf{Z} \Psi = \mathbf{O}_f \mathbf{T}$, donde \mathbf{T} es una matriz de rango n , luego el producto $\mathbf{O}_f \mathbf{T}$ es también una matriz de observabilidad del sistema. Teniendo esto en cuenta, es posible extraer, por descomposición en valores

¹ Ljung proporciona una visión más general que la que aquí resumimos, e incluye variables exógenas.

singulares de la matriz $\mathbf{Y}_f \mathbf{\Psi}$, un estimador consistente de una matriz de observabilidad del sistema.

Podemos elegir como matriz $\mathbf{\Psi}$ de instrumentos las observaciones pasadas \mathbf{Y}_p , que están incorrelacionadas con las perturbaciones futuras¹. De este modo, para N observaciones, la matriz \mathbf{G} definida por $\mathbf{G} = \frac{1}{N} \mathbf{Y}_f \mathbf{Y}'_p$ puede constituir el estimador ruidoso que buscábamos para obtener una matriz de observabilidad del sistema.

Para aportar flexibilidad al algoritmo, se puede trabajar con la matriz $\mathbf{W}_1 \mathbf{G} \mathbf{W}_2$. De igual modo a como se vio previamente, la descomposición en valores singulares $\mathbf{W}_1 \mathbf{G} \mathbf{W}_2 = \mathbf{U} \mathbf{\Sigma} \mathbf{V}' \approx \mathbf{U}_1 \mathbf{\Sigma}_1 \mathbf{V}'_1$ (truncando las matrices conforme al orden estimado del sistema) proporcionaría, mediante la fórmula $\mathbf{O}_f = \mathbf{W}_1^{-1} \mathbf{U}_1 \mathbf{T}^{-1}$, un estimador de una matriz de observabilidad del sistema, para cualquier matriz \mathbf{T}^{-1} que se elija libremente².

Obsérvese que la proyección lineal $\mathbf{Y}_f \mathbf{Y}'_p (\mathbf{Y}_p \mathbf{Y}'_p)^{\dagger} \mathbf{Y}_p$ con la que trabajábamos en el enfoque anterior es en particular una matriz de la forma $\mathbf{W}_1 \mathbf{G} \mathbf{W}_2$, tomando $\mathbf{W}_1 = \mathbf{I}$ y $\mathbf{W}_2 = \mathbf{N} (\mathbf{Y}_p \mathbf{Y}'_p)^{\dagger} \mathbf{Y}_p$. Esto implica que, desde el punto de vista de su aplicación práctica, este algoritmo centrado en la estimación de una matriz de observabilidad se trata en definitiva del mismo algoritmo que habíamos estudiado anteriormente.

Desde el punto de vista de la interpretación teórica, es un enfoque de variables instrumentales centrado en el cálculo de un estimador consistente de una matriz de observabilidad extendida del sistema \mathbf{O}_f . Encontramos que la proyección ortogonal de las observaciones futuras sobre los datos históricos puede proporcionar ese estimador

¹ Al hacer esta elección, con la matriz \mathbf{Y}_p tal como se definió previamente, estamos tomando como instrumentos un número i de observaciones pasadas. Ljung hace notar que no hay razón particular para tomar ese número igual a i y que podría tomarse otro valor s , aunque la mayoría de los algoritmos de subespacios toman $s = i$.

² La elección libre de las matrices \mathbf{W}_2 (de rango completo) y \mathbf{T}^{-1} (invertible) determina la base de representación para la secuencia de estados.

buscado¹. Bajo este punto de vista, los métodos de subespacios estiman en primer lugar, de forma consistente y mediante procedimientos de proyección lineal a partir de las observaciones entrada-salida, una matriz de observabilidad extendida del sistema, \mathbf{O}_f .

A partir de aquí seguiríamos el mismo tratamiento que en el caso anterior:

- Podemos utilizar la matriz de observabilidad para obtener estimadores de las matrices \mathbf{A} y \mathbf{C} .

- Si queremos modelo para el ruido, estimaríamos una secuencia de estados basándonos en la relación que existe entre la proyección lineal $\mathbf{Y}_f / \mathbf{Y}_p$ y el producto $\mathbf{O}_f \mathbf{Z}$. Con la secuencia de estados, estimaríamos por mínimos cuadrados el resto de las matrices.

- Alternativamente, podemos estimar desde el principio la secuencia de estados \mathbf{y} , a partir de ésta, todas las matrices del modelo por mínimos cuadrados.

Nótese que la matriz \mathbf{G} del enfoque de Ljung está formada por los mismos bloques de covarianzas muestrales que la matriz \mathbf{H} del método de Aoki. De hecho, si en este método que estamos estudiando tomamos $\mathbf{W}_1 = \mathbf{I}_{mi}$, y $\mathbf{W}_2 = \mathbf{I}_{mi}$, estaríamos trabajando con la misma matriz que el método de Aoki (método de componentes principales), y si elegimos $\mathbf{T} = \boldsymbol{\Sigma}_1^{1/2}$, estaríamos eligiendo la misma matriz de observabilidad. Del mismo modo (Ljung 1999, p. 351), si elegimos las matrices

$$\mathbf{W}_1 = (\mathbf{j}^{-1} \mathbf{Y}_f \mathbf{Y}'_f)^{-1/2}$$

$$\mathbf{W}_2 = (\mathbf{j}^{-1} \mathbf{Y}_p \mathbf{Y}'_p)^{-1/2},$$

estaríamos empleando el método de correlaciones canónicas CCA.

Resumiendo, bajo este enfoque alternativo, las principales etapas de los métodos de subespacios

son:

¹ En realidad, basta con el término $\mathbf{Y}_f \mathbf{Y}'_p$ de la expresión matricial de cálculo de la proyección ortogonal (Ljung, 1999, p. 349). Los términos finales pueden interpretarse como la multiplicación por una matriz de acondicionamiento, que fija la base del espacio de estados.

1 Cálculo de una matriz \mathbf{G} como estimación (ruidosa) de una matriz de observabilidad extendida del sistema \mathbf{O}_f . \mathbf{G} es calculada a partir de observaciones entrada-salida y contiene información sobre el orden del sistema y sobre las matrices de su representación en el espacio de estados. Permitirá además calcular la secuencia de estados. Para sistemas estocásticos puede tomarse la matriz $\mathbf{G} = \frac{1}{N} \mathbf{Y}_f \mathbf{Y}'_p$

$$\mathbf{G} = \frac{1}{N} \mathbf{Y}_f \mathbf{Y}'_p$$

2 Estimación del orden del sistema mediante una descomposición en valores singulares de la matriz acondicionada $\mathbf{W}_1 \mathbf{G} \mathbf{W}_2 = \mathbf{U} \mathbf{\Sigma} \mathbf{V}' \approx \mathbf{U}_1 \mathbf{\Sigma}_1 \mathbf{V}'_1$ (truncadas conforme al orden estimado).

3 Elección de una matriz de observabilidad $\mathbf{O}_f = (\mathbf{W}_1)^{-1} \mathbf{U}_1 \mathbf{T}$, donde \mathbf{T} es una matriz invertible de orden adecuado elegida arbitrariamente.

4 Estimación de las matrices $\{\mathbf{A}, \mathbf{C}\}$ del sistema. Para ello pueden seguirse dos caminos:

4.1 Estimar \mathbf{A} y \mathbf{C} directamente a partir de \mathbf{O}_f .

4.2 Estimar la secuencia de estados a partir de \mathbf{O}_f . Con la secuencia de estados, las matrices \mathbf{A} y \mathbf{C} pueden calcularse por mínimos cuadrados.

5 Estimación de las matrices $\{\mathbf{Q}, \mathbf{R}, \mathbf{S}\}$ de covarianzas de las perturbaciones. Para ello debe calcularse la secuencia de estados (si no se hizo ya en 4.2).

Referencias

Anderson, B.D.O. y Moore, J.B. (1979). *Optimal Filtering*. Englewood Cliffs (N.J.), Prentice Hall.

Aoki, M. (1976). *Optimal Control and System Theory in Dynamic Analysis*. North-Holland.

Aoki, M. (1983). *Note on Economic Time Series Analysis: System Theoretic approach*. Springer-Verlag. Heidelberg.

Aoki, M. (1990). *State Space Modeling of Time Series*. Second, Revised and Enlarged Edition. Springer-Verlag.

Aoki, M. (1997). "The SSATS Algorithm and Subspace Methods". *Applications of Computer Aided Time Series Modeling*, pp. 3-13. Springer-Verlag New York.

Aoki, M. y Havenner, A. (1991). "State Space Modeling of Multiple Time Series", *Econometric Reviews*, 10 (1), pp. 1-59.

Arndt, C. y Foster, K. (1997). "A State Space Model of Monthly US Wheat Prices". *Applications of Computer Aided Time Series Modeling*, pp. 91-105. Springer-Verlag New York.

Banerjee, A., Dolado, J., Galbraith, J.W. y Hendry, D.F. (1993). *Co-integration, Error-Correction, and the Econometric Analysis of Non-stationary Data*. Advanced Texts in Econometrics. Oxford University Press.

Bauer, D. y Ljung, L. (2002). "Some Facts about the Choice of the Weighting Matrices in Larimore Type of Subspace Algorithms". *Automatica* 38, Issue 5, 763-773

Bauer, D. y Wagner, M. (2002). "Estimating Cointegrated Systems using Subspace Algorithms". *Journal of Econometrics* 111, 47-84.

Bauer, D. y Wagner, M. (2003). "A Canonical Form for Unit Root Processes in the State Space Framework". Documento de trabajo de la Universidad de Berna, Suiza. Actualizado el 1 de Julio de 2003.

Bauer, D. y Wagner, M. (2003b). "The Performance of Subspace Algorithm Cointegration Analysis: A Simulation Study". Documento de trabajo de la Universidad de Berna, Suiza. Actualizado mayo 2003.

Bennet, R.J. (1979). *Spatial Time Series*. Pion Limited.

Clark, P.K. (1987). "The Cyclical Component of U.S. Economic Activity". *The Quarterly Journal of Economics*, 102, 797-814.

Deistler, M. (1996). "Time Series Econometrics". *CWI Quarterly* 9, 3, pp. 165-179.

De Cock, K. y De Moor, B. (2003). "Subspace Identification Methods", contribución a la sección 5.5, "Control systems robotics and automation" de *EOLSS, UNESCO Encyclopedia of Life Support Systems*, (Unbehauen H.D.), vol. 1 de 3, Eolss Publishers Co., Ltd. (Oxford, UK), 2003, pp. 933-979.

De la Fuente, D., Hernández, C. y Del Olmo, R. (1989). "Control Adaptativo de Calidad en Procesos Industriales". *QÜESTIÓ*, vol. 13, pp. 105-125.

Del Hoyo, J.L. y Cendejas, J.L. (2003). "El Ciclo Económico de las Islas Baleares en relación con los Ciclos Español y Europeo". I Congreso de Economía de las Islas Baleares. Accesible en http://conecoib.caib.es/comunicacions/23_analisis_ciclo_economico.pdf a fecha 5 de octubre de 2004.

Del Olmo, R. y Hernández, C. (1991). "Estabilidad Paramétrica y Riesgo Percibido: Comportamiento del Mercado Español de Capitales". *Anales de Estudios Económicos y Empresariales*, vol. 6, pp. 163-176.

Diebold, F.X. y López, J.A. (1995). "Forecast Evaluation and Combination". Escrito para G.S. Maddala y C.R. Rao (eds.), *Handbook of Statistics*. Amsterdam: North-Holland. Accesible en www.frbsf.org/economics/economists/jlopez/forecast_eval_combo.pdf a fecha 1 de junio de 2004.

Durbin, J. y Koopman, S. J. (2001). *Time Series Analysis by State Space Methods*. Oxford University Press.

Elliot, G., Rothenberg, T.J. y Stock, J.H. (1996). "Efficient Tests for an Autoregressive Unit Root". *Econometrica* 64(4): 813-836.

Referencias

- Engel, R. y Granger, C. (1987). "Co-integration and Error Correction: Representation, Estimation and Testing". *Econometrica* 35: 251-276.
- Escribano Sáez, A. (1990). "Introducción al Tema de Cointegración y Tendencias". *Cuadernos Económicos de ICE*, N° 44, 1990/1, 7-42.
- Escribano, A. y Peña, D. (1994). "Cointegration and Common Factors". *Journal of Time Series Analysis* 15: 577-86.
- Fuertes, A. M. (1997). *Desarrollos Recientes en la Teoría de la Cointegración: Contraste de la Hipótesis PPA*. Tesis doctoral, Universidad de Valladolid.
- Gantmacher, R.R. (1960). *The Theory of Matrices, Volume 2*. New York: Chelsea Publishing Company.
- García-Ferrer, A. y del Hoyo, J. (1992). "On Trend Extraction Models: Interpretation, Empirical Evidence and Forecasting Performance". *Journal of Forecasting*, 11, 645-665.
- García-Ferrer, A. y Poncela, P. (2002). "Forecasting European GNP Data through Common Factor Models and other Procedures". *Journal of Forecasting*, 21, 4, 225-244.
- Gilbert, P.D. (1993). "State Space and ARMA Models: An Overview of the Equivalence". ISBN 0-662-20494-8.
- Golub, G.H. y Van Loan, C.F. (1983). *Matrix Computations*. Johns Hopkins University Press. Baltimore.
- Gonzalo, J. (1994). "Five Alternative Methods of Estimating Long-run Equilibrium Relationships", *Journal of Econometrics*, 60, pp. 203-233.
- Gourieroux, C. y Monfort, A. (1997), *Time Series and Dynamic Models*. Cambridge University Press.
- Greene, W. H. (1999). *Análisis Económico*. 3ª edición, Prentice Hall.
- Grewal, M.S. y Andrews, A.P. (2001). *Kalman Filtering. Theory and Practice Using Matlab*, 2nd Edition, John Wiley & Sons.

- Hamilton, J.D. (1994). *Time Series Analysis*. Princeton University Press.
- Hannan, E. y Deistler, M. (1988). *The Statistical Theory of Linear Systems*. Wiley, New York.
- Hansen, B. y Phillips, P.C.B. (1990). "Estimation and Inference in Models of Cointegration: A Simulation Study". *Advances in Econometrics* 8, 225-248.
- Harvey, A.C. y Peters, S. (1984). "Estimation Procedures for Structural Time-Series Models". London School of Economics, Discussion Paper n° A28.
- Harvey, A.C. (1989). *Forecasting, Structural Time Series Models and the Kalman filter*, Cambridge University Press.
- Havenner, A. (1997). "A Guide to State Space Modeling of Multiple Time Series". *Applications of Computer Aided Time Series Modeling*, pp. 15-72. Springer-Verlag New York.
- Harris, T.J., Seppala, C.T., y Bacon, D.W. (1999). *Dynamic Analysis of Variance Methods for Multivariate Control System Data*. 1st International Symposium on Industrial Statistics (Understanding Variation: A Key to Successful Quality Improvement), August 19-21 Linköping, Sweden. Disponible en <http://appsci.queensu.ca/tjh/tjhpapers.asp> a fecha 26 de septiembre de 2004.
- Hernández, C. (1975). *Métodos Cuantitativos para la Decisión Económica*. Tesis Doctoral. U.P. Barcelona.
- Ho, B. y Kalman, R.E. (1966). Efficient Construction of Linear State Variable Models from Input/output Functions. *Regelungstechnik*, 14:545-548.
- Intriligator, Bodkin y Hsiao (1996). *Econometric Models, Techniques and Applications*. 2nd Edition. Prentice Hall.
- Johansen, S. (1988). "Statistical Analysis of Cointegration Vectors". *Journal of Economic Dynamics and Control*, 12: 231-254.
- Kalman, R.E. (1960). "A New Approach to Linear Filtering and Prediction Problems", *J. Basic Engineering, Transactions ASMA, Series D*, 82, 35-45.

Referencias

Kalman, R.E. (1966). "On Structural Properties of Linear Constant, Multivariate Systems", paper 6A *Proceedings, International Federation of Automatic Control Congress, Prague*.

Larimore, W.E. (1996). "Statistical Optimality and Canonical Variate Analysis System Identification". *Signal Processing* 52, Issue 2, 131-144

Larimore, W.E. (2000). "Identification of Collinear and Cointegrated Multivariable Systems using Canonical Variate Analysis". IFAC SYSID'99.

Luenberger, D.G. (1969). *Optimization by Vector Space Methods*. John Wiley & Sons.

Lütkepohl, H. (1991). *Introduction to Multiple Time Series Analysis*. Springer-Verlag.

Ljung, L. (1999). *System Identification. Theory for the User*. 2nd edition. Prentice Hall.

Ljung, L. (2002). *System Identification Toolbox For Use with Matlab. User's Guide*. Version 5. The MathWorks.

Maybeck, P.S. (1979). *Stochastic Models, Estimation and Control*. Academic Press, Inc.

Nelson, C.R. y Plosser, C.I. (1982). "Trends and Random Walks in Macroeconomic Time Series: Some Evidence and Implications". *Journal of Monetary Economics* 10: 139-62.

Östermark, R. (1997). "Modeling Cointegrated Processes by a Vector-Valued State Space Algorithm". *Applications of Computer Aided Time Series Modeling*, pp. 3-13. Springer-Verlag New York.

Östermark, R. y Aoki, M. (1992). "Time Series Evidence of Impacts of the US Economy on the Scandinavian Economy". IFAC Workshop on Economic Time Series Analysis and System Identification, Vienna.

Peña, D. (1990). "Cointegración y Reducción de Dimensionalidad en Series Temporales Multivariantes". *Cuadernos Económicos de ICE*, N° 44, 1990/1, 109-126.

Peña, D. y Poncela, P. (2005). "Nonstationary Dynamic Factor Analysis". *Journal*

of Statistical Planning and Inference (en prensa, disponible en www.sciencedirect.com)

Pindyck y Rubinfeld (1998). *Econometric Models and Economic Forecasts*, 4th edition, Mc Graw-Hill.

Poskitt, D.S. (2000). “Strongly Consistent Determination of Cointegrating Rank Via Canonical Correlations”. *Journal of Business and Economic Statistics* 18, 77-90.

Siegel, S. (1985). *Estadística No Paramétrica Aplicada a las Ciencias de la Conducta*. Novena impresión, Trillas.

Sorenson, H.W. (1970). “Least-squares Estimation: from Gauss to Kalman”. *IEEE Spectrum*, vol. 7, pp. 63-68.

Stock, J.H. y Watson, M.W. (1988). “Testing for Common Trends”. *Journal of the American Statistical Association* 83, pp. 1097-1107.

Stock, J.H. y Watson, M.W. (1991). “A Probability Model of the Coincident Economic Indicators” in Lahiri, K. y Moore, G.H. eds., *Leading Economic Indicators: New Approaches and Forecasting Records*. Cambridge University Press.

Stock, J.H. y Watson, M.W. (2003). *Introduction to Econometrics*. Addison Wesley.

Terceiro, J. (1990). *Estimation of Dynamic Econometric Models with Errors in variables*. Berlin: Springer-Verlag.

Terceiro, J., Casals, J.M., Jerez, M., Serrano, G.R., Sotoca, S. (2000). *Time Series Analysis using MATLAB. Including a complete MATLAB Toolbox*. Reference Manual of E4. Accesible en <http://www.ucm.es/info/icae/e4> a fecha 11 de mayo de 2004.

Vaccaro, R.J. y Vukina, T. (1993). “A Solution to the Positivity Problem in the State-space Approach to Modeling Vector-valued Time Series”. *Journal of Economic Dynamics and Control* 17, pp. 401-421

Van Overschee, P. y De Moor, B. (1996). *Subspace Identification for Linear Systems: Theory – Implementation - Applications*. Dordrecht, The Netherlands: Kluwer Academic Publishers.

Referencias

Vargas, M. (1999). "Modelización de Series Temporales Estacionarias en el Espacio de Estados". Documento de trabajo 2/1999/4. Fac. CC.EE. y EE. Albacete, U.C.L.M.

Vargas, M. (2000). "Análisis de Tendencias Comunes y Cointegración en el Espacio de Estados". Documento de trabajo 2/2000/3. Fac. CC.EE. y EE. Albacete, U.C.L.M.

Viberg, M. (1994). "Subspace Methods in System Identification". *Proc. of SYSID '94*, Vol. 1, 4-6 July, Copenhagen, Denmark, pp. 1-12.

Verhaegen, M. (1991). "A Novel Non-iterative MIMO State Space Model Identification Technique". *Proc. 9th IFAC/IFORS Symposium on Identification and System Parameter Estimation*, 1453-1458, Budapest, Hungary.

Wagner, M. (1999). "VAR Cointegration in VARMA Models". *Economics Series 65*. Institut for Advanced Studies, Viena.

Wagner, M. (2000). *The Analysis of Cointegrated Systems: Structural Properties, Statistical Theory and Performance*. Dissertation. Institut für Ökonometrie, Operations Research und Systemtheorie. Technischen Universität Wien.

Wells, C. (1996). *The Kalman Filter in Finance*. Advanced Studies in Theoretical and Applied Econometrics, Kluwer Academic Publishers.

Notación. Principales símbolos y abreviaturas

a (minúscula)	-escalar
\mathbf{a} (minúscula negrita)	-vector (columna)
\mathbf{A} (mayúscula negrita)	-matriz
\mathbf{a}' , \mathbf{A}'	vector traspuesto de \mathbf{a} , matriz traspuesta de \mathbf{A}
\mathbf{A}^{-1}	matriz inversa de \mathbf{A}
\mathbf{A}^\dagger	matriz inversa de Moore-Penrose de \mathbf{A}
ACCA	algoritmo de correlaciones canónicas adaptado
CCA	algoritmo de correlaciones canónicas
CCA2	algoritmo de correlaciones canónicas bietápico
DVS	descomposición en valores singulares
δ_{kj}	función delta de Kronecker
\mathbf{e}_t	vector de perturbaciones o ruido. Salvo indicación expresa, supondremos ruido blanco gaussiano ($\mathbf{0}$, \mathbf{R}).
i.i.d.	variables independientes igualmente distribuidas
MCG	mínimos cuadrados generalizados
MCO	mínimos cuadrados ordinarios
ML	máxima verosimilitud
VAR	vector autorregresivo
VEC	vector autorregresivo en forma de “corrección del error”
\mathbf{x} , \mathbf{u}	vectores de variables exógenas o variables de control
\mathbf{y}	vector de variables endógenas o variables observadas
y_t	serie temporal, o valor de una serie temporal en el periodo t
$\hat{y}_{t+k t}$	predicción del valor y_{t+k} utilizando observaciones hasta el periodo t
\mathbf{z}	vector de variables de estado o vector de estados