

REVISTA DE ESTUDIOS REGIONALES

I.S.S.N.: 0213-7585

2ª EPOCA Septiembre-Diciembre 2020



119

SUMARIO

Esther López-Vizcaino, Patricio Sanchez-Fernandez y Carlos L. Iglesias Patiño. Monitorización de la coyuntura económica regional a través de un indicador sintético.

Mónica Benito y Rosario Romera. ¿Cantidad o calidad de la educación? Un análisis por Comunidades Autónomas.

Miguel González-Mohino Sánchez, María Ángeles Rodríguez Domenech y Ana Isabel Callejas Albiñana. Patrones de conocimiento escolar sobre el patrimonio local en ciudades de mediano tamaño de Castilla-La Mancha. El caso de Ciudad Real.

Juan José Díaz Hernández y José Ignacio Estrán Ramírez. Patrón de especialización productiva y Valor Añadido en el sistema portuario español.

Antonio Fernández Morales y María Cruz Mayorga Toledano. Caracterización y concentración de la oferta de Airbnb en Málaga.

María Jesús García García. Direct democracy and lawmaking: Initiatives and referenda at local level in the Usa.

Textos

I. Artículos

Monitorización de la coyuntura económica regional a través de un indicador sintético

Monitoring of the regional economic situation through a synthetic indicator

Esther López-Vizcaíno¹

Instituto Galego de Estatística

Patricio Sanchez-Fernandez

GEN - Universidad de Vigo

Carlos L. Iglesias Patiño

Instituto Galego de Estatística

Recibido, Agosto de 2018; Versión final aceptada, Junio de 2019.

PALABRAS CLAVE: Indicadores sintéticos, Coyuntura económica, Economía regional, Análisis factorial dinámico, Modelos aditivos generalizados

KEYWORDS: Synthetic indicators, Business cycle, Regional economy, Dynamic factor analysis, Generalized additive models.

Clasificación JEL: O18, R11, C23.

RESUMEN:

En el presente trabajo se obtiene un indicador sintético que pretende proporcionar una herramienta para el seguimiento de la coyuntura económica de una región. Siguiendo la literatura existente el indicador sintético se construye mediante la aplicación a los datos coyunturales del modelo factorial dinámico que permite reducir la dimensionalidad de los datos iniciales. De este modo se obtiene una interpretación más simple y compacta a través de un conjunto reducido de factores comunes. El indicador obtenido se utiliza para el caso de Galicia, sus resultados permiten seguir la evolución, con periodicidad mensual, de la economía gallega de manera consistente con la evolución trimestral del PIB calculado por la oficina de estadística regional. Además, el estimador sintético calculado se puede utilizar para predecir el PIB a través de un modelo aditivo generalizado (GAM) en el que se tiene en cuenta la estructura de correlación de los datos.

1 Los autores desean agradecer la ayuda inestimable de los profesores Santiago Lago Peñas, Fernando González Laxe y dos evaluadores/as anónimos quienes, con sus aportaciones y sugerencias han contribuido de manera determinante a la mejora del presente trabajo.

ABSTRACT:

This paper builds a synthetic indicator that aims to provide a tool for monitoring the economic situation of a region. Thus, a synthetic indicator is constructed by applying to the economic data a dynamic factorial model that allows reducing the dimensionality of the initial data.

To guarantee its technical solvency, the synthetic indicator developed considers the methodological developments pointed out by Stock & Watson (1991). In this way, we proceed by applying the dynamic factorial model to the economic series. This sort of indicators constructed using dynamic common factor models aims to represent a relatively large set of initial series by means of a smaller set that achieves a simpler and more compact interpretation.

These indicators must fulfill the characteristics pointed out by Burns & Mitchell (1946), Artís *et al.* (1997) or Cuevas & Quilis (2010, 2015). Those are the following characteristics: i.- Length of the series; ii.- Economic significance; iii.- Statistical quality; iv.- Smooth profile; v.- Speed in the availability of information, and vi.- Monthly frequency. By doing this, a simpler and more compact interpretation is got through a reduced set of common factors.

The variables finally included in the indicator are:

- 1.- Consumption of gasoline and diesel;
- 2.- Exports of goods deflated by the Export Unit Value Index;
- 3.- Imports of goods deflated by the Import Unit Value Index;
- 4.- Surface of the houses to create a new plant;
- 5.- Registration of tourism vehicles;
- 6.- Production of tourism vehicles;
- 7.- Air transport of passengers;
- 8.- Maritime transport of goods (loaded and unloaded);
- 9.- Housing visas;
- 10.- Index of turnover in the industry deflated by the Industrial Price Index (IPRI);
- 11.- People registered in the Social Security System;
- 12.- Industrial Production Index deflated by the IPRI (linked series);
- 13.- Retail trade sales index;
- 14.- Travelers entering the hotels; and
- 15.- Index of turnover in the services sector deflated by the services Consumer Price Index (CPI).

The MARSS R package is used to estimate the indicator (Holmes *et al.*, 2014). The functions of the package estimate the parameters of the model (Z and R) by maximum likelihood using the Expectation Maximization (EM) algorithm. The EM algorithm of Dempster *et al.* (1977) was used for this type of models by Shumway & Stoffer (1982) & Watson & Engle (1983). For the estimation of the factors, the Kalman filter is used once the model is represented in the state space.

The results show the number of common factors that are needed to represent most of the variability of the original series. To do this, the eigenvalues of the sample autocovariance matrices are calculated. Likewise, with the objective of finding the co-integration range “s”, the Johansen test (1991) was carried out to the nine series in which there is information since 1995. This test concludes that, with a significance level of 5%, the co-integration range is 8 and, consequently, the series share a total of 1 trend in common. Then, the factorial loads are obtained in the series used, which are all positive and are used to calculate the common factor. The variable with the highest load is relative to the people registered in the Social Security System. Once the estimation of the synthetic indicator has been carried out, it is important to analyze the explanatory capacity of the indicator on quarterly GDP growth as a commonly accepted benchmark indicator. This analysis shows the existence of cyclical coherence and stability in the relationship that will allow accepting the common factor as a valid synthetic indicator.

Finally, once the indicator has good properties to be a good predictor of GDP, a model is established that relating these two indicators and allows GDP to be predicted from the synthetic indicator.

The results obtained allow us to monitor the monthly evolution of a regional economy in a way consistent with the quarterly evolution of GDP. In addition, the estimated synthetic estimator can be used to predict GDP through a generalized additive model (GAM) in which the correlation structure of the data is taken into account.

This work has several advantages when comparing to similar proposals in previous literature. We can summarize this advantages in the following four. Firstly, it is necessary to point out the good performance of the indicator when compared with the evolution of GDP. The high degree of coupling between the evolution of both magnitudes (synthetic indicator and GDP), determined by the correlation coefficient, shows their cyclical coherence and stability. The second advantage lies in the simplicity of the procedure used. It is a simpler procedure with fewer hypothesis assumptions than in most papers with similar characteristics and, on the other hand, the empirical results are very similar. A third advantage of the calculated synthetic indicator is that it serves as an indicator of the evolution of the economy, nowadays the periodicity of GDP is quarterly, so in any quarter, two months in advance and by calculating the synthetic indicator and establishing the function of relationship between the two indicators by the proposed GAM model, it will be possible to anticipate the evolution of GDP in that quarter. Finally, the fitting of the polynomial of degree three, that is slightly worse than that made by “splines”, has the advantage that allows to give an approximation of the inflection point, in this case when $IS = 1,77$, which can be used as a control value in the monitoring of the conjuncture.

Therefore, an easily replicable methodology is presented, both with the data available in another region or with the data available at national level.

1. INTRODUCCIÓN

Una de las consecuencias de la reciente crisis económica viene dada por la necesidad de realizar análisis de coyuntura que permitan superar los inevitables retrasos y pérdidas de información que se producen con el uso de medidas estándar de la actividad económica, basados en los macroagregados económicos de la contabilidad nacional. Como señalan Cuevas & Quilis (2012) se requiere información oportuna y precisa sobre las condiciones macroeconómicas generales que evite el obligado retraso de las medidas macroeconómicas más estándares. Por ejemplo, en el caso de Galicia los datos de las cuentas económicas son trimestrales y se publican con dos meses de desfase. Un mes antes de que los datos se publiquen, ya estarían disponibles muchos de los indicadores mensuales que se utilizan para estimar el dato proporcionado por la contabilidad.

En este sentido, cada vez es más frecuente el diseño de indicadores a muy corto plazo que permitan evaluar la evolución de la economía con los últimos datos disponibles hasta el momento actual. Para ello se debe ampliar la base sobre la que descansa el análisis, utilizando un conjunto suficientemente amplio de variables de actividad económica que permitan su actualización continua, como recomiendan Stock & Watson (2002).

El problema de trabajar con muchas variables o series de tiempo es que los datos están correlacionados dentro de la misma serie y entre las series. Si se quieren modelizar utilizando los modelos de vectores autorregresivos de medias móviles (VARMA), el número de parámetros crece con el cuadrado del número de series considerado, por lo que es habitual que se necesite un número de parámetros demasiado grande (Peña & Poncela, 2004). Por tanto, si se logra reducir el número de series, se puede simplificar considerablemente la estructura del modelo (Zhang, 2009). Así mismo, resulta importante descubrir los factores latentes en series económicas para la monitorización de la coyuntura económica de un territorio.

Este tipo de indicadores tienen su precedente a mediados del siglo XX, con los indicadores cíclicos de Burns & Mitchell (1946), que fueron continuados más recientemente por Stock & Watson (1991) aunque también conviene mencionar como pioneros, los barómetros económicos de entre guerras de los institutos de coyuntura con sus éxitos y fracasos (Gabisch & Lorenz, 2013). Desde entonces, son frecuentes los trabajos de este tipo realizados por instituciones tanto oficiales como académicas. Entre estos hay que des-

tacar los trabajos para la zona euro de Marcellino *et al.* (2003) y Angelini *et al.* (2008), para la economía EE.UU. de Evans *et al.* (2002), Giannone *et al.* (2008) y Aruoba *et al.* (2009) y para la economía argentina de Muñoz & Trombetta (2015). Para el caso español destacan las propuestas de Camacho & Pérez-Quirós (2009), que es ampliada para el contexto europeo, y las de Cuevas & Quilis (2010, 2015) y Camacho & Doménech (2011).

En la línea de estos indicadores, se enmarca el presente trabajo de diseño y construcción de un indicador sintético a escala regional. Su pretensión final reside en proporcionar una herramienta para el seguimiento en el corto plazo de la coyuntura económica de una región mediante la extracción de las señales comunes de un conjunto de variables de la economía (Cendejas *et al.* 2014). En concreto, para este trabajo se toma como referencia territorial Galicia donde se consigue representar las series iniciales mediante un conjunto reducido de factores comunes, lo que permite ganar interpretabilidad. Se trata, en definitiva, como apuntan Cendejas *et al.* (2011), de extraer para la economía regional una señal cíclica, no directamente observable, compartida por un conjunto de indicadores de actividad observados.

A diferencia de otros trabajos con un objetivo similar (Cuevas & Quilis, 2012, Camacho *et al.*, 2009, 2011), la metodología empleada en este trabajo es más sencilla y los resultados empíricos son muy similares, lo que da, en definitiva, validez al indicador calculado. También, en este trabajo se presentan sendas expresiones para el cálculo de tasas de variación interanuales e intertrimestrales del indicador trimestralizado, que permiten su comparación con el Producto Interior Bruto (PIB) regional calculado por la oficina de estadística regional, indicador de evolución comúnmente aceptado en la literatura económica. Además, como novedad en este trabajo, se permite predecir el PIB mediante la relación de éste con el indicador sintético a través de un modelo aditivo generalizado (GAM), que tiene en cuenta la estructura de correlación de los datos.

Teniendo en cuenta estos precedentes, lo que resta del trabajo se organiza de la manera siguiente. En la sección 2, se desarrolla la metodología empleada en este trabajo, especialmente la relacionada con el cálculo del indicador sintético mediante el análisis factorial dinámico. A continuación, se exponen los resultados de la aplicación práctica con los datos de Galicia y, por último, se presentan las conclusiones más relevantes del estudio realizado.

2. DESARROLLO METODOLÓGICO

Con el objetivo de obtener un indicador alternativo y a la vez complementario a las Cuentas económicas de Galicia, las series que se emplearán en este trabajo son series mensuales, por tanto, el indicador coyuntural también será mensual.

Para garantizar su solvencia técnica, el indicador sintético diseñado parte de los desarrollos metodológicos señalados en el apartado precedente y, de manera particular del trabajo de Stock & Watson (1991). De este modo se procede mediante la aplicación a las series económicas coyunturales del modelo factorial dinámico. Como señalan Stock & Watson (2010), los indicadores construidos mediante modelos de factores comunes dinámicos tienen como objetivo representar un conjunto relativamente grande de series iniciales mediante un conjunto de tamaño inferior que logra una interpretación más simple y compacta.

Por lo que se refiere a la selección de las variables, en línea con lo apuntado por Mondéjar & Vargas (2008), este paso constituye una de las principales dificultades a la hora de construir un indicador sintético como el que aquí se pretende. En este sentido es preciso disponer de una amplia base de datos (indicadores parciales) para el territorio de referencia que permitan captar las fluctuaciones periódicas del corto plazo de la economía objeto de análisis (Artís *et al.*, 1997). Estos indicadores parciales deben cumplir las características apuntadas por Burns & Mitchell (1946), sobre las que insisten autores como Artís *et al.* (1997) o Cuevas & Quilis (2010, 2015), que se concretan en las características siguientes:

1. Longitud de las series,
2. Significación económica,
3. Calidad estadística,
4. Perfil suave,
5. Rapidez en la disponibilidad de la información,
6. Frecuencia mensual

Las series iniciales de las que se parte se encuentran recogidas en el Cuadro 1. Todas ellas cumplen los requisitos señalados, proporcionando una medida sintética de algún aspecto de la evolución de la coyuntura económica. En general se eligió como periodo de inicio en la mayor parte de las series el mes de enero del año 1995. De todos modos, en algunas

de las series el período de inicio disponible en las fuentes de información es posterior. Por tanto, en este trabajo hay un problema de paneles de datos no equilibrados (Cuevas & Quilis, 2015).

CUADRO 1
SERIES UTILIZADAS EN EL ANÁLISIS²

Variable	Abrev.	Fuente	Año de inicio	Unidad
Consumo de gasolina y gasóleo	GASOL	Ministerio de Industria	1995	Miles de toneladas
Exportaciones de bienes deflactadas por el Índice de valor unitario de exportación	CXX	AEAT/IGE	1995	Euros del 2010
Importaciones de bienes deflactadas por el Índice de valor unitario de exportación	CXM	AEAT/IGE	1995	Euros del 2010
Superficie de las viviendas a crear de nueva planta	EDVIV	IGE	1995	Metros cuadrados
Matriculación de vehículos turismo	MAT	DGT	1995	Número
Producción de vehículos turismo	PROD-VE	PSA/IGE	2003	Número
Transporte aéreo de pasajeros	TAP	MF	1995	Número
Transporte marítimo de mercancías (cargadas + descargadas)	TMM	Puertos del Estado	1995	Miles toneladas
Viviendas visadas	VIV	MF	2000	Número
Índice de cifra de negocios en la industria deflactado por el Índice de Precios Industriales (IPRI)	ICNI	INE	2002	Índice deflactado
Afiliaciones a la Seguridad Social	ASS	MTAS/IGE	1995	Número
Índice de Producción Industrial deflactado por el IPRI (serie enlazada)	IPI	INE/IGE	1995	Índice deflactado
Viajeros entrados en los hoteles	VIAJ	INE/IGE	1995	Número
Índice de cifra de negocios en el sector servicios deflactado por el IPC de los servicios	ICNSS	INE	2005	Índice deflactado

Fuente: Elaboración propia.

2 Nota: IGE: Instituto Galego de Estatística, AEAT: Agencia Española de Administración Tributaria, MF: Ministerio de Fomento, DGT: Dirección General de Tráfico, INE: Instituto Nacional de Estadística, MTAS: Ministerio de Trabajo y Asuntos Sociales

Para que la participación de las series originales en el indicador sintético tenga sentido económico, todas las series deben estar expresadas en términos reales. Por tanto, las variables que vienen expresadas en precios corrientes se deflactan por su correspondiente índice de precios para que queden expresadas en términos reales.

Las series empleadas en este trabajo están sujetas a variaciones estacionales. Como el indicador sintético tiene como objetivo funcionar como una *proxy* del PIB, es necesario eliminar esta estacionalidad y el efecto calendario. Este proceso se hace utilizando el filtro X12-ARIMA desarrollado por el United States Census Bureau (2002) y de uso habitual en la literatura. Después de aplicar el filtro nos quedaremos con la ciclo-tendencia obtenida de este procedimiento. Se utiliza este filtro porque ejerce poca suavización sobre los datos, lo que lleva a que no se pierda la variabilidad existente en las series originales (Muñoz & Trombetta, 2015).

Con el objetivo de emplear en el análisis series estacionarias se tomaron las primeras diferencias de las series corregidas en logaritmos. Estas diferencias se aproximan a la tasa de variación mensual. A todas estas series diferenciadas en logaritmos se les aplicó el test de Dickey y Fuller obteniendo que todas son estacionarias.

El objetivo consiste en explicar la variación temporal de un conjunto de n series observadas usando una combinación lineal de m paseos aleatorios desconocidos, donde $m \ll n$. Por tanto, el modelo de análisis factorial dinámico (DFA) empleado tiene la siguiente estructura:

Fórmula 1:

$$\begin{aligned} y_t &= Zx_t + v_t, v_t \sim \text{MVN}(0, R) \\ x_t &= x_{t-1} + w_t, w_t \sim \text{MVN}(0, I) \end{aligned}$$

donde *MVN* denota la distribución normal multivariante

La idea general es que las observaciones (y) se modelen como una combinación lineal de tendencias desconocidas (x) de dimensión m y de cargas factoriales (Z) de dimensión $n \times m$. Por tanto, la estructura dinámica viene reflejada a través de los factores comunes x y w se refiere a la dinámica particular de cada serie. El elemento Z_{ij} de la matriz Z representa el peso del j -ésimo factor sobre la i -ésima serie observada. Al igual que en el análisis fac-

torial clásico la primera parte de la ecuación 1 no está determinada de manera unívoca, pues para cualquier matriz invertible de tamaño $m \times m$, H , si se define $Z=ZH$ y $x_t^*=H^{-1}x_t$ se tiene que $y_t=Zx_t^*+w_t$ y se logra la misma representación con nuevos parámetros. Para evitar este problema de identificación se toma Z ortogonal ($Z^T Z=I$) y la matriz R se asume con valores desconocidos.

El problema que presentan las series que se están empleando es que no tienen la misma amplitud, es decir, muchas series empiezan demasiado tarde y algunas series finalizan demasiado pronto. Así, por ejemplo, mientras el Índice de Producción Industrial está disponible desde enero de 1995, el Índice de Cifra de negocios del Sector Servicios sólo está disponible desde enero del año 2005. Un posible ejemplo de la disponibilidad de la información se encuentra en el Cuadro 2

CUADRO 2
EJEMPLO DE DISPONIBILIDAD DE LA INFORMACIÓN EN LAS SERIES

Observaciones	1	2	3	4	T_0	T_1	T_2
Series									
y_1	d	d	d	d	d	d	..
y_2	d	d	d	d	d
y_3	d	d	d	d	d	d	d	d	d
y_4	d	d	d	d	d	d	d	d	..
y_5	d	d	d	d	d	d	d	d	d
...	..	d	d	d	d	d	d
y_n	..	d	d	d	d	d	d	d	d

Nota: d = dato disponible, “..” = dato no disponible

Fuente: Elaboración propia.

La zona en gris es el panel transversal (o común) para el que se dispone de la información completa para todas las series (desde $t=3$ hasta $t=T_0$) y la zona horizontal de las series y_3, y_4 e y_5 es el panel longitudinal para el cual se dispone de información desde $t=1$ hasta $t=T_1$.

Por tanto, para la estimación del indicador sintético final se seguirá un procedimiento iterativo que está basado parcialmente en el empleado por Cuevas & Quilis (2012), aunque el modelo factorial dinámico empleado es más sencillo. Los pasos que se siguen en el procedimiento iterativo son los siguientes:

1. Se estima el modelo factorial dinámico (1) para las series del panel longitudinal. En estas series se dispone de datos desde el año 1995 ($t=1, \dots, t=T_1$), y son 9. El uso de estas 9 series implica una pérdida de información que se compensará en los siguientes pasos.
2. Del paso anterior se obtienen los factores para el periodo $t=1$ hasta $t=T_1$. Con estos factores se hacen regresiones lineales utilizando como variable respuesta cada una de las series no consideradas en el paso anterior, y como variable explicativa los factores. Los parámetros estimados de la regresión se usan para retroproyectar las series que no formaban parte del panel longitudinal desde $t=1$.
3. Se calculan, de nuevo, los factores usando las series retroproyectadas del paso 2 y utilizando el mismo procedimiento que en el paso 1.
4. Se proyectan los factores obtenidos en el paso anterior desde $t=T_1$ hasta $t=T_2$ utilizando la metodología ARIMA.
5. Se utilizan los factores obtenidos en el paso anterior como regresores en aquellas series en las que no se dispone de información para el período más reciente y se proyecta este periodo empleando los parámetros de la regresión. Se vuelve a calcular los factores mediante el DFA. Los pasos 2 a 5 se repiten hasta la convergencia. La convergencia se determina por los cambios en el Criterio de Información de Akaike (AIC)

Para la estimación del modelo DFA se utiliza el paquete MARSS de R (Holmes *et al.*, 2014). Las funciones del paquete estiman los parámetros del modelo (Z y R) por máxima verosimilitud utilizando el algoritmo de Expectation-Maximization (EM). El algoritmo EM de Dempster *et al.* (1977) fue utilizado para este tipo de modelos por Shumway & Stoffer (1982) y Watson & Engle (1983). Para la estimación de los factores se emplea el filtro de Kalman, una vez que el modelo se representa en el espacio de estados.

Uno de los principales contratiempos que se plantean al hacer el DFA es la decisión del número de factores que es necesario considerar. En este caso para tomar esta decisión se utilizan las matrices de autocovarianzas muestrales de las series observadas $C(h)$, mediante la observación de los valores propios.

Fórmula 2:

$$\hat{C}(h) = \frac{1}{T} \sum_{t=1}^{n-h} (y_{t+h} - \bar{y})(y_t - \bar{y})' \quad h = 1, 2, 3, 4 \dots$$

El problema de las matrices de autocovarianzas muestrales es que no son simétricas y por tanto los valores propios pueden tomar valores negativos e, incluso, complejos. Esto dificulta la determinación del número de factores puesto que no es correcto emplear las magnitudes de los valores propios. Se puede resolver calculando el número de valores singulares no nulos de las matrices de autocovarianzas muestrales, tal y como se muestra en Zhang (2010). Los valores singulares de la matriz se corresponden con las raíces cuadradas de los valores propios asociados a $C(h)'C(h)$.

Por último, se relacionará el indicador sintético obtenido con el procedimiento anterior con el PIB, utilizando para ello dos procedimientos, un polinomio de grado 3 y un modelo GAM.

Los modelos lineales son métodos de regresión simple que modelan la variable respuesta (dependiente) en función de las covariables. Se estima mediante el método de los mínimos cuadrados. Cuando se quiere modelar de forma lineal una variable respuesta que no tiene distribución normal, por ejemplo, una variable binaria (modelo logit) o discreta (Poisson), podemos usar una generalización del modelo lineal: el modelo lineal generalizado (GLM). Este modelo se estima con el método de los mínimos cuadrados iterativamente ponderados (IRLS). El modelo aditivo significa que la variable respuesta depende linealmente de funciones desconocidas pero suaves. Es decir, el objetivo de los GAM es modelar la variable respuesta mediante variables independientes que son un conjunto de funciones suaves.

El modelo GAM se escribe formalmente como:

$$g(E(y_i)) = a_0 + f_1(x_{i1}) + \dots + f_p(x_{ip}) + e_i$$

donde $i=1, \dots, N$, g es la función link (identidad, logaritmo, etc), y es la variable respuesta y x_1, \dots, x_p son las variables independientes. a_0 es el término independiente y f_1, \dots, f_p son funciones desconocidas suaves y e es el término de error.

La función suave f está compuesta por la suma de una base de funciones b y sus correspondientes coeficientes de regresión

$$f(x) = \sum_{i=1}^q b_i(x) \beta_i$$

donde q es la dimensión de la base.

Las funciones suaves que utilizaremos en este trabajo son splines. Los splines son funciones reales que están definidas a trozos por funciones polinomiales (funciones básicas). Los lugares donde se conectan las piezas polinomiales se denominan nodos. En los GAM, las splines de regresión penalizadas se utilizan para regularizar la suavidad de una spline.

Por tanto, el modelo se puede escribir como un modelo lineal de la siguiente forma:

$$g(E(y)) = \beta X^* + E$$

donde X^* es la matriz del diseño y β es el vector de coeficientes de la regresión.

Entonces, la función objetivo que es necesario minimizar es

$$\|y - \beta X^*\|^2 + \lambda \int_0^1 (f''(x))^2 dx$$

donde λ es el parámetro de suavización y

$$\int_0^1 (f''(x))^2 dx = \beta S \beta'$$

donde S es una matriz de parámetros desconocidos. Esto implica que la estimación de los coeficientes de regresión se puede hacer utilizando la siguiente ecuación:

$$\beta = (X^{*'} X^* + \beta S)^{-1} X^{*'} y$$

Para obtener la estimación de β se utiliza el método de los mínimos cuadrados penalizados e iterativamente reponderados (P-IRLS). En este artículo se emplean como bases los P-splines y en cuanto al parámetro de suavizado y al número de bases se usan las que por defecto nos ofrece el paquete *mgcv* de R (Wood, 2006).

La expresión del modelo GAM que se utiliza en este trabajo se presenta en la Fórmula 3 (Wood, 2006). En él se tiene en cuenta la estructura

de correlación de los datos a través del tiempo y considerando un modelo AR(2) para los errores

Fórmula 3:

$$PIB_T = f_1 (IS_T) + f_2 (T) + e_T$$

donde

PIB: tasa de variación interanual del PIB

IS: tasa de variación interanual del indicador sintético

f_1 : función suave que modeliza el Indicador sintético

f_2 : función suave que modeliza el tiempo

$e \sim AR(2)$

Se utiliza esta estructura de correlación para los errores porque en un primer paso se ajustó el modelo sin estructura de correlación, se calcularon los residuos y se comprobó que estos seguían un modelo AR(2).

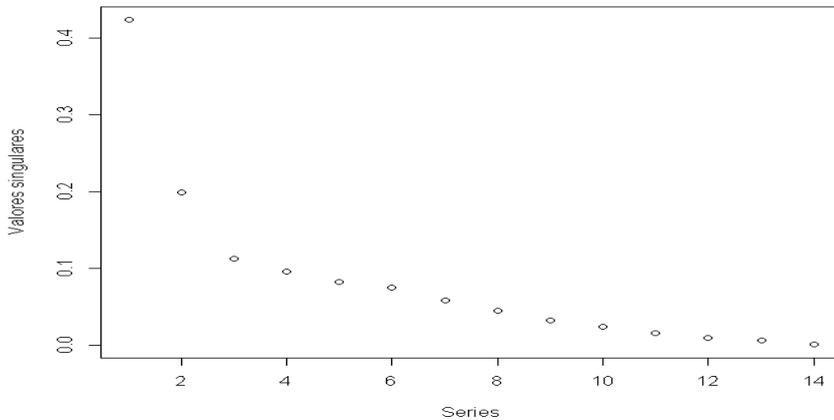
Todos los pasos anteriores están programados utilizando el software R y las API's que proporciona el IGE para la captura automática de las series iniciales del estudio. De esta manera se puede obtener la evolución del indicador en tiempo real simplemente con la ejecución del algoritmo programado en R. Este algoritmo consulta directamente las bases de datos públicas disponibles en la web del IGE a través de las API's y, por tanto, captura la última información disponible.

3. RESULTADOS

En esta sección se presentan los resultados de la aplicación del DFA a las series descritas en la sección anterior. El primer paso a seguir es determinar el número de factores comunes que se necesitan en el DFA para representar la mayor parte de la variabilidad de las series originales.

Para determinar el número de factores comunes, se calculan los valores propios de las matrices de autocovarianzas muestrales. Estos valores propios son complejos. Por tanto, es necesario calcular los valores singulares. En la Figura 1 se presentan los valores singulares de la matriz de autocovarianzas muestrales de retardo 1 de las series, para el período transversal común. Este gráfico parecería indicar que es necesario considerar dos factores.

FIGURA 1
VALORES SINGULARES ASOCIADOS A CADA UNA DE LAS SERIES



Fuente: Elaboración propia.

Por simplicidad para esta aplicación práctica, se decidió tomar un solo factor, que será el indicador sintético de coyuntura. Teniendo en cuenta esto y una vez aplicado el modelo factorial dinámico, siguiendo los pasos 1-5 expuestos en la sección anterior, se obtienen las cargas factoriales del Cuadro 2 en las series empleadas.

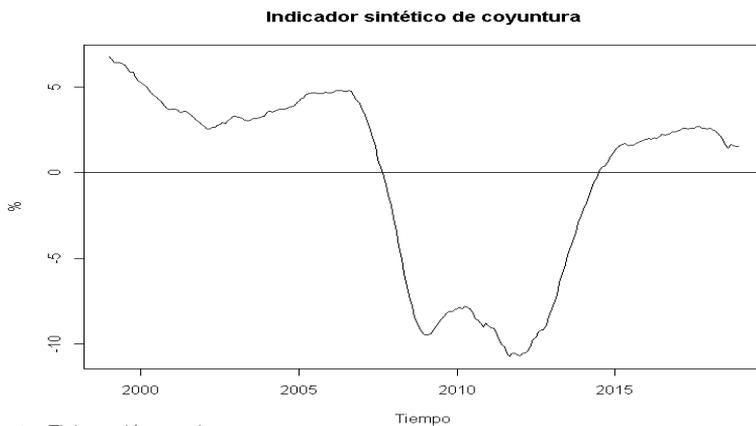
Las cargas factoriales estimadas para cada una de las series originales miden el grado de correlación dinámica entre el factor común y cada una de ellas. Como se puede observar en el Cuadro 2, todos ellos son positivos. En este sentido se puede considerar a todos los indicadores utilizados en el modelo como procíclicos. Aunque todas las series contienen información sobre la evolución de la coyuntura económica de Galicia, hay algunas diferencias en las cargas factoriales. Así, estas estimaciones muestran que las afiliaciones a la Seguridad Social presentan el mayor valor de las cargas factoriales, seguidas del Transporte aéreo de pasajeros, la cifra de negocios en el sector servicios, el consumo de combustibles, el índice de producción industrial y el de cifra de negocios en la industria. Estos son los 6 indicadores que consideraremos finalmente para estimar nuestro modelo factorial. Con estas cargas factoriales se obtiene el factor común que se presenta en la Figura 2.

CUADRO 2
CARGAS FACTORIALES

Variables	Cargas factoriales
ASS	0,093
TAP	0,022
ICNSS	0,016
GASOL	0,012
ICNI	0,012
IPI	0,011
CXX	0,007
EDVIV	0,006
VIAJ	0,006
CXM	0,005
MAT	0,004
TMM	0,003
VIV	0,001
PRODVE	0,001

Fuente: Elaboración propia.

FIGURA 2
INDICADOR SINTÉTICO DE COYUNTURA



Fuente: Elaboración propia.

Una vez realizada la estimación del indicador sintético es importante hacer un análisis de la capacidad explicativa del mismo sobre el crecimiento del PIB trimestral, como indicador coyuntural de referencia comúnmente aceptado. El PIB se obtiene a partir de una estadística de síntesis, resultado de combinar indicadores coyunturales trimestrales y mensuales con fuentes de información estructural. Por tanto, el PIB funcionalmente es equivalente a un factor común, aunque para su cálculo no se utilicen modelos factoriales. Por tanto, la existencia de coherencia cíclica y de estabilidad en la relación permitirá aceptar el factor común como indicador sintético válido. En este trabajo se usa la serie del PIB corregida de estacionalidad y calendario.

En la Figura 3 se presenta el indicador sintético trimestralizado y la variación intertrimestral del PIB obtenida del IGE. Para la trimestralización del factor se ha utilizado la expresión de Mariano & Murasawa (2003) quienes demostraron que si la media muestral de las tasas de crecimiento mensual se puede aproximar por la media geométrica, la tasa de crecimiento trimestral se puede aproximar con la media de las tasas de crecimiento mensual empleando la siguiente expresión:

Fórmula 4:

$$x_T = \left(\frac{1}{3} + \frac{2}{3}B + B^2 + \frac{2}{3}B^3 + \frac{1}{3}B^4 \right) x_t$$

donde x_t representa el factor mensual dinámico común y x_T es el agregado trimestral, por tanto, $T=3t$ y B es el operador de retardos.

En la Figura 3 se ve un comportamiento parecido entre la tasa de variación intertrimestral del indicador sintético y del PIB, aunque se observa un cierto sesgo en la última parte de la Figura. El coeficiente de correlación de Pearson entre las tasas de variación intertrimestral del indicador sintético y el PIB es del 0,86 con un intervalo de confianza del 95% de (0,78, 0,92). La función de correlación cruzada también muestra un alto grado de relación entre el indicador sintético y el PIB (Figura 4). La función tiene un máximo en el cero, lo que confirma la naturaleza coincidente de los dos indicadores y además el perfil asimétrico de la Figura 4 da idea de que el indicador sintético va ligeramente adelantado con respecto al PIB, propiedad deseable para hacer predicciones.

FIGURA 3
CRECIMIENTO INTERTRIMESTRAL DEL INDICADOR SINTÉTICO Y DEL PIB (%)

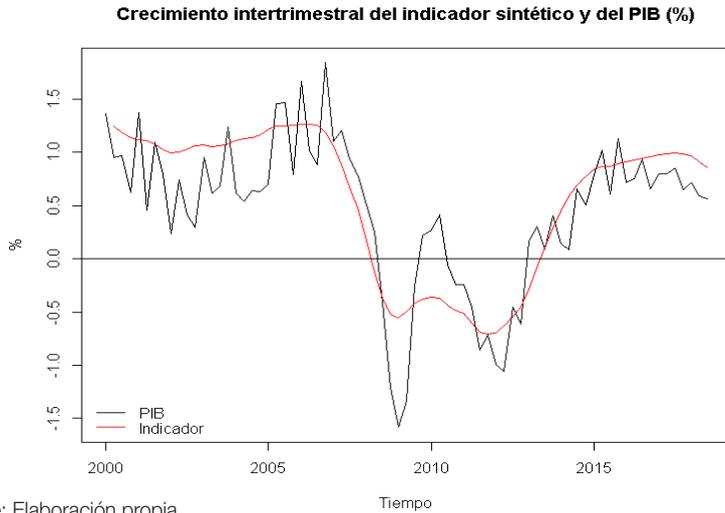
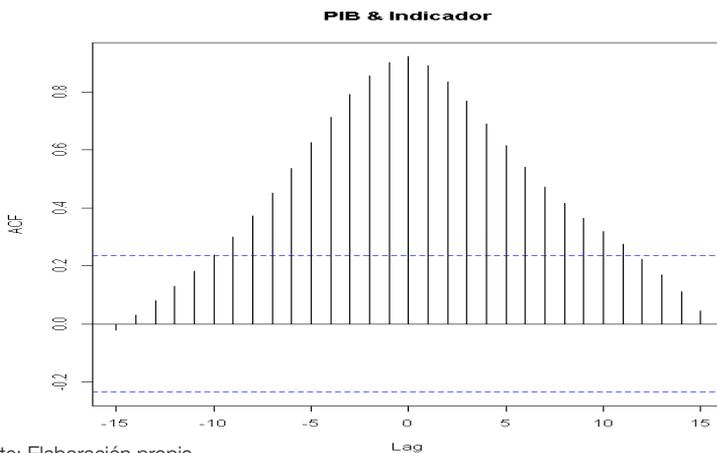


FIGURA 4
FUNCIÓN DE CORRELACIÓN CRUZADA ENTRE EL INDICADOR SINTÉTICO Y EL PIB



También es deseable observar el comportamiento de la evolución de la tasa de variación interanual del PIB y de la tasa de variación interanual del indicador. Para calcular la tasa de variación interanual del indicador se utiliza la siguiente aproximación:

Fórmula 5

$$\ln(x_t) - \ln(x_{t-1}) \cong \frac{x_t - x_{t-1}}{x_{t-1}}$$

Entonces se tiene que:

Fórmula 6

$$\ln(x_t) - \ln(x_{t-1}) + \ln(x_{t-1}) - \ln(x_{t-2}) + \dots + \ln(x_{t-11}) - \ln(x_{t-12}) \cong \frac{x_t - x_{t-12}}{x_{t-12}}$$

Por tanto, se calcularán las tasas de variación interanual del indicador aproximadas por la suma de las tasas de variación mensual para los 12 meses anteriores (Fórmula 6).

En la Figura 5 se puede comprobar la coherencia de los dos estimadores, aunque al igual que pasaba en la taxa de variación intertrimestral se observa un cierto sesgo en la última parte de gráfico. En la última parte de este trabajo se verá como este sesgo se corrige al utilizar un modelo que permitirá relacionar el Indicador sintético con el PIB. El coeficiente de correlación de Pearson entre las tasas de variación interanual del indicador sintético y del PIB es del 0,92 con un intervalo de confianza del 95% de (0,87, 0,95).

Por último, y una vez visto que el indicador tiene buenas propiedades para ser un buen predictor del PIB, se establece un modelo que relacione estos dos indicadores y que permita predecir el PIB a partir del indicador sintético. En primer lugar, es necesario comentar que a las series de tasas de variación interanual del PIB y del indicador sintético se les aplicó el test de Johansen (1991) y se obtuvo que, con una confianza del 95%, las dos series están cointegradas, lo que en principio, es una buena propiedad para predecir el PIB a partir del indicador.

En la Figura 6 se presenta un gráfico de dispersión que refleja la relación existente entre la tasa de variación interanual del indicador sintético (IS) y el PIB. A la vista de la Figura podemos concluir que la relación no es lineal,

lo que justifica el uso de un modelo no lineal para modelar esta relación. Además, también se observa que la relación se aleja más de la linealidad cuando los crecimientos del PIB son negativos. Es decir, el indicador refleja con menor exactitud las fases de caída brusca del PIB.

FIGURA 5
CRECIMIENTO INTERANUAL DEL INDICADOR SINTÉTICO Y DEL PIB (%)



Fuente: Elaboración propia.

Desde el punto de vista cualitativo, si se comparan los signos de ambas tasas de variación, se obtiene igualdad en 63 de los pares y discordancia en tres. Por tanto, un indicador de concordancia será:

Fórmula 7:

$$\frac{C - D}{C + D} = \frac{63}{69} = 0,913$$

donde

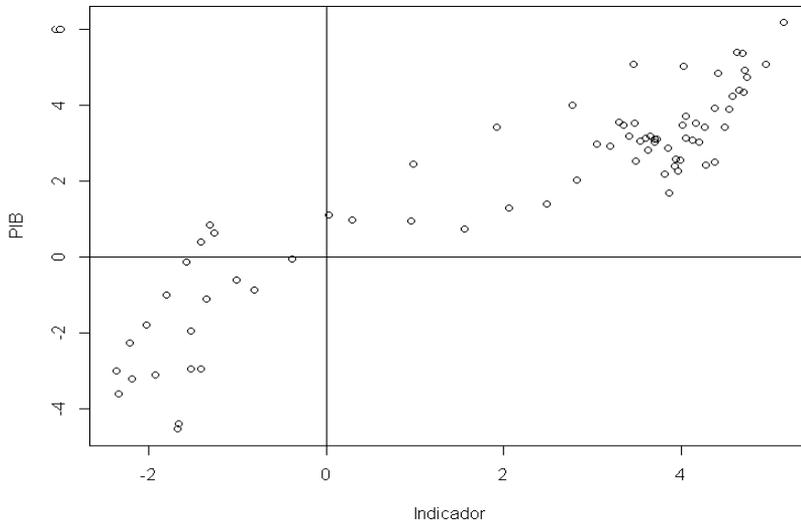
C signos concordantes de las tasas de variación

D signos discordantes de las tasas de variación

En valor absoluto, la tasa de variación del PIB es, grosso modo, inferior a la del indicador sintético. Aunque como ya se ha mencionado se observa un comportamiento diferenciado entre las fases de crecimiento y las de decrecimiento de la economía gallega. En las primeras, el indicador sobrevalora la tasa de variación al igual que en algunas de las segundas.

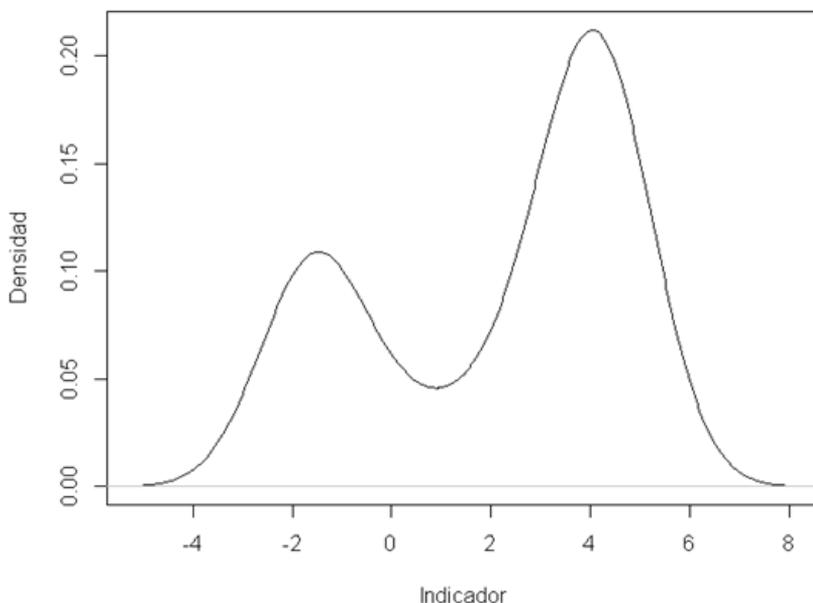
Con las observaciones disponibles hasta el momento, en el intervalo (0, 2), el indicador es poco informativo sobre el comportamiento de la variación del PIB como se puede comprobar en la Figura 6 o en la densidad de la tasa de variación del propio indicador (Figura 7) con una bimodalidad clara (cerca de -2 y de 4), que ofrece el aspecto de una mixtura de dos distribuciones campaniformes.

FIGURA 6
RELACIÓN ENTRE LA TASA DE VARIACIÓN INTERANUAL DEL INDICADOR SINTÉTICO Y DEL PIB (%)



Fuente: Elaboración propia.

FIGURA 7
**FUNCIÓN DE DENSIDAD DE LA TASA DE VARIACIÓN
 INTERANUAL DEL INDICADOR SINTÉTICO**

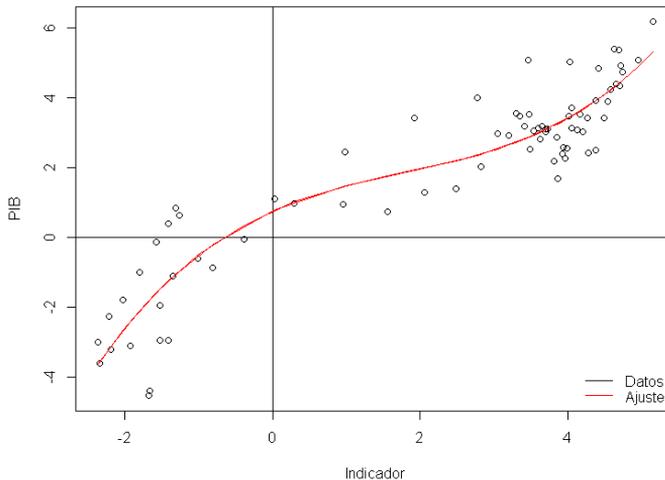


Fuente: Elaboración propia.

En un primer momento, y después de ver la Figura 6, se establece la relación entre las series de tasas de variación del PIB y del indicador a través de un polinomio de grado 3, que se presenta en el Gráfico 8. El ajuste mediante este polinomio tiene la expresión $PIB=0,76+0,94*IS-0,27*IS^2+0,05*IS^3$, donde todos los coeficientes son significativamente distintos de 0 con una confianza del 95% y la relación tiene un $R^2=0,86$.

Con el objetivo de mejorar el ajuste anterior, se relacionan las series de tasas de variación del PIB y del indicador sintético a través del modelo GAM propuesto en la sección anterior. Este modelo emplea como covariable el indicador sintético y el periodo y como variable respuesta el PIB. Además, se asume que los residuos están correlacionados siguiendo un AR(2). El modelo ajustado tiene un R^2 de 0,92.

FIGURA 8
**RELACIÓN ENTRE LA TASA DE VARIACIÓN INTERANUAL DEL
 INDICADOR SINTÉTICO Y DEL PIB Y EL AJUSTE MEDIANTE UN
 POLINOMIO DE GRADO 3**



Fuente: Elaboración propia.

En la Figura 9 se presenta la variación interanual del PIB de la economía gallega y su predicción empleando el modelo GAM. Se puede observar una buena predicción en la última parte de la serie. Teniendo en cuenta esta predicción podemos decir que previsiblemente el crecimiento del PIB de la economía de Galicia se situará en el cuarto trimestre del año 2018 en el 2,4%.

Desde el punto de vista de los resultados obtenidos se puede comentar que desde el año 2000 al año 2008 la economía gallega presenta uno de sus períodos de crecimiento sostenido más largo de su historia, esta etapa coincide con valores positivos del indicador sintético. En ella, el indicador resulta demasiado plano, quizás como resultado del proceso de imputación realizado por el panel incompleto de las series utilizadas en el análisis.

A finales del año 2008 y coincidiendo con la recesión internacional que sufren la mayoría de los países industrializados, Galicia muestra las tasas de crecimiento del PIB más negativas de su historia reciente. Esta recesión está bien identificada por el indicador sintético hasta finales del año 2014. En el año 2014 se empieza a ver una recuperación en las tasas de crecimiento del PIB, recuperación que también se puede identificar en el indicador sintético.

FIGURAR 9 VARIACIÓN INTERANUAL DEL PIB Y SU PREDICCIÓN EMPLEANDO EL MODELO GAM



Fuente: Elaboración propia.

Para valorar la capacidad predictiva del indicador, respecto al PIB, se ha diseñado un pequeño estudio de simulación en el que se calculó la raíz del error cuadrático medio (RMSE) de la predicción del PIB con el modelo GAM en comparación con el verdadero valor del crecimiento del PIB obtenido de la Contabilidad Trimestral de Galicia. Estas valoraciones se han hecho para dos escenarios del horizonte de predicción comprendido entre el cuarto trimestre del 2015 y el tercer trimestre del 2018:

- Escenario 1: se ha ejecutado todo el procedimiento en los meses de enero para el cuarto trimestre del año anterior, en marzo, para el primer trimestre, julio, para el segundo trimestre y octubre, para el tercer trimestre (dentro del horizonte de predicción)
- Escenario 2: se ha ejecutado todo el procedimiento en los meses de febrero, para el cuarto trimestre del año anterior, abril para el primer trimestre, agosto, para el segundo trimestre, noviembre, para el tercer trimestre.

La diferencia entre los dos escenarios es que en el segundo tenemos más información disponible de los indicadores parciales para el DFA.

Para el escenario 1 hemos obtenido un RMSE del 0,114% y para el escenario 2 un error del 0,112%, valores similares o incluso inferiores a otros estudios semejantes (Cuevas & Quilis, 2012).

4. CONCLUSIONES

El trabajo realizado proporciona un indicador sintético que permite el seguimiento en el corto plazo de la evolución de la coyuntura económica regional. De este modo se está en condiciones de satisfacer las necesidades de información cada vez más inmediatas que se precisan para el análisis económico.

Las ventajas que presenta este trabajo con respecto a propuestas semejantes existentes en la literatura son cuatro. En primer lugar, y como principal rasgo diferencial del procedimiento metodológico de este trabajo, hay que señalar el buen comportamiento del indicador cuando se compara con la evolución del PIB, calculado por el IGE. El buen grado de acoplamiento entre la evolución de ambas magnitudes (indicador sintético y PIB), determinado por el coeficiente de correlación, evidencian su coherencia cíclica y estabilidad. Ambos indicadores tienen un comportamiento muy similar en todas las fases del ciclo económico comprendido en la ventana temporal considerada, aunque también hay que destacar que el indicador sintético es más suave, como cabía esperar. En particular, es interesante notar que el indicador captura perfectamente la crisis del año 2008 y su correspondiente salida en forma de W. Se trata en definitiva de una validación empírica del indicador sintético calculado.

La segunda de las aportaciones del trabajo realizado reside en la simplicidad del procedimiento empleado. En este trabajo se utiliza un procedimiento más sencillo y, por tanto, con menos asunciones de hipótesis que en la mayoría de trabajos de características similares y, en cambio, los resultados empíricos son similares. Por ejemplo, en Cuevas & Quilis (2012) utilizan un algoritmo iterativo similar pero el DFA y su representación en el espacio de estados es más complicada. En este trabajo se emplea el proceso no estacionario más sencillo, el paseo aleatorio.

Una tercera ventaja del indicador sintético calculado es que sirve como indicador de la evolución de la economía, en este momento la periodicidad del PIB es trimestral, por lo que en un trimestre cualquiera, con dos meses

de antelación y mediante el cálculo del indicador sintético y el establecimiento de la función de relación entre los dos indicadores por el modelo GAM propuesto, se podrá anticipar cuál va a ser la evolución del PIB en ese trimestre. Esta propuesta metodológica es alternativa a la utilizada en otros trabajos de similares características (Cuevas & Quilis, 2012, Camacho *et al.*, 2009, 2011)

Por último, el ajuste del polinomio de grado tres que, en principio, resulta ligeramente peor que el realizado mediante *splines* tiene la ventaja de que permite, de un modo intuitivo, dar una aproximación del punto de inflexión, en este caso cuando $IS=1,77$, que puede emplearse como un valor testigo en el seguimiento de la coyuntura. Cuando el indicador se aproxime a él o quizás, a uno más redondeado (1,5) debe comenzar a preocupar su evolución. Y esta preocupación debe ser seria cuando crucemos el cero de la cúbica, $IS=-0,67$, o si se prefiere más redondeado (-0,6).

En añadidura, y a modo de conclusión final, se presenta una metodología fácilmente replicable, tanto con los datos disponibles en otra región como con los datos disponibles a escala nacional o incluso en otros países. No obstante, hay que tener en cuenta que a escala nacional o de país hay una mayor disponibilidad de series relacionadas con el crecimiento de la economía, con lo que convendría incrementar las series iniciales de las que se parte en este trabajo.

BIBLIOGRAFÍA

- ANGELINI, E., CAMBA-MÉNDEZ, G., GIANNONE, D., REICHLIN, L., RUNSTLER, G. (2008): *Short-term forecasts of Euro area GDP growth*, CEPR Discussion Paper N° 6746.
- ARUOBA, S. B., DIEBOLD, F. X., SCOTTI, C. (2009): "Real-time measurement of business conditions". *Journal of Business & Economic Statistics*, 27(4).
- ARTÍS, M., PONS, J., SIERRA, M. Á., SURIÑACH, J. (1997): "Nivel de actividad mediante indicadores de coyuntura". *Revista de economía aplicada*, 13, 129-147.
- BURNS, A., MITCHELL, W. (1946): *Measuring business cycles*. National Bureau of Economic Research.
- CAMACHO, M., DOMÉNECH, R. (2011): "MICA-BBVA: A Factor Model of Economic and Financial Indicators for Short-term GDP Forecasting", *Journal of the Spanish Economic Association*.
- CAMACHO, M., PEREZ-QUIROS, G. (2009): Ñ-Sting: *España Short Term Indicator of Growth*, Banco de España, Working Paper N°. 0912.
- CENDEJAS, J. L., DE LUCAS, S., DELGADO, M. J., ÁLVAREZ I. (2011): "Testing for structural breaks in factor loadings: An application to international business cycles", *Economic Modelling*, 28 (1-2), 259-263.
- CENDEJAS BUENO, J. L., DE LUCAS, S., DELGADO RODRÍGUEZ, M. J. (2014): Evaluación del cambio estructural en los indicadores cíclicos. *Estadística española*, 56(183), 39-60.
- CUEVAS, Á., QUILIS, E. M. (2012): "A factor analysis for the Spanish economy". *SERIEs*, 3(3), 311-338.
- CUEVAS, A., QUILIS, E. M., ESPASA, A. (2015): "Quarterly Regional GDP Flash Estimates by Means of Benchmarking and Chain Linking". *Journal of Official Statistics*, 31(4), 627-647.
- DEMPSTER, A. P., LAIRD, N. M., RUBIN, D. B. (1977): Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society*, B, 39, 1-38
- EVANS, C. L., LIU, C. T., PHAM-KANTER, G. (2002): "The 2001 recession and the Chicago Fed National Activity Index: Identifying business cycle turning points". *Economic Perspectives-Federal Reserve Bank Of Chicago*, 26(3), 26-43.
- GABISCH, G., LORENZ, H. (2013): *Business Cycle Theory: A Survey of Methods and Concepts*. Springer Science & Business Media.
- GARRIDO YSERTE, R., GALLO RIVERA, M. T., MARTÍNEZ GAUTIER, D. (2016): "La crisis económica y la geografía del tejido empresarial español: 2000-2013". *Revista de Estudios Regionales*, 106, 165-195.
- GIANNONE, D., REICHLIN, L., SMALL, D. (2008): "Nowcasting: The real-time informational content of macroeconomic data". *Journal of Monetary Economics*, 55, 665-676.
- JOHANSEN, S. (1991): "Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models". *Econometrica: Journal of the Econometric Society*, 1551-1580.
- HOLMES, E. E., WARD, E. J., SCHEUERELL, M.D. (2014): *Analysis of multivariate time series using the MARSS package*. <http://cran.r-project.org/web/packages/MARSS/vignettes/UserGuide.pdf>
- MARCELLINO, M., STOCK, J. H., WATSON, M. W. (2003): "Macroeconomic forecasting in the euro area: Country specific versus area-wide information". *European Economic Review*, 47(1), 1-18.
- MARIANO, R. S., MURASAWA, Y. (2003): "A New coincident index of business cycles based on monthly and quarterly series". *Journal of Applied Econometrics*. 18, (4), 427-443.

- MONDÉJAR-JIMÉNEZ, J., VARGAS-VARGAS, M. (2008): "Indicadores sintéticos: una revisión de los métodos de agregación". *Economía, sociedad y territorio*, 8(27), 565-585.
- MUÑOZ, F., TROMBETTA, M. (2015): "Indicador Sintético de Actividad Provincial (ISAP): un Aporte al Análisis de las Economías Regionales argentinas". *Investigaciones Regionales*, 33, 71-96
- PEÑA, D., PONCELA, P. (2004): "Nonstationary dynamic factor analysis". *Journal of statistical planning and inference*, 136(4), 1237-1257.
- SÁNCHEZ FERNÁNDEZ, P., GALLEGO RODRÍGUEZ, E., RIVERO FERNÁNDEZ, D., LAGO PEÑAS, S. (2014): "El impacto de la actividad exportadora sobre las empresas: Evidencia empírica para el caso gallego". *Revista de Estudios Regionales*, 100, 223-243.
- SHUMWAY, R. H., STOFFER, D.S. (1982): "An approach to time series smoothing an forecasting using the EM algorithm". *Journal of time series analysis*, 3, 253-264.
- STOCK, J. H., WATSON, M. W. (1991): "A probability model of the coincident economic indicators", K. Lahiri, G.H. Moore (eds.): *Leading economic indicators. New approaches and forecasting records*. Cambridge, Cambridge University Press,
- STOCK, J. H., WATSON, M. W. (2002): "Macroeconomic forecasting using diffusion indexes". *Journal of Business & Economic Statistics*, 20(2), 147-162.
- TRUJILLO, F., BENÍTEZ, M. D., LÓPEZ, P. (2000): "Trimestralización de los valores añadidos sectoriales mediante indicadores. Aplicación al caso de Andalucía". *Revista de Estudios Regionales*, 57, 59-97.
- U.S. CENSUS BUREAU (2002): *X-12-ARIMA Reference Manual Version 0.2.10*. September.
- ZHANG, H. (2009): "Comparación entre dos métodos de reducción de dimensionalidad en series de tiempo". *Revista Colombiana de Estadística*, 32 (2), 189-212.
- WATSON, M.W. ENGLE, R.F. (1983): "Alternative algorithms for the estimation of dynamic, mimic and varying coefficient regression models". *Journal of Econometrics*, 23, 385-400.
- WOOD, S. (2006): "Generalized Additive Models: an introduction with R", Chapman and Hall/CRC. <http://reseau-mexico.fr/sites/reseau-mexico.fr/files/igam.pdf>.

