Notas metodológicas. Coyuntura Ciudad de Zaragoza

fundear

1. Índice de Actividad Económica de Zaragoza

El nivel de desagregación utilizado no debería afectar al método empleado para la construcción de indicadores de actividad[1]. Sin embargo, la información disponible a nivel regional, provincial o municipal sí introduce una serie de características que es preciso tener en cuenta a la hora de diseñar una técnica adecuada[2]. El número, tamaño muestral y calidad de los indicadores factibles de ser incluidos en el indicador sintético es inferior a medida que descendemos en el nivel de desagregación. Todos estos factores condicionan de manera importante la técnica econométrica a utilizar. Finalmente, en la mayoría de los casos no es posible disponer de un agregado macroeconómico de referencia con las cualidades deseadas, y sobre el que basar la comparación del indicador, lo que dificulta la construcción de indicadores adelantados, debiendo limitarnos a elaborar un indicador coincidente que permita realizar predicciones a partir del mismo, e identificar puntos de giro.

La metodología utilizada se basa en la extracción de señales de un conjunto de variables de actividad y demanda de la economía de Zaragoza, para posteriormente sintetizar la información que contienen en un solo indicador. Aunque la finalidad del índice sintético es acercarse a la evolución de la tasa real de crecimiento del VAB y adelantar la evolución del ciclo económico, no es posible hacer diferencias directamente sobre todas las series utilizadas. Cada una de ellas se analiza de forma individual atendiendo a su naturaleza, bien se trate de datos en niveles nominales o reales, porcentajes o datos puntuales procedentes de una encuesta. Ello implica que es necesario realizar un tratamiento previo adecuado para cada caso: deflactando, convirtiendo la serie en continua a través de una

---

1 Una buena síntesis de las diferentes metodologías y su aplicación al ámbito regional se encuentra en Sansó (2001).

2 Una detallada descripción de toda la información estadística disponible para Zaragoza se encuentra en el Anexo.

transformación logística o estudiando los posibles datos atípicos que contiene. Asimismo, aquellas series con periodicidad inferior a la mensual –trimestral se desagregan con el fin de incorporar su información.

Una vez tratadas cada una de las series de este modo, es preciso eliminar sus componentes estacionales, efectos calendario y elementos irregulares al objeto de extraer el componente ciclo-tendencia. Es decir, se sustraen aquellos elementos ruidosos de cada serie original, que impiden observar su comportamiento tendencial –evolución del crecimiento a largo plazo- y sus variaciones en el corto plazo -efecto cíclico-, mediante la aplicación de filtros. La representación de cada proceso mediante el modelo ARIMA más adecuado, permite realizar además predicciones sobre su evolución en un futuro próximo.

Preparadas de este modo las series, en el siguiente paso se sintetiza la información que contienen en un único indicador, para lo que se utiliza, como es habitual en la mayor parte de los estudios de coyuntura, el análisis factorial dinámico. Los resultados obtenidos permiten ponderar el peso de cada uno de los indicadores según el porcentaje de varianza explicada. Esta metodología tiene, sin embargo, el inconveniente de amplificar la volatilidad de las series, por lo que es preciso corregir sus ponderaciones por la inversa de la varianza. Finalmente, el indicador sintético obtenido, al carecer de escala, debe ser ajustado a la tasa media de crecimiento del VAB para facilitar su interpretación. O en su defecto, si no se dispone de este u otro agregado, referir el índice a un año base. Cada una de estas etapas se describen con detalle a continuación.

1.1 Análisis previo de cada serie

Aunque el objetivo final es lograr una aproximación a la tasa de crecimiento real del VAB total o sectorial, no es posible diferenciar directamente las series sin realizar un análisis previo de su naturaleza y características; veamos algunos ejemplos. En algunos casos nos encontraremos con cambios metodológicos que será preciso corregir realizando en la medida de lo posible un enlace de las series. Sobre la posible presencia de *outliers* se realiza un estudio gráfico de las series para reconocer su presencia, aunque su análisis formal se relega a fases posteriores, en concreto antes de realizar la descomposición ciclo-tendencia. Es preciso, además, deflactar adecuadamente las series que se encuentran en términos nominales seleccionando un deflactor adecuado, utilizando en la mayoría de los casos el Índice de Precios al Consumo –IPC- regional o provincial o el deflactor sectorial nacional, al no contar con el correspondiente a nivel municipal.

Un último problema que puede aparecer en algunas series, como las procedentes de encuestas, es que se trata de observaciones de tipo de discreto. La conversión de este tipo de series en continúas se realiza aplicando una transformación logística. Siendo Y_t una serie de tipo discreto, es posible conseguir su transformación continua del siguiente modo:

$$\tilde{Y}_t = \ln \frac{Y_t}{1 - Y_t} \qquad (1.1.1)$$

Finalmente, en otro tipo de series como índices ó personas ocupadas,..., pueden considerarse directamente sus tasas de crecimiento tomadas como diferencias logarítmicas.

1.2 Homogeneización de las frecuencias de las series

Para construir el indicador sintético de la economía de Zaragoza se dispone de un conjunto de variables de coyuntura que, al margen de carecer de un periodo muestral suficientemente amplio, cuentan con la dificultad añadida de no ser coincidentes en cuanto a su desagregación temporal[3]. Esto nos conduce a tener que homogeneizar las frecuencias de cada una de las variables que vamos a utilizar. En concreto, dado que el propósito es construir un indicador sintético mensual, parece razonable conseguir una adecuada mensualización de aquellos indicadores parciales que hayan sido

[3] Por ejemplo, de variables como los afiliados a la seguridad social podemos obtener datos de forma mensual pero, por el contrario, las variables procedentes de la encuesta de población activa sólo están disponibles de forma trimestral.

obtenidos para una frecuencia distinta a la mensual, de forma que su información pueda ser empleada, conjuntamente con la del resto de indicadores parciales, para construir el indicador sintético.

La "mensualización" de series trimestrales o, en términos mucho más generales, la generación de series de alta frecuencia a partir de la información que proporcionan series de una frecuencia inferior es una cuestión que ha sido estudiada desde hace bastante tiempo por parte de los economistas. Nosotros nos decantamos por el método de Chow-Lin (1971), en virtud de las características de las series que vamos a utilizar. El procedimiento de Chow-Lin de desagregación temporal se basa en la existencia de una relación entre la variable que queremos desagregar y una serie de indicadores parciales de la actividad económica. Por tanto, necesitamos información tanto de la variable a desagregar, como de estos indicadores de actividad. En la medida que estos últimos sean capaces de explicar el comportamiento de nuestra variable objetivo, el procedimiento puede proporcionar resultados altamente satisfactorios.

De forma más analítica, el procedimiento de Chow-Lin se describe de la siguiente manera. Supongamos que se dispone de una muestra de tamaño T de una variable $y_B = (y_{B1}, y_{B2}, ..., y_{BT})$ que queremos desagregar temporalmente. La variable desagregada se denota como $y_A = (y_{A1}, y_{A2}, ..., y_{An})$, donde $n = \upsilon$ T. El valor del parámetro υ depende del tipo de desagregación a realizar. Por ejemplo, para transformar datos anuales a mensuales, entonces $\upsilon=12$, mientras que si se desagrega información trimestral en mensual, entonces $\upsilon=3$. El objetivo es el de hallar este vector y_A. El problema es cómo obtenerlo de una forma eficiente. Para ello es necesario formular una serie de requisitos previos.

El primero de ellos es definir la relación entre las variables y_A e y_B. De forma analítica, representamos esta relación mediante la siguiente ecuación:

$$y_B = C\, y_A = (I_T \otimes c)\, y_A \qquad (1.2.1)$$

donde $\otimes$ representa el producto de Kronecker y el valor del vector c, de dimensión $1x\upsilon$, depende del tipo de variable que estamos desagregando. En la mayoría de las ocasiones, y suponiendo una desagregación de datos trimestrales en mensuales, el vector c será igual a c= (1, 1, 1), pero no es la única posibilidad[4]. Por tanto, la matriz de transformaciones C tiene una dimensión T x (T+υ).

Por otro lado, en el procedimiento de Chow-Lin es clave la existencia de una relación entre la variable y el conjunto de indicadores de la actividad económica. Esta relación se explicita como sigue:

$$y_A = X_A \beta + u_A \qquad (1.2.2)$$

donde X_A es una matriz de orden n x k de indicadores de la actividad económica, β es un vector de k parámetros de posición y u_A es un vector de perturbaciones aleatorias de orden n, cuya matriz de varianzas y covarianzas la denotamos como Σ_A. Entonces, sin más que premultiplicar la ecuación anterior por C tenemos que:

$$y_B = X_B \beta + u_B \qquad (1.2.3)$$

donde la matriz de varianzas y covarianzas del vector de perturbaciones u_B la denotamos como Σ_B y es igual a:

$$\Sigma_B = Var(u_B) = Var(C\, u_A) = C\, \Sigma_A\, C' \qquad (1.2.4)$$

En este modelo la estimación de los parámetros de posición es posible, dado que tanto y_B como X_B son observables. No obstante, la restricción longitudinal que implica (1.2.1) supone que la simple aplicación de la técnica de mínimos cuadrados ordinarios no proporciona estimadores con propiedades óptimas. Si bien, no resulta excesivamente complicado encontrar un estimador eficiente del vector β que tenga en cuenta dicha restricción. Dicho vector de estimadores es igual a:

$$\hat{\beta} = (X_B'\, \Sigma_B^{-1}\, X_B)^{-1}\ X_B'\, \Sigma_B^{-1}\, y_B \qquad (1.2.5)$$

lo que nos permite definir el vector de residuos

$$\hat{u}_B = y_B - X_B\, \hat{\beta} \qquad (1.2.6)$$

A partir del mismo, es directo encontrar la estimación de la variable desagregada:

[4] Por ejemplo, si estamos utilizando variables que representan el valor medio del periodo o si estamos desagregando variables de stock, la forma que adopta el vector c sería distinta.

$$\hat{y}_A = X_A\ \hat{\beta} + \Sigma_A C'\ \Sigma_B^{-1}\ \hat{u}_B \qquad (1.2.7)$$

Por tanto, el valor de la variable desagregada es igual a un primer término que depende linealmente del conjunto de indicadores de la actividad económica seleccionados más un segundo que recoge el residuo de la estimación, convenientemente desagregado.

El estimador $\hat{\beta}$ no es sino el estimador generalizado del vector de parámetros β. Por tanto, comparte todas sus propiedades, pero también sus problemas. En consecuencia, su aplicabilidad está supeditada al conocimiento de la matriz de varianzas y covarianzas Σ_A. Si ésta es conocida, algo que no va a suceder en la práctica, se puede proceder a desagregar la variable. De otra forma, hay que establecer diversos supuestos sobre el comportamiento de esta matriz. En el artículo original de Chow y Lin (1971) se admite tanto que la perturbación u_A pueda seguir un proceso ruido blanco[5], como que siga un proceso autorregresivo de primer orden. Dadas las características de las series a emplear, nos decantamos por la primera opción, por lo que se deben incluir las restricciones:

$\Sigma_A = I_n$

$\Sigma_B = C\ C'$

en (2.2.5) y (2.2.7). A pesar de reconocer que este supuesto es altamente restrictivo, la escasez de la información con la que se trabaja obliga a inclinarse por esta opción en contra de otras, más interesantes desde el punto de vista teórico, pero no tan viables desde el práctico.

1.3 Extracción del componente ciclo tendencia

Una vez confeccionado de forma acertada el conjunto de indicadores de la actividad económica que van a ser la base del análisis, el siguiente paso dentro de la metodología es eliminar del conjunto de información disponible

5 Un ruido blanco es un proceso estocástico que se distribuye de forma independiente con media 0 y varianza constante. Aunque no es imprescindible, añadimos el supuesto de que dicho proceso sigue una distribución normal.

aquellos efectos que pueden ser nocivos, o simplemente irrelevantes, para la construcción de nuestro indicador sintético. La justificación para realizar este ajuste previo reside en el hecho de que las variables que vamos a utilizar para medir el estado de la coyuntura económica pueden estar contaminadas por la presencia de efectos que no proporcionan información apropiada de cara a conocer la evolución del ciclo económico de la economía que queremos estudiar. Los ejemplos más relevantes en este sentido son la existencia de un componente estacional, la presencia de *outliers* u observaciones anómalas, o también la existencia de efectos calendario, aunque este último tipo de efectos no son habituales para las periodicidades que vamos a utilizar en la generación del índice, aunque sí en otros escenarios.

En consecuencia, los indicadores parciales que se utilizan se pueden expresar de la siguiente manera:

$$y_t = c_t + s_t + p_t + u_t \qquad (1.3.1)$$

Donde c_t es el componente ciclo-tendencia, que es el que proporciona información útil sobre la evolución de largo plazo de la variable; s_t es el componente estacional, mientras que p_t y u_t son los componentes transitorio e irregular, respectivamente. De todos estos ellos, el único que debemos utilizar para la creación del índice sintético de la evolución de la economía aragonesa es el componente cíclico-tendencial. Por tanto, es preciso distinguirlo y de extraerlo. Esta tarea se dificulta por el hecho de que ninguno de los componentes incluidos en (1.3.1) se puede observar. No obstante, existen en la literatura diversas técnicas que pueden ayudar a la extracción del componente ciclo-tendencia de cada una de las variables de coyuntura que se usan en este trabajo[6]. De entre todos los métodos disponibles, se utilizado la metodología propuesta en el programa SEATS, acrónimo de *Signal Extraction in ARIMA Time Series*[7].

---

[6] Al margen del método que será empleado aquí, podemos señalar el uso de tendencias deterministas o el muy popular filtro de Hodrick-Prescott.

[7] Este programa ha sido desarrollado por Agustín Maravall y Víctor Gómez y se puede descargar libremente desde la página web del Banco de España, http://www.bde.es.

El planteamiento de estos autores se fundamenta en el hecho de que las variables a las que se les va a extraer el componente cíclico-tendencial se pueden expresar mediante un modelo ARIMA. Estos modelos nos ayudan a describir el comportamiento de las variables económicas interpretándolas a partir de procesos estocásticos. Muy brevemente, un modelo general ARIMA(p,d,q) x ARIMA(P,D,Q)$_s$ se puede escribir de la siguiente manera:

$$\phi_p(L)\ \Phi_P(L^s)\ \Delta^d\ \Delta^D\ x_t = \theta_q(L)\ \Theta_Q(L^s)\ u_t + \mu \qquad (1.3.2)$$

Donde L es el operador matemático de retardos, Δ=1-L es el operador de primeras diferencias, s indica la frecuencia en la que están medidos los datos[8], μ es un parámetro y u_t es ruido blanco. Por otro lado, $\phi_p(L)$ y $\Phi_P(L^s)$ son sendos polinomios de retardos de orden p y P, respectivamente, que representan el componente autorregresivo de la parte regular y de la estacional, también respectivamente. Del mismo modo $\theta_q(L)$ y $\Theta_Q(L^s)$ son sendos polinomios de retardos de orden q y Q, que se asocian a la parte de medias móviles del modelo regular y estacional, respectivamente. De una forma más compacta, un proceso ARIMA se puede expresar así[9]:

$$\Gamma(L)\ x_t = \vartheta(L)\ u_t + \mu \qquad (1.3.3)$$

donde $\Gamma(L)$ y $\vartheta(L)$ son sendos polinomios de retardos de orden (p+ P x s + d + D x s) y (q+Q), respectivamente.

Los valores de los diversos parámetros del proceso ARIMA son desconocidos, por lo que se hace necesario, como primer paso, determinar el orden del proceso ARIMA que mejor se adecua a los datos para, en segundo lugar, otorgar un valor numérico a todos los parámetros del modelo. Dicha tarea puede llevarse a cabo a través del empleo del programa TRAMO ("Times Series Regression with ARIMA Noise, Missing Observations and Outliers" diseñado también por los anteriormente mencionados profesores Agustín Maravall y Víctor Gómez). Aunque no es la única posibilidad disponible, el hecho de que TRAMO y SEATS se ejecuten

---

8 Por ejemplo, si disponemos de datos mensuales s=12, mientras que para el caso de datos trimestrales usaremos s=4.

9 Como es tradicional, hemos de incluir el supuesto de que las raíces características de los polinomios $\phi_p(L)$, $\Phi_P(L^s)$, $\theta_q(L)$ y $\Theta_Q(L^s)$ están todas fuera del círculo unidad.

de forma cuasi-automática, a lo que se añaden los buenos resultados que ofrecen dichos programas y el mayoritario empleo que hacen de los mismos tanto la comunidad científica como diversos centros estadísticos, nos conducen a elegir este método para estimar el modelo ARIMA más apropiado para cada uno de los indicadores parciales[10].

Una vez caracterizado dicho modelo ARIMA para un determinado indicador parcial, podemos pasar a la extracción de su componente cíclico-tendencial. El método utilizado por el programa SEATS consiste en admitir que cada uno de los posibles componentes del indicador expresados en (1.3.1) se puede especificar a partir de un modelo ARIMA. Entonces, podemos descomponer el modelo ARIMA asociado al indicador de la siguiente manera:

$$\Gamma_c(L)\ c_t = \vartheta_c(L)\ e_{ct} \qquad (1.3.4)$$

$$\Gamma_s(L)\ s_t = \vartheta_s(L)\ e_{st} \qquad (1.3.5)$$

$$\Gamma_p(L)\ p_t = \vartheta_c(L)\ e_{pt} \qquad (1.3.6)$$

donde asumimos que $\Gamma(L) = \Gamma_c(L)\ \Gamma_s(L)\ \Gamma_p(L)$, siendo e_{ct}, e_{st} y e_{pt} diversas perturbaciones aleatorias, que cumplen los requisitos necesarios para ser considerados ruidos blancos y que, además, son independientes entre sí. La forma concreta que adoptan los distintos polinomios de retardos incluidos en la expresión anterior depende del proceso ARIMA que sigue el indicador, ya que se obtienen como una descomposición de aquél. Una vez que disponemos de esta descomposición, podemos estimar cada uno de los componentes mediante el uso de procedimientos como el filtro de Kalman. No obstante, para una realización determinada de un indicador, $x_t = (x_1, x_2, ..., x_T)$, resulta más apropiado estimar cada uno de los componentes mediante la minimización del error cuadrático medio de cada uno de ellos. Este es el procedimiento que está implementado en el programa SEATS y, en consecuencia, es el que se adopta aquí[11].

[10] Se puede consultar Gómez y Maravall (1996) para ver detalles sobre ambos programas.

[11] Una descripción más detallada de este método se ofrece en Kaiser y Maravall (2001), entre otros.

1.4 Estimación del índice sintético: métodos basados en la extracción de factores comunes

En las secciones anteriores se ha preparado una base de datos que ofrezca información acertada sobre la evolución de la economía objeto de estudio. Con esta información se puede elaborar un indicador sintético que refleje cómo se ha comportado la economía a lo largo de un determinado periodo de tiempo. Para ello, se dispone nuevamente de diversas opciones, aunque nos guiaremos por el método habitualmente utilizado por el NBER (National Bureau of Economic Research). Sus puntos clave se presentan en la siguiente sección, que particularizamos para la economía de Zaragoza.

1.4.1 Metodología general

Investigadores y grupos de trabajo han ideado diversos procedimientos para construir indicadores sintéticos que recojan fidedignamente la evolución de una determinada economía. El método seleccionado, de gran repercusión en el ámbito de los estudios de coyuntura tiene como origen los trabajos de Stock y Watson (1989,1993)[12]. El espíritu que subyace en su propuesta es el siguiente. Supongamos que disponemos de una serie de n indicadores parciales de la actividad de una determinada economía (como producción, desempleo, consumo, etc), tales que admitimos que cada uno de ellos está compuesto por dos partes claramente diferenciadas. La primera representa un conjunto de p factores comunes a todos estos indicadores, mientras que la segunda representa el comportamiento propio o factor idiosincrático de estas series. Claramente, podemos asociar estos factores comunes con los co-movimientos de los que hemos hablado con anterioridad, por lo que para la estimación del indicador sintético basta con la extracción de estos factores comunes. Desde el punto de vista técnico, esto tan sólo cuenta con el inconveniente de que dichos factores comunes no son directamente observables. No obstante, Stock y Watson (1989) resuelven el problema de la siguiente manera. Primero, su planteamiento es como sigue:

$$y_t = \beta + g(L)\, F_t + u_t \qquad (1.4.1)$$

12 Ver también Crone y Babyak (1996).

$D(L)\, u_t = \varepsilon_t$ (1.4.2)

$F(L)\, F_t = \delta + \eta_t$ (1.4.3)

donde y es el conjunto de indicadores parciales utilizados, F representa el conjunto de factores comunes y u es una perturbación aleatoria, que refleja el factor idiosincrásico. Del mismo modo, L es el operador matemático de retardos, D(L), g(L) y F(L) representan los correspondientes matrices de polinomios de retardos, siendo ε y η sendos ruidos blancos. La única restricción que se impone en el sistema es que los factores comunes y los idiosincrásicos están mutuamente incorrelacionados.

Una vez planteado el problema, la estimación de este sistema es sencilla a partir de la aplicación de métodos como el conocido por filtro de Kalman [ver Harvey (1989) para una descripción de esta metodología]. No resulta tan sencillo determinar el número de factores comunes que existen. Esta cuestión se puede resolver empleando diferentes alternativas[13]. Sin embargo, la propuesta de Stock-Watson parte de la consideración de un único factor común, que es precisamente el ciclo económico que queremos estudiar. Por tanto, resulta aconsejable mantener dicha restricción e imponer la existencia de un único factor común a los diversos indicadores de la actividad económica que vamos a utilizar.

Hasta aquí se ha realizado una descripción de los aspectos fundamentales de la metodología utilizada para la construcción del indicador sintético de la economía de Zaragoza. No obstante, debemos tener en cuenta que este método está pensado para un escenario como la economía norteamericana donde la disponibilidad muestral no es un problema. Desgraciadamente, como se ha comentado, no podemos decir lo mismo de una economía tan pequeña como es la de Zaragoza. En consecuencia, la aplicación directa de esta metodología al ámbito que nos ocupa no parece apropiada por cuanto nos podríamos encontrar con casos en los que los algoritmos de convergencia que hay que utilizar no convergiesen hacia ningún valor, lo que haría que lo estimadores perdiesen sus buenas propiedades. Para evitar este problema, se propone una ruta alternativa que, respetando el espíritu

[13] Ver Doz y Lenglart (1999).

de la metodología presentada, permita una más fácil aplicación para este caso específico.

1.4.2 Aplicación a la ciudad de Zaragoza

Siguiendo el esquema anterior, el comportamiento de los indicadores parciales (y_1, y_2, ..., y_n), que a partir de ahora suponemos que están tipificados, se puede representar mediante el siguiente sistema:

$$y_{1t} = \lambda_{11} F_{1t} + \lambda_{12} F_{2t} + \ldots + \lambda_{1p} F_{pt} + u_{1t}$$

$$y_{2t} = \lambda_{21} F_{1t} + \lambda_{22} F_{2t} + \ldots + \lambda_{2p} F_{pt} + u_{2t}$$

$$\ldots.$$

$$y_{nt} = \lambda_{n1} F_{1t} + \lambda_{n2} F_{2t} + \ldots + \lambda_{np} F_{pt} + u_{nt}$$

lo cuál no es sino una expansión de (1.4.1), donde F_1, F_2, ..., F_p son los p factores comunes a estos indicadores parciales. De forma compacta, el sistema de ecuaciones anterior lo podemos representar mediante:

$$y_{it} = \lambda_1 F_{1t} + \lambda_2 F_{2t} + \ldots + \lambda_p F_{pt} + u_{it}; \quad i=1, 2, .., n, \quad t = 1, 2, \ldots T \qquad (1.4.4)$$

o, de forma matricial:

$$y = \Lambda F + u \qquad (1.4.5)$$

En el sistema anterior, suponemos que los factores están incorrelacionados entre sí, de forma que toda la correlación común a los n indicadores parciales ha de venir exclusivamente explicada por los p factores. Además, incluimos el supuesto de que poseen varianza unitaria. De esta forma, la importancia relativa de cada factor no se ve alterada por el cambio de unidades de medida. Por tanto, el conjunto de restricciones consideradas son:

$$E(F)=0$$

$$E(u) = 0$$

$$E(u\,u') = D = diag(d_1,d_2,\ldots,d_n)$$

$$E(F\,F') = I$$

$$E(F_t\,u_\tau') = 0$$

Donde y_t = (y_{1t}, y_{2t}, ..., y_{nt})', F_t = (F_{1t}, F_{2t}, ..., F_{pt}) u_t=(u_{1t}, u_{2t}, ..., u_{nt})'.

Bajo estas restricciones, debemos estimar la matriz de ponderaciones Λ. Nuevamente, es preciso señalar que no existe un único método en la literatura. Al contrario, disponemos de diversas alternativas. Una excelente revisión de las mismas la tenemos en Uriel (1995), entre otros. Por ejemplo, una técnica muy empleada para la obtención de estos pesos es la estimación máximo verosímil. Sin embargo, un procedimiento alternativo, menos costoso en términos de disponibilidad muestral es el basado en el uso de componentes principales, que será el que se utilizará.

Para aplicar este método, basta con tener en cuenta que los factores comunes que intentamos extraer, también se pueden expresar de la siguiente manera:

$$F_{it} = \kappa_{i1}\, y_{1t} + \kappa_{i2}\, y_{2t} + \ldots + \kappa_{in}\, y_{nt} \qquad i = 1,2, \ldots, n \qquad t = 1,2, \ldots, T \tag{1.4.6}$$

El primero de estos componentes, que coincide con el factor común que nos interesa, lo podemos expresar matricialmente así:

$$F_1 = Y\, \kappa_i \tag{1.4.7}$$

donde F_1 es un vector de orden Tx1, Y es una matriz de orden T x n y $\kappa_{\square}$ es un vector de orden nx1. Esta primera componente se obtiene maximizando su varianza, condicionada al cumplimiento de que la suma de los cuadrados de los pesos es igual a la unidad. Entonces, resulta inmediato probar que la varianza del primer componente vendrá dada por:

$$\mathrm{Var}(F_1) = \frac{1}{T} F_1^{'} F_1 = \frac{1}{T}\, \kappa_1^{'}\, Y' Y\, \kappa_1 = \kappa_1^{'} \left[\frac{1}{T}\, Y' Y \right] \kappa_1 = \kappa_1^{'}\, R\, \kappa_1 \tag{1.4.8}$$

donde, dado que los indicadores parciales están todos tipificados, la matriz R es la matriz de correlaciones entre los mismos. Entonces, el objetivo es maximizar la cantidad anterior, sujeto a la restricción:

$$\sum_{i=1}^{n} \kappa_{1i}^{2} = \kappa_1^{'}\, \kappa_1 = 1 \tag{1.4.9}$$

Para ello, basta con formar la siguiente función de multiplicadores de Lagrange:

$$L = \kappa_1^{'}\, V\, \kappa_1 - \theta \left(\kappa_1^{'}\, \kappa_1 - 1 \right) \tag{1.4.10}$$

El valor óptimo lo obtenemos derivando el valor del lagrangiano con respecto a κ_1 e igualando a 0:

$$\frac{\partial L}{\partial \kappa_1} = 2\,V\kappa_1 - 2\theta\,\kappa_1 = 0 \qquad (1.4.11)$$

con lo que resulta que

$$(V - \theta I)\kappa_1 = 0 \qquad (1.4.12)$$

De la resolución de la ecuación |V-θ☐I|=0 se obtienen p raíces características. Tomando la raíz característica mayor, se halla el vector característico asociado a mismo vector κ_1. Así, el vector de ponderaciones coincide con este vector característico. A partir de este vector de ponderaciones se puede construir el factor común que estamos buscando a partir de la siguiente relación:

$$\tilde{F}_1 = y\,\tilde{\kappa}_1 \qquad (1.4.13)$$

Por tanto, la multiplicación del vector de indicadores parciales por el vector de componentes estimado proporciona el valor del factor común o, lo que es lo mismo, el valor del indicador sintético de la economía objeto de estudio.

1.5 Corrección de la varianza

Con la metodología aplicada hasta este momento se obtiene un primer resultado del indicador sintético de coyuntura. Sin embargo, es preciso hacer una serie de correcciones. La técnica del análisis factorial dinámico, en la medida que selecciona las series de acuerdo con su contribución a la varianza explicada, tiende a amplificar la volatilidad de las mismas. Este es un problema relevante al trabajar con series en niveles de desagregación bajos y para un periodo muestral reducido, elementos ambos que provocan una elevada erraticidad en las series originales. Por ello, una vez aplicadas las ponderaciones obtenidas en el análisis factorial dinámico hay que corregirlas a través un factor de acuerdo a la inversa de la varianza, y suavizarlas mediante medias móviles. De este modo, el indicador sintético original, una vez corregido, ofrecerá una aproximación más adecuada e informativa de la posición cíclica de la economía zaragozana.

1.6 Ajuste al ciclo económico

El indicador sintético obtenido refleja la evolución del ciclo económico zaragozano pero carece de magnitud. Para ello es preciso ajustarlo a las cifras de crecimiento del VAB real de dicha economía. De nuevo en este punto reaparecen las dificultades de no contar con una serie larga del agregado macroeconómico de referencia. Mientras no se disponga de estimaciones de dichas macromagnitudes una alternativa es tomar un año base donde el índice tomaría el valor de 100. Es preciso señalar, no obstante, que el principal objetivo del indicador sintético no es la estimación del VAB o el PIB de Zaragoza, para lo cual existen técnicas más adecuadas, sino situar la posición cíclica de la economía de la ciudad y avanzar su evolución en un futuro próximo. Como se ha detallado en líneas anteriores su construcción se basa en el concepto de co-movimiento entre los diferentes indicadores de actividad, producción y consumo. De este modo, el ajuste a una determinada tasa de crecimiento debe tomarse como una referencia relativa y nunca como una cifra precisa de crecimiento real.

1.7 Métodos de predicción

El indicador sintético que se puede elaborar siguiendo el procedimiento anteriormente descrito ofrece información atractiva de cara a seguir la evolución de la economía a estudiar. Pero, además de ello, también se puede usar para avanzar cuál será el devenir de la economía analizada en los próximos periodos. Para ello, basta con predecir el valor del índice cierto número de periodos hacia delante.

Metodológicamente existen diversas alternativas para obtener dicha predicción. Por ejemplo, una primera posibilidad es la de utilizar simplemente los valores del índice de cara a conocer su evolución futura, por ejemplo, estimando la siguiente relación:

$$IS_t = \sum_{i=1}^{p} \gamma_i IS_{t-i} + v_t \qquad t= p+1, ..., T \qquad (1.7.1)$$

donde IS_t es la estimación del indicador con la información disponible hasta el periodo T, γ_i son los parámetros que hay que estimar, donde es necesario determinar previamente el valor de p, y v_t es un ruido blanco.

Una segunda alternativa tiene como punto de partida el procedimiento mediante el que se ha construido el propio indicador sintético. Para su obtención, se ha estimado una matriz de pesos que, multiplicada por los indicadores de la actividad económica, proporciona el resultado final del indicador. Entonces, si se conoce el comportamiento futuro de los indicadores, es directo obtener el valor del índice para dichos periodos. De forma analítica, esto se puede expresar como sigue:

$$IS_t = \hat{\kappa}_{11} y_{1t} + \hat{\kappa}_{12} y_{2t} + \ldots + \hat{\kappa}_{1n} y_{nt} \qquad t = T+1, T+2, \ldots, T+h$$

donde $\hat{\kappa}_{1j}$ son las estimaciones a partir de (1.4.12). En este escenario resulta clave disponer de predicciones fiables sobre el comportamiento de los indicadores de coyuntura, ya que son necesarias las observaciones futuras de estas variables. De otra forma, el uso de esta estrategia nos ha de conducir necesariamente a resultados poco fiables. Para obtener dichas predicciones de los valores futuros de los indicadores parciales, recurrimos nuevamente al mencionado programa TRAMO que, a partir de los modelos ARIMA previamente estimados, ofrece de forma automática estos valores.

Por último, señalar que alternativas mixtas, donde se tiene en cuenta tanto la propia correlación del indicador sintético como la propia evolución de los indicadores parciales pueden conducirnos a resultados igualmente satisfactorios, como se comprueba en, por ejemplo, Megna y Xu (2002) o Clayton-Matthews y Stock (1999).

2. Estimación del PIB

El método empleado se puede definir analíticamente de la siguiente manera. Supongamos que y^A representa la macromagnitud de una economía, siendo Δy^A su correspondiente tasa de crecimiento. Admitamos que dicha tasa de crecimiento viene explicada en función del comportamiento de una serie de indicadores.

Entonces, resulta cierto que:

$$\Delta y^A_{t-j} = \phi_{oo} + \sum_{i=1}^{\ell} \sum_{j=0}^{k} \phi_{ij} \Delta x^A_{i,t-j} + u^A_t \quad (2.1)$$

donde $\Delta x^A_{i,t}$ es la tasa de crecimiento del indicador x_i de la economía en el periodo t, u^A_t es una perturbación aleatoria que asumimos que sigue una distribución niid(0, σ^2) y ϕ_{ij} (i=0, 1, 2, … ℓ; j=0,1,2, …, k) son los parámetros del modelo que determinan la influencia que el retardo j-ésimo del indicador i-ésimo tiene sobre la evolución de la tasa de crecimiento de la macromagnitud y^A.

Si se traslada el modelo anterior al caso de la economía de la ciudad de Zaragoza, el comportamiento de las macromagnitudes zaragozanas se puede expresar de la siguiente manera:

$$\Delta y^Z_t = \gamma_{oo} + \sum_{i=1}^{\ell} \sum_{j=0}^{k} \gamma_{ij} \Delta x^Z_{i,t-j} + u^Z_t \quad (2.2)$$

donde Δy^Z_t es la tasa de crecimiento de la macromagnitud y en la economía zaragozana en el periodo t, $\Delta x^z_{i,t}$ es la tasa de crecimiento del indicador x_i en la economía zaragozana en el periodo t, u^Z_t es la perturbación aleatoria que tiene las misma propiedades que en el modelo precedente y γ_{ij} son los parámetros del modelo. Si tenemos en cuenta la hipótesis de que existe una elevada correlación entre las economías zaragozana y aragonesa (que se podría representar según el primer modelo), se puede considerar que los parámetros de los modelos (2.1) y (2.2) son idénticos. Si introducimos la restricción $\gamma_{ij}=\phi_{ij}\ \forall$ i,j, entonces es directo demostrar que:

$$\Delta y_t^Z = \Delta y_t^A + \sum_{i=1}^{\ell} \sum_{j=0}^{k} \phi_{ij} \left(\Delta x_{i,t-j}^Z - \Delta x_{i,t-j}^A\right) + \left(u_t^Z - u_t^A\right) \tag{2.3}$$

Este modelo nos permite obtener el valor de la tasa de crecimiento de la macromagnitud y^Z en el periodo t en función del valor de la tasa de crecimiento de esta macromagnitud en la primera economía y del diferencial de la evolución de las tasas de crecimiento de los diferentes indicadores. Si, como es habitual, desconocemos los valores de los parámetros, estos deben ser estimados. Para ello, basta con la simple aplicación de mínimos cuadrados ordinarios al modelo (2.1).

Finalmente, para obtener la predicción de la macromagnitud en el periodo T+h, donde T es el año base (el que marca el límite de nuestra información muestral), mientras que h es el horizonte temporal (número de periodos hacia delante sobre los que queremos obtener la predicción), se debe tener en cuenta que el predictor óptimo de z_{T+h} es siempre $E(z_{T+h}/I_T)$, siendo I_T el conjunto de información disponible. En consecuencia, la mejor predicción de los valores futuros de la tasa de crecimiento de la macromagnitud y para la economía aragonesa viene dada por:

$$\Delta \hat{y}_{T+h}^Z = \Delta \hat{y}_{T+h}^A + \sum_{i=1}^{\ell} \sum_{j=0}^{k} \phi_{ij} \left(\Delta \hat{x}_{i,T+h-j}^Z - \Delta \hat{x}_{i,T+h-j}^A\right) \tag{2.4}$$

Por tanto, para obtener la predicción para el periodo T+h de la tasa de crecimiento de la macromagnitud y para la economía zaragozana es preciso conocer también el valor de la tasa de crecimiento para el periodo T+h de esta macromagnitud para la economía aragonesa además de los valores de los indicadores (zaragozanos y aragoneses) en el periodo T+h y en sus diversos retardos. En ocasiones, es posible que estos valores no sean conocidos, en cuyo caso se hace imprescindible su predicción. Esta se puede llevar a cabo mediante la estimación de un modelo ARIMA o de un VAR.

Como se puede apreciar, la estimación del modelo (2.1) resulta clave en el esquema. Una vez obtenida una buena estimación de este modelo, el resto del procedimiento resulta inmediato. Sin embargo, hay factores que pueden alterar el grado de confianza en los resultados de la estimación del modelo (2.1). El primero de ellos es el hecho de que no siempre se dispone de toda la información necesaria. En este sentido, debemos señalar que resulta habitual no disponer de series suficientemente amplias de todos los indicadores que pensamos que pueden incluirse en la especificación del modelo. Ante esta disyuntiva, caben dos salidas. Omitir estos indicadores, con el consiguiente riesgo de que los estimadores del modelo pierdan parte de sus buenas propiedades (insesgadez) o bien utilizar solamente aquel periodo muestral que permite la inclusión de todos los indicadores que, a nuestro juicio, tienen influencia sobre la evolución de la macromagnitud que queremos estudiar. En este último caso, se corre el riesgo de disminuir seriamente los grados de libertad del modelo, reduciendo la confianza en los resultados.

Otro problema que puede plantearse a la hora de estimar el modelo (2.1) es la aparición de problemas graves de multicolinealidad. Este problema se presenta cuando los regresores del modelo están muy correlacionados entre sí. Sus consecuencias son una falta de precisión en la estimación de los parámetros del modelo. En este tipo de estudios en los que se tiende a utilizar un elevado número de regresores que miden una misma parte de la realidad, la presencia de problemas de multicolinealidad grave es bastante frecuente. De nuevo, las soluciones no son sencillas. Si eliminamos un conjunto de regresores en aras de disminuir el grado de correlación entre los regresores, entonces podemos estar omitiendo una variable relevante, lo que implica que los estimadores dejan de tener buenas propiedades. Por otro lado, ignorar los problemas de multicolinealidad implica que no podemos confiar en los resultados de la estimación del modelo.

Para solucionar estos problemas, se ha optado por seguir el método propuesto en Granger y Newbold (1986), que también ha sido aplicado por Raymond, Matas y Pujolar (1998,1999) para la predicción de la evolución de las macromagnitudes de la economía catalana.

Bajo este método, en lugar de estimar un solo modelo en el que la evolución de la tasa de crecimiento de la macromagnitud de la economía aragonesa es función de ℓ indicadores más sus respectivos retardos, estimamos ℓ modelos (uno por cada indicador) donde esta tasa de crecimiento es función exclusiva del i-ésimo indicador (i=1, 2, ..., ℓ) y de sus retardos. Es decir, estimamos el siguiente modelo:

$$\Delta y_t^A = \alpha_i + \sum_{s=1}^{3} \delta_{i,s} D_s + \sum_{j=0}^{k} \phi_{ij} \Delta x_{i,t-j}^A + u_t \qquad (2.5)$$

donde D_s (s=1,2,3) es una variable dummy estacional y la inclusión del subíndice i en los parámetros del modelo (2.5), indica que estamos considerando únicamente la presencia del indicador i-ésimo (i=1,2,…, ℓ). La estimación del modelo (2.5) proporciona una estimación de Δy_t^A, que se denota como $\Delta \hat{y}_{i,t}^A$, mientras que la varianza estimada del modelo se representa mediante $\hat{\sigma}_i^2$. Este procedimiento lo debemos repetir para los distintos indicadores, lo que da lugar a un conjunto de estimaciones de los valores de la tasa de crecimiento de la macromagnitud y de desviaciones típicas, cada uno de ellos asociado a uno de los indicadores. A partir de esta información, se puede obtener la estimación promedio Δy_t^A a partir de la siguiente fórmula:

$$\Delta \hat{y}_t^A = \sum_{i=1}^{\ell} b_i^A \, \Delta \hat{y}_{i,t}^A \qquad (2.6)$$

donde

$$b_j^A = \frac{1/\hat{\sigma}_j^2}{\sum_{i=1}^{\ell} 1/\hat{\sigma}_i^2} \qquad (2.7)$$

Los valores predichos para la evolución de las tasas de crecimiento de la economía zaragozana se asume de nuevo la hipótesis de que la economía aragonesa y la zaragozana están altamente correlacionadas. Entonces, tenemos que:

$$\Delta y_t^Z = \alpha_i + \sum_{s=1}^{3} \delta_{i,s} D_s + \sum_{j=0}^{k} \phi_{ij} \Delta x_{i,t-j}^Z + u_t \qquad (2.8)$$

y, adicionalmente,

$$\Delta y_t^Z = \Delta y_t^A + \sum_{j=0}^{k} \phi_{ij} \left(\Delta x_{i,t-j}^Z - \Delta x_{i,t-j}^A\right) + \left(u_t^Z - u_t^A\right) \qquad (2.9)$$

A partir de (2.9) resulta directo obtener la estimación de $\Delta \hat{y}_{i,t}^Z$ sin más que considerar que en muestras suficientemente grandes la influencia de las perturbaciones se anula y, adicionalmente, las estimaciones de los parámetros del modelo se aproximan a los valores reales de los parámetros. En estas circunstancias, la predicción del valor futuro de la macromagnitud de la economía zaragozana viene dada por:

$$\Delta \hat{y}_t^Z = \sum_{i=1}^{\ell} b_i^A \Delta \hat{y}_{i,t}^Z \qquad (2.10)$$

donde b_i^A se obtiene a partir de (2.7).

Referencias bibliográficas

Chow G. and A.L. Lin (1971), Best linear unbiased interpolation, distribution and extrapolation of time series by related series, *The Review of Economics and Statistics*, 53: 372-375.

Clayton-Matthews, A. y Stock, J.H. (1999). "An application of the Stock-Watson index methodology to the Massachussets economy", Journal of Economic and Social Mesaurement, 25, 183-223.

Crone, T. M., y K. J. Babyak. (1996). "Looking Ahead: Leading Indexes for Pennsylvania and New Jersey." Federal Reserve Bank of Philadelphia *Business Review*,May/June, pp. 3-14.

Doz, C. y Lenglart, F.,(1999), Analyse factorielle dynamique: test du nombre de facteurs, estimation et application l'enqu^ete de conjoncture dans l'industrie, Annales d'Economie et de Statistique, 54, 91-127.

Gómez, V. y Maravall, A. (1996). "Programs TRAMO (Time series Regression with Arima noise, Missing observations, and Outliers) and SEATS (Signal Extraction in Arima Time Series). Instructions for the User", Working Paper 9628, Servicio de Estudios, Banco de España.

Granger, C.W. and Newbold, P. (1986). *Forecasting economic time series*. Academic Press.

Harvey, A. (1989), Forecasting, structural time series models and the Kalman filter, Cambridge University Press.

Kaiser, R. y Maravall, A. (2001), *Measuring Business Cycles in Economic Time Series*, Lecture Notes in Statistics 154, NY: Springer-Verlag.

Megna, R. y Xu, Q. (2002). Forecasting the New York State Economy: The Coincident and Leading Indicators Approach. Mimeo.

Raymond, J. L., A. Matas y D. Pujolar (1998). Estudis económics, Documento de trabajo, Generalitat de Catalunya.

Raymond, J. L., A. Matas y D. Pujolar (1999). Evolución económica en Cataluña en 1998. Cuadernos de Información Económica, 147, 88-95.

Stock, J.H. y Watson, M.W. (1989) - New indexes of coincident and leading indicators, in NBER Macroeconomics Annual, Blanchard and Fisher Ed. MIT Press, Cambridge.

Stock, J.H. y Watson, M.W. (1993) - A procedure for predicting recessions with leading indicators: econometric issues and recent experience, in Business cycles, indicators and forecasting, J.H. Stock and M.W. Watson Ed., University of Chicago Press.

Anexo

Datos estadísticos de la ciudad de Zaragoza

Mercado de trabajo			
Variable	**Descripción**	**Periodicidad**	**Primer dato disponible**
Afiliados	Afiliados a la Seguridad Social según sector de actividad económica	Mensual	Enero de 2002
Paro	Paro registrado según sector de actividad económica	Mensual	Enero de 1997
Contratos	Contratos registrados según sector de actividad económica	Mensual	Mayo de 2005
EPA1	Población de 16 o más años según sexo	Trimestral	Primer trimestre 2000
EPA2	Activos según sexo	Trimestral	Primer trimestre 2000
EPA3	Ocupados según sexo	Trimestral	Primer trimestre 2000
EPA4	Ocupados según sector económico	Trimestral	Primer trimestre 2000
EPA5	Parados según sexo	Trimestral	Primer trimestre 2000
Suelo y vivienda			
Variable	**Descripción**	**Periodicidad**	**Primer dato disponible**
Transacciones	Transacciones inmobiliarias y de suelo	Trimestral	Primer trimestre 2004
Tasaciones	Número de tasaciones	Trimestral	Primer trimestre 2005
Precio1	Precio de la vivienda libre	Trimestral	Primer trimestre 2005
Precio2	Precio medio del metro cuadrado de suelo urbano	Trimestral	Primer trimestre 2004
Hipotecas	Número e importe de hipotecas constituidas	Mensual	Enero de 2007
Licencias	Licencias de construcción de viviendas	Mensual	Enero de 2006
Turismo y transporte			
Variable	**Descripción**	**Periodicidad**	**Primer dato disponible**
EHO	Viajeros entrados, pernoctaciones, grado de ocupación, estancia	Mensual	Algunas enero de 2009
	media, personal empleado, número de establecimientos abiertos		y otras enero de 2005
	estimados y número de plazas estimadas		
Aeropuerto1	Aeropuerto de Zaragoza. Tráfico de pasajeros	Mensual	Enero de 1991
Aeropuerto2	Aeropuerto de Zaragoza. Tráfico de aeronaves	Mensual	Enero de 1991
Aeropuerto3	Aeropuerto de Zaragoza. Tráfico de mercancías	Mensual	Enero de 1991
TransptPco	Viajeros transportados y kilometros recorridos por los autobuses público	Mensual	Enero de 2006
Medio ambiente			
Variable	**Descripción**	**Periodicidad**	**Primer dato disponible**
Agua1	Volúmenes de agua potable puestos en la red	Mensual	Enero de 2006
Agua2	Consumo de agua según uso	Mensual	Enero de 2007
Residuos	Recogida total y selectiva de residuos urbanos	Mensual	Enero de 2008