



REVISION DEL SISTEMA DE INDICADORES CICLICOS

S.G. Cuentas Nacionales
Instituto Nacional de Estadística

Marzo 2001

1. INTRODUCCION

En esta nota se presentan las primeras modificaciones del Sistema de Indicadores Cíclicos (SIC) que viene elaborando el INE desde 1994, véase INE (1994). La concurrencia de cambios en las fuentes básicas, la modificación de la estructura dinámica de los indicadores elementales y la disponibilidad de nuevos procedimientos computacionales han aconsejado realizar una revisión profunda del SIC. A continuación se presentan los primeros resultados, detallándose la metodología empleada (sección 2), los resultados empíricos básicos (sección 3) y el esquema de difusión de los mismos (sección 4).

A lo largo de 2001 y 2002 se incorporarán sucesivas ampliaciones, tales como un nuevo índice difusión que revise el anteriormente publicado (véase INE (1995)) y un indicador de probabilidad de puntos de giro dependiendo, especialmente este último, de los desarrollos informáticos del proyecto BUSY (*Tools and Practices For Business Cycle Analysis in National Statistical Institutes of EU*).

2. METODOLOGIA ESTADISTICA

La metodología que se expone a continuación es una extensión al caso dinámico del conocido modelo factorial del análisis multivariante, véase Peña y Box (1987) y Peña y Poncela (1999). Esta metodología permite tanto la identificación de patrones comunes como la determinación de la estructura de desfases (clasificación cíclica). Asimismo, puede ser utilizada para el diseño de índices sintéticos. Un tratamiento detallado se ofrece en Melis (1985), Cristóbal *et al.* (1996), Cristóbal y Quilis (1998) y Abad *et al.* (2000).

El principio básico de la metodología se puede expresar mediante el siguiente ejemplo.

Sea $z_{2,t}$ un indicador adelantado de $z_{1,t}$ con un desfase de 1 mes. Entonces $z_{2,t-1}$ es un indicador coincidente de $z_{1,t}$ y la relación que vincula a ambas series podrá ser expresada de la siguiente forma:

$$[1] \quad \begin{bmatrix} z_{1,t} \\ z_{2,t-1} \end{bmatrix} = \begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix} f_t + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix}$$

donde f_t es el factor común que vincula a ambas series y que explica la mayor parte de su variabilidad, λ_i son las cargas correspondientes y ε_i son los factores idiosincrásicos que representan las variaciones de z_i que no pueden atribuirse al patrón común del sistema. De esta forma, una vez identificada la estructura de desfases existente entre los indicadores (en el ejemplo, el desfase de un mes entre $z_{2,t}$ y $z_{1,t}$), se estima un modelo factorial estático para el vector de series adecuadamente puesto en fase. Naturalmente, las medidas habituales de bondad de ajuste y de diagnóstico de esta clase de modelos resultan aplicables, por lo que se dispone de un potente instrumento para cuantificar la hipótesis factorial.

En consecuencia, en esta clase de modelos existen dos ecuaciones: una expresa la relación dinámica existente entre el vector de series temporales y otra que representa la estructura factorial. La primera es:

$$[2] \quad W_t = H(B)Z_t \quad \text{con } H(B) = \text{diag}(B^{d_1} \dots B^{d_k})$$

donde Z_t es el vector de k series observadas, d_i son los desfases correspondientes y W_t es el vector de series puestas en fase que, por lo tanto, evolucionan de manera coincidente. La expresión [2] actúa como "ecuación de transición", reflejando la estructura dinámica del sistema.

Para completar el modelo se necesita una segunda ecuación que relaciona W_t con sus factores comunes inobservables:

$$[3] \quad W_t = PF_t + E_t$$

donde F_t y E_t son vectores que representan, respectivamente, los r factores comunes del sistema ($r < k$) y los k factores idiosincrásicos. La especificación se completa asumiendo ortogonalidad entre F_t y E_t e imponiendo una estructura diagonal a los componentes idiosincrásicos:

$$[4] \quad E_t \sim \text{iid } N(0, \Sigma_E) \quad \Sigma_E = \text{diag}(\sigma_{E1} \dots \sigma_{Ek})$$

La identificación empírica de este modelo requiere, como pasos esenciales, la determinación de la estructura de desfases $d=[d_1 \dots d_k]$ y la estimación de la correspondiente matriz de cargas P . Con el fin de realizar este procedimiento de identificación empírica se propone el siguiente algoritmo:

1. Extensión del vector de series observadas

Se amplía el vector original Z_t incluyendo sus valores desfasados Z_{t-j} con $j=0..M$. El correspondiente vector de $k(M+1)$ elementos es:

$$[5] \quad Y_{M,t} = \begin{bmatrix} Z_t \\ Z_{t-1} \\ \dots \\ Z_{t-M} \end{bmatrix}$$

La determinación del máximo valor del retardo M se puede realizar bien mediante consideraciones *a priori*, bien mediante algún criterio de penalización como el AIC o el BIC. La elección de un valor apropiado para M requiere ponderar dos consideraciones contrapuestas: parsimonia (para evitar la sobreparametrización del modelo y facilitar su interpretación) y resolución (para identificar correctamente la estructura de desfases).

La experiencia acumulada con este método sugiere que seleccionar $M=12$ o $M=4$ para series mensuales o trimestrales, respectivamente, es una elección razonable. Asimismo, si el valor de k es elevado la dimensión de $Y_{M,t}$ deviene inmanejable, por lo que se su composición se efectúa considerando $j=0..M$ de manera discontinua. Por ejemplo, con $M=12$:

$$[6] \quad Y_{M,t} = \begin{bmatrix} Z_t \\ Z_{t-3} \\ Z_{t-6} \\ Z_{t-9} \\ Z_{t-12} \end{bmatrix}$$

2. Estimación del modelo factorial estático

Se identifica y estima un modelo factorial estático para el vector ampliado $Y_{M,t}$, mediante el análisis de los autovalores y autovectores de la matriz de correlaciones de dicho vector. Nótese que, si M es un valor suficientemente grande, dicha matriz encapsula toda la dinámica del vector original Z_t , al contener todas las matrices de correlaciones cruzadas de dicho vector hasta el retardo M :

$$[7] \quad \Gamma_Y(0) = \begin{bmatrix} \Gamma_Z(0) & \Gamma_Z(1) & \dots & \Gamma_Z(M) \\ \Gamma_Z(1) & \Gamma_Z(0) & \dots & \Gamma_Z(M-1) \\ \dots & \dots & \dots & \dots \\ \Gamma_Z(M) & \Gamma_Z(M-1) & \dots & \Gamma_Z(0) \end{bmatrix}$$

Esta matriz es de tipo Toeplitz (simétrica por bloques y en su conjunto) y condensa toda la información dinámica del vector de series observadas Z_t .

Usualmente, por razones prácticas y también con el fin de facilitar la interpretación del modelo, se considera un modelo unifactorial: $r=1$. No obstante, el algoritmo es aplicable al caso general $r>1$ sin problema alguno.

$$[8] \quad Y_{M,t} = P f_t + E_t$$

La estimación del vector de cargas P se realiza calculando el autovector asociado al mayor autovalor de la matriz de correlaciones contemporáneas de $Y_{M,t}$.

3. Simplificación de la matriz de cargas

Habitualmente, la dimensión de P es muy elevada, por lo que su interpretación no es sencilla. Por ello es conveniente idear algún criterio de simplificación que resalte los aspectos más importantes contenidos en P .

Sea $\lambda_{i,j}$ la carga de la variable z_i en el retardo j . Se considerará, en primer lugar, sólo aquellos valores que sean mayores, en valor absoluto, que un cierto umbral o "valor crítico" λ_c . Como la distribución de las cargas en el muestreo no es fácil de derivar, se recomienda seleccionar un valor para λ_c proporcional a $n^{-1/2}$, donde n es el número de observaciones. En consecuencia:

$$[9] \quad \lambda_{i,j}^* = \begin{cases} \lambda_{i,j} & \leftrightarrow |\lambda_{i,j}| \geq \lambda_c \\ 0 & \text{en los demás casos} \end{cases}$$

La identificación del índice de desfase d_i de cada serie se obtiene calculando el máximo correspondiente:

$$[10] \quad \lambda_{i,j}^{**} = \begin{cases} \lambda_{i,j}^* & \leftrightarrow \lambda_{i,j}^* = \max_j (|\lambda_{i,j}^*|) \\ 0 & \text{en los demás casos} \end{cases}$$

El valor del subíndice j para el que se maximiza la carga de la serie i constituye la estimación del retardo d_i correspondiente.

4. Generación del modelo

Una vez que se ha simplificado la matriz de cargas P y que se conoce la estructura de desfases $d=[d_1..d_k]$, se puede formar el modelo completo antes definido:

$$[2] \quad W_t = H(B)Z_t \quad \text{con } H(B) = \text{diag}(B^{d_1}..B^{d_k})$$

$$[3] \quad W_t = Pf_t + E_t$$

5. Determinación de la serie de referencia

Una de las ventajas de esta técnica es que ofrece una manera puramente empírica de establecer cuál es la serie de referencia. Para ello conviene recordar que el factor común resume toda la información dinámica de las series consideradas, ya que se ha estimado considerando los sucesivos desfases de Z_t (al operar sobre el vector extendido $Y_{M,t}$) y que, en consecuencia, las cargas $\lambda_{i,j}$ especifican la vinculación de cada serie en cada desfase con el factor. De esta manera, podremos identificar como serie de referencia aquella que se encuentra más relacionada con el factor común obtenido. Formalmente:

$$[11] \quad z_{h,t} \text{ es la serie de referencia } \leftrightarrow h = \max_i (|\lambda_{i,j}|)$$

Naturalmente, la elección de la serie de referencia puede hacerse *a priori*, en cuyo caso esta etapa se omite.

6. Clasificación cíclica

La serie de referencia $z_{h,t}$ está máximamente identificada con el factor común f_t en un retardo d ($0 \leq d \leq M$). Para cada variable $i \neq h$ se considera el retardo j correspondiente a su carga simplificada no nula $\lambda_{i,j}$ y se calcula su desfase característico ($de_i = d - j$). De esta manera tenemos el siguiente criterio de clasificación:

$$[12] \quad z_{i,t} \text{ es } \begin{cases} < a > \\ < c > \\ < r > \end{cases} \text{ respecto a } z_{h,t} \leftrightarrow \begin{cases} de_i < -L \\ -L \leq de_i \leq L \\ de_i > L \end{cases}$$

donde L es un parámetro fijado *a priori*. Naturalmente, la clasificación puede ser más detallada, incluyendo, si es posible, series muy adelantadas y/o muy retrasadas. El caso $< i >$ (aciclicidad) aparece cuando $\lambda_{i,j} = 0 \forall j$, esto es, cuando la serie considerada no está identificada de forma alguna con el patrón dominante y, en consecuencia, tampoco lo está con la serie de referencia $z_{h,t}$.

El siguiente ejemplo ilustra el método propuesto. Se considera: $k=4$, $M=2$, $r=1$ y $\lambda_c=0.2$. El modelo factorial para el vector ampliado $Y_{2,t}$ es:

$$[13] \quad \begin{bmatrix} z_{1,t} \\ z_{2,t} \\ z_{3,t} \\ z_{4,t} \\ \dots \\ z_{1,t-1} \\ z_{2,t-1} \\ z_{3,t-1} \\ z_{4,t-1} \\ \dots \\ z_{1,t-2} \\ z_{2,t-2} \\ z_{3,t-2} \\ z_{4,t-2} \end{bmatrix} = \begin{bmatrix} +0.91 \\ +0.72 \\ -0.43 \\ +0.11 \\ \dots \\ +0.73 \\ +0.98 \\ -0.52 \\ +0.09 \\ \dots \\ +0.61 \\ +0.82 \\ -0.64 \\ +0.12 \end{bmatrix} f_t + E_t$$

La aplicación a [13] del algoritmo de simplificación antes expuesto arroja los siguientes resultados:

$$[14] \quad \begin{bmatrix} z_{1,t} \\ z_{2,t} \\ z_{3,t} \\ z_{4,t} \\ \dots \\ z_{1,t-1} \\ z_{2,t-1} \\ z_{3,t-1} \\ z_{4,t-1} \\ \dots \\ z_{1,t-2} \\ z_{2,t-2} \\ z_{3,t-2} \\ z_{4,t-2} \end{bmatrix} = \begin{bmatrix} +0.91 \\ 0 \\ 0 \\ 0 \\ \dots \\ 0 \\ +0.98 \\ 0 \\ 0 \\ \dots \\ 0 \\ 0 \\ -0.64 \\ 0 \end{bmatrix} f_t + E_t$$

En consecuencia, cabe establecer lo siguiente:

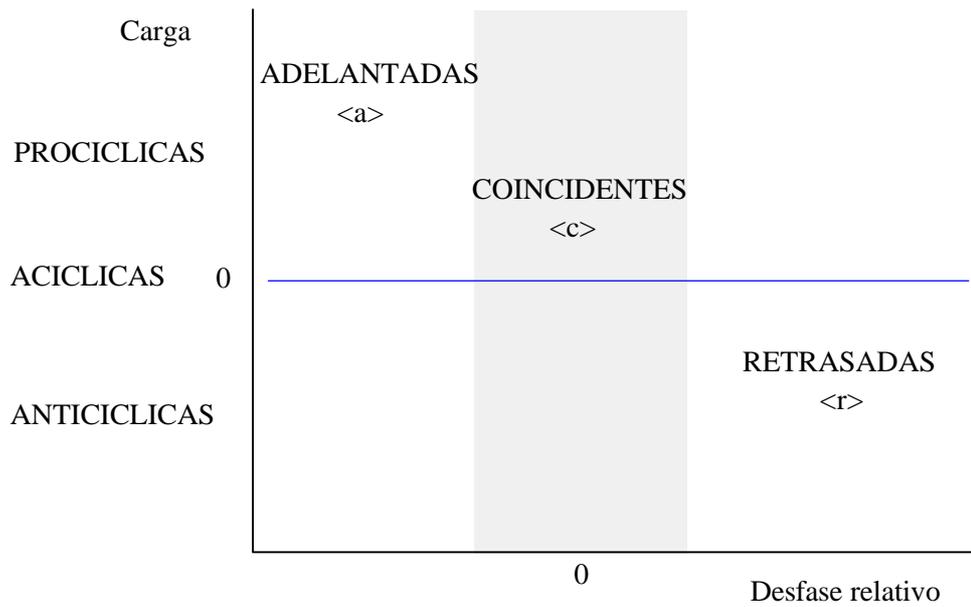
- La serie z_1 se comporta de forma procíclica y está identificada con el factor común de forma contemporánea.
- Por su parte, la serie z_2 también es procíclica pero su identificación con el factor se produce con un retraso de un período.
- La serie z_3 es anticíclica y está vinculada con el factor común con un desfase de dos períodos.
- Finalmente, z_4 no está identificada con el factor común que vincula a las restantes series, por lo que puede ser considerada como acíclica.

Como ya se ha comentado, una de las ventajas de este enfoque es que permite elegir una serie de referencia en función de las propias características empíricas de los datos, sin necesidad de realizar dicha elección *a priori*. En el ejemplo anterior, la serie z_2 podría actuar como referente por ser la más identificada con el factor que describe el ciclo común del vector de series. En consecuencia, como $z_{1,t}$, $z_{2,t-1}$ y $z_{3,t-2}$ son coincidentes entre sí, se puede establecer que z_1 y z_3 son, respectivamente, indicadores retrasado y adelantado de z_2 :

$$[15] \quad \begin{bmatrix} z_{1,t} \\ z_{2,t-1} \\ z_{3,t-2} \end{bmatrix} \text{son } < c > \Rightarrow \begin{bmatrix} z_{1,t} < r > z_{2,t} \\ z_{3,t} < a > z_{2,t} \end{bmatrix}$$

De esta manera, combinando la información suministrada por las cargas con la proporcionada por los desfases, se puede realizar una clasificación cíclica completa, como la que se aprecia en el siguiente gráfico:

Gráfico 1: Clasificación cíclica mediante análisis factorial



Finalmente, a partir de los grupos de series adelantadas ($Z_{<a>}$), coincidentes ($Z_{<c>}$) y retrasadas ($Z_{<r>}$) se pueden formar índices sintéticos de adelanto (a_t), coincidencia (c_t) y retraso (r_t) como combinaciones lineales basadas en la estructura de cargas y teniendo en cuenta la estructura de desfases relativos existente dentro de cada grupo:

$$\begin{aligned}
 a_t &= A(B)Z_{<a>,t} \\
 [16] \quad c_t &= C(B)Z_{<c>,t} \\
 r_t &= R(B)Z_{<r>,t}
 \end{aligned}$$

Dichos índices son una combinación convexa y normalizada de las series que integran cada categoría dinámica, tomando como base la matriz de cargas simplificadas. Hay que advertir que los grupos no están, por lo común, ubicados en un punto único sino en una banda temporal. Usualmente, se elige como centro de cada grupo el desfase que, comenzando en cero, se encuentra separado tres meses por cada lado del correspondiente al siguiente grupo: 0, 3, 6 y 9. A continuación se colapsa cada banda a dichos puntos reteniendo la estructura interna de desfases.

3. RESULTADOS EMPIRICOS BASICOS

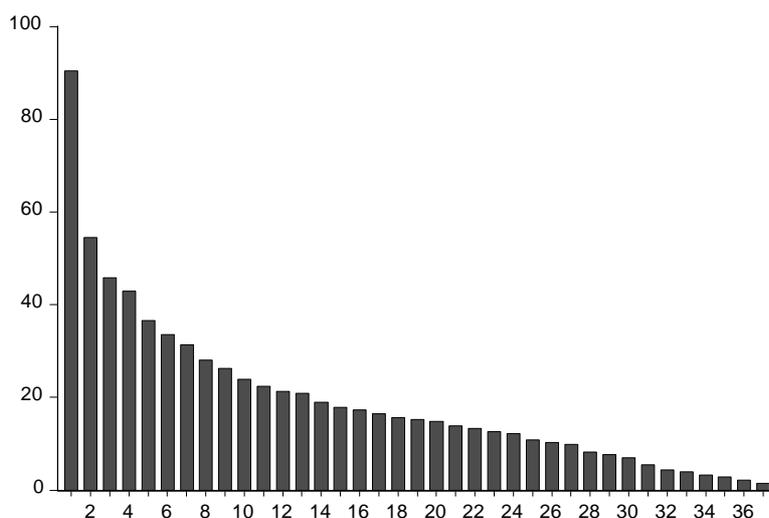
La metodología expuesta en la sección anterior va a ser aplicada a 133 Indices de la Producción Industrial (IPI), elaborados mensualmente por el INE. El período muestral comienza en enero de 1975.

En general, las series de los IPI están caracterizadas por una amalgama de elementos tendenciales, estacionales e irregulares que dificultan la estimación de su perfil cíclico. Por ello, en este trabajo se han filtrado dichas series por medio de la combinación de un filtro de paso en banda (la tasa interanual) y un suavizador de paso bajo (el filtro AR(2) descrito en Melis, 1991). El resultado es una estimación bastante nítida del ritmo de avance de la serie temporal objeto de estudio.

La resolución de desfases seleccionada es $M=12$, que se ha mostrado como una elección razonable en los diversos ensayos realizados con esta técnica.

En el gráfico 1 se muestran los 37 primeros autovalores de la matriz de correlaciones de Y_t que, como ya se ha comentado, encapsula la dinámica de los 133 IPI considerados. La elevada distancia entre el primer y el segundo autovalor así como la presencia de un fenómeno de morrena (*scree*) a partir del segundo, sugiere que el número apropiado de factores es uno ($r=1$).

Gráfico 2: Autovalores de la matriz Γ_Y



Una vez determinado el número adecuado de factores se puede estimar la matriz de cargas correspondiente y, como ya se ha comentado, especificar la estructura de desfases. No se ha considerado un umbral de significación *a priori* para las cargas individuales habida cuenta del desconocimiento práctico de su distribución. La dimensión de la matriz de tipo Toeplitz Γ_Y es 1729×1729 , lo que proporciona una medida de los requisitos de cómputo de esta clase de técnicas.

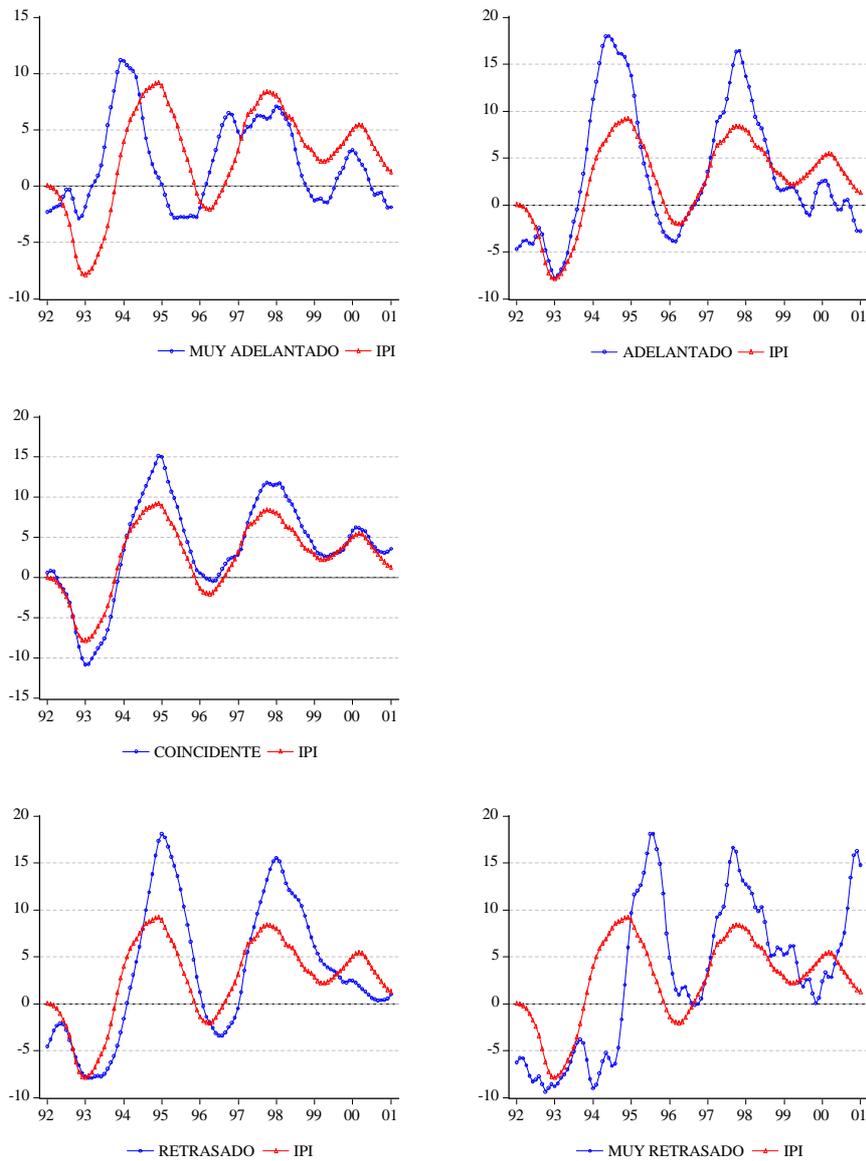
Como referencia temporal se ha tomado el punto medio de la retícula de desfases ($d=6$), formándose cinco grupos en función de la distancia de cada desfase característico, de_i . De esta manera se tiene:

- $\langle aa \rangle$: series muy adelantadas: $de_i \in [-6, -5]$

- <a>: series adelantadas: $de_i \in [-4,-2]$
- <c>: series coincidentes: $de_i \in [-1,1]$
- <r>: series retrasadas: $de_i \in [2,4]$
- <rr>: series muy retrasadas: $de_i \in [5,6]$

Habida cuenta de la elevada dimensión del sistema aquí modelizado, no se detalla la matriz de cargas correspondiente sino que se ofrecen, en el gráfico 3, los correspondientes cinco índices sintéticos. Dichos índices son una combinación convexa y normalizada de los IPI que integran cada categoría dinámica, tomando como base la matriz de cargas simplificadas. Hay que advertir que los grupos no están, por lo común, ubicados en un punto único sino en una banda temporal. Se ha elegido como centro de cada grupo el desfase que, comenzando en cero, se encuentra separado un trimestre por cada lado del correspondiente al siguiente grupo: 0, 3, 6 y 9. A continuación se colapsa cada banda a dichos puntos reteniendo la estructura interna de desfases. En Cristóbal y Quilis (1996) se encuentra una exposición detallada.

Gráfico 3: Indicadores sintéticos del SIC e IPI general



Finalmente, cabe preguntar si estas propiedades dinámicas se encuentran ligadas a alguna característica estructural de los IPI. A modo de ilustración, en la tabla 1 se compara la agrupación cíclica obtenida con la correspondiente a la Clasificación Nacional de Actividades Económicas (CNAE-74).

Tabla 1: Clasificación cíclica y agrupaciones CNAE

División CNAE-74	Clasificación cíclica				
	<aa>	<a>	<c>	<r>	<rr>
1: Energía y agua	1	1	3	1	2
2: Extracción de minerales. Química	6	7	10	8	13
3: Transformación de metales	15	5	13	1	4
4: Otras manufacturas	8	6	8	10	11

En primer lugar, no existe un patrón de concentración claro en la tabla anterior, de manera que los IPI de las distintas ramas de actividad no se agrupan de manera muy señalada en alguna de las cinco categorías dinámicas. No obstante, los índices correspondientes a las industrias básicas y de otras manufacturas tienden a ser retrasados, mientras que los de las transformadoras son predominantemente coincidentes o muy adelantados. Dada la gran incertidumbre existente acerca de los fenómenos dinámicos, parece razonable emplear ambas clasificaciones con el fin de asentar el diagnóstico coyuntural sobre una base más firme, puesto que cada una de ellas aporta información sobre aspectos no contemplados en la otra clasificación.

4. DIFUSION DE LOS RESULTADOS

Acompañando los datos de la Contabilidad Nacional Trimestral (CNTR) se ofrecerán los resultados del SIC, en <http://www.ine.es/daco/daco42/daco4214/btcmenu.htm>, cuyo análisis se efectuará regularmente en el informe *El Momento Económico* que se publica en el *Boletín Trimestral de Coyuntura* del INE.

REFERENCIAS

- Abad, A., Cristóbal, A. y Quilis, E.M. (2000) "Fluctuaciones económicas, puntos de giro y clasificación cíclica", *Boletín Trimestral de Coyuntura*, n. 78, p. 135-171.
- Cristóbal, A., Melis, F. y Quilis, E.M. (1996) "Las series de cartera de pedidos de la industria y su clasificación dinámica", Instituto de Estudios Fiscales, Papeles de Trabajo n. 27/96.
- Cristóbal, A. y Quilis, E.M. (1998) "Análisis del ciclo económico mediante modelos de índices dinámicos. Aplicación a los Índices de la Producción Industrial", *Boletín Trimestral de Coyuntura*, n. 68, p. 73-93.
- INE (1994) *Sistema de Indicadores Cíclicos de la Economía Española. Metodología e índices sintéticos de adelanto, coincidencia y retraso*, Instituto Nacional de Estadística, Madrid, España.
- INE (1995) "La elaboración del índice de difusión del empleo", *Boletín Trimestral de Coyuntura* n. 58, p. 61-76.
- Melis, M. (1985) "Caracterización y clasificación de series económicas. Aplicación al sistema de números índices de la producción industrial", Instituto Nacional de Estadística, Documento Interno.
- Melis, F. (1991) "La estimación del ritmo de variación en series económicas", *Estadística Española*, n. 126, p. 7-56.
- Peña, D. y Box, G.E.P (1987) "Identifying a simplifying structure in time series", *Journal of the American Statistical Association*, vol. 82, n. 399, p. 836-842.
- Peña, D. y Poncela, P (1999) "Nonstationary dynamic factor analysis", Universidad Carlos III, Documento Interno.