

Nowcasting con Google Trends: Dinámica de la Actividad Económica Mensual, el Consumo Privado y la Inversión basada en datos de Google Trends y un Modelo Bayesiano Estructural de Series de Tiempo

Lisette Josefina Santana Jiménez

Banco Central de la República Dominicana

Diciembre 2017

Resumen

Las técnicas de proyección basadas en *nowcasting*¹ han adoptado cada vez mayor preponderancia, gracias a su bondad para pronosticar determinadas variables macroeconómicas por medio de información de alta frecuencia y dada la posibilidad de publicar dicha información sin rezagos. El objetivo primordial de este estudio es generar proyecciones basadas en datos nano-económicos procedentes de la plataforma *Google Trends*, con frecuencia mensual y semanal, y empleando un Modelo Bayesiano Estructural de Series de Tiempo (BSTS, por sus siglas en inglés) para el Índice de Actividad Económica (IMAE), el Consumo Privado y la Inversión, los cuales son publicados por el Banco Central de la República Dominicana con periodicidad mensual y trimestral, respectivamente. Los resultados exhiben una convergencia entre los indicadores adelantados estimados a través de la metodología propuesta y las variables observadas, *i.g.* . IMAE, Consumo Privado e Inversión. Se llega a la conclusión de que la concatenación de la información de *Google Trends* con un modelo especializado como el BSTS, para el manejo de grandes conjuntos de datos propicia una mejora en la exactitud de los resultados en comparación con los valores efectivamente observados. La importancia y el valor de las herramientas inherentes a la teoría de *machine learning* continúan en ascenso y su desempeño refleja su alto potencial para reforzar los procesos de toma de decisiones.

Palabras Claves: Actividad Económica, Inversión, Consumo Privado, Google Trends, Modelo Bayesiano Estructural.

Clasificación JEL: E2, C11, C14, C19, E27, C69.

¹ En economía, el término *nowcasting* se refiere a la predicción del presente o del futuro muy cercano.

*Subdirección de Estudios Económicos, Departamento de Programación Monetaria y Estudios y Económicos. Para preguntas y comentarios escribir a lj.santana@bancentral.gov.do. El resultado de esta investigación no representa la visión del Banco Central de la República Dominicana. Cualquier error u omisión es estricta responsabilidad de los autores.

Nowcasting with Google Trends: Dynamics of the Monthly Economic Activity, Private Consumption and Investment based on Google Trends Data and a Bayesian Structural Time Series Model

Lisette Josefina Santana Jiménez

December 2017

Abstract

Forecasting techniques based in *nowcasting*² have become more relevant during the last years, given their advantages to predict the behavior of macroeconomic variables using high frequency information and also the possibility of making this information available without any lags. The main objective of this research is to generate predictions based on nano-economic data proceeding from the Google Trends platform, with monthly and weekly frequency, through a Bayesian Structural Time Series Model (BSTS) for the Monthly Economic Activity Index, Private Consumption and Investment, which are published by the Central Bank of the Dominican Republic with monthly and quarterly frequency, respectively. The results exhibit a convergence between the leading indicators estimated through the proposed methodology and the observed variables, *i.e.* the Monthly Economic Activity Index, Private Consumption and Investment. It is concluded that the concatenation of the information obtained from Google Trends with a specialized model as the BSTS, for managing big data sets, provides an improvement in the accuracy of the forecasted values with the effectively observed variables. The importance and value of the tools inherent to machine learning theory keep rising and reflecting their potential to reinforce decision making processes.

Key Words: Economic Activity, Private Consumption, Google Trends, Bayesian Structural Time Series Model.

JEL Classification: E2, C11, C14, C19, E27, C69.

² In economics, the term *nowcasting* refers to the prediction of the present or forecasting of a very close future.

I. Introducción

La medición eficiente y la generación de pronósticos referentes a la trayectoria futura de las principales variables macroeconómicas, a nivel de diferentes horizontes temporales, constituyen el núcleo de la toma de decisiones desde la óptica de los bancos centrales y de los distintos agentes económicos.

Las técnicas documentadas en la literatura empírica para proyectar la trayectoria futura de la actividad económica son diversas, así como también los métodos utilizados para evaluar la coyuntura prevaleciente y para construir indicadores adelantados que permitan extraer señales sobre el posible comportamiento de las principales variables macroeconómicas.

No obstante, uno de los inconvenientes que se plantean para el uso de estas herramientas está relacionado a la falta de datos de alta frecuencia y, en ocasiones, a los rezagos existentes en la información publicada por algunas entidades. La posibilidad de manejar un espectro de datos mayor, disponible en una frecuencia más alta y de manera oportuna, se presenta como una opción atractiva desde la perspectiva de la modelación econométrica.

La literatura económica moderna permite abordar nuevas técnicas y alternativas al problema de falta de información de alta periodicidad, empleando datos procedentes de grandes compañías de internet que han invertido recursos en el procesamiento y almacenamiento de dicha información. Este es el caso de Google Inc., con su plataforma *Google Trends*, que ha revolucionado la teoría económica en los años recientes, explotando las ventajas de esos conjuntos de datos que capturan las huellas de los usuarios de internet, poniéndolas a disposición del público de manera gratuita.

La plataforma *Google Trends* se ha convertido en una valiosa fuente de datos, proporcionando información acerca de la frecuencia de búsqueda o popularidad de determinados tópicos. El potencial de esta información para mejorar la precisión de proyección en los modelos económicos ha sido explorado en la literatura reciente (Varian y Choi, 2012; Bortoli y Combes, 2015), conjuntamente con el surgimiento de nuevas herramientas para llevar a cabo análisis basado en esta metadata.

Esta investigación está orientada a evaluar los resultados de las proyecciones que se obtienen usando estos datos “nano-económicos”, en el caso de una economía emergente como la de República Dominicana, para hacer inferencia sobre el comportamiento de la actividad económica, medida a través de un conjunto de indicadores: Índice Mensual de Actividad Económica (IMAE), Inversión y Consumo Privado.

Para este propósito, se emplean datos de búsquedas relacionados a los componentes de inversión, consumo y al IMAE y se hace *nowcasting* empleando técnicas inherentes a la teoría de *machine learning*, específicamente un Modelo Bayesiano Estructural de Series de Tiempo, que concatena un módulo estructural de series de tiempo para la serie objetivo con un componente de regresión (Scott y Varian, 2014), utilizando técnicas bayesianas cimentadas en *priors* de tipo “*Spike & Slab*”. Esta clase de *prior* reduce la dimensión del conjunto de predictores dentro de las posibles variables, evitando problemas relacionados a la “*maldición de la dimensionalidad*”. A fin de verificar el desempeño de los indicadores derivados de este modelo, en términos de su error de pronóstico, se hacen proyecciones fuera de muestra, las cuales permiten calcular el Error Cuadrático Medio (ECM).

El resto de este trabajo está estructurado de la manera siguiente: en la sección subsiguiente se lleva a cabo una revisión de literatura de trabajos empíricos basados en *nowcasting* y proyecciones de corto plazo con datos de *Google Trends*; en la Sección III se mencionan las fuentes de datos y se provee una explicación sobre el funcionamiento del Índice de *Google Trends*, basado en la popularidad de búsquedas *online*; en la Sección IV se presenta el Modelo Bayesiano Estructural de Series de tiempo (BSTS), usado para realizar *nowcasting*, así como también los aspectos metodológicos inherentes a este proceso. En la sección V se incluyen los indicadores derivados de este modelo para las variables consideradas y, finalmente, se presentan las conclusiones de este trabajo.

II. Revisión de Literatura

La trascendencia del fenómeno de *Big Data* (Bholat, 2016), desde la óptica de los bancos centrales, permite visualizar cómo los nuevos métodos para obtener y analizar datos no sólo suponen un incremento cuantitativo en la cantidad de información disponible, sino que también representan un novedoso enfoque para analizar el sistema económico y financiero.

Visto de manera aislada, el interés en el tema de *Big Data* puede ser considerado como un fenómeno coyuntural, en respuesta a las crisis financieras; sin embargo, visto de otra manera, refleja una estructura más sólida con el poder de incrementar el volumen, la velocidad y la variedad de datos a menores costos. Por otra parte, considerando este fenómeno desde el punto de vista de la demanda, existe un creciente interés por parte de los agentes económicos en entender cómo el análisis de datos puede beneficiar su productividad y sus ingresos.

Los trabajos pioneros en esta línea de investigación (Varian y Choi, 2008; Carriere-Swallow y Labbe, 2010) han llevado al surgimiento de una nueva tendencia de investigación que emplea la

data de *Google Trends* como fuente primaria para obtener información procedente de búsquedas de internet, demostrando cómo esta plataforma se ha posicionado como una herramienta valiosa con un alto potencial para mejorar el proceso de *nowcasting* y proyecciones de corto plazo de variables macroeconómicas.

Guzmán (2011) emplea datos de *Google Trends* para pronosticar expectativas de inflación, mientras que Artola y Galán (2012) los utilizan para generar predicciones del turismo en España, a nivel de provincias. Choi y Varian (2012) presentan uno de los primeros trabajos que hace el *nowcasting* basado en *Google Trends* para variables tales como compra de motores de vehículos, partes de autos y beneficios de desempleo en Estados Unidos. Se encuentra que un simple modelo AR, que incluye información exógena de *Google Trends*, tiende a minimizar el error de pronóstico de otros modelos que no contemplan la incorporación de dichos regresores de un 5% a un 10%.

Schmidt y Vosen (2009) analizan el comportamiento del consumo privado basándose en series de tiempo provistas por *Google Trends*; construyen un indicador fundado en factores extraídos de búsquedas relacionadas al consumo. El desempeño del nuevo indicador, en términos de su poder predictivo, es examinado con respecto a los indicadores basados en encuestas más comúnmente usados: el Indicador de Sentimiento del Consumidor de la Universidad de Michigan y el Índice de Confianza del Consumidor del *Conference Board*. Los resultados muestran que, prácticamente en todos los experimentos conducidos tanto fuera como dentro de la muestra, el indicador de *Google Trends* tiene mejor desempeño que los índices construidos en base a encuestas. Estas conclusiones sugieren que la incorporación de la información de *Google Trends* ofrece beneficios significativos para la proyección del consumo privado.

Otros trabajos se enfocan en el estudio de los precios de viviendas (Kulkarni *et al.*, 2009), expectativas de inflación (Guzmán, 2011), turismo (Artola y Galán, 2012; Bangwayo-Skeete y Skeete, 2015), ventas de automóviles (Carriere-Swallow y Labbe, 2010) y en la construcción de indicadores relacionados al sentimiento económico y a la confianza del consumidor (Heimer *et al.*, 2015; Levenberg *et al.*, 2014).

Para el caso de la República Dominicana, Santana (*no prelo*) pronostica la dinámica futura del sector turístico, medida en términos de la potencial llegada de pasajeros no residentes, haciendo *nowcasting* con datos nano-económicos procedentes de *Google Trends*. Los resultados derivados de estas proyecciones, además de lograr una significativa minimización del error de pronóstico con respecto a otros modelos alternativos, corroboran la utilidad de las variables de alta frecuencia asociadas a la llegada de turistas para pronosticar la dinámica de este sector.

El enfoque Bayesiano se presenta como una alternativa lógica de modelación, considerando su eficiencia para llevar a cabo la selección de las variables explicativas. Brodersen *et al.* (2015) proponen un modelo Bayesiano estado-espacio para obtener predicciones contrafactuales de la actividad de mercados. Sus resultados reflejan que el modelo seleccionado ayuda a lidiar con algunos de los problemas que se presentan en los modelos estadísticos tradicionales, los cuales carecen de ciertas características y que, por tanto, no resultan apropiados para predecir procesos estocásticos.

Ludwig *et al.* (2016) analizan el desempeño de un modelo BSTS para predecir precios en el sector eléctrico, encontrando que dicho modelo es adecuado para la selección de las variables explicativas.

Scott y Varian (2014) presentan un sistema para hacer *nowcasting* de series de tiempo de carácter económico, basándose en *Google Trends* y *Google Correlates*, combinando modelos estructurales de series de tiempo con regresiones “*Spike & Slab*” para promediar sobre un conjunto de predictores disponibles. Los autores llevan a cabo aplicaciones para el caso del desempleo y ventas minoristas, arribando a la conclusión de que los predictores de *Google Trends* no solo resultan útiles para el proceso de *nowcasting*, sino también para predecir los puntos de giro de las series de tiempo.

III. Datos

Los datos procedentes de *Google Trends* (véase apéndice II, gráficos 4 y 5, para ejemplos) fueron la fuente primaria de información usada en esta investigación. Esta plataforma provee un índice normalizado de búsquedas, de manera que el valor máximo sea 100 y el valor mínimo sea 0, basándose en la popularidad relativa de la búsqueda de determinado término en el tiempo. El período de análisis considerado para esta investigación es de 2004-2017 para el caso de la inversión y el consumo privado, y de agosto 2012 a agosto 2017 para el IMAE. Los datos de *Google Trends* son publicados con un rezago de dos días, siendo el domingo el primer día de cada semana.

La identificación de las palabras claves para las búsquedas se llevó a cabo de dos maneras distintas: en primer lugar, para el caso del consumo, se buscaron términos relacionados a las necesidades de adquisición de los agentes económicos; para el caso de la inversión, la cual está compuesta por la formación bruta de capital fijo y la importación de bienes de capital, las palabras buscadas están relacionadas al sector construcción (principalmente búsquedas relacionadas a materiales de construcción como cemento y varilla, compras en ferreterías, etc.) y al conjunto de

bienes de capital importados por el país que se presentan en el listado de la Dirección General de Aduanas (DGA). Para el caso del IMAE, se hace una conjunción de las variables empleadas para proyectar la inversión y el consumo (con frecuencia semanal) considerando que dichos renglones representan la mayor ponderación para la construcción de este índice. La segunda fuente empleada es una réplica de los algoritmos de *Assymetric Hashing* y *K-Nearest Neighborhood* (véase Vanderkam *et al.*, 2013, para detalles metodológicos), empleados por *Google Correlates*³ para identificar las palabras claves relacionadas a una variable particular, ya que la plataforma *Google Correlates* no está habilitada para el caso de República Dominicana.

Las series mensuales y trimestrales del IMAE, la inversión y el consumo privado fueron obtenidas de la página web del Banco Central de la República Dominicana.

IV. Metodología

4.1. Nowcasting y Proyecciones de Corto Plazo

El método utilizado está fundamentado en un modelo BSTS con dos componentes: un módulo de series de tiempo que captura la tendencia general y los patrones estacionales en los datos, y un componente de regresión que permite la incorporación de datos exógenos procedentes de búsquedas de *Google Trends*.

4.1.1. Modelo Estado-Espacio

Los modelos estado-espacio son representaciones de sistemas dinámicos que proveen suficiente flexibilidad para la modelación y el tratamiento de un extenso rango de problemas en el marco de series temporales, simplificando el análisis con una metodología unificada, como ha sido

³ La herramienta *Google Correlates* considera millones de candidatos para identificar las series que tengan una mayor correlación, medida en términos del coeficiente de Pearson, con la serie provista por el usuario.

ampliamente documentado en la literatura económica (Durbin y Koopman, 1999). El principal objetivo de este tipo de modelo es cuantificar la incertidumbre inherente a una variable no observada, dada alguna información perteneciente a dicha variable.

Es lógico pensar que la complejidad de la representación estado-espacio generalmente depende de la envergadura de la aplicación considerada, la naturaleza del modelo y las necesidades individuales. Un amplio rango de modelos puede ser expresado en forma estado-espacio, permitiendo la selección de los componentes para modelar la tendencia, la estacionalidad, el módulo de regresión y otros componentes de estado que pueden ser considerados como relevantes.

Un modelo estructural de series de tiempo tiene la particularidad de que puede ser expresado en formato estado-espacio. Como se ha señalado previamente, bajo este enfoque, se asume que la evolución del sistema a través del tiempo está determinada por una serie de vectores no observados $\alpha_1, \dots, \alpha_n$, los cuales están asociados a un conjunto de observaciones y_1, \dots, y_n . Una representación estado-espacio simple y generalizada que refleja la relación entre el vector de y_t y α_t está dada por un sistema de ecuaciones diferenciales de primer orden, que puede ser expresado de la manera siguiente:

$$(1) \quad \begin{aligned} y_t &= Z_t \alpha_t + \varepsilon_t, & \varepsilon_t &\sim N(0, H_t), \\ \alpha_{t+1} &= T_t \alpha_t + R_t \eta_t, & \eta_t &\sim N(0, Q_t) \quad t = 1, \dots, n, \end{aligned}$$

donde y_t es un vector $p \times 1$ de observaciones denominado vector de observación y α_t es un vector $m \times 1$ no observado denominado vector de estado. La primera ecuación del sistema (1) se denomina ecuación de observación, y la segunda ecuación se denomina ecuación de estado. Se asume que los estados siguen un proceso de transición Markoviano.

Dado que α_t no puede ser observado directamente, el análisis debe estar basado en las observaciones y_t . Las matrices Z_t, T_t, R_t, H_t y Q_t inicialmente se asumen como conocidas y los términos de error ε_t, η_t se suponen serialmente y mutuamente independientes en todos los momentos del tiempo. Las matrices Z_t y T_{t-1} pueden depender de y_1, \dots, y_{t-1} . El vector inicial de estado α_1 se espera que sea $N(\alpha_1, P_1)$ e independiente de $\varepsilon_1, \dots, \varepsilon_n$ y η_1, \dots, η_n , donde se asume que, inicialmente, α_1 y P_1 son conocidos. La primera ecuación en (1) tiene la estructura de un modelo de regresión lineal, donde el vector de coeficientes α_t varía en el tiempo. La segunda ecuación representa un Vector Autorregresivo (VAR) de primer orden.

Para el propósito de este trabajo, la representación estado-espacio considerada es la sugerida por Scott y Varian (2014):

$$(2) \quad \begin{aligned} y_t &= \mu_t + \tau_t + \beta^T x_t + \varepsilon_t \\ \mu_t &= \mu_{t-1} + \delta_{t-1} + u_t \\ \delta_t &= \delta_{t-1} + v_t \\ \tau_t &= - \sum_{s=1}^{S-1} \tau_{t-s} + \omega_t \end{aligned}$$

donde se indica que $\eta_t = (u_t, v_t, \omega_t)$ contiene componentes independientes de ruido Gaussiano. En este caso, Q_t es una matriz diagonal constante, con elementos diagonales $\sigma_\varepsilon^2, \sigma_v^2$ and σ_ω^2 ; H_t es una constante σ_ε^2 . La bondad de este modelo radica en el hecho de que, no solamente captura el componente de tendencia y los patrones estacionales, sino que también incluye un módulo de regresión que permite incorporar información exógena, en este caso, procedente de las búsquedas de *Google Trends*. El nivel de la tendencia en el tiempo t está representado por μ_t y la pendiente es δ_t .

El componente estacional, τ_t , puede ser representado como un conjunto S de variables *dummy* con coeficientes dinámicos que se ajustan a la restricción de tener valor esperado equivalente a cero durante el ciclo completo de S estaciones. Los parámetros en (2) son las varianzas $\sigma_\varepsilon^2, \sigma_u^2, \sigma_v^2, \sigma_\omega^2$ y el vector de coeficientes de regresión β , el cual está representado por los parámetros de búsquedas de *Google Trends*.

Además de la aparente complejidad que implica la implementación de la técnica de modelación planteada, las metodologías usadas para el tratamiento de este tipo de modelos son relativamente simples, siendo el filtro de Kalman, el suavizador de Kalman y el incremento bayesiano de datos las técnicas más recurrentes.

No obstante, es necesario destacar que tanto el filtro de Kalman como el suavizador de Kalman proveen soluciones analíticas y son apropiados para la estimación de modelos Gaussianos, en los cuales se asume normalidad y linealidad, mientras que los modelos no-Gaussianos requieren un tratamiento más avanzado, como el incremento bayesiano de datos u otras técnicas de suavizamiento.

El filtro de Kalman (Kalman, 1960) es un algoritmo recursivo que incorpora la información provista para procesar todas las medidas disponibles, con el propósito de estimar el valor actual de las variables de interés (Maybeck, 1979). Se dice que es recursivo puesto que no requiere que toda la información previa sea guardada, almacenada y reprocesada cada vez que se introduzca una nueva medida. Denotando $y_{1:t} = y_1, \dots, y_t$ como un vector que representa la variable dependiente, el filtro de Kalman computa recursivamente la distribución predictiva $p(\alpha_{t+1}|y_{1:t})$, combinando $p(\alpha_t|y_{1:t-1})$ con y_t y usando un conjunto standard de fórmulas equivalentes a una regresión lineal.

Por otra parte, el suavizador de Kalman se mueve en retrospectiva a través del tiempo, distribuyendo información sobre las últimas observaciones a pares anteriores (μ_t, P_t) , donde μ_t representa la media de la distribución y P_t es la varianza. El suavizador de Kalman regresa las T medias y varianzas de las variables no observables, en cada momento del tiempo dentro de la muestra y condicionada a la información de la muestra completa.

Es evidente que el filtro y el suavizamiento son las técnicas computacionales más usadas para lidiar con modelos estado-espacio. No obstante, cuando los supuestos de normalidad y linealidad no son plausibles, no es posible generar cálculos exactos basadas en el filtro de Kalman y el suavizador de Kalman; estos métodos no resultan pragmáticos en el tratamiento de problemas de alta dimensión (Pnevmatikakis *et al.*, 2012). En este caso, es necesario apelar a otras metodologías que contemplan supuestos no-Gaussianos. En este sentido, se han propuesto diversos enfoques (Fahrmeir y Wagenpfeil, 1997; Fahrmeir y Tutz, 1994, Jungbacker y Koopman, 2007), incluyendo un suavizador bayesiano que emplea cadenas de Markov usando simulaciones Monte Carlo.⁴

La computación basada en incremento bayesiano de datos (Neal y Kypraios, 2015; Taylor *et al.*, 2015) produce simulaciones a partir de $p(\boldsymbol{\alpha}_t | \mathbf{y})$, donde $\mathbf{y} = y_{1:n}$ y $\boldsymbol{\alpha} = \alpha_{1:n}$ denotan los conjuntos completos de datos observados y latentes. Esta alternativa se presenta dado que no es posible derivar cada $\boldsymbol{\alpha}_t$ a partir de $p(\boldsymbol{\alpha}_t | \mathbf{y})$, ya que se debe respetar la correlación serial entre α_t y α_{t+1} .

El enfoque bayesiano es inherente a una teoría donde los parámetros de estado son variables aleatorias y las observaciones son fijas; para un tratamiento bayesiano se necesitan métodos basados en simulaciones para la estimación de parámetros adicionales. Uno de estos métodos

⁴ Es un método numérico que permite resolver problemas físicos y matemáticos mediante la simulación de variables aleatorias.

basados en simulaciones es el algoritmo de Durbin y Koopman (2002), el cual es usado en este trabajo (Scott y Varian, 2014). La ventaja de este enfoque es que extiende el alcance del suavizador de Kalman, haciendo posible la respuesta a preguntas que conciernen las covarianzas de variables no observables a través del tiempo (que es una limitación del suavizador de Kalman). El suavizador de simulaciones de Durbin y Koopman se encuentra entre los más rápidos y convenientes de implementar en términos computacionales y es sintetizado brevemente por Jarocinski (2015) de la manera siguiente:

- **Paso 1:** Derivar α^+ y y^+ por medio de la recursión en (1), donde dicha recursión se inicializa $\alpha_1^+ \sim N(0, P_1)$.
- **Paso 2:** Construir la serie artificial $y^* = y - y^+$ y derivar $\widehat{\alpha}^* = E(\alpha|y^*)$ pasando y^* a través del filtro de Kalman y del suavizador de Kalman.
- **Paso 3:** Construir $\widehat{\alpha} = \widehat{\alpha}^* + \alpha^+$, donde $\widehat{\alpha}$ se deriva de la distribución de α condicional en y .

4.1.2. Componente de Regresión

La representación estado-espacio del sistema (2), considerada para llevar a cabo las estimaciones en esta investigación, contiene un componente de regresión que permite la inclusión de un conjunto de factores exógenos con potencial para contribuir en la minimización de los errores de pronóstico. Para indagar sobre la forma de estructurar las matrices del modelo para incorporar el componente de regresión en la representación estado-espacio, véase Scott y Varian (2014).

La cantidad de búsquedas potenciales procedentes de *Google Trends* aunada al hecho de que las series económicas generalmente no son extensas, hace que, en general, la cantidad de regresores sea superior al número de observaciones disponibles. En este sentido, el enfoque basado en *priors*

“*Spike y Slab*” se presenta como una opción lógica para la estimación de los coeficientes de regresión, ya que minimiza el problema de dispersión, disminuyendo el tamaño del problema de regresión. Este *prior* asume que los coeficientes de regresión son mutuamente independientes con una distribución bimodal consistente de una distribución uniforme (*Slab*) y una distribución que se degenera en la vecindad de cero (*Spike*). La idea fundamental es hacer cero los coeficientes no significativos dentro del conjunto de datos considerados, haciendo los valores de su media posterior lo suficientemente pequeños. Esta propiedad del proceso de selección basado en selección a través de *priors Spike & Slab*⁵ hace este enfoque atractivo y popular en el paradigma Bayesiano.

De manera sucinta, dado un subconjunto β_γ de elementos y sea $\gamma_k = 1$ si $\beta_k \neq 0$ y $\gamma_k = 0$ si $\beta_k = 0$. Matemáticamente, un *prior Spike & Slab* puede ser expresado de la manera siguiente:

$$(3) \quad p(\beta, \gamma, \sigma_\varepsilon^2) = p(\beta_\gamma | \gamma, \sigma_\varepsilon^2) p(\sigma_\varepsilon^2 | \gamma) p(\gamma)$$

La distribución marginal $p(\gamma)$ en (3) es el *Spike*, ya que asigna una probabilidad positiva en la ϵ -vecindad de cero.⁶ En la práctica, resulta conveniente simplificar (3) empleando un *prior* independiente de Bernoulli (Scott y Varian, 2014):

$$(4) \quad \gamma \sim \prod_{k=1}^K \pi_k^{\gamma_k} (1 - \pi_k)^{1-\gamma_k}$$

La ecuación (4) puede ser reducida asumiendo que todos los ϕ_k tienen el mismo valor ϕ . El valor de ϕ puede ser determinado apelando al criterio experto con respecto al tamaño esperado del

⁵ Para más detalles matemáticos sobre los *priors Spike & Slab*, consultar Mitchel y Beauchamp (1988) o Ishwaran *et al.* (2010).

⁶ Como en el caso de la función Delta de Dirac.

modelo. Por tanto, si se espera una cantidad p de predictores distintos de cero, entonces $\phi = p/K$, $\pi = p/K$, donde K es la dimensión de ϕ_t .

Por otra parte, la ecuación (5) se denomina *Slab*, ya que es posible seleccionar los *priors* que la hagan débilmente informativa (cercana a una superficie plana), condicional en γ :

$$(5) \quad \beta_\gamma \mid \sigma_\varepsilon^2, \gamma \sim N\left(b_\gamma, \sigma_\varepsilon^2 (\Omega_\gamma^{-1})^{-1}\right) \quad \frac{1}{\sigma_\varepsilon^2} \mid \gamma \sim Ga\left(\frac{v}{2}, \frac{ss}{2}\right),$$

donde $Ga(r, s)$ denota la distribución gamma con media r/s y varianza r/s^2 . El parámetro de mayor dimensión en la ecuación (5) es la matriz de información completa de los *priors* del modelo Ω^{-1} .

4.1.3. Estimación de la distribución posterior: Algoritmo Cadena de Markov Monte Carlo

Básicamente, la idea del algoritmo Cadena de Markov Monte Carlo (MCMC, por sus siglas en inglés) consiste en evaluar la media posterior de $x(\alpha)$ del vector de parámetros $\boldsymbol{\phi} = (\theta, \beta, \sigma_\varepsilon^2, \alpha)$ por medio de simulaciones, seleccionando muestras de una densidad conjunta aumentada $p(\boldsymbol{\phi}, \alpha \mid \mathbf{Y})$. En el procedimiento MCMC, el muestreo realizado a partir de la densidad conjunta se realiza como una cadena de Markov.⁷

Tras la inicialización para $\boldsymbol{\phi}$, dígase $\boldsymbol{\phi} = \boldsymbol{\phi}^{(0)}$, en el ciclo de las simulaciones subsiguientes, se repiten los pasos:

1. Simular el estado latente $\boldsymbol{\alpha}$ a partir de $p(\boldsymbol{\alpha} \mid y, \theta, \beta, \sigma_\varepsilon^2)$.
2. Simular $\theta \sim p(\theta \mid y, \boldsymbol{\alpha}, \beta, \sigma_\varepsilon^2)$.

⁷ En la teoría de la probabilidad, se conoce como cadena de Markov o modelo de Markov a un tipo especial de proceso estocástico discreto en el que la probabilidad de que ocurra un evento depende del evento inmediatamente anterior.

3. Simular β y σ_ε^2 a partir de una cadena de Markov con distribución estacionaria $p(\beta, \sigma_\varepsilon^2 | \mathbf{y}, \boldsymbol{\alpha}, \theta)$.

Una secuencia $\phi^{(1)}, \phi^{(2)}, \dots$ es generada a partir de una cadena de Markov con distribución estacionaria $p(\phi | \mathbf{y})$, la distribución posterior de ϕ dado \mathbf{y} .

V. Proyecciones

La proyección de la dinámica futura de una variable económica implica el uso de la información disponible para hacer inferencia sobre el posible curso de los eventos; es decir “condicionar lo que no se conoce, sobre lo que se conoce para generar conclusiones relevantes” (Geweke y Whiteman, 2006; Blake y Mumtaz, 2012).

Bajo el paradigma bayesiano, es pertinente generar simulaciones a partir de la distribución posterior predictiva (Scott y Varian, 2014), dados los *draws* de los parámetros del modelo y el estado de la distribución predictiva.

Las iteraciones pueden realizarse directamente del sistema de las ecuaciones dado por la especificación (2), partiendo de la especificación $\alpha_n^{(g)}$ con parámetros $\theta^{(g)}, \beta^{(g)}, \sigma_\varepsilon^{2(g)}$, con el propósito de generar un muestreo de $p(\tilde{\mathbf{y}} | \mathbf{y})$ a través de $p(\tilde{\mathbf{y}}^{(g)} | \phi^{(g)})$ y, dado un conjunto de *draws* de la simulación Monte Carlo, los *draws* de la distribución predictiva ayudan a corregir automáticamente los coeficientes irrelevantes. Este método de proyecciones genera una muestra de *draws* de la distribución predictiva que pueden resumirse por sus respectivas medias, que es una estimación Monte Carlo de $E(\tilde{\mathbf{y}} | \mathbf{y})$.

VI. Resultados

Las proyecciones realizadas en este trabajo, tanto dentro como fuera de muestra, se llevaron a cabo empleando la metodología descrita en la sección V, a través de un Modelo Bayesiano Estructural de Series de Tiempo (BSTS), que considera *priors* “*Spike & Slab*”, a fin de depurar las variables que no resultan significativas y minimizar el problema de dispersión. Los intervalos de tiempo considerados para la estimación, en el caso del consumo privado y la inversión, que son variables publicadas con periodicidad trimestral, por el Banco Central de la República Dominicana, son enero 2004 - septiembre 2017, contemplándose un conjunto de regresores con frecuencia mensual para llevar a cabo los pronósticos de dichas variables dependientes, teniéndose un conjunto de 140 y 78 regresores, respectivamente (véase cuadros 2 y 3, apéndice I, para el listado de variables).

Por otra parte, las estimaciones generadas para el IMAE, que es un indicador publicado frecuencia mensual, se circunscribieron a un intervalo de tiempo más limitado, a fin de aprovechar la disponibilidad semanal de las variables explicativas publicadas por la plataforma *Google Trends* a partir del año 2012, considerándose 230 variables (véase cuadro 1, apéndice I, para listado de variables).

Las proyecciones correspondientes a cada variable macroeconómica son evaluadas en términos de su error de pronóstico, medido a través del error cuadrático medio (ECM), la bondad de ajuste y el desempeño relativo con relación a otros modelos tomados como referencia. Los modelos tomados como puntos de comparación fueron un Vector Autorregresivo (VAR) de Frecuencia Mixta y una mera representación estado espacio.

La intención de seleccionar estos dos modelos como puntos de comparación subyace en el hecho de que es posible verificar dos aspectos con respecto al Modelo Bayesiano Estructural: En primer lugar, por medio del VAR de frecuencia mixta se puede constatar el aporte adicional de la información de *Google Trends* al desempeño de un modelo de *nowcasting* pero que, en este caso, no emplea técnicas de *shrinkage* que permitan depurar regresores. Por otra parte, la representación general Estado-Espacio solamente captura en este caso la tendencia lineal local de la variable dependiente y el componente estacional, sin contemplar la contribución adicional de regresores externos.

La última fase de esta investigación consistió en la construcción de un indicador adelantado, a través de una estimación con regresión rodante para los años 2015 - Septiembre 2017, considerándose un horizonte temporal equivalente a uno ($h = 1$), con el propósito de examinar la convergencia entre el indicador generado a través del BSTS con respecto al dato observado, para cada período de tiempo.

Tras el proceso de selección de las variables, la primera fase del proceso de estimación consistió en mezclar la frecuencia de los datos, dado que la información procedente de *Google Trends* se presenta con periodicidad mensual para las variables de IMAE, y con frecuencia mensual para el consumo privado y la inversión.

En una segunda etapa, se lleva a cabo la especificación y estimación del componente de series de tiempo del modelo, dada por una tendencia lineal y un componente estacional, así como también la inclusión del componente de regresión, donde los datos de *Google Trends* entran a formar parte del modelo. Los paneles (a), (b) y (c) de los gráficos 6, 7 y 8 (apéndice II) muestran la contribución

de cada componente particular para los indicadores de IMAE, consumo privado e inversión, respectivamente.

Como se señala en la sección metodológica, el *prior* utilizado fue el *Spike y Slab prior*, como sugieren Scott y Varian (2014); se puede inferir información útil concerniente al *prior* a partir del tamaño esperado del modelo, el R^2 esperado, el tamaño de la muestra v y usando shrinkage diagonal, que corresponde al *prior g* de Zellner⁸.

El tamaño esperado del modelo fue establecido en cinco predictores para el IMAE, Consumo Privado e Inversión, lo que lleva a un umbral de inclusión de 0.0173, 0.028, 0.051, respectivamente (número de predictores esperados/cantidad total de variables). Los otros *priors* fueron establecidos de acuerdo a los valores que por *default* entrega el paquete BSTS (Scott, 2016), *i.e.* $R^2 = 0.5, v = 0.01$. Las condiciones consideradas en términos de selección de *priors*, número de iteraciones y cantidad de iteraciones descartadas, fueron las mismas para todos los modelos.

El algoritmo MCMC (sección IV) se usa para hacer el muestreo a partir de la distribución posterior del modelo BSTS, con 1,500 iteraciones, descartando las primeras 500 en la fase de *burn-in*; se utilizó el método *Stochastic Search Variable Selection (SSVS)* (George y McCulloch, 1997) para llevar a cabo promedio bayesiano de modelos.

Tras la estimación del modelo, se obtuvieron los predictores más probables para cada variable macroeconómica considerada (*i.e.* IMAE, Consumo Privado, Inversión) con sus respectivos coeficientes. En el caso del IMAE, las variables que exhiben una probabilidad de inclusión que supera el umbral son: “sector construcción” (8.56%), “precios gasolina” (4.10%), “compra de

⁸ Para detalles técnicos sobre el *shrinkage* diagonal, véase Scott (2016).

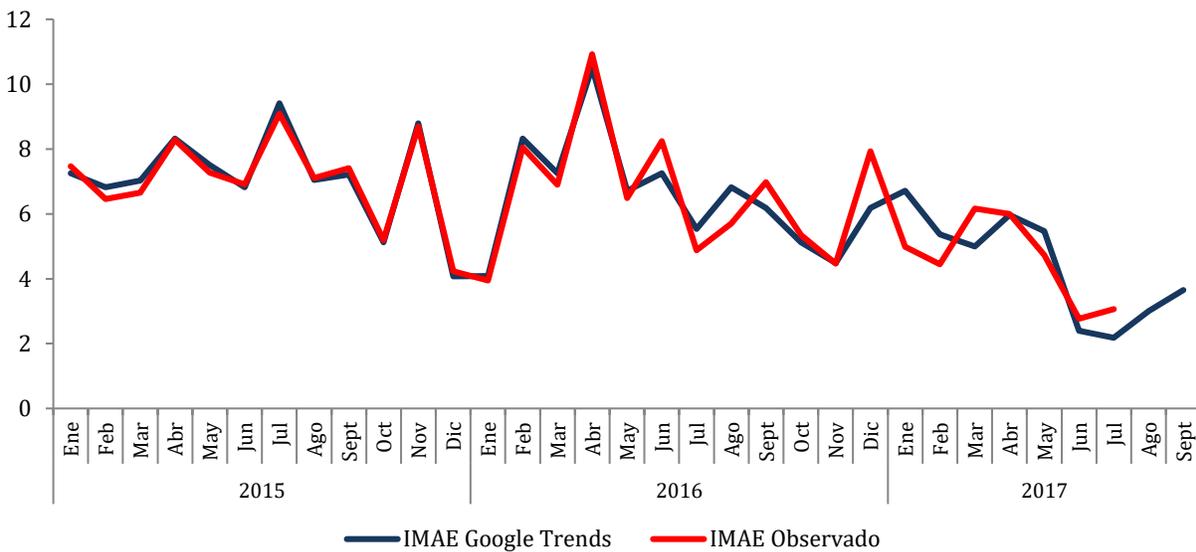
automóviles" (2.0%) y "precios de *commodities*" (1.83%). Otras variables asociadas a compra de bienes de consumo, como ropa, accesorios y contratación de servicios, no sobrepasan el umbral de inclusión, no obstante, la brecha entre sus respectivas probabilidades y el umbral de inclusión establecido de 0.0173 es ínfima.

Los predictores más probables determinados para el Consumo Privado fueron "Reservaciones de Hoteles" (69.7%), "Compras de Supermercado" (12%), "Precio Canasta Básica" (10.5%), y "Expectativas de Mercado" (5.2%). Pese a que otras variables vinculadas a factores tales como precios de bienes básicos ("precios plátano", "precios arroz"), servicios de aseguradoras ("MAPFRE Seguros"), acarreo de mercancías desde el exterior ("CPS servicio"), compra de bienes de consumo no durables ("Victoria's Secret shop online", "Compra de accesorios del Hogar", "Forever21") se posicionaron por encima del umbral de inclusión, no pueden ser incluidas en el modelo, ya que se sobrepasaría la cantidad de regresores pautados *a priori* y sus probabilidades de inclusión son menores que las de las variables explicativas consideradas.

En el caso de la inversión, las variables que exhiben una mayor probabilidad de inclusión son: "Compra de Maquinarias" (37.5%), "Ferreterías" (11.0%), "Precios de aparatos quirúrgicos" (5.3%), y "Precios Materiales Construcción" (4.4%). Resulta lógico pensar que estos regresores sobrepasen el umbral de inclusión, considerando que la mayor proporción de la dinámica de la inversión en la República Dominicana se explica por la construcción, que a su vez se apoya en la compra de materiales. Por otra parte, la importación de bienes de capital constituye un factor de alta ponderación en el comportamiento de esta variable, lo que explica la presencia de variables tales como "Precios de Aparatos Quirúrgicos", las cuales son mayormente bienes importados.

Es importante enfatizar que los coeficientes para estas variables, así como también las probabilidades de inclusión en el modelo, pueden variar de acuerdo al número de iteraciones seleccionadas, así como también de los datos; cada resolución del modelo puede generar cambios en las variables que se ubican por encima del umbral de inclusión. Una posible solución a este problema es fijar las variables en el componente de regresión, no obstante, esta fijación no permitiría que el modelo incorpore cambios derivados de la nueva información procedente de los datos más recientes.

Gráfico 1. Indicador Adelantado del IMAE en base a *Google Trends*



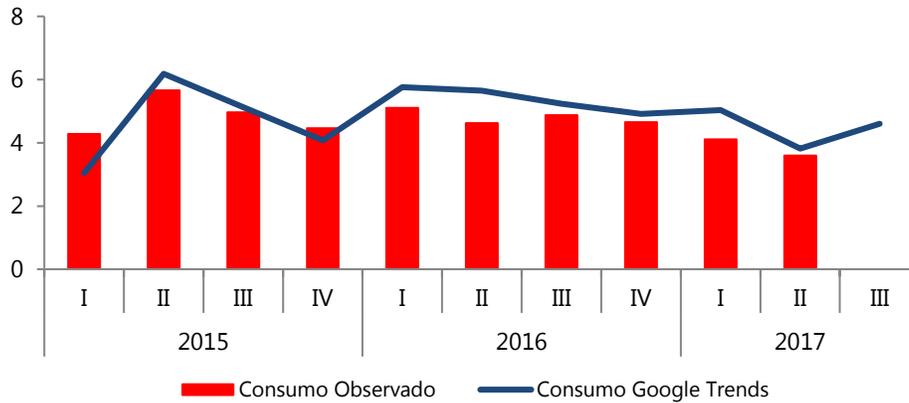
*Fuente: BSTS y elaboración propia.

Una vez que se identificaron los regresores relevantes para cada modelo, se generaron las proyecciones derivadas del Modelo BSTS. Las predicciones fueron realizadas dentro de muestra para los años 2016 y 2017 (hasta septiembre) y evaluadas en términos del error de pronóstico (ECM) y la bondad de ajuste.

Al comparar los resultados del BSTS con los modelos alternativos se constata una disminución de más de un 100% en el error de pronóstico con respecto al VAR de Frecuencia Mixta (*Overall*

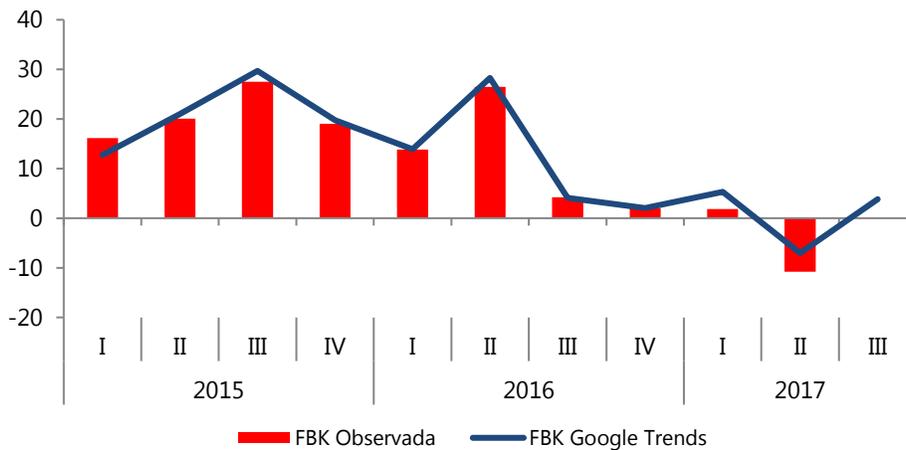
RMSE), para las tres variables consideradas, y en aproximadamente un 60% en comparación a la representación estado-espacio. Los resultados obtenidos justifican la selección del modelo BSTS para lidiar con grandes conjuntos de datos, donde el número de regresores potenciales es mayor que la cantidad de observaciones.

Gráfico 2. Indicador Adelantado Consumo Privado en base a *Google Trends*



*Fuente: BSTS y elaboración propia.

Gráfico 3. Indicador Adelantado Inversión en base a *Google Trends*



*Fuente: BSTS y elaboración propia.

Los gráficos 1, 2 y 3 muestran la coherencia entre los indicadores construidos en base a los datos procedentes de *Google Trends* con las variables efectivamente observadas de IMAE, Consumo Privado e Inversión. Los resultados generados a través del modelo base se tornan prometedores

ya que, aparentemente, las señales provistas por los indicadores de *Google Trends* no solamente resultan efectivas para proyectar las variables con un error de pronóstico mínimo, sino que también permiten predecir los puntos de giro de la serie, lo que se visualiza en la convergencia entre el indicador construido en base a dichas señales y los valores observados.

VII. Conclusiones

Los avances en términos de modelación hacen plausible la evaluación de la dinámica futura de las principales variables macroeconómicas, incorporando información en tiempo real y disponible de manera oportuna que, indiscutiblemente, está orientada a reforzar los procesos de toma de decisiones, a través de un conjunto de información que se hace cada vez más amplio. En este sentido, las técnicas inherentes a la teoría de *machine learning* para el manejo de grandes conjuntos de datos, se perfilan como herramientas de gran valor para este propósito y con alto potencial para continuar desarrollándose. El verdadero reto para las entidades públicas y privadas consistirá en la explotación apropiada de este tipo de datos, con la intención de escrudiñar y optimizar el beneficio derivado de las mismas.

El objetivo de esta investigación se centraba en verificar cómo la inclusión de nuevas técnicas de modelación, específicamente un Modelo Bayesiano Estructural de Series de Tiempo y la inclusión de datos procedentes de plataformas de internet, pueden coadyuvar a mejorar los pronósticos de determinadas variables macroeconómicas, en este caso el Índice Mensual de Actividad Económica (IMAE), el Consumo Privado y la Inversión. Es pertinente preguntarse cómo la estructura y desarrollo de este trabajo permite responder estas preguntas de investigación y, más aún, cómo los resultados obtenidos se concilian con las expectativas de desempeño de estas nuevas herramientas. En este sentido, este trabajo no solamente explica detalladamente cómo es

posible extraer señales sobre el comportamiento de variables como el IMAE, el Consumo Privado y la Inversión a través de los datos nano-económicos disponibles en *Google Trends*, sino que también se explica cómo utilizar esta información para realizar proyecciones y construir indicadores adelantados.

La metodología considerada para el manejo y realización de las proyecciones permite seleccionar los indicadores que se posicionan como los predictores más probables y relevantes en términos de su probabilidad de inclusión en el modelo de proyección, gracias al Modelo Bayesiano Estructural de Series de Tiempo, que emplea *priors Spike & Slab* para descartar regresores irrelevantes y disminuir la dispersión, simplificando el proceso de selección de variables, que puede representar un problema en el caso de grandes conjuntos de datos.

Los resultados muestran que el Modelo Bayesiano Estructural tiene mejor desempeño que los modelos alternativos seleccionados como *benchmarks*, evaluados en función del error de pronóstico (ECM), observándose una reducción de más de un 100% (*Overall RMSE*) con respecto al VAR de Frecuencia Mixta y en aproximadamente 60% en comparación con el Modelo Estado-Espacio. Los resultados para los indicadores construidos a partir del modelo seleccionado, basado en el *nowcasting* del momento t , evidencian que las señales capturadas a través de *Google Trends* no solamente se tornan eficientes para realizar las proyecciones, sino que también son útiles para predecir los puntos de giro de las series.

Las futuras extensiones de este trabajo contemplan la realización de proyecciones de mediano plazo a través de métodos iterativos y técnicas de Múltiples Insumos-Múltiples Resultados (MIMO, por sus siglas en inglés), así como también la inclusión de ejercicios de impulso-respuesta aplicando choques a las variables seleccionadas y comparando estos resultados con el escenario

base, que es análogo al ejercicio de impulso-respuesta realizado en el contexto de un Vector Autorregresivo (VAR).

VIII. Referencias

- Artola, C. y Galán, E. (2012). "Tracking the future on the Web: Construction of Leading Indicators using Internet Searches". Bank of Spain, Working Paper No. 1203.
- Bholat, D. (2016). "Big Data and Central Banks". Bank of England, Advanced Analytics Division, Quarterly Bulletin.
- Blake, A. y Mumtaz, H. (2012). "Applied Bayesian Econometrics for Central Bankers". Centre for Central Banking Studies Technical Handbook, No.4. Bank of England.
- Bortoli, C. y Combes, S. (2015). "Contribution from Google Trends for forecasting the short-term economic outlook in France: limited avenues". National Institute of Statistics and Economic Studies (INSEE).
- Brodersen, K., Gallusser, F., Koehler, J., Remy, N. y Scott, S. (2015). "Inferring causal impact using Bayesian structural time-series model". *Annals of Applied Statistics*, Vol. 9, No. 1, pp. 247-274.
- Carriere-Swallow, Y. y Labbe, F. (2010). "Nowcasting with Google Trends in an Emerging Market". Working Paper, Central Bank of Chile, No.588.
- Choi, H. and Varian, H. (2008). "Predicting the present with Google Trends". Technical Report, Google Inc.
- Choi, H. y Varian, H. (2012). "Predicting the present with Google Trends". *Economic Record* 88, pp. 2-9.
- Durbin, J. y Koopman, S. J. (2002). "A simple and efficient simulation smoother for state space time series analysis". *Biometrika* 89, pp. 603- 616.
- Fahrmeir, L. y Tutz, G. (1994). "Multivariate Statistical Modelling based on Generalized Linear Models". Springer Series in Statistics, New York.
- Fahrmeir, L. y Wagenpfeil, S. (1997). "Penalized likelihood estimation and iterative Kalman smoothing for non-Gaussian dynamic regression models". *Comp. Stat. Data Analysis* 24, 295- 320.
- George, E. y McCulloch, R. (1997). "Approaches for Bayesian Variable Selection". *Statistica Sinica*, 7, pp. 339-373.
- Geweke, J. y Whiteman, C. (2006). "Bayesian Forecasting". Chapter 1 in *Handbook of Economic Forecasting*, 2006, vol. 1, pp 3-80 from Elsevier.
- Guzmán, G. (2011). "Internet search behavior as an economic forecasting tool: The case of inflation expectations". *The Journal of Economic and Social Measurement*, Vol. 36, No. 3.
- Heimer, R., Stehulak, T. y Kolliner, D. (2015). "Assessing Consumer Confidence with Google Search Terms". *Economic Trends*, Federal Reserve Bank of Cleveland.

- Ishwaran, H., Kogalur, U., y Rao, J. (2010). "Spikeslab: Prediction and variable selection using spike and slab regression". *R Journal*, 2(2), 68-73.
- Kalman, R. (1960). "A new approach to linear filtering and prediction problems". *Journal of Basic Engineering* 82, pp. 35-45.
- Kulkarni, R., Haynes, K., Stough, R. y Paelinck, J. (2009). "Forecasting Housing Prices with Google Econometrics". GMU School of Public Policy Research Paper No. 2009-10.
- Jarocinski, M. (2015). "A note on implementing the Durbin and Koopman simulation smoother". Working Paper, European Central Bank, No.1867.
- Jungbacker, B. y Koopman, S. (2014). "Likelihood-based Dynamic Factor Analysis for Measurement and Forecasting". *Econometrics Journal*, Vol.17, pp. 1-21.
- Levenberg, A., Pulman, S., Moilanen, K., Simpson, E. y Roberts, S. (2014). "Predicting Economic Indicators from Web Text Using Sentiment Composition". *International Journal of Computer and Communication Engineering*, Vol. 3, No. 2.
- Ludwig, N., Feuerriegel, S. y Neumann, D. (2016). "Time Series Analysis for Big Data: Evaluating Bayesian Structural Time Series Using Electricity Prices". *Multikonferenz Wirtschaftsinformatik (MKWI) 2016, Volume III*, pp. 1569-1580.
- Maybeck, P. (1979). "Stochastic models, estimation, and control". *Mathematics in Science and Engineering Series*, Vol. 141.
- Mitchell, T. y Beauchamp, J. (1988). "Bayesian variable selection in linear regression". *J. Amer. Stat. Assoc.*, 83, pp. 1023-1036.
- Neal, P. y Kypraios, T. (2015). "Exact Bayesian inference via data augmentation". *Statistics and Computing*, Vol. 25, Issue 2, pp 333-347.
- Pnevmatikakis, E., Rad, K., Huggins, J. y Paninski, L. (2014). "Fast Kalman Filtering and Forward-Backward Smoothing via a Low-Rank Perturbative Approach". *Journal of Computational and Graphical Statistics*, Vol. 23, Issue 2. pp. 316-339.
- Santana, L. (No prelo). "Nowcasting y Proyecciones de Corto y Mediano Plazo de la Dinámica del Sector Turístico de la República Dominicana: Un enfoque basado en Probabilidades de Google Trends y un Modelo Bayesiano Estructural de Series de Tiempo".
- Schmidt, T. y Vosen, S. (2009). "Forecasting private consumption: Survey-based indicators vs. Google Trends". *Ruhr Economic Papers* 0155.
- Scott, L., y Varian, H. (2014). "Predicting the Present with Bayesian Structural Time Series". *International Journal of Mathematical Modelling and Numerical Optimisation*, Vol. 5 (1/2), pp. 4-23.
- Scott, S. (2016). "Package BSTS for R: Time series regression using dynamic linear models using MCMC". Cran Project.
- Taylor, B., Davies, T., Rowlingson, B. y Diggle, P. (2015). "Bayesian Inference and Data Augmentation Schemes for Spatial, Spatiotemporal and Multivariate Log-Gaussian Cox Processes in R". *Journal of Statistical Software*, Vol. 63, Issue 7.
- Vanderkam, D., Schonberger, R., Rowley, H. y Kumar, S. (2013). "Nearest Neighbor Search in Google Correlate". Google Research

IX. Apéndice I

Cuadro 1. Listado de Variables Explicativas para el Índice Mensual de Actividad Económica (IMAE)

Búsquedas <i>Google Trends</i> - Bloque FBK			
No.	Variable	No.	Variable
x1	Teléfonos Ferreterías	x23	Salarios Construcción
x2	Materiales Construcción	x24	Compra de pino
x3	Compra Cemento	x25	Precios Madera
x4	Construcción Viviendas	x26	Importadora Madera
x5	Costo metro cuadrado construcción	x27	Precios Plywood
x6	Precio mano de obra construcción	x28	Precios Hormigón
x7	Precios materiales construcción	x29	Precios Arena
x8	Construcción_Noticias	x30	Precios Grava
x9	Constructoras República Dominicana	x31	Precios Pintura
x10	Precios blocks	x32	Compra Sanitarios
x11	Sector construcción	x33	Compra Pisos
x12	Precios varillas	x34	Precios Ferretería Americana
x13	Ferretería Ramón Corripio	x35	Construcción
x14	Ferreterías República Dominicana	x36	Yeso Precios
x15	Ferretería Maderera Central	x37	Precios Material Eléctrico
x16	Ferretería Popular	x38	Precios Níquel
x17	Ferretería Cima	x39	Precios Hierro
x18	Ferretería Hermanos Pappaterra	x40	Precios Cobre
x19	Caribe Group Santo Domingo	x41	Precios Aluminio
x20	Ferretería Felimón	x42	Precios Plomo
x21	Ferretería Americana	x43	Precios Zinc
x22	Ferretería Ochoa	x44	Precios Estaño

Fuente: *Google Trends*.

Cuadro 1 (cont.). Listado de Variables Explicativas para el Índice Mensual de Actividad Económica (IMAE)

Búsquedas Google Trends - Importaciones Bienes Capital			
No.	Variable	No.	Variable
x45	Precios Calderas	x60	Importaciones desde Japón
x46	Máquinas Empaquetadoras	x61	Equipos Eléctricos
x47	Aparatos Mecánicos	x62	Precios Hierro Compra
x48	Precios Tractores	x63	Reactores Nucleares Compra
x49	Instrumentos Médicos	x64	Compra Calderas
x50	Comprar Equipos Médicos	x65	Tiendas Armas Online EEUU
x51	Comprar Instrumentos Opticos	x66	Compra Máquinas Mecánicas
x52	Importar Instrumentos Opticos	x67	Compra Máquinas Eléctricas
x53	Equipos Médicos	x68	Repuestos Maquinarias
x54	Medical Equipment_Import	x69	Eléctricos
x55	Precios Equipos Médicos	x70	Exportaciones Bienes de Capital (EEUU)
x56	Precios Instrumentos Opticos	x71	Precios Tractores Caterpillar
x57	Instrumentos Oftalmológicos	x72	Precios Aparatos Quirúrgicos
x58	Precios Máquinas Eléctricas	x73	Precios John Deere
x59	Importaciones desde EEUU	x74	Catálogo John Deere

Fuente: *Google Trends*.

Cuadro 1 (cont.). Listado de Variables Explicativas para el Índice Mensual de Actividad Económica (IMAE)

No.	Búsquedas <i>Google Trends</i> - Consumo Bienes y Servicios		
	Variable	No.	Variable
x75	Automóvil	x101	Ilumel
x76	Automóviles Precios	x102	Casa Cuesta
x77	Ferías Vehículos	x103	Accesorios del Hogar
x78	Ventas Vehículos	x104	Complementos Hogar
x79	Carro Toyota	x105	Autoparts
x80	Carro Honda	x106	Agencias Bienes Raíces
x81	Carro Hyundai	x107	Precios Bebidas Alcohólicas
x82	Mazda	x108	Precios Alimentos
x83	Supercarro	x109	Tipo de Cambio Compra
x84	Corotos.com	x110	Tipo de Cambio Venta
x85	Seguros Vehículos	x111	Ofertas Supermercados
x86	Precios Computadoras	x112	Precios Supermercados
x87	Precios Electrodomésticos	x113	Precios Arroz
x88	Préstamos Personales	x114	Precios Canasta Básica
x89	Muebles Hogar	x115	Precios Cebolla
x90	Ikea	x116	Precios Carne
x91	Servicios Reparaciones	x117	Precios Aceite
x92	Decoraciones	x118	Precios Habichuela
x93	Ebay	x119	Precios Pollo
x94	Comprar Iphone	x120	Precios Víveres
x95	Venta Celulares	x121	Precios Plátano
x96	Shop online	x122	Retail Prices
x97	Amazon	x123	Precio Minorista
x98	Piezas Computadoras	x124	Blusas Damas
x99	Piezas Celulares	x125	Pantalones Damas
x100	Servicios Internet	x126	Ropa Interior Damas

*Fuente: *Google Trends*.

Cuadro 1 (cont.). Listado de Variables Explicativas para el Índice Mensual de Actividad Económica (IMAE)

No.	Búsquedas <i>Google Trends</i> - Consumo Bienes y Servicios		
	Variable	No.	Variable
x127	Victoria's Secret shop	x153	Spa
x128	Zara	x154	Comprar Maquillaje
x129	Pantalones Hombres	x155	Beauty Supply
x130	Camisas Hombres	x156	Clínica Estética
x131	Ropa Interior Hombres	x157	Farmacia
x132	Old Navy	x158	Farmax
x133	Gap Shop Online	x159	Farmacia Carol
x134	Jumbo Precios	x160	Laboratorio Referencia
x135	La Sirena Precios	x161	Laboratorio Amadita
x136	AlSuper Precios	x162	Plaza de la Salud
x137	Especiales Supermercados	x163	Cedimat
x138	Supermercado Bravo	x164	Centro Ginecología y Obstetricia
x139	Supermercado Pola	x165	Clínica San Rafael
x140	Zapatos Hombres	x166	Centro Médico Dominicano
x141	Zapatos Mujeres	x167	Clínica Abreu
x142	Women Shoes	x168	Préstamos Personales
x143	Women Clothes	x169	Electricistas Servicios
x144	Women Lingerie	x170	Doncella Servicios
x145	Men Clothes	x171	Agencia Servicios Domésticos
x146	Men Shoes	x172	Plomeros Servicios
x147	Forever21 Shop	x173	Remax
x148	Energía Costos	x174	Supercasas
x149	Pago Online Edeeste	x175	Seguros Autos
x150	Gasolina Precios	x176	Seguros Pepín
x151	Combustible Precios	x177	La Colonial Seguros
x152	Salón de Belleza y Spa	x178	Mapfre

Fuente: *Google Trends*.

Cuadro 1 (cont.). Listado de Variables Explicativas para el Índice Mensual de Actividad Económica (IMAE)

No.	Búsquedas <i>Google Trends</i> - Consumo Bienes y Servicios		
	Variable	No.	Variable
x179	Seguros Universal	x204	Resort
x180	Senasa	x205	Courier Servicios
x181	ARS Plan Salud	x206	Vimenca
x182	ARS Humano	x207	Moneygram remesas
x183	ARS Palic	x208	Aeropack
x184	Claro Pago Online	x209	Tasa de Interés
x185	Edeeste Pago Online	x210	Precios
x186	Edenorte Pago Online	x211	Expectativas Mercado
x187	Edesur Pago Online	x212	Discurso Banco Central
x188	Windtelecom Servicio	x213	Dollar
x189	Caasd Pago	x214	Gold Gym
x190	Caasd	x215	Smartfit Gym
x191	Cine Tickets	x216	Taxi
x192	Caribbean Cinema	x217	Accesorios Hogar
x193	Palacio de Cine Tickets	x218	Precios Alimentos
x194	Cartelera Cine	x219	Precios Viviendas
x195	Boletería Teatro Nacional	x220	Materiales Construcción
x196	Ticket Express	x221	Ferretería
x197	Internet Banking	x222	Casas
x198	Reservar Hotel	x223	Apartamentos
x199	Millas Banco Popular	x224	Solares
x200	Banco BHD	x225	Costos
x201	Banreservas	x226	Restaurants
x202	Reservar Vuelos	x229	Tipo de Cambio Nominal Extrabancario Compra
x203	Cheap Flights	x230	Tipo de Cambio Nominal Extrabancario Venta

Fuente: *Google Trends*.

Cuadro 2. Listado de Variables Explicativas para la Inversión

Búsquedas Google Trends - Bloque FBK			
No.	Variable	No.	Variable
x1	Teléfonos Ferreterías	x25	Precios Madera
x2	Materiales Construcción	x26	Importadora Madera
x3	Compra Cemento	x27	Precios Plywood
x4	Construcción Viviendas	x28	Precios Hormigón
x5	Costo metro cuadrado construcción	x29	Precios Arena
x6	Precio mano de obra construcción	x30	Precios Grava
x7	Precios materiales construcción	x31	Precios Pintura
x8	Construcción_Noticias	x32	Compra Sanitarios
x9	Constructoras República Dominicana	x33	Compra Pisos
x10	Precios blocks	x34	Precios Ferretería Americana
x11	Sector construcción	x35	Construcción
x12	Precios varillas	x36	Yeso Precios
x13	Ferretería Ramón Corripio	x37	Precios Material Eléctrico
x14	Ferreterías República Dominicana	x38	Precios Níquel
x15	Ferretería Maderera Central	x39	Precios Hierro
x16	Ferretería Popular	x40	Precios Cobre
x17	Ferretería Cima	x41	Precios Aluminio
x18	Ferretería Hermanos Pappaterra	x42	Precios Plomo
x19	Caribe Group Santo Domingo	x43	Precios Zinc
x20	Ferretería Felimón	x44	Precios Estaño
x21	Ferretería Americana	x45	Gasolina Premium
x22	Ferretería Ochoa	x46	Gasolina Regular
x23	Salarios Construcción	x47	Tipo de Cambio Nominal Extrabancario Compra
x24	Compra de pino	x48	Tipo de Cambio Nominal Extrabancario Venta

*Fuente: Google Trends.

Cuadro 2 (Cont.). Listado de Variables Explicativas para la Inversión

Búsquedas Google Trends - Importaciones Bienes Capital			
No.	Variable	No.	Variable
x49	Precios Calderas	x64	Importaciones desde Japón
x50	Máquinas Empaquetadoras	x65	Equipos Eléctricos
x51	Aparatos Mecánicos	x66	Precios Hierro Compra
x52	Precios Tractores	x67	Reactores Nucleares Compra
x53	Instrumentos Médicos	x68	Compra Calderas
x54	Comprar Equipos Médicos	x69	Tiendas Armas Online EEUU
x55	Comprar Instrumentos Opticos	x70	Compra Máquinas Mecánicas
x56	Importar Instrumentos Opticos	x71	Compra Máquinas Eléctricas
x57	Equipos Médicos	x72	Repuestos Maquinarias
x58	Medical Equipment_Import	x73	Eléctricos
x59	Precios Equipos Médicos	x74	Exportaciones Bienes de Capital (EEUU)
x60	Precios Instrumentos Opticos	x75	Precios Tractores Caterpillar
x61	Instrumentos Oftalmológicos	x76	Precios Aparatos Quirúrgicos
x62	Precios Máquinas Eléctricas	x77	Precios John Deere
x63	Importaciones desde EEUU	x78	Catálogo John Deere

*Fuente: Google Trends.

Cuadro 3. Listado de Variables Explicativas para el Consumo Privado

<i>Búsquedas Google Trends-Consumo Bienes y Servicios</i>			
<i>No.</i>	<i>Variable</i>	<i>No.</i>	<i>Variable</i>
x1	Automóvil	x25	Accesorios del Hogar
x2	Automóviles Precios	x26	Complementos Hogar
x3	Ferias Vehículos	x27	Autoparts
x4	Ventas Vehículos	x28	Agencia Bienes Raíces
x5	Carro Toyota	x29	Precios Bebidas Alcohólicas
x6	Carro Honda	x30	Precios Alimentos
x7	Carro Hyundai	x31	Tipo de Cambio Extrabancario Compra
x8	Mazda	x32	Tipo de Cambio Extrabancario Venta
x9	Supercarro	x33	Ofertas Supermercados
x10	Precios Computadoras	x34	Precios Supermercados
x11	Precios Electrodomésticos	x35	Precios Arroz
x12	Préstamos Personales	x36	Precio Canasta Básica
x13	Muebles Hogar	x37	Precio Cebolla
x14	Ikea	x38	Precio Carne
x15	Servicios Reparaciones	x39	Precios Aceite
x16	Decoraciones	x40	Precios Habichuela
x17	Ebay	x41	Precio Pollo
x18	Comprar Iphone	x42	Precios Víveres
x19	Comprar Celulares	x43	Precio Plátano
x20	Shop online	x44	Retail Prices
x21	Amazon	x45	Precio Minorista
x22	Piezas Computadoras	x46	Blusas Damas
x23	Piezas Celulares	x47	Pantalones Damas
x24	Servicios Internet	x48	Ropa Interior Damas

Fuente: *Google Trends*.

Nota: Variables observadas; no representan búsquedas procedentes de *Google Trends*.

Cuadro 3 (cont.). Listado de Variables Explicativas para el Consumo Privado

Búsquedas <i>Google Trends</i> - Consumo Bienes y Servicios			
No.	Variable	No.	Variable
x49	Victoria's Secret Shop	x73	Clínica Estética
x50	Zara	x74	Farmacia
x51	Pantalones Hombres	x75	Farmacia Carol
x52	Camisas Hombres	x76	Laboratorio Referencia
x53	Ropa Interior Hombres	x77	Laboratorio Amadita
x54	Old navy shop online	x78	Plaza de la Salud
x55	Gap Shop	x79	Cedimat
x56	Jumbo Precios	x80	Centro Ginecología y Obstetricia
x57	La Sirena Precios	x81	Centro Médico Dominicano
x58	Especiales Supermercados	x82	Clínica Abreu
x59	Supermercado Bravo	x83	Préstamos Personales
x60	Precios Supermercado Pola	x84	Servicios Electricistas
x61	Zapatos hombres	x85	Doncella
x62	Zapatos Damas	x86	Agencia Servicios
x63	Women Shoes	x87	Servicios Plomeros
x64	Women Clothes	x88	REMAX
x65	Women Lingerie	x89	Supercasas
x66	Men Clothes	x90	Seguros Autos
x67	Men Shoes	x91	Seguros Pepín
x68	Forever 21 Shop	x92	Seguros La Colonial
x69	Salón de Belleza	x93	Seguros Mapfre
x70	Spa	x94	Seguros Universal
x71	Comprar Maquillaje	x95	Seguros Senasa
x72	Beauty Supply	x96	Seguros ARS

Fuente: *Google Trends*.Nota: Variables observadas; no representan búsquedas procedentes de *Google Trends*.

Cuadro 3 (cont.). Listado de Variables Explicativas para el Consumo Privado

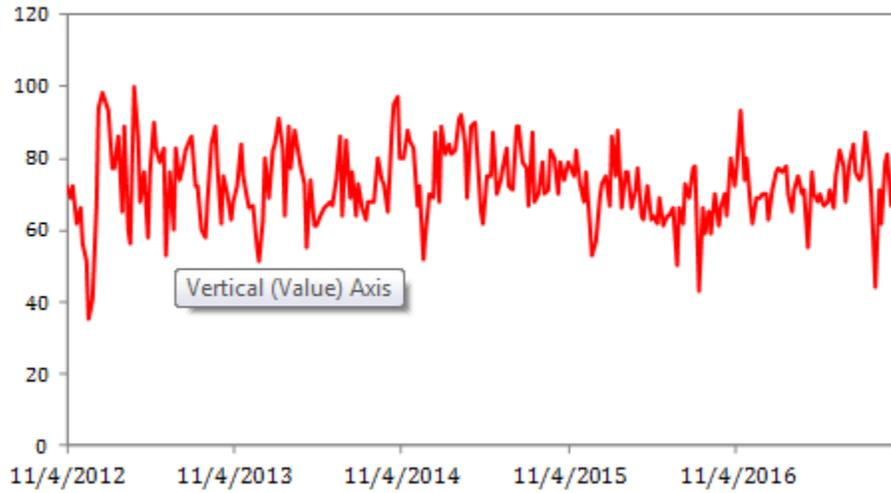
Búsquedas Google Trends-Consumo Bienes y Servicios			
No.	Variable	No.	Variable
x97	ARS Humano	x119	Tasa de Interés
x98	Seguros Palic	x120	Precios
x99	Teléfono Edeeste	x121	Expectativas Mercado
x100	Teléfono Edenorte	x122	Precio Dólar
x101	Teléfono Edesur	x123	Gold Gym
x102	Pago Caasd	x124	Taxi Teléfono
x103	Tickets Cine	x125	Accesorios del Hogar
x104	Caribbean Cinema	x126	Precios Alimentos
x105	Tickets Palacio del Cine	x127	Precios Viviendas
x106	Cartelera de Cine	x128	Materiales Construcción
x107	Ticket Express	x129	Ferretería
x108	Reservar Hotel	x130	Precios Casas
x109	Millas Banco Popular	x131	Precios Apartamentos
x110	Banco BHD León	x132	Precios Solares
x111	Banreservas	x133	Costos
x112	Reservar Vuelo	x134	Restaurants
x113	Cheap Flights	x135	Gasolina Regular*
x114	Resorts	x136	Gasolina Premium*
x115	Servicios Courier	x137	Precios Gasolina
x116	Servicios Vimenca	x138	Precios Combustibles
x117	Servicios EPS	x139	Costos Energía
x118	Servicios CPS	x140	Pago Edeeste

Fuente: *Google Trends*.

Nota: Variables observadas; no representan búsquedas procedentes de *Google Trends*.

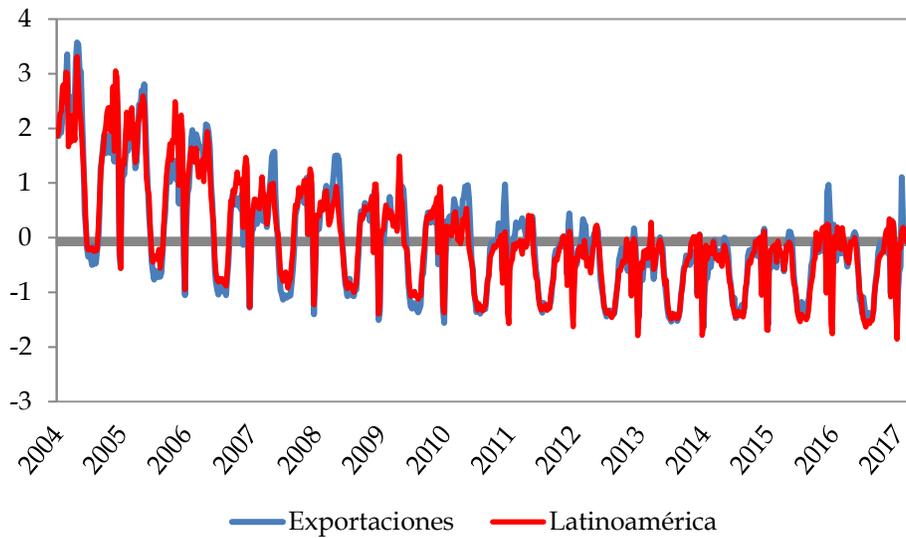
X. Apéndice II

Gráfico 4. Ejemplo de Popularidad de Búsqueda de la palabra Crédito en *Google Trends*, República Dominicana



Fuente: *Google Trends*.

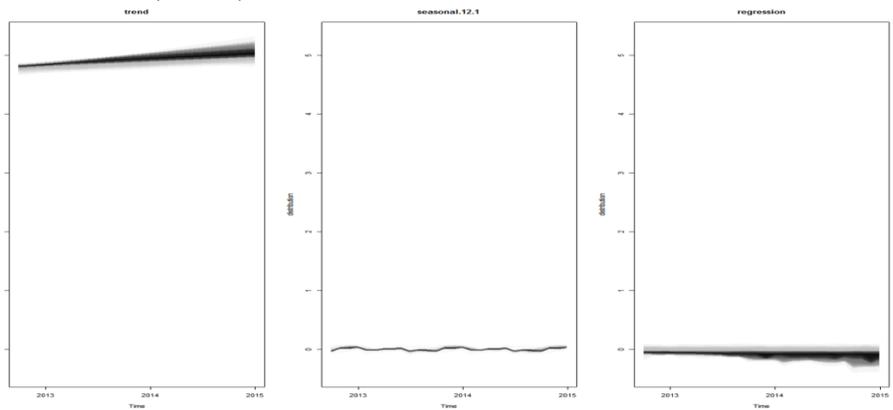
Gráfico 5. Ejemplo de Correlación encontrada a través de *Google Correlates*, EEUU



Fuente: *Google Correlates*.

Nota: Coeficiente de Correlación de Pearson=0.956.

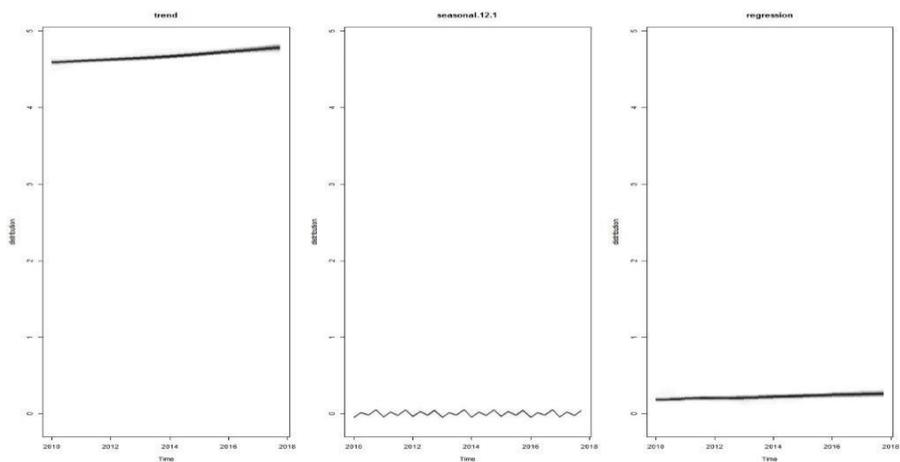
Gráfico 6. Componentes del Modelo Bayesiano Estructural para la estimación del Índice Mensual de Actividad Económica (IMAE)



Fuente: Modelo Bayesiano Estructural (BSTS).

Nota: Contribuciones por componente a la estimación del modelo estado-espacio BSTS para IMAE. En el Panel (a) se muestra el comportamiento de la tendencia; el Panel (b) muestra el patrón estacional y el Panel (c) presenta el componente de regresión.

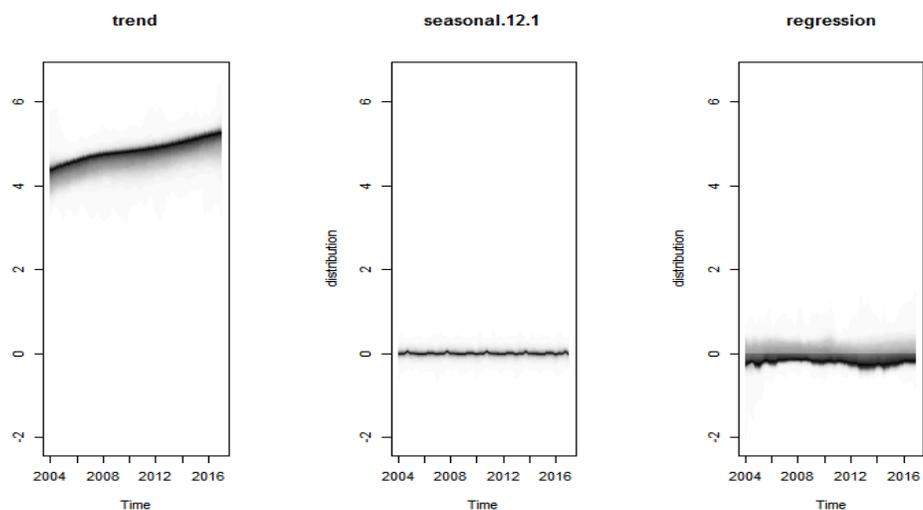
Figura 7. Componentes del Modelo Bayesiano Estructural para la estimación del Consumo Privado



*Fuente: Modelo Bayesiano Estructural (BSTS).

**Nota: 1/ Contribuciones por componente a la estimación del modelo estado-espacio BSTS para Consumo Privado. En el Panel (a) se muestra el comportamiento de la tendencia; el Panel (b) muestra el patrón estacional y el Panel (c) presenta el componente de regresión.

Figura 8. Componentes del Modelo Bayesiano Estructural para la estimación de la Inversión 1/



*Fuente: Modelo Bayesiano Estructural (BSTS).

**Nota: 1/ Contribuciones por componente a la estimación del modelo estado-espacio BSTS para Inversión. En el Panel (a) se muestra el comportamiento de la tendencia; el Panel (b) muestra el patrón estacional y el Panel (c) presenta el componente de regresión.

Cuadro 4. Resumen Error Cuadrático Medio (ECM) Modelo Bayesiano Estructural

	Período	IMAE	Consumo	
			Privado 1/	Inversión 2/
2015	Ene	0.22	-	-
	Feb	0.36	-	-
	Mar	0.38	1.24	3.38
	Abr	0.04	-	-
	May	0.23	-	-
	Jun	0.09	0.53	0.96
	Jul	0.32	-	-
	Ago	0.06	-	-
	Sept	0.20	0.17	2.18
	Oct	0.07	-	-
	Nov	0.10	-	-
	Dic	0.16	0.38	0.69
2016	Ene	0.13	-	-
	Feb	0.27	-	-
	Mar	0.35	0.66	0.05
	Abr	0.41	-	-
	May	0.23	-	-
	Jun	1.00	1.03	1.66
	Jul	0.66	-	-
	Ago	1.12	-	-
	Sept	0.79	0.36	0.14
	Oct	0.23	-	-
	Nov	0.02	-	-
	Dic	1.75	0.26	0.02
2017	Ene	1.73	-	-
	Feb	0.92	-	-
	Mar	1.17	0.93	3.78
	Abr	0.03	-	-
	May	0.74	-	-
	Jun	0.37	-	-
	ECM General	0.472	0.616	1.428
	Bondad de Ajuste	0.945	0.915	0.937

Fuente: BSTS

Notas: 1/ Evaluado Trimestralmente.

2/ Evaluado Trimestralmente.