

# Notas Fiscales

**Indicadores ISAAC:  
Siguiendo la actividad sectorial a partir  
de *Google Trends***

Luis Fernando Mejía, Daniel Monsalve, Yesid Parra,  
Santiago Pulido, Ángela María Reyes

**No. 22, Noviembre 2013**



**Ministerio de Hacienda y Crédito Público**  
**Dirección General de Política Macroeconómica**  
Centro de Estudios Fiscales

Notas Fiscales, 2013  
ISSN 2027-8365

Noviembre de 2013

2013, Ministerio de Hacienda y Crédito Público- Dirección General de Política Macroeconómica  
Centro de Estudios Fiscales  
Cra 8 no. 6-64  
Bogotá, D.C., Colombia  
Tel: 3811700 ext. 3361  
[indipolmacro@minhacienda.gov.co](mailto:indipolmacro@minhacienda.gov.co)  
<http://www.minhacienda.gov.co/MinHacienda/haciendapublica/CentroEstudios/publicaciones/notas>

## **Indicadores ISAAC:**

### **Siguiendo la actividad sectorial a partir de *Google Trends***

Luis Fernando Mejía<sup>1</sup>

Daniel Monsalve<sup>2</sup>

Yesid Parra<sup>2</sup>

Santiago Pulido<sup>2</sup>

Ángela María Reyes<sup>2</sup>

## **RESUMEN EJECUTIVO**

Este documento presenta la primera aplicación al caso colombiano de la metodología de pronóstico de corto plazo de series económicas diseñada por Choi y Varian (2011) con base en el sistema de estadísticas de búsquedas por Internet llamado “*Google Trends*”. A partir de esta metodología se construyen los Indicadores Sectoriales Adelantados de Actividad (ISAAC), que permiten anticipar las tendencias del PIB de las nueve grandes ramas de actividad nacional en el corto plazo.

*Boletín No. 22*

*Códigos JEL: C1, C8, C53*

*Palabras Claves: Pronósticos PIB, Google Trends, Métodos de Pronóstico y Predicción*

## **ABSTRACT**

This document presents the first application for Colombia of Choi and Varian’s (2011) methodology for short term economic forecasting. This methodology is based on the Google search statistical tool known as “*Google Trends*” which provides weekly reports on the volume of queries related to different economic sectors in real time. Taking this data as an input, several anticipated sector indicators, named ISAAC for their corresponding acronym in Spanish, were constructed in order to analyze the short-run trends for the nine GDP sectors.

*Bulletin No. 22*

*JEL codes: C1, C8, C53*

*Keywords: GDP Forecasts, Google Trends, Forecasting and Prediction Methods*

---

<sup>1</sup> Director General de Política Macroeconómica del Ministerio de Hacienda y Crédito Público.

<sup>2</sup> Asesores de la Dirección General de Política Macroeconómica. Daniel Monsalve es actualmente funcionario del Departamento Nacional de Planeación.



## 1 Introducción

Para el diseño de la política económica es deseable contar con estadísticas confiables que permitan monitorear constantemente, y en tiempo real, la actividad sectorial. El sistema de cuentas nacionales sigue siendo la principal herramienta para esta tarea pues, en efecto, permite hacer un seguimiento periódico a las 49 ramas en que se descompone la actividad económica nacional. Sin embargo, la publicación de los indicadores de actividad con los que el DANE construye el PIB por el lado de la oferta tiene un rezago promedio de diez semanas. Dada esta importante limitación, surge la necesidad de contar con indicadores líderes que permitan hacer un seguimiento de las tendencias de corto plazo de la actividad económica. Este es precisamente el objetivo de este documento, que presenta la primera aplicación conocida para Colombia de una herramienta que permite un seguimiento de bajo costo y en tiempo real de la producción: el buscador *Google Trends*.

La literatura colombiana ha sido prolífica en la construcción de una gran diversidad de indicadores líderes, entre los que se cuentan los ejercicios de Maurer et. al (1996), Nieto y Melo (2001), Rozo (2008) y Kamil et. al (2010). Estos últimos diseñaron el índice IMACO, un ejercicio mensual que se anticipa seis meses a la publicación oficial del PIB de Colombia. Además de su rigurosidad estadística, el indicador cuenta con ventajas de relevancia, como su permanente actualización y publicación por parte del Banco de la República. En el plano de los indicadores líderes se destaca también el ALI (Indicador Líder de ANIF, por sus siglas en inglés), en donde se analiza el comportamiento de información del sector real y de variables monetarias y crediticias, directamente relacionadas con el producto interno.

El análisis sectorial, en cambio, no ha recibido la misma atención en la literatura. En efecto, existe escasa información secundaria sobre las ramas de actividad económica y la escasa información existente se utiliza más como una herramienta para pronosticar el producto agregado. La metodología que aquí se propone contribuye a la creación de indicadores sectoriales para reducir la incertidumbre que genera la ausencia de datos al interior de los sectores productivos.

Sin duda, una generación óptima de datos debería ser de carácter semanal, registrando por lo menos sólo unos cuantos días de rezago. Ello sucede con la mayoría de variables monetarias, cambiarias y crediticias, que en general proceden de sistemas de información que reportan en tiempo real el desempeño de los mercados. Sin embargo, la compilación de la información en sectores reales como la industria, el comercio, la construcción y la minería se puede establecer mucho después de generada la actividad. Una primera estrategia para este efecto es el uso de variables de desempeño empresarial. Éstas, aunque pueden provenir de firmas con aceptables sistemas de información, como es el caso de las compañías mineras y petroleras, pueden no ser generalizables ni comparables con el resto de sectores.

En este documento se propone una nueva estrategia consistente en la extracción de estadísticas a partir de la frecuencia de búsqueda en Internet de determinadas palabras, mediante la utilización de *Google Trends (GT)*, herramienta creada por *Google Labs* para la realización de investigaciones de mercado de este conocido buscador. Esta utilidad produce estadísticas desde el año 2004 y fue utilizada recientemente por Choi y Varian (2011) para el pronóstico de la situación presente de diversas series económicas (*nowcast*).

El uso de GT es justificable, por ejemplo, en el caso en que una caída súbita en la búsqueda de una determinada marca de maquinaria industrial pueda estar relacionada con un descenso de su demanda y un consecuente menor registro de la producción del subsector. Estas estadísticas no solo pueden ser comparables y generalizables para un conjunto amplio de sectores, sino que su disponibilidad es inmediata y con un nivel mínimo de rezago.

En este documento se expone la metodología de construcción de los indicadores ISAAC con base en los datos de GT, los cuales se utilizan posteriormente en el pronóstico de series de tiempo de PIB sectorial, para cada una de las nueve ramas de actividad. Con base en este comportamiento sectorial, se construye un indicador agregado del PIB total, el ISAAC+, se analiza su capacidad predictiva y se compara con otros indicadores líderes comúnmente utilizados para Colombia. Luego, con base en Diebold y Mariano (1995) se prueba para el caso colombiano la hipótesis de que las variables GT mejoran la capacidad predictiva de los modelos de series de tiempo. El ejercicio permite conocer en cuáles sectores las variables GT tienen la mayor relevancia estadística para anticipar de manera adecuada el PIB. Finalmente, el documento concluye con algunas consideraciones.

## 2 El uso de *Google Trends* en el análisis de coyuntura económica

De acuerdo con el sistema *NetMarketShare* de *Net Applications*, a diciembre de 2012 el 83,8% de las búsquedas por Internet en el mundo se realizaron a través de Google. En Colombia, esta proporción ascendió a 96,1% en este mismo período. Es decir, prácticamente todas las palabras que describen lo que hoy se busca por Internet en el país pasan a través de este famoso buscador. Como el algoritmo con el que opera la herramienta (*PageRank*) obliga a Google a guardar toda información de búsquedas e indexarla de cierta manera, lo que esta página ha ido consolidando es quizás la fuente de información comercial más grande que existe. Y es que si bien la herramienta cuenta una muy incompleta historia acerca de las transacciones comerciales que se generan a diario por Internet, el volumen de información permite monitorear en tiempo real las tendencias de diversos mercados.

De allí que en los últimos años hayan surgido en la literatura científica y económica una serie de preguntas sobre el potencial uso de esta aplicación. Tales interrogantes se tradujeron en documentos como los de Ettredge et al. (2005), Cooper et al. (2005), Askitas y Zimmermann (2010), D'Amuri y Marcucci (2010), entre otros reseñados en Choi y Varian (2011).

En América Latina, entre tanto, los ejercicios no han sido tan numerosos. Ello puede ser en parte atribuible al menor uso que se hace de Internet en estos países, frente al de Europa y EE.UU. (la penetración en América Latina es de 40%, mientras que en Europa y EE.UU. se llega a 63% y 79%). Sin embargo, existe un primer ejercicio para el caso chileno desarrollado en Carrière-Swallow y Labbé (2010), en el cual se utiliza GT para predecir cambios de tendencia en las ventas de automóviles.

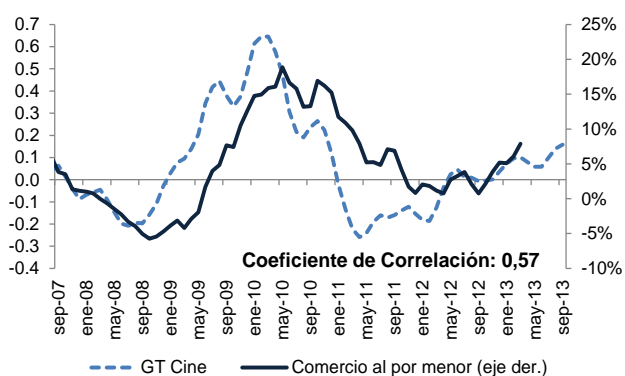
En Colombia aún no hay literatura sobre el tema, lo que da cabida a la formulación de varias preguntas. Por ejemplo, ¿es posible utilizar las búsquedas de Google para conocer el comportamiento presente del PIB colombiano u otras variables económicas?, ¿es posible mejorar la predictibilidad de un modelo de series de tiempo introduciendo variables que

resuman búsquedas de Google? Son este tipo de interrogantes los que dan origen a este documento.

## 2.1 Algunos ejemplos de *Google Trends* (GT) en la actividad económica colombiana

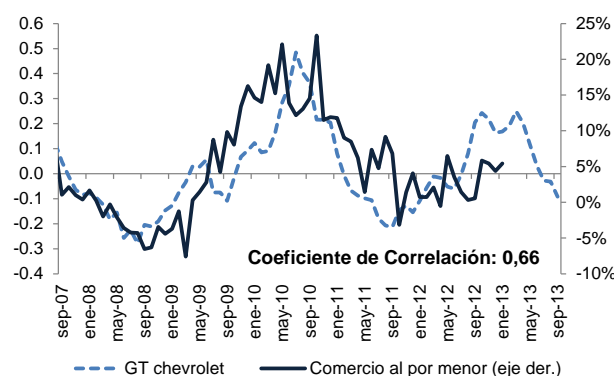
Sin duda, el análisis de GT ofrece algunos ejemplos más obvios que otros en sus relaciones con la actividad económica. Uno que resulta evidente es la relación entre la búsqueda de la palabra “Cine” con la variable de actividad económica “Comercio minorista”. En efecto, para este caso, el coeficiente de correlación se estima en 0,57, donde lo más importante es el hecho de que el movimiento de la variable GT está ligado a la tendencia del comercio minorista (Gráfico 1).

**Gráfico 1. Búsqueda de “Cine” vs. Comercio Minorista (promedio trimestral)**



\*GT Cine adelantado 6 meses  
Fuente: Google, Dane.

**Gráfico 2. Búsqueda de “Chevrolet” vs. Comercio Minorista (promedio trimestral)**

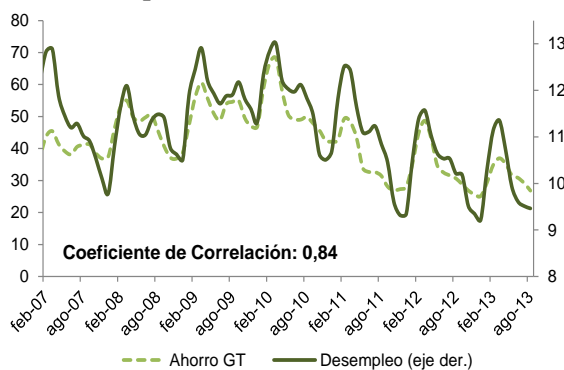


\*GT Chevrolet adelantado 6 meses  
Fuente: Google, Dane.

En consecuencia, se intuye que uno de los sectores que genera mayor dinamismo al comercio es el del cine y por lo tanto, las búsquedas de la palabra “cine” están asociadas a un eventual gasto en entretenimiento, lo cual impulsa el consumo interno. Así mismo, la búsqueda de la palabra “Chevrolet” tiene una alta correlación (0,66) con el comercio minorista. Datos del sector automotriz indican que gran parte de las ventas de automóviles nuevos en Colombia son de la marca Chevrolet. Por lo tanto, la búsqueda de esta palabra en Internet estaría vinculada con la posible compra de vehículo, lo cual estimularía el comercio automotriz y toda la cadena productiva que de este sector se deriva ( Gráfico 2).

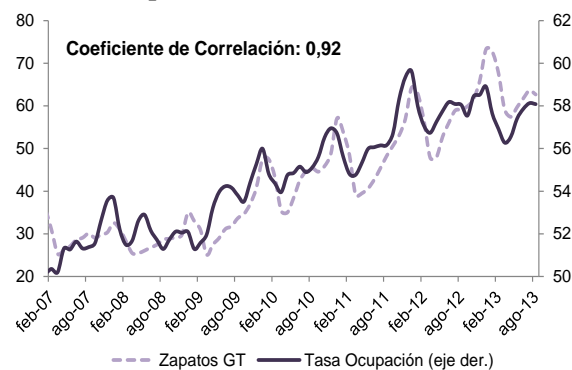
De otro lado, la correlación entre la búsqueda de la palabra “ahorro” con la tasa de desempleo arroja un resultado de 0,84 (Gráfico 3). Esto tiene su explicación en que cuando una persona se encuentra desempleada, la consecuencia inmediata es que disminuya su gasto. Por esta razón, un desempleado intentará ahorrar parte de sus ingresos para poder solventar por el mayor tiempo posible su período de paro. Por su parte, la correlación entre la tasa de ocupación y la palabra “zapatos” es de 0,92. Lo anterior podría ser consecuencia de que el sector de vestuario y calzado es una fuente importante de generación de mano de obra calificada. Entonces, una posible hipótesis es que la frecuencia de la búsqueda de la palabra “zapatos” por Internet es un reflejo del dinamismo de las ventas del sector, lo cual se traduce en una mayor tasa de ocupación. En conclusión, las búsquedas en Internet de las palabras “zapatos” y “ahorro” contribuyen a anticipar el comportamiento de la actividad económica (Gráfico 4).

**Gráfico 3. Búsqueda de “Ahorro” vs. Tasa de desempleo (%) (promedios trimestrales)**



Fuente: Google, Dane.

**Gráfico 4. Búsqueda de “Zapatos” vs. Tasa de Ocupación (promedios trimestrales)**



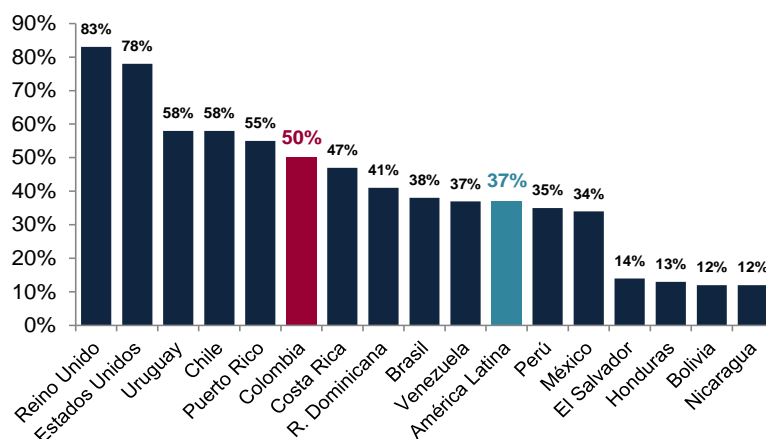
Fuente: Google, Dane.

## 2.2 Consideraciones para Colombia

La importante dinámica económica del país en los últimos años se ha reflejado en el desarrollo económico, en el incremento del empleo y la reducción de la pobreza. De esta forma, la población colombiana tiene acceso a más y mejores bienes y servicios para su consumo y bienestar. El acceso a las tecnologías de la información, como el Internet, permite a los habitantes del país una mayor interacción con lo que pasa, no solo en Colombia, sino a nivel mundial.

En la actualidad, Colombia tiene una penetración de Internet del 50% de la población, superior al promedio de América Latina, el cual se ubica en 37% (Gráfico 5). Es importante anotar que a comienzos de la década anterior, la penetración de Internet en el país era inferior al 20% de la población, lo que demuestra los grandes avances que se han logrado en este ámbito.

**Gráfico 5. Penetración de Internet en el mundo (% de la población)**



Fuente: Tendencias Digitales. Cálculos propios.

El grado de penetración a Internet por parte de los colombianos permite extraer información sobre lo que es más buscado en la red y su posible relación con el ambiente económico del país. Dicha información es importante como indicador de la percepción de los consumidores sobre temas como la crisis financiera mundial, la política económica del gobierno y la actividad económica en general.

### 3 Utilizando GT para construir Indicadores Sectoriales Adelantados de Actividad (ISAAC)

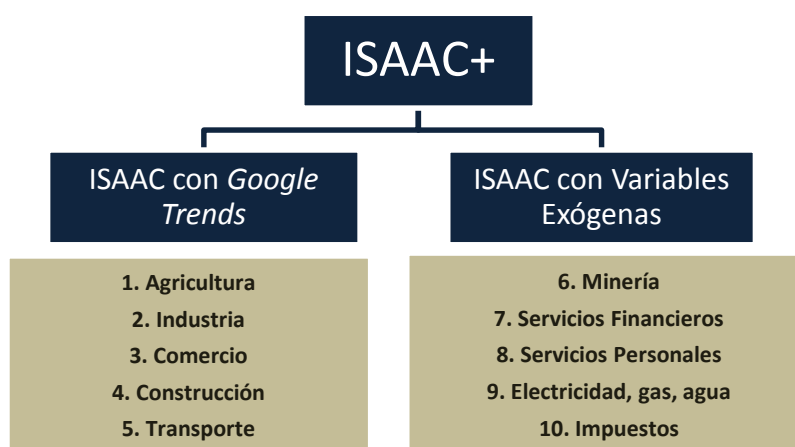
El objetivo de los Indicadores Sectoriales Adelantados de Actividad (ISAAC) en cada rama de actividad es monitorear el comportamiento que en cada momento del tiempo podría presentar el PIB sectorial. Para ello, se construyeron series de tiempo con base en la aplicación *Google Trends*.

La información que aporta *Google Trends* corresponde a un índice de frecuencia de búsqueda de determinadas palabras en Google durante un lapso de tiempo específico y en una determinada zona geográfica, como proporción del total de búsquedas realizadas con las mismas características de espacio y tiempo. De esta forma, el rango del indicador va de 0 a 100, donde 100 es el punto de mayor frecuencia de búsqueda. Esta herramienta comenzó a operar desde enero de 2004, y aunque para el caso colombiano existen algunos datos originados en esta fecha, las series resultan confiables desde 2005.

Es importante mencionar que el comportamiento de algunos sectores no se puede estimar utilizando búsquedas de Internet, debido a la naturaleza de las actividades. En efecto, sectores como la minería, los servicios sociales (de los cuales casi un 60% corresponde a servicios de la administración pública), la provisión de energía, gas y agua, el sector financiero, y los impuestos, no tienen necesariamente una relación lógica derivable de un sistema de búsquedas. Por fortuna, estos sectores cuentan con algunas variables exógenas que permiten tener información actualizada acerca de su desempeño con muy poco rezago.

De esta manera, aunque para la construcción de los indicadores ISAAC se sigue una misma metodología, el mecanismo de obtención de las variables no es igual para todos los sectores. En particular, la información de GT fue relevante en los casos de los sectores de agricultura, industria, comercio, construcción y transporte. Para el resto, fueron consideradas variables sectoriales exógenas con el mínimo rezago posible en su fecha de publicación. A partir de los ISAAC sectoriales se construye un indicador para el PIB total, el cual se denomina ISAAC+. En el siguiente diagrama se muestra la forma como se distribuyen los sectores de acuerdo a la información utilizada para construir el indicador ISAAC+.

**Gráfico 6. Composición del ISAAC+**



Fuente: Elaboración propia.

Posterior a la recolección, las series de tiempo se usaron como insumo para una medida del crecimiento del PIB mensualizado de cada sector. De esta manera los indicadores ISAAC para cada uno de los sectores en los que se utilizan los resultados de las búsquedas en *Google Trends* (del 1 al 5, según Gráfico 6) tienen la siguiente forma:

$$ISAAC_{it}|(i = 1, \dots, 5) = \sum_{j=1}^n \beta_j GT_{jt} \quad (1)$$

Cada sector económico está descrito por variables de caracterización  $GT_j$  ( $j = 1, \dots, n$ ), donde  $GT$  se refiere a la serie de búsquedas en *Google Trends* de palabras relacionadas con la actividad<sup>3</sup>.

Por su parte, el ISAAC para aquellos sectores que no utilizan variables de *Google Trends* sino variables líderes exógenas (del 6 al 10) se determina según la siguiente ecuación:

$$ISAAC_{it}|(i = 6, \dots, 10) = \sum_{j=1}^n \gamma_j Ex_{jt} \quad (2)$$

Donde  $Ex$  ( $j = 1, \dots, n$ ), se refiere a las variables líderes exógenas utilizadas.

**Tabla 1. Resultados sectoriales - Crecimiento del PIB vs. *Google Trends***

	Agro	Industria	Comercio	Construcción	Transporte	Minería	Servicios Financieros	Servicios Personales	Electricidad, gas y agua	Impuestos
<b>Var. Sector</b>						0.56*** (0.07)	0.15*** (0.01)	0.05*** (0.01)	0.43*** (0.08)	0.12*** (0.03)
<b>GT1</b>	-0.02** (0.01)	0.09** (0.04)	0.03*** (0.01)	-0.00 (0.02)	-0.01 (0.02)					
<b>GT2</b>	-0.09*** (0.01)	-0.02 (0.02)	0.03* (0.02)	0.10* (0.05)	0.02* (0.01)					
<b>GT3</b>	0.08** (0.03)	0.14** (0.06)	-0.06*** (0.01)	-0.25*** (0.06)	-0.19*** (0.03)					
<b>GT4</b>	0.01 (0.02)	0.34*** (0.08)	-0.04*** (0.01)							
<b>GT5</b>		-0.15*** (0.04)								
<b>GT6</b>		-0.20*** (0.06)								
<b>Constante</b>	0.01*** (0.00)	0.01 (0.01)	0.05*** (0.00)	0.10*** (0.01)	0.04*** (0.00)	0.03*** (0.01)	0.02*** (0.00)	0.04*** (0.00)	0.02*** (0.00)	0.04*** (0.00)
<b>Obs.</b>	54	64	83	78	80	61	147	35	148	148
<b>R<sup>2</sup></b>	0.47	0.54	0.57	0.29	0.42	0.50	0.78	0.24	0.12	0.12

Nota: errores estándar robustos entre paréntesis. \*, \*\*, \*\*\*, significativo al 1%, 5% y 10%.

Fuente: Google, DANE, elaboración propia.

<sup>3</sup> En general, las palabras buscadas se refieren a marcas de insumos o productos del sector, gremios y actividades relacionadas, las cuales pueden afectar positiva o negativamente el desempeño del sector. De allí que el signo esperado de estas variables en la explicación de la variable dependiente (el PIB) sea positivo o negativo, dependiendo de cada caso.

La Tabla 1 muestra las ecuaciones descriptivas de cada uno de los ISAAC. Tal como se esperaba, la mayor parte de variables buscadas en GT tuvo una relación estadísticamente significativa con la variable dependiente de cada una de las actividades sectoriales. Además, los modelos que utilizaron información de GT tuvieron en general una aceptable bondad de ajuste. Por su parte, como era de esperarse, los modelos basados en fuentes exógenas que se emplearon en la construcción del ISAAC también exhibieron un buen desempeño.

Cabe anotar que todos los modelos presentaron evidencia de variables omitidas, lo cual es previsible por tratarse de una sola fuente de información, lo cual puede tener efectos sobre el sesgo de los estimadores.

#### **4 Resultados ISAAC: pronósticos de crecimiento y tendencias del PIB sectorial**

En esta sección se presentan los resultados de la construcción de los indicadores de los nueve sectores estudiados, así como un ejercicio de pronóstico del PIB para el tercer trimestre de 2013.

Como se mencionó, los indicadores ISAAC tienen dos funciones: por un lado, señalar cuál es el crecimiento puntual esperado para cada sector y, por otro lado, identificar el momento del tiempo en el cual se presenta un cambio (quiebre) en la tendencia del crecimiento.

##### **4.1 Índices sectoriales**

Tal como se observa en el Gráfico 7, cada ISAAC se comporta con la misma dinámica del sector. Dado que la información de búsqueda tiene un rezago casi nulo, este resultado permite prever que es posible utilizar el ISAAC para pronosticar el comportamiento presente de cada sector de oferta (*nowcast*).

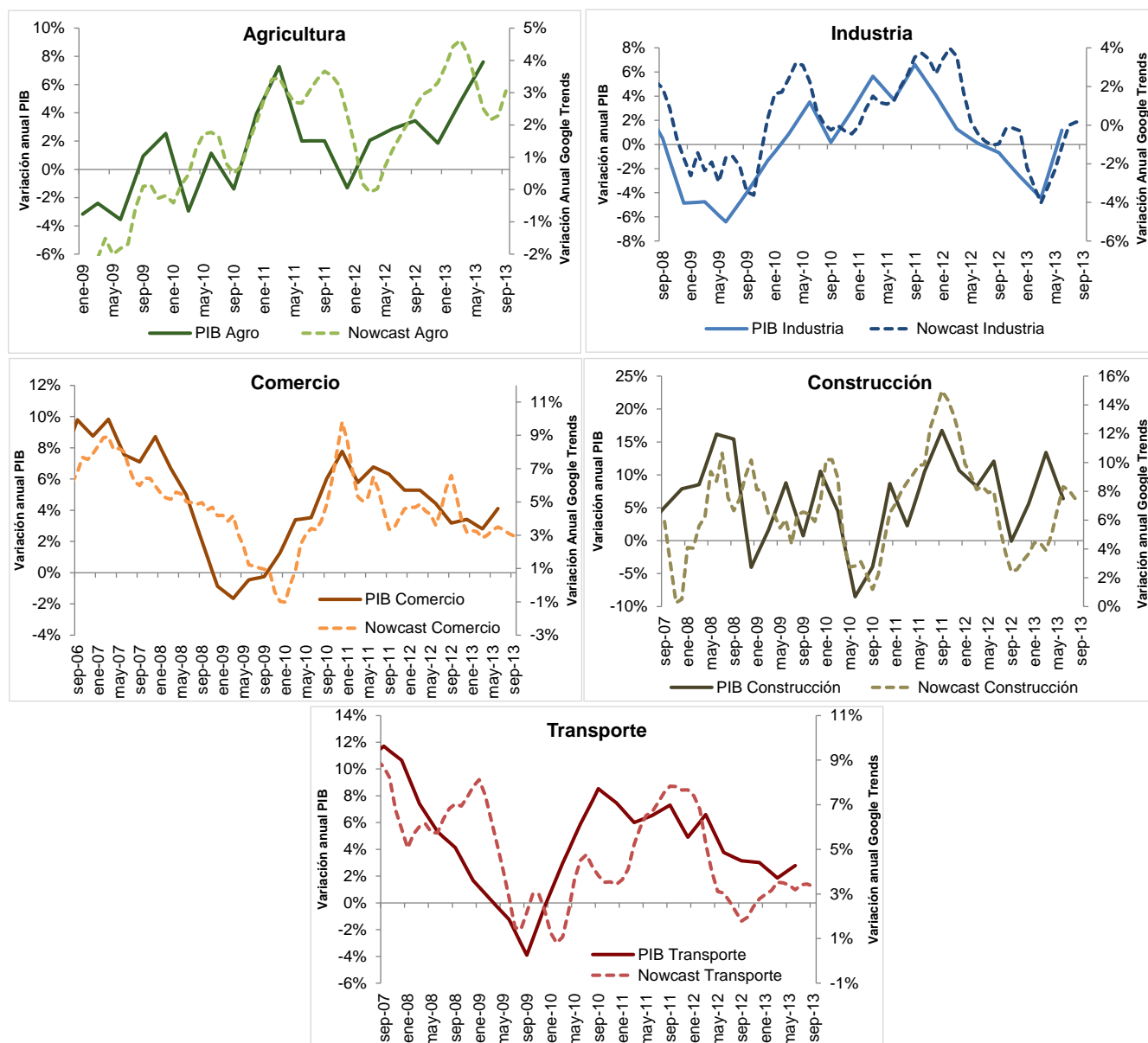
Es importante anotar que los resultados del ISAAC revelan la tendencia del PIB sectorial, y por lo tanto se pueden prever quiebres en la tendencia y comportamientos inesperados. Esto es evidente en el Panel A. del Gráfico 7, en donde se observa que los ISAAC muestran los cambios en la tendencia del crecimiento de cada uno de los sectores.

En el sector industrial, por ejemplo, el ISAAC predijo el quiebre en la tendencia observado en el segundo trimestre de 2013, cuando el sector salió de la recesión y registró un crecimiento positivo. Este comportamiento del PIB industrial durante el segundo trimestre del año responde a las políticas de impulso al sector implementadas durante la primera mitad de 2013. El indicador ISAAC no solo mostró la caída que registró la industria desde el inicio de 2012, sino que también anticipó los resultados del PIPE<sup>4</sup> sobre la producción industrial en el segundo trimestre de 2013. Teniendo en cuenta la información que provee el ISAAC para el 3T2013, se puede anticipar que el crecimiento de la industria continuará ubicándose en terreno positivo o cercano a cero durante ese periodo.

---

<sup>4</sup> PIPE son las siglas para Plan de Impulso a la Productividad y el Empleo, el cual es un programa del Gobierno durante el 2013 compuesto por distintas políticas transversales y sectoriales, encaminadas a la recuperación de la actividad industrial.

**Gráfico 7. Correlación entre las series ISAAC y el PIB sectorial**  
Panel a. ISAAC con *Google Trends*



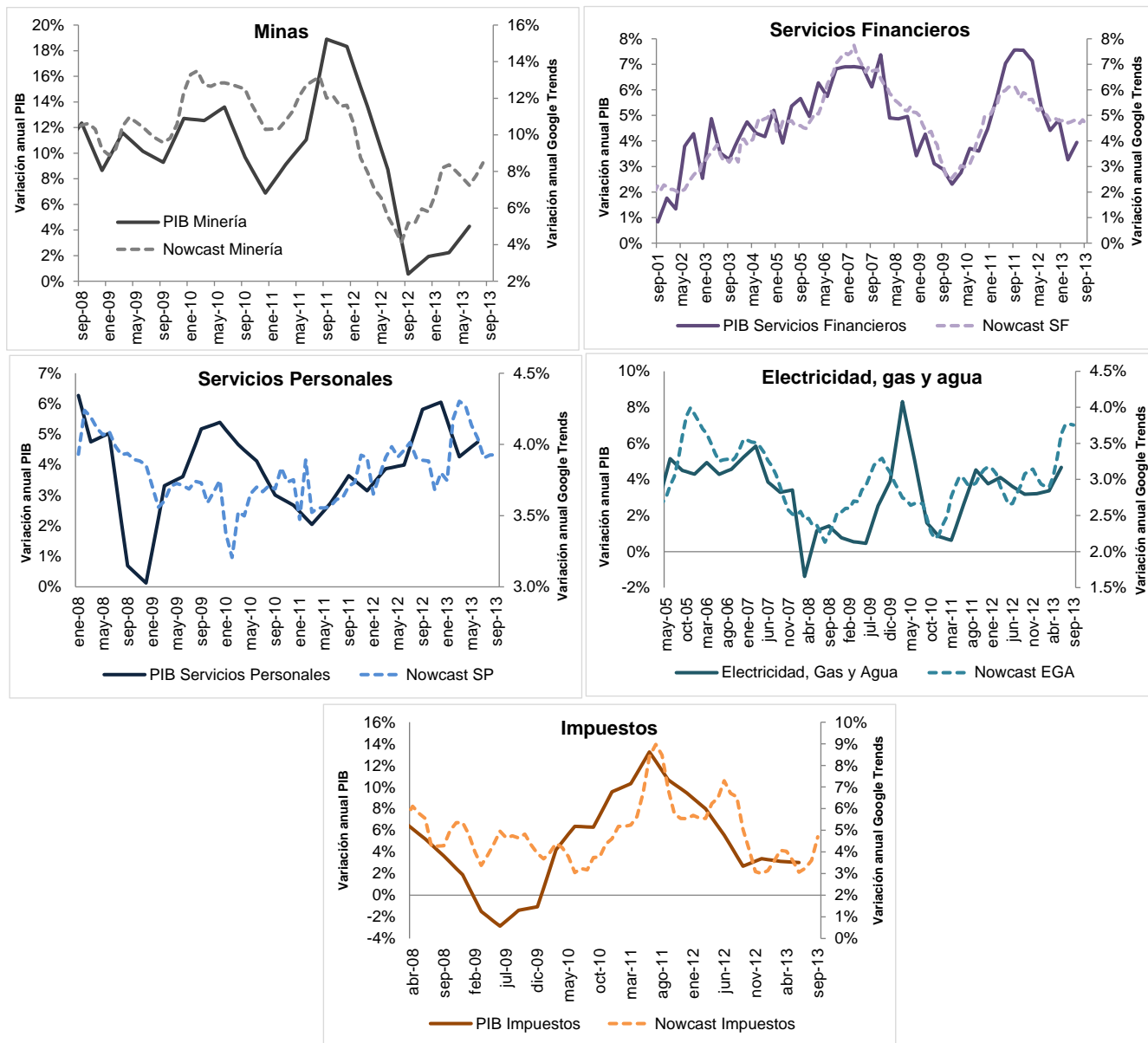
Fuente: Dane, Google, cálculos propios.

El ISAAC del sector de construcción también refleja de manera satisfactoria la tendencia y los quiebres del PIB de construcción. De acuerdo con los resultados del indicador, para el tercer trimestre de 2013 se puede esperar que el crecimiento del sector continúe siendo favorable, ubicándose por encima del 6%.

El ISAAC del comercio al por menor ha reflejado la tendencia a la baja del PIB de comercio. En efecto, esta variable tuvo una desaceleración durante el 2012, debido principalmente a factores externos, a un auge en la acumulación de bienes durables y semidurables, y a un cambio de expectativas de compra por los distintos tratados de libre comercio. Esta situación se revirtió en el 2T2013, cuando el crecimiento del sector mostró una aceleración al compararse con trimestres anteriores. De acuerdo con el ISAAC, el comercio debería continuar exhibiendo tasas de crecimiento positivas durante el tercer trimestre del año. Por su parte, el ISAAC de

agricultura muestra que durante el 3T2013 el sector va a continuar creciendo por encima del 5%.

**Gráfico 7 (cont.)**  
Panel b. ISAAC con Variables Exógenas



Fuente: Dane, Google, cálculos propios.

Es necesario resaltar que los ISAAC construidos a partir de variables exógenas también logran predecir la tendencia y los quiebres del PIB del sector relacionado. Por ejemplo, el ISAAC de minería mostró de manera anticipada la caída en el PIB del sector que se observó durante la primera mitad de 2012. Posteriormente predijo el cambio de tendencia que se presentó en el último trimestre de 2012.

## 4.2 Desempeño predictivo

### 4.2.1 Pronóstico sectorial

Uno de los mayores aportes de la metodología de *nowcast* es la posibilidad de generar predicciones de manera anticipada. Como se mencionó anteriormente, uno de los objetivos fundamentales en la construcción del ISAAC es el de generar pronósticos de corto plazo. Teniendo en cuenta que el dato del PIB trimestral se obtiene con 10 semanas de rezago, el ISAAC permite determinar el comportamiento trimestral de cada sector, tan pronto como el trimestre cierra, incluso sin ningún tipo de información líder del sector.

Bajo este enfoque, se realizaron ejercicios de estimación a lo largo de tres trimestres desde 4T2012 hasta 2T2013, evaluando el desempeño predictivo de los ISAAC para cada sector por el lado de la oferta. La Tabla 2 muestra un resumen de esta comparación, así como el pronóstico para el 3T2013, teniendo en cuenta que a la fecha en que se escribe este documento este trimestre ya había cerrado, pero aún quedaban más de 4 semanas para la publicación de esta cifra.

**Tabla 2. Resumen de pronósticos de PIB sectorial 2012 y 2013- Observado y pronóstico**

	4T2012		1T2013		2T2013		3T2013
	Pronóstico ISAAC	Observado	Pronóstico ISAAC	Observado	Pronóstico ISAAC	Observado	Pronóstico ISAAC
Agricultura	2,8	1,9	3,2	4,8	5,0	7,6	5,5
Industria	0,3	-2,7	-1,5	-4,4	-0,5	1,2	2,5
Comercio	3,8	3,4	3,8	2,8	3,5	4,1	3,4
Construcción	5,8	5,6	9,0	13,4	10,0	6,4	7,3
Transporte	3,8	3,0	3,8	1,9	1,1	2,8	3,1
Minería	0,3	1,9	2,0	2,2	5,7	4,3	6,3
Servicios financieros	4,2	4,8	4,2	3,3	3,5	3,9	4,7
Servicios	5,4	6,0	5,4	4,3	3,9	4,7	5,2
Electricidad, gas y agua	2,7	3,2	2,7	3,4	4,4	4,7	6,0
Impuestos	1,6	3,4	1,6	3,1	1,1	3,0	4,6
PIB	3,2	3,3	3,3	2,7	3,6	4,2	4,6

Fuente: DANE, Google y elaboración propia.

En términos generales, aunque la diferencia entre el pronóstico y el observado parece ser importante, vale la pena recordar que para el caso de los primeros 5 sectores, no se utiliza ningún tipo de variable líder, sino únicamente el *nowcast* de GT en tiempo real. Al tener esto presente, es importante resaltar que los pronósticos capturan la tendencia de corto plazo de la serie, así como los puntos de quiebres. Este caso es particularmente relevante en el caso de la industria, donde el ISAAC permitió prever la recuperación del 2T2013.

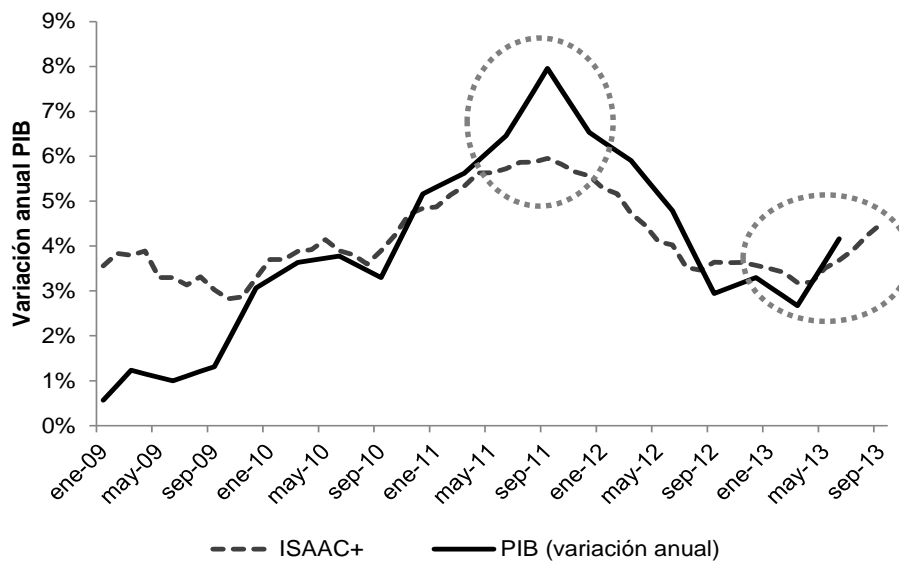
En este orden de ideas, para el 3T2013 el ISAAC+ estima un crecimiento de 4,6% explicado en una importante dinámica en el crecimiento agrícola y del sector financiero, así como en un buen desempeño de la construcción. Esto además, teniendo en consideración el deficiente desempeño

de la economía en 3T2012, que resultará en un mayor crecimiento este año dada la baja base de comparación.

#### 4.2.2 ISAAC+ y PIB Total

Utilizando los distintos ISAAC se realizó la construcción del ISAAC+ desde el 2009, teniendo en cuenta la disponibilidad de la información de todos los sectores desde esta fecha. El Gráfico 8 muestra el comportamiento de este indicador frente al comportamiento de la variación anual del PIB total observado mensualizado.

**Gráfico 8. ISAAC+ y PIB Total**

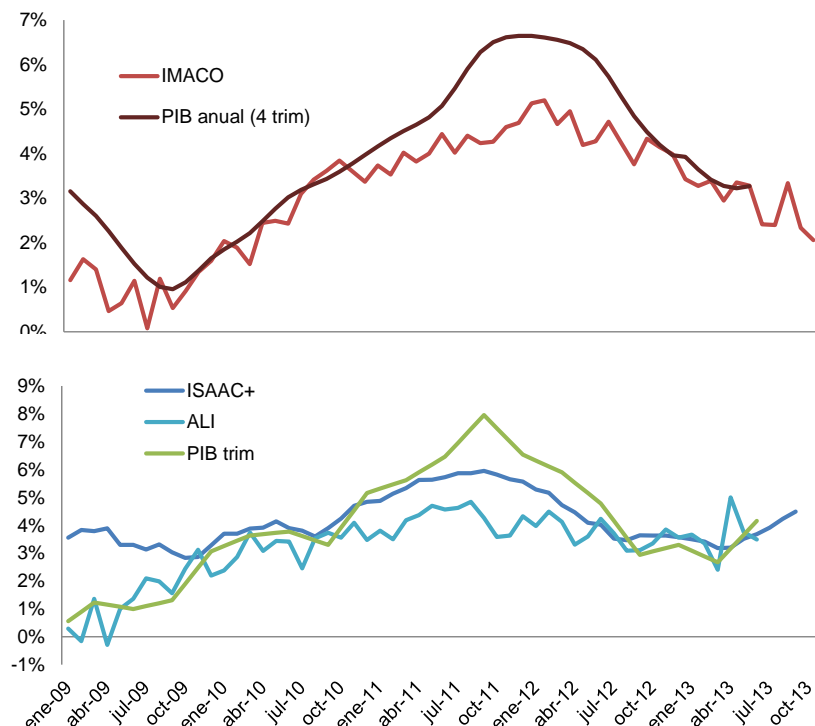


Fuente: DANE, Google y elaboración propia.

En primer lugar, es importante resaltar que el ISAAC+ refleja la tendencia de la serie del crecimiento del PIB, mostrando un coeficiente de correlación de 0,89 entre las dos series. No obstante, el aspecto más relevante es el comportamiento de los puntos de quiebre (señalados con un círculo punteado). Como se evidencia en el Gráfico 8, el ISAAC+ capturó el quiebre presentado el 3T2011 así como el más reciente quiebre del 2T2013. Esto es de suma importancia en términos del desempeño predictivo de este indicador, pues permite anticipar casi en tiempo real un cambio en el comportamiento de corto plazo de la economía. Como ya se mencionó, para el 3T2013 el ISAAC+ estima una recuperación de la actividad económica con un crecimiento de 4,6%. Esto último reafirma el quiebre en la tendencia negativa de la economía y la recuperación hacia la senda de crecimiento de largo plazo.

#### 4.2.3 Comparación con otros indicadores líderes

Adicionalmente, se realizó el ejercicio de comparar el comportamiento del ISAAC+ frente al de otros indicadores líderes de la actividad económica. Entre ellos, los más destacados son el IMACO, elaborado por el Banco de la República, y el ALI, elaborado por ANIF. En el Gráfico 9 se muestra el comportamiento de estos indicadores comparado con el crecimiento del PIB. En el caso del IMACO, este se muestra en una gráfica separada, pues para este indicador la variable de referencia es el crecimiento del PIB anual (suma de los 4 trimestres), mientras que para el ALI, así como para el ISAAC+, es el crecimiento anual del PIB trimestral.

**Gráfico 9. Comparación del comportamiento de diferentes Indicadores Líderes**

Fuente: DANE, ANIF, Banco de la República, Google y elaboración propia.

Como era de esperarse, se observa que cada indicador refleja el comportamiento de su variable de referencia. Tal es el caso del IMACO el cual predice la tendencia del PIB anual, mostrando un coeficiente de correlación de 0,91. Por su parte, el ALI y el ISAAC+ anticipan el comportamiento para el PIB trimestral, evidenciando una correlación de 0,84 y 0,89 respectivamente, para el periodo comprendido entre 2009 y lo corrido de 2013. Finalmente, en esta comparación el ISAAC+ muestra un buen desempeño, mostrando que es un indicador que está al mismo nivel de estos últimos.

**Tabla 3. Comparativo Coeficiente de correlación entre Indicadores Líderes y PIB**

	Coeficiente de Correlación	
	PIB trim	PIB anual (4 trim)
ISAAC+	0,89	0,72
IMACO	0,81	0,91
ALI	0,84	0,62

Fuente: DANE, ANIF, Banco de la República, Google y elaboración propia.

No obstante, es relevante mencionar que el ISAAC+ presenta una ventaja frente a los indicadores IMACO y ALI en el sentido en que permite tener la información del crecimiento desagregada por sector. Esto es de gran importancia pues permite establecer las causas del comportamiento pronosticado de la actividad económica, lo cual resulta de gran utilidad para la

formulación de políticas económicas encaminadas a impulsar el crecimiento del PIB colombiano.

## **5 Probando la eficiencia de los indicadores ISAAC**

En esta sección se presenta un análisis de las variables ISAAC estimadas de acuerdo a la metodología descrita. En primer lugar se examina la relación causal entre los ISAAC y el PIB sectorial y finalmente se mide su efecto sobre la precisión de modelos de pronóstico de series de tiempo con base en Diebold y Mariano (1995).

### **5.1 Causalidad con las variables dependientes**

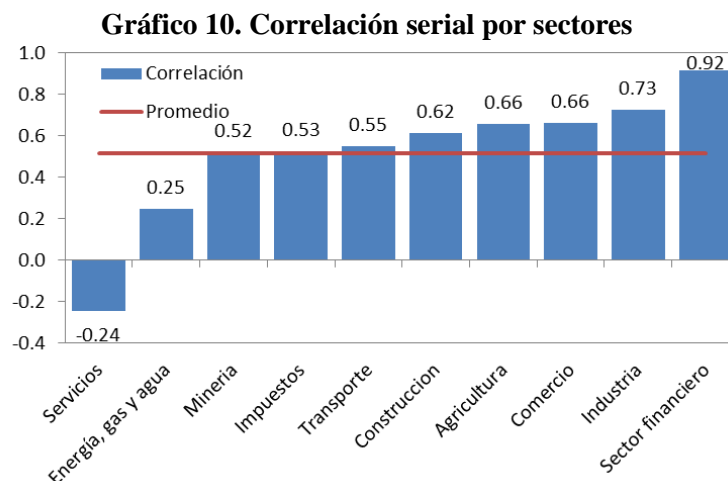
Para este efecto, se realizaron pruebas estadísticas con dos objetivos específicos:

1. Correlación con el PIB sectorial mensualizado: Se utilizó esta prueba para determinar la correlación lineal entre dos variables cuantitativas, en este caso cada una de las variables con la respectiva variable líder de cada sector.
2. Test de Granger: Con el fin de determinar si los valores de cada una de las variables sirven para predecir el comportamiento del PIB, se utilizó este test que arroja resultados de forma tal que permite establecer si existe una causalidad unidireccional, bidireccional o si no existe ningún tipo de causalidad.

El período de referencia de las pruebas econométricas se estableció de acuerdo a la disponibilidad de la información. Para la mayor parte de las variables se utilizaron series desde 2007 hasta los últimos datos disponibles de 2013. Posteriormente, con las pruebas mencionadas se procedió a determinar el primer análisis que consistió en establecer qué tipo de indicador era cada una de las variables: líder, coincidente o rezagada.

Es preciso anotar que un indicador líder es aquel cuyo valor predice anticipadamente el comportamiento de una variable. A su vez, un indicador coincidente es aquel cuyo comportamiento es el mismo o tiene la misma tendencia de una variable, es decir, se mueven en la misma dirección sin ningún tipo de rezago o adelanto. Finalmente, un indicador rezagado es aquel en el cual el comportamiento de la variable dependiente predice su comportamiento. Para los efectos de este trabajo, se descartaron todas aquellas variables que fueran indicadores rezagados por cuanto no permiten predecir el comportamiento de la actividad económica con anticipación, que es el objetivo principal de esta investigación.

A criterio propio y para facilitar la construcción de modelos de predicción de la actividad económica, en la selección inicial de palabras de GT se tuvo en cuenta únicamente aquellas que presentaban correlaciones superiores al 50%. Por ello, los ISAAC finales que utilizan variables provenientes del buscador, presentan unos notables índices de correlación con las variables dependientes (Gráfico 10), particularmente en industria y comercio. En general, la correlación promedio entre todos los sectores se ubicó en el 52%.



Fuente: Elaboración propia.

En cuanto a las pruebas de causalidad, se observa para tres de los cinco sectores que utilizaron datos de GT (industria, comercio y transporte) en la construcción de los ISAAC, la evidencia estadística de una causalidad serial (en el sentido Granger) que va del indicador GT al PIB, lo cual representa una primera prueba de su importancia en el *nowcast* del PIB sectorial (Tabla 4).

**Tabla 4. Pruebas de causalidad de Granger**  
ISAAC con GT (ISAAC\_GT) vs. PIB sectorial

	(1) causa ganger a (2)		P > chi2
	(1)	(2)	
Agricultura	PIB	ISAAC_GT	0.01
	ISAAC_GT	PIB	0.49
Industria	PIB	ISAAC_GT	0.15
	ISAAC_GT	PIB	0.00
Comercio	PIB	ISAAC_GT	0.49
	ISAAC_GT	PIB	0.01
Transporte	PIB	ISAAC_GT	0.02
	ISAAC_GT	PIB	0.00
Construcción	PIB	ISAAC_GT	0.04
	ISAAC_GT	PIB	0.40

Fuente: Elaboración propia.

## 5.2 Efecto de los indicadores ISAAC en la predicción del PIB

Un resultado importante de este ejercicio es la comprobación de la hipótesis planteada por Choi y Varian (2011) en el sentido en que GT contribuye de manera efectiva a la estimación de modelos de series de tiempo utilizados para pronosticar variables económicas.

En este apartado se comprueba esta hipótesis mediante la comparación de dos tipos de modelos. El primero es el mejor modelo autorregresivo (ARIMA) que se puede construir sobre las series de tiempo del crecimiento del PIB para cada uno de los sectores. El segundo es un modelo ARMAX especificado con las variables ISAAC como exógenas.

En consecuencia, por medio de este ejercicio se busca determinar cómo la introducción de los indicadores ISAAC puede mejorar la predictibilidad de los modelos de series de tiempo. Para ello se empleó el estadístico creado por Diebold y Mariano (1995), que evalúa una hipótesis de no diferencia significativa entre la predictibilidad de dos modelos de series de tiempo. Una gran ventaja del uso de este estadístico es la posibilidad de incorporar cualquier medida de predictibilidad en la comparación.

Para este ejercicio se emplean dos tipos de medidas: el Error Medio Absoluto (MAE) y el Error Medio Absoluto Porcentual (MAPE). Este último tiene en cuenta la dimensión de las variables, en caso de que ello sea relevante (v.g. tasas de crecimiento demasiado elevadas o volátiles). Las ecuaciones 3 y 4 presentan la forma de ambas medidas:

$$MAE = \frac{\sum_{t=T}^{T+h} |\text{observado}_{t+h} - \text{pronóstico}_{t+h}|}{h}, \quad (3)$$

$$MAE = \frac{100}{h} \sum_{t=T}^{T+h} \left| \frac{\text{observado}_{t+h} - \text{pronóstico}_{t+h}}{\text{observado}_{t+h}} \right|, \quad (4)$$

donde  $h$  es el número de periodos estimados y  $t$  el tamaño de cada submuestra.

Diebold y Mariano (1995) parten de los dos modelos,  $y_{t+h|t}^1$  e  $y_{t+h|t}^2$  que describen dos pronósticos diferentes a partir de  $y_{t+h}$ . De esta manera, los errores de pronóstico para estos dos modelos serían,  $\varepsilon_{t+h|t}^1 = y_{t+h} - y_{t+h|t}^1$  y  $\varepsilon_{t+h|t}^2 = y_{t+h} - y_{t+h|t}^2$ . Por tanto, la capacidad predictiva en ambos casos estaría medida por una función de pérdida de la forma  $L(y_{t+h}, y_{t+h|t}^i) = L(\varepsilon_{t+h|t}^i)$ ,  $i = 1, 2$  donde  $L(\cdot)$  tiene la forma descrita en las ecuaciones 3 y 4.

Para determinar si uno de los modelos predice mejor que el otro, se procede a probar la siguiente hipótesis nula:

$$H_0: E[L(\varepsilon_{t+h|t}^1)] = E[L(\varepsilon_{t+h|t}^2)], \quad (5)$$

frente a

$$H_1: E[L(\varepsilon_{t+h|t}^1)] \neq E[L(\varepsilon_{t+h|t}^2)]. \quad (6)$$

El estadístico Diebold-Mariano (DM) se basa en el siguiente diferencial de pérdidas:

$$d_t: L(\varepsilon_{t+h|t}^1) - L(\varepsilon_{t+h|t}^2). \quad (7)$$

con lo cual la hipótesis nula se transformaría en:

$$H_0: E[d_t] = 0. \quad (8)$$

Finalmente, la ecuación 9 presenta el estadístico (DM) que establece el diferencial de pérdidas como proporción de la raíz de la varianza asintótica estimada para dicho diferencial, lo que se puede expresar a su vez en términos de su covarianza:

$$S_1 = \frac{\bar{d}}{(\widehat{\text{var}}(\bar{d}))^{1/2}} = \frac{\bar{d}}{\left(\widehat{\text{LRV}}_{\bar{d}/T}\right)^{1/2}}, \quad (9)$$

donde:

$$\bar{d} = \frac{1}{T_0} \sum_{t=t_0}^T d_t, \quad (10)$$

$$\widehat{\text{LRV}}_{\bar{d}} = \gamma_0 + 2 \sum_{j=1}^{\infty} \gamma_j, \gamma_j = \text{cov}(d_t, d_{t-j}). \quad (11)$$

Este es un estimador consistente de la varianza asintótica de  $\sqrt{T}\bar{d}$ . La necesidad de la utilización del estadístico DM en la evaluación de predicciones radica en el hecho de que se utilice la varianza asintótica, pues la muestra de diferenciales de pérdida presenta una fuerte correlación serial para  $h > 1$ .

Diebold y Mariano (1995) prueban que el estadístico tiene una distribución normal estándar, con lo cual se rechaza la hipótesis nula cuando:

$$|S| > 1.96.$$

La Tabla 5 muestra los resultados de los ejercicios de estimación de los dos modelos de pronóstico, ARIMA vs. ARIMAX (incluyendo variables exógenas), cada uno con las especificaciones necesarias para el cumplimiento de la metodología Box-Jenkins.

**Tabla 5. Evaluación de los pronósticos sectoriales, estadístico DM y conclusiones**

Sectores	Criterio	ARMA	ARMAX (GT)	Diferencia	S(1)	Conclusión
Agricultura	MAE	0.006	0.013	0.007	7.803***	(1)
	MAPE	0.478	0.944	0.466	2.389**	
Minería	MAE	0.008	0.006	-0.002	-3.016***	(2)
	MAPE	0.127	0.082	-0.045	-1.589*	
Industria	MAE	0.010	0.005	-0.005	3.462***	(1)
	MAPE	0.362	1.269	0.907	2.919***	
Electricidad, gas y agua	MAE	0.003	0.003	0.000	-2.174**	(2)
	MAPE	0.389	0.262	-0.128	-1.729*	
Construcción	MAE	0.025	0.024	-0.001	-0.46	(3)
	MAPE	1.125	0.935	-0.190	0.81	
Comercio	MAE	0.004	0.003	-0.001	-2.606***	(2)
	MAPE	0.229	0.151	0.078	-1.954**	
Transporte	MAE	0.004	0.018	0.013	5.893***	(1)
	MAPE	0.183	0.913	0.731	1.643*	
Servicios financieros	MAE	0.002	0.002	0.000	-0.19	(3)
	MAPE	0.072	0.078	0.005	-0.94	
Servicios	MAE	0.003	0.003	0.000	1.15	(3)
	MAPE	0.440	0.484	0.044	1.35	
Impuestos	MAE	0.006	0.004	-0.002	-4.525***	(2)
	MAPE	0.261	0.146	-0.115	-2.115**	

Nota: Estadísticos significativos al \* 1%, \*\* 5% y \*\*\* 10%.

Fuente: Elaboración propia.

De acuerdo con los resultados, si bien los modelos presentan diferenciales de pérdida estadísticamente significativos, los resultados benefician el uso de *Google Trends* para el caso de ciertos sectores específicos, en los cuales mejora la predictibilidad de las series.

La última columna de la tabla resume las conclusiones de las estimaciones. Al grupo (1) pertenecen aquellos sectores en los que se encontró una evidencia estadísticamente significativa a favor de los modelos de series de tiempo (ARIMA) que no incluyen variables GT ni exógenas. En el grupo (2) están aquellos en los que GT contribuye positiva y significativamente al pronóstico de la serie. A esta categoría pertenecen sectores como la minería, el comercio, los servicios públicos domiciliarios y los impuestos. En el último grupo (3) están los sectores en donde no se encontró evidencia estadística sobre la diferencia de ambas especificaciones de modelos.

No obstante lo anterior, es preciso destacar dos elementos. En primer lugar, que la conclusión se basó, en esencia, en los resultados de ambas funciones de pérdida (MAE y MAPE). Sin embargo, en la industria, por ejemplo, sector en donde se pronosticó con series de GT, hubo fuerte evidencia estadística a favor de su uso tomando únicamente en cuenta el MAE. En segundo lugar, se destaca que en los nueve casos en los que la evaluación benefició el uso de GT (contando las dos diferentes medidas), la mejora en predictibilidad promedio fue justamente la obtenida por Choi y Varian (1995), es decir, de 5% a 20%.

## 6 Conclusiones

Este documento presenta la primera aplicación al caso colombiano de la metodología de pronóstico de corto plazo de series económicas diseñada por Choi y Varian (2011) con base en el sistema de estadísticas de búsquedas por Internet llamado "*Google Trends*". A partir de esta metodología se construyen los Indicadores Sectoriales Adelantados de Actividad (ISAAC), que permiten anticipar las tendencias del PIB de las nueve grandes ramas de actividad nacional en el corto plazo.

Al construir los ISAAC, se encuentra, en primer lugar, que las búsquedas efectuadas en GT efectivamente son significativas a la hora de explicar el PIB sectorial para ciertos casos, como lo son: agricultura, industria, comercio, construcción y transporte. Para los demás sectores, no es conveniente utilizar esta metodología pues no muestran una relación lógica derivable de un sistema de búsquedas por Internet.

En este orden de ideas, las búsquedas en GT permiten la construcción del ISAAC que determina en tiempo real, sin ningún rezago, la tendencia de corto plazo de estos sectores económicos, así como los puntos de quiebre en su tendencia. Esto se evidencia al comparar el comportamiento del ISAAC frente a la dinámica del PIB sectorial correspondiente. Tal vez, el mejor ejemplo de este es la predicción del quiebre en la tendencia del sector industrial evidenciado para el 2T2013, cuando el sector parece haber tocado fondo.

Mediante la agregación de los distintos ISAAC se construyó el indicador ISAAC+ buscando reflejar la tendencia de crecimiento del PIB total. En este documento se mostró que el ISAAC+ presenta una alta correlación con su variable de referencia, la variación anual del PIB trimestral total, capturando sus quiebres y comportamiento de corto plazo. Al comparar el ISAAC+ con otros indicadores como el IMACO y el ALI se evidencia un buen desempeño comparativo, mostrando que es un indicador que está al mismo nivel de estos últimos. Adicionalmente, el

ISAAC+ presenta ventajas frente al IMACO y el ALI, pues permite obtener información económica desagregada por sector, lo cual resulta de gran utilidad en la formulación de políticas económicas que busquen impulsar la actividad económica en Colombia.

Por otra parte, al someter los indicadores a las pruebas de causalidad, se observa para tres de los cinco sectores que utilizaron datos de GT en la construcción de los ISAAC (industria, comercio y transporte), que hay evidencia estadística de una causalidad serial (en el sentido Granger) que va del indicador GT al PIB, lo cual representa una primera prueba de su importancia en el *nowcast* del PIB sectorial.

Finalmente, es importante reconocer que los ISAAC, así como el ISAAC+, tienen algunas limitaciones en su desempeño predictivo. El hecho de que el nivel de penetración de Internet en Colombia sea del 50% de la población permite recoger información aproximada de solo una parte de los consumidores; aquellos que tienen acceso al servicio. Esto definitivamente restringe la capacidad predictiva de los indicadores. No obstante, en este momento el país adelanta una de las políticas de comunicaciones más agresivas, mediante el programa Vive Digital donde se busca la masificación del Internet. En la medida en que se avance en este sentido, el análisis de pronóstico mediante GT hará que los ISAAC sean cada vez más relevantes.

Más aún, gracias a la masificación de las TICs, en muy poco tiempo se dará mucha más importancia a la información en tiempo real. El rezago en la publicación de las cifras oficiales dificulta la toma de decisiones y retrasa la ejecución. Por este motivo, contar con una herramienta como el ISAAC permitirá contar con información en tiempo real del comportamiento de la economía, facilitando así la implementación rápida de políticas públicas. En ese caso perderá sentido “pronosticar” cifras de periodos pasados, para realmente pronosticar el futuro basado en *nowcasts* confiables.

## Bibliografía

- Abdi. H., y Williams, L.J. (2010), "Principal component analysis," *Wiley Interdisciplinary Reviews: Computational Statistics*, 2: 433–459.
- Askitas, N. y K. F. Zimmermann (2009) "Google Econometrics and Unemployment Forecasting," *Applied Economics Quarterly*, 55(2), 107-120.
- Choi, H. y H. Varian (2011), "Predicting the Present with Google Trends," *The Economic Society of Australia*, The Economic Record.
- Diebold, F.X., y R.S.Mariano (1995), "Comparing Predictive Accuracy," *Journal of Business and Economic Statistics*, 13, 253–263.
- Ettredge, J. Gerdes, and G. Karuga (2005). "Using web-based search data to predict macroeconomic statistics," *Communications of the ACM*, 48(11):87-92.
- D'Amuri F. y J. Marcucci (2010) "Google it! Forecasting the US unemployment rate with a Google job search index," SSRN Working Paper.
- Jolliffe I.T. (2002) "Principal Component Analysis, Series: Springer Series in Statistics", 2nd ed., Springer, NY, XXIX, 487 p. 28 illus. ISBN 978-0-387-95442-4
- Maurer, M., M. Uribe y J. Birchernal (1996), "El sistema de indicadores líderes para Colombia," *Archivos de Macroeconomía*, No. 49, Departamento Nacional de Planeación, Colombia
- Melo, L.F., F. Nieto, C. Posada, Y. Bentacourt y J. Baron (2001), "Un índice coincidente para la actividad económica colombiana," *Borradores de Economía* 195, Banco de la República de Colombia.
- Rozo, S. (2008), "Nuevo enfoque para la construcción de un único indicador líder de la actividad económica colombiana", Ministerio de Hacienda y Crédito Público.

**Boletines publicados**

- No. 1. Estrategia Económica y Fiscal 2010-2014.
- No. 2. Por qué es necesaria la creación de un Sistema General de Regalías.
- No. 3. La Sostenibilidad Fiscal, un principio para alcanzar los fines del Estado Social de Derecho.
- No. 4. Regla fiscal para Colombia.
- No. 5. Crecimiento de la productividad y costos derivados de la regulación: El efecto de aranceles y precios de la energía sobre los establecimientos manufactureros colombianos.
- No. 6. El Régimen de Seguridad Social en Salud: problemas financieros, medidas adoptadas y retos para el futuro próximo.
- No. 7. Régimen de transición pensional de los servidores públicos, Sistema General de Pensiones.
- No. 8. El impacto de la reforma tributaria en la demanda de efectivo.
- No. 9. Los retos de la economía informal en Colombia.
- No. 10. Mecanismo de cobertura condicionada de tasa de interés.
- No. 11. Ciclos económicos en Colombia: Bonanzas y recesión.
- No. 12. Esquema de fijación de precios de los combustibles en Colombia
- No. 13. Efecto de las expectativas sobre la estabilidad de la tasa de cambio, Colombia 2003-2009
- No. 14. Red para la inclusión de una economía formal
- No. 15. Documentación de modelos pertinentes para el diseño del Sistema General de Regalías
- No. 16. Tres Complementos Institucionales para el Sistema General de Regalías
- No. 17. El FLAR, una alternativa para un Fondo Regional de Unasur
- No. 18. ¿Se encuentran las reservas internacionales en un nivel óptimo?
- No. 19. Informalidad y Monotributo
- No. 20. Implicaciones de la adopción del Manual de Estadísticas de Finanzas Públicas de 2001 del Fondo Monetario Internacional sobre las estadísticas fiscales.
- No. 21. Informalidad empresarial y laboral en pequeños comerciantes: nueva evidencia para Colombia
- No. 22. Indicadores ISAAC: Siguiendo la actividad sectorial a partir de Google Trends



Ministerio de Hacienda y Crédito Público  
[www.minhacienda.gov.co](http://www.minhacienda.gov.co)

Este documento puede ser consultado en  
<http://www.minhacienda.gov.co/MinHacienda/haciendapublica/CentroEstudios/publicaciones/notas>