

Estudios de Economía Aplicada
Nº 15, 2000. Págs. 125-162

Comparación de la capacidad predictiva de los modelos de coeficientes fijos frente a variables en los modelos econométricos regionales: un análisis para Cataluña

RAMOS LOBO, R.

CLAR LÓPEZ, M.

SURIÑACH CARALT, J.

*Grupo de Investigación Anàlisi Quantitativa Regional
Departamento de Econometría, Estadística y Economía Española
Facultad de Ciencias Económicas y Empresariales
Universidad de Barcelona*

Los autores agradecen el soporte recibido de a DGICYT proyecto SEC))-0700 y del Plan Nacional de I+D proyecto 2FD97-1004-C03-01. Así como los comentarios recibidos de un evaluador anónimo.

RESUMEN

Durante las últimas décadas se ha producido un creciente interés en nuestro país en relación a las economías regionales dada la necesidad de los gobiernos regionales de obtener información sobre sus economías para así llevar a cabo actuaciones de política económica más efectivas y eficientes. En este marco, los modelos econométricos constituyen una herramienta de utilidad puesto que ofrecen información sobre las relaciones estructurales que se dan en una economía y permiten predecir su evolución. Sin embargo, la utilización de dichos modelos con finalidad predictiva se enfrenta al inconveniente de la elevada inestabilidad a corto plazo que se produce en las relaciones entre variables económicas a nivel regional. Por este motivo, en el presente trabajo se propone la utilización de un modelo de coeficientes variables para recoger dicha inestabilidad y mejorar las predicciones sobre la evolución de las variables del bloque de producción de la economía catalana. Para contrastar la mejora obtenida a partir de la aplicación de dicho modelo, se compara su capacidad predictiva con la de un modelo de coeficientes fijos. Los resultados muestran un mejor comportamiento del modelo de coeficientes variables frente al modelo de coeficientes fijos.

Palabras clave: Modelos econométricos regionales, predicción, modelos de coeficientes variables, cointegración, filtro de Kalman.

ABSTRACT

During the last decades, there has been a growing interest in Spain in the issues related to regional economies. This interest is related to the need of regional governments to obtain information about their economies and act efficiently. In this context, econometric models are useful tools to offer information about the structural relationships of the economy and also to predict their evolution. However, the predictive capacity of these models is adversely affected by the potential instability of the relationships between economic variables at a regional level. For this reason, in this article we present a time-varying coefficient model to improve the predictions of an econometric model for the production block variables of the Catalan economy. To validate the proposed methodology predictions with a fixed coefficient model and with the varying coefficient one are compared. The results obtained show a better behaviour of the last one in changing situations for economic relationships.

Keywords: Regional econometric models, forecasting, time-varying coefficient models, cointegration, Kalman filter.

Clasificación AMS: 62M10, 62M20, 90A17, 90A20, 93E24 y 93E25.

Código UNESCO: 530201, 530202, 530203 y 530205.

Artículo recibido en marzo de 1999. Revisado en enero de 2000.

1. Introducción y objetivos¹

En nuestro país, a lo largo de las últimas décadas, se ha producido un creciente interés relacionado con las economías regionales. Entre los factores que de una u otra manera han colaborado en la dinamización de dicho interés cabe destacar los siguientes:

- a) el nuevo escenario político configurado tras la creación del Estado de las Autonomías y, en concreto, la redefinición de las funciones de los gobiernos regionales ha generado una necesidad de estadísticas regionales como herramienta para el seguimiento de la evolución de la realidad económica regional;
- b) la entrada en la Comunidad Europea en 1986 ha supuesto, a su vez, un incentivo para aproximar las estadísticas a la realidad regional como una de las vías para maximizar la eficiencia en la distribución de los fondos estructurales comunitarios;
- c) la creación de los Institutos de Estadística Regionales;
- d) la continuidad en el tiempo y el (relativo) buen comportamiento de modelos macroeconómicos nacionales como el Wharton-UAM y el MOISEES para analizar y predecir la evolución de la economía española;
- e) el desarrollo del *hardware* y del *software*, que ha hecho posible llevar a cabo estudios y proyectos que tan sólo hace unos años o bien eran muy costosos de realizar o no era posible abordarlos; y,
- f) la extensa tradición en otros países occidentales en lo que al análisis de cuestiones y problemas regionales se refiere.

En este contexto, los modelos econométricos constituyen una herramienta de gran utilidad puesto que ofrecen información sobre las relaciones estructurales que se dan en una economía y permiten obtener predicciones así como simular su evolución a corto y medio plazo. Dicha información puede ser utilizada por los *policy-makers* para adoptar aquellas medidas que más se adecuen a las necesidades reales de la economía en cuestión.

Desde este punto de vista, un objetivo primordial es, por tanto, la elaboración de un modelo econométrico que permita predecir la evolución de las economías de

1. Una versión previa de este trabajo fue presentada en la XXIII Reunión de Estudios Regionales celebrada en Valencia del 18 al 21 de noviembre de 1997.

ámbito regional. Sin embargo, uno de los principales inconvenientes a los que se enfrenta el modelizador es que las relaciones entre las variables económicas no son ni fijas ni invariables respecto al tiempo, especialmente a corto plazo y a nivel regional (Ramajo, 1996). De hecho, la evidencia empírica existente, así como la obtenida por Stock y Watson (1996), muestra la elevada frecuencia con que varían las relaciones entre diferentes variables económicas especialmente a corto plazo. Una posible solución a este problema consistiría en revisar de manera continuada la especificación del modelo econométrico y obtener únicamente predicciones a muy corto plazo, lo cual sin embargo, disminuiría su utilidad. De este modo, la solución óptima sería aquella que permitiese modelizar correctamente las relaciones a largo plazo entre las variables económicas de las que se pretende obtener predicciones e incorporar la posibilidad de que las relaciones existentes a corto plazo puedan variar a lo largo del tiempo para recoger la inestabilidad existente.

En este sentido, en este artículo se propone una estrategia de modelización econométrica basada en la teoría de la cointegración que hace posible la identificación y estimación de las relaciones a largo plazo entre las variables analizadas, pero donde también se intenta recoger la inestabilidad temporal en las relaciones a corto plazo a partir de un modelo de coeficientes variables con la finalidad de mejorar la capacidad predictiva del modelo considerado. De este modo, el objetivo concreto del trabajo consiste en comparar la capacidad predictiva del modelo propuesto con la capacidad predictiva de un modelo de coeficientes fijos a partir del análisis del bloque de producción de un modelo econométrico para Cataluña. En concreto, la variable endógena modelizada como *proxy* de la producción ha sido el Valor Añadido Bruto (VAB) anual para un nivel de desagregación sectorial de las cuatro principales ramas de actividad (agricultura, industria, construcción y servicios)² a precios de mercado y a pesetas constantes de 1986³.

La estructura del artículo es la siguiente: en primer lugar, en el siguiente apartado, se consideran aspectos relacionados con los modelos de coeficientes variables a

2. El motivo que ha llevado a abordar el estudio del bloque de producción es que es el bloque central del modelo, en el sentido que el resto de macromagnitudes dependen del comportamiento de las variables de producción.

3. Cabe señalar que una de las líneas de trabajo del equipo de investigación *Anàlisi Quantitativa Regional* de la *Universitat de Barcelona* es, desde hace algún tiempo, la elaboración de un modelo econométrico para Cataluña. Dicha línea de trabajo ha dado sus frutos. Actualmente existe un modelo econométrico para un nivel de desagregación sectorial de ocho ramas de actividad elaborado en el marco de la teoría de la cointegración. El modelo que se presenta en este trabajo no debe considerarse, por tanto, como el modelo de Cataluña sino como un estudio para avanzar en el entorno de dicha línea de investigación, incorporando nuevas técnicas econométricas de especificación y estimación con el fin de mejorar el modelovigente.

lo largo del tiempo como su especificación, su estimación a través del filtro de Kalman y, por último, cómo pueden ser utilizados para mejorar las predicciones de un modelo econométrico en el que también se tienen en cuenta las relaciones a largo plazo entre las variables. A continuación, en el tercer apartado, se presenta un modelo econométrico subsectorial para el bloque de producción de la economía catalana y se compara la capacidad predictiva de un modelo "tradicional" de coeficientes fijos en el que se incorpora información sobre los posibles cambios estructurales con la capacidad del modelo de coeficientes variables propuesto. Por último se presentan las principales conclusiones.

2. Los modelos de coeficientes variables: estrategia de modelización

2.1. Introducción

Los modelos de coeficientes variables a lo largo del tiempo pretenden incorporar al proceso de estimación la existencia de inestabilidad entre las relaciones consideradas. Estos modelos comenzaron a utilizarse a finales de los cincuenta y principios de los sesenta aunque hasta los últimos años y, gracias en parte a los avances en *software* y *hardware*, no ha sido cuando han recibido el impulso definitivo. Los trabajos realizados por Rubin (1950) y Quandt (1958, 1960) sentaron las bases del desarrollo de una línea de investigación que tenía como principal objetivo encontrar nuevos métodos de tratamiento de la inestabilidad de las relaciones entre las variables económicas. Durante principios de los setenta, los modelos de coeficientes variables a lo largo del tiempo fueron adquiriendo relevancia como consecuencia del sentimiento generalizado de que el supuesto clásico de estabilidad de los parámetros en los modelos de regresión a menudo no era válido. En la actualidad, este tipo de modelos son ampliamente utilizados como alternativa a los modelos de coeficientes fijos en aquellos casos en que estos últimos no ofrecen resultados satisfactorios. En concreto, los principales motivos que se argumentan para utilizar dichos modelos se pueden agrupar en tres categorías (Dzciechciarz, 1989; Engle y Watson, 1987; Min y Zellner, 1993):

- a) las relaciones entre las variables económicas consideradas pueden ser diferentes para los distintos subperíodos muestrales considerados como consecuencia de cambios estructurales debidos al progreso técnico o cambios institucionales;
- b) es posible que la teoría económica postule una relación cambiante entre las variables consideradas, por ejemplo, en términos de expectativas (Lucas, 1976); y,
- c) la existencia de errores de especificación en un modelo de regresión tales como la omisión de un regresor relevante o la utilización de una variable *proxy* en vez del regresor real, llevará generalmente a la obtención de unos residuos que no son ruido

blanco, y que pueden ser explicados parcialmente dejando que los coeficientes varíen a lo largo del tiempo. En este caso, sería necesario contrastar si el modelo obtenido mejora los resultados en términos predictivos en presencia de errores de especificación, tanto dentro como fuera de la muestra.

Así pues, en términos generales, un modelo de coeficientes variables a lo largo del tiempo está formado por dos ecuaciones: una primera ecuación que recoge el comportamiento a lo largo del tiempo de los coeficientes \mathbf{b}_t considerados y que viene dada por la siguiente expresión:

$$\mathbf{b}_t = \mathbf{f}_t \cdot \mathbf{b}_{t-1} + \mathbf{W}_t \cdot \mathbf{q}_t + \mathbf{h}_t, \quad (1)$$

donde \mathbf{f}_t representa la magnitud del cambio del coeficiente en cada instante del tiempo, es decir, en qué medida depende del valor anterior, \mathbf{W}_t denota un conjunto de variables que pueden aportar información sobre el comportamiento de \mathbf{b}_t , \mathbf{q}_t representa los coeficientes asociados a dichas variables y \mathbf{h}_t es un término de perturbación que sigue una distribución normal con esperanza cero y varianza \mathbf{s}_h^2 . Por su parte, la segunda ecuación recoge la relación existente entre los coeficientes \mathbf{b}_t , las variables endógenas Y_t y las variables exógenas X_t y Z_t :

$$Y_t = X_t \cdot \mathbf{b}_t + Z_t \cdot \mathbf{g} + \mathbf{e}_t, \quad (2)$$

donde \mathbf{g} representa los coeficientes fijos asociados a las variables exógenas Z_t y \mathbf{e}_t denota un término de perturbación que sigue una distribución normal con esperanza cero y varianza \mathbf{s}_e^2 .

En el modelo formado por ambas ecuaciones, los valores de \mathbf{f}_t y \mathbf{q}_t son los que determinan la evolución temporal de los coeficientes. En relación a este aspecto, la experiencia modelizadora en el entorno de la Economía así como argumentos teóricos en la línea de los ofrecidos por Engle y Watson (1987) sugieren utilizar en la mayoría de los casos valores para $\mathbf{f}_t = 1$ y $\mathbf{q}_t = 0$. Este caso particular de modelos de coeficientes variables se conocen como modelos de coeficientes sistemáticamente variables y, de hecho, en este tipo de modelos se supone que los coeficientes se comportan siguiendo un camino aleatorio (Shively y Kohn, 1997). También es importante destacar que en este modelo es posible permitir que las variaciones de un coeficiente sean totalmente independientes de las de los demás.

La principal ventaja de este tipo de modelos es que permiten que la relación existente entre las variables endógenas y exógenas de una ecuación de comportamiento de variables económicas puedan variar a lo largo del tiempo superando así parte de las desventajas de los modelos de coeficientes fijos en presencia de inestabilidad temporal.

En lo que se refiere a la estimación de este tipo de modelos se han propuesto diferentes técnicas como, por ejemplo, modelos de tipo bayesiano (Min y Zellner, 1993), regresiones *spline* (Engle y Watson, 1987), o la estimación por mínimos cuadrados flexibles (Kalaba y Testfasion, 1990; Lütkepohl y Herwartz, 1996), pero sin duda la más extendida es la utilización de los modelos *state-space* y el filtro de Kalman. En el siguiente apartado se profundiza en este aspecto.

2.2. Estimación de los modelos de coeficientes variables a través del filtro de Kalman

2.2.1. Los modelos state-space

Los modelos *state-space*, desarrollados inicialmente en la Ingeniería de Control, ofrecen la posibilidad de formular modelos dinámicos más generales que los utilizados habitualmente en diferentes entornos de la Economía, siendo éste el principal motivo que ha llevado a su mayor difusión y utilización entre los economistas durante los últimos años y, especialmente, en relación a la utilización de modelos de coeficientes variables.

El principal objetivo que se persigue con este tipo de modelos es obtener información sobre el comportamiento de \mathbf{a}_t , una variable no observable, que en el caso que nos ocupa se trataría de los coeficientes variables a lo largo del tiempo, a partir de su relación con Y_t , una variable de la que sí se dispone de observaciones. El punto de partida del modelo es suponer que el comportamiento dinámico de \mathbf{a} , sigue un proceso de Markov de primer orden que vendría dado por la siguiente expresión:

$$\mathbf{a}_t = T_t \cdot \mathbf{a}_{t-1} + c_t + R_t \cdot \mathbf{h}_t, \quad (3)$$

donde T_t es conocida como la matriz de transición, c_t es un vector de variables exógenas que pueden afectar el comportamiento dinámico de \mathbf{a}_t y R_t y \mathbf{h}_t denotan un término de perturbación con media cero y matriz de varianzas y covarianzas Q_t .

El modelo se completa con el supuesto de que la variable no observable \mathbf{a}_t y la variable observable Y_t se relacionan a partir de la siguiente expresión:

$$Y_t = Z_t \cdot \mathbf{a}_t + d_t + \mathbf{e}_t, \quad (4)$$

donde Z_t es la matriz que relaciona ambas variables, d_t es un vector de variables exógenas y \mathbf{e}_t es un término de perturbación aleatorio con media cero y matriz de varianzas H_t .

La ecuación (4) se conoce como ecuación de medida y junto a la ecuación (3), denominada ecuación de transición, forman el modelo *state-space*. La especifica-

ción del modelo *state-space* se completa con dos supuestos adicionales que hacen referencia a los valores iniciales del vector de estado y a la matriz de varianzas y covarianzas de los términos de perturbación:

$$\begin{aligned} E(\mathbf{a}_0) &= \mathbf{a}_0; \\ \text{var}(\mathbf{a}_0) &= \mathbf{P}_0; \end{aligned} \quad (5)$$

$$E(\mathbf{e}_t \cdot \mathbf{h}_s) = 0 \quad s, t = 1, \dots, T; \quad (6)$$

$$\begin{aligned} E(\mathbf{e}_t \cdot \mathbf{a}_0) &= 0 \quad t = 1, \dots, T; \mathbf{y}, \\ E(\mathbf{h}_t \cdot \mathbf{a}_0) &= 0 \quad t = 1, \dots, T. \end{aligned} \quad (7)$$

La principal dificultad para aplicar los modelos *state-space* en el entorno económico está relacionada con la dependencia del sistema de los parámetros Z_t, H_t, T_t, R_t, Q_t , conocidos como hiperparámetros y que son los que determinan las propiedades estocásticas del sistema, así como los valores iniciales de la variable no observable. En Ingeniería, dichos hiperparámetros acostumbran a ser conocidos ya que se tiene un buen conocimiento de las características de los sistemas físicos estudiados, pero en Economía se requiere un esfuerzo adicional para obtener estimaciones de los mismos, dado que generalmente son desconocidos, siendo ésta la principal crítica a la utilización de estos modelos en este ámbito.

2.2.2. El filtro de Kalman

El filtro de Kalman es un algoritmo recursivo que permite obtener las estimaciones óptimas de las variables de estado en el instante t , utilizando la información disponible en el instante $t-1$ y actualizar estas estimaciones a medida que se dispone de información adicional. Para aplicar el filtro de Kalman, el sistema considerado debe estar expresado en forma de modelo *state-space*. De hecho, en la mayoría de las aplicaciones económicas el filtro de Kalman no es más que una herramienta para obtener estimaciones de las variables latentes utilizando la información de las variables observables relacionadas.

El filtro de Kalman, propuesto inicialmente por Kalman (1960) y Kalman y Bucy (1961), está formado por dos conjuntos de ecuaciones que se aplican secuencialmente. La primera etapa del procedimiento de estimación consiste en obtener el predictor óptimo de la siguiente observación de la variable de estado (instante t) utilizando toda la información disponible (hasta $t-1$). La segunda etapa actualiza esta predicción incorporando la información adicional disponible en el instante t .

La primera etapa, en el caso del modelo *state-space* general presentado en las ecuaciones (3) a (7), puede resumirse de la siguiente manera. Si se denota por a_{t-1} el estimador óptimo de \mathbf{a}_{t-1} basado en las observaciones disponibles hasta Y_{t-1} incluida, la matriz, de orden $m \times m$, de varianzas y covarianzas del error de estimación P_{t-1} asociado al estimador anterior viene dada por:

$$P_{t-1} = E[(\mathbf{a}_{t-1} - a_{t-1}) \cdot (\mathbf{a}_{t-1} - a_{t-1})'], \quad (8)$$

de manera que, una vez a_{t-1} y P_{t-1} son conocidos, el estimador óptimo de \mathbf{a}_t restringido a estos valores se define como:

$$a_{t/t-1} = \mathbf{a}_t / a_{t-1} = E(\mathbf{a}_t / a_{t-1}) = E(\mathbf{T}_t \cdot a_{t-1} + \mathbf{c}_t + \mathbf{R}_t \cdot \mathbf{q}) = \mathbf{T}_t \cdot a_{t-1} + \mathbf{c}_t, \quad (9)$$

con una matriz de varianzas y covarianzas asociada al error de estimación igual a:

$$P_{t/t-1} = E[(\mathbf{a}_t - a_{t/t-1}) \cdot (\mathbf{a}_t - a_{t/t-1})'] = \mathbf{T}_t \cdot P_{t-1} \cdot \mathbf{T}_t' + \mathbf{R}_t \cdot \mathbf{Q}_t \cdot \mathbf{R}_t'. \quad (10)$$

Las ecuaciones (9) y (10) resumen la primera etapa del procedimiento de estimación y se conocen como ecuaciones de predicción.

La segunda etapa comienza una vez está disponible la observación Y_t y, por tanto, el estimador de \mathbf{a}_t , $a_{t/t}$, puede ser actualizado. Las ecuaciones que integran esta segunda etapa se conocen como ecuaciones de actualización y vienen dadas por las siguientes expresiones:

$$a_t = a_{t/t} = a_{t/t-1} + P_{t/t-1} \cdot \mathbf{Z}_t' \cdot \mathbf{F}_t^{-1} \cdot (Y_t - \mathbf{Z}_t \cdot a_{t/t-1} - d_t); \quad (11)$$

$$P_t = P_{t/t-1} - P_{t/t-1} \cdot \mathbf{Z}_t' \cdot \mathbf{F}_t^{-1} \cdot \mathbf{Z}_t \cdot P_{t/t-1}; \quad (12)$$

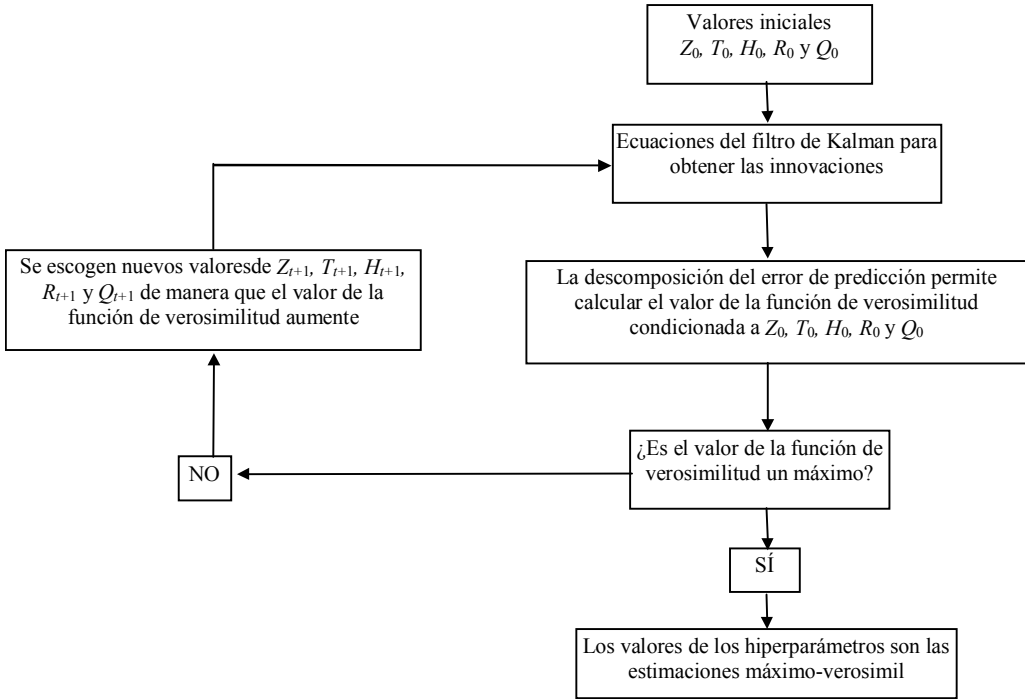
$$\mathbf{F}_t = \mathbf{Z}_t \cdot P_{t/t-1} \cdot \mathbf{Z}_t' + \mathbf{H}_t. \quad (13)$$

Sin embargo, y tal y como se ha comentado anteriormente, es importante destacar que el filtro de Kalman sólo puede ser aplicado si los valores iniciales del vector de estado a_0 (o equivalentemente $a_{1/0}$) y la matriz de varianzas y covarianzas del error de estimación asociado a los mismos P_0 (o $P_{1/0}$) y los hiperparámetros son conocidos. Si no lo son, es necesario proceder a su estimación antes de aplicar las ecuaciones del filtro de Kalman.

2.2.3. Estimación máximo-verosímil de los hiperparámetros

El procedimiento a seguir para estimar los hiperparámetros del modelo por máxima verosimilitud puede resumirse en las cuatro etapas siguientes (véase el gráfico 1):

Gráfico 1. Estimación máximo-verosímil de los hiperparámetros de un modelo *state-space*



Fuente: Adaptado a partir de Cuthbertson *et al.* (1992, p. 214).

a) la primera etapa implica analizar si el modelo está identificado y determinar cuáles son los hiperparámetros desconocidos: Z_t, T_t, H_t, R_t y/o Q_t . A continuación, se fijan unos valores iniciales *a priori* para cada uno de ellos;

b) utilizando estos valores iniciales, es posible aplicar las ecuaciones del filtro de Kalman para obtener los valores de las innovaciones v_t (los errores de predicción para un período hacia delante);

c) a partir de la descomposición del error de predicción (Harvey, 1981 y 1984), es posible obtener el valor numérico de la función de verosimilitud a partir de los valores de las innovaciones y utilizar un procedimiento de optimización numérico para determinar si el valor de dicha función es un máximo o no; y,

d) en caso de que el valor sea un máximo, el procedimiento de estimación de los hiperparámetros habrá finalizado. Si el valor no es un máximo, a partir del procedimiento de optimización utilizado será posible escoger unos nuevos valores iniciales para los hiperparámetros que incrementen el valor de la función de verosimilitud reiterándose el proceso desde la segunda etapa.

Ahora bien, una de las principales críticas a la aplicación de los modelos *state-space* en Economía está relacionada con la inestabilidad asociada a dicho procedimiento de estimación. En este sentido, Hackl y Westlund (1996) muestran que las estimaciones obtenidas a partir de la aplicación del filtro de Kalman son muy sensibles a la especificación del modelo *state-space* y al modo en que se determinan los valores de los hiperparámetros. Su conclusión es que es necesario tener un cuidado sumo cuando se obtienen estimaciones a partir de la aplicación del filtro de Kalman. Una protección razonable en este sentido, consistiría en especificar el modelo lo más sencillo posible y escoger algunos de los valores de los hiperparámetros *a priori* en vez de estimarlos conjuntamente con el vector de estado.

2.2.4. La inicialización del filtro de Kalman

Para resolver el problema de la inicialización del filtro de Kalman existen básicamente dos procedimientos en función de si el vector de estado es o no es estacionario (Reiter, 1995; Snyder y Saligari, 1996).

Un modelo *state-space* es estacionario si los valores propios de la matriz T_t en la ecuación (3) se sitúan dentro del círculo unidad o bien hay suficientes observaciones disponibles para poder afirmar que el sistema considerado ha alcanzado la estacionariedad. En esta situación, es posible aproximar los valores iniciales del vector de estado a partir de la media incondicional del propio proceso. Siguiendo a Gardner *et al.* (1980), entre otros, estos valores pueden obtenerse utilizando las primeras m observaciones disponibles para estimar la ecuación (4) por MCO y a continuación inicializar el filtro de Kalman para el instante $m+1$ con dichos valores. El principal inconveniente de este método es que cuando se dispone de un número reducido de observaciones, los grados de libertad serán muy reducidos. Otra alternativa consiste en considerar los valores iniciales como hiperparámetros desconocidos y estimarlos a través de los procedimientos descritos en el apartado anterior (Rosenberg, 1973).

En cambio, cuando el modelo no es estacionario, las condiciones iniciales no están bien definidas y, por tanto, las soluciones señaladas en el párrafo anterior no deberían aplicarse. En este supuesto, en la literatura existen diferentes propuestas para inicializar el filtro de Kalman. Harvey y Phillips (1979) proponen inicializar el filtro suponiendo que la varianza del error de estimación de los valores iniciales es

muy elevada pero finita. De este modo, a través de las ecuaciones de predicción y actualización, existiría un proceso de convergencia que llevaría a disminuir dicho error, de manera que se podrían obtener las estimaciones deseadas del vector de estado. Sin embargo, la utilización de este método podría llevar a obtener estimaciones muy poco estables. Otra solución propuesta por Anderson y Moore (1979) y Kitagawa y Gersch (1984) consiste en utilizar un algoritmo alternativo conocido como filtro de información que consiste en calcular la inversa de la matriz del error de estimación (conocida como matriz de información) en vez de intentar obtener una aproximación a través de un mayor número de iteraciones. Sin embargo, esta propuesta no es aplicable en el caso en que la matriz de transición sea singular o en el que haya un conjunto de valores iniciales que sean conocidos (en cuyo caso, la varianza del error de estimación sería cero).

Para solucionar los inconvenientes derivados de las propuestas anteriores, diferentes autores proponen tratar los valores iniciales del vector de estado como difusos introduciendo ecuaciones complementarias a las del filtro de Kalman y poder estimar así de manera óptima el vector de estado. En este sentido, se han propuesto diferentes maneras de introducir estas ecuaciones (entre otros: Ansley y Kohn, 1989; de Jong, 1991; Gomez y Maravall, 1994; Kohn y Ansley, 1986). El principal inconveniente de esta alternativa radica en que supone un coste computacional importante y su aplicación con un número reducido de observaciones no es adecuada.

De este modo, y a pesar de que ninguna de las propuestas realizadas hasta el momento consigue resolver completamente el problema de la inicialización del filtro de Kalman, las aproximaciones existentes permiten abordar dicho problema de manera satisfactoria.

2.2.5. El modelo de coeficientes sistemáticamente variables en forma de modelo state-space

En el caso particular del modelo de coeficientes sistemáticamente variables más sencillo posible, la ecuación de transición vendría dada por la siguiente expresión:

$$\mathbf{b}_t = \mathbf{b}_{t-1} + \mathbf{h}_t, \quad (14)$$

mientras que la ecuación de medida sería:

$$\mathbf{Y}_t = \mathbf{X}_t \cdot \mathbf{b}_t + \mathbf{e}_t. \quad (15)$$

En este modelo, los únicos hiperparámetros a estimar por máxima verosimilitud antes de poder aplicar las ecuaciones del filtro de Kalman para estimar los valores de los coeficientes variables son las varianzas de los términos de perturbación \mathbf{h}_t y \mathbf{e}_t , ya que los valores del resto de parámetros del modelo *state-space* general vienen dados por la propia especificación del modelo. En lo que se refiere a los valores iniciales, sería posible aplicar cualquiera de las diferentes propuestas existentes en la literatura y resumidas en el subapartado anterior.

2.3. Combinación de relaciones a largo plazo con relaciones a corto plazo variables a lo largo del tiempo

2.3.1. Especificación

Granger y Newbold (1974) fueron los primeros que señalaron con claridad la importancia de los problemas que podría provocar sobre la estimación la existencia de relaciones espúreas entre las variables consideradas (correlaciones engañosamente elevadas). La solución más obvia, propuesta también por ellos, consiste en diferenciar las series hasta lograr su estacionariedad y a continuación trabajar con esas series diferenciadas. Sin embargo, dicha solución no es del todo satisfactoria: Davidson *et al.* (1978) y Hendry y Mizon (1978) hicieron notar que al expresar el modelo en diferencias, no era posible inferir su solución a largo plazo a partir del modelo estimado. La teoría de la cointegración pretende superar tanto el inconveniente de no aceptar como verdaderas relaciones espúreas como el de considerar tanto las relaciones a largo plazo existentes entre las variables como las relaciones a corto. Por este motivo, la teoría de la cointegración ofrece un marco idóneo para la especificación de modelos econométricos donde ambos efectos adversos puedan ser relevantes.

El concepto de series cointegradas fue desarrollado por Engle y Granger (1987). Dicho concepto está basado en la idea que en el largo plazo dos (o más) variables presentan la misma tendencia estocástica y en consecuencia aunque no sean estacionarias, existe una (o más) combinación lineal entre ellas en la que se cancelan dichas tendencias, siendo, por tanto, estacionaria. Se dice que las variables que componen un vector X_t están cointegradas si todas ellas son integradas de orden d y existe un vector \mathbf{a} distinto de cero tal que cumple que:

$$Z_t = \mathbf{a}' \cdot X_t, \quad (16)$$

donde Z_t es una variable integrada de orden $(d-b)$ siendo b mayor que cero. La existencia del vector \mathbf{a} , indicaría la existencia de una relación no espúrea a largo plazo (véase Suriñach *et al.*, 1995).

La importancia de determinar si un conjunto de variables están cointegradas radica en el hecho que si existen, hay una relación de equilibrio a largo plazo entre ellas, lo cual significa que las desviaciones respecto del equilibrio son transitorias (se dan sólo en el corto plazo). En definitiva se trata de "encontrar" relaciones estacionarias entre variables que no lo son. De este modo, tal y como demuestra Granger (1981), cuando un conjunto de variables están cointegradas es posible expresar dicha relación en forma de Mecanismo de Corrección del Error (MCE) (Sargan, 1964) de manera que se recogerían de manera explícita las relaciones a largo plazo y las relaciones a corto. La expresión en forma de MCE de una relación de cointegración entre dos variables X_t e Y_t , integradas de orden uno, es la siguiente:

$$\Delta Y_t = a + b \cdot \Delta X_t + c \cdot [Y_{t-1} - a_0 - b_0 \cdot X_{t-1}] + e_t, \quad (17)$$

donde Δ representa el operador diferencia y e_t denota un término de perturbación con esperanza cero y varianza conocida. Dicha expresión resulta muy atractiva desde el punto de vista de la interpretación económica de las relaciones entre las variables consideradas. El término $[Y_{t-1} - a_0 - b_0 \cdot X_{t-1}]$ recoge la desviación respecto al equilibrio a largo plazo en el instante $t-1$ mientras que el resto de términos recogen las relaciones a corto plazo. De este modo, se permite que existan situaciones de desequilibrio a corto plazo pero siempre se tiende a volver a una situación de equilibrio a largo plazo.

En este sentido, diversos estudios muestran que la principal fuente de inestabilidad estructural en las relaciones económicas son debidas a fluctuaciones en el corto plazo, mientras que las relaciones a largo plazo son más estables por lo que, en principio, la utilización de MCE resultaría especialmente útil. Sin embargo, existe una segunda fuente de inestabilidad temporal que no quedaría suficientemente bien recogida a través de los MCE. Se trata del supuesto de estabilidad temporal del coeficiente asociado a la relación a corto plazo entre las variables, es decir, del coeficiente b de (17). Tal y como se ha expuesto anteriormente, si se permite la variabilidad de los coeficientes que recogen las relaciones a corto plazo, es posible mejorar considerablemente los resultados de la estimación y predicción del modelo.

De este modo, combinando la estrategia de modelización propuesta por la teoría de la cointegración con la posibilidad de que los coeficientes que recogen las relaciones a corto plazo (las más inestables) entre las variables consideradas puedan variar de acuerdo con la exposición realizada, las relaciones existentes entre una variable endógena Y y una variable exógena X cointegradas vendrían dadas por la siguientes expresiones:

$$\Delta Y_t = a_t + b_t \cdot \Delta X_t + c_t \cdot [Y_{t-1} - a_0 - b_0 \cdot X_{t-1}] + e_t, \quad (18)$$

$$\mathbf{a}_t = \mathbf{a}_{t-1} + \mathbf{h}_{1t}, \quad (19)$$

$$\mathbf{b}_t = \mathbf{b}_{t-1} + \mathbf{h}_{2t}, \text{ y}, \quad (20)$$

$$\mathbf{c}_t = \mathbf{c}_{t-1} + \mathbf{h}_{3t}. \quad (21)$$

De acuerdo con las características de dicha especificación, la estimación del sistema de ecuaciones formado por las ecuaciones (18) a (21) debe considerar necesariamente aspectos propios de la teoría de la cointegración y de los modelos de coeficientes variables en el tiempo.

2.3.2. Estimación

En la literatura se han propuesto diferentes métodos para estimar relaciones de cointegración expresadas en términos de MCE como, por ejemplo, el procedimiento bietápico de Engle y Granger, la estimación directa por mínimos cuadrados no lineales, la estimación directa por mínimos cuadrados ordinarios dinámicos y la estimación máximo-verosímil de Johansen. Sin embargo, y tal y como se expone en el próximo apartado, el reducido número de observaciones disponibles hace más aconsejable utilizar el primero de ellos.

El método propuesto por Engle y Granger (1987, pp. 260-264) consta de dos etapas. La primera consiste en estimar por MCO la relación a largo plazo entre las variables, es decir:

$$\mathbf{Y}_t = \mathbf{a} + \mathbf{b} \cdot \mathbf{X}_t + \mathbf{Z}_t, \quad (22)$$

donde \mathbf{Z}_t denota la serie de residuos de dicha regresión y representa las desviaciones existentes en cada instante del tiempo respecto a la situación de equilibrio a largo plazo, es decir, el término $[\mathbf{Y}_t - \mathbf{a}_0 - \mathbf{b}_0 \cdot \mathbf{X}_t]$ que desfasado un período sería el término del MCE que recoge las relaciones a largo plazo. De este modo, la segunda etapa consiste en introducir la variable \mathbf{Z}_{t-1} como explicativa en una regresión donde se consideran las relaciones a corto plazo y estimar por MCO la ecuación:

$$\mathbf{DY}_t = \mathbf{a} + \mathbf{b} \cdot \mathbf{DX}_t + \mathbf{c} \cdot \mathbf{Z}_{t-1} + \mathbf{e}_t. \quad (23)$$

Teniendo en cuenta las características de la especificación formada por las ecuaciones (18) a (21), en este trabajo únicamente es necesario modificar la segunda etapa del procedimiento propuesto por Engle y Granger (1987). En concreto, una

vez obtenidos los residuos que recogen las desviaciones a largo plazo, se introducen retardados un período como variables explicativas en la expresión del MCE pero se permite que los coeficientes de la regresión varíen en el tiempo. De este modo, la ecuación que se propone estimar en la segunda etapa es la siguiente:

$$DY_t = a_t + b_t \cdot DX_t + c_t \cdot Z_{t-1} + e_t, \quad (24)$$

y, teniendo en cuenta que el comportamiento de los coeficientes a_t , b_t y c_t vendrían recogidos por las ecuaciones (19) a (21), las ecuaciones (24), (19), (20) y (21) forman un modelo *state-space* (la ecuación (24) es la ecuación de medida mientras que las ecuaciones (19) a (21) forman la ecuación de estado) a partir del cual es posible aplicar el filtro de Kalman y obtener estimaciones de los coeficientes deseados.

3. Análisis de la capacidad predictiva de la metodología propuesta a partir de su aplicación al bloque de producción de la economía catalana

El objetivo de este apartado es especificar y estimar un modelo para el bloque de producción subsectorial de la economía catalana con la finalidad de comparar la capacidad predictiva de los modelos de coeficientes variables como el presentado en el apartado anterior con los modelos tradicionales de coeficientes fijos.

Las fuentes utilizadas para la especificación y para la estimación del modelo así como los datos utilizados se presentan en los anexos 1 y 2. Las principales características del modelo que se presenta a continuación son las siguientes:

a) la variable endógena modelizada es el VAB a precios de mercado y a pesetas constantes de 1986 para los diferentes sectores de actividad;

b) el nivel de desagregación sectorial considerado es de cuatro ramas de actividad. En consecuencia, el modelo consiste en un bloque de ecuaciones simultáneas formado por cuatro ecuaciones de comportamiento y una identidad contable;

c) el objetivo perseguido con el modelo es predecir la evolución del VAB subsectorial de la economía catalana a corto y medio plazo. Por ello se ha adoptado, en parte condicionados por la información estadística disponible, un enfoque *top-down* en el cual la causalidad es unidireccional, de la nación a la región;

d) se trata de un modelo mixto en el sentido que entre las variables explicativas, en las ecuaciones de comportamiento, se consideran indicadores;

e) el área geográfica considerada se refiere únicamente a la Comunidad Autónoma de Cataluña, sin tener en consideración las relaciones que se dan con otras regiones. Así pues, se trata de un modelo regional simple;

f) es un modelo anual (la información estadística utilizada es de periodicidad anual); y,

g) la metodología econométrica utilizada en la especificación de las ecuaciones es, como se ha señalado anteriormente, la teoría de la cointegración y los modelos *state-space*.

El resto del apartado se estructura como sigue: en primer lugar, se presenta la especificación teórica del modelo econométrico considerado; a continuación, se procede a la validación de dicha especificación a partir de la teoría de la cointegración; y, por último, se presentan los resultados de estimar el modelo bajo los supuestos de coeficientes fijos y de coeficientes variables comparándose la capacidad predictiva de ambos.

3.1. Especificación del modelo econométrico

En la especificación de las ecuaciones de comportamiento se ha seguido la teoría clásica de base económica^{4,5} que postula que los sectores productivos de una región pueden clasificarse en dos grandes categorías en función de la relación entre la actividad económica regional y la nacional. Dichos sectores son conocidos como básicos (o preponderantes) y subsidiarios. En cualquier caso, es importante señalar que se ha considerado esta teoría únicamente como punto de partida para determinar las variables explicativas de las ecuaciones de comportamiento con el fin de disponer de materia prima a partir de la cual estudiar si la aproximación basada en los modelos de coeficientes variables a lo largo del tiempo permite mejorar las predicciones obtenidas con los modelos econométricos regionales "clásicos". Así pues, tal y como ya se ha señalado anteriormente (véase nota número 3), el modelo que aquí se presenta no es el modelo correspondiente al bloque de producción de la economía catalana vigente actualmente, sino que sólo debe considerarse como un mero instrumento para alcanzar el objetivo perseguido en el presente trabajo: estudiar la potencialidad del instrumental relacionado con la metodología *state-space* y el filtro de Kalman en el ámbito de la modelización econométrica regional.

Las principales hipótesis en que se sustentan los modelos especificados a partir de la teoría de base económica son, de acuerdo con Serrano (1995, p. 434) las tres siguientes:

a) los sectores preponderantes son los determinantes del desarrollo económico de la región y su evolución condiciona la del resto de sectores de la economía regional;

4. Véase, por ejemplo, Richardson (1985) o Nijkamp *et al.* (1986).

5. Para un detalle sobre las principales ventajas e inconvenientes de esta teoría véase, por ejemplo, Richardson (1985, pp. 607-661).

b) la relación entre los sectores básicos y subsidiarios de la región se mantiene estable a lo largo de todo el período considerado; y,

c) la producción depende completamente de la demanda existente.

Por tanto, uno de los inconvenientes que se presenta a la hora de especificar un modelo de estas características consiste en identificar los sectores preponderantes y los sectores no básicos. En la literatura se pueden encontrar diversas propuestas para afrontar dicho problema. Entre ellas cabe destacar las siguientes:

a) Lesage y Reed (1989, p. 617) proponen determinar dichos sectores a partir de una encuesta a los establecimientos de la región. Sin embargo, el elevado coste de esta propuesta dificulta en gran medida poderla llevar a la práctica;

b) adoptar una solución *ad-hoc* consistente en considerar como sector básico el sector manufacturero (Isserman, 1980); y,

c) determinar dichos sectores a partir de un análisis descriptivo de la actividad económica regional, basándose, por ejemplo, en el coeficiente de localización regional.

En este trabajo, se ha optado por la tercera de las estrategias señaladas. El coeficiente de localización regional, L_i , como es sabido, se define de la forma:

$$L_i = \frac{Lreg_i / Lreg}{Lnac_i / Lnac} \quad (25)$$

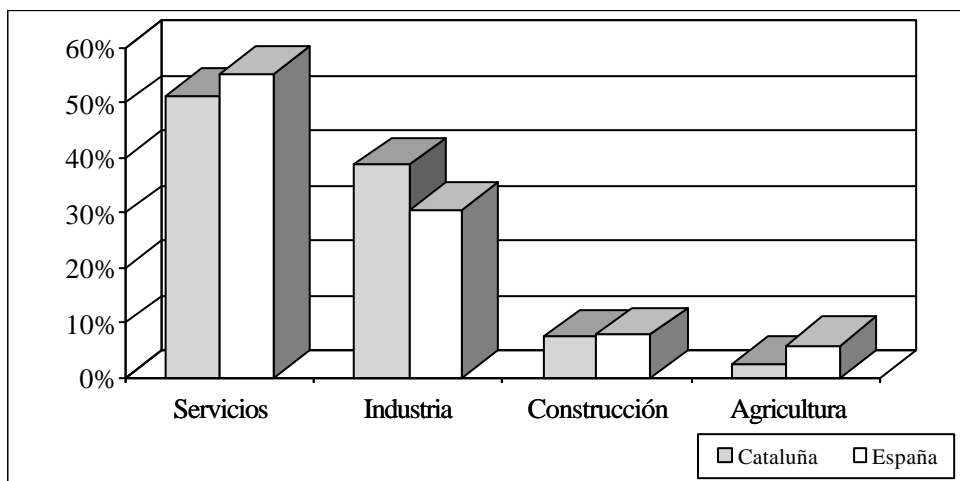
donde $Lreg_i$ representa el número de ocupados en la actividad i en la región, $Lreg$ el total de ocupados en la región, $Lnac_i$ el número de ocupados en la actividad i en la nación y $Lnac$ el total de ocupados en la nación. En consecuencia, este coeficiente puede interpretarse como una medida del grado de especialización regional y, por tanto, es un instrumento a partir del cual es posible clasificar los sectores productivos: si el coeficiente toma un valor superior a uno para un sector i , la región está más especializada que la nación en dicho sector, por lo que puede considerarse que se trata de un sector preponderante en la región, mientras que si por el contrario para un determinado sector el coeficiente es inferior a la unidad se puede concluir que se trata de un sector subsidiario en la región.

En el caso de la comunidad catalana, la estructura productiva media para el período 1972-94, muestra que el sector terciario es el más importante (tiene un peso del 51.15% sobre el total del VAB generado por la economía catalana). Asimismo, el sector secundario presenta un peso importante (38.65%), mientras que el sector de la construcción y el primario tienen un peso bastante reducido (7.68% y 2.51% respectivamente). Por otro lado, como se puede observar en el gráfico 2, la estructura productiva de Cataluña no difiere sustancialmente de la del conjunto del Estado.

Los valores del coeficiente de localización regional obtenidos para los cuatro grandes sectores de actividad de la economía catalana muestran la importancia del sec-

tor secundario catalán en el del conjunto del Estado. De hecho, este sector es el único para el que el coeficiente de localización regional toma un valor superior a la unidad. En concreto, los valores obtenidos para dicho coeficiente han sido los siguientes: agricultura, 0.41; industria, 1.26; construcción, 0.95; y servicios, 0.92.

Gráfico 2. Comparación estructuras productivas Cataluña-España. Período 1972-94



En función de los resultados anteriores, el VAB del sector industrial de la economía catalana ($VINCT86$) dependería únicamente de variables de fuera de la región. Esta demanda externa se ha aproximado por el VAB total de la economía española ($VATOT86$). Asimismo, también se ha considerado la productividad aparente de la industria española ($PROIND$) como indicador de la competitividad del sector secundario a nivel nacional y como *proxy* de la probabilidad que la demanda de productos industriales a nivel estatal se desplace de Cataluña hacia otras regiones. Así pues, el signo esperado *a priori* para estas variables explicativas debería ser positivo para la primera y negativo para la segunda:

$$VINCT86_t = f(VATOT86_t, PROIND_t). \quad (26)$$

Para el resto de sectores productivos la demanda provendría del interior de la propia región. Así, la producción del sector agrícola estaría orientada a satisfacer el mercado catalán, por lo que el principal determinante del VAB del sector primario catalán ($VACT86$) sería la demanda interna ($VTOTCT86$) cuyo signo, de acuerdo

con los postulados de teoría económica, es de esperar que sea negativo. En cualquier caso, el sector agrícola dependería además de otra serie de factores que es necesario tener en cuenta: en primer lugar hay un fuerte componente de intervención pública que determina la producción de este sector y, en segundo lugar, las rigideces de oferta debidas a las condiciones climáticas que pueden afectar de manera importante la producción agrícola. Estos determinantes han sido considerados en la especificación de la ecuación de este sector aproximados respectivamente por el nivel de precios de los productos agrícolas (*PPRIAG*) y por el nivel de precipitaciones anual en la comunidad catalana (*LLUVIA*):

$$VACT\ 86_t = f(VTOTCT\ 86_t, PPRIAG_t, LLUVIA_t). \quad (27)$$

El VAB del sector de la construcción (*VBCT86*) también vendría explicado por variables relativas a la comunidad catalana. En concreto, la demanda de este sector se explicaría por el VAB total de la economía catalana (*VTOTCT86*), el número de viviendas finalizadas (*VIV*) como *proxy* de la inversión residencial y la diferencia entre las nuevas inversiones y las registradas (*INV*):

$$VBCT\ 86_t = f(VTOTCT\ 86_t, VIV_t, INV_t). \quad (28)$$

Por su parte el VAB del sector terciario (*VSERCT86*) estaría determinado por la actividad turística desarrollada en la región, aproximada por el número de pernoctaciones en Cataluña (*PERN*) y por el VAB total de la economía catalana (*VTOTCT86*):

$$VSERCT\ 86_t = f(PERN_t, VTOTCT\ 86_t). \quad (29)$$

Por último, el VAB total de la economía catalana se obtiene como suma del VAB generado por cada uno de los cuatro sectores productivos:

$$VTOTCT\ 86_t = VACT\ 86_t + VINCT\ 86_t + VBCT\ 86_t + VSERCT\ 86_t. \quad (30)$$

Así, el modelo propuesto puede sintetizarse en las siguientes cinco ecuaciones (cuatro de comportamiento y una identidad), donde las relaciones entre las variables se han considerado lineales y las variables vienen expresadas en logaritmos:

$$\ln VACT\ 86_t = b_{11} + b_{12} \cdot \ln VTOTCT\ 86_t + b_{13} \cdot \ln PPRIAG_t + b_{14} \cdot \ln LLUVIA_t + u_{1t}; \quad (31)$$

$$\ln VINCT_{86_t} = \mathbf{b}_{21} + \mathbf{b}_{22} \cdot \ln VATOT_{86_t} + \mathbf{b}_{23} \cdot \ln PROIND_t + \mathbf{u}_{2t}; \quad (32)$$

$$\ln VBCT_{86_t} = \mathbf{b}_{31} + \mathbf{b}_{32} \cdot \ln VTOTCT_{86_t} + \mathbf{b}_{33} \cdot \ln VIV_t + \mathbf{b}_{34} \cdot \ln INV_t + \mathbf{u}_{3t}; \quad (33)$$

$$\ln VSERCT_{86_t} = \mathbf{b}_{41} + \mathbf{b}_{42} \cdot \ln PERN_t + \mathbf{b}_{43} \cdot \ln VTOTCT_{86_t} + \mathbf{u}_{4t}; \quad (34)$$

$$\ln VTOTCT_{86_t} = \ln VACT_{86_t} + \ln VINCT_{86_t} + \ln VBCT_{86_t} + \ln VSERCT_{86_t}. \quad (35)$$

3.2. Validación de la especificación propuesta a partir de la teoría de la cointegración

Tal y como se ha señalado anteriormente, la técnica econométrica que se ha utilizado en la especificación y estimación del modelo es la teoría de la cointegración con la finalidad de recoger adecuadamente las relaciones a largo plazo entre las variables consideradas. Así, una vez especificado el modelo en términos teóricos, el primer paso consiste en estudiar el orden de integrabilidad de las variables consideradas y, a continuación, analizar la existencia de relaciones de cointegración con el fin de validar la especificación realizada en el subapartado anterior⁶.

3.2.1. Análisis de integrabilidad

La presencia de tendencias estocásticas en los procesos generadores de variables de serie temporal pueden provocar, si no son tratadas adecuadamente, efectos negativos en la modelización como, por ejemplo, aceptar como ciertas relaciones espúreas. Para evitar caer en este tipo de especificaciones se han aplicado los contrastes propuestos por Dickey y Fuller -DF- (1979) y la corrección paramétrica de éste, denominada contraste Dickey y Fuller ampliado -DFA- (1981)⁷, con el fin de determinar el orden de integrabilidad de cada una de las variables del modelo.

Como es sabido, el test DF consiste en contrastar la hipótesis nula $H_0: \mathbf{g}=0$ frente a la alternativa $H_A: \mathbf{g}<0$ en los modelos:

$$\mathbf{Dx}_t = \mathbf{m} + \mathbf{b} \cdot t + \mathbf{g} \cdot x_{t-1} + \mathbf{e}_t;$$

$$\mathbf{Dx}_t = \mathbf{m} + \mathbf{g} \cdot x_{t-1} + \mathbf{e}_t; \text{ y,}$$

$$\mathbf{Dx}_t = \mathbf{g} \cdot x_{t-1} + \mathbf{e}_t.$$

6. Esta estrategia ha sido utilizada en diversos trabajos. Véase, por ejemplo, Brown *et al.* (1992, pp. 623-627), Lesage (1990, pp. 310-315), Mur y Trivez (1996, pp. 464-466).

7. Entre otros contrastes para determinar el orden de integrabilidad de una determinada variable también cabe señalar el propuesto por Phillips y Perron (1988) que no es más que una corrección no paramétrica del contraste de Dickey y Fuller. Sin embargo, presenta el inconveniente que no es del todo satisfactorio cuando la muestra es pequeña. Por esta razón en este trabajo hemos optado por no aplicarlo.

Si no es posible rechazar la hipótesis nula la variable presenta tendencias estocásticas. Debe tenerse en cuenta que bajo la hipótesis nula los tres modelos anteriores son caminos aleatorios y, en consecuencia, son no estacionarios por lo que los estadísticos del contraste no se distribuyen asintóticamente según una distribución normal, invalidándose por tanto la inferencia estándar. Por esta razón, los valores críticos a utilizar para contrastar la hipótesis nula no pueden ser los proporcionados por la distribución *t*-Student estándar. Los valores críticos apropiados (τ_t , τ_m y τ respectivamente), fueron tabulados por Fuller (1976) y Mackinnon (1991), y los valores críticos para contrastar la significación individual de los componentes deterministas (τ_{bt} , τ_m y τ_{am}) por Dickey y Fuller (1981).

En cualquier caso, para que la inferencia sea válida en los modelos anteriores es necesario que el término de perturbación se comporte según un ruido blanco. Así pues, si los residuos de la estimación presentan algún esquema de autocorrelación se deben incorporar como regresores en el modelo un número suficiente de retardos de la variable endógena con el fin de eliminar (capturar) dicha autocorrelación^{8,9}. Esta corrección paramétrica a la que se somete el test DF es conocida como test DFA.

De acuerdo con Dickey y Pantula (1987) la estrategia más adecuada para determinar el orden de integrabilidad de una variable serie temporal consiste en iniciar la secuencia de contrastes por el grado más elevado posible que pueda presentar dicha serie (*d*) y, en caso de rechazar la hipótesis nula, contrastar si la serie es integrable de orden inmediatamente inferior (*d*-1). En este trabajo, de acuerdo con la mayoría de estudios empíricos en los que se pone de manifiesto que las series macroeconómicas presentan como máximo dos raíces unitarias¹⁰, se ha iniciado el proceso contrastando la hipótesis *d*=2 y, en caso de rechazarla, se ha contrastado si la serie es *I*(1) contra la alternativa que no presenta ninguna tendencia estocástica.

De acuerdo con todo lo anterior, la regresión auxiliar del test DFA es:

$$\mathbf{D}^d x_t = \gamma \cdot \mathbf{D}^{d-1} x_{t-1} + \sum_{j=1}^k \delta_j \cdot \mathbf{D}^d x_{t-j} + \mu + \beta \cdot t + \varepsilon_t, \quad d = 1, 2. \quad (36)$$

8. Para contrastar la significación individual de los parámetros asociados a los retardos de la variable endógena introducidos en el modelo como regresores pueden utilizarse los valores críticos proporcionados por la distribución *t*-Student estándar.

9. Respecto al número de retardos de la variable endógena a incluir como regresores en el modelo, Said y Dickey (1984) proponen que como máximo sea igual a la raíz cúbica del número de observaciones disponibles dado que de lo contrario la potencia del contraste se reduce considerablemente.

10. Por ejemplo, en Molinas *et al.* (1990), en Andrés *et al.* (1990a y 1999b), así como en diferentes modelos econométricos regionales que actualmente están siendo utilizados en nuestro país.

En la tabla 1 se muestran los principales resultados obtenidos al aplicar los contrastes descritos para las variables consideradas en el modelo. El análisis de dichos resultados muestra que la hipótesis nula de una raíz unitaria no puede ser rechazada para ninguna de las variables consideradas.

Tabla 1. Resultados de los tests DF y DFA

Variable	Modelo	t_g	valor crítico (5%)	DW	Nº de observaciones
<i>lnVACT86</i>	C,2	-1.80	-3.01	2.02	21
<i>lnVINCT86</i>	C,1	-0.94	-3.00	1.95	22
<i>lnVBCT86</i>	C,2	-1.70	-3.00	2.06	22
<i>lnVSERCT86</i>	T,0	-1.03	-3.64	2.39	21
<i>LnVATOT86</i>	T,2	-2.67	-3.63	2.08	22
<i>LnPPRIAG</i>	C,2	-2.67	-3.00	1.72	22
<i>LnLLUVIA</i>	T,2	-3.25	-3.63	1.83	22
<i>LnPROIND</i>	T,1	-2.06	-3.64	1.80	21
<i>LnINV</i>	T,1	-3.09	-3.62	1.98	23
<i>LnVIV</i>	T,1	-1.84	-3.62	2.14	23
<i>LnPERN</i>	T,0	-3.30	-3.62	1.89	23

Las letras T y C hacen referencia, respectivamente, a la consideración de tendencia y constante o sólo constante en la regresión auxiliar del test. El número que les acompaña indica el número de retardos de la variable endógena incluidos en la regresión del test.

3.2.2. Análisis de las relaciones de cointegración

Una vez determinado el orden de integrabilidad de las variables, es importante estudiar la posible existencia de relaciones de cointegración entre ellas. En este trabajo, para determinar la existencia de relaciones estacionarias a largo plazo se ha seguido una estrategia de lo particular a lo general, en el sentido que se ha partido de la relación más sencilla posible (aquella en la cual se consideran un menor número de variables)^{11,12} y se ha aplicado el test de cointegración propuesto por Engle y Granger (1987), que consiste en analizar el orden de integrabilidad de los residuos de la estimación por MCO de las ecuaciones (31) a (34) utilizando el contraste DFA.

11. De hecho, esta estrategia es la seguida en numerosos estudios. Véase, por ejemplo, Andrés *et al.* (1990a y b).

12. Cabe señalar que a la hora de interpretar los resultados hay que tener en cuenta que la potencia de los tests de integrabilidad se reduce de manera muy significativa cuando se trabaja, como es el caso, con muestras pequeñas. En concreto, los valores críticos pueden estar ligeramente sesgados al alza (en valor absoluto) con lo cual la probabilidad de no rechazar la hipótesis nula de no cointegración aumenta.

Los resultados obtenidos (recogidos en la tabla 2) muestran que la hipótesis nula de no cointegración es rechazada en todos los casos.

Tabla 2. Resultados del test de cointegración de Engle y Granger (1987)

<i>Sector</i>	<i>Modelo</i>	t_g	<i>valor crítico (5%)</i>	<i>Nº de variables</i>	<i>Nº de observaciones</i>
Agricultura	C,0	-3.22	-4.66	3	22
Manufacturas	C,0	-2.66	-4.17	2	22
Construcción	C,0	-2.51	-4.66	3	22
Servicios	C,0	-3.29	-4.17	2	22

La letra *C* hace referencia a la consideración de una constante en la regresión auxiliar del test. El cero que le acompaña indica el número de retardos de la variable endógena incluidos en la regresión del test.

3.3. Especificación de las relaciones entre las variables consideradas en forma de mecanismo de corrección del error

Tal y como se ha comentado en el apartado anterior, cuando dos variables están cointegradas, es decir, cuando existe una relación a largo plazo entre ellas, el teorema de la representación de Granger permite expresar dichas relaciones en términos dinámicos mediante mecanismos de corrección del error (MCE), lo cual permite considerar conjuntamente la existencia de relaciones a corto (en diferencias) y a largo plazo (en niveles). Así, tal y como se muestra en la tabla 3, es posible reescribir las ecuaciones (31) a (34) en forma de MCE.

A partir de los resultados de la estimación uniecuacional¹³ por Mínimos Cuadrados No Lineales, recogidos en la tabla 4, se puede afirmar que los resultados del ajuste obtenido son satisfactorios. En este sentido cabe señalar que:

13. Se ha optado por estimar cada ecuación por separado por mínimos cuadrados no lineales (método de estimación directo) en lugar de métodos de información limitada (como, por ejemplo, mínimos cuadrados en dos etapas o máxima verosimilitud información limitada) o de información completa (mínimos cuadrados en tres etapas o máxima verosimilitud información completa) a pesar de que garantizan la consistencia (en el caso de los primeros) y además la eficiencia (en el de los segundos) de los estimadores ha estado motivada por las siguientes razones: *a*) el reducido número de observaciones disponibles (en este sentido existe evidencia empírica que muestra que en caso de trabajar con series cortas los métodos de estimación conjunta no presentan ventajas -mejores resultados- respecto a los métodos directos); *b*) los métodos de estimación conjunta son más sensibles que los directos ante problemas de multicolinealidad y de errores de medida en las variables; y, *c*) cuando se estima el modelo de forma conjunta no es posible evitar que los errores que puedan haberse cometido en una ecuación del modelo se trasladen (y por tanto afecten) al resto de ecuaciones del modelo.

Tabla 3. Ecuaciones (31) a (34) términos de MCE

$$\Delta \ln VACT86_t = \gamma_{11} \Delta \ln VTOTCT86_t + \gamma_{12} \Delta \ln PPRIAG_t + \gamma_{13} \Delta \ln LLUVIA_t + \gamma_{14} [\ln VACT86_{t-1} - \beta_{11} - \beta_{12} \ln VTOTCT86_{t-1} - \beta_{13} \ln PPRIAG_{t-1} - \beta_{14} \ln LLUVIA_{t-1}]$$

$$\Delta \ln VINCT86_t = \gamma_{21} \Delta \ln VATOT86_t + \gamma_{22} \Delta \ln PROIND_t + \gamma_{23} [\ln VINCT86_{t-1} - \beta_{21} - \beta_{22} \ln VATOT86_{t-1} - \beta_{23} \ln PROIND_{t-1}]$$

$$\Delta \ln VBCT86_t = \gamma_{31} \Delta \ln VTOTCT86_t + \gamma_{32} \Delta \ln VIV_t + \gamma_{33} \Delta \ln INV_t + \gamma_{34} [\ln VBCT86_{t-1} - \beta_{31} - \beta_{32} \ln VTOTCT86_{t-1} - \beta_{33} \ln VIV_{t-1} - \beta_{34} \ln INV_{t-1}]$$

$$\Delta \ln VSERCT86_t = \gamma_{41} \Delta \ln PERN_t + \gamma_{42} \Delta \ln VTOTCT86_t + \gamma_{43} [\ln VSERCT86_{t-1} - \beta_{41} - \beta_{42} \ln PERN_{t-1} - \beta_{43} \ln VTOTCT86_{t-1}]$$

Tabla 4. Resultados de la estimación uniecuacional del modelo con coeficientes fijos

$$\Delta \ln VACT86_t = -0.42 \cdot \Delta \ln VTOICT86_t + 0.27 \cdot \Delta \ln PPRIAG_t + 0.02 \cdot \Delta \ln LLUVIA_t - 1.16 \cdot res_{t-1}$$

(-1.60) (2.59) (0.78) (-10.63)

$R^2 = 0.93; DW = 1.97$

$$\Delta \ln VINCT86_t = 1.39 \cdot \Delta \ln VATOT86_t - 0.42 \cdot \Delta \ln PROIND_t - 0.64 \cdot res_{t-1}$$

(6.63) (-2.80) (-2.74)

$R^2 = 0.66; DW = 1.52$

$$\Delta \ln VBCT86_t = 1.06 \cdot \Delta \ln VTOTCT86_t + 0.37 \cdot \Delta \ln VIV_t + 0.03 \cdot \Delta \ln INV_t - 0.38 \cdot res_{t-1}$$

(3.56) (3.64) (5.42) (-2.26)

$R^2 = 0.78; DW = 1.32$

$$\Delta \ln VSERCT86_t = 0.12 \cdot \Delta \ln PERN_t + 0.86 \cdot \Delta \ln VTOICT86_t - 0.63 \cdot res_{t-1}$$

(3.16) (7.93) (-2.78)

$R^2 = 0.82; DW = 1.63$

a) el período utilizado para estimar las ecuaciones del modelo va desde 1973 hasta 1994, en consecuencia, sólo se dispone de veintidós observaciones para la estimación;

b) el R^2 es en todos los casos aceptable;

c) los parámetros estimados son en todos los casos significativos (excepto en el caso de la variable *LLUVIA* en la ecuación del sector agrícola) y las relaciones de comportamiento son adecuadas en el sentido que el contraste de significación global del modelo lleva a rechazar la hipótesis nula en todos los casos; y,

d) los parámetros asociados a las relaciones a largo plazo (los residuos retardados un período) son todos ellos significativos dado que los t -estadísticos superan los valores críticos tabulados por Banerjee *et al.* (1993). Este hecho permite afirmar, de acuerdo con el teorema de la representación de Granger, que las variables están cointegradas.

Teniendo en cuenta, por tanto, que es posible expresar las relaciones consideradas en forma de MCE, la especificación de cada una de las ecuaciones sectoriales en forma de modelo *state-space* permitiendo que los coeficientes que recogen las relaciones a corto plazo varíen a lo largo del tiempo y, teniendo en cuenta que la primera etapa del procedimiento de estimación propuesto ya ha sido llevada a cabo, viene dada por:

a) *Agricultura*: ecuación de transición (37) y ecuación de medida (38):

$$\begin{pmatrix} \mathbf{g}_{11t} \\ \mathbf{g}_{12t} \\ \mathbf{g}_{13t} \\ \mathbf{g}_{14t} \end{pmatrix} = \begin{pmatrix} \mathbf{g}_{11t-1} \\ \mathbf{g}_{12t-1} \\ \mathbf{g}_{13t-1} \\ \mathbf{g}_{14t-1} \end{pmatrix} + \begin{pmatrix} \mathbf{q}_{11t} \\ \mathbf{q}_{12t} \\ \mathbf{q}_{13t} \\ \mathbf{q}_{14t} \end{pmatrix}; \quad (37)$$

$$\mathbf{DlnVACT86}_t = (\mathbf{g}_{11t} \quad \mathbf{g}_{12t} \quad \mathbf{g}_{13t} \quad \mathbf{g}_{14t}) \cdot \begin{pmatrix} \mathbf{DlnVTOTCT86}_t \\ \mathbf{DlnPPRIAG}_t \\ \mathbf{DlnLLUVIA}_t \\ \mathbf{res}_{1t-1} \end{pmatrix} + \mathbf{e}_{1t}. \quad (38)$$

b) *Manufacturas*: ecuación de transición (39) y ecuación de medida (40):

$$\begin{pmatrix} \mathbf{g}_{21t} \\ \mathbf{g}_{22t} \\ \mathbf{g}_{23t} \end{pmatrix} = \begin{pmatrix} \mathbf{g}_{21t-1} \\ \mathbf{g}_{22t-1} \\ \mathbf{g}_{23t-1} \end{pmatrix} + \begin{pmatrix} \mathbf{q}_{21t} \\ \mathbf{q}_{22t} \\ \mathbf{q}_{23t} \end{pmatrix}; \quad (39)$$

$$DlnVINCT86_t = (\mathbf{g}_{21t} \quad \mathbf{g}_{22t} \quad \mathbf{g}_{23t}) \cdot \begin{pmatrix} DlnVATOT86_t \\ DlnPROIND_t \\ res_{2t-1} \end{pmatrix} + \mathbf{e}_{2t} . \quad (40)$$

c) *Construcción*: ecuación de transición (41) y ecuación de medida (42):

$$\begin{pmatrix} \mathbf{g}_{31t} \\ \mathbf{g}_{32t} \\ \mathbf{g}_{33t} \\ \mathbf{g}_{34t} \end{pmatrix} = \begin{pmatrix} \mathbf{g}_{31t-1} \\ \mathbf{g}_{32t-1} \\ \mathbf{g}_{33t-1} \\ \mathbf{g}_{34t-1} \end{pmatrix} + \begin{pmatrix} \mathbf{q}_{31t} \\ \mathbf{q}_{32t} \\ \mathbf{q}_{33t} \\ \mathbf{q}_{34t} \end{pmatrix} ; \quad (41)$$

$$DlnVBCT86_t = (\mathbf{g}_{31t} \quad \mathbf{g}_{32t} \quad \mathbf{g}_{33t} \quad \mathbf{g}_{34t}) \cdot \begin{pmatrix} DlnVTOTCT86_t \\ DlnVIV_t \\ DlnINV_t \\ res_{3t-1} \end{pmatrix} + \mathbf{e}_{3t} . \quad (42)$$

d) *Servicios*: ecuación de transición (43) y ecuación de medida (44):

$$\begin{pmatrix} \mathbf{g}_{41t} \\ \mathbf{g}_{42t} \\ \mathbf{g}_{43t} \end{pmatrix} = \begin{pmatrix} \mathbf{g}_{41t-1} \\ \mathbf{g}_{42t-1} \\ \mathbf{g}_{43t-1} \end{pmatrix} + \begin{pmatrix} \mathbf{q}_{41t} \\ \mathbf{q}_{42t} \\ \mathbf{q}_{43t} \end{pmatrix} ; \quad (43)$$

$$DlnVSERCT86_t = (\mathbf{g}_{41t} \quad \mathbf{g}_{42t} \quad \mathbf{g}_{43t}) \cdot \begin{pmatrix} DlnPERN_t \\ DlnVTOTCT86_t \\ res_{4t-1} \end{pmatrix} + \mathbf{e}_{4t} . \quad (44)$$

Para estimar los valores de los coeficientes de corto plazo del modelo formado por las ecuaciones (37) a (44) mediante el filtro de Kalman, es necesario en primer lugar obtener los valores iniciales del vector de estado (en este caso, los coeficientes) así como los hiperparámetros del modelo (las matrices de varianzas y covarianzas de los términos de perturbación de las ecuaciones de medida y de transición). Los primeros han sido aproximados siguiendo la propuesta de Harvey (1981 y 1989) que, brevemente, consiste en utilizar una distribución impropia como punto de partida. En

concreto, se trata de un proceso recursivo en el que las iteraciones para $t=0$ comienzan con $\alpha_0=0$ y $P_0=\kappa \cdot I$, siendo κ un número positivo suficientemente grande (en este trabajo, como es habitual en la mayor parte de trabajos empíricos realizados en este entorno, se ha aproximado por 10^6).

Por su parte, los hiperparámetros han sido estimados por máxima verosimilitud utilizando la descomposición del error de predicción de la función de verosimilitud propuesta por Harvey (1984) y el procedimiento de optimización numérico de Broyden-Fletcher-Goldfarb-Shanno (BFGS)¹⁴ usando las expresiones numéricas para el gradiente de la función de verosimilitud. Una vez realizada esta etapa previa, la aplicación de las ecuaciones del filtro de Kalman permiten obtener los valores de los coeficientes.

Para analizar la capacidad predictiva del modelo propuesto, se ha utilizado como punto de referencia un modelo de coeficientes fijos donde se han recogido los principales cambios estructurales detectados mediante el contraste de Chow a través de variables ficticias¹⁵. Este tipo de modelos permite, por tanto, recoger parcialmente los efectos de los cambios estructurales sobre la estimación siendo uno de los modelos más utilizados con este fin. Los resultados de la estimación MCO del modelo con variables ficticias son los siguientes:

$$\ln VACT86_t = 24.24 - 0.77 \cdot \ln VTOTCT86_t + 0.07 \cdot \ln PPRIAG_t - 0.09 \cdot \ln LLUVIA_t + 0.13 \cdot D80_t - 0.16 \cdot D8586_t + e_t$$

(11.08) (-5.38) (2.19) (-1.95) (2.43) (-3.15)

$R^2 = 0.79$; $DW = 2.23$

$$\ln VINCT86_t = 3.24 + 1.14 \cdot \ln VATOT86_t - 0.38 \cdot \ln PROIND_t + 0.03 \cdot D7677_t - 0.05 \cdot D8182_t + e_t$$

(4.01) (13.49) (-6.35) (2.22) (-3.41)

$R^2 = 0.97$; $DW = 1.79$

$$\ln VBCT86_t = -3.52 + 0.73 \cdot \ln VTOTCT86_t + 0.44 \cdot \ln VIV_t + 0.03 \cdot \ln INV_t - 0.11 \cdot D7880_t + 0.15 \cdot D9192_t + e_t$$

(-1.82) (6.68) (10.31) (4.24) (-3.09) (3.13)

$R^2 = 0.93$; $DW = 1.66$

$$\ln VSERCT86_t = -2.78 + 0.06 \cdot \ln PERN_t + 1.09 \cdot \ln VTOTCT86_t + 0.05 \cdot D78_t + 0.04 \cdot D9294_t + e_t$$

(-4.95) (2.41) (26.36) (3.87) (2.98)

$R^2 = 0.99$; $DW = 1.98$

14. Sobre dicho procedimiento de optimización numérica véase Davidon (1968), Fletcher (1987) y Fletcher y Powell (1963).

15. Variables que toman el valor 1 para un año concreto (aquel en el que se produce el cambio estructural) y 0 para el resto de los años del período considerado.

3.3. Análisis de la capacidad predictiva del modelo: coeficientes fijos versus coeficientes variables

Dado que el principal objetivo perseguido es comparar la capacidad predictiva de ambos modelos se ha evaluado la capacidad predictiva de cada uno de ellos vía simulación histórica para todo el período considerado y predicción *ex-post* para 1993 y 1994. En el anexo 3 se comparan gráficamente los resultados obtenidos al estimar ambos modelos. Así pues, en este apartado únicamente presentaremos los principales resultados obtenidos al comparar la capacidad predictiva de los diferentes modelos. Los resultados obtenidos en términos del error porcentual absoluto medio (EPAM)¹⁶, tanto para el modelo de coeficientes fijos (con ficticias) como el de coeficientes variables se muestran en la tabla 5.

Teniendo en cuenta que existe el consenso que valores del EPAM inferiores al 3% son indicativos de una buena capacidad predictiva, entre el 3% y el 5% de una capacidad predictiva no demasiado buena y que valores superiores al 5% son indicativos de una mala capacidad predictiva, los resultados obtenidos muestran que la capacidad predictiva de ambos modelos es, en términos generales, aceptable excepto para el caso del sector de la agricultura.

Tabla 5. Análisis de la capacidad predictiva del modelo de coeficientes fijos (con ficticias) y del modelo de coeficientes variables

	<i>Simulación histórica 1972-94</i>		<i>Predicción ex-post 1993-94</i>	
	<i>EPAM Coef. Fijos</i>	<i>EPAM Coef. Var.</i>	<i>EPAM Coef. Fijos</i>	<i>EPAM Coef. Var.</i>
Agricultura	4.87%	4.21%	7.41%	6.90%
Manufacturas	1.60%	1.72%	2.58%	0.71%
Construcción	3.80%	4.98%	4.18%	1.21%
Servicios	1.24%	1.50%	1.40%	0.95%
Total	0.63%	1.05%	0.79%	0.48%

Comparando los resultados obtenidos con ambos modelos, y más allá de si las ecuaciones presentadas son las más adecuadas para explicar las diferentes variables endógenas, se puede concluir que la capacidad predictiva del modelo de coeficientes variables es mejor que la del modelo con coeficientes fijos, especialmente si se tienen en cuenta los problemas asociados al tratamiento de los valores iniciales en el modelo de coeficientes variables que no hace sino aumentar "artificialmente" el valor del EPAM cosa que se pone de manifiesto en el análisis de la simulación histórica.

16. $EPAM = \left[\frac{\sum_{t=1}^T \frac{|Y_t - \hat{Y}_t|}{Y_t}}{T} \right] \cdot 100$.

4. Conclusiones

La modelización econométrica regional en nuestro país ha adquirido una relevancia importante en las últimas décadas dado el nuevo marco político-económico existente desde la creación del Estado de las Autonomías en el que los gobiernos regionales juegan un papel clave en tanto que sus decisiones afectan no sólo a las regiones vecinas sino también al conjunto del Estado como consecuencia de las externalidades existentes.

Ante este marco, en este trabajo se ha presentado una aproximación a la modelización del VAB subsectorial de la economía catalana para un nivel de desagregación sectorial de cuatro ramas de actividad con el objetivo de analizar la bondad de la consideración de coeficientes variables en el tiempo en la predicción con los modelos econométricos regionales. La especificación de las ecuaciones del modelo se ha llevado a cabo tomando como punto de partida la teoría de base económica siguiendo un enfoque *top-down* desde el punto de vista de la oferta. La clasificación efectuada entre sectores preponderantes y subsidiarios se ha validado a través de la teoría de la cointegración. A continuación, las ecuaciones del modelo se han expresado en términos de MCE donde, teniendo en cuenta la finalidad predictiva del modelo, los coeficientes de las relaciones a corto plazo varían a lo largo del tiempo para recoger los efectos de los cambios estructurales detectados. La estimación del modelo (especificado en términos de un modelo *state-space*) se ha realizado mediante el filtro de Kalman.

La evaluación de la capacidad predictiva del modelo ha puesto de manifiesto que las estimaciones obtenidas ofrecen mejores resultados en comparación a los obtenidos con un modelo de coeficientes fijos en el que los cambios estructurales son tratados mediante la incorporación en las ecuaciones de comportamiento de variables ficticias.

En este sentido, el análisis realizado abre una línea de trabajo en torno a la consideración de coeficientes variables en el tiempo en los modelos econométricos regionales como instrumento para mejorar su capacidad predictiva.

Bibliografía

- ANDERSON, B. y MOORE, J. (1979): *Optimal Filtering*, Prentice-Hall, Englewoods Cliffs.
- ANDRÉS, J., ESCRIBANO, A., MOLINAS, C. y TAGUAS, D. (1990a): *La Inversión en España. Econometría con Restricciones de Equilibrio*, Antoni Bosch editor e IEF, Barcelona.
- ANDRÉS, J., MOLINAS, C. y TAGUAS, D. (1990b): "Una Función de Consumo Privado para la Economía Española: Aplicación del Análisis de Cointegración", *Cuadernos Económicos de ICE*, 1990/1, 44, pp. 173-212.

- ANSLEY, C. y KOHN, R. (1989): "Filtering and Smoothing in State-Space Models with Partially Diffuse Initial Conditions", *Journal of Time Series Analysis*, 11, pp. 275-293.
- ARTÍS, M., CLAR, M., del BARRIO, T., GARCÍA, J.R., SANSÓ, A. y SURIÑACH, J. (1997): *Metodologia per a l'Anàlisi de les Previsions Econòmiques en l'Àmbit Regional: Una Aplicació al Cas de Catalunya*, Col.lecció Estudis Econòmics, 3, Direcció General de Programació Econòmica, Departament d'Economia i Finances, Generalitat de Catalunya, Barcelona. También *Documento de Trabajo 96R07*, Grup d'Anàlisi Quantitativa Regional, Universitat de Barcelona.
- BANERJEE, A., DOLADO, J. y MESTRE, R. (1993): "On Some Tests from Cointegration: the Cost of Simplicity", *Banco de España, Documento de Trabajo 9302*.
- BROWN, S.I., COULSON, N.E. y ENGLE, R.F. (1992): "On the Determination of Regional Base and Regional Base Multipliers", *Regional Science and Urban Economics*, 22, pp. 619-635.
- CHOW, G.C. (1960): "Tests of Equality Between Sets of Coefficients in Two Linear Regressions", *Econometrica*, 28, pp. 591-605.
- CUTHBERSON, K., HALL, S. y TAYLOR, M. (1992): *Applied Econometric Techniques*, Phillip Allan, Nueva York.
- DAVIDON, W. (1963): "Variance Algorithm for Minimization", *Computer Journal*, 10, pp. 406-410.
- DAVIDSON, J., HENDRY, D., SRBA, F. y YEO, S. (1978): "Econometric Modelling of Aggregate Time-Series Relationship Between Consumer's Expenditure and Income in the United Kingdom", *The Economic Journal*, 88, pp. 661-692.
- DE JONG, P. (1991): "The Diffuse Kalman Filter", *The Annals of Statistics*, 19, pp. 1073-1083.
- DICKEY, D.A. y FULLER, W. (1979): "Distribution of the Estimators for Autoregressive Time Series with a Unit Root", *Journal of American Statistical Association*, 74, pp.427-431.
- DICKEY, D.A. y FULLER, W. (1981): "Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root", *Econometrica*, 49, pp. 1057-1072.
- DICKEY, D.A. y PANTULA, S.G. (1987): "Determining the Order of Differencing in Autoregressive Processes", *Journal of Business and Economics Statistics*, 5, 4, pp. 455-461.
- DZIECHCIARZ, J. (1989): "Changing and Random Coefficient Models. A Survey", en Hackl, P. (ed.), *Seasonal Analysis and Forecasting of Economic Structural Change*, Springer Verlag, Berlin.
- ENGLE, R.F. y GRANGER, C.W.J. (1987): "Cointegration and Error Correction: Representation, Estimation and Testing", *Econometrica*, 55, 2, pp. 251-276.

- ENGLE, R.F. y WATSON, M. (1987): "The Kalman Filter: Applications to Forecasting and Rational-Expectations Models", en Bewley, T. (ed.), *Advances in Econometrics, Fifth World Congress, Econometric Society Monograph*, 13, pp. 245-283, Cambridge University Press, Cambridge.
- FLETCHER, R. (1987): *Practical Methods of Optimization*, John Willey, Nueva York.
- FLETCHER, R. y POWELL, M. (1963): "A Rapidly Convergent Descent Method for Minimization", *Computer Journal*, 6, pp. 163-168.
- FULLER, W. (1976): *Introduction to Statistical Time Series Analysis*, John Willey & Sons, Nueva York.
- GARDNER, G., HARVEY, A.C. y PHILLIPS, G.D.A. (1980): "An Algorithm for Exact Maximum Likelihood Estimation of Autoregressive-Moving Average Models by Means of Kalman Filtering", *Applied Statistics*, 29, pp. 311-312.
- GÓMEZ, V. y MARAVALL, V. (1994): "Estimations, Prediction and Interpolation for Nonstationary Series with the Kalman Filter", *Journal of the American Statistical Association*, 89, pp. 611-624.
- GONZALO, J. (1994): "Five Alternative Methods of Estimating Long-Run Equilibrium Relationships", *Journal of Econometrics*, 16, pp. 121-130.
- GRANGER, C. W. J. (1981): "Some Properties of Time Series Data and their Use in Econometric Model Specification", *Journal of Econometrics*, 16, pp. 121-130.
- GRANGER, C. W. J. Y NEWBOLD, P. (1974): "Spurious Regressions in Econometrics", *Journal of Econometrics*, 2, pp. 111-120.
- HACKL, P. y WESTLUND, A.H. (1996): "Demand for International Telecommunication: Time-Varying Price Elasticity", *Journal of Econometrics*, 70, pp. 243-260.
- HAMILTON, J.D. (1994): *Time Series Analysis*, Princeton University Press, Princeton.
- HARVEY, A.C. (1981): *Time Series Models*, Phillip Allan, Oxford.
- HARVEY, A.C. (1984): "Dynamic Models, the Prediction Error Decomposition and State Space Models", en Hendry, D.F. y Wallis, K.F. (eds.) *Econometrics and Quantitative Economics*, pp. 37-59, Basil Blackwell, Oxford.
- HARVEY, A.C. (1989): *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge University Press, Cambridge.
- HARVEY, A.C. y PHILLIPS, G. (1979): "Maximum Likelihood Estimation of Regression Models with Autoregressive-Moving Average Disturbances", *Biometrika*, 66, pp. 69-58.
- HENDRY, D. F. Y MIZON, G. (1978): "Serial Correlation as a Convenient Simplification, not a Nuisance: A Comment on a Study of the Demand for Money By the Bank of England", *The Economic Journal*, 88, pp. 549-563.

- ISSERMAN, A. (1980): "Estimating Export Activity in a Regional Economy: A Theoretical and Empirical Analysis of Alternative Methods", *International Regional Science Review*, 5, pp. 155-184.
- JOHANSEN, S. (1988): "Statistical Analysis of Cointegration Vectors", *Journal of Economic Dynamics and Control*, 12, pp. 231-254.
- KALABA, R. y TESTFASION, L. (1990): "A Further Note on Flexible Least Squares and Kalman Filtering", *Journal of Economics Dynamics and Control*, 14, pp. 183-185.
- KALMAN, R. (1960): "A New Approach to Linear Filtering and Prediction Problems", *Transactions ASME, Journal of Basic Engineering*, 82, pp. 35-45.
- KALMAN, R. y BUCY, R. (1961): "New Results in Linear Filtering and Prediction Theory", *Transactions ASME, Journal of Basic Engineering*, 83, pp. 95-108.
- KITAGAWA, G. y GERSCH, W. (1984): "A Smoothness Prior-State Space Modelling of Time Series with Trend and Seasonality", *Journal of the American Statistical Association*, 82, pp. 1032-1063.
- KOHN, R. y ANSLEY, C. (1986): "Estimation, Prediction and Interpolation for ARIMA Models with Missing Data", *Journal of the American Statistical Association*, 81, pp. 751-761.
- KREMERS, J., ERICSSON, N. y DOLADO, J. (1992): "The Power of Cointegration Tests", *Banco de España, Documento de Trabajo 9218*.
- LESAGE, J.P. (1990): "Forecasting Metropolitan Employment Using an Export-Base Error-Correction Model", *Journal of Regional Science*, 30, pp. 307-323.
- LESAGE, J.P. y REED, J.D. (1989): "The Dynamic Relationship Between Export, Local and Total Area Employment", *Regional Science and Urban Economics*, 19, pp. 615-636.
- LUCAS, R. (1976): "Understanding Business Cycles", en Brunner, K. y Meltzer, A. (eds.), *Stabilization of Domestic and International Economy*, North-Holland, Amsterdam.
- LÜTKEPOHKL, H. y HERWARTZ, H. (1996): "Specification of Varying Coefficient Time Series Models via Generalized Flexible Least Squares", *Journal of Econometrics*, 70, pp. 261-290.
- MACKINNON, J. (1991): "Critical Values for Cointegration Test", en Engle, R.F. y Granger, C.W.J. (eds.), *Long-Run Economic Relationships*, pp. 267-276, Oxford University Press, Oxford.
- MIN, C. y ZELLNER, A. (1993): "Bayesian and Non-Bayesian Methods for Combining Models and Forecasts with Applications to Forecasting International Growth Rates", *Journal of Econometrics*, 56, pp. 89-118.
- MOLINAS, C., BALLABRIGA, F.C., CANADELL, E., ESCRIBANO, A., LÓPEZ, E., MANZANEDO, L., MESTRE, R., SEBASTIÁN, M. y TAGUAS, D. (1990): *MOISEES: Un Modelo de Investigación y Simulación de la Economía Española*, Antoni Bosch editor e IEF, Barcelona.

- MUR, J. y TRÍVEZ, F.J. (1996): "Dynamic Modeling of Interregional Economic Activity: An Application to the Spanish Labour Market", *Papers in Regional Science*, 75, pp. 463-481.
- NIJKAMP, P., RIETVELD, P. y SNICKARS, F. (1986): "Regional and Multiregional Economic Models: A Survey", en Nijkamp, P. (ed.), *Handbook of Regional and Urban Economics*, vol.I, North-Holland, Amsterdam.
- PHILLIPS, P.C.B. y PERRON, P. (1988): "Testing for a Unit Root in Time Series Regression", *Biometrika*, 75, pp. 335-346.
- QUANDT, R. (1958): "The Estimation of the Parameters of a Linear Regression System Obeying Two Separate Regimes", *Journal of the American Statistical Association*, 53, pp. 873-880.
- QUANDT, R. (1960): "Tests of the Hypothesis that a Linear Regression System Obeys Two Separate Regimes", *Journal of the American Statistical Association*, 55, pp. 324-330.
- RAMAJO, J. (1996): "Cointegración y Filtro de Kalman: Aplicación a la Demanda de M1 en Venezuela, 1983.I-1994.IV", Departamento de Economía Aplicada y Organización de Empresas, Documento de Trabajo de la Universidad de Extremadura.
- REITER, M. (1995): *The Dynamics of Business Cycles, Stylized Facts, Economic Theory, Econometric Methodology and Applications*, Physica-Verlag, Heidelberg.
- RICHARDSON, H. (1985): "Input-Output and Economic Base Multipliers: Looking Backward and Forward", *Journal of Regional Science*, 25, pp. 607-661.
- ROSENBERG, B. (1973): "Random Coefficient Models: The Analysis of a Cross Section of Time Series by Stochastically Convergent Parameter Regression", *Annals of Economic and Social Measurement*, 2, pp. 399-427.
- RUBIN, H. (1950): "Note on Random Coefficient", en Koopmans, T.C. (ed.), *Statistical Inference in Dynamic Economic Models*, John Wiley, Nueva York.
- SAID, S. y DICKEY, D.A. (1984): "Testing for Unit Roots in Autoregressive-Moving Average Models of Unknown Order", *Biometrika*, 71, pp. 599-607.
- SARGAN, J. D. (1964): "Wages and Prices in the United Kingdom: A Study of Econometric Methodology" en HART, P. y WHITAKER, J. (eds): *Econometric Analysis for Natural Economic Planning*, London, Butterworths.
- SERRANO, G. (1995): "Modelización Regional: Una Aplicación a la Economía Valenciana", en *IX Reunión Asepelt España*, Santiago de Compostela, 1, pp. 433-447.
- SHIVELY, T. y KOHN, R. (1997): "A Bayesian Approach to Model Selection in Stochastic Coefficient Regression Models and Structural Time Series Models", *Journal of Econometrics*, 76, pp. 39-52.
- SNYDER, R.D. y SALIGARI, G.R. (1996): "Initialisation of the Kalman Filter with Partially Diffuse Initial Conditions", *Journal of Time Series Analysis*, 17, pp. 409-424.
- STOCK, J. y WATSON, M. (1996): "Evidence on Structural Instability in Macroeconomic Time Series Relations", *Journal of Business and Economic Statistics*, 14, pp. 11-30.

SURIÑACH, J., ARTÍS, M., LÓPEZ-BAZO, E. y SANSÓ, A. (1995): *Análisis Económico Regional. Nociones Básicas de la Teoría de la Cointegración*, Antoni Bosch editor e IEF, Barcelona.

Anexo 1: Información estadística disponible

La reducida información estadística de que actualmente se dispone en nuestro país a nivel regional, a pesar de los importantes esfuerzos realizados en los últimos años por diversas instituciones tanto públicas como privadas, tanto en lo que se refiere a alargar las series como a las variables disponibles, es el principal problema con el que se encuentran los investigadores a la hora de realizar estudios de ámbito regional. Este hecho hace necesario que en muchas ocasiones se tenga que recurrir a variables *proxy*, lo cual conlleva una serie de efectos negativos en las propiedades de los estimadores¹⁸.

Más aún, el problema al que se hace referencia en el párrafo anterior es tanto más importante cuanto mayor es el nivel de desagregación sectorial utilizado en el estudio. De hecho, es realmente difícil encontrar variables disponibles que puedan ser potencialmente candidatas para explicar la evolución de los diferentes subsectores de las economías regionales que no sean la propia dinámica de la región. Ante este marco, en este trabajo tal y como se ha comentado en la introducción, se ha optado por trabajar con un nivel de desagregación sectorial de cuatro ramas de actividad.

Lógicamente como consecuencia de todo lo comentado, las especificaciones que se presentan en este trabajo no pueden ser consideradas como definitivas sino que están sujetas a revisión a medida que se disponga de información oficial adicional. En cualquier caso, el interés del modelo que se presenta radica en el tratamiento al que son sometidos los datos así como en la consideración de las técnicas de modelización utilizadas para el ámbito regional.

Las fuentes de información estadística utilizadas han sido la Contabilidad Regional y Nacional del INE, la base de datos HISPADAT, el Anuario Estadístico de España, el *Anuari Estadístic de Catalunya* y las memorias de las Cámaras Oficiales de Comercio, Industria y Navegación de Barcelona y Cataluña. La frecuencia de los datos es anual y el período considerado ha sido 1972-94, años para los que se dispone de información definitivas para todas las variables.

18. Como es sabido la omisión de variables relevantes lleva, por lo general, a obtener estimadores sesgados, inconsistentes e ineficientes.

Anexo 2: Datos utilizados

	VACT86	VINCT86	VBCT86	VSERCT86	VTOTCT86	VAAG86	VAIND86	VACST86	VASER86	VATOT86
1972	138	1786	472	2077	4473	1541	6819	2204	11712	22277
1973	152	1993	514	2267	4926	1598	7518	2392	12497	24005
1974	152	2014	547	2393	5106	1701	7929	2480	13155	25266
1975	139	2135	522	2469	5265	1698	7820	2381	13466	25365
1976	128	2184	504	2615	5430	1773	8096	2288	13992	26148
1977	129	2181	457	2720	5485	1698	8509	2244	14480	26932
1978	146	2132	434	2678	5390	1808	8696	2135	14700	27340
1979	147	2135	427	2815	5524	1733	8729	2056	14775	27293
1980	169	2080	374	2608	5231	1883	8752	2017	14905	27557
1981	144	1961	337	2666	5108	1704	8686	2013	15093	27497
1982	160	1935	369	2616	5081	1678	8540	2067	15540	27826
1983	147	2014	375	2609	5145	1783	8763	2069	15829	28444
1984	150	2053	328	2720	5250	1937	8716	1941	16224	28818
1985	141	2078	277	2762	5257	1997	8995	1985	16595	29572
1986	126	2168	346	2895	5536	1815	9427	2103	17182	30527
1987	146	2291	360	3069	5866	2025	9874	2278	17974	32152
1988	134	2409	399	3300	6242	2092	10322	2509	18828	33751
1989	156	2537	469	3508	6669	1953	10699	2848	19803	35304
1990	152	2609	512	3634	6906	2013	10918	3139	20588	36658
1991	131	2578	550	3782	7042	2006	11069	3233	21151	37459
1992	127	2561	525	3931	7144	1962	11005	3062	21665	37694
1993	122	2480	477	3935	7014	1988	10637	2891	21847	37363
1994	114	2603	472	4039	7227	1789	11076	2926	22233	38024

	PROIND	PPRIAG	LLUVIA	VIV	PERN	INV	INVO	DEFPIB 86
1972	2.10	23.37	3804.80	58260	10838	168623	25230	0.15
1973	2.25	26.34	1692.30	66174	12147	103388	17320	0.17
1974	2.27	28.83	2335.90	80189	9623	86490	16847	0.19
1975	2.20	33.66	2927.90	80979	10541	75946	17271	0.23
1976	2.39	37.15	2684.00	63387	10096	188056	49832	0.26
1977	2.52	45.14	3939.40	58081	11112	111038	36241	0.33
1978	2.63	50.20	2095.80	59805	12926	88478	34829	0.39
1979	2.70	54.60	2496.60	53881	13825	24704	11389	0.46
1980	2.81	57.41	2092.60	49290	11105	34900	18370	0.53
1981	2.91	65.51	2087.80	35022	15660	45903	27057	0.59
1982	3.03	75.79	3298.80	35007	15898	65519	43946	0.67
1983	3.18	82.53	2055.80	29953	17263	143361	107327	0.75
1984	3.24	90.84	2259.30	27022	18567	31355	26042	0.83
1985	3.47	92.70	1228.40	29136	17852	34145	30784	0.90
1986	3.55	100.00	1640.10	29665	19304	35930	35930	1.00
1987	3.59	105.05	2737.60	31735	19800	42411	44959	1.06
1988	3.70	109.98	1867.70	34573	20036	47857	53609	1.12
1989	3.69	114.90	2316.50	37838	19809	57596	69163	1.20
1990	3.67	119.59	2025.40	37034	19091	77968	100411	1.29
1991	3.83	123.80	1522.20	33971	20685	63220	87321	1.38
1992	3.92	126.68	2682.10	33847	18675	76103	112367	1.48
1993	4.19	128.37	2202.70	32385	19277	170978	263101	1.54
1994	4.48	119.09	2342.60	31982	23060	81915	131477	1.61

Anexo 3: Comparación entre los valores reales, los estimados y los predichos

