

Pronósticos de corto plazo en tiempo real para la actividad económica colombiana*

Deicy J. Cristiano[†], Manuel D. Hernández[†], José David Pulido[‡]
Banco de la República

Resumen

La toma de decisiones de política económica requiere estimaciones del comportamiento de la actividad económica en tiempo real. Sin embargo, la información utilizada solo está disponible a nivel de indicadores de actividad y de encuestas de opinión, los cuales suelen tener distintas frecuencias y rezagos de publicación, además de choques idiosincráticos. En este trabajo se adaptan para la economía colombiana los esquemas de pronóstico de Camacho y Perez-Quiros (2009,2010) que producen estimaciones del crecimiento del PIB en tiempo real. El modelo de factores dinámicos adaptado involucra series de actividad de diferente frecuencia, disponibilidad y procedencia, empleadas con la información disponible en el momento de cada publicación. La evaluación de pronóstico sugiere que el modelo presenta un mejor desempeño frente a otros esquemas de referencia, y que la precisión de los pronósticos aumenta al incorporar el flujo de información en tiempo real de los indicadores de actividad.

Clasificación JEL: C22, C53, E27

Palabras clave: crecimiento del producto, pronóstico en tiempo real, modelo de factores dinámicos.

* Las opiniones expresadas en este documento son responsabilidad exclusiva de los autores y no comprometen ni al Banco de la República ni a su Junta Directiva. Así mismo los errores que pueda tener el documento son su responsabilidad exclusiva. Agradecemos los comentarios de Eliana González y Daniel Parra y la colaboración de Alejandra Hernández y Evelyn Tique.

[†] Profesional del Departamento Técnico y de Información Económica

[‡] Profesional especializado del Departamento de Programación Macroeconómica e Inflación

1. Introducción

La toma de decisiones de política monetaria requiere estimaciones del comportamiento de la actividad económica con la mayor oportunidad posible. En especial, dados los rezagos con que las medidas de política se transmiten a la economía, resulta indispensable contar con diagnósticos sobre la coyuntura económica en tiempo real, aprovechando de la mejor manera posible el conjunto de información disponible en cada momento del tiempo.

En Colombia, así como en la mayoría de países emergentes, las series relevantes para evaluar oportunamente el estado de la economía provienen principalmente de indicadores de actividad sectorial, variables monetarias y financieras y de los resultados de encuestas de opinión y expectativas. Para integrar estos indicadores en la generación de pronósticos es necesario tener en cuenta distintos aspectos del conjunto de información en tiempo real: las diferencias en frecuencias y en rezagos de publicación, la heterogeneidad de las series por cuenta de su construcción metodológica y la presencia de factores idiosincráticos debido al carácter sectorial o parcial de las series.

En primer lugar, existen diferencias notorias en la frecuencia y los rezagos de publicación de las series colombianas. Por ejemplo, mientras que la demanda de energía eléctrica se divulga a los cinco días de finalizado el mes, el índice de producción industrial, otro indicador mensual, se produce con dos meses de rezago. Por otra parte, el censo de la actividad edificadora, de frecuencia trimestral, se difunde dos meses y medio después de finalizado el trimestre, justo dos semanas antes de que se dé a conocer el PIB.

En segundo lugar, la construcción metodológica de algunos indicadores lleva a que estos muestren un grado de asociación únicamente con determinadas medidas de crecimiento de la actividad económica. Por ejemplo, los balances de las encuestas de opinión, están correlacionados solamente con la variación anual de sus agregados económicos asociados¹. Por último, algunas series están sujetas tanto a choques idiosincráticos como a revisiones de diferente magnitud, que pueden distorsionar los juicios realizados sobre la evolución de la economía en su conjunto.

Aunque el estudio del impacto de estas dificultades no es nuevo², recientemente la literatura académica y las autoridades de política han avanzado en el tratamiento econométrico de la información disponible en tiempo real. Por ejemplo, en 2008 se creó por parte de la *Federal Reserve Bank of Philadelphia* el *Real-Time Data Research*

¹Al respecto, Julio y Grajales (2011, p.17) afirman: “*Los promedios trimestrales de los índices agregados de confianza son indicadores contemporáneos moderadamente buenos del crecimiento anual del PIB... (pero) ...NO son indicadores contemporáneos ni líderes del crecimiento trimestral del PIB*”.

²Quizá el primer trabajo en dar un tratamiento estadístico al problema de información incompleta propio de la información en tiempo real (“*the ‘ragged edge’ problem*”) sea el de Wallis (1986).

Center, con el fin de estudiar los problemas asociados con dicho tipo de información³. Bouwman y Jacobs (2005) señalan tres áreas principales en donde se han concentrado la mayoría de los esfuerzos: revisiones de datos, pronóstico y análisis de política. Este trabajo se enmarca en la segunda de estas categorías⁴.

En lo que respecta a pronóstico, existen varias técnicas para aprovechar conjuntos de información de grandes dimensiones. Stock y Watson (2006) señalan que las más habituales se basan en modelos de factores dinámicos, regresiones tipo *bridge*, técnicas bayesianas y combinación de pronósticos. Sin embargo, en el contexto de información de frecuencia mixta, no balanceada y sujeta a choques idiosincráticos, la alternativa más utilizada suele ser los esquemas de factores dinámicos (Forni *et al.* 2000, Kapetianos, 2006, Doz *et al.* 2006, 2007). En particular, el uso del filtro de Kalman en la estimación por máxima verosimilitud de estos modelos permite tratar de manera directa el problema de observaciones faltantes, propio de conjuntos de información no balanceados.

Haciendo uso de este tipo de modelos este trabajo adapta para la economía colombiana los esquemas de pronóstico de corto plazo de Camacho y Perez-Quiros (2009, 2010 – en adelante CP), que generan estimaciones del crecimiento del PIB a partir de la información disponible al momento de ser divulgado un indicador. Su estructura permite aprovechar la información parcial o completa de los indicadores de actividad económica para obtener estimaciones del crecimiento de la economía en el mismo periodo, lo que es relevante dado el mayor rezago de publicación del PIB. Esta técnica, que se conoce en la literatura como *nowcast* (Giannone *et al.* 2005, 2008) permite actualizar en tiempo real las estimaciones de crecimiento, una vez se encuentra disponible nueva información. Además, posibilita la cuantificación del impacto de las nuevas cifras divulgadas de los indicadores de actividad económica sobre los pronósticos de crecimiento.

Este trabajo entra a hacer parte de una serie de esfuerzos relativamente recientes en la región por contar con esquemas para producir *nowcasts* de crecimiento. Liu *et al.* (2010) evalúan modelos univariados, de vectores autoregresivos y de factores dinámicos para generar estimaciones de crecimiento en tiempo real para diez países latinoamericanos. D'Amato *et al.* (2010) usan un modelo de factores dinámicos y combinaciones de pronósticos para obtener *nowcasts* de crecimiento para Argentina. Pedersen (2010) emplea datos en tiempo real para extraer señales sobre el crecimiento del índice coincidente de la economía chilena. Respecto a otras economías emergentes, los trabajos de Maier (2011) y Bhattacharya *et al.* (2011) para China e India

³ Véase <http://www.phil.frb.org/research-and-data/real-time-center/>

⁴ Para Colombia, la modelación de revisiones de datos en un contexto de información en tiempo real ha sido realizada por Julio (2011), que extiende en varios frentes el trabajo de Jacobs y van Norden (2010).

respectivamente, que utilizan enfoques univariados y regresiones tipo *bridge*, se constituyen en otras referencias recientes⁵.

Este documento está organizado como sigue. En la segunda sección se describe la especificación del modelo empleado. En la tercera se comentan los criterios para la selección de variables y se caracteriza el tipo de información disponible en tiempo real para la economía colombiana. En la cuarta se muestra la evaluación del desempeño del modelo y en la quinta se concluye.

2. Especificación del modelo

En esta sección se adapta la representación estado-espacio del modelo de factores dinámicos empleado en CP para la economía colombiana. La principal diferencia radica en que para el caso colombiano no es posible incorporar la modelación de las revisiones entre las distintas versiones de publicación de cada dato del PIB, ya que en Colombia se hace una sola divulgación por cada dato y las revisiones se hacen a partir de la publicación del siguiente dato. En ese contexto, el tratamiento de las revisiones de la serie del PIB de Colombia mediante un modelo de factores dinámicos es realizada por Julio (2011).

En primer lugar, se debe considerar cómo incorporar frecuencias mixtas en la representación estado espacio. Si bien en Colombia la mayoría de series de actividad económica poseen frecuencia mensual, el PIB y unos pocos (pero relevantes) indicadores de actividad son trimestrales⁶. Sea G_T un dato del trimestre T , y X_t su realización en el último mes (t) del trimestre, desconocida. La media geométrica de los X que conforman a G_T multiplicada por 3 puede ser una buena aproximación a G_T si las tasas de crecimiento mensuales son pequeñas, esto es⁷:

$$G_T \approx 3(X_t X_{t-1} X_{t-2})^{1/3} \quad (1)$$

El crecimiento trimestral de la variable de interés, $g_T = \ln G_T - \ln G_{T-1}$, queda definido entonces en términos de los crecimientos mensuales $x_t = \ln X_t - \ln X_{t-1}$ mediante la relación⁸:

⁵ En las economías desarrolladas los esquemas para producir *nowcasts* tienen una difusión mucho más amplia. Véase por ejemplo, Evans (2005), Angelini *et al.* (2008), Barhoumi *et al.* (2008), Giannone *et al.* (2009), Banbura y Giannone (2010), Kuzin *et al.* (2011), Lahiri y Monokroussos (2011) entre otros.

⁶ Por ejemplo, el censo de edificaciones o el indicador de obras civiles son variables cruciales para el pronóstico de PIB con frecuencia trimestral.

⁷ Técnicamente, el empleo de la media geométrica en lugar de la aritmética se prefiere para evitar la introducción de no linealidades en el modelo estado espacio. La aproximación a través de la media geométrica es propuesta por Mariano y Murasawa (2003).

⁸ Ya que $g_T \approx \ln 3 + \frac{1}{3}(\ln X_t + \ln X_{t-1} + \ln X_{t-2}) - \ln 3 - \frac{1}{3}(\ln X_{t-3} + \ln X_{t-4} + \ln X_{t-5})$, esta expresión se puede manipular sumando y restando $\ln X_{t-i}$ con $i = 1 \dots, 4$ para obtener (2).

$$g_T \approx \frac{1}{3}x_t + \frac{2}{3}x_{t-1} + x_{t-2} + \frac{2}{3}x_{t-3} + \frac{1}{3}x_{t-4} \quad (2)$$

Expresión que permite incorporar en el modelo los crecimientos de las variables trimestrales en términos de variaciones mensuales ponderadas, tasas en las que estarán expresados la mayor parte de indicadores de actividad.

El siguiente paso consiste en definir un método para utilizar las series cuya asociación con el PIB solo se evidencie respecto a una determinada tasa de crecimiento. Es el caso de los balances de respuesta de las encuestas de opinión y expectativas, que, como lo documentan Julio y Grajales (2011), son indicadores contemporáneos solamente del crecimiento anual del PIB, por lo que no son útiles para predecir su crecimiento trimestral. Siguiendo a CP a dichas variables las denominaremos indicadores “*soft*”, para distinguirlas del resto de series de actividad económica (“*hard*”)⁹.

Sea entonces Z_t el conjunto de todas las series de actividad económica mensuales en el mes t , que está dividido entre indicadores *soft* (Z_t^s) y *hard* (Z_t^h). El dato en el mes t del i -ésimo indicador *soft* (Z_{it}^s) se relaciona con el crecimiento anual en dicho mes de un agregado X_t asociado, esto es $Z_{it}^s = \ln X_t - \ln X_{t-12}$. Un poco de algebra¹⁰ lleva a una relación entre Z_{it}^s y los crecimientos mensuales de X , $x_t = \ln X_t - \ln X_{t-1}$:

$$Z_{it}^s = \sum_{j=0}^{11} x_{t-j} \quad (3)$$

Expresión que, al igual que lo hecho con los crecimientos de las variables trimestrales en (2), permite incorporar en el modelo las series provenientes de los balances de respuesta de las encuesta de opinión en términos de variaciones mensuales, tasas en las que también se expresarán los indicadores *hard*.

Para la representación de estado espacio del modelo se utiliza junto con el crecimiento trimestral del PIB, y_t , el conjunto de indicadores mensuales Z_t y un grupo adicional de variables trimestrales expresadas en tasas de crecimiento, Z_t^q . Sea r la cantidad de indicadores en Z_t ($r = r^h + r^s$, r^h cantidad de *hard* y r^s de *soft*) y r^q el número de series en Z_t^q . Todas las variables a incluir en el esquema son modeladas a partir de la suma de dos componentes, el factor común a todas las series f_t , y el componente idiosincrático para cada una de las series, u_{1t} para y_t , $U_t^h = (v_{1,t}, v_{2,t} \dots v_{r^h,t})$ para Z_t^h , $U_t^s = (v_{r^h+1,t}, v_{r^h+2,t} \dots v_{r^h+r^s,t})$ para Z_t^s y $U_t^q = (u_{2,t}, u_{3,t} \dots u_{r^q+1,t})$ para Z_t^q .

La ecuación de medida debe ser formulada teniendo en cuenta la distinta frecuencia de publicación de cada grupo de series y la clasificación de los indicadores mensuales entre

⁹ La denominación *soft* hace referencia a la menor volatilidad que presentan los crecimientos anuales (con los que dichas series tienen correlación) en comparación con los crecimientos de mayor frecuencia.

¹⁰ La definición de $Z_{it}^s = \ln X_t - \ln X_{t-12}$ se puede manipular sumando y restando $\ln X_{t-i}$ con $i = 1 \dots 11$ para obtener (3).

hard y *soft*. En particular, f_t medirá el componente común de la evolución mensual de todas las series, por lo que cada conjunto de series debe expresarse en términos de crecimientos mensuales. Para las series en crecimientos trimestrales (y_t y Z_t^q) se emplea la expresión (2), para Z_t^s se emplea la expresión (3), mientras que las variables en Z_t^h , que están en crecimientos mensuales desestacionalizados, no requieren transformación. Por tanto, la ecuación de medida es de la forma:

$$\begin{pmatrix} y_t \\ Z_t^h \\ Z_t^s \\ Z_t^q \end{pmatrix} = \begin{pmatrix} \beta_1 \left(\frac{1}{3} f_t + \frac{2}{3} f_{t-1} + f_{t-2} + \frac{2}{3} f_{t-3} + \frac{1}{3} f_{t-4} \right) \\ \beta_2 f_t \\ \beta_3 \sum_{j=0}^{11} f_{t-j} \\ \beta_4 \left(\frac{1}{3} f_t + \frac{2}{3} f_{t-1} + f_{t-2} + \frac{2}{3} f_{t-3} + \frac{1}{3} f_{t-4} \right) \end{pmatrix} + \begin{pmatrix} \frac{1}{3} u_{1t} + \frac{2}{3} u_{1t-1} + u_{1t-2} + \frac{2}{3} u_{1t-3} + \frac{1}{3} u_{1t-4} \\ U_t^h \\ U_t^s \\ \frac{1}{3} u_{lt} + \frac{2}{3} u_{lt-1} + u_{lt-2} + \frac{2}{3} u_{lt-3} + \frac{1}{3} u_{lt-4} \end{pmatrix} \quad (4)$$

Y la dinámica del modelo se completa al suponer que:

$$f_t = a_1 f_{t-1} + \dots + a_{m_1} f_{t-m_1} + \epsilon_t^f \quad (5)$$

$$u_{1t} = b_1 u_{1t-1} + \dots + b_{m_2} u_{1t-m_2} + \epsilon_t^{u_1} \quad (6)$$

$$v_{jt} = c_{j1} v_{jt-1} + \dots + c_{jm_3} v_{jt-m_3} + \epsilon_t^{v_j} \quad (7)$$

$$u_{lt} = d_{l1} u_{lt-1} + \dots + d_{lm_4} u_{lt-m_4} + \epsilon_t^{u_l} \quad (8)$$

Con $\epsilon_t^f \stackrel{iid}{\sim} N(0, \sigma_f^2)$, $\epsilon_t^{u_1} \stackrel{iid}{\sim} N(0, \sigma_{u_1}^2)$, $\epsilon_t^{v_j} \stackrel{iid}{\sim} N(0, \sigma_{v_j}^2)$, $\epsilon_t^{u_l} \stackrel{iid}{\sim} N(0, \sigma_{u_l}^2)$, $j = 1, \dots, r$, $l = 2, \dots, (r^q + 1)$, $m_1 = m_2 = m_4 = 6$ y $m_3 = 2$ y todas las covarianzas iguales a cero¹¹.

Dicha ecuación de medida puede ser reescrita en la forma compacta:

$$Y_t = H h_t + w_t \quad (9)$$

Con $w_t \stackrel{iid}{\sim} N(0, R)$, donde Y_t, h_t, H, w_t y R vienen dadas por:

$$Y_t = (y_t, Z_t^h, Z_t^s, Z_t^q)'$$

$$h_t = (f_t, \dots, f_{t-11}, u_{1t}, \dots, u_{1t-5}, v_{1t}, v_{1t-1}, \dots, v_{rt}, v_{rt-1}, u_{2t}, \dots, u_{2t-5}, \dots, u_{r^q+1,t}, \dots, u_{r^q+1,t-5})'$$

¹¹ El modelo queda identificado asumiendo $\sigma_f^2 = 1$.

$$H = \begin{pmatrix} H_{11} & 0_{1,6} & H_{12} & 0_{1,2r^h} & 0_{1,2r^s} & 0_{1,6r^q} \\ H_{21} & 0_{r^h,6} & 0_{r^h,6} & H_{22} & 0_{r^h,2r^s} & 0_{r^h,6r^q} \\ H_{31} & H_{31} & 0_{r^s,6} & 0_{r^s,2r^h} & H_{32} & 0_{r^s,6r^q} \\ H_{41} & 0_{r^q,6} & 0_{r^q,6} & 0_{r^q,2r^h} & 0_{r^q,2r^s} & H_{42} \end{pmatrix}$$

dónde:

$$\begin{aligned} H_{11} &= \begin{pmatrix} \frac{\beta_1}{3} & \frac{2\beta_1}{3} & \beta_1 & \frac{2\beta_1}{3} & \frac{\beta_1}{3} & 0 \end{pmatrix} \\ H_{12} &= \begin{pmatrix} \frac{1}{3} & \frac{2}{3} & 1 & \frac{2}{3} & \frac{1}{3} & 0 \end{pmatrix} \\ H_{21} &= 0_{r^h,6} \text{ con primera columna } \beta_2 \\ H_{22} &= I_{r^h} \otimes (1 \ 0) \\ H_{31} &= \beta_3 * 1_{r^s,6} \\ H_{32} &= I_{r^s} \otimes (1 \ 0) \\ H_{41} &= 1_{r^q,1} \otimes \begin{pmatrix} \frac{\beta_4}{3} & \frac{2\beta_4}{3} & \beta_4 & \frac{2\beta_4}{3} & \frac{\beta_4}{3} & 0 \end{pmatrix} \\ H_{42} &= I_{r^q} \otimes \begin{pmatrix} \frac{1}{3} & \frac{2}{3} & 1 & \frac{2}{3} & \frac{1}{3} & 0 \end{pmatrix} \end{aligned}$$

$$w_t = 0_{1+r+r^q,1} \text{ y } R = 0_{(1+r+r^q),(1+r+r^q)},$$

A su vez, la forma compacta de la ecuación de transición viene dada por:

$$h_t = Fh_{t-1} + \xi_t, \quad (10)$$

con $\xi_t \stackrel{iid}{\sim} N(0, Q)$, donde F y Q vienen dadas por:

$$F = \begin{pmatrix} a & 0_{12,6} & 0_{12,2r^h} & 0_{12,2r^s} & 0_{12,6r^q} \\ 0_{6,12} & b & 0_{6,2r^h} & 0_{6,2r^s} & 0_{6,6r^q} \\ 0_{2r^h,12} & 0_{2r^h,6} & c_{ih} & 0_{2r^h,2r^s} & 0_{2r^h,6r^q} \\ 0_{2r^s,12} & 0_{2r^s,6} & 0_{2r^s,2r^h} & c_{is} & 0_{2r^s,6r^q} \\ 0_{6r^q,12} & 0_{6r^q,6} & 0_{6r^q,2r^h} & 0_{6r^q,2r^s} & d \end{pmatrix}$$

con:

$$a = \begin{pmatrix} a_1 & \dots & a_6 & \dots & 0 & 0 \\ 1 & \dots & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & \dots & 0 & \dots & 1 & 0 \end{pmatrix}, \quad b = \begin{pmatrix} b_1 & \dots & b_5 & b_6 \\ 1 & \dots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \dots & 1 & 0 \end{pmatrix}$$

$$c_{ih} = \begin{pmatrix} c_{1,1} & c_{1,2} & \dots & 0 & 0 \\ 1 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & c_{r^h,1} & c_{r^h,2} \\ 0 & 0 & \dots & 1 & 0 \end{pmatrix}, \quad c_{is} = \begin{pmatrix} c_{r^h+1,1} & c_{r^h+1,2} & \dots & 0 & 0 \\ 1 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & c_{r,1} & c_{r,2} \\ 0 & 0 & \dots & 1 & 0 \end{pmatrix}$$

$$d_l = \begin{pmatrix} d_{l,1} & \dots & d_{l,5} & d_{l,6} \\ 1 & \dots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \dots & 1 & 0 \end{pmatrix}, d = \begin{pmatrix} d_2 & 0_{6,6} & \dots & \dots & 0_{6,6} \\ 0_{6,6} & \ddots & 0_{6,6} & \dots & 0_{6,6} \\ \vdots & 0_{6,6} & d_l & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0_{6,6} \\ 0_{6,6} & \dots & \dots & 0_{6,6} & d_{r^q+1} \end{pmatrix}$$

Q es una matriz diagonal cuya diagonal viene dada por:

$$(\sigma_f^2 \ 0_{1,11} \ \sigma_{u_1}^2 \ 0_{1,5} \ \sigma_{v_1}^2 \ 0 \ \dots \ \sigma_{v_r}^2 \ 0 \ \sigma_{u_2}^2 \ 0_{1,5} \dots \ \sigma_{u_{r^q+1}}^2 \ 0_{1,5})$$

Para el tratamiento de los datos faltantes ocasionados por la combinación de frecuencias entre las series o por la naturaleza no balanceada del conjunto de información, CP siguen a Mariano y Murasawa (2003). El procedimiento consiste en remplazar los datos faltantes por valores aleatorios provenientes de una distribución normal $\theta_t \sim N(0, \sigma_\theta^2)$, los cuales son independientes de los parámetros del modelo y permiten que las matrices sean conformables. Dado que el filtro de Kalman utiliza para ellos un proceso generador de datos distribuido normalmente, este tratamiento no tiene impacto sobre la estimación del modelo pues las observaciones faltantes solo añaden una constante en la función de verosimilitud que se maximiza.

Sea Y_{it} el i – ésimo elemento del vector Y_t y R_{ii} su varianza. Sea H_i la i – esima fila de la matriz H , la cual tiene α columnas y sea $0_{i\alpha}$ un vector fila con α ceros. Para el tratamiento de datos faltantes la ecuación de medida (9) es remplazada por la expresión:

$$Y_t^* = H^* h_t + w_t^* \tag{11}$$

Donde, si Y_{it} es observado, $Y_{it}^* = Y_{it}$, $H_i^* = H_i$, $w_{it}^* = 0$ y $R_{ii}^* = 0$, y de lo contrario, $Y_{it}^* = \theta_t$, $H_i^* = 0_{i\alpha}$, $w_{it}^* = \theta_t$ y $R_{ii}^* = \sigma_\theta^2$. De esta manera la representación espacio estado no tiene observaciones faltantes, y el filtro de Kalman se puede aplicar directamente sobre (11) y (10) (para más detalles véase CP, 2010, pg. 15).

3. Selección de series y caracterización en tiempo real

En Colombia, existe una buena cantidad de indicadores de actividad económica relevantes para el pronóstico del PIB. Su naturaleza es bastante heterogénea, no solo por sus distintas fechas de disponibilidad y rezagos de publicación, sino también por las diferencias en su cubrimiento (nacional, urbano, principales ciudades) y fuentes¹². Así las cosas, es relevante la elección de un conjunto representativo de series, puesto que esto condicionará el desempeño predictivo del modelo empleado.

¹² Si bien en su mayoría son producidos por la agencia nacional de estadística – DANE –, existen otros que provienen de centros de investigación independientes, gremios, asociaciones, etc.

La selección del número óptimo de series a ser incluidas en la estimación de modelos de factores ha sido un aspecto al que se han dedicado algunos trabajos. Anteriormente se suponía, de manera tácita, que entre mayor fuera el número de series empleadas la precisión en la estimación de los factores comunes se incrementaría. Sin embargo, Boivin y Ng (2006) demostraron que incrementar la cantidad de series empleadas más allá de un número determinado podría resultar en pérdidas de eficiencia, y, lo más importante, podría empeorar el desempeño predictivo del modelo. Intuitivamente, la calidad del estimador se afectaría al utilizar series con poca información y muchas señales falsas.

Por ello, varios trabajos sugieren procedimientos para seleccionar solo un pequeño grupo de variables para estimar el modelo. En muchos casos, se utiliza un sistema subjetivo de puntajes con base a características deseables de las series (Bravo y Franken, 2001), o algoritmos soportados en distintos criterios que evalúan el potencial predictivo de los modelos (para el caso colombiano véase Kamil *et al.* 2010). En este trabajo se parte de considerar un modelo de factores dinámicos con un pequeño grupo de series económicas, con las que se suele generar consenso entre los analistas sobre su alta capacidad para predecir el desempeño de la economía: la producción industrial, las ventas minoristas del comercio, la confianza de los consumidores y la demanda de energía.

Este conjunto de variables, de forzosa inclusión, es ampliado con las series resultantes de un proceso de búsqueda heurística que intenta encontrar el modelo cuyo primer factor dinámico explique el mayor porcentaje posible de la variabilidad del PIB. Este procedimiento calcula el porcentaje de varianza del PIB explicado por cada modelo al incorporar una serie en el conjunto inicial de indicadores. De esta manera, una nueva variable es incluida solo si el porcentaje de varianza explicado por el modelo se incrementa. Las series adicionales que se consideraron se presentan en el Cuadro A1 del anexo, que incluyen indicadores mensuales y trimestrales de actividad de varios sectores productivos, series provenientes de encuestas de opinión (a empresarios y hogares), variables de demanda externa, fiscales, monetarias, financieras y del mercado laboral. En el Cuadro 1 se muestran las series escogidas, con sus respectivos coeficientes estimados (factores de carga). La inicialización de los parámetros del modelo se realizó siguiendo el procedimiento propuesto por Doz *et al.* (2007) y que utiliza Liu *et al.* (2010)¹³. El modelo especificado, que utiliza solo un factor dinámico¹⁴, logra explicar en promedio 53% de la varianza observada del PIB¹⁵. La Gráfica A2 del anexo muestra la relación entre dicho factor común y el crecimiento del PIB.

¹³ En este procedimiento se utiliza componentes principales en el panel balanceado de las series para estimar un factor común, con el que posteriormente se aplican regresiones OLS para obtener las condiciones iniciales de los parámetros.

¹⁴ Siguiendo a CP, el modelo emplea un solo factor dinámico para evitar una fuerte pérdida de grados de libertad dada su alta dimensionalidad. Sin embargo, existen criterios de información (Bai y Ng, 2002) que permiten elegir un número óptimo de factores dinámicos.

¹⁵ El porcentaje de varianza explicada por el modelo es variable en el tiempo. En la gráfica A1 del anexo se muestra cómo cambia dicho porcentaje, que oscila entre 47% y 61%.

Cuadro 1. Series y factores de carga del modelo

Tipo de Indicador	Serie	Fuente	Acrónimo	Factor de carga
Hard	Índice de Producción Industrial	DANE	IPI	0.225
	Ventas del comercio al por menor	DANE	VENTAS	0.108
	Demanda de Energía Total	XM	ENER	0.083
	Horas extras trabajadas en la industria	DANE	HORAS	0.170
	M1 + Ahorros	Superfinanciera	M1YAHO	0.282
	Producción Cemento	DANE	CEMENTO	0.175
Soft	Índice de Confianza Industrial	Fedesarrollo	ICI	0.072
	Índice de situación económica de la empresa en el comercio	Fedesarrollo	ISEC	0.063
	Índice de confianza del consumidor	Fedesarrollo	ICC	0.054
Trimestral	Censo de Edificaciones	DANE	CENSO	0.106

Fuente: Cálculos propios de los autores.

Es importante resaltar que en la utilización en tiempo real de los indicadores seleccionados en el cuadro 1, se presentan los problemas discutidos en cuanto al tratamiento de este tipo de información. Como se mencionó, el PIB colombiano del trimestre T se anuncia dos meses y tres semanas después de haberse finalizado dicho trimestre. Así, al momento de la divulgación del PIB del trimestre T se tiene información de algunos indicadores de actividad mensuales para uno o dos meses del trimestre T+1, dependiendo de sus rezagos de publicación. En la Figura 1 se evidencia este hecho, para los indicadores de actividad del Cuadro 1. Adicionalmente, en la misma figura se observa cómo el conjunto de información relevante para estimar el crecimiento del trimestre T+1 se va enriqueciendo con los datos mensuales de los trimestres T+1 y T+2 conforme transcurren uno y dos meses después de divulgado el PIB del trimestre T. Esto es, muestra cómo en el transcurso de T+1 se obtiene información del mismo trimestre para calcular *nowcasts*.

Figura 1. Caracterización de la información en tiempo real

Trimestre	Mes	PIB	IPI	VENTAS	ENER	HORAS	M1YAHO	CEMENTO	ICI	ISEC	ICC	CENSO
Trimestre T-1	Mes 1	--										--
	Mes 2	--										--
	Mes 3											
Trimestre T	Mes 1	--										--
	Mes 2	--										--
	Mes 3											
Trimestre T+1	Mes 1											
	Mes 2											
	Mes 3											
Trimestre T+2	Mes 1											
	Mes 2											
	Mes 3											

Información disponible 1 semana antes de publicarse el PIB del trimestre T

 Información disponible 1 mes después de publicado el PIB del trimestre T

 Información disponible 2 meses después de publicado el PIB del trimestre T

 -- Información no disponible debido a la frecuencia trimestral de las variables

Fuente: Cálculos propios de los autores.

Lo anterior muestra la posibilidad de producir *nowcasts* cada vez que se publican nuevos datos de las series de actividad. El modelo especificado proporciona pronósticos de corto plazo y *nowcasts* para el crecimiento del PIB de cada trimestre, dependiendo del conjunto de series disponibles. Esto permite cuantificar el impacto que tiene la información contenida en las nuevas publicaciones de cada indicador de actividad sobre la estimación del crecimiento del PIB.

4. Desempeño predictivo del modelo

Se realizó un ejercicio de evaluación de pronóstico fuera de muestra utilizando información en “pseudo-tiempo real”, esto es, teniendo en cuenta la serie del PIB disponible en cada trimestre del periodo considerado, pero empleando los indicadores de actividad disponibles en cada momento del ejercicio con las revisiones incorporadas hasta 2011-Q3 (cabe decir que de los 10 indicadores de actividad considerados solo 3 se revisan, con magnitudes de menor valor relativo respecto a las del PIB). La evaluación de pronóstico se realiza para el período de tiempo 2005-Q1 a 2011-Q3 y se calculan los errores de pronóstico respecto a los crecimientos anuales del PIB de las publicaciones inmediatamente posteriores.

En los Cuadros 2 y 3 se presentan las principales medidas de evaluación de pronóstico a horizontes de uno y dos trimestres, utilizando el modelo de factores dinámicos planteado en este documento (en adelante MFD). Adicionalmente, su desempeño predictivo se compara con el de los siguientes modelos de referencia, bajo los mismos supuestos del ejercicio de evaluación de pronóstico fuera de muestra realizado:

- **Modelo ARX.** Su forma funcional viene dada por:

$$y_t = \alpha + \beta(L)y_{t-1} + \delta_i(L)x_{i,t} + \mu_t$$

donde y_t y $x_{i,t}$ representan los crecimientos trimestrales del PIB y de la i -ésima variable exógena. $\beta(L)$ y $\delta_i(L)$ son polinomios de rezagos que se seleccionan de acuerdo al algoritmo de estimación *stepwise*¹⁶ con orden máximo 3. Se utilizan como variables exógenas las mismas empleadas en el cuadro 1 y una significancia de 5%¹⁷.

¹⁶ El algoritmo de estimación *stepwise* parte del modelo de regresión más simple (con variable exógena la de mayor correlación con la endógena) e incluye nuevas variables explicativas (de acuerdo a los valores de sus estadísticos t, de los incrementos del R^2 y de las correlaciones parciales), cotejando en cada paso si alguna de las variables que ya están presentes en el modelo puede ser eliminada.

¹⁷ El modelo elegido de esa manera fue: $y_t = c + \beta_1 IPI_t + \beta_2 Censo_{t-1} + \beta_3 Horas_t + \beta_4 Horas_{t-3} + \beta_5 Energia_{t-2} + \varepsilon_t$

- **Modelo ARMAX con IPI.** Se considera un modelo ARMA(1,1) para el crecimiento trimestral del PIB, que incluye al IPI como variable exógena, puesto que es el indicador con la mayor correlación contemporánea con el PIB¹⁸.
- **Modelo AR(1).** Estimado sobre el crecimiento del PIB.
- **Promedio histórico del PIB (MEDIA).** Se considera como pronóstico del crecimiento anual del PIB el promedio de sus variaciones anuales para el período de estudio.

Cuadro 2. Evaluación de pronóstico un paso adelante para diferentes modelos

	MFD	ARX	ARMAX	AR(1)	MEDIA
EAM	0.925	1.001	1.144	1.187	2.735
EAPM	0.469	0.563	0.640	0.554	1.744
RECM	1.098	1.225	1.399	1.441	3.169
RECMF	0.947	1.234	1.335	1.086	3.978
U-Theil	0.392	0.447	0.510	0.522	1.150

Fuente: Cálculos propios de los autores.

Cuadro 3. Evaluación de pronóstico dos pasos adelante para diferentes modelos

	MFD	ARX	ARMAX	AR(1)	MEDIA
EAM	1.300	1.432	1.738	1.819	2.745
EAPM	0.532	0.821	1.042	0.854	1.832
RECM	1.612	1.822	2.162	2.226	3.276
RECMF	1.076	1.647	2.099	1.683	4.129
U-Theil	0.581	0.656	0.774	0.798	1.168

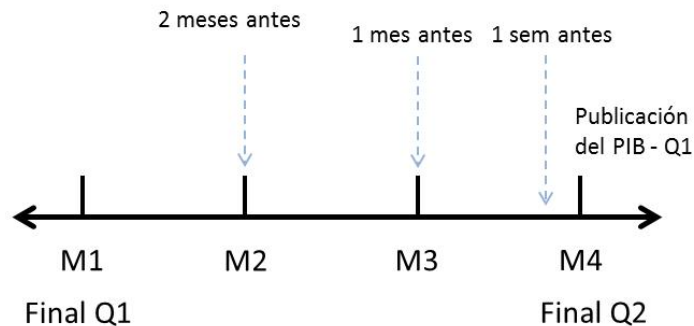
Fuente: Cálculos propios de los autores.

Los resultados de los Cuadros 2 y 3 sugieren que todas las medidas de error de pronóstico del MFD son menores respecto a los de los otros modelos considerados, tanto para horizontes de uno y dos trimestres adelante.

A continuación se procede a evaluar el desempeño predictivo del MFD para el *nowcast* del crecimiento anual del PIB. El ejercicio considera tres cortes en el tiempo para cada trimestre, tal como se establece en la Figura 2. Esta indica, que para obtener el *nowcast* del trimestre Q1 se tiene en cuenta la información disponible de las series explicativas dos meses antes, un mes antes y finalmente una semana antes de la publicación del PIB de dicho periodo, todo lo cual se realiza durante el trimestre Q2.

¹⁸ Esto es, se considera el modelo $y_t = c + \phi y_{t-1} + \beta IPI_t + \theta \varepsilon_{t-1} + \varepsilon_t$

Figura 2. Tiempos de *nowcast* para un trimestre arbitrario



Teniendo en cuenta estos puntos de referencia, en el Cuadro 4 se presentan los resultados de la evaluación de pronóstico de los *nowcasts* del MFD para el crecimiento del PIB. Estos sugieren que incorporar nueva información de las variables exógenas disminuye en promedio el error de pronóstico del modelo.

Cuadro 4. Evaluación del MFD para *nowcast* en diferentes momentos del trimestre

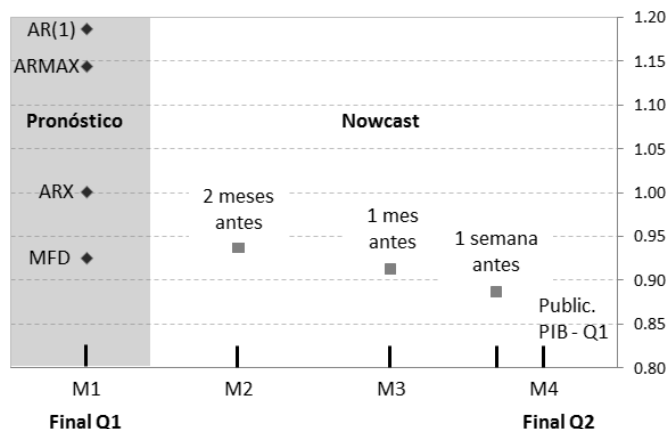
	1 semana	1 mes	2 meses
EAM	0.887	0.913	0.937
EAPM	0.435	0.446	0.445
RECM	1.032	1.048	1.063
RECMF	0.804	0.820	0.805
U-Theil	0.364	0.369	0.377

Fuente: Cálculos propios de los autores.

Adicionalmente, se compara el desempeño predictivo de los *nowcast* del MFD en comparación con el de los pronósticos realizados por el mismo modelo sin contar con información del trimestre en curso. Para ello, en la Gráfica 1 se contrastan los resultados del error absoluto medio (EAM) de los pronósticos un paso adelante para los diferentes modelos y los EAM producidos por el *nowcast* del MFD en los tres momentos de tiempo considerados (2 meses antes, 1 mes antes y 1 semana antes de la publicación del PIB)¹⁹.

¹⁹ Los resultados se mantienen utilizando otras medidas de error de pronóstico.

Gráfica 1. EAM para pronósticos un paso adelante y *nowcast*



Fuente: Cálculos propios de los autores.

La Gráfica 1 muestra en primer lugar que el *nowcast* produce menores errores de pronóstico frente a las estimaciones de los otros modelos utilizados. Y en segundo lugar, que el EAM de los *nowcasts* decrece conforme se acerca el momento de la publicación del PIB, de tal manera que, en promedio, el menor error de pronóstico se comete una semana antes de la divulgación de PIB. Esto confirma que la precisión de los pronósticos se incrementa al incorporar el flujo de información en tiempo real de los indicadores de actividad.

Finalmente, las Gráficas A3 y A4 del anexo proporcionan una caracterización visual del desempeño del modelo en tiempo real. La Gráfica A3 presenta el ajuste de los *nowcasts* del modelo, una semana antes de la entrega del PIB, a los crecimientos anuales divulgados en las primeras publicaciones, durante el periodo de tiempo en el que se realizó el ejercicio de evaluación de pronóstico. Según este ejercicio, a ese horizonte se encuentra el menor error de pronóstico en promedio. Por su parte, la Gráfica A4 muestra la evolución en tiempo real de los pronósticos y los *nowcasts* del modelo durante todo 2011, respecto a los crecimientos observados del PIB en los trimestres de ese año. En particular, se aprecia la influencia positiva del flujo de información de las variables explicativas en la actualización de las estimaciones. La Gráfica A4 sugiere que, a excepción del tercer trimestre de 2011, donde se evidenció un choque importante en un sector que no está representado en las variables explicativas del modelo²⁰, conforme se acercaron las fechas de publicación del PIB, los errores de las estimaciones tendieron a disminuir.

²⁰ Este choque se presentó en la construcción de obras civiles. Aunque el indicador de actividad asociado a este sector se consideró en el proceso de selección de series (Cuadro A1), no fue escogido bajo el criterio empleado (maximización del porcentaje de varianza del PIB explicada por el factor dinámico)

5. Conclusiones

En este trabajo se adaptaron para la economía colombiana los esquemas de pronóstico de corto plazo de Camacho y Perez-Quiros (2009,2010), que, a través de un modelo de factores dinámicos, incorpora la información disponible en tiempo real de distintos indicadores de actividad económica para obtener estimaciones del crecimiento del PIB del mismo periodo (*nowcasting*) y de periodos subsiguientes. La versión adaptada del modelo se constituye en una herramienta para tratar los diferentes aspectos de los conjuntos de información en tiempo real en Colombia: las diferencias en frecuencias y en rezagos de publicación entre series, la heterogeneidad de las mismas por cuenta de su construcción metodológica (por ejemplo entre balances de encuestas e indicadores de actividad real) y la presencia de factores idiosincráticos debido al carácter sectorial o parcial de cada serie. Así mismo, posibilita la cuantificación del impacto de cada nueva publicación de los indicadores de actividad sobre los pronósticos de crecimiento.

La evaluación del desempeño predictivo del modelo frente a otras alternativas de pronóstico usuales señaló que a horizontes de uno y dos trimestres el modelo de factores dinámicos produce el menor error de pronóstico respecto al de los otros esquemas considerados. Además, los resultados establecen que la incorporación del flujo de información en tiempo real de los indicadores de actividad mejora la estimación del crecimiento del PIB, de tal manera que los errores de pronóstico decrecen conforme se acerca el momento de su publicación, por la extensión de los conjuntos de información.

El modelo adaptado es susceptible a varias extensiones. En primer lugar, se podría considerar la selección de más de un factor dinámico siguiendo los criterios de información de Bai y Ng (2002). En este caso se debe tener en cuenta la importante afectación que tendrían los grados de libertad del modelo debido a su elevada dimensionalidad. Una segunda extensión consistiría en integrar al esquema utilizado el tratamiento a las revisiones del PIB colombiano que realiza Julio (2011) a través de un modelo de factores dinámicos. Esto controlaría una posible fuente de error de los pronósticos de crecimiento. Adicionalmente, se podría incluir un tratamiento a la estacionalidad de las series, que tenga en cuenta además el calendario colombiano. Todas estas alternativas pueden convertirse en materia de agenda de investigación en el futuro.

6. Bibliografía

Angelini, E., Camba-Mendez, G., Giannone, D., Reichlin, L. y Rünstler, G. (2008), “Short-term forecasts of Euro Area GDP growth”, *Working Paper* No. 949, European Central Bank.

Banbura, M. y Giannone, D. (2010), “Nowcasting” *ECARES Working Paper* No. 2010-021.

Barhoumi, K., Rünstler, G., Cristadoro, R., Den Reijer, A., Jakaitiene, A., Jelonek, P., Rua, A., Ruth, K., Benk, S. y Van Nieuwenhuyze, C. (2008), “Short-Term Forecasting of GDP Using Large Monthly Datasets: A Pseudo Real-Time Forecast Evaluation Exercise”, *Occasional Paper Series* No. 84, *European Central Bank*.

Bhattacharya, R., Pandey, R. y Veronese, G. (2011), “Tracking India Growth in Real Time”. *Working Paper* No. 90, Indian National Institute of Public Finance and Policy.

Bai, J. y Ng, S. (2002), “Determining the Number of Factors in Approximate Factor Models”, *Econometrica*, 70 (1), 191-221.

Boivin, J. y Ng, S. (2006), “Are more data always better for factor analysis?”, *Journal of Econometrics*, 132 (1), 169 - 194.

Bouwman, K. y Jacobs, J. (2005), “Forecasting with real-time macroeconomic data: the ragged-edge problem and revisions”. University of Groningen.

Bravo, F. y Franken, H. (2001), “Un indicador líder del IMACEC”, *Working Papers* No. 99, Banco Central de Chile.

Camacho M., Perez-Quiros G. (2009), “Ñ-Sting: España short term indicator of growth”. *Documentos de Trabajo* 0912, Banco de España.

Camacho M., Perez-Quiros G. (2010), “Introducing the Euro-Sting: Short term indicator of Euro Area Growth”, *Journal of Applied Econometrics*, 25 (4), 663-694.

D’Amato, L. Garegnani, L. y Blanco, E. (2010), “Using the flow of conjunctural information for short term forecasting of economic activity in Argentina”. Banco Central de Argentina.

Doz, C., Giannone, D. y Reichlin L. (2007), “A two-step estimator for large approximate dynamic factor models based on Kalman filtering”. *CEPR Discussion Paper* No. 6043.

Doz, C., Giannone, D. y Reichlin, L. (2006), “A quasi maximum likelihood approach for large approximate dynamic factor models”. *CEPR Discussion Paper* No. 5724.

Evans, M. (2005), “Where are we now? Real-time estimates of the macro economy”, *International Journal of Central Banking*, 2, 127-175.

Forni, M., Hallin, M., Lippi, M. y Reichlin, L. (2000), “The generalized dynamic factor model: Identification and estimation”. *The Review of Economics and Statistics*, 82 (4), 540-554.

Giannone, D., Reichlin, L. y Simonelli, S. (2009), “Nowcasting Euro Area Economic Activity in Real-Time: The Role of Confidence Indicators” *ECARES Working Paper* No. 2009-021.

Giannone, D., Reichlin, L. y Small, D. (2005), "Nowcasting GDP and Inflation: The Real-Time Informational Content of Macroeconomic Data Releases" *Working Paper* No. 2005-42, European Central Bank.

Giannone, D., Reichlin, L. y Small, D. (2008), "Nowcasting: The real-time informational content of macroeconomic data", *Journal of Monetary Economics*, 55, 665-676.

Jacobs, J. y S. van Norden, S. (2010), "Modeling data revisions: Measurement errors and dynamics of "true" values". Próximo *Journal of Econometrics*. Disponible en: <http://www.eco.rug.nl/jacobs/jjdownload/jacobsvannordenfinalerfeb2010.pdf>.

Julio, J. M. (2011), "Modeling Data Revisions". *Borrador de Economía* No. 641, Banco de la República de Colombia.

Julio J. M. y Grajales A. (2011), "¿Qué nos dicen los índices de confianza?". *Borrador de Economía* No. 659., Banco de la República de Colombia.

Kamil, H., Pulido, J. y Torres, J. (2010), "El "IMACO": un índice mensual líder de la actividad económica en Colombia". *Monetaria*, CEMLA, Vol. XXXIII, 4, 495-598.

Kapetanios, G. y Marcellino, M. (2006), "A parametric estimation method for dynamic factors models of large dimensions". *IGIER Working paper* No. 305.

Kuzin, V., Marcellino, M. y Schumacher, C. (2011), "MIDAS vs Mixed-Frequency VAR for Now-casting GDP in the Euro Area", *International Journal of Forecasting*, 27, 529-542.

Lahiri, K. y Monokroussos, G. (2011), "Nowcasting US GDP: The role of ISM Business Surveys", *Albany Discussion Papers* No. 11-01

Liu, P., Matheson, T. y Romeu, R. (2010), "Real-Time forecast of economic activity for Latin American countries". *Working Paper* No. 11-98, International Monetary Fund.

Mariano, R. y Murasawa, Y. (2003), "A new coincident index of business cycles based on monthly and quarterly series", *Journal of Applied Econometrics*, 18, 427-443.

Maier, P. (2011), "Mixed Frequency Forecasts for Chinese GDP" *Working Paper* No. 2011-11. Bank of Canada

Pedersen, M. (2010), "Extracting GDP signals from the monthly indicator of economic activity: evidence from Chilean real-time data", *Working Papers* No. 595. Banco Central de Chile.

Stock, J. y Watson, M. (2006), "Forecasting with many predictors" en Graham, E., Granger, C., y Timmermann, A. (editores) *Handbook of Economic Forecasting*, vol. 1, 516-550, North Holland.

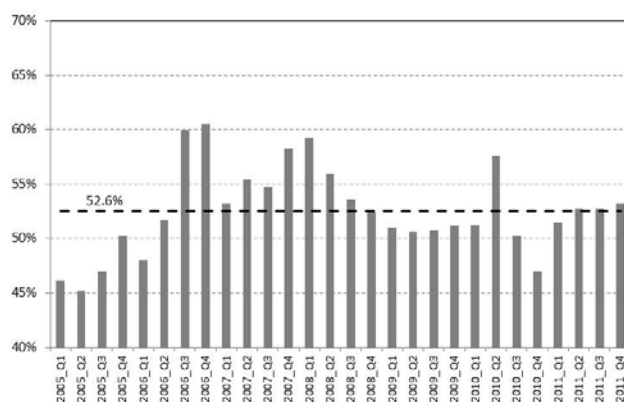
Wallis, KF. (1985), "Forecasting with an Econometric model: The ragged edge problem" *Journal of Forecasting*, 5, 1-13.

ANEXOS

Cuadro A1. Total de series consideradas para la estimación del modelo

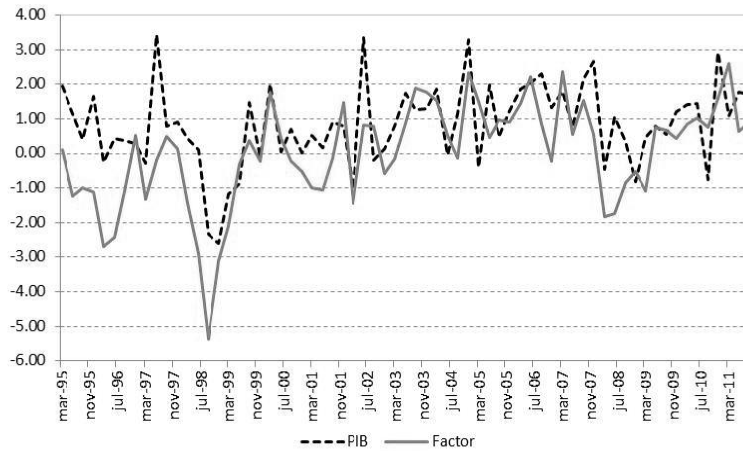
Tipo de indicador	Serie	Fuente
Hard	Demanda de energía total	XM
	Exportaciones totales	DANE
	Horas extras trabajadas en la industria	DANE
	Importaciones totales	DANE
	Índice general de la bolsa de Colombia (IGBC)	BVC
	Índice de producción industrial sin trilla de café (IPI)	DANE
	Licencias de construcción	DANE
	M1 + Ahorros	Superfinanciera
	M3	Superfinanciera
	Producción de café	Federación de cafeteros
	Producción de cemento	DANE
	Producción de petróleo	ANH
	Productividad media en la industria	DANE
	Recaudo de IVA	DIAN
	Sacrificio de ganado	DANE
	Terminos de intercambio	DANE - Cálculos propios
	Ventas del comercio al por menor	DANE
Soft	Demanda como principal problema en la industria	ANDI
	Índice de confianza del consumidor	Fedesarrollo
	Índice de confianza industrial	Fedesarrollo
	Índice de situación económica de la empresa en el comercio	Fedesarrollo
	Utilización de la capacidad instalada	ANDI
Trimestral	Censo de edificaciones	DANE
	Indicador de obras civiles	DANE
	Ocupados en las 7 áreas	DANE
	Tasa de desempleo en las 7 áreas	DANE

Gráfica A1. Porcentaje de varianza explicada por el MFD para el *nowcast* una semana antes de la publicación del PIB



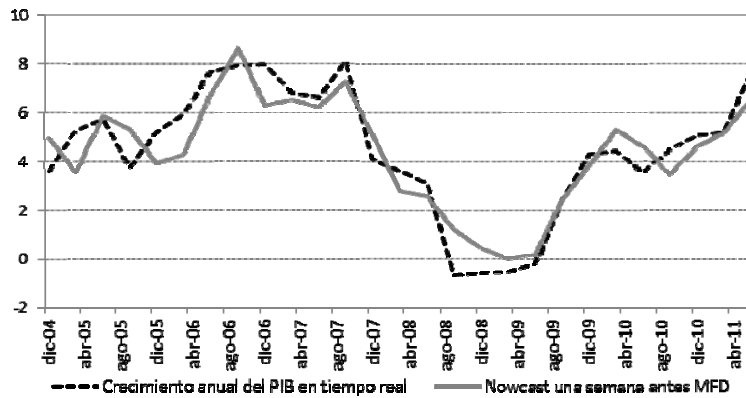
Fuente: Cálculos propios de los autores.

Gráfica A2. Componente común estimado usando el modelo de factores dinámicos y crecimiento trimestral del PIB.



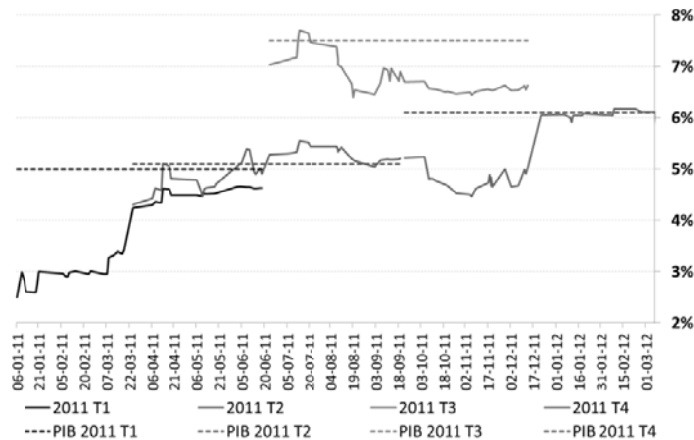
Fuente: Cálculos propios de los autores.

Gráfica A3. Nowcast una semana antes de la publicación del PIB y crecimiento observado del PIB en tiempo real.



Fuente: Cálculos propios de los autores.

Gráfica A4. Pronósticos en tiempo real para los cuatro trimestres de 2011



Fuente: Cálculos propios de los autores.