

DOCUMENTS DE TREBALL
DE LA DIVISIÓ DE CIÈNCIES JURÍDIQUES,
ECONÒMIQUES I SOCIALS

Col·lecció d'Economia

**POTENCIALIDAD DE LA MODELIZACIÓN *STATE-SPACE* Y EL
FILTRO DE KALMAN PARA EL ANÁLISIS REGIONAL. UNA
APLICACIÓN PARA EL ÍNDICE DE ACTIVIDAD INDUSTRIAL ***

Miquel Clar López
Raúl Ramos Lobo
Jordi Suriñach Caralt

Adreça correspondència

Grup de Recerca de Qualitat "Anàlisi Quantitativa Regional"
Departament d'Econometria, Estadística i Economia Espanyola
Facultat d'Econòmiques
Universitat de Barcelona
Avda. Diagonal 690 - 08034 Barcelona, Espanya
Tel: 934024320
Fax: 934021821
e-mail: mclar@eco.ub.es / rrlobo@eco.ub.es / surinach@eco.ub.es

Rebut: Juny 1998

* Los autores agradecen los comentarios y sugerencias realizadas por un evaluador anónimo así como de los participantes en el 38th. Congress of the European Regional Science Association celebrado en Viena del 28 de agosto al 1 septiembre de 1998.

ABSTRACT: The deficit existent in our country in order of the available of quantitative indicators to made a regional industrial activity conjuncture analysis have initiated a discussion in different forums about which is the best methodology for elaborating indicators of this characteristics.

In this context, in this paper we summarise and extend the main conclusions obtained in previous studies (Clar, *et al.*, 1997a, 1997b and 1998) about the possibility of extending the indirect methods that nowadays have been used in the Spanish regions to analyse the short term evolution of regional industrial production. This conclusions takes us to propose a different strategy for elaborating indirect quantitative indicators. In particular, (following Israilevich and Kuttner, 1993) a latent variable model for estimating the regional production activity is proposed.

This kind of models can be specified in terms of *state-space* model and estimated by the Kalman filter.

To validate the proposed methodology an indicators are estimated following this methodology for three of the four Spanish regions that have an Industrial Production Index (IPI) obtained by direct methods (*Andalucía*, *Asturias* and *País Vasco*) and we compare them with the published IPIs. The results obtained shows the good accuracy of the proposed strategy, opening a working line for surpassing the deficit existent nowadays in Spanish regions.

KEY WORDS: *State-space* models, Kalman filter, regional indicators, conjuncture, industrial activity, Production Industrial Index.

JEL Classification: L60, R11, R12.

RESUM: El dèficit existent al nostre país pel que fa a la disponibilitat d'indicadors quantitius amb els que dur a terme una anàlisi conjuntural de l'activitat industrial regional ha obert un debat centrat en l'estudi de quina és la metodologia més adient per a elaborar indicadors d'aquestes característiques.

Dins d'aquest marc, a aquest treball es presenten les principals conclusions obtingudes en anteriors estudis (Clar, *et al.*, 1997a, 1997b i 1998) sobre la idoneïtat d'estendre les metodologies que actualment s'estan aplicant a les regions espanyoles per a elaborar indicadors de l'activitat industrial mitjançant mètodes indirectes. Aquestes conclusions porten a plantejar una estratègia distinta a les que actualment es venen aplicant. En concret, es proposa (seguint a Israilevich i Kuttner, 1993) un model de variables latents per a estimar l'indicador de la producció industrial regional.

Aquest tipus de model pot especificar-se en termes d'un model *state-space* i estimar-se mitjançant el filtre de Kalman.

Per a validar la metodologia proposada s'estimen uns indicadors d'acord amb ella per a tres de les quatre regions espanyoles que disposen d'un Índex de Producció Industrial (IPI) elaborat mitjançant el mètode directe (Andalusia, Astúries i el País Basc) i es comparen amb els IPIs publicats (oficials). Els resultats obtinguts mostren el bon comportament de l'estratègia proposada, obrint així una línia de treball amb la que subsanar el dèficit al que es feia referència anteriorment.

PARAULES CLAU: Models *state-space*, filtre de Kalman, indicadors regionals, conjuntura, activitat industrial, Índex de Producció Industrial.

Classificació JEL: L60, R11, R12.

1. INTRODUCCIÓN

Las estadísticas cuantitativas se utilizan para una gran variedad de finalidades, relacionadas tanto con el funcionamiento corriente del sistema económico como con la formulación de políticas económicas a más largo plazo. En este sentido, la disponibilidad de un indicador cuantitativo de la actividad industrial permite entre otras cuestiones: *a)* llevar a cabo un seguimiento de la producción industrial en volumen excluyendo la incidencia de los precios; *b)* tener un conocimiento descriptivo del sector industrial de la economía analizada; *c)* disponer de uno de los principales instrumentos para el análisis económico coyuntural, especialmente si se utiliza en combinación con otros indicadores; *d)* disponer de un indicador de referencia en relación al que se juzga la evolución de otros indicadores como por ejemplo indicadores relativos al comercio exterior o a la ocupación; *e)* analizar la evolución de la oferta o la demanda agregadas según si se dispone de él desagregado por ramas de actividad¹ o por destino económico de los bienes; *f)* comparar (a los empresarios) la evolución de su producción con la del resto de empresas de su mismo sector o para hacer un seguimiento de la evolución de su sector en el conjunto de la industria; *g)* utilizarlo como variable *proxy* del valor de la producción industrial en modelos de crecimiento regional; y, *h)* realizar un seguimiento de la actividad económica general, ya sea en sí mismo o formando parte de indicadores sintéticos de actividad como por ejemplo el elaborado para la economía catalana². Además, también juega un papel relevante

¹ En este sentido, Revilla (1997) califica el Índice de Producción Industrial (IPI) como el principal indicador económico de oferta.

² El indicador sintético de la actividad económica de Cataluña, ISAEC, elaborado por el grupo de investigación *Anàlisi Quantitativa Regional* de la *Universitat de Barcelona* para la *Cambra Oficial de Comerç, Indústria i Navegació de Barcelona*, cuantifica la evolución mensual del VAB generado por los sectores industrial, de la construcción y de los servicios. Además, también, ofrece información sobre el VAB no agrario a partir de la agregación de los tres anteriores. La metodología para elaborarlo se basa en cuatro etapas: *a)* selección de los indicadores parciales (aquellos que mejor recogen el comportamiento de la actividad económica); *b)* estimación de la señal tendencia-ciclo aplicando el filtro de líneas aéreas

en la elaboración de Contabilidades Trimestrales por métodos indirectos. Todo ello no hace sino poner de manifiesto el interés de disponer de un indicador cuantitativo de la actividad industrial.

A la hora de elaborar un indicador cuantitativo con el que aproximar la evolución de la producción industrial de una determinada economía existen dos vías claramente diferenciadas entre sí desde el punto de vista metodológico dependiendo del método utilizado (directo o indirecto) para elaborarlo.

Por un lado, los indicadores cuantitativos directos se elaboran a partir de una estimación directa de la variable de interés, la producción industrial³. El adjetivo directo hace referencia al hecho que son indicadores que se construyen tomando como única fuente de información datos correspondientes a la producción industrial realizada en un determinado período. La información de base utilizada para elaborarlos proviene de una encuesta que suele ser especialmente diseñada para este fin. En este caso, el proceso de recopilación de la información implica necesariamente diseñar un cuestionario apropiado y definir una muestra de unidades productivas y productos que represente correctamente la composición sectorial y geográfica de la producción industrial. Así pues, es de vital importancia seleccionar adecuadamente las unidades productivas y los productos que representen los sectores más relevantes de la economía.

Sin duda alguna, este método permite obtener los mejores índices cuantitativos para realizar un seguimiento de la evolución de la producción industrial. A pesar de ello, sin embargo, presenta el inconveniente que el coste asociado a su elaboración es muy elevado: seleccionar las unidades productivas

modificado (LAM); *c*) trimestralización del VAB de los sectores productivos utilizando la misma metodología que el INE basada en el método de Chow-Lin (véase Suriñach *et al.*, 1996 y Artís *et al.*, 1997c); *y, d*) aplicación de técnicas multivariantes para el tratamiento de la información. Para un mayor detalle véase Artís *et al.* (1994, 1997a y 1997b).

³ Véase EUROSTAT (1978).

a encuestar, diseñar la cesta de productos representativos, elaborar la encuesta, recoger la información, comprobar la calidad de los datos, tratarlos, ...

Por otro lado, los indicadores cuantitativos indirectos de la actividad industrial se caracterizan por aproximar la producción industrial realizada en un determinado período de tiempo en una economía a partir de información pre-existente y no directamente a partir de datos de producción provenientes de las unidades productivas de la economía investigada. En consecuencia, la aproximación a la variable de interés no es (generalmente) tan exacta como la que se logra con los indicadores directos, pero tiene la ventaja que los costes que se han de soportar son mucho más reducidos. Por este motivo, los indicadores indirectos han estado (y están) siendo muy utilizados en un gran número de economías, principalmente de ámbito regional, que acostumbran a enfrentarse a mayores restricciones presupuestarias para dedicar a la información estadística.

En la literatura se han propuesto diferentes aproximaciones para obtener indicadores indirectos cuantitativos de la actividad industrial. Entre ellas las más utilizadas son las siguientes: *a)* utilizar como indicador el consumo de energía eléctrica para usos industriales; *b)* utilizar como información de base la producción que se realiza en otra economía (relacionada con la economía objeto de estudio, normalmente de un ámbito geográfico superior) adaptándola previamente a la economía considerada; *c)* partir de toda una batería de indicadores como por ejemplo registros administrativos de ventas, indicadores de ocupación, ...; y, *d)* estimar los parámetros de una función de producción para la economía considerada a nivel anual y extrapolar los resultados a nivel mensual, relacionando estas estimaciones con datos mensuales para los factores de producción (capital y trabajo).

Actualmente, la disponibilidad de información estadística existente en nuestro país a nivel regional, a pesar de la evidente mejora experimentada en los últimos años (tanto en lo que se refiere a su calidad como al número de

magnitudes consideradas y a su longitud) sigue presentado algunas lagunas⁴: hay una cierta escasez de indicadores⁵, hasta hace poco tiempo no existían deflatores regionales (como mínimo oficiales) lo cual no permitía transformar los indicadores en términos corrientes en constantes⁶, algunas fuentes estadísticas ven reducida muy sensiblemente su significación cuando se utilizan a nivel regional (por ejemplo, la Encuesta de Población Activa, EPA), ...

Uno de los ámbitos donde se pone de manifiesto este déficit de estadísticas regionales es el sector industrial: con la información estadística (cuantitativa) existente hoy en día no es posible analizar el comportamiento de la actividad industrial regional de manera homogénea para todas las Comunidades Autónomas (CC.AA.) españolas a corto plazo puesto que si bien el INE elabora un IPI de periodicidad mensual para el conjunto del Estado (general y sectorial) mediante el método directo no ofrece ninguna información de estas características desagregada a nivel regional, lo cual supone un grave inconveniente de acuerdo con lo señalado anteriormente referente a la importancia de disponer de indicadores cuantitativos de la actividad industrial⁷.

⁴ A pesar de todo ello, es necesario reconocer los importantes esfuerzos realizados en los últimos años por diferentes instituciones, tanto públicas como privadas, para subsanar este déficit. En este sentido, se ha de destacar, entre otros, el impulso dado desde los Institutos de Estadística Regionales, así como diversos estudios como los elaborados en el seno del Instituto de Estudios Valencianos (IVIE), del FIES, del BBV, del Proyecto HISPALINK, del Ministerio de Economía y Hacienda, de los Gobiernos Locales, de diversas entidades financieras, ..., encaminados todos ellos a elaborar información estadística regional más precisa, homogénea y coherente.

⁵ De hecho nuestro país es, juntamente con Portugal y Grecia, uno de los países que menos indicadores coyunturales facilita a la Oficina de Estadística de la Unión Europea (EUROSTAT).

⁶ El INE en la Contabilidad Regional de España base 1986, hasta 1996 únicamente publicaba información a precios corrientes sin facilitar información referente a los deflatores. Sin embargo, ello ha sido rectificado recientemente con la publicación por el INE de la serie homogénea del Producto Interior Bruto (PIB) a precios de mercado en términos corrientes y constantes para el período 1980-95.

⁷ Sin embargo, cabe señalar que recientemente el INE ha comenzado a publicar unos indicadores de la producción industrial regional elaborados a partir de métodos indirectos. En este sentido, en Clar *et al.* (1998) se puede encontrar un análisis de las ventajas e

Ante este marco, en nuestro país diversas instituciones (públicas y privadas) a lo largo de los últimos años han puesto en marcha diferentes iniciativas para superar las deficiencias existentes hasta hace poco tiempo en lo que a las estadísticas industriales regionales se refiere. Sin embargo, a pesar del importante esfuerzo realizado la realidad es que hasta hace pocos meses no existía un indicador cuantitativo a partir del cual llevar a cabo un seguimiento de la actividad industrial regional homogéneo para todas las comunidades españolas. Más aún, incluso había CC.AA. que ni tan solo disponían de un indicador de dichas características.

En concreto, de las diecisiete regiones españolas únicamente diez (País Vasco, Cataluña, Madrid, Andalucía, Asturias, Navarra, La Rioja, Baleares, Canarias y Extremadura) disponían de un indicador cuantitativo de la actividad industrial de su región, elaborados en unos casos a partir del método directo y en otros por métodos indirectos.

Las experiencias en nuestro país en lo que a la elaboración de un indicador de la producción industrial a partir directamente de los datos sobre el volumen de producción realizada por las unidades productivas de los diferentes sectores se refiere son más bien escasas. De hecho, de las diez comunidades señaladas en el párrafo anterior, las únicas que disponen de un índice obtenido mediante el método directo son el País Vasco, Asturias, Andalucía⁸ y Extremadura.

El método que utilizan para estas comunidades el Instituto Vasco de Estadística (EUSTAT), la Sociedad Asturiana de Estudios Económicos Industriales (SADEI), el Instituto de Estadística de Andalucía (IEA) y el Gobierno de Extremadura sigue las prácticas internacionales estándares. Como ya se ha comentado anteriormente, a pesar que este método es el más adecuado para

inconvenientes de dicho método así como una discusión referente a la fiabilidad de los indicadores elaborados de acuerdo con el método utilizado.

elaborar un indicador cuantitativo de la actividad industrial puesto que utiliza directamente información sobre la producción industrial realizada en la economía en cuestión, diversas causas hacen que sea (muy) difícil su aplicación de manera generalizada a áreas geográficas inferiores a un país. Así, entre otras, cabe destacar: *a)* el elevado coste (que en muchas ocasiones no puede ser asumido por los Gobiernos regionales debido a las restricciones presupuestarias a las que se enfrentan); *b)* el hecho que supone, en última instancia, una duplicación de la operación que ya hace el INE y por tanto un coste suplementario a añadir al realizado por el INE; y, *c)* un incremento de la carga administrativa que han de soportar los empresarios al haber de responder a dos encuestas de similares características (la del INE y la de la institución regional correspondiente).

Por su parte, en las CC.AA. de Madrid, Navarra, La Rioja, Baleares, Canarias y Cataluña (y Andalucía) se elaboran indicadores cuantitativos para conocer la evolución a corto plazo de su producción industrial mediante distintos métodos indirectos.

Las cinco primeras (y Andalucía) utilizan el consumo de energía eléctrica para usos industriales como *proxy* de la producción industrial⁹. Por su parte, el *Institut d'Estadística de Catalunya* (IEC) para elaborar el indicador (cuantitativo) de la actividad industrial catalana utiliza como información de base, de acuerdo con los principios básicos que rigen la *Llei del Pla Estadístic de Catalunya 1992-95*¹⁰ (que se pueden sintetizar en minimizar los costes y las molestias a los ciudadanos no duplicando operaciones estadísticas), los IPIs para las diferentes ramas de actividad a un nivel de desagregación sectorial de cuatro dígitos de la

⁸ En Andalucía, sin embargo, el Instituto de Estadística de Andalucía además de elaborar este indicador directo también elabora un indicador a partir de métodos indirectos llamado Índice General de Actividad Industrial de Andalucía (IGAIA).

⁹ De hecho, en la literatura desde hace más de dos décadas se viene proponiendo la validez de esta estrategia. Así, por ejemplo, en nuestro país en esta línea cabe destacar los trabajos de Sanz (1979) y Molina y Sanz (1985).

¹⁰ Ley 30/91 de 13 de diciembre de la *Generalitat de Catalunya*. Para una presentación de esta ley véase Oliveras (1992).

Clasificación Nacional de Actividades Económicas 1974 -CNAE74- (esto es, la máxima desagregación sectorial posible) elaborados por el INE para la economía española¹¹.

Ante esta heterogeneidad de indicadores existentes, las dificultades para compararlos resultan evidentes: la metodología, el número de sectores industriales considerados, la fecha de inicio de las series de los indicadores, el año base seleccionado, ..., son diferentes.

Ante este marco, en diversos trabajos anteriores (Clar, *et al.*, 1997a, 1997b y 1998) se analizó dichas metodologías, las hipótesis en que se basan y la idoneidad de extenderlas al resto de regiones españolas con el fin de disponer de un indicador cuantitativo homogéneo con el que poder llevar a cabo un seguimiento coyuntural de la actividad industrial para todas las CC.AA. españolas. La principal conclusión a la que se llegó es que a nivel regional, si bien la aplicación de estas metodologías presentan la ventaja de tener asociado un (muy) reducido coste, los indicadores obtenidos no recogen de manera totalmente correcta la evolución de la producción industrial dado que la fiabilidad de dichas metodologías depende de todo un conjunto de hipótesis que no se cumplen en muchas regiones españolas (para un detalle véase Clar, *et al.* 1997a, 1997b y 1998).

Así pues, continuando con la misma línea de investigación, en este trabajo se plantea una estrategia alternativa (que no se aplica actualmente en ninguna región española) para elaborar indicadores de la actividad industrial regional. En concreto, siguiendo a Israilevich y Kuttner (1993), se propone un modelo de variables latentes. Brevemente, la idea básica que hay detrás de esta propuesta consiste en considerar la variable que se desea estimar (el indicador de la actividad industrial regional en este caso) como una variable no observable (variable latente, variable de estado) que depende de un conjunto de

¹¹ Para un detalle sobre la metodología seguida por el IEC en la elaboración del indicador de

variables que sí lo son. De esta manera, se especifican dos ecuaciones: la ecuación de medida (que relaciona las variables conocidas con la variable de estado desconocida) y la ecuación de transición (que contempla la posibilidad que los diferentes estados de la naturaleza puedan variar con el paso del tiempo de acuerdo con un proceso autorregresivo de tipo Markov de primer orden). Estas dos ecuaciones forman lo que se conoce como modelo *state-space*. Este tipo de modelos pueden ser estimados de manera óptima (bajo unos supuestos básicos) mediante el filtro de Kalman. El filtro de Kalman es un algoritmo que consiste en un conjunto de ecuaciones que, aplicadas secuencialmente, permiten obtener la estimación óptima (en términos de error cuadrático medio) del vector de estado en el instante t teniendo en cuenta toda la información disponible.

Para validar la metodología propuesta se calculan de acuerdo con ella indicadores de la actividad industrial para tres de las cuatro regiones españolas que disponen de un IPI elaborado mediante el método directo (País Vasco, Asturias y Andalucía) y, a continuación, se lleva a cabo un estudio comparativo entre los indicadores elaborados y los índices regionales publicados (oficiales)¹².

Así pues, en el siguiente apartado se presenta el modelo teórico y se especifica en términos de un modelo *state-space*. A continuación se aborda la estimación del modelo para el caso de las comunidades del País Vasco, Asturias y Andalucía obteniéndose como *output* los índices (indirectos) de la producción industrial de dichas regiones y se validan los resultados obtenidos. Finalmente, se presentan las principales conclusiones del trabajo.

la actividad industrial para la comunidad catalana véase Costa y Galter (1994).

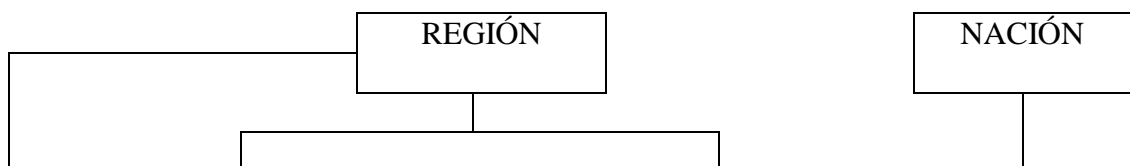
¹² El hecho de no aplicar la metodología propuesta a Extremadura (cuarta región que dispone de un índice obtenido con el método directo) es debido a que en el momento de realizar este trabajo no se disponía de los datos correspondientes al IPI de dicha comunidad.

2. UN MODELO DE VARIABLES LATENTES PARA ESTIMAR LA ACTIVIDAD INDUSTRIAL REGIONAL

2.1. Introducción

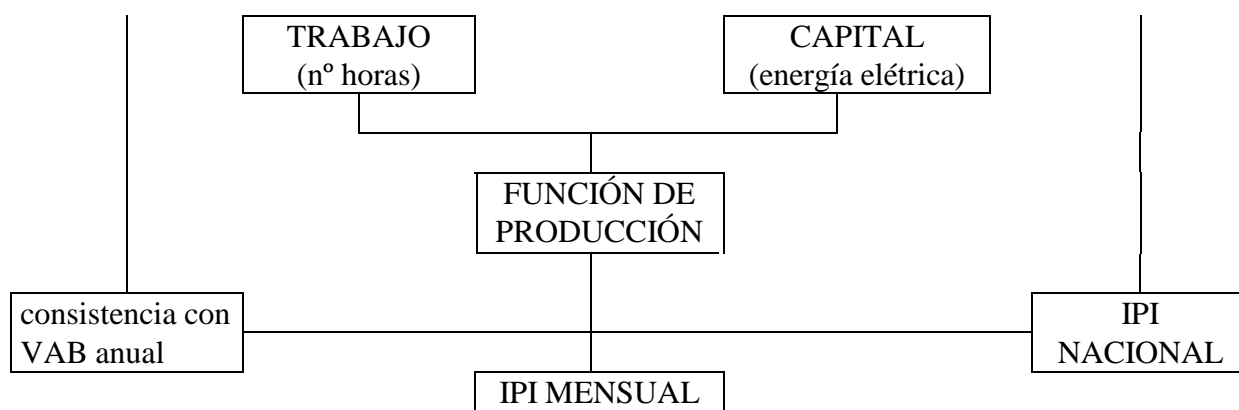
Tal y como se ha apuntado anteriormente, el indicador regional de la producción industrial puede interpretarse como una variable latente (no observable). Desde esta perspectiva, es posible utilizar el instrumental relacionado con los modelos *state-space* y el filtro de Kalman (juntamente con un algoritmo de alisado) para estimarlo¹³. En concreto, siguiendo a Israilevich y Kuttner (1993) se modeliza la producción mensual regional como una variable latente que depende de diferentes variables observables regionales (el trabajo y el capital) y nacionales (el IPI del INE). Además se impone la consistencia entre el índice estimado mensual y las series anuales observadas (o predichas) del VAB regional (véase cuadro 1). De esta manera, se solucionan dos problemas que son comunes en la mayoría de métodos indirectos: *a)* los indicadores mensuales no acaban de recoger correctamente la evolución anual de las series macroeconómicas; y, *b)* las estimaciones del índice regional se basan enteramente en el supuesto de la validez del índice nacional como un perfecto indicador del *output* regional¹⁴. Así pues, el indicador propuesto es de tipo indirecto y econométrico.

Cuadro 1. Esquematización del modelo propuesto



¹³ Véase Harvey (1982, 1987), Engle y Watson (1987) o Aoki (1990) para un detalle sobre la estimación de variables latentes a partir de modelos *state-space* y el filtro de Kalman.

¹⁴ Nótese que precisamente esta es una de las hipótesis intrínsecas en la metodología utilizada por el IEC.



Fuente: Elaboración propia a partir de Israilevich y Kuttner (1993).

2.2. El modelo teórico

Siguiendo a Israilevich y Kuttner la estimación del *output* industrial mensual regional puede llevarse a cabo a partir de un modelo de frecuencia mixta, esto es combinando información de periodicidad anual y mensual. En concreto, la estrategia que proponen estos autores para obtener un indicador de la producción industrial se fundamenta en dos hipótesis:

- a) la variable de interés no observable (el *output* industrial mensual regional) puede modelizarse como una variable estocástica latente cuyo comportamiento viene determinado por un conjunto de indicadores regionales (capital y trabajo) sobre los que se dispone de información mensual; y,
- b) existe una relación entre las fluctuaciones del *output* industrial regional y nacional, por lo que éstas pueden aportar información (indirecta) sobre las primeras.

La primera de las hipótesis anteriores permite especificar una función de producción paramétrica de tipo Cobb-Douglas mensual regional donde el *output*

industrial (mensual y regional) es función de los *inputs* capital (aproximado por el consumo de energía eléctrica para usos industriales realizado en la región¹⁵) y el trabajo (aproximado por las horas de trabajo efectivo -directo- en la producción):

$$\Delta x_{t,s}^{reg} = \gamma + \phi \Delta e_{t,s}^{reg} + \theta \Delta l_{t,s}^{reg} + \eta_{t,s}, \quad [2.1]$$

donde el primer subíndice, t , hace referencia al año mientras que el segundo, s , representa el mes; $\Delta(\cdot)_{t,s}$ denota el operador diferencias mensual: $\Delta(\cdot)_{t,s} = (\cdot)_{t,s} - (\cdot)_{t,s-1}$; x^{reg} representa (el logaritmo de) la producción regional (no observable); e^{reg} el (logaritmo del) consumo de energía eléctrica regional para usos industriales; l^{reg} el (logaritmo del) número de horas trabajadas efectivas en la región en el período considerado¹⁶; ϕ y θ son, respectivamente, la participación de los *inputs* energía (capital) y (horas de) trabajo; γ mide el progreso tecnológico, dado que es la tasa mensual del cambio tecnológico (de la eficiencia productiva); y, η es el término de perturbación que recoge los *shocks* en la función de producción. Además, como es habitual en la literatura se supone que la función de producción [2.1] es neutra en el sentido de Hicks, es decir, que un cambio en la tecnología repercute de la misma manera en la relación capital-producto que en la relación trabajo-producto.

El problema que se plantea a la hora de estimar [2.1] es evidente: al no disponer de información mensual sobre el *output* industrial regional no es posible estimarla directamente. En este punto es donde entra en juego la

¹⁵ De hecho, aproximar el *input* capital por el consumo de energía eléctrica es una práctica habitual en muchos trabajos empíricos debido a que, como señalan Griliches y Jorgenson (1966, pp. 50-51) y Moody (1974, pp. 45-46), existen diversidad de problemas tanto de tipo conceptual como práctico que son prácticamente insuperables y que hacen que sea (muy) difícil medir este *input*. Por otro lado, tal y como muestra Moody, existen razones de tipo teórico y práctico que justifican aproximarlos por el consumo de energía eléctrica para usos industriales.

segunda de las hipótesis señaladas anteriormente: si las fluctuaciones (económicas) nacionales y regionales están relacionadas, los indicadores nacionales proporcionarán una medida indirecta de la actividad económica regional. Así pues, en el caso analizado en concreto, puede considerarse que el IPI nacional es una medida indirecta de la actividad industrial regional. En otras palabras, estadísticamente el IPI mensual nacional puede considerarse un indicador con ruido del indicador regional subyacente. En consecuencia, la forma natural de especificar lo anterior consiste en relacionar las fluctuaciones mensuales del IPI nacional con las del *output* regional más un término de perturbación estocástico:

$$\Delta x_{t,s}^{nac} = \mu + \delta \Delta x_{t,s}^{reg} + v_{t,s}, \quad [2.2]^{17}$$

donde $x_{t,s}^{nac}$ representa (el logaritmo del) IPI nacional correspondiente al mes s del año t ; v el término de perturbación; y, μ una constante que se incluye en la ecuación para permitir que el ratio de crecimiento en el IPI nacional (mensual) y en el *output* regional (mensual) puedan ser diferentes. En concreto, μ es la

¹⁶ Nótese por tanto que, como es habitual en la literatura, las variables en [2.1] están expresadas en tasas de crecimiento (en diferencias sobre logaritmos).

¹⁷ A pesar del parecido entre esta ecuación y una ecuación de regresión clásica es importante remarcar la principal diferencia existente entre ambas: en el lado derecho de [2.2] aparece una variable no observable. Precisamente, el hecho que $\Delta x_{t,s}^{reg}$ sea no observable junto con los términos de perturbación de las ecuaciones [2.1] y [2.2] hace que no sea posible expresar el modelo formado por estas dos ecuaciones en términos de una única regresión. Ello sólo sería posible en los dos siguientes supuestos:

a) cuando la varianza del término de perturbación de la ecuación [2.1], σ_{η}^2 , fuese cero. En este supuesto la función de producción regional sería determinista, con lo cual podría sustituirse directamente en [2.2]; y,

b) cuando el coeficiente de correlación entre el IPI (mensual) nacional y el *output* (mensual) regional fuese igual a uno. En este caso la varianza del término de perturbación de la ecuación [2.2], σ_v^2 , sería cero, con lo cual podría utilizarse $\Delta x_{t,s}^{nac}$ en lugar de $\Delta x_{t,s}^{reg}$ en la función de producción quedando el modelo reducido a una única ecuación.

diferencia entre ambos ratios de crecimiento respecto a un año base. Así pues, valores positivos de μ están asociados a un crecimiento más lento en la región que en la nación¹⁸.

Para mostrar más claramente lo anterior, considérese la derivación siguiente: en el supuesto que todas las regiones integrantes de un país dispusiesen de un indicador (cuantitativo) de la actividad industrial elaborado de acuerdo con la misma metodología, el indicador para el conjunto del Estado podría obtenerse simplemente como la agregación de los regionales ponderados adecuadamente para recoger el diferente peso (importancia) de la industria de cada región en el ámbito nacional. Esto es, en términos de la nomenclatura que se está utilizando y, suponiendo M regiones:

$$\Delta x_{t,s}^{nac} = \sum_{i=1}^M \alpha_i \Delta x_{t,s}^{region i}, \quad [2.3]$$

donde $\sum_{i=1}^M \alpha_i = 1$.

Por otro lado, si se considera el supuesto que el comportamiento de la actividad industrial de las regiones depende, por un lado, de todo un conjunto de factores específicos de la propia región y, por otro, de unos factores comunes para todas ellas resultado de la dinámica económica del Estado común para todas las regiones (mismo marco institucional: política fiscal, monetaria, ..., son las mismas para todas las regiones), el $\Delta x_{t,s}^{region i}$ puede descomponerse como sigue:

¹⁸ En términos del trabajo de Norrbin y Schlagenhauf (1988) la ecuación [2.2] se interpretaría de la siguiente manera: las fuentes de las fluctuaciones del *output* son factores no observables específicos de la industria y de la región. De acuerdo con esta interpretación, δ es el peso asociado a las fluctuaciones del *output* regional mientras que el término de perturbación, v ,

$$\Delta x_{t,s}^{region\ i} = \Delta x_{t,s}^{comun} + \Delta x_{t,s}^{especifico\ region\ i} . \quad [2.4]$$

Sustituyendo [2.4] en [2.3] y operando se tiene que:

$$\Delta x_{t,s}^{nac} = \sum_{i=1}^M \alpha_i \left(\Delta x_{t,s}^{comun} + \Delta x_{t,s}^{especifico\ region\ i} \right) = \sum_{i=1}^M \alpha_i \Delta x_{t,s}^{comun} + \sum_{i=1}^M \alpha_i \Delta x_{t,s}^{especifico\ region\ i} = \quad [2.5]$$

$$= \Delta x_{t,s}^{comun} \sum_{i=1}^M \alpha_i + \sum_{i=1}^M \alpha_i \Delta x_{t,s}^{especifico\ region\ i} = \Delta x_{t,s}^{comun} + \sum_{i=1}^M \alpha_i \Delta x_{t,s}^{especifico\ region\ i} .$$

Si se considera una región j aisladamente del resto, [2.5] puede de la forma:

$$\Delta x_{t,s}^{nac} = \Delta x_{t,s}^{comun} + \alpha_j \Delta x_{t,s}^{especifico\ region\ j} + \sum_{i=1}^{M-j} \alpha_i \Delta x_{t,s}^{especifico\ region\ i} . \quad [2.5bis]$$

Si en [2.5bis] se suma y resta la constante $\alpha_j \Delta x_{t,s}^{comun}$ se llega a:

$$\Delta x_{t,s}^{nac} = (1 - \alpha_j) \Delta x_{t,s}^{comun} + \alpha_j \left(\Delta x_{t,s}^{especifico\ region\ j} + \Delta x_{t,s}^{comun} \right) + \sum_{i=1}^{M-j} \alpha_i \Delta x_{t,s}^{especifico\ region\ i} \quad [2.6]$$

El tercer sumando de [2.6] recoge la variación en $\Delta x_{t,s}^{nac}$ que es consecuencia (que tiene su origen) en la variación de la producción regional específica del resto de regiones (todas menos la j -ésima). En términos de la ecuación [2.2] es pues el ruido $\mathbf{n}_{t,s}$. Lo que queda recogido dentro de los corchetes

representa aquellas fuentes de las fluctuaciones del IPI nacional que no están relacionadas con el *output* de las regiones.

del segundo sumando no es más que la variación total experimentada por la región j -ésima, $\Delta x_{t,s}^{reg\ j}$. Por último, el primer sumando es una constante. De acuerdo con lo anterior, [2.6] puede escribirse como sigue:

$$\Delta x_{t,s}^{nac} = \mu + \delta \Delta x_{t,s}^{reg\ j} + v_{t,s}, \quad [2.7]$$

donde $\mu = (1 - \alpha_j) \Delta x_{t,s}^{comun}$ y $\delta = \alpha_j$. De esta manera se llega a la ecuación [2.2] propuesta por Israilevich y Kuttner.

El análisis de [2.2] (o [2.7]) pone de manifiesto que la variación mensual en el logaritmo (la tasa de crecimiento mensual) del IPI nacional cambia con la del *output* de la región j de acuerdo con una constante δ . Por consiguiente, $n_{t,s}$ recoge los movimientos (las fluctuaciones) mensuales de (la tasa de crecimiento del) IPI nacional que no tienen su origen en la región considerada.

De acuerdo con lo anterior pues, δ y σ_v^2 son dos medidas que permiten cuantificar el *link* entre (las fluctuaciones de) la región j y la nación. En concreto, δ es un factor de escala relativo a la dirección y la magnitud de las fluctuaciones regionales-nacionales. Valores positivos de δ son indicativos de una relación directa entre las fluctuaciones de la región j y las nacionales: al aumentar (disminuir) $\Delta x_{t,s}^{reg\ j}$ aumenta (disminuye) $\Delta x_{t,s}^{nac}$ ¹⁹.

Por lo que se refiere a la magnitud de las fluctuaciones, cuanto mayor sea δ mayor es el efecto de los movimientos de la región j sobre los nacionales, en otras palabras, cuanto mayor sea el valor del parámetro δ mayor es la correlación entre las fluctuaciones nacionales y regionales. Si $(0 <) \delta < 1$ las fluctuaciones nacionales

¹⁹ Nótese que este parámetro no puede tomar valores negativos dado que ello significaría que hay una relación inversa entre (las fluctuaciones del) *output* industrial de la región j y el IPI nacional lo cual por definición no es posible dado que ambas variables se mueven en la misma dirección.

son menores que las regionales: una variación del *output* regional provoca una variación proporcionalmente menor en el índice nacional (así, por ejemplo, si δ es 0,5 una variación del 10% en el *output* regional provoca una variación del 5% en el índice nacional) y viceversa si $\delta > 1$. Desde este punto de vista, por tanto, δ tiene una interpretación análoga a la pendiente de una ecuación de regresión clásica (con la restricción que no puede tomar valores negativos): determina el tamaño relativo de las fluctuaciones entre el *output* regional y el índice nacional.

Sin embargo, aunque no hay (al menos desde un punto de vista econométrico) razón alguna por la que el rango de valores plausibles del parámetro δ esté acotado entre cero y uno, es de esperar (desde un punto de vista económico) que así sea. Piénsese que valores de δ dentro del círculo unidad (positivo) implica que las fluctuaciones regionales son mayores que las nacionales, lo cual es consistente con el hecho que el IPI nacional se obtiene a partir de la producción realizada en todas las regiones que integran la nación y no sólo a partir de la producción de la región considerada, por lo que a la hora de elaborar el IPI nacional la producción de cada región tiene que estar ponderada por el peso de ésta en el total producido a nivel nacional. En el límite, δ sería uno si el IPI nacional estuviese determinado únicamente a partir del *output* industrial de la región considerada. De todo lo anterior puede concluirse, por tanto, que el valor de δ depende en buena parte del peso relativo que la producción de la región considerada represente en la producción total de la nación. Todo ello puede comprobarse fácilmente a partir de [2.6] y [2.7]:

si $\delta = \alpha_j = 0$, entonces $\mu = \Delta x_{t,s}^{comun}$ y, por tanto, $\Delta x_{t,s}^{nac} = \Delta x_{t,s}^{comun} + v_{t,s}$ con

$$v_{t,s} = \sum_{i=1}^{M-j} \alpha_i \Delta x_{t,s}^{especifico\ region\ i}; \text{ y,}$$

si $\delta = \alpha_j = 1$, entonces $\mu = 0$ y, por tanto, $\Delta x_{t,s}^{nac} = \Delta x_{t,s}^{reg\ j}$.

Por otro lado, σ_v^2 es equivalente a la varianza del término de perturbación de un modelo de regresión clásico puesto que no es más que una medida de la cantidad de ruido que hay en el *link* región-nación. Por tanto, cuanto menor sea menor es el ruido, es decir, menos importancia tienen las variaciones en el *output* industrial regional de otras regiones en el indicador nacional; en otras palabras, mayor es la dependencia de las fluctuaciones del IPI nacional de las fluctuaciones del *output* industrial de la región considerada.

De hecho, σ_v^2 está acotada inferiormente por cero y superiormente por $var(\Delta x_{t,s}^{nac})$. La cota inferior está asociada al supuesto de ausencia de ruido lo cual quiere decir que existe una correlación perfecta entre el *output* industrial regional y el IPI nacional: las fluctuaciones en el IPI nacional vienen completamente determinadas por las fluctuaciones en el *output* industrial de la región considerada. Por el contrario, si σ_v^2 alcanza la cota superior todo es ruido: las fluctuaciones en el IPI nacional no tienen ninguna relación con las fluctuaciones en el *output* industrial de la región en cuestión.

Resumiendo, pues, puede afirmarse, de acuerdo con todo lo anterior que en líneas generales cuanto mayor sea δ y menor sea σ_v^2 mayor es el *link* de las fluctuaciones región-nación.

Israilevich y Kuttner para medir el *link* región-nación proponen normalizar σ_v^2 por la varianza de la variable endógena (el índice nacional). Así, obtienen un estadístico al que denominan *pseudo-R*² dada su similitud (en términos de interpretación) con el *R*² de un modelo de regresión lineal estándar:

$$pseudo - R^2 = 1 - \frac{\sigma_v^2}{var(\Delta x_{t,s}^{nac})}, \quad [2.8]$$

de manera que cuanto mayor es el *pseudo-R*² mejor es el ajuste. Ello significa que mayor es el *link* entre las fluctuaciones del *output* regional y nacional. Desde este punto de vista, por tanto, el *pseudo-R*² es una medida para evaluar cómo son de informativas las fluctuaciones del indicador nacional para inferir el tamaño y la dirección de las fluctuaciones regionales: si el *pseudo-R*² es cero (o lo que es lo mismo, si σ_v^2 es igual a la varianza del IPI nacional), las fluctuaciones en el IPI nacional no aportan ninguna información sobre las fluctuaciones en el *output* industrial de la región considerada dado que no hay ninguna relación entre ellas (todo es ruido). Por el contrario, si el *pseudo-R*² toma valores cercanos a uno el modelo propuesto permite estimar el *output* regional de forma mucho más precisa que la que se conseguiría a partir de un modelo que no considerase la información nacional (mensual)²⁰.

Por otro lado, aunque no existe información sobre el *output* regional con frecuencia mensual sí existe información sobre el *output* con frecuencia anual. Para que las estimaciones del *output* regional mensual sean consistentes con el (único) indicador de la producción regional disponible (el dato anual del VAB) es necesario imponer en el modelo que las estimaciones mensuales cumplan la siguiente condición:

$$\Delta^A x_t^{reg} = \frac{1}{12} \left(\sum_{s=1}^{12} \sum_{j=0}^{11} \Delta x_{t,s-j}^{reg} \right), \quad [2.9]$$

donde $\Delta^A x_t^{reg}$ representa la variación anual del (logaritmo) de la producción regional: $x_t^{reg} - x_{t-1}^{reg}$. Precisamente, tal y como se ha dicho anteriormente, una de

las ventajas del modelo propuesto es que al incorporar [2.9] se garantiza la consistencia entre el índice mensual regional estimado y la serie observada (o predicha) del VAB regional anual²¹.

2.3. Especificación del modelo en términos de un modelo *state-space*

Las relaciones expresadas en las ecuaciones [2.1], [2.2] (o [2.7]) y [2.9] pueden considerarse conjuntamente (facilitando su estimación) especificándolas en forma de un modelo *state-space*.

De hecho, muchos modelos dinámicos convencionales pueden expresarse fácilmente en términos de un modelo *state-space*. Los modelos *state-space* ofrecen una mayor flexibilidad a la hora de identificar y estimar modelos dinámicos, lo que ha hecho que en los últimos años hayan sido a menudo utilizados en el campo de la Economía para afrontar problemas que con especificaciones más clásicas o bien eran complicadas o, en el peor de los casos, no se podían abordar.

El punto de partida de los modelos *state-space* es un conjunto de valores observados a lo largo del tiempo denotados por $(Y_1, Y_2, \dots, Y_{t-1}, Y_t)^{22}$, que pueden representar un escalar o un vector. La idea que hay detrás de los modelos *state-space* es que estas variables observables Y_t , sujetas a un ruido simbolizado por e_t , permiten captar la dinamicidad (el comportamiento) de una (o unas) variable(s)

²⁰ Nótese por tanto que una ventaja del modelo propuesto para estimar el *output* industrial regional es que como subproducto se obtienen un conjunto de medidas sobre el grado y la naturaleza de los *linkages* entre la actividad económica regional y nacional.

²¹ Alternativamente, podría especificarse la ecuación [2.9] en términos estocásticos sin más que incorporando un término de perturbación. En tal caso, la tasa anual derivada del indicador estimado podría considerarse como un indicador “con ruido” de la tasa de crecimiento del VAB industrial:

$$\Delta^A x_t^{reg} = \frac{1}{12} \left(\sum_{s=1}^{12} \sum_{j=0}^{11} \Delta x_{t,s-j}^{reg} \right) + \xi_t^A. \quad [2.9bis]$$

²² La notación que se va a utilizar es, básicamente, la propuesta por Harvey (1989).

no observable(s), \mathbf{a}_t , conocidas con el nombre de vector de estado. El principal objetivo de expresar un modelo en forma *state-space* consiste en obtener información acerca del comportamiento de \mathbf{a}_t teniendo en cuenta (en base a) su relación con Y_t .

La relación entre el vector Y_t , de orden $nx1$, y el vector \mathbf{a}_t , de orden $mx1$, se supone (para simplificar) que es de tipo lineal, de forma que puede ser expresada mediante la siguiente ecuación conocida con el nombre de ecuación de medida:

$$Y_t = Z_t \alpha_t + d_t + \varepsilon_t, \quad [2.10]$$

donde Z_t es una matriz de orden nxm , d_t es un vector de orden $nx1$ de variables exógenas y ε_t es un vector de orden $nx1$ de términos de perturbación que se supone que se distribuyen idéntica e independientemente según una normal con valor esperado cero, incorrelacionados y matriz de varianzas y covarianzas conocida, H_t : $\varepsilon_t \sim Niid(0_{nx1}, H_{nxn})$.

A pesar que en general los elementos de \mathbf{a}_t no son observables (conocidos), se supone que se comportan de acuerdo con un proceso autorregresivo de Markov de primer orden:

$$\alpha_t = T_t \alpha_{t-1} + c_t + R_t \theta_t, \quad [2.11]$$

donde T_t es una matriz de orden mxm , c_t es un vector de orden $mx1$ de variables exógenas que influyen en el comportamiento de \mathbf{a}_t , R_t es una matriz de orden mxg y θ_t es un vector de orden $gx1$ de términos de perturbación incorrelacionados con valor esperado cero y matriz de varianzas y covarianzas Q_t : $\theta_t \sim Niid(0_{mx1}, Q_{mxm})$.

La ecuación [2.11] es conocida como ecuación de transición (o ecuación de estado o ecuación del sistema) y junto con la ecuación de medida [2.10] forman el

modelo *state-space*. Z_t , d_t , H_t , T_t , c_t , R_t y Q_t son las denominadas matrices del sistema. Por último, la especificación del modelo *state-space* se completa con dos hipótesis adicionales referentes a los valores iniciales del vector de estado y a las matrices de varianzas y covarianzas de los términos de perturbación. En concreto, se supone, por un lado, que el valor esperado y la varianza de α_t para $t=0$ vienen dados respectivamente por:

$$\begin{aligned} E(\alpha_0) &= a_0, y \\ \text{var}(\alpha_0) &= P_0; \end{aligned} \quad [2.12]$$

y, por otro, que no existe correlación entre los términos de perturbación de las ecuaciones de medida y estado, ni entre los términos de perturbación de ambas ecuaciones y el vector de valores iniciales de las variables no observables:

$$E(\varepsilon_t \cdot \theta'_s) = 0 \quad s, t = 1, \dots, T; \quad [2.13]$$

$$\begin{aligned} E(\varepsilon_t \cdot \alpha_0) &= 0 \quad t = 1, \dots, T; \\ E(\theta_t \cdot \alpha_0) &= 0 \quad t = 1, \dots, T. \end{aligned} \quad [2.14]$$

La definición de \mathbf{a}_t depende de las características del sistema considerado, pero es importante señalar que generalmente es posible especificar más de un modelo *state-space* para cada sistema. En cualquier caso, de acuerdo con el principio de parametrización escueta es preferible formular aquella especificación que contenga un menor número de parámetros.

La principal dificultad a la hora de aplicar los modelos *state-space* en el ámbito de la Economía (y de las Ciencias Sociales en general) radica en su dependencia de los parámetros del sistema Z_t , H_t , T_t , R_t y Q_t que son desconocidos. Estos elementos son conocidos con el nombre de hiperparámetros y son los que determinan las características estocásticas del modelo. Sin embargo,

en el ámbito de la Ingeniería, donde se desarrollaron inicialmente estos modelos, los hiperparámetros suelen ser conocidos puesto que se conocen las leyes que rigen los sistemas. En consecuencia, la aplicación de los modelos *state-space* en Economía requiere de un esfuerzo adicional para estimarlos previamente.

Una posible especificación de las ecuaciones [2.1], [2.2] y [2.9] en forma de modelo *state-space* se presenta en las ecuaciones [2.15] y [2.16]²³.

²³ En el supuesto de considerar en el modelo la ecuación [2.9bis] (véase nota a pie número 21) en lugar de la ecuación [2.9], el último sumando de [2.15] quedaría como sigue:

$$\begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 1 \end{pmatrix} \begin{pmatrix} \xi_t^A \\ v_{t,12} \\ \dots \\ v_{t,1} \end{pmatrix}.$$

2.4. Estimación del modelo para las comunidades del País Vasco, Asturias y Andalucía. Validación de los resultados

2.4.1. El filtro de Kalman

Para aplicar el filtro de Kalman el modelo considerado debe estar especificado en términos de un modelo *state-space*. Así, como en muchas aplicaciones en el ámbito económico, en este trabajo se utiliza el filtro de Kalman como herramienta para estimar una variable latente (no observable) a partir de información sobre variables que sí son observables.

Así pues, la estimación del modelo formado por las ecuaciones [2.15] y [2.16] puede llevarse a cabo fácilmente a través del filtro de Kalman. El filtro de Kalman es un algoritmo recursivo que permite obtener la estimación óptima (bajo unos supuestos básicos) del vector de estado (de la variable de interés) en el período t utilizando toda la información disponible hasta el período $t-1$ y actualizando dicha estimación cuando se dispone de información adicional sobre la(s) variable(s) observables.

El filtro de Kalman, propuesto originalmente por Kalman (1960) y Kalman y Bucy (1961), no es más que un conjunto de ecuaciones que se aplican secuencialmente. En una primera etapa se obtiene la predicción (estimación) óptima del vector de estado para el período (t) usando para ello toda la información disponible (hasta el período $t-1$). A continuación, en una segunda etapa, se actualiza dicha predicción incorporando la información disponible para el período t . Así pues, en esta segunda etapa se mejora la estimación obtenida en la etapa anterior.

En concreto²⁴, dado el modelo *state-space* presentado (formado por) las ecuaciones [2.10] a [2.14], suponiendo que a_{t-1} es el predictor (estimador) óptimo de \mathbf{a}_{t-1} obtenido a partir de la información (observaciones) de Y hasta el período $t-1$ incluido, y que la matriz de varianzas y covarianzas de orden $m \times m$ del error de estimación, P_{t-1} , asociada a dicho estimador viene dada por:

$$P_{t-1} = E[(\alpha_{t-1} - a_{t-1}) \cdot (\alpha_{t-1} - a_{t-1})'], \quad [2.17]$$

una vez conocidos a_{t-1} y P_{t-1} , puede obtenerse el estimador óptimo del vector de estado, \mathbf{a}_t , condicionado a estos valores de la forma:

$$a_{t/t-1} = \alpha_t / a_{t-1} = E(\alpha_t / a_{t-1}) = E(T_t a_{t-1} + c_t + R_t \theta_t), \quad [2.18]$$

y dado que $E(\theta_t) = 0$, se tiene que $a_{t/t-1}$, viene dado por:

$$a_{t/t-1} = T_t a_{t-1} + c_t, \quad [2.18bis]$$

siendo la matriz de varianzas y covarianzas asociada al error de estimación de [2.18bis]:

$$P_{t/t-1} = E[(\alpha_t - a_{t/t-1}) \cdot (\alpha_t - a_{t/t-1})'] = T_t P_{t-1} T_t' + R_t Q_t R_t'. \quad [2.19]$$

Las ecuaciones [2.18bis] y [2.19], conocidas como ecuaciones de predicción, permiten obtener el predictor óptimo del vector de estado para el

²⁴ En la derivación que a continuación se presenta de las ecuaciones del filtro de Kalman se trabaja con las hipótesis de que los hiperparámetros y los valores iniciales del vector de estado y de la matriz de varianzas y covarianzas del error de estimación son conocidas para todos los períodos y, en consecuencia, no es necesario incluirlas entre los parámetros a estimar.

período t con toda la información disponible hasta ese instante de tiempo ($t-1$). Son, por tanto, las ecuaciones correspondientes a la primera de las etapas a las que se ha hecho referencia anteriormente.

Una vez se dispone de la (nueva) observación de Y para el período t , Y_t , se incorpora al conjunto de información disponible hasta entonces ($0,1,\dots,t-1$) y se actualiza el estimador (óptimo) de \mathbf{a}_t , $a_{t/t}$, obtenido en la etapa anterior. Las ecuaciones que permiten llevar a cabo esta segunda etapa, conocidas como ecuaciones de actualización, son las dos siguientes:

$$a_t = a_{t/t} = a_{t/t-1} + P_{t/t-1} Z_t' F_t^{-1} (Y_t - Z_t a_{t/t-1} - d_t), \text{ y} \quad [2.20]$$

$$P_t = P_{t/t-1} - P_{t/t-1} Z_t' F_t^{-1} Z_t P_{t/t-1}, \quad [2.21]$$

donde $F_t = Z_t P_{t/t-1} Z_t' + H_t$ es la matriz de varianzas y covarianzas del vector de innovaciones, v_t , (vector que recoge la nueva información que se incorpora en el estimador, $Y_t - Z_t a_{t/t-1} - d_t$).

Pues bien, el filtro de Kalman está formado por las ecuaciones de predicción (las de la primera etapa) y las de actualización (las de la segunda), es decir, el conjunto de ecuaciones [2.18bis] a [2.21]^{25 y 26}.

²⁵ La formulación presentada es la más general posible. En este sentido, existen diferentes maneras de interpretar el algoritmo del filtro dependiendo de cómo se deriven las ecuaciones de predicción y actualización: a partir del supuesto de normalidad (Harvey, 1989; Tanizaki, 1996), por proyección ortogonal (Brockwell y Davis, 1991), a partir del estimador Theil-Golberger (Cooley, 1977; Harvey, 1989; Cuthbertson *et al.*, 1992; Tanizaki, 1996), a partir del estimador de mínimos cuadrados generalizados (Sant, 1977), ...

²⁶ En cualquier caso, cabe señalar que existe la posibilidad de pasar directamente de $a_{t/t-1}$ a $a_{t+1/t}$ mediante la expresión siguiente:

$$a_{t+1/t} = (T_{t+1} + \vartheta_t Z_t) a_{t/t-1} + \vartheta_t Y_t + (c_{t+1} - \vartheta_t d_t),$$

donde la matriz ϑ_t , conocida como matriz de ganancia, viene dada por:

$$\vartheta_t = T_{t+1} P_{t/t-1} Z_t' F_t^{-1} \quad t = 1, 2, \dots, T,$$

y la recursión para la matriz de varianzas y covarianzas es la llamada ecuación de Riccati:

Sin embargo, para poder aplicar las ecuaciones del filtro de Kalman es necesario conocer los hiperparámetros del sistema (en el marco de este trabajo, los parámetros γ , ϕ , θ , μ , δ y las varianzas de los términos de perturbación de las ecuaciones [2.15] y [2.16]) así como los valores iniciales del vector de estado a_0 (o, $a_{1,0}$) y su matriz de varianzas y covarianzas P_0 (o, $P_{1,0}$). En el supuesto que no lo sean, cosa habitual en la práctica en la mayor parte de aplicaciones en el entorno de la Economía (y en el caso tratado en particular también), se tienen que estimar previamente.

2.4.2. Hiperparámetros y valores iniciales

El método de estimación que permite obtener la mejor aproximación posible a los valores (desconocidos) de los hiperparámetros es el de la máxima verosimilitud. La aplicación de este método de estimación consiste únicamente en resolver el sistema de ecuaciones que se obtiene igualando a cero las derivadas parciales (del logaritmo) de la función de verosimilitud respecto a cada uno de los (hiper)parámetros desconocidos (condición necesaria). Para garantizar que las soluciones encontradas cumplen además la condición suficiente, la matriz de segundas derivadas (del logaritmo) de la función de verosimilitud ha de ser definida negativa. Dicha función, bajo la hipótesis de distribución normal de los términos de perturbación de las ecuaciones de medida y de transición, para el caso de un modelo *state-space* viene dada por la siguiente expresión:

$$\log L(Y; \lambda) = -\frac{T-M}{2} \log 2\pi - \frac{T-M}{2} \log H_t - \frac{1}{2} \sum_{t=1}^T |F_t| - \frac{1}{2} \sum_{t=1}^T \mathbf{v}_t' F_t^{-1} \mathbf{v}_t, \quad [2.22]$$

$$P_{t+1/t} = T_{t+1} \left(P_{t/t-1} - P_{t/t-1} Z_t' F_t^{-1} P_{t/t-1} \right) T_{t+1}' + R_{t+1} Q_{t+1} R_{t+1}' \quad t = 1, 2, \dots, T.$$

donde λ representa el conjunto de hiperparámetros desconocidos.

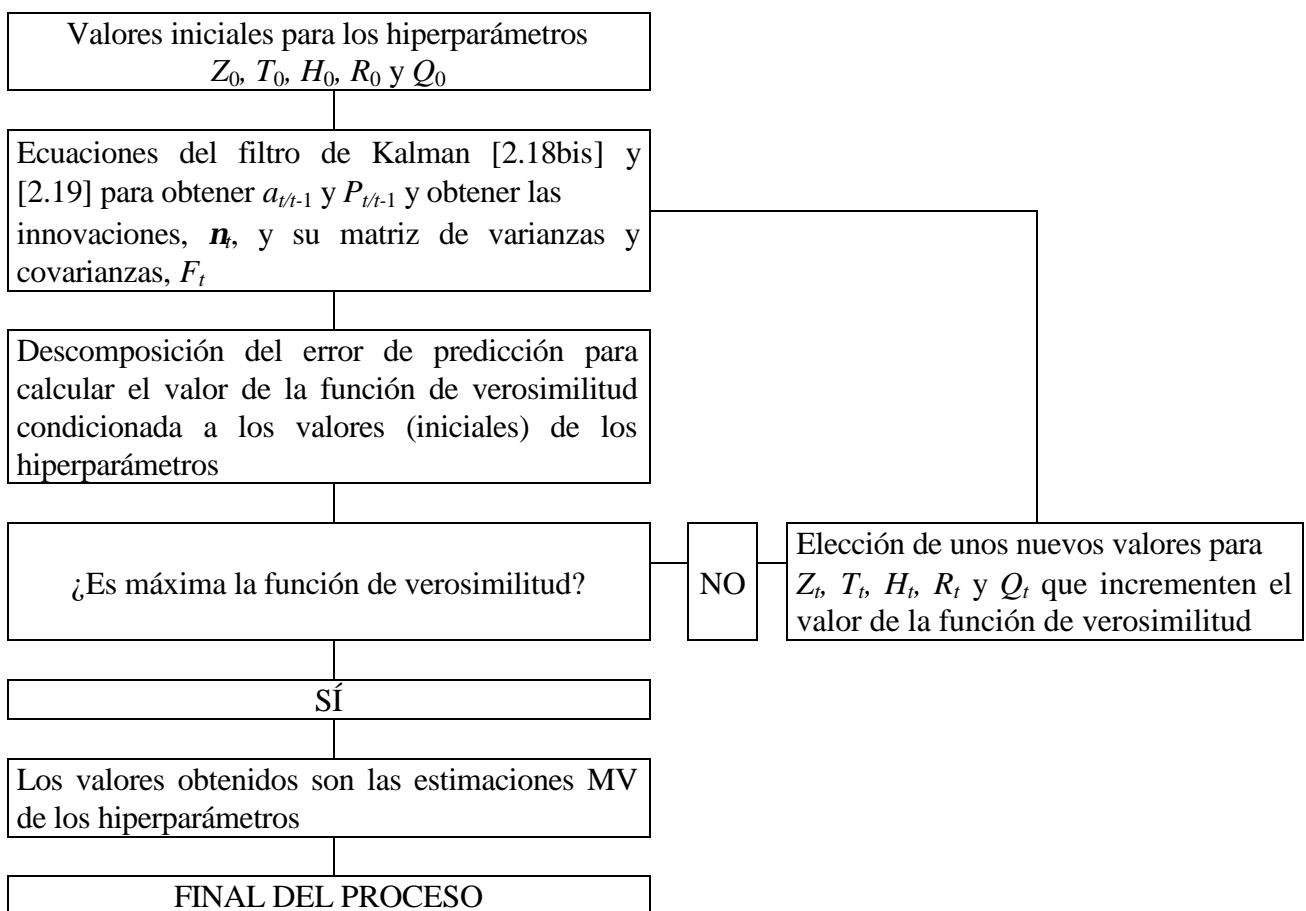
Ahora bien, el hecho que la expresión de la función de verosimilitud sea (generalmente) demasiado complicada como para obtener los valores deseados a partir de su expresión analítica (es decir, que la resolución del sistema de ecuaciones que se deriva de la condición necesaria de optimalidad es como mínimo muy complicada²⁷) hace que se acostumbren a utilizar las expresiones numéricas de la función de verosimilitud conjuntamente con procedimientos de optimización numérica para encontrar el valor en el que alcanza el máximo. Así pues, brevemente, el procedimiento de estimación máximo verosímil de los hiperparámetros de un modelo *state-space* se puede resumir en las siguientes tres etapas (véase el cuadro 2):

- a) en una primera etapa se estudia si el modelo está identificado, se determinan cuáles son los hiperparámetros desconocidos que hay que estimar (Z_t , T_t , H_t , R_t y Q_t) y se fijan unos valores iniciales para dichos parámetros;
- b) a partir de estos valores iniciales y mediante las ecuaciones del filtro del Kalman se obtienen los valores de las innovaciones, v_t , y de su matriz de varianzas y covarianzas, F_t , (condicionados a los valores asignados a los hiperparámetros en la etapa anterior); y,
- c) se obtiene un valor para la función de verosimilitud (a partir de la descomposición del error de predicción propuesta por Harvey, 1984b y 1990- pp. 106-111-) y, utilizando algún procedimiento de optimización

²⁷ De hecho, como señala Harvey (1990, pág. 88), la solución directa de las ecuaciones de verosimilitud (las resultantes de igualar la primera derivada del logaritmo de la función de verosimilitud respecto de los hiperparámetros desconocidos a cero) es más bien una excepción y no la regla general.

numérica, se evalúa si este valor es máximo o no lo es. En el primer caso finaliza el proceso y, en el segundo, el procedimiento de optimización facilita unos nuevos valores (que se toman como iniciales) para los hiperparámetros que incrementan el valor de (el logaritmo de) la función de verosimilitud y se empieza de nuevo el proceso a partir de la etapa anterior.

Cuadro 2. Procedimiento de estimación máximo-verosímil de los hiperparámetros de un modelo *state-space*



Fuente: Elaboración propia a partir de Cuthbertson *et al.* (1992, pág. 214).

El principal inconveniente que presenta este procedimiento recursivo de estimación es su elevada sensibilidad respecto del procedimiento de optimización numérica y del período muestral utilizados. Efectivamente, dependiendo del

procedimiento de optimización numérica y/o del período muestral las estimaciones obtenidas para los hiperparámetros pueden variar significativamente.

Un procedimiento alternativo para estimar los hiperparámetros de un modelo *state-space* consiste en utilizar el denominado algoritmo *EM* desarrollado inicialmente por Dempster *et al.* (1977) e introducido en el ámbito de este tipo de modelos por Shumway y Stoffer (1982) y Watson y Engle (1983). Muy brevemente, este procedimiento consta de dos etapas. En la primera, llamada etapa de expectativas, se calcula la esperanza de la función de verosimilitud condicionada a los valores observados de las variables y a unos valores iniciales (cualesquiera) de los hiperparámetros mediante la descomposición del error de predicción en combinación con un algoritmo de alisado (generalmente el alisado de intervalo fijo²⁸). A continuación, en una segunda etapa (denominada maximización) se maximiza la función obtenida en la etapa anterior respecto a los parámetros desconocidos.

Una de las principales críticas a la aplicación de los modelos *state-space* en campos como la Economía donde los valores de los hiperparámetros no se conocen *a priori*, se centra en la inestabilidad de ambos procedimientos para estimar estos hiperparámetros. En este sentido, Hackl y Westlund (1996) muestran que los resultados del filtro de Kalman son muy sensibles a la especificación del modelo *state-space*. Su conclusión es que para protegerse de este problema en la mayor parte de casos la solución consiste en especificar un modelo tan simple como sea posible y estimar *a priori* los hiperparámetros en lugar de hacerlo conjuntamente con el vector de estado.

²⁸ Véase Anderson y Moore (1979, pp. 187-190) o Jazwinski (1970, pp. 216-218) para un detalle sobre este algoritmo de alisado.

Por otra parte, para solventar el problema de la inicialización del filtro de Kalman existen dos tipos de estrategias dependiendo de si el modelo *state-space* especificado es o no es estacionario²⁹.

Así, en el supuesto de estacionariedad, los valores iniciales del vector de estado pueden aproximarse a partir de la media y la varianza incondicionadas del propio proceso. En concreto, en la literatura pueden encontrarse, entre otras las siguientes propuestas: *a*) estimar la ecuación de medida [2.10] por *MQO* utilizando las primeras *m* observaciones disponibles y utilizar la media y la varianza de las estimaciones obtenidas como valores iniciales del filtro en el período *m+1* (Harvey, 1984); y, *b*) considerar los valores iniciales como hiperparámetros desconocidos y estimarlos por cualquiera de los dos procedimientos explicados anteriormente³⁰.

Sin embargo, si el modelo es no estacionario las condiciones iniciales no están bien definidas por lo que las propuestas señaladas en el párrafo anterior no pueden aplicarse. En este caso, la solución más habitual en la práctica consiste en considerar las condiciones iniciales como difusas, introduciendo una serie de ecuaciones complementarias en el filtro de Kalman “usual”. En la literatura se pueden encontrar diferentes propuestas sobre cómo introducir estas ecuaciones complementarias (Harvey y Phillips, 1979; Anderson y Moore, 1979-Kitagawa y Gersch, 1984; Harvey 1981 y 1989; Ansley y Kohn, 1989-de Jong, 1991-Kohn y Ansley, 1986; ...), sin embargo, ninguna de ellas soluciona de forma completamente satisfactoria el problema³¹.

²⁹ Se dice que un modelo *state-space* es estacionario si: *a*) los valores propios de la matriz de transición de la ecuación de estado, T_t , están dentro del círculo unidad, y *b*) se dispone de suficientes observaciones como para poder afirmar que el modelo ha alcanzado la estacionariedad.

³⁰ En cualquier caso, esta opción sólo es posible llevarla a cabo si las condiciones iniciales son deterministas (Rosenberg, 1973).

³¹ En este sentido cabe señalar que muchas de las investigaciones que actualmente se están llevando a cabo relacionadas con el filtro de Kalman están centradas en solucionar el problema de la inicialización del filtro cuando el modelo es no estacionario.

2.4.3. Información estadística disponible

Para el caso del modelo propuesto para las regiones españolas consideradas (País Vasco, Asturias y Andalucía)³², la información referente a las horas de trabajo directo (efectivo) en la producción (*proxy* del trabajo) no está disponible con periodicidad mensual, sino que sólo lo está trimestralmente³³. En consecuencia se ha tenido que aproximar el *input* trabajo por otra variable que sí estuviese disponible para todas las CC.AA. con una periodicidad mensual. Esta variable es el número de afiliados al (régimen general) de la Seguridad Social que ha sido facilitada por la Subdirección General de Estadísticas Sociales y Laborales del Ministerio de Trabajo y Asuntos Sociales.

Por otro lado, el consumo de energía eléctrica para usos industriales (*proxy* del capital) con una periodicidad mensual para las tres regiones consideradas únicamente está disponible desde enero de 1993 (en adelante)³⁴ y ha sido facilitado por IBERDROLA (en el caso del País Vasco), el SADEI (en el de Asturias) y el IEA (en el de Andalucía)³⁵.

³² No se ha considerado la región de Extremadura a pesar de que dispone de un índice de producción industrial elaborado a partir del método directo, puesto que en el momento de llevar a cabo el presente estudio no se disponía de los datos referentes al IPI de esta región.

³³ De hecho, ello es así no sólo para estas tres comunidades sino para todas las CC.AA. españolas.

³⁴ En este trabajo hasta diciembre de 1996.

³⁵ Hay que señalar, sin embargo, que en el caso de la serie de consumo de energía eléctrica para usos industriales del País Vasco se observa un comportamiento claramente atípico en los meses de mayo y junio de 1995. Por tanto, la consideración de esta serie como *input* en el modelo comportaría graves errores derivados de la atipicidad de la información disponible para estos dos meses. Así pues se ha corregido la serie mediante un tratamiento clásico univariante, consistente en predecir el valor para dichos dos meses de acuerdo con el patrón de comportamiento histórico de la serie. Es por tanto esta serie corregida y no la facilitada por IBERDROLA la que se ha utilizado como *input* a la hora de elaborar el indicador indirecto de la actividad industrial para el País Vasco.

2.4.4. Estimación del modelo y validación de los resultados³⁶

Tal y como se ha señalado en el subapartado anterior, se dispone de datos para los dos *inputs* regionales para el período 1993-96, con lo cual únicamente se dispone de treinta y seis observaciones (una vez diferenciados los datos). Esta circunstancia supone una restricción insalvable que hace que la estimación de los hiperparámetros no sea adecuado llevarla a cabo por máxima verosimilitud. Para solucionar este problema, de acuerdo con la propuesta de Hackl y Westlund (1996), se ha utilizado información *a priori* (datos de panel) para estimar los valores de los hiperparámetros en las tres regiones consideradas.

En concreto, los parámetros correspondientes a la ecuación de estado (γ , ϕ y θ) han sido estimados a partir de una función de producción de tipo Cobb-Douglas para el período 1964-91³⁷ considerando efectos fijos (el estadístico del test de Hausman para contrastar la hipótesis de efectos fijos contra efectos aleatorios es 38,89). Los resultados de esta estimación así como los efectos fijos estimados para cada una de las tres comunidades consideradas se presentan en la tabla 1.

En ella se puede apreciar la elevada significación estadística de los *inputs* energía (capital) y afiliados a la Seguridad Social (trabajo). Además, los resultados obtenidos por lo que se refiere a los valores estimados para ambos *inputs* son consistentes con los obtenidos en otros trabajos: aproximadamente suman uno (rendimientos constantes a escala), y la participación del trabajo en

³⁶ Los resultados referidos a la estimación del modelo *state-space* presentados en este apartado han sido obtenidos mediante el paquete SAS versión 6.12. Agradecemos a Philip Israilevich que nos facilitase parte de las codificaciones en Gauss que utilizó en su trabajo.

³⁷ La información estadística utilizada para llevar a cabo esta estimación proviene, por lo que se refiere al VAB (a coste de factores) y a los ocupados, de las series bianuales del BBV, y en cuanto al *stock* de capital del IVIE (BBV). En total, se dispone de doscientas treinta y ocho observaciones (diecisiete CC.AA. por catorce años).

el *output* es aproximadamente una tercera parte y la del trabajo dos terceras partes.

Por su parte, los parámetros de la ecuación de medida (μ y δ) han sido estimados a partir de la participación en el total del conjunto del Estado en términos de VAB. Por último, las varianzas de los términos de perturbación de las dos ecuaciones se han estimado por máxima verosimilitud.

Tabla 1. Estimación de la función de producción y efectos fijos de las comunidades consideradas

Variable	Coefficient	Std. Error	t-ratio	Prob.	Mean & S.D. of Var.
LIN	0.69499	0.2631E-01	26.420	0.0000	11.6987 0.9467
KPIND	0.38304	0.2342E-01	16.355	0.0000	13.2882 0.9955
Constant	-0.61157	0.2404	-2.544	0.0110	
Estimated fixed effects:					
Individual					
	Andalucía		-0.21167		
	Asturias		0.11637		
	País Vasco		-0.16833		

LIN y KPIN simbolizan los ocupados y el capital industriales.

En la tabla 2 se presentan los resultados de las estimaciones de los hiperparámetros para cada una de las tres comunidades consideradas.

Tabla 2. Estimaciones *a priori* de los hiperparámetros del modelo

	País Vasco	Asturias	Andalucía	
γ	-0,77	-0,49	-0,82	$X_{jt} = \mathbf{g} + \mathbf{f}K_{jt} + \mathbf{q}L_{jt}$ $R^2 = 0,92$
ϕ	0,38	0,38	0,38	
θ	0,69	0,69	0,69	
μ	4,01	4,85	1,58	$VAB_{Estado} = \mathbf{m} + \mathbf{c}VAB_j$ $R^2_{P.Vasco} = 0,72; R^2_{Asturias} = 0,72; R^2_{Andalucía} = 0,81$
δ	0,65	0,68	0,85	

Por último, en cuanto a la cuestión de los valores para inicializar el filtro de Kalman, se ha solucionado mediante la aproximación propuesta por Harvey (1981 -pág. 113- y 1989 -pág. 134-) apuntada en el apartado 2.4.2 y que, brevemente, consiste en calcular una distribución *a priori* para los valores iniciales del vector de estado a partir de las primeras h observaciones. Las

recursiones comienzan para $t=0$ con $\alpha_0=0$ y $P_0=\kappa \cdot I$, donde $\kappa \rightarrow \infty$ (aunque en la práctica se suele utilizar un número positivo muy grande, por ejemplo, 10^6 -que ha sido el utilizado en este trabajo-) y, los valores del vector de estado obtenidos para la observación h son los que se utilizan como valores iniciales del filtro para $t>h$ ³⁸.

Una vez estimados los hiperparámetros y solventado el problema de la inicialización del filtro de Kalman, es inmediato (una vez programado) obtener la estimación de los índices de producción industrial regionales. Así, los resultados obtenidos en términos trimestrales y anuales (los primeros en niveles y tasas de crecimiento y los segundos únicamente en niveles) se presentan en los gráficos 1 a 3.

Tal y como se observa en dichos gráficos, los resultados obtenidos proporcionan una buena aproximación a la evolución de los índices elaborados por métodos directos para las comunidades del País Vasco, Asturias y Andalucía.

³⁸ Nótese que de esta manera se utiliza una distribución impropia como punto de partida para el proceso de estimación. La principal ventaja de esta propuesta radica en su sencillez, pero presenta el inconveniente que los resultados que se obtienen no siempre son estables desde un punto de vista numérico.

Gráfico 1. Evolución del Índice de Producción Industrial del País Vasco (índice trimestral, tasa de crecimiento e índice anual)

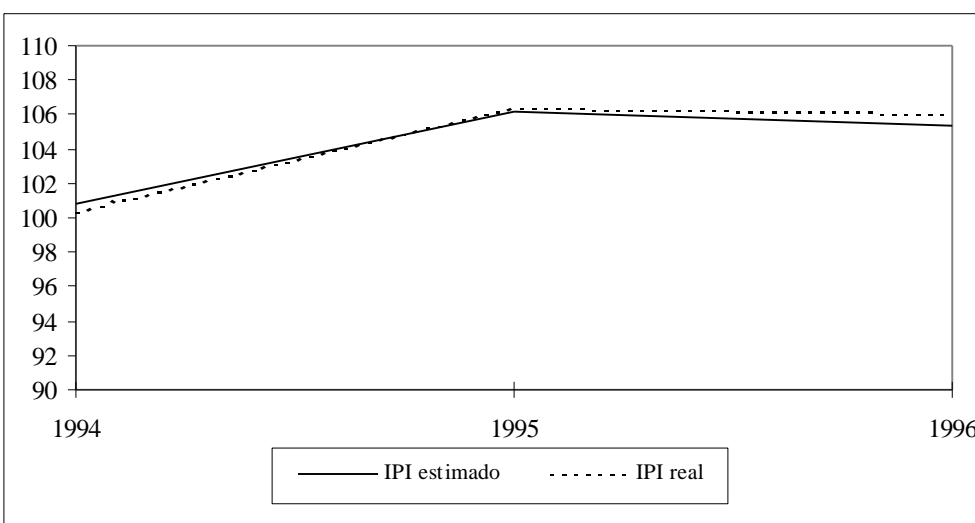
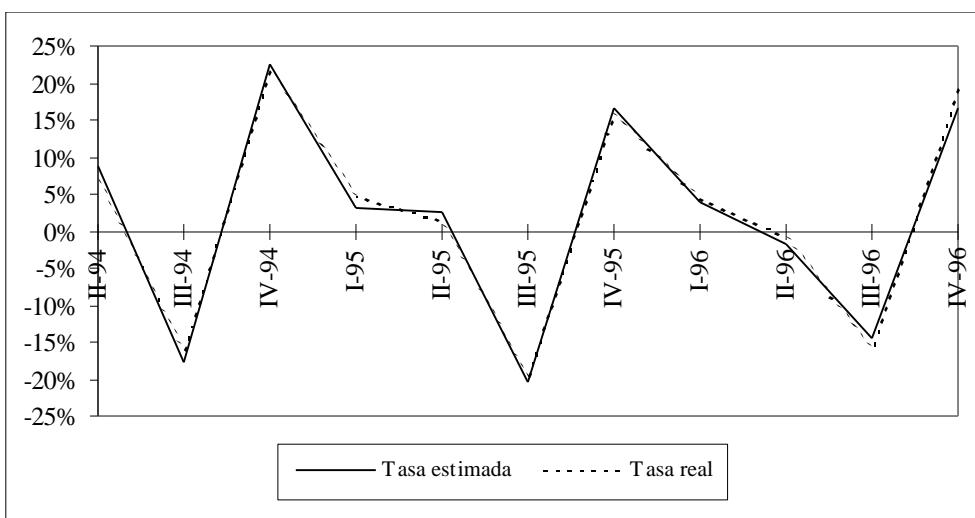
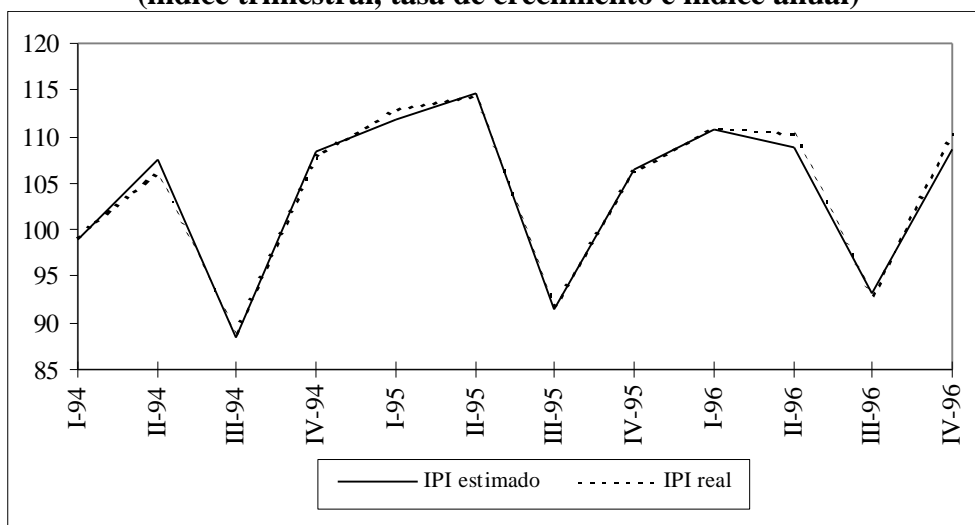


Gráfico 2. Evolución del Índice de Producción Industrial de Asturias (índice trimestral, tasa de crecimiento e índice anual)

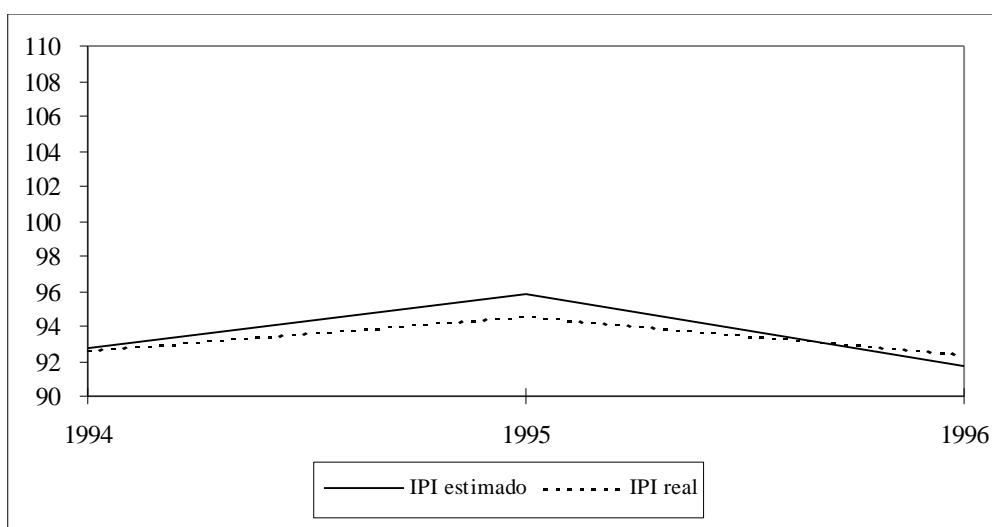
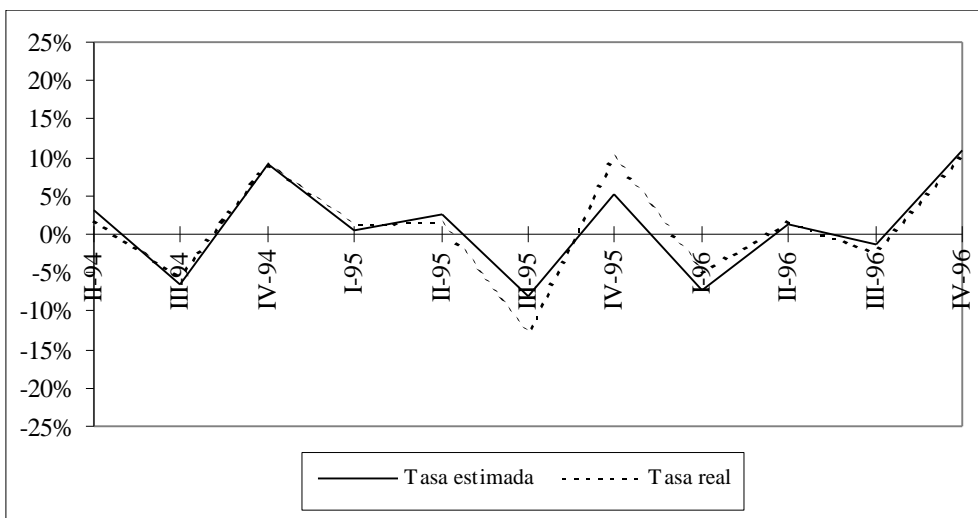
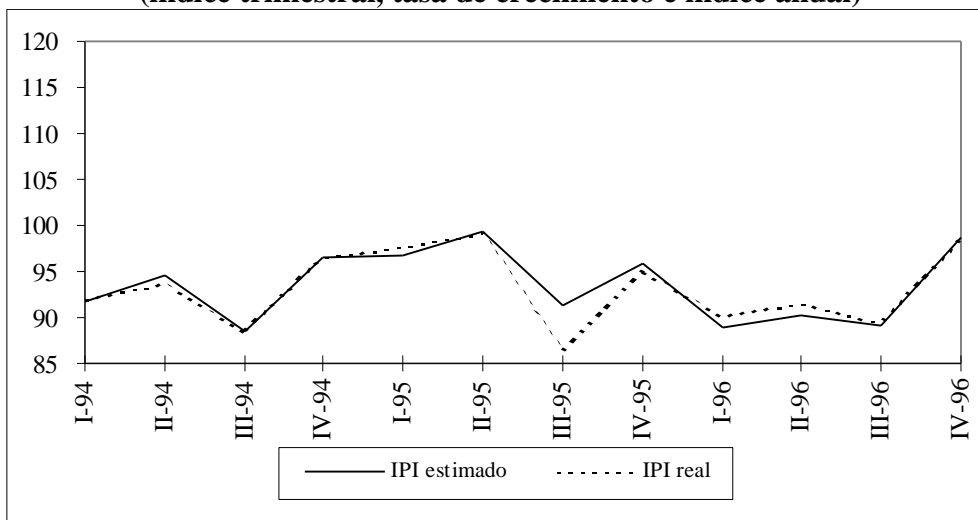
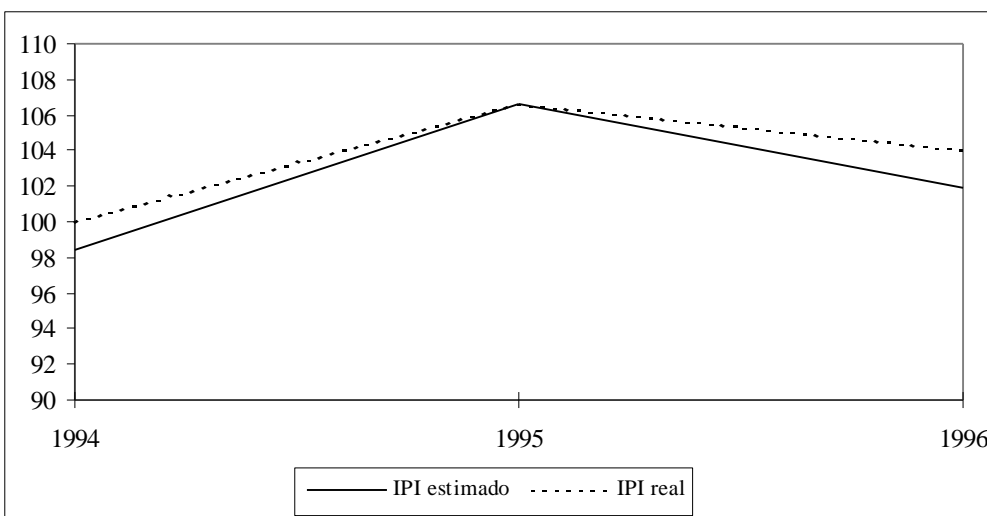
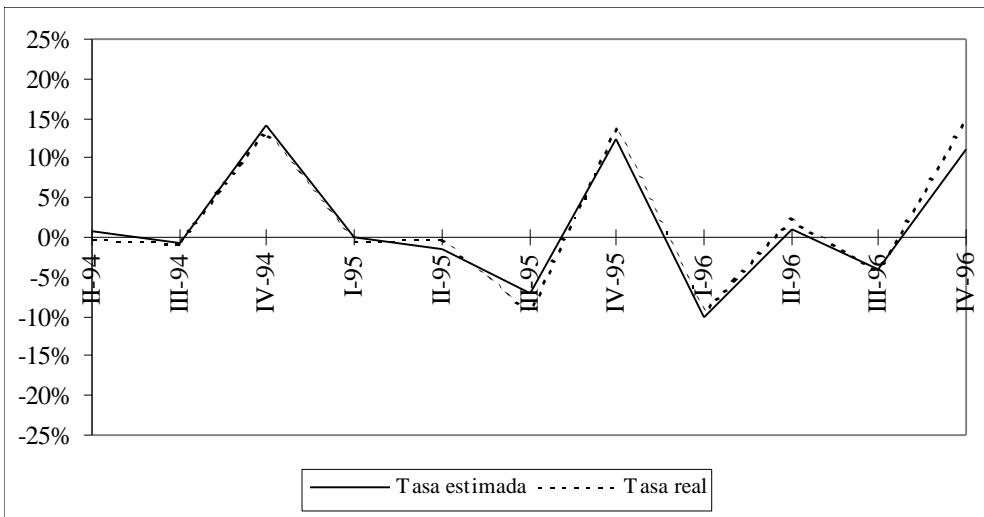
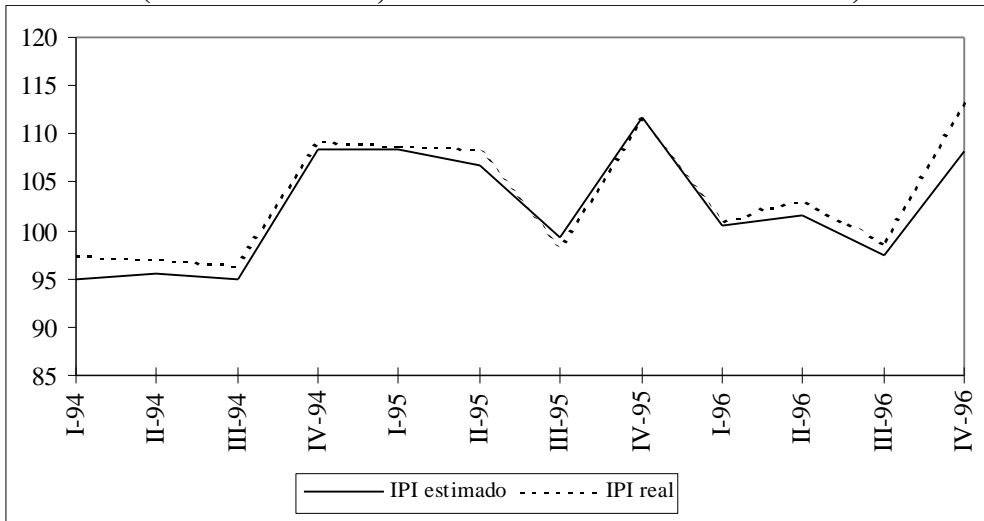


Gráfico 3. Evolución del Índice de Producción Industrial de Andalucía (índice trimestral, tasa de crecimiento e índice anual)



Como elemento adicional de validación se ha calculado el *EPAM*³⁹ entre ambas series así como el coeficiente *pseudo-R*² propuesto por Israilevich y Kuttner (1993) para conocer cuánto de informativas son las fluctuaciones nacionales para las fluctuaciones regionales. Los resultados obtenidos (véase tabla 3) muestran que, efectivamente, el ajuste obtenido es satisfactorio, confirmando, en consecuencia la conclusión que se derivaba del análisis gráfico realizado anteriormente.

Lo anterior pone de manifiesto que, a diferencia del método utilizado por el IEC que es válido sólo bajo ciertas hipótesis (entre ellas el parecido entre la estructura industrial de la región y de la nación, el peso que la industria de la región tenga en la del conjunto del Estado, ...), el método que aquí se presenta es más general e independiente de las limitaciones sobre la estructura productiva señaladas. Un ejemplo se tiene con las tres CC.AA. para las que se ha desarrollado el estudio, que tienen estructuras industriales y pesos muy diferentes entre sí. A pesar de ello, en todas ellas los resultados obtenidos son satisfactorios.

Tabla 3. Valores del *pseudo-R*² y *EPAMs* asociados al modelo de variables latentes. Período 1994-96

	<i>pseudo-R</i> ²	<i>EPAM</i>		
		Mensual	Trimestral	Anual
País Vasco	0,50	6,34%	0,66%	0,45%
Asturias	0,48	3,10%	1,09%	0,77%
Andalucía	0,66	4,89%	1,48%	1,24%

³⁹ El *EPAM* (Error Porcentual Absoluto Medio) se define, como es sabido, de la forma:

$$EPAM = \frac{1}{T} \sum_{t=1}^T \frac{|Y_t - y_t|}{Y_t} 100,$$

donde Y_t son los índices elaborados por el EUSTAT, el SADEI y el IEA, e y_t son los estimados a partir del modelo propuesto.

Adicionalmente, con el fin de analizar si efectivamente el modelo de variables latentes propuesto proporciona unos índices más fiables (en el sentido que se aproximan más a los índices publicados) que los que se obtienen a partir de otros métodos indirectos que actualmente se aplican a nivel regional en nuestro país, se han elaborado unos índices para el mismo período que el considerado en este trabajo (1994-96) de acuerdo con la metodología que utiliza el IEC para elaborar el indicador de la actividad industrial de la comunidad catalana y se han comparado los resultados obtenidos en términos de *EPAMs* (en la tabla 4 se presentan los *EPAMs* asociados a los indicadores elaborados según la metodología del IEC para el período señalado). La conclusión que se deriva de este análisis es que el modelo de variables latentes propuesto proporciona mejores resultados para las tres comunidades tanto a nivel anual como trimestral y mensual⁴⁰.

Tabla 4. *EPAMs* asociados a los indicadores calculados siguiendo la metodología del IEC. Período 1994-96

	Mensual	Trimestral	Anual
País Vasco	5,74%	1,48%	1,14%
Asturias	5,04%	3,43%	1,84%
Andalucía	8,52%	7,16%	4,94%

En cualquier caso, sin embargo, a pesar de los satisfactorios resultados obtenidos es necesario hacer un comentario adicional: los *EPAMs* obtenidos a partir del modelo de variables latentes en términos mensuales son ligeramente elevados (principalmente en el caso de Andalucía y sobre todo en el del País Vasco). Este hecho puede explicarse por tres factores:

- a) porque el número de horas de trabajo efectivo no está disponible a nivel regional con periodicidad mensual. Ello ha obligado a aproximar el *input*

⁴⁰ Excepto para el País Vasco a nivel mensual.

trabajo por el número de afiliados a la Seguridad Social. Esta variable, si bien permite superar los problemas que plantea las horas de trabajo dado que está disponible para todas las CC.AA. con periodicidad mensual, presenta el inconveniente que no recoge las fluctuaciones estacionales que se producen en la producción industrial como consecuencia de los períodos vacacionales.

En este sentido, una posible solución para obtener mejores resultados consistiría en corregir la serie de afiliados a la Seguridad Social para que recogiera este efecto estacional. Para ello una opción sería considerar las preferencias de los trabajadores a la hora de hacer sus vacaciones. Una segunda posibilidad pasaría por mensualizar la serie de horas de trabajo efectivo.

Respecto a la primera de las posibles soluciones apuntadas, corregir la serie de afiliados en función de las preferencias de los trabajadores en lo que a las vacaciones se refiere para, de esta forma, inducirle la estacionalidad que se observa en la producción industrial en los períodos vacacionales, dada la falta de información estadística existente a nivel regional en nuestro país actualmente sobre este punto, no es posible llevarla a cabo. De hecho, únicamente se ha encontrado información estadística referente al comportamiento de los españoles frente a las vacaciones para el año 1995⁴¹. Sin embargo, esta información no se ofrece desagregada por meses y, en cualquier caso, aunque lo fuese, utilizarla significaría suponer que el comportamiento vacacional de los españoles es el mismo en todas las comunidades, lo cual no tiene porque ser cierto, de manera que, bajo este supuesto, considerarla podría introducir un sesgo: todas las CC.AA. presentarían la misma estacionalidad en el *input* trabajo.

En cuanto a la mensuralización de la serie trimestral de horas de trabajo efectivas, la falta de indicadores mensuales adecuados a nivel regional hace que no sea posible llevarla a cabo;

- b) el reducido número de observaciones disponibles repercute negativamente en el comportamiento de la metodología en la que se sustenta la especificación y estimación del modelo propuesto. En este sentido, podrían obtenerse mejores resultados al incorporar en la muestra nuevas observaciones cuando éstas estén disponibles; y,
- c) en el período considerado la economía española experimentó una corta pero profunda crisis seguida del inicio de una etapa expansiva (en la que todavía hoy nos encontramos); ello no hace sino añadir un grado adicional de dificultad en la estimación de este tipo de índices. En este sentido, en un trabajo anterior (Clar, *et al.*, 1998) se elaboraron unos indicadores para el seguimiento de la actividad industrial para las tres comunidades consideradas de acuerdo con la metodología del IEC, para un período más amplio del que ha podido ser considerado en este trabajo. Los *EPAMs* entre las series de dichos indicadores y los publicados (véase tabla 5) son inferiores a los que se obtienen aplicando la misma metodología para el período 1994-96 (tabla 4), ello muestra la especial dificultad asociada al período considerado en este trabajo.

Tabla 5. *EPAMs* a partir de los indicadores elaborados siguiendo la metodología del IEC y los publicados

Indicador	País Vasco		Asturias		Andalucía	
	Período	<i>EPAM</i>	Período	<i>EPAM</i>	Período	<i>EPAM</i>

⁴¹ Centro de Investigaciones Sociológicas: Comportamiento de los Españoles ante las Vacaciones, Opiniones y Actitudes, 11.

Mensual	1986/01-1996/12	6,83%	1990/01-1996/12	4,32%	1986/01-1996/12	5,67%
Trimestral	1986/I-1996/IV	2,54%	1990/I-1996/IV	2,33%	1986/I-1996/IV	4,63%
Anual	1986-1996	0,67%	1990-1996	1,29%	1986-1996	3,36%

3. CONCLUSIONES

Durante los últimos años diversas instituciones públicas y privadas regionales vienen publicando unos índices para el seguimiento de la actividad industrial de sus regiones. El principal problema de estos índices es que no son directamente comparables puesto que no se elaboran siguiendo una metodología común.

En este trabajo se ha estudiado la potencialidad de la modelización *state-space* y el filtro de Kalman para el análisis de series económicas en el marco de la Economía Regional en general y, en particular, para la elaboración de un indicador (indirecto) cuantitativo de la actividad industrial (regional). En concreto, se ha aplicado esta metodología a tres de las cuatro regiones españolas que actualmente disponen de un IPI elaborado a partir directamente de información sobre su producción industrial obtenida a partir de una encuesta a una muestra representativa de establecimientos y productos industriales (método directo).

La principal conclusión que se deriva del análisis realizado es que la estrategia planteada permite aproximar suficientemente bien la evolución de la producción industrial regional. Este resultado hace pensar que la aplicación de la técnica considerada en el campo de la Economía Regional puede abrir un camino para la solución de problemas que con otras metodologías (más habituales, clásicas) resultan difíciles de alcanzar.

Sin embargo, hay diversos problemas de tipo práctico que aparecen a la hora de aplicar esta metodología dadas las limitaciones en la disponibilidad de información estadística regional con la que nos encontramos hoy en día en

nuestro país. Ello pone de manifiesto que es necesario seguir trabajando no sólo en la mejora de la calidad de la información estadística existente actualmente, sino también en la elaboración de nuevas estadísticas regionales.

4. REFERENCIAS BIBLIOGRÁFICAS

Anderson, B.D. y Moore, J.B. (1979): *Optimal Filtering*, Prentice-Hall, Englewoods Cliffs.

Ansley, C.F. 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.

Aoki, M. (1990): *State Space Modelling of Time Series*, Springer-Verlag, Berlin.

Artís, M., Pons, J., Sierra, M.A. y Suriñach, J. (1994): “Elaboració d’un Sistema d’Indicadors d’Activitat per a l’Economia Catalana”, *Perspectiva Econòmica de Catalunya*, 176, pp. 83-102.

Artís, M., Pons, J., Sierra, M.A. y Suriñach, J. (1997a): “Estimación de la Actividad Económica a Corto Plazo Mediante Indicadores de Coyuntura”, *Revista de Economía Aplicada*, 13, pp. 129-147.

Artís, M., Pons, E., Pons, J. y Suriñach, J. (1997b): *Evolución Cíclica de las Comunidades Autónomas y Análisis Cíclico*, Escuela de Economía Regional, Universidad Internacional Menéndez Pelayo, Santander.

Artís, M., Pons, E., Pons, J. y Suriñach, J. (1997c): “Comptabilitat Econòmica de Catalunya i Mètodes de Trimestralització. Components de la Demanda”, *Document de Treball 97R02*, Grup d’Anàlisi Quantitativa Regional, UB.

Brockwell, P. y Davis, R. (1991): *Time Series: Theory and Methods*, Springer-Verlag, Berlin, 2ª edición.

Clar, M., Ramos, R. y Suriñach, J. (1997a): “A Methodological Proposal for Elaborating Regional Manufacturing Production Indices”, Comunicación presentada al 37th. Congress of the European Regional Science Association, Roma, 26-29 de agosto.

Clar, M., Ramos, R. y Suriñach, J. (1997b): “Análisis Comparativo de Métodos Indirectos para la Elaboración de un Indicador de la Actividad Industrial Regional”, Comunicación presentada a la XXIII Reunión de Estudios Regionales, Valencia, 18-21 de noviembre.

Clar, M., Ramos, R. y Suriñach, J. (1998): “Algunes Reflexions sobre la Construcció d’Indicadors Indirectes pel Seguiment de l’Activitat Industrial Regional”, *Documento de Trabajo de la División II de la UB* (en prensa). También *Document de Treball 98R15*, Grup d’Anàlisi Quantitativa Regional, UB.

Cooley, T.F. (1977): “Generalized Least Squares Applied to Time Varying Parameters Model: A Comment”, *Annals of Economic and Social Measurement*, 6, pp. 313-314.

Costa, A. y Galter, J. (1994): “L’IPPI, un Indicador molt Valuós per Mesurar l’Activitat Industrial Catalana”, *Revista d’Indústria*, 3, 2^a etapa, pp. 6-15. Generalitat de Catalunya, Departament d’Indústria i Energia.

Cuthberson, K., Hall, S. y Taylor, M. (1992): *Applied Econometric Techniques*, Phillip Allan, Nueva York.

de Jong, P. (1991): “The Diffuse Kalman Filter”, *The Annals of Statistics*, 19, pp. 1073-1083.

Dempster, A., Laird, N. y Rubin, D. (1977): “Maximum Likelihood from Incomplete Data via the EM Algorithm”, *Journal of the Royal Statistical Society, Serie B*, 39, pp. 1-38.

Engle, R.F. y Watson, M.W. (1987): “The Kalman Filter: Applications to Forecasting and Rational-Expectations Models”, en Bewley, T.F. (ed.), *Advances in Econometrics: Fifth World Congress*, Econometric Society Monograph, vol. 13, pp. 245-283, Cambridge University Press, Cambridge.

EUROSTAT (1978): *L’Indice de la Production Industrielle de la Communauté Européenne*, Suplement Metodològic 1/78.

Griliches, Z. y Jorgenson, D.W. (1966): “Sources of Measured Productivity Change: Capital Input”, *American Economic Review*, 56, pp. 50-61.

Hackl, P. y Westlund, A.H. (1996): “Demand for International Telecommunications: Time Varying Price Elasticity”, *Journal of Econometrics*, 70, pp. 243-260.

Harvey, A.C. (1981): *Time Series Models*, Phillip Allan, Oxford.

Harvey, A.C. (1982): “The Kalman Filter and its Applications in Econometrics and Time Series Analysis”, *Methods of Operational Research*, 44, pp. 3-18.

Harvey, A.C. (1984a): “A Unified View of Statistical Forecasting Procedures”, *Journal of Forecasting*, 3, pp. 245-275.

Harvey, A.C. (1984b): “Dynamic Models, the Prediction Error Decomposition and State-Space Models”, en Hendry, D.F. y Wallis, K.F. (eds.), *Econometrics and Quantitative Econometrics*, pp. 37- 59, Basil Blackwell, Oxford.

Harvey, A.C. (1987): “Applications of the Kalman Filter in Economics”, en Bewley, T.F. (ed.), *Advances in Econometrics: Fifth World Congress*, Econometric Society Monograph, vol. 13, pp. 285-313, Cambridge University Press, Cambridge.

Harvey, A.C. (1989): *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge University Press, Cambridge.

Harvey, A.C. (1990): *The Econometric Analysis of Time Series*, Phillip Allan, Nueva York, 2ª edición.

Harvey, A.C. y Phillips, G.D.A. (1979): “Maximum Likelihood Estimation of Regression Models with Autoregressive-Moving Average Disturbances”, *Biometrika*, 66, pp. 69-58.

Israilevich, P.R. y Kuttner, K.N. (1993): “A Mixed Frequency Model of Regional Output”, *Journal of Regional Science*, 33, pp. 321-342.

Jazwinski, A.H. (1970): *Stochastic Processes and Filtering Theory*, Academic Press, Nueva York.

Kalman, R.E. (1960): “A New Approach to Linear Filtering and Prediction Problems”, *Transactions ASME, Journal of Basic Engineering*, 82, pp. 35-45.

Kalman, R.E. y Bucy, R.S. (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 Priors 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.F. (1986): “Estimation, Prediction and Interpolation for ARIMA Models with Missing Data”, *Journal of the American Statistical Association*, 81, pp. 751-761.

Molina, A. y Sanz, R. (1985): “Un Indicador Mensual del Consumo de Energía Eléctrica para Usos Industriales 1976-84”, *Documento de Trabajo 8510*, Servicio de Estudios, Banco de España.

Moody, C.E. (1974): “The Measurement of Capital Services by Electrical Energy”, *Oxford Bulletin of Economics and Statistics*, 36, pp. 42-52.

Norrbin, S.C. y Schlagenhauf, D.E. (1988): “An Inquiry into the Sources of Macroeconomic Fluctuations”, *Journal of Monetary Economics*, 22, pp. 43-70.

Oliveras, J. (1992): “El Pla Estadístic de Catalunya: La Llei 30/1991”, *Nota d’Economia*, 43, pp. 33-42.

Revilla, P. (1997): “El IPI como Principal Indicador Económico de Oferta”, *Fuentes Estadísticas*, 30, pág. 11.

Rosenberg, B. (1973): “Commentaire a l’Exposé de M. Dunn”, *Economie Appliquée*, 4, pp. 531-534.

Sant, D. (1977): “Generalized Least Squares Applied to Time Varying Parameters Models”, *Annals of Economic and Social Measurement*, 6, pp. 301-312.

Sanz, R. (1979): “Modelización del Índice de Producción Industrial y su Relación con el Consumo de Energía Eléctrica”, *Cuadernos Económicos del ICE*, 11-12, pp. 227-259.

SAS Inst. Inc, (1988): *SAS/IML User’s Guide 6.12 version*.

Shumway, R.H. y Stoffer, D.S. (1982): “An Approach to Time Series Smoothing and Forecasting Using the EM Algorithm”, *Journal of Time Series Analysis*, 3, pp. 253-264.

Suriñach, J., Pons, E. y Pons, J. (1996): *Comptabilitat Econòmica de Catalunya i Mètodes de Trimestralització*, Institut d’Estadística de Catalunya, Barcelona. Una versió reducida puede encontrarse en *Revista Econòmica de Catalunya*,

1996, 30, pp. 38-56.

Tanizaki, H. (1996): *Nonlinear Filters. Estimation and Inference*, Springer-Verlag, Heidelberg.

Watson, M.W. y Engle, R.F. (1983): “Alternative Algorithms for the Estimation of Dynamic Factor, MIMIC and Varying Coefficient Regression”, *Journal of Econometrics*, 23, pp. 385-400.