
Construcción de un índice coincidente para la actividad económica Colombiana

Construction of a coincident index for Colombian economic shorttitle

Fabio H. Nieto^a
fhnietos@unal.edu.co

Nicolás Chudt^b
nchudtr@unal.edu.co

Resumen

Usando factores comunes de un conjunto de variables macroeconómicas y el enfoque de perfil coincidente para seleccionar uno de ellos, como el más cercano al estado de la economía, se propone en este trabajo un índice coincidente para la actividad económica Colombiana.

Palabras clave: factores comunes dinámicos, índice coincidente, perfil coincidente.

Abstract

We propose a construction of coincident index for Colombian economic by using common factors of a set of macroeconomic variables and the coincident profile approach to select one of them, as the closest to the state of the economy.

Keywords: coincident index, coincident prole, dynamic common factors..

1. Introducción

(Martínez et al. 2016) diseñaron un procedimiento para obtener índices coincidentes del estado no observable de un sistema. Como ejemplos de estos sistemas se tiene el económico de un país, el del mercado financiero de una bolsa de valores o el del clima de una región geográfica. Considerado el estado del sistema como un proceso estocástico, un índice coincidente es una estimación de él. (Martínez et al. 2016) ilustraron su metodología con un ejemplo para la economía colombiana y otro para el sector financiero colombiano. En el caso de la economía, utilizaron nueve variables observadas mensualmente en el período muestral enero,

^aProfesor PhD, Universidad Nacional de Colombia

^bEstudiante del programa de Maestría en Estadística, Universidad Nacional de Colombia

1980-diciembre, 2005 y una estimación del PIB mensual, obtenida con el método de (Nieto et al. 2004), para obtener los llamados perfiles coincidentes de cada factor común encontrado.

Con el propósito de mejorar el contenido económico del grupo de variables usado por (Martínez et al. 2016) y actualizar el periodo muestral hasta una fecha reciente (en la medida de lo posible), en este trabajo se seleccionan seis variables, a saber: índice de los ingresos reales de la industria (IND) con base 2014, índice de oferta de energía (ENE) con base 2005, producción total de caña de azúcar (CANA) en toneladas, producción de cemento (CEM) en toneladas, índice del comercio sin vehículos (COMCV) con base 2005 y tasa de desempleo (TD) en unidades porcentuales. El período muestral es enero, 2000-junio, 2017.

Como es bien sabido, actualmente se dispone en la literatura de una variedad de métodos para calcular índices coincidentes, específicamente para el sistema económico de un país. Ellos están basados fundamentalmente en modelos de series de tiempo y el rango puede ir desde (Stock & Watson 1989), (Stock & Watson 1991) hasta (Martínez et al. 2016). En este último trabajo, se puede consultar una lista de los más relevantes métodos al respecto.

El documento se organiza de la siguiente forma: en la segunda sección se incluye una breve descripción del modelo factorial de (Peña & Poncela 2016) y de (Lam & Yao 2012) y del perfil coincidente de Martínez *et al* (2016); en la tercera sección se hace una descripción del conjunto inicial de variables considerado y se justifica la selección final de las seis variables indicadas antes. ; la aplicación del método propuesto se presenta en la cuarta sección, en donde además, se hace un análisis del ciclo determinado por el factor común escogido como índice; la quinta sección está conformada por algunas conclusiones y recomendaciones para una futura investigación.

2. Especificación del modelo factorial y del perfil coincidente

2.1. Modelo de factores comunes

Sea $\{y_t\}$ un proceso estocástico observable de dimensión m , el cual es generado por un proceso latente r -dimensional $\{f_t\}$, con $r \leq m$, via la ecuación

$$y_t = P f_t + e_t, \quad t \in \mathbb{Z}, \quad (1)$$

Donde \mathbb{Z} denota el conjunto de números enteros, P es una matriz $m \times r$ llamada la matriz de pesos y $\{e_t\}$ es un proceso ruido blanco multivariado gaussiano m -dimensional, con media 0 y matriz de varianzas y covarianzas de rango completo Σ_e . Los procesos componentes de $\{\mathbf{f}_t = (f_{1t}, \dots, f_{rt})'\}$ donde el símbolo $'$ significa transposición de matrices, se llaman los factores comunes y ellos pueden ser no estacionarios o estacionarios. En este trabajo se supondrá además que

no son estacionales (para un estudio de factores comunes estacionales se puede consultar el trabajo de (Nieto et al. 2016)).

La metodología estadística para determinar el número de factores comunes r y estimarlos en cada tiempo t , se puede consultar en los artículos de (Peña & Poncela 2016) y de (Lam & Yao 2012), entre muchos otros. Siguiendo a (Peña & Poncela 2016), la especificación del número total r de factores comunes, se hace a través de una prueba estadística basada en los valores propios de las matrices de autocorrelación canónica. La idea es: la matriz de correlación canónica de rezago k está dada por

$$M(k, N) = \left[\sum_{t=k+1}^N y_t y_t^T \right]^{-1} \sum_{t=k+1}^N y_t y_{t-k}^T \left[\sum_{t=k+1}^N y_{t-k} y_{t-k}^T \right]^{-1} \sum_{t=k+1}^N y_{t-k} y_t^T ,$$

para cada $k = 1, \dots, N - 1$, donde N denota el tamaño de muestra. Sea

$$S_{m-r,k} = -(N - k) \sum_{j=1}^{m-r} \ln(1 - \hat{\lambda}_j) ,$$

Donde $\hat{\lambda}_1 \leq \dots \leq \hat{\lambda}_m$ son los valores propios ordenados de $M(k, N)$. Entonces, $S_{m-r,k}$ es una estadística de prueba para examinar la hipótesis nula de que a lo más $m - r$ valores propios, de la matriz límite de la sucesión de matrices $M(k, N)$ cuando N tiende a ∞ , son iguales a cero. La distribución límite de $S_{m-r,k}$ bajo esta hipótesis nula es $\chi^2_{(m-r)2}$, para todo $k = 1, 2, \dots$

En este trabajo, además de la prueba estadística de (Peña & Poncela 2016) descrita antes, se utilizará como criterio complementario para determinar el número de factores comunes, el procedimiento de (Lam & Yao 2012). Este se basa en las matrices muestrales de autocovarianza generalizada definidas por (Peña & Poncela 2016), las cuales están dadas por la expresión

$$C(k, N) = \frac{1}{N^{2d}} \sum_{t=k+1}^N (y_{t-k} - \bar{y})(y_t - \bar{y})' ,$$

Donde el símbolo $'$ significa transposición de matrices, $k = 0, 1, 2, \dots, N - 1$, $\bar{y} = \frac{1}{N} \sum_{t=1}^N y_t$ y d denota el grado de integración de los factores comunes no estacionarios. Estas matrices son llamadas por (Peña & Poncela 2016), matrices muestrales de autocovarianza generalizada. Ahora, para $k_0 \geq 1$ conocido, sean

$$M = \sum_{k=1}^{k_0} C(k, N) C(k, N)' ,$$

y $\lambda_1 \geq \dots \geq \lambda_p$ los valores propios de M ordenados en forma descendente de magnitud. Entonces, un estimador de r está dado por

$$\hat{r} = \operatorname{argmin}\{\lambda_{i+1}/\lambda_i | 1 \leq i \leq R\} ,$$

donde R es un valor conocido tal que $r < R < m$. En la práctica, para un valor de m moderadamente grande, se toma $R = m/2$.

Determinado r , se procede a la estimación de la matriz P , Σ_e y los factores comunes f_{1t}, \dots, f_{rt} para cada $t = 1, \dots, n$, para ello, en este trabajo se usa la metodología de (Peña & Poncela 2016), la cual está basada en un modelo de estados. En este contexto, la estimación (predicción) de los factores comunes se realiza por medio del suavizador de punto fijo. Sin embargo, una forma alterna de obtener estas estimaciones es el procedimiento de (Lam & Yao 2012). Es importante anotar también, que se utilizan las condiciones de identificabilidad del modelo factorial de (Peña & Poncela 2016).

2.2. Perfil coincidente

Siguiendo el enfoque de (Martínez et al. 2016), pero precisando los conceptos y formalizando el método estadístico utilizado por esos autores, se presenta a continuación el enfoque del perfil coincidente. Los planteamientos esgrimidos en este texto extienden su método y son, por lo tanto, una contribución teórica y metodológica al tema de construcción de índices coincidentes, en particular de actividad económica, por esta razón, a partir de ahora solo se hará al contexto económico.

Definición 1. Sean $\{C_t\}$ e $\{I_t\}$ procesos estocásticos que representan, respectivamente, el estado de la economía y un estimador de $\{C_t\}$. Decimos que $\{I_t\}$ es l -coincidente con $\{C_t\}$ para algún $l \in \{0, \pm 1, \pm 2, \dots, \pm c\}$, con c número entero positivo conocido, si para cada trayectoria de $\{C_t\}$ e $\{I_t\}$, cada uno de los puntos de giro de la trayectoria de $\{I_t\}$ está rezagado o adelantado con respecto a cada uno de los correspondientes de la trayectoria de $\{C_t\}$, por $|l|$ periodos de tiempo.

Observaciones: 1. Estrictamente, los puntos de giro (o de inflexión) hacen referencia a valores máximos y mínimos locales de las trayectorias bajo consideración.

2. Si $l < 0$, se tiene un rezago. Si $l > 0$, se tiene un adelanto. El caso $l = 0$ será interpretado como coincidencia exacta o, simplemente, coincidencia.

3. En economía, y para variables mensuales, usualmente es razonable tomar $c \leq 3$. Este valor de c ha sido utilizado antes en la literatura, por ejemplo, por (Melo et al. 2001)

De acuerdo con la definición anterior, el objetivo es tener un índice 0-coincidente para $\{C_t\}$. Sin embargo, un índice l -coincidente con l muy cerca a cero podría ser muy útil para estimar el estado actual de la economía.

Para determinar el valor de l , se propone examinar la hipótesis nula $H : l = l_0$ v.s. $K : l \neq l_0$, para l_0 un valor conocido en $\{0, \pm 1, \dots, \pm c\}$. Inicialmente, se supone $l_0 = 0$. Sean $\{c_t\}$ e $\{i_t\}$ series de tiempo (trayectorias) de $\{C_t\}$ e $\{I_t\}$, respectivamente, en el período muestral $1, \dots, N$. Bajo coincidencia exacta, habría Q puntos de giro en los puntos temporales t_1, \dots, t_Q , para algún $Q \geq 1$, en las series de tiempo $\{c_t\}$ e $\{i_t\}$. Pero, en virtud de la variabilidad muestral, los puntos

de giro en $\{i_t\}$ podrían ocurrir en t'_1, \dots, t'_Q . Así, se tiene un pareamiento natural de los puntos de giro, así $\{(t_j, t'_j) : j = 1, \dots, Q\}$. Se espera que $d_j = t_j - t'_j \approx 0$ para todo $j = 1, \dots, Q$ y, en consecuencia, $\sum_{j=1}^Q d_j \approx 0$. Obviamente, esta suma puede ser positiva o negativa.

Con el objeto de construir una estadística que permita decidir si la suma observada es un valor extremo y, en consecuencia, tener evidencia empírica para rechazar la hipótesis nula, se considera el siguiente procedimiento estadístico basado en las pruebas de permutación (un marco teórico formal aparece en el apéndice).

Sean d_1, \dots, d_Q las diferencias observadas en los puntos de giro de las series de tiempo $\{c_t\}$ e $\{i_t\}$ y sea s_0 su suma. Se reorganizan las diferencias de la siguiente forma: inicialmente, es considerado el valor absoluto de estas, esto es, $|d_1|, \dots, |d_Q|$. Luego, se obtienen las 2^Q permutaciones de longitud Q de los signos $+, -$ y cada una de ellas es asignada a la sucesión $|d_1|, \dots, |d_Q|$. A manera de ejemplo, una de las sucesiones de signos es $+, +, \dots, +$, otra $-, -, \dots, -$ y otra la de las diferencias observadas.

Ahora, sea s_j el valor de la suma para la j -ésima sucesión signada de las diferencias absolutas, $j = 1, \dots, Q$. Entonces, la suma observada coincide con s_{j_0} para algún $j_0 = 1, \dots, Q$. Sea

$$p_0 = \frac{\#\{s_j : |s_j| \geq |s_{j_0}|, j = 1, \dots, Q\}}{2^Q},$$

Donde “#” denota el cardinal del conjunto en referencia. Este número es el p -valor de la sucesión signada observada, bajo la distribución de permutación para la variable aleatoria S , cuyas realizaciones son todas las posibles sumas consideradas en la permutación.

Al nivel de significancia α , $0 < \alpha < 1$, la regla de decisión es rechazar la hipótesis nula y aceptar la alterna si $p_0 \leq \alpha$. Un rechazo de la hipótesis nula significa que s_{j_0} es un valor extremo de S bajo su distribución de permutación o que la sucesión observada d_1, \dots, d_Q es un resultado extremo bajo una distribución equivalente. Estos hechos están formalizados en el apéndice.

Supóngase que H es rechazada. Esto significa que $l_0 \neq 0$, luego $l_0 < 0$ ó $l_0 > 0$ y el interés ahora es identificar el valor de l_0 . Con ese fin, se consideran las hipótesis condicionales sucesivas

$$H^l : l_0 = l | H^{l'} \text{ ha sido rechazada}$$

$$K^l : l_0 \neq l | H^{l'} \text{ ha sido rechazada,}$$

donde $l = -l'$ si $l' > 0$ o $l = -l' + 1$ si $l' < 0$. Nótese que $l \in \{\pm 1, \dots, \pm c\}$.

Equivalentemente, las hipótesis anteriores pueden ser replanteadas como

$$H^l : l_0 - l = 0 | H^{l'} \text{ ha sido rechazada}$$

v.s.

$K^l : l_0 - l \neq 0 | H^l$ ha sido rechazada,

con l como antes.

Dado $l \in \{\pm 1, \dots, \pm c\}$, se propone como estadística de prueba a $S^l = \sum_{i=1}^Q d_i^l$, con $d_i^l = d_i - l$, $i = 1, \dots, Q$. Luego, son organizadas nuevamente las diferencias d_i^l y se procede como en el caso $l = 0$, definiendo

$$p_l = \frac{\#\{s_j^l : |s_j^l| \geq |s_{j_0}^l|, j = 1, \dots, Q\}}{2^Q},$$

donde s_j^l denota la suma de la j -ésima sucesión signada de $d_1 - l, \dots, d_Q - l$.

Definición 2. Al $\{p_l : l = 0, \pm 1, \dots, \pm c\}$ se lo denomina el *perfil coincidente* de la serie de tiempo $\{i_t\}$ con respecto a $\{c_t\}$.

Definición 3. Los procesos $\{C_t\}$ e $\{I_t\}$ son l_m -coincidentes al nivel α , $0 < \alpha < 1$, si $l_m = \operatorname{argmax}\{p_l : l = 0, \pm 1, \dots, \pm c\}$ y $p_{l_m} > \alpha$.

Un gráfico cartesiano de $\{p_l : l = 0, \pm 1, \dots, \pm c\}$, donde l es la abscisa, ilustra visualmente el significado y utilidad en la práctica del perfil coincidente. Nótese que para l dado en el $\{0, \pm 1, \dots, \pm c\}$, si $p_l \geq \alpha$, entonces la hipótesis nula H^l no es rechazada. En este caso puede interpretarse p_l como un nivel de confianza para un intervalo aleatorio que incluya a l .

En la vida real, la serie de tiempo $\{c_t\}$ no es observada y esto es un inconveniente para hallar un perfil coincidente. Para dar una solución a este problema, (Martínez et al. 2016) proponen utilizar una variable observable que permita rastrear la dinámica del estado de la economía $\{C_t\}$. Esa serie de tiempo se considera una estimación preliminar del estado de la economía, en el período muestral considerado, y con ella se construye el perfil coincidente para un índice propuesto.

Habiendo precisado conceptos y formalizado el método estadístico de (Martínez et al. 2016), la metodología para la construcción del índice consiste de las siguientes etapas:

ETAPA 1. Desestacionalizar las series de tiempo cuando todas o algunas de ellas sean estacionales y ajustar por datos atípicos e intervenciones si es del caso. Transformaciones para ajustar por escala (incluyendo logaritmos) también son apropiadas.

ETAPA 2. Determinar el número de factores comunes de las variables utilizadas y obtener sus estimaciones, usando la metodología de Peña y Poncela (2006) o la de Lam y Yao (2012).

ETAPA 3. En el entendido que el grado de coincidencia entre dos procesos lo determina el valor l_m , se obtienen los perfiles coincidentes al nivel α de cada factor común. Luego se selecciona el factor común con grado de coincidencia más cercano a cero y mayor p -valor.

ETAPA 4. Asignar la base del índice y obtener los puntos de giro de este, con el fin de determinar el ciclo económico.

3. Selección de variables

Una fase muy importante en la construcción de un índice coincidente de actividad económica, es la selección de los indicadores (variables) observables que lo constituirán, los cuales deben, en tiempo real, señalar razonablemente bien la evolución dinámica del estado de la economía. Para este ejercicio, y siguiendo el enfoque de (Bujosa et al. 2013), se tomaron inicialmente un total de 29 variables o indicadores. De ellas, finalmente se seleccionaron seis variables. Como se mencionó, el período muestral a utilizar es enero, 2000-junio, 2017.

La Tabla 1 presenta la lista de las 29 variables utilizadas inicialmente, junto con los sectores de la economía a los que pertenecen. Hay una primera sección donde están las variables asociadas a las actividades económicas de la producción, como son la producción de café y caña de azúcar, representando la actividad agrícola; el índice de la producción industrial representando el sector manufacturero; la demanda de energía del Sistema Interconectado Nacional (SIN) caracterizando el sector de servicios públicos; la producción de crudo, como variable del sector de minas y canteras; la producción de cemento como variable del sector de la construcción; el comercio minorista total sin combustibles, el comercio al por menor sin los combustibles ni los vehículos y la venta de vehículos, los cuales hacen parte del sector comercial.

En la segunda sección, hay dos tipos de variables: las primeras están asociadas al funcionamiento operacional del gobierno, entre las cuales se encuentran la remuneración a los asalariados, gasto general del gobierno central y el sistema general de participaciones; las segundas corresponden al recaudo, y son el impuesto al valor agregado (IVA), otros tipos de impuestos (aranceles) y el gravamen al margen financiero (conocido como 4 por mil).

Para la tercera sección se incluyen las exportaciones en cuatro diferentes unidades de medida: precios FOB dólares, precios FOB pesos, volumen peso neto kilo y volumen peso bruto kilo. asimismo, se incluyen las importaciones en precios CIF en dólares, precios CIF en pesos, volumen peso neto kilo y volumen peso bruto kilo.

La última sección incluye la tasa de desempleo que hace parte del sector de mercado laboral y las variables tasa de cambio, el índice de precios a las exportaciones y el índice de precios a las importaciones, para el sector que representa los precios.

En Colombia, actualmente, el Departamento Administrativo Nacional de Estadística (DANE) produce el Índice de Seguimiento Económico (ISE) desde enero del 2000, el cual es una estimación *mensual* de la actividad económica de Colombia. Como tal, se puede considerar una estimación razonable del estado de la economía casi en tiempo real, pues su cálculo tiene un rezago de dos meses, aproximadamente. Los detalles de su construcción se pueden consultar en su sitio web (DANE 2016). Ese punto de inicio del ISE es la razón por la cual, el rango muestral en este trabajo comienza en enero del 2000.

Se tenían procedimientos previos para la construcción de un índice coincidente, es

el caso por ejemplo de los trabajos de (Melo et al. 2001) y (Nieto 2004); (Nieto et al. 2004) entre muchos otros. En particular, (Nieto 2004);(Nieto et al. 2004) obtiene una estimación óptima del PIB mensual, la cual se considera un índice coincidente.

Tabla 1: *Tabla 1: Conjunto de variables. Fuente: DANE*

SECTOR	VARIABLES
I. ACTIVIDAD ECONÓMICA	
(a) AGROPECUARIO	Producción de miles de sacos de 60 Kg de café verde Producción de Caña de Azúcar Índice de Producción Real
(b) INDUSTRIA	Índice de Producción Real
(c) SERVICIOS PÚBLICOS	Demanda Energía (SIN) medido en kWh
(d) MINAS Y CANTERAS	Producción de millones de barriles de Crudo
(e) CONSTRUCCIÓN	Producción de Cemento
(f) COMERCIO	Total Comercio Minorista sin Combustibles Total Comercio Minorista sin Combustibles ni Vehículos Vehículos automotores y motocicletas
II. GOBIERNO	
(g) ADMINISTRACIÓN PÚBLICA	Remuneración a los a salarizados Gastos Generales del Gobierno Central Sistema general de participaciones
(h) IMPUESTOS	Impuesto al Valor Agregado Otros Impuestos Gravamen al Margen Financiero
III. COMERCIO EXTERIOR	
(i) EXPORTACIONES	Valores FOB Dólares y Valores FOB Pesos Cantidades Peso Neto Kilo y Peso Bruto Kilo
(j) IMPORTACIONES	Valores CIF Dólares y Valores CIF Pesos Cantidades Peso Neto Kilo y Peso Bruto Kilo
IV. MERCADO LABORAL Y PRECIOS	
(k) MERCADO LABORAL	Tasa de Desempleo
(l) INDICE DE PRECIOS	Tasa Representativa del Mercado Índice de Precios a las Importaciones Índice de Precios a las Exportaciones

Antes de continuar con la descripción del procedimiento usado para seleccionar las variables finales, se debe anotar lo siguiente; la definición de perfil coincidente dada en la segunda sección, está basada, filosófica y conceptualmente hablando, en la presencia de máximos y mínimos locales en los valores de las variables bajo estudio. En ocasiones, no es fácil detectar estos puntos extremos locales, debido a series de tiempo muy suaves o presencia de demasiado ruido en ellas, se recomienda entonces, transformar las series originales por medio de los filtros de primera diferencia finita (cambios absolutos) y, con base en los nuevos datos, encontrar el perfil coincidente. Al hacerlo así, el procedimiento de detección de puntos de giro en las primeras diferencias, produce aproximadamente los denominados valles y picos del ciclo económico; El perfil coincidente que así se obtenga, mostrará solo la coincidencia entre estos puntos de las variables transformadas. Un valle puede ser interpretado como el punto de fondo de una recesión, mientras que un pico señalaría el punto de techo de una expansión. Los puntos de giro de la variable original, corresponderían a aquellos en donde la primera diferencia es cero o cambia de signo. Así, de negativo a positivo, se tendría un mínimo local y, al contrario, un máximo local. En el caso que nos ocupa, el análisis de coincidencia se hizo con las series de tiempo diferenciadas, tanto de las 29 variables como del ISE, luego estaremos apareando valles con valles y picos con picos.

La selección de las variables finales se hizo de la siguiente manera. Inicialmente, se determinó el grado de coincidencia de cada una de las variables con respecto al ISE. Paralelamente, se calculó la función muestral de correlación cruzada entre cada variable y el ISE (en primeras diferencias) y se examinó la significancia del coeficiente de correlación cruzada en los rezagos $0, \pm 1, \pm 2, \pm 3$. Una variable con correlación cruzada significativa en alguno de esos rezagos, se puede considerar “coincidente” con el estado de la economía. Luego, tomando sector por sector, se hizo una evaluación de la importancia económica de cada variable en la dinámica del sistema económico; de esta manera, se llegó a un conjunto final de seis variables, cuya gráfica aparece en la Figura 1. Las letras SSA al final del nombre nemotécnico dado antes, indican series de tiempo desestacionalizadas (trabajo hecho con el paquete X13-ARIMA). Se puede observar, en términos muy generales, la presencia de dos movimientos seculares comunes, uno de ellos es sugerido por la variable CANASSA y el otro por las restantes cinco variables.

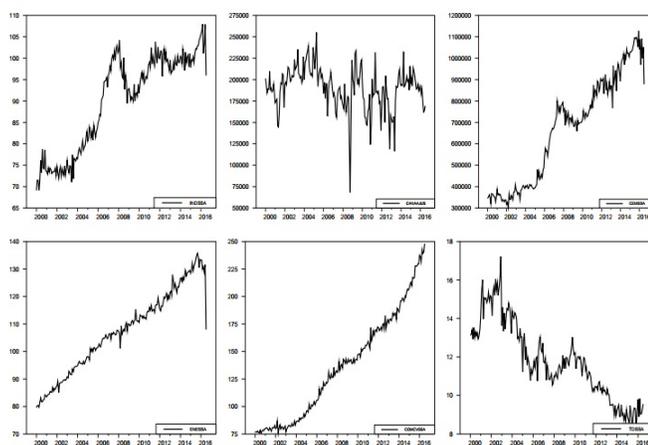


Figura 1: *Series temporales seleccionadas.*

Desde el punto de vista económico su descripción es la siguiente.

1. Producción industrial (IND). En muchos países, la industria es el sector de mayor calidad y confiabilidad para captar los ciclos de la actividad económica (Bujosa & García Ferrer 2000a). En Colombia, la actividad industrial representa uno de los componentes que más contribuye a la evolución del PIB. Para el 2000, la actividad manufacturera participaba con el 3,6 % del PIB en términos corrientes. Esta cifra cambió para el 2016, donde su participación creció 12,6 %. La Figura 2 muestra su perfil coincidente con respecto al ISE, indicando 0-coincidencia con un p -valor (nivel de confianza) del 100 %. Este es un resultado muy deseable.

2. Energía eléctrica (ENE). La generación, administración y distribución de energía eléctrica está directamente relacionada con el funcionamiento de la actividad económica. Su importancia se manifiesta en la contribución que esta actividad tiene sobre el crecimiento económico y su influencia en los diferentes sectores de la economía.

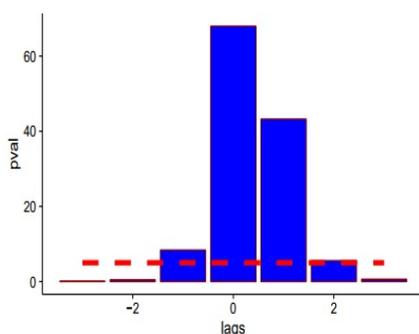


Figura 2: *Perfil coincidente de la industria.*

En Colombia, la producción de energía eléctrica ha tenido en promedio una participación del 2,3 % del PIB desde el 2000. En el análisis de coincidencia, la Figura 3 muestra -1 -coincidencia con un p -valor del 85 %. Este hecho puede interpretarse como un rezago de un mes entre los valles y picos del ciclo de esta variable y los del ISE. Cabe destacar que a pesar de la menor participación de esta actividad comparada con la industria, es imperante su inclusión pues es un soporte trascendental de la economía.

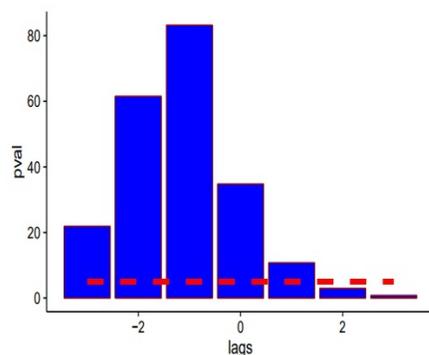


Figura 3: *Perfil coincidente de la producción de Energía.*

3. Producción de caña de azúcar (CANA). La inclusión de esta actividad puede generar cierta discusión, dado que los niveles de producción son bajos con respecto a otras variables de interés y, adicional, muestra una volatilidad muy alta. Sin embargo, durante los últimos años, el sector pasó de abastecer la demanda de los hogares, a canalizar sus recursos hacia la acumulación de materias primas para la producción de biocombustibles y otros derivados manufacturados. Si bien es una serie con presencia de heterocedasticidad y valores atípicos, su dinámica puede ser muy útil para reflejar los ciclos de la producción. Esto se demuestra en la Figura

4, donde se observa que el grado de coincidencia es 0 con un nivel de confianza del 75 %.

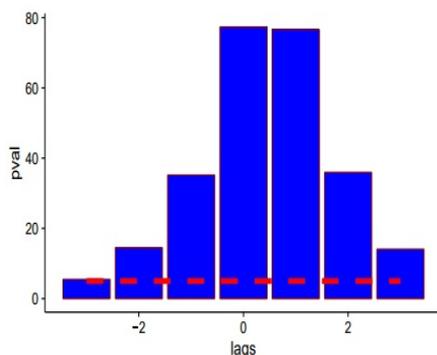


Figura 4: Perfil coincidente de la producción de caña.

4. Comercio al por menor sin combustible y sin vehículos (COMCV). En general, las ventas comerciales al por menor sin combustible y sin vehículos tienen la característica de reflejar una relación directa con el comportamiento del gasto de los hogares, ya que esta variable muestra la evolución de las líneas de mercancía que los hogares más consumen, como los alimentos, las bebidas alcohólicas, las prendas de vestir, los equipos electrónicos y electrodomésticos, entre otros. Esta relación permite ver, por ejemplo, cómo los ingresos reales pueden estar disminuyendo debido a la contracción en el gasto de los hogares y, por ende, que esta disminución sea el reflejo de una inminente inflación. En Colombia, la actividad comercial representó 7,4 % del PIB en precios corrientes en 2010 y aumentó a 7,7 % en 2016. Se espera que su contribución siga creciendo con el incremento de las importaciones y el debilitamiento de la industria nacional. La Figura 5 indica -2 -coincidencia con un nivel de confianza del 63,28 %, lo que implica que su ciclo se rezaga dos meses con respecto al del ISE.

5. Producción de cemento (CEM). Para la planificación gubernamental y la gestión fiscal, la construcción en obras civiles y edificaciones es una de las actividades que más repercute en el crecimiento económico. En esta actividad se mide, no solamente el grado de inversión por parte del sector oficial y privado, sino también su contribución a la actividad económica. La producción de cemento representa un insumo de vital importancia para la construcción de las edificaciones residenciales, no residenciales y obras civiles. Observando la Figura 6, la variable es 1-coincidente con un nivel de confianza del 95 %, lo cual revela un adelanto promedio de un mes de sus valles y picos con respecto a los del ISE.

6. Tasa de desempleo (TD). Esta variable es un indicador muy importante para el análisis macroeconómico, ya que tiene la característica de indicar las principales señales del estado de la economía (Leamer 2009). Por ejemplo, una fuerte alza en la tasa de desempleo muestra un claro síntoma de recesión. Asimismo, cuando la fase

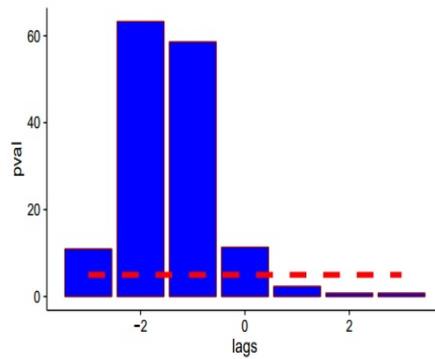


Figura 5: *Perfil coincidente del comercio.*

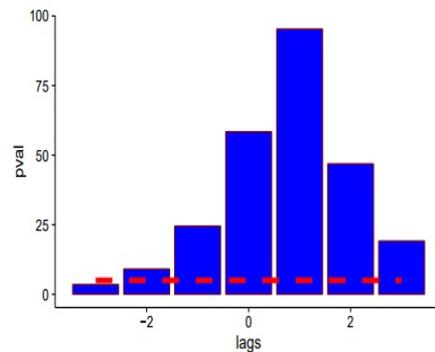


Figura 6: *Perfil coincidente de la producción de cemento.*

de recesión culmina, el nivel de desocupados se mantiene y se toma cierto tiempo para que empiece una fase de recuperación, pues, aunque las condiciones sean favorables, las empresas prefieren no comprometerse aún en contratar empleados. Ya para los primeros periodos de la recuperación, la tasa de desocupación puede disminuir y la economía puede repuntar. Como se puede apreciar en la Figura 7, esta variable es 0-coincidente con un nivel de confianza del 80%; en consecuencia, el indicador está sincronizado con la evolución del ciclo para el ISE.

4. El índice para Colombia

4.1. Proceso para su construcción

Con las variables coincidentes seleccionadas en la sección anterior, procedemos a encontrar un índice para la economía Colombiana.

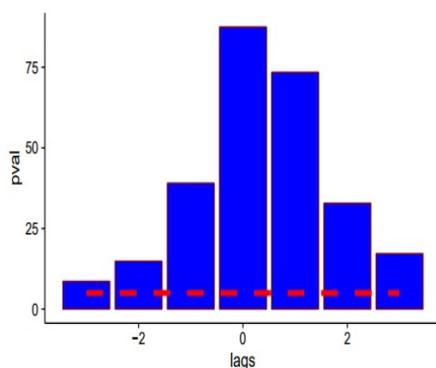


Figura 7: Perfil coincidente de la tasa de desempleo.

Etapa 1. Como se mencionó antes, cada variable fue desestacionalizada y, además, caña de azúcar y cemento son ajustadas por escala multiplicando por 10^{-3} la primera y por 10^{-4} la segunda. Aunque la variable CANASSA es la que presenta mayor número de datos atípicos, según un análisis hecho en el paquete estadístico TSW (Caporello & Maravall 2003), se decide no corregirla y utilizarla con sus valores originales en el análisis factorial, al igual que el resto de variables.

Etapa 2. Usando el método de (Lam & Yao 2012) se identificaron dos factores comunes para las variables en niveles. Luego, obtenemos estimaciones preliminares de los dos factores usando combinaciones lineales de las variables. Los coeficientes de estas combinaciones lineales son las componentes de los vectores propios, asociados a los dos primeros valores propios de la matriz M . En seguida, se especificaron modelos ARIMA para cada uno de los factores estimados preliminarmente y ellos fueron ARIMA(0,1,1) para el primer factor (valor propio mayor) y ARIMA(0,2,0) para el segundo (segundo mayor valor propio). Nótese que este último factor es un proceso de caminata aleatoria integrado (IEW por sus siglas en inglés), el cual es el modelo básico en el trabajo de (Bujosa et al. 2013) para la actividad económica de España. Utilizando la forma de espacio-estado del modelo factorial de (Peña & Poncela 2016), descrito completamente por (Nieto et al. 2016), se estimaron los hiperparámetros del modelo factorial y los factores. La estimación de la transpuesta de la matriz de carga, P^T , y de las varianzas de los ruidos intrínsecos fue la siguiente:

$$\hat{P}^T = \begin{bmatrix} 0.76 & 0.88 & 1.93 & 0.76 & 0.37 & 0.14 \\ 0.00 & 0.01 & -0.16 & 0.19 & 0.09 & -0.02 \end{bmatrix},$$

y $\hat{\Sigma}_e = \text{diag}\{21.66, 1.27, 489.88, 120.31, 20.83, 0.42\}$. Todos estos últimos parámetros son significativos al 1%. Además, la estimación del parámetro de promedio móvil del factor 1 es $\hat{\theta} = -0.18(0.13)$, error estándar en paréntesis, el cual, aunque es no significativo al 5%, se mantiene en el modelo. Según estas estimaciones, el primer factor es un promedio ponderado de las variables, mientras que el segundo refleja un contraste entre el grupo de variables conformado por energía, comercio

sin vehículos y cemento y el grupo integrado por caña de azúcar y tasa de desempleo. En la estimación de la matriz P se fijó el valor de la entrada (1, 2) en cero, con el objeto de completar las condiciones de identificabilidad del modelo factorial, como lo proponen (Peña & Poncela 2016).

Teniendo en cuenta que hemos seleccionado al ISE como la estimación preliminar del estado de la economía, en las Figuras 8 y 9 se presenta una comparación visual de este con cada factor común estimado, con el fin de tomar una idea acerca de las dinámicas del ISE y los factores comunes. Se observa en la Figura 8 que la dinámica del ISE y del factor 1 es similar, mientras que hay mucha disimilitud entre las dinámicas del ISE y del factor 2, como se aprecia en la Figura 9.

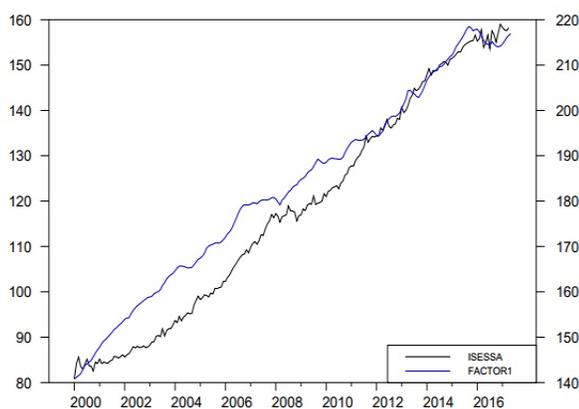


Figura 8: Comparación entre el ISE y el Factor 1 estimado.

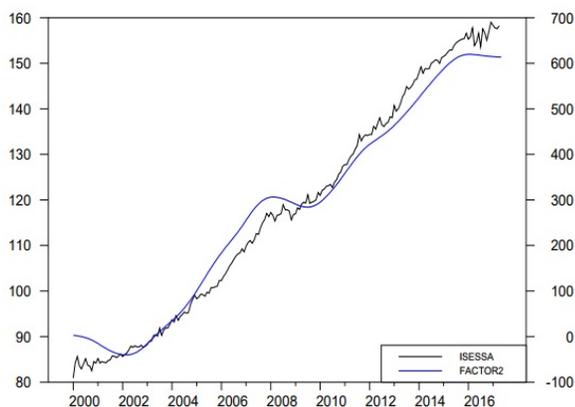


Figura 9: Comparación entre el ISE y el Factor 2 estimado.

Etapas 3. Para escoger el factor común buscando más coincidencia con el ISE, se calcula el perfil coincidente para cada uno de ellos. Siguiendo la literatura sobre

el tema, específicamente el método de (Bujosa et al. 2013), se obtuvo la primera derivada (diferencia finita) de las *tendencias* del ISE y de los factores comunes. Estas tendencias fueron extraídas usando el filtro de Hodrick y Prescott con una constante de suavizamiento $\lambda = 600$. Como lo explican (Bujosa et al. 2013), se busca extraer el ciclo de los negocios y, de esta manera, los perfiles coincidentes se obtienen para los ciclos. Este procedimiento es una modificación al de (Martínez et al. 2016), quienes usaron las series temporales en *niveles* (datos originales), para el cálculo de los perfiles coincidentes. En ese trabajo, la estimación preliminar del PIB mensual fue obtenida con el método de (Nieto et al. 2004).

En las Figuras 10 y 11 se presentan los perfiles coincidentes obtenidos para los ciclos del ISE y de cada factor. Como el factor 1 es 0-coincidente y el factor 2 no lo es, tomamos el primero como el índice coincidente. Nótese que, en términos de las estimaciones de las filas de P^T , el primer factor tiene más interpretabilidad que el segundo, en términos económicos.

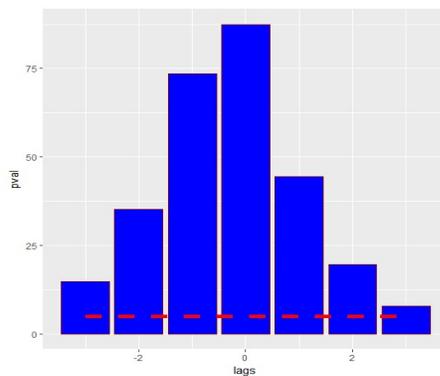


Figura 10: *Perfil coincidente entre el ciclo del ISE y del factor 1.*

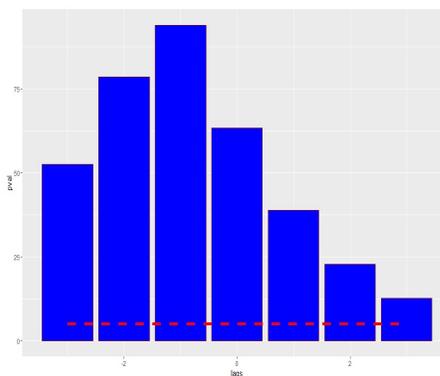


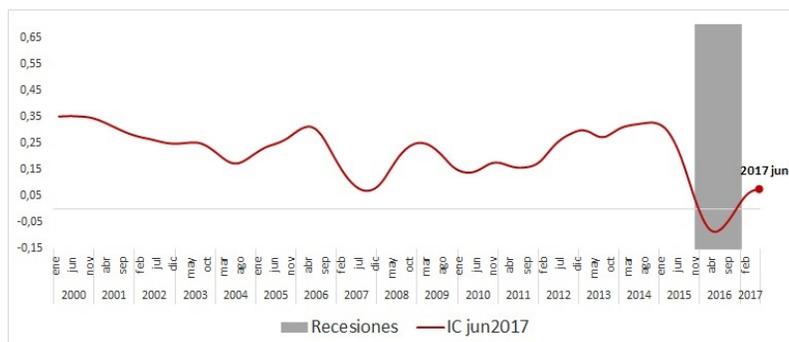
Figura 11: *Perfil coincidente entre los ciclos del ISE y del factor 2.*

Etapa 4. Con el fin de facilitar la comparación con respecto al ISE, el factor

Tabla 2: *Valles y picos de los ciclos del ISE y del índice coincidente*

Puntos de Giro	Posición - ISE	Posición - Factor 1	Diferencia Meses
Min	Feb, 2001	Feb, 2001	0
Max	Ago, 2006	Jun, 2006	2
Min	Ene, 2009	Nov, 2008	2
Max	Jul, 2011	Ene, 2011	6
Min	Feb, 2012	Jul, 2012	-5
Max	Jul, 2014	Abr, 2014	3

seleccionado es transformado a un índice con base promedio al 2005 y, a partir de ahí, se hace el siguiente análisis para la actividad económica colombiana. En la Tabla 2 presentamos los picos y valles del ciclo del índice coincidente y del ISE, obtenidos con el procedimiento de (Bry & Boschan 1971). Resultados similares fueron obtenidos con el enfoque de (Harding & Pagan 1999). En la segunda y tercera columnas, se presentan las localizaciones temporales de ellos (año-mes) y en la cuarta, se incluye la diferencia en meses entre las localizaciones temporales del índice y el ISE. Se puede observar que las diferencias en valor absoluto son menores o iguales a siete meses y que hay más coincidencia en los puntos temporales izquierda para el ISE y derecha para el índice (julio, 2003 y julio, 2003; marzo, 2008; y marzo, 2008; julio, 2009 y septiembre, 2009; febrero, 2012 y enero, 2012). En la Figura 12 se presenta la gráfica del ciclo del índice.

Figura 12: *Ciclo del índice*

A partir de los criterios de duración de las fases de un ciclo de (Bujosa et al. 2013) y comparando con los picos y valles de la Tabla 2, se identificó una señal para una eventual fase de recesión en el período enero, 2015-marzo, 2017, la cual se muestra en la Figura 12 usando un área sombreada. Estos puntos de la Tabla 2, junto con el ciclo en la Figura 12, permiten destacar varios periodos de pequeñas desaceleraciones, como es el caso de (1) octubre, 2003-junio, 2004; (2) junio, 2006-septiembre, 2007; (3) enero, 2011-julio, 2011; y una importante desaceleración en (4) abril, 2009-enero, 2010. La *presunta* fase de recesión y las fases de desaceleración son comparadas con los crecimientos anuales del PIB de Colombia para el

Tabla 3: *Crecimientos anuales PIB. Fuente: DANE*

Año	2001	2002	2003	2004	2005	2006	2007	2008	2009
%	1,7	2,5	3,9	5,3	4,7	6,7	6,9	3,5	1,7
Año	2010	2011	2012	2013	2014	2015	2016	2017	
%	4,0	6,6	4,0	4,9	4,4	3,1	2,0	1,2	

periodo 2000 - 2017², los cuales se muestran en la Tabla 3. Aunque no hay tasas de crecimiento negativas para este periodo, algunos años muestran tasas de crecimiento pequeño como es el caso de 2001, 2009, 2016 y 2017. Estos años de menor crecimiento coinciden precisamente con la fase (4) de alta desaceleración y la *eventual* fase de recesión.

4.2. Análisis de Coyuntura

Además de las interpretaciones económicas del índice presentadas en la etapa 4, puede usarse también para un análisis de coyuntura, el cual está basado en las variaciones anuales del índice. Estas están definidas por la expresión

$$VA_{m,t} = \left(\frac{I_{m,t}}{I_{m,t-1}} - 1 \right) \times 100 ,$$

Donde, en general, $I_{m,t}$ es el valor del índice en el mes m del año de referencia t , con $m \in \{Enero, \dots, Diciembre\}$. En la Figura 13 se observa un crecimiento en su variación anual de 1,5%; mientras que para el mismo periodo del 2015 presentaba una caída de 0,7%, lo que refleja un escenario de parsimonia de la economía colombiana. En un contexto general, desde el 2015 la economía viene presentando malos síntomas que están afectando su dinámica. El índice coincidente refleja estos indicios y avisa sobre la preocupante situación en 2016 y todavía más complejo para el primer semestre de 2017. Sin embargo, hay que destacar que el segundo trimestre de 2017 muestra un escenario mas alentador pero a tasas de crecimiento lentas.

4.2.1. Contexto regional

El ciclo del índice se puede utilizar en el análisis del contexto regional latinoamericano. Para el período junio, 2017, las principales economías latinoamericanas mostraron un escenario parecido al de la economía nacional. En efecto, el índice coincidente para Colombia tuvo un crecimiento de 1,5%; el índice mensual de la actividad económica para Argentina (EMAE), producido por el INDEC, creció 3,6% y el índice de la actividad económica del Banco Central del Brasil (IBC) creció 0,9%. De la misma manera, el índice mensual de la actividad económica

²La tasa de crecimiento de 2017 corresponde al periodo Enero-Junio

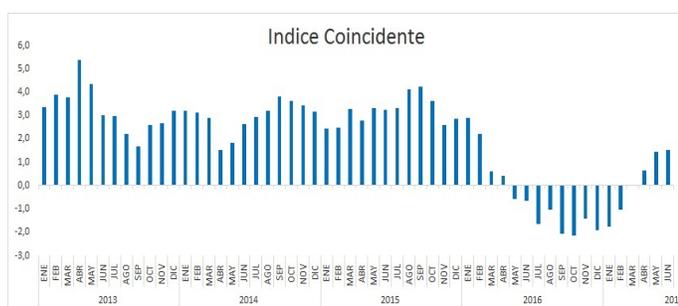


Figura 13: *Evolución de las variaciones anuales hasta junio, 2017.*

para Chile (IMAEC) creció 1,4 %. Para el caso de México y Perú, se mantiene una estabilidad positiva, donde el índice global de la actividad económica para México (IGAE) reportó un crecimiento de 2,5 % y el índice mensual de la producción nacional de Perú (IMPV) creció 2,8 % para el mismo periodo. En la Figura 14 se presentan las gráficas de estas variaciones anuales. El análisis muestra que para este periodo, hay una similitud en las dinámicas de las economías de la región.



Figura 14: *Evolución de las variaciones anuales regionales hasta junio, 2017.*

Con el ciclo del índice se puede realizar también un análisis sectorial de la actividad económica colombiana, a través de los indicadores utilizados para la construcción del índice. Por ejemplo, cabe destacar que aunque la producción de cemento y el índice de producción industrial tuvieron tasas de crecimiento anual negativas (-1,8 % y -2,1 % respectivamente), los demás sectores tuvieron un comportamiento favorable, lo que no permitió que el ciclo tomara valores negativos.

5. Conclusiones y recomendaciones

En el presente trabajo, se ha construido un índice coincidente para la actividad económica de Colombia, siguiendo el enfoque de perfil coincidente de Martínez *et al* (2016). Dentro de este proceso, también se ha mejorado el procedimiento de

los anteriores autores tanto en el marco teórico como en el metodológico; adicionalmente, se utilizó un conjunto nuevo de variables macroeconómicas y el periodo muestral analizado cubre hasta junio de 2017, mes para el cual se tienen los últimos datos de las variables involucradas. Como líneas futuras de investigación en el tema, se propone (a) utilizar este índice como el ciclo de referencia para construir un índice líder para la actividad económica de Colombia y explorar la posibilidad de incluir variables climatológicas, en el grupo de variables seleccionadas en este trabajo y (b) determinar las propiedades estadísticas de la prueba de permutación, que está detrás de la construcción de los perfiles coincidentes, en este caso de datos autocorrelacionados.

AGRADECIMIENTOS

Los autores de este trabajo agradecen a la Universidad Autónoma de Madrid (UAM) y al Banco Santander, quienes a través del proyecto CEAL-AL/2015-11, financiaron el desarrollo del mismo. A su vez, nuestros sinceros agradecimientos a los profesores Antonio García-Ferrer, Pilar Poncela y Aranzazu de Juan Fernández, de la UAM y miembros del equipo de investigación, por sus valiosos aportes al tema. A la Universidad Nacional de Colombia por facilitar la conformación conjunta del equipo español-colombiano de investigación y proporcionarnos el tiempo laboral para trabajar en el proyecto. Asimismo,, agradecemos inmensamente al evaluador anónimo de la revista por sus valiosos comentarios, los cuales nos permitieron mejorar la presentación del presente trabajo.

Recibido: 7 de Septiembre de 2017

Aceptado: 10 de Abril de 2018

Referencias

- Bry, G. & Boschan, E. (1971), 'Programmed selection of cyclical turning points in cyclical analysis of time series: Selected procedures and computer programs', *NBER* pp. 7-63.
- Bujosa, M. & García Ferrer, A. (2000a), 'Forecasting oecd industrial turning points using unobserved component models with business survey data', *International Journal of Forecasting* pp. 207-227.
- Bujosa, M. & García Ferrer, A. (2000b), 'Forecasting oecd industrial turning points using unobserved component models with business survey data', *International Journal of Forecasting* pp. 207-227.
- Bujosa, M., García Ferrer, A. & de Juan, A. (2013), 'Predicting recessions with factor linear dynamic harmonic regressions', *Journal of Forecasting* pp. 481-499.
- Caporello, G. & Maravall, A. (2003), 'Tsw program', *Banco de España* .

- DANE (2016), ‘Metodología general indicador de seguimiento a la economía ise’.
*<http://dane.gov.co/files/investigaciones/fichas/cuentas-nacionales/DSO-CT-MET-Metodologia-General-Indicador-de-Seguimiento-a-la-Economia-ISE-Agosto-2016.pdf>
- Harding, D. & Pagan, A. (1999), ‘Dissecting the cycle’, *Melbourne Institute Working paper* (13).
- Lam, C. & Yao, Q. (2012), ‘Factor modeling for high-dimensional time series: inference for the number of factors’, *The Annals of Statistics* (40), 694–726.
- Leamer, E. (2009), ‘Macroeconomic patterns and stories’, *Springer: Berlin* .
- Martínez, W., Nieto, F. & Poncela, P. (2016), ‘Choosing a dynamic common factor as a coincident index’, *Statistics and Probability Letters* (109), 89–98.
- Melo, L., Nieto, F., Posada, c., Betancourt, Y. & Barón, J. (2001), ‘Un índice coincidente para la actividad económica de colombia’, *ENSAYOS sobre Política Económica* (40), 46–88.
- Nieto, F. (2004), ‘Ex-post and ex-ante prediction of unobserved economic time series: a case study’, *Journal of Forecasting* (17), 35–58.
- Nieto, F. H., Peña, D. & Saboya, D. (2004), ‘A coincident index for the state of the economy’, *International Statistical Review*, (72), 1237–1257.
- Nieto, F. H., Peña, D. & Saboya, D. (2016), ‘Common seasonality in multivariate time series’, *Statistica Sinica* pp. 1389–1410.
- Peña, D. & Poncela, P. (2016), ‘Nonstationary dynamic factor analysis’, *Journal of Statistical Planning and Inference* pp. 1237–1257.
- Stock, J. & Watson, M. (1989), ‘New indexes of coincident and leading economic indicators’, *JNBER Macroeconomics Annual* (4), 351–394.
- Stock, J. & Watson, M. (1991), ‘A probability model of the coincident economic indicators. in lahiri, k., moore, g.h. (eds.), leading economic indicators: New approaches and forecasting records’, *Cambridge University Press: New York* (4), 63–85.

A. Apéndice

Con el fin de darle un soporte formal a la prueba de permutación que fundamenta la construcción del índice y a partir del investigar las propiedades estadísticas de la prueba de permutación, se considera el siguiente marco teórico:

Sean Y y Z los tiempos de ocurrencia (por ejemplo, meses, trimestres, etc.) de los puntos de giro de $\{C_t\}$ e $\{I_t\}$, respectivamente. Ellos son variables aleatorias sobre un espacio de probabilidad, por ejemplo (Ω, \mathcal{A}, P) , con valores en \mathbb{Z}^+ , el conjunto

de los números enteros positivos. Considérese el vector aleatorio (Y, Z) el cual está definido sobre (Ω, \mathcal{A}, P) y toma valores en $\mathbb{Z}^+ \times \mathbb{Z}^+$, el producto cartesiano de \mathbb{Z}^+ consigo mismo. Así, dado $\omega \in \Omega$, $(Y, Z)(\omega) = (Y(\omega), Z(\omega)) = (t, s)$, donde $(t, s) \in \mathbb{Z}^+ \times \mathbb{Z}^+$. Ahora, se define la variable aleatoria $D = Y - Z$ sobre el espacio muestral de (Y, Z) con valores en \mathbb{Z} . Acá, $D(t, s) = t - s$ para todo $(t, s) \in \mathcal{R}_{(Y, Z)}$, el recorrido o espacio muestral de (Y, Z) . Es claro que la distribución de D , representada por la probabilidad inducida $P_{(Y, Z)}$ de (Y, Z) , no es fácilmente conocida, a menos que se consideren supuestos adicionales. Ni tampoco las marginales de Y y Z .

Sea m la mediana de D . El conjunto de hipótesis propuesto en la subsección 2.2. puede ser re-escrito como $H : m = 0$ v.s. $K : m \neq 0$. Como una estadística de prueba, se propone a $S = \sum_{i=1}^Q D_i$, donde D_i denota la diferencia D en el par i , $i = 1, \dots, Q$. Bajo la nula, las diferencias de pares están alrededor de cero en una forma simétrica aproximada, entonces su suma es cercana a cero. En consecuencia, valores de S lejos de cero señalarán un rechazo de H .

Ahora, para analizar formalmente las permutaciones de las diferencias signadas como se indicó en la Subsección 2.2, se considera el siguiente esquema:

El mecanismo probabilístico que genera los signos $+, -$ puede ser representado por el espacio de probabilidad $(\Omega_g = \{+, -\}, \mathbb{F} = \mathcal{P}(\Omega_g), P_g)$, donde $\mathcal{P}(\Omega_g)$ es el conjunto de todos los subconjuntos de Ω_g . Bajo equiprobabilidad, $P_g(\{+\}) = 1/2$. Ahora, se considera el espacio de medida producto $(\Omega_g^Q, \mathcal{F}, \bar{P})$, donde $\Omega_g^Q = \times_{i=1}^Q \Omega_g$ es el producto cartesiano de Ω_g consigo mismo Q veces, \mathcal{F} es la σ -álgebra generada por el producto cartesiano de \mathbb{F} consigo mismo Q veces y \bar{P} es la medida producto. Nótese que $\bar{P}(\{\delta_1, \dots, \delta_Q\}) = \prod_{i=1}^Q P_g(\{\delta_i\}) = 1/2^Q$, con $\delta_i \in \{+, -\}$, $i = 1, \dots, Q$. Fijando $|d_1|, \dots, |d_Q|$, se define el elemento aleatorio A de $(\Omega_g^Q, \mathcal{F}, \bar{P})$ en el espacio medible $(\mathbb{Z}^Q, \mathcal{P}(\mathbb{Z}^Q))$, donde $\mathbb{Z}^Q = \times_{i=1}^Q \mathbb{Z}$ y, como antes, \mathcal{P} indica el conjunto de partes, tal que $A(\delta_1, \dots, \delta_Q) = (\delta_1 |d_1|, \dots, \delta_Q |d_Q|)$ con $\delta_i \in \{+, -\}$. Nótese que esta imagen por A es una sucesión signada de las diferencias observadas correspondiendo a la permutación $(\delta_1, \dots, \delta_Q)$. Sea \bar{P}_A la medida inducida por A , entonces $\bar{P}_A(A(\delta_1, \dots, \delta_Q)) = 1/2^Q$, para todo $(\delta_1, \dots, \delta_Q) \in \Omega_g^Q$. Esta medida de probabilidad será considerada la distribución de permutación.

Sea s_j el valor observado de la estadística S en la j -ésima sucesión signada en el espacio muestral de A , $j = 1, \dots, Q$. Entonces $s_0 = s_{j_0}$ para algún $j_0 = 1, \dots, Q$. También, sea

$$p_0 = \frac{\#\{s_j : |s_j| \geq |s_{j_0}|, j = 1, \dots, Q\}}{2^Q},$$

Donde $\#$ denota el cardinal del conjunto en referencia. Este número es el p -valor de la sucesión signada observada bajo la distribución de permutación \bar{P}_A .

Observación. S es un elemento aleatorio de $(\mathbb{Z}^Q, \mathcal{P}(\mathbb{Z}^Q))$ en el espacio medible $(\mathbb{Z}, \mathcal{P}(\mathbb{Z}))$. Sea $\bar{P}_{A, S}$ su medida de probabilidad inducida. Esta no es necesariamente equiprobable pero si $B = \{y \in \mathbb{Z} : |y| \geq |s_{j_0}|\}$, entonces $\bar{P}_{A, S}(B) = \bar{P}_A(S^{-1}(B)) = p_0$, donde $S^{-1}(B)$ es la imagen inversa de B por S .

Al nivel de significancia α , $0 < \alpha < 1$, la regla de decisión es rechazar la hipótesis nula y aceptar la alternativa si $p_0 \leq \alpha$. Un rechazo de la hipótesis nula significa que

s_{j_0} es un valor extremo de S bajo la distribución $P_{A,S}$ o que la sucesión observada d_1, \dots, d_Q es un resultado extremo bajo la distribución P_A .