



**6to Simposio Internacional de Investigación en Ciencias Económicas,
Administrativas y Contables - Sociedad y Desarrollo**
*2do Encuentro Internacional de Estudiantes de Ciencias Económicas,
Administrativas y Contables*



Bogotá, 12, 13 y 14 de septiembre de 2019

**MODELO DE PRONÓSTICO PARA LA DEMANDA DE TURISTAS EN
COLOMBIA A PARTIR DE CRITERIOS DE BÚSQUEDA EN GOOGLE, UNA
APROXIMACIÓN UTILIZANDO LA METODOLOGÍA MIDAS**

Eje temático: Teoría Económica y desarrollo

Alexander Correa Ospina, Economista, Magister en Economía. Cargo Actual: Docente-Investigador área de mercadeo. Universitaria Agustiniiana
Email: alexander.correa@uniagustiniana.edu.co

Luis Enrique Caballero Andrade, Administrador de Empresas, Magister en Gerencia de Mercadeo. Cargo Actual: Docente-Coinvestigador área de mercadeo Universitaria Agustiniiana. Email: luis.caballero@uniagustiniana.edu.co

RESUMEN

Es razonable asumir que lo que las personas buscan hoy en internet es predictivo de lo que han hecho recientemente o harán en el futuro cercano, siendo la razón por la cual Google se ha convertido en una de las compañías más valoradas en el mundo. Este estudio analiza la capacidad de los datos producidos por Google Trends para predecir la llegada de turistas a Colombia. En primer lugar, el estudio analiza si Google Trends permite mejorar la capacidad de pronóstico. En segundo lugar, se evalúa si una variable de alta frecuencia (Google Trends semanal) es mejor para realizar pronósticos que una variable de baja frecuencia (llegadas mensuales de turistas). En este estudio se introduce un nuevo indicador para pronosticar la demanda de turistas construido a partir de los criterios de búsqueda de Google Trends. El indicador es basado en un criterio de búsqueda compuesto "hoteles-vuelos-Colombia" realizados en Estados Unidos, Canadá y el Reino Unido. Siguiendo a Bangwayo-Skeete y Skeete (2015) se evalúa la capacidad predictiva del indicador usando regresiones econométricas del tipo Mixed-Data Sampling (MIDAS). La metodología MIDAS, desarrollada inicialmente por Ghysels et al. (2007) y



**6to Simposio Internacional de Investigación en Ciencias Económicas,
Administrativas y Contables - Sociedad y Desarrollo**
*2do Encuentro Internacional de Estudiantes de Ciencias Económicas,
Administrativas y Contables*



Bogotá, 12, 13 y 14 de septiembre de 2019

posteriormente por Andreou et al. (2010) permite estimar modelos que explican una variable de baja frecuencia por medio de una variable de alta frecuencia y sus respectivos rezagos. Los resultados sugieren que la información de Google Trends ofrece beneficios significativos para pronosticar la llegada de turistas. Por lo tanto, los involucrados en el diseño de la política pública de turismo y empresas en este sector pueden aprovechar la capacidad de pronóstico de Google Trends para mejorar sus procesos de planeación.

Abstract

It is reasonable to assume that what people search the Internet today is predictive of what they have done recently or will do in the near future, which is why Google has become one of the most valued companies in the world. This study analyzes the capacity of the data produced by Google Trends to predict the arrival of tourists to Colombia. In the first place, the study analyzes whether Google Trends improves the forecasting capacity. Second, we evaluate whether a high frequency variable (Google Trends weekly) is better for forecasting than a low frequency variable (monthly arrivals of tourists). In this study a new indicator is introduced to forecast the demand of tourists built from the Google Trends search criteria. The indicator is based on a composited search criterion of "hotels-flights-Colombia" made in the United States, Canada and the United Kingdom. Following Bangwayo-Skeete and Skeete (2015) the predictive capacity of the indicator is evaluated using econometric regressions of the Mixed-Data Sampling (MIDAS) type. The MIDAS methodology developed initially by Ghysels et al. (2007) and later by Andreou et al. (2010) allows to estimate models that explain a low frequency variable by means of a high frequency variable and its respective lags. The results suggest that Google Trends information offers significant benefits to forecast the arrival of tourists. Therefore, those involved in the design of public tourism policy and companies in this sector can take advantage of Google Trends' forecasting ability to improve their planning processes.



**6to Simposio Internacional de Investigación en Ciencias Económicas,
Administrativas y Contables - Sociedad y Desarrollo**
*2do Encuentro Internacional de Estudiantes de Ciencias Económicas,
Administrativas y Contables*



Bogotá, 12, 13 y 14 de septiembre de 2019

1. Introducción

Las personas revelan información útil acerca de sus necesidades, deseos, intereses, y preocupaciones a través de sus búsquedas de internet. Esta puede ser la mejor explicación del éxito de Google ya que ha desarrollado de manera rápida información útil de carácter público. Es un supuesto razonable que lo que las personas buscan hoy es predictivo de lo que ellos han hecho recientemente o planean hacer en el futuro cercano.

Diferentes estudios se han enfocado en el tema de “predecir el presente” y han mostrado que los criterios de búsqueda se correlacionan con actividades contemporáneas (Askitas y Zimmermann, 2009; Hong, 2011; Choi y Varian, 2012). De hecho, Choi y Varian (2012) muestran cómo usar los datos producidos por un motor de búsqueda para predecir el valor presente (nowcasting) de indicadores económicos, tales como desempleo, ventas de automóviles, confianza del consumidor y tendencias de viaje.

Varios estudios han mostrado que los datos de Google Trends son útiles como indicador económico. Investigadores han usado el Índice de Automóviles de Google Trends para intentar mejorar el ajuste y eficiencia de los modelos de nowcasting para ventas de carros en Chile (Carriere-Swallow y Labbe, 2013), también se ha mostrado que existe una correlación fuerte entre los criterios de búsqueda de internet y las tasas de desempleo en Alemania (Askitas y Zimmermann, 2009); proyectar el precio del petróleo usando indicadores macroeconómicos y búsquedas de Google (Fantazzini y Fomichev, 2014); usar el Google Flu Trends para describir la proliferación de influenza en los Estados Unidos durante 2003-2009 (Dukic et al., 2012); determinar si los criterios de Google pueden mejorar las predicciones de desempleo juvenil en Francia (Fondeur y Karam, 2013); ofrecer beneficios significativos a las proyecciones de los indicadores de consumo privado utilizando series de tiempo proporcionadas por Google Trends (Vosen y Schmidt, 2011); usar el volumen de criterios de búsqueda para proyectar resultados tales como



**6to Simposio Internacional de Investigación en Ciencias Económicas,
Administrativas y Contables - Sociedad y Desarrollo**
*2do Encuentro Internacional de Estudiantes de Ciencias Económicas,
Administrativas y Contables*



Bogotá, 12, 13 y 14 de septiembre de 2019

niveles de desempleo, ventas de carros y casas, y prevalencia de enfermedades en tiempo casi real (Goel et al., 2010); analizar factores que influyen la información de los inversionistas cuando se producen anuncios de ganancias a través de las búsquedas de Google (Drake et al., 2012); enfatizar una aproximación a la diversificación de portafolio utilizando como referencia la popularidad de una acción medida por los criterios de búsqueda de Google Trends (Kristoufek, 2013), y han mostrado que Google Trends mejora los métodos de pronóstico usando datos sobre consumo de petróleo

En este estudio se evalúa la relación entre Google Trends y las llegadas de turistas a Colombia durante el periodo 201X-201X. Pronosticar la llegada de turistas no solo desempeña un papel fundamental para empresarios y aquellos que diseñan la política pública, sino también puede ayudar en el desarrollo de metodologías en la literatura sobre turismo. En ese sentido el principal objetivo de este artículo es identificar si Google Trends tiene valor agregado en pronosticar la demanda de turismo al tiempo que adicionalmente contribuye al campo de las siguientes formas: Primero, el artículo está enfocado en estudiar la posible relación entre búsquedas de internet y llegadas de turistas en tiempo real. Google Trends tiene potencial para el mercado empresarial definiendo actividades de turistas en tiempo presente evitando de esta manera meses de espera para obtener información sobre llegadas de turistas del ministerio de industria y comercio. Segundo, el artículo proporciona un procedimiento paso a paso para pronosticar la llegada de turistas ignorando modelación de frecuencias del mismo orden. La metodología Mixed Data Sampling (MIDAS) permite estimar modelos que explican una variable de baja frecuencia por medio de una variable de alta frecuencia y sus respectivos rezagos.

El artículo está organizado de la siguiente manera. En la sección 2 se revisa la literatura relacionado con el pronóstico de llegada de turistas y Google Trends. La sección 3 discute la metodología y la muestra de datos. La sección 4 discute los resultados



**6to Simposio Internacional de Investigación en Ciencias Económicas,
Administrativas y Contables - Sociedad y Desarrollo**
*2do Encuentro Internacional de Estudiantes de Ciencias Económicas,
Administrativas y Contables*



Bogotá, 12, 13 y 14 de septiembre de 2019

empíricos del modelo MIDAS realizado sobre llegada de turistas. La sección 5 presenta las principales conclusiones.

2. Revisión de la literatura

El pronóstico de turistas ha sido un tema de interés en la literatura. Diferentes investigadores han analizado la demanda de turistas usando índices de precios de los países de origen y destino para evaluar el desempeño de pronóstico de las preferencias de los turistas, usando llegadas de turistas a España (Gonzalez y Moral, 1995), y desarrollando modelos de pronóstico basados en diferentes métodos de series de tiempo usando flujos de turistas hacia Hong Kong y provenientes de China, Corea del Sur, Reino Unido y los Estados Unidos (Song et al., 2011). Otros autores han usado diferentes modelos de series de tiempo para evaluar los determinantes de llegadas de turistas (Athanasopoulos et al., 2011; Akin, 2015), al igual que han propuesto métodos de redes neuronales artificiales (Hadavandi et al., 2011; Claveria y Torra, 2014). El principal objetivo en el estudio de Claveria y Torra (2014) es definir que método proporciona información más adecuada sobre el número de turistas y encuentran que modelos ARIMA superan a los modelos SETAR al igual que a los modelos ANN.

El uso de datos de Google Trends para pronosticar el turismo ha sido de igual forma sujeto de estudio. Bangwayo-Skeete y Skeete (2015) sugieren que los volúmenes de las búsquedas de Google proporcionan ventajas para pronosticar la demanda de turistas para ciertos destinos en el Caribe. Otros autores han argumentado que Google Trends podría realizar pronósticos más precisos en Suiza si se usará como un indicador concurrente (Silverstone y Wochner, 2017), y que existe una fuerte correlación entre visitantes de hoteles y criterios de búsqueda en Google en Puerto Rico (Rivera, 2016). Park et al. (2017) realizan pronósticos de corto plazo para los flujos de turistas de Corea del Sur hacia Japón. Estos autores argumentan que Google Trends no solo mejora la



**6to Simposio Internacional de Investigación en Ciencias Económicas,
Administrativas y Contables - Sociedad y Desarrollo**
*2do Encuentro Internacional de Estudiantes de Ciencias Económicas,
Administrativas y Contables*



Bogotá, 12, 13 y 14 de septiembre de 2019

precisión de los pronósticos de demanda de turistas, sino también que los pronósticos fuera de la muestra superan a los pronósticos dentro de la muestra cuando se usa Google Trends.

Google Trends proporciona información gratuita, abundante y prácticamente en tiempo real aunque con algunas desventajas. En primer lugar, Google únicamente muestra datos absolutos, proporcionando un índice que es relativo a todas las búsquedas. En segundo lugar, los usuarios de internet podrían digitar palabras similares incluso si están buscando temas diferentes, o diferentes palabras, incluso si están buscando temas similares. En tercer lugar, los criterios de búsqueda de internet están relacionados a características personales tales como educación, ingreso, edad, entre otros. Claramente, los datos de Google son imperfectos; sin embargo, tomado como referencia el hecho que es una de las mejores bases de datos en tiempo real tiene el potencial de actuar como un indicador líder.

MIDAS es un método para estimar y pronosticar el impacto de variables de alta frecuencia sobre variables dependientes de baja frecuencia por lo que es posible ignorar el requerimiento tradicional que las variables deben estar en la misma frecuencia. MIDAS usa rezagos distribuidos de los polinomios para garantizar especificaciones parsimonias que permitan manejar series de tiempo cuyas muestras son de frecuencias diferentes. La metodología MIDAS propuesta por Ghysels et al. (2006) fue adicionalmente desarrollada por Andreou et al. (2010) quienes propusieron una nueva descomposición para la regresión MIDAS. Estudios empíricos en la literatura MIDAS han analizado la dinámica en la microestructura del ruido y la volatilidad (Ghysels et al., 2007), pronosticado el crecimiento del PIB (Ghysels y Wright, 2009; Andreou et al., 2012), realizado pronósticos en tiempo presente (nowcasting) del crecimiento de PIB en la zona Euro (Kuzin et al., 2011), y la volatilidad del mercado accionario y actividad macroeconómica (Engle et al., 2013; Girardin y Joyeux, 2013).



**6to Simposio Internacional de Investigación en Ciencias Económicas,
Administrativas y Contables - Sociedad y Desarrollo**
*2do Encuentro Internacional de Estudiantes de Ciencias Económicas,
Administrativas y Contables*



Bogotá, 12, 13 y 14 de septiembre de 2019

Este artículo analiza la potencialidad de Google Trends para pronosticar la llegada de turistas a Colombia. El mismo reporta si datos semanales de Google Trends pueden mejorar los pronósticos cuando se usan regresiones MIDAS. Primero, el estudio analiza si Google Trends ofrece mejoras significativas en los pronósticos. Segundo, valora si una variable explicativa de alta frecuencia es mejor para pronosticar adecuadamente cuando se compara con datos semanales y mensuales de Google Trends usando regresiones MIDAS.

3. Metodología y datos

3.1. Metodología

Este estudio considera cómo obtener mejores análisis de llegadas de turistas usando MIDAS y esta dirigido a detectar si los criterios de Google Search pueden agregar una visión útil en la predicción de llegadas de turistas a Colombia. La metodología de pronóstico empieza eligiendo un modelo de línea base que tenga un poder predictivo significativo. Después de eso, se corre el modelo de línea base tanto con datos de Google como sin ellos para poder determinar si Google mejora los pronósticos de llegada de turistas.

La metodología MIDAS fue usada por Bangwayo-Skeete y Skeete (2015) para estudiar la llegada de turistas de los Estados Unidos, Canadá y el Reino Unido a cinco destinos en el Caribe. Estos autores enfatizan que la información que Google Trends proporciona sobre turistas beneficia de manera significativa los pronósticos y que la metodología MIDAS supera otros métodos econométricos.

La metodología usada en este estudio sigue a Ghysels et al. (2007) y Andreou et al. (2010) y ha sido organizada específicamente para este estudio:



6to Simposio Internacional de Investigación en Ciencias Económicas,
Administrativas y Contables - Sociedad y Desarrollo
2do Encuentro Internacional de Estudiantes de Ciencias Económicas,
Administrativas y Contables



Bogotá, 12, 13 y 14 de septiembre de 2019

$$turistas_t = \alpha + \sum_{i=1}^n \beta_i L^i turistas_t + \gamma \sum_{i=1}^m B(k; \theta) L^{k/m} google_t^{(m)} + \epsilon_t^{(m)} \quad (1)$$

para $t = 1, \dots, T$, donde la función $B(k; \theta)$ es una especificación polinómica que determina las ponderaciones para la agregación temporal. $L^{k/m}$ representa el operador de rezagos tal que $L^{k/m} google_t = google_{t-k/m}(m)$. En el modelo, mientras que $turistas_t$ representa la variable dependiente de baja frecuencia, $google_t$ representa la variable independiente de alta frecuencia. L es el operador de rezagos. β representa el efecto de los valores rezagados de los turistas y γ representa el efecto de las búsquedas de Google.

La parametrización de la función de ponderación es una de las principales contribuciones de la regresión MIDAS. Ghysels et al. (2007) propone dos parametrizaciones diferentes. La primera es

$$B(k; \theta) = \frac{\epsilon^{\theta_1 k + \dots + \theta_Q k^Q}}{\sum_{k=1}^m \epsilon^{\theta_1 k + \dots + \theta_Q k^Q}} \quad (2)$$

que sugiere una especificación Almon exponencial (Almon, 1965). Ghysels et al. (2006) usa la forma funcional expresada en (2) con dos parámetros ($\theta = [\theta_1; \theta_2]$). La especificación resulta en ponderaciones iguales cuando $\theta_1 = \theta_2 = 0$; de otro modo las ponderaciones pueden caer de manera rápida o lenta con el número de rezagos. La tasa de caída determinada por el número de rezagos es incluida en el modelo. La función exponencial puede producir formas de joroba y una caída en la ponderación es garantizada en la medida que $\theta_2 \leq 0$.



**6to Simposio Internacional de Investigación en Ciencias Económicas,
Administrativas y Contables - Sociedad y Desarrollo**
*2do Encuentro Internacional de Estudiantes de Ciencias Económicas,
Administrativas y Contables*



Bogotá, 12, 13 y 14 de septiembre de 2019

La segunda parametrización es una formulación Beta

$$B(k; \theta_1, \theta_2) = \frac{f(k/m, \theta_1; \theta_2)}{\sum_{k=1}^m f(k/m, \theta_1; \theta_2)} \quad (3)$$

donde

$$f(i, \theta_1; \theta_2) = \frac{i^{\theta_1-1}(1-i)^{(\theta_2-1)}\Gamma(\theta_1+\theta_2)}{\Gamma(\theta_1)\Gamma(\theta_2)} \quad (4)$$

θ_1 y θ_2 son hiperparámetros que gobiernan la forma de la función de ponderación, y

$$\Gamma(\theta_p) = \int_0^\infty e^{-i} i^{\theta_p-1} di \quad (5)$$

es la función gamma estándar. La especificación Beta también resulta en una ponderación igual cuando $\theta_1 = \theta_2 = 0$. La tasa de caída determina de qué manera los rezagos son incluidos en el modelo como en el caso de Almon. La ponderación cae lentamente en la medida que $\theta_1 = 1$ y $\theta_2 > 1$. En tanto que si θ_2 aumenta, la ponderación cae rápidamente.

La evaluación de la calidad de un pronóstico requiere comparar los valores pronosticados con los valores actuales y los modelos alternativos. El test Diebold-Mariano compara dos modelos de pronóstico para evaluar si estos tienen igual exactitud predictiva o si un modelo es más adecuado que el otro. El test Diebold-Mariano se especifica a continuación

$$DM = \frac{\tilde{d}}{s_d} \quad (6)$$



**6to Simposio Internacional de Investigación en Ciencias Económicas,
Administrativas y Contables - Sociedad y Desarrollo**
*2do Encuentro Internacional de Estudiantes de Ciencias Económicas,
Administrativas y Contables*



Bogotá, 12, 13 y 14 de septiembre de 2019

donde \tilde{d} y s_d son el promedio y la desviación estándar de la muestra de d . En tanto que d estima

$$d = \epsilon_1 - \epsilon_2 \quad (7)$$

donde ϵ_i representa la diferencia cuadrada o absoluta entre el pronóstico y el valor actual para los dos modelos ($i = 1, 2$). Este estudio se concentra en valores absolutos definidos como $\epsilon_i = \hat{y}_i - y_i$, donde \hat{y}_i representa el valor pronosticado y y_i representa el valor observado real. La hipótesis nula del test Diebold-Mariano es que ambos pronósticos tienen la misma exactitud, la hipótesis alternativa es que el modelo 2 (el modelo de Google Trends) es más adecuado que el modelo de línea base (modelo sin Google Trends).

3.2. Datos y estadísticas descriptivas

Datos mensuales de llegadas de turistas por avión de diferentes países a Colombia de Enero de 2010 a Diciembre de 2018 se obtuvieron del Centro de Información Turística de Colombia (CITUR). Ambas series de tiempo muestran una tendencia ascendente y variaciones estacionales. Existen múltiples métodos para pronosticar series de tiempo basados en tendencias al igual que estacionalidad. La diferencia del logaritmo natural ha sido usada para obtener la tasa de crecimiento interanual y así eliminar tendencias lineales y variaciones estacionales.

Las series históricas de los volúmenes de búsqueda para el término combinado vuelos+hoteles+Colombia fueron recogidas usando Google Trends. Las series semanales y mensuales cubren el mismo periodo de tiempo. Google Trends mide con qué frecuencia un término de búsqueda en particular es ingresado en relación con el



**6to Simposio Internacional de Investigación en Ciencias Económicas,
Administrativas y Contables - Sociedad y Desarrollo**
*2do Encuentro Internacional de Estudiantes de Ciencias Económicas,
Administrativas y Contables*

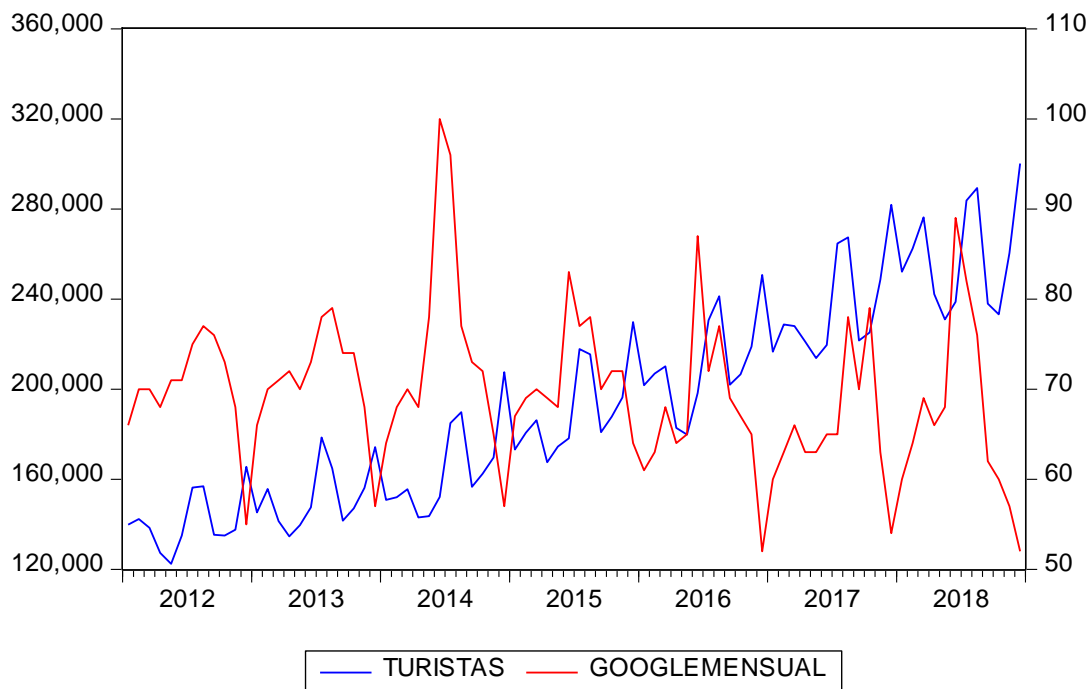


Bogotá, 12, 13 y 14 de septiembre de 2019

volumen total de búsqueda de Google en varios países (regiones) y en varios idiomas. Google Trends ajusta los datos de búsqueda para hacer comparaciones. Cada punto es dividido entre el total de búsquedas por la geografía y el rango de tiempo. Los números resultantes son entonces escalados en un rango de 0 a 100 tomando como base la proporción del tema con respecto a todas las búsquedas en todos los temas.

La gráfica 1 muestra las llegadas mensuales de turistas por avión a Colombia y los resultados mensuales de las búsquedas de las palabras clave en Google. Inspección visual de la gráfica indica una fuerte correlación entre llegadas mensuales de turistas y las búsquedas de Google.

Gráfica 1. Llegadas mensuales de turistas a Colombia y búsquedas mensuales en Google para Colombia



Fuente: Cálculos propios, CITUR y Google Trends. El lado izquierdo representa el número de turistas en miles, el lado derecho representa el índice de Google Trends.



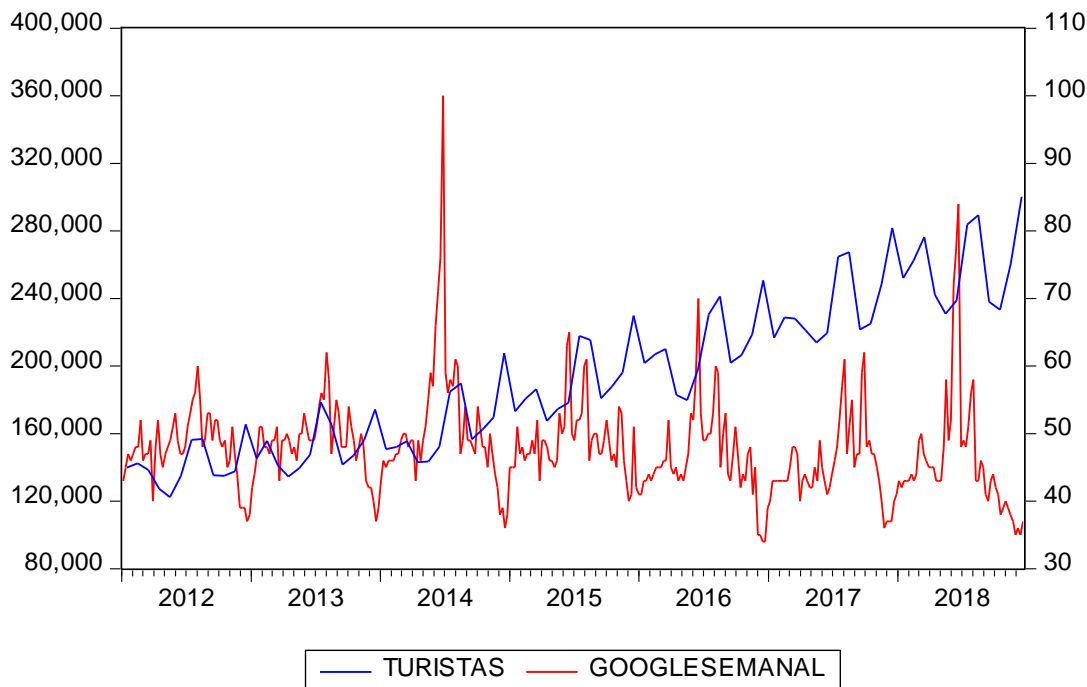
**6to Simposio Internacional de Investigación en Ciencias Económicas,
Administrativas y Contables - Sociedad y Desarrollo**
**2do Encuentro Internacional de Estudiantes de Ciencias Económicas,
Administrativas y Contables**



Bogotá, 12, 13 y 14 de septiembre de 2019

La gráfica 2 muestra la llegada mensual de turistas y los resultados de búsqueda semanales de Google. Aunque aparecen algunos datos atípicos, en general se aprecia una asociación cercana entre las dos variables. Estas evaluaciones visuales dan fundamento a investigar y desarrollar modelos para analizar si Google Trends puede mejorar el pronóstico de llegada de turistas a Colombia.

Gráfica 2. Llegadas mensuales de turistas a Colombia y búsquedas semanales en Google para Colombia



Fuente: Cálculos propios, CITUR y Google Trends. El lado izquierdo representa el número de turistas en miles, el lado derecho representa el índice de Google Trends.



**6to Simposio Internacional de Investigación en Ciencias Económicas,
Administrativas y Contables - Sociedad y Desarrollo**
*2do Encuentro Internacional de Estudiantes de Ciencias Económicas,
Administrativas y Contables*



Bogotá, 12, 13 y 14 de septiembre de 2019

Tabla 1. Estadísticas descriptivas de llegadas mensuales de turistas por país.

País	Promedio	Desviación Estándar	Mínimo	Máximo
Total Mensual	193433.32	45407.94	122437	300231
Estados Unidos	37697.98	11835.9	21510	72680
Venezuela	25550.9	7648.57	14184	47587
Brasil	12254.83	4186.98	5696	21598
Ecuador	12066.18	2789.52	7609	20246
Argentina	11943.85	3712.09	7364	26018
México	11369.39	3320.45	6008	18904
Perú	10467.27	2180.1	6540	14806
Chile	9390.64	2345.75	5131	15929
España	8778.27	1380.85	6572	13898
Panamá	6090.57	3214.62	2057	14307
Otros	47823.43	11778.98	29085	79526

Fuente: Cálculos propios

La tabla 1 representa las estadísticas descriptivas de llegadas de turistas a Colombia por país de origen, entre Enero de 2012 y Diciembre de 2018. La tabla muestra los diez países más importantes, los cuales tienen un impacto significativo en la llegada de turistas a Colombia. Esos diez países representan el 75% del total de llegadas de turistas.

Adicionalmente, este estudio aplica el test Augmented Dickey-Fuller (ADF) y el test Kwiatkowski-Schmidt-Shin (KPSS). El test ADF prueba la hipótesis de raíz unitaria tanto en los niveles (Tabla 2) como en las diferencias internacionales del logaritmo natural (Tabla 3) de la llegada de turistas a Colombia. El test KPSS prueba estacionariedad tanto en niveles como en diferencias internacionales (Tabla 2 y 3).

Como se muestra en la tabla 2, para todos los países de origen, no es posible rechazar la hipótesis nula de una raíz unitaria al nivel del 5%. Resultados similares son obtenidos



**6to Simposio Internacional de Investigación en Ciencias Económicas,
Administrativas y Contables - Sociedad y Desarrollo**
*2do Encuentro Internacional de Estudiantes de Ciencias Económicas,
Administrativas y Contables*



Bogotá, 12, 13 y 14 de septiembre de 2019

para el test KPSS, donde la hipótesis de estacionariedad es rechazada en muchos casos. Cuando las pruebas son aplicadas a la diferencia internaual del logaritmo natural de las series individuales (Tabla 3), la hipótesis nula de una raíz unitaria es altamente rechazada en muchos casos. En el caso del test KPSS, no es posible rechazar la hipótesis nula de estacionariedad al 5% para ningún país. Estos resultados implican que se requiere diferenciar las series y mostrando la importancia de desestacionalizar y eliminar la tendencia en la llegada de turistas antes de modelar y pronosticar.

Tabla 2. Test de raíz unitaria para llegada de turistas a Colombia - Niveles

<u>País</u>	<u>ADF</u>	<u>KPSS</u>
Total Mensual	1.51	1.28
Estados Unidos	0.91	1.22
Venezuela	-0.46	1
Brasil	-1.65	1.03
Ecuador	-1.17	1.13
Argentina	-0.76	0.7
México	-0.84	1.23
Perú	-1.89	1.17
Chile	-1.48	1.22
España	1.76	0.98
Panamá	-0.46	1.24
Otros	1.9	1.24

Fuente: Cálculos propios. La estimación representa datos mensuales para Enero 2012 – Diciembre de 2018. Test para raíz unitaria: ADF, el valor crítico al 5% es -2.90. Test para estacionariedad: KPSS, el valor crítico al 5% es 0.46.



**6to Simposio Internacional de Investigación en Ciencias Económicas,
Administrativas y Contables - Sociedad y Desarrollo**
*2do Encuentro Internacional de Estudiantes de Ciencias Económicas,
Administrativas y Contables*



Bogotá, 12, 13 y 14 de septiembre de 2019

Tabla 3. Test de raíz unitaria para la llegada de turistas a Colombia – Diferencias interanuales del logaritmo natural

País	ADF	KPSS
Total Mensual	-3.01	0.2
Estados Unidos	-2.49	0.21
Venezuela	-3.16	0.08
Brasil	-3.06	0.23
Ecuador	-5.69	0.47
Argentina	-0.78	0.17
México	-2.85	0.27
Perú	-3.78	0.52
Chile	-6.11	0.34
España	-4.57	0.4
Panamá	-2.2	0.27
Otros	-2.84	0.3

Fuente: Cálculos propios. La estimación representa datos mensuales para Enero 2012 – Diciembre de 2018. Test para raíz unitaria: ADF, el valor crítico al 5% es -2.90. Test para estacionariedad: KPSS, el valor crítico al 5% es 0.46.

4. Resultados

En esta sección se presentan los modelos MIDAS usando la llegada mensual de turistas por avión a Colombia. Los datos oficiales de CITUR son usados para evaluar el desempeño de pronóstico de los modelos de regresión MIDAS que usan datos semanales de Google Trends. Todos los modelos fueron estimados usando datos de Enero de 2012 a Diciembre de 2018 con información semanal de Google Trends.

La Tabla 4 representa resultados para tres diferentes modelos MIDAS que usan datos semanales de Google, datos mensuales de Google, y un modelo sin la información de Google Trends. Los resultados confirman que los rezagos de un mes y doce meses están significativamente correlacionados con cambios en la llegada de turistas. A modo de



**6to Simposio Internacional de Investigación en Ciencias Económicas,
Administrativas y Contables - Sociedad y Desarrollo**
*2do Encuentro Internacional de Estudiantes de Ciencias Económicas,
Administrativas y Contables*



Bogotá, 12, 13 y 14 de septiembre de 2019

ilustración, la llegada de turistas es mensual, en tanto que la información de Google Trends es semanal. Se usan 16 rezagos (semanas) de Google Trends para explicar cada mes de llegadas de turistas. Los resultados muestran que el índice de búsqueda semanal en Google Trends tiene un impacto significativo en la llegada de turistas. Estos resultados son comparables a los obtenidos por Bangwayo-Skeete y Skeete (2015), Siliverstovs y Wochner (2017) y Park et al. (2017), quienes encuentran evidencia de que Google Trends ofrece beneficios significativos para pronosticar la llegada de turistas.

De igual forma, se realizó un modelo que incluía datos mensuales de Google Trends. Estos resultados indican que información de cuatro meses por adelantado es útil para pronosticar el número actual de llegada de turistas en Colombia. En ese sentido, los datos mensuales de Google ofrecen información importante para entender la llegada de turistas a Colombia. Esto confirma que índices de actividad de búsqueda web ofrecen señales tempranas que pueden ayudar significativamente a predecir la llegada de turistas a Colombia con cuatro meses de anticipación.



**6to Simposio Internacional de Investigación en Ciencias Económicas,
Administrativas y Contables - Sociedad y Desarrollo**
**2do Encuentro Internacional de Estudiantes de Ciencias Económicas,
Administrativas y Contables**



Bogotá, 12, 13 y 14 de septiembre de 2019

Tabla 4. Estimaciones modelos MIDAS para llegada de turistas: Enero 2012 – Diciembre 2018

	Google Semanal			Google Mensual	Sin Google
	Beta MIDAS	Step MIDAS	Almon MIDAS	ARIMA	ARIMA
DLTURISTAS(-1)	0.565***	0.568***	0.564***	0.554***	0.553***
DLTURISTAS(-12)	0.320***	-0.346***	0.321***	-0.341***	-0.343***
Beta01	1.666**	0.005	0.03		
Beta02	1.825*	-0.005	-0.013**		
Beta03	-0.056***	0.029*	0.00**		
GoogleMensual(-4)				0.109***	
CONSTANTE	0.08***	0.082***	0.08***	0.082***	0.083***
R2	0.529	0.53	0.531	0.522	0.461

Notas: La variable dependiente es el cambio interanual del logaritmo natural de la llegada de turistas; la ecuación estimada es $turistas_t = \alpha + \sum_{i=1}^n \beta_i L^i turistas_t + \gamma \sum_{i=1}^m B(k; \theta) L^{k/m} google_t^{(m)} + \epsilon_t^{(m)}$. Las columnas (2)-(4) son estimaciones MIDAS usando datos semanales de Google. La columna (5) usa datos mensuales de Google. La columna (6) usa un modelo ARIMA sin información de Google Trends. ***, **, y * representan significancia estadística al 1%, 5% y 10%, respectivamente.

A continuación, se realizan pronósticos por fuera-de-la-muestra para evaluar el desempeño para cada modelo. De esta manera, estimaciones por dentro-de-la-muestra fueron realizadas cubriendo el periodo Enero 2012 a Septiembre 2016, y pronósticos fuera-de-la muestra para el periodo Octubre 2016 a Diciembre 2018.

Los métodos más comunes usados para determinar la precisión de los pronósticos son las funciones de error de pronóstico. Para evaluar el desempeño de pronóstico de los modelos MIDAS que usan datos semanales de Google Trends, fueron utilizados los test Root Squared Forecast Error (RMSFE) y Mean Absolute Percentage Error (MAPE). Los resultados se muestran en la tabla 6. Valores bajos de MAPE y RMSE implican que los métodos de pronóstico de MIDAS que usan datos semanales ofrecen un desempeño de pronóstico mucho mejor que los modelos con datos mensuales de Google y sin datos de Google. Así, la utilidad de un modelo de pronóstico debe ser evaluada por su desempeño de pronóstico por fuera-de-la-muestra. Los resultados muestran que el modelo MIDAS-



**6to Simposio Internacional de Investigación en Ciencias Económicas,
Administrativas y Contables - Sociedad y Desarrollo**
*2do Encuentro Internacional de Estudiantes de Ciencias Económicas,
Administrativas y Contables*



Bogotá, 12, 13 y 14 de septiembre de 2019

BETA tiene un mejor desempeño que los otros modelos ya que tiene el menor error de pronóstico en todos los test (exceptuando MAE).

Tabla 6. Evaluaciones de pronóstico de las estimaciones MIDAS para llegada de turistas

	RMSE	MAE	MAPE	SMAPE	Theil U1
MIDASBETA	0.037887*	0.030748	31.86329*	28.34072*	0.17009*
MIDASSTEP	0.059746	0.04442	54.04465	42.40197	0.274468
MIDASALMON	0.03837	0.032248	33.45233	31.27529	0.179969
MENSUALGOOGLE	0.042157	0.030362	36.56531	28.90561	0.196788
SINGOOGLE	0.039269	0.031016	35.97854	30.25358	0.183489
Simple mean	0.039811	0.031124	35.4955	29.91345	0.185026
Mean square error	0.038521	0.030105*	33.98163	28.99445	0.178602

Notas: Los modelos MIDAS representan datos semanales de Google con diferentes funciones de ponderación. MensualGoogle representa regresiones con datos mensuales de Google y SinGoogle representa los resultados sin la información de Google Trends. * muestra las mejores evaluaciones para los modelos de pronóstico.

La gráfica 3 muestra las evaluaciones de pronóstico para la llegada de turistas usando diferentes regresiones MIDAS. Para la llegada de turistas MIDAS-BETA es el que muestra mejor ajuste.



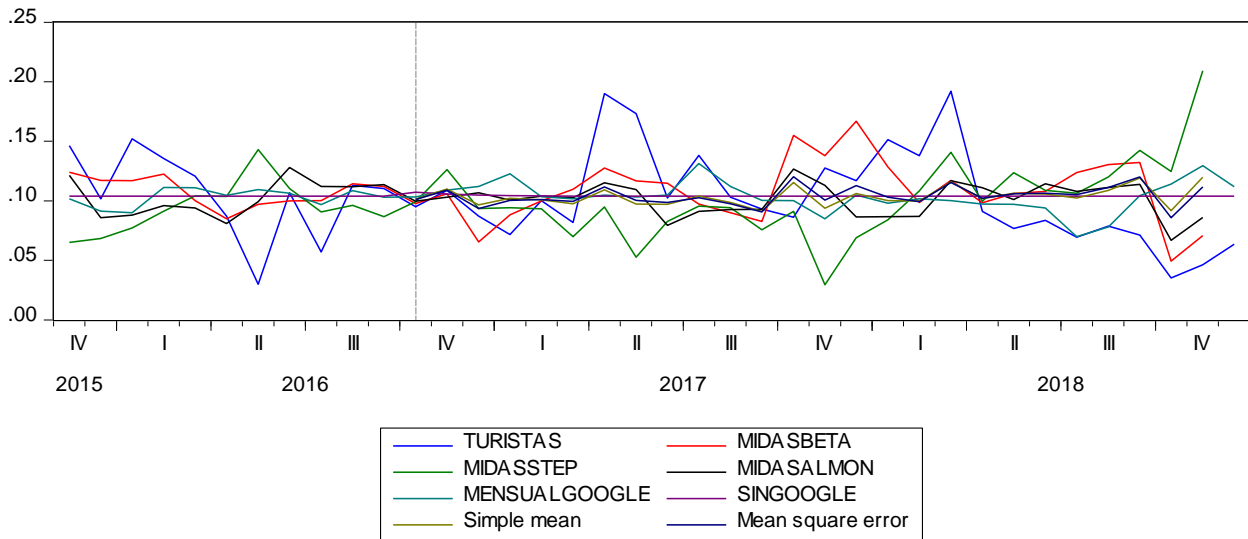
6to Simposio Internacional de Investigación en Ciencias Económicas,
Administrativas y Contables - Sociedad y Desarrollo
2do Encuentro Internacional de Estudiantes de Ciencias Económicas,
Administrativas y Contables



Bogotá, 12, 13 y 14 de septiembre de 2019

Gráfico 3. Pronóstico de llegada de turistas a Colombia.

Forecast Comparison Graph



En resumen, comparando el desempeño de pronóstico de diferentes modelos de series de tiempo aumentados con datos semanales y mensuales de Google Trends y un modelo sin Google Trends, los resultados confirman que los modelos MIDAS con datos semanales de Google Trends tienen un mejor desempeño que el modelo mensual y el modelo sin datos de Google.



**6to Simposio Internacional de Investigación en Ciencias Económicas,
Administrativas y Contables - Sociedad y Desarrollo**
*2do Encuentro Internacional de Estudiantes de Ciencias Económicas,
Administrativas y Contables*



Bogotá, 12, 13 y 14 de septiembre de 2019

5. Conclusiones

El principal objetivo de este estudio es el de desarrollar modelos econométricos que permitan realizar un mejor pronóstico de la llegada de turistas a Colombia. Un adecuado pronóstico de la demanda de turistas es importante debido al rápido crecimiento del turismo global. Las búsquedas de internet desempeñan un papel importante en el turismo y evaluando la dinámica del consumo de turismo. En este estudio se analiza si las búsquedas semanales de Google Trends pueden pronosticar adecuadamente la llegada mensual de turistas a Colombia usando regresiones MIDAS, las cuales permiten ignorar el supuesto de series de la misma frecuencia.

En este estudio se han empleado tres diferentes modelos MIDAS que usan datos de Google semanales y los mismos se han comparado con un modelo mensual que usa datos de Google mensuales y con un modelo que no utiliza la información de Google. El principal objetivo fue el de evaluar si la información de Google Trends produce beneficios significativos para la evaluación y pronóstico de la llegada de turistas a Colombia, al igual que determinar si los datos de alta frecuencia (datos semanales) superan los métodos que usan datos de la misma frecuencia.

Los resultados muestran un potencial innegable para Google Trends como herramienta para mejorar la evaluación y pronóstico de la llegada de turistas. MIDAS permite la evaluación de series de diferente frecuencia tales como los datos semanales de Google y los datos mensuales de llegada de turistas. El desempeño de pronóstico de los indicadores usando el modelo MIDAS-BETA superan los modelos que usan datos de Google mensuales y al modelo sin Google Trends. Los resultados confirman que usar criterios de búsqueda en Google enriquece la información para los que definen la política pública de turismo y para las empresas que operan en el sector. Una adecuada



**6to Simposio Internacional de Investigación en Ciencias Económicas,
Administrativas y Contables - Sociedad y Desarrollo**
*2do Encuentro Internacional de Estudiantes de Ciencias Económicas,
Administrativas y Contables*



Bogotá, 12, 13 y 14 de septiembre de 2019

proyección de la llegada de turistas desempeña un papel importante debido al alto impacto económico en los destinos dependientes de turistas.



**6to Simposio Internacional de Investigación en Ciencias Económicas,
Administrativas y Contables - Sociedad y Desarrollo**
*2do Encuentro Internacional de Estudiantes de Ciencias Económicas,
Administrativas y Contables*



Bogotá, 12, 13 y 14 de septiembre de 2019

6. Bibliografía

Akin, M. (2015). A novel approach to model selection in tourism demand modeling. *Tourism Management*, 48(C):64-72.

Almon, S. (1965). The distributed lag between capital appropriations and expenditures. *Econometrica: Journal of the Econometric Society*, pages 178-196.

Andreou, E., Ghysels, E., and Kourtellos, A. (2010). Regression models with mixed sampling frequencies. *Journal of Econometrics*, 158(2):246-261.

Andreou, E., Ghysels, E., and Kourtellos, A. (2012). Forecasting with mixed-frequency data. In *The Oxford Handbook of Economic Forecasting*. Oxford University Press.

Askitas, N. and Zimmermann, K. F. (2009). Google Econometrics and Unemployment Forecasting. *Applied Economics Quarterly (formerly: Konjunkturpolitik)*, Duncker & Humblot, Berlin, 55(2):107-120.

Athanasopoulos, G., Hyndman, R. J., Song, H., and Wu, D. C. (2011). The tourism forecasting competition. *International Journal of Forecasting*, 27(3):822-844.

Bangwayo-Skeete, P. F. and Skeete, R. W. (2015). Can google data improve the forecasting performance of tourist arrivals? mixed-data sampling approach. *Tourism Management*, 46(C):454-464.

Carriere-Swallow, Y. and Labbe, F. (2013). Nowcasting with Google Trends in an Emerging Market. *Journal of Forecasting*, 32(4):289-298.

Choi, H. and Varian, H. (2012). Predicting the Present with Google Trends. *The Economic Record*, 88(1):2-9.



**6to Simposio Internacional de Investigación en Ciencias Económicas,
Administrativas y Contables - Sociedad y Desarrollo**
*2do Encuentro Internacional de Estudiantes de Ciencias Económicas,
Administrativas y Contables*



Bogotá, 12, 13 y 14 de septiembre de 2019

Claveria, O. and Torra, S. (2014). Forecasting tourism demand to catalonia: Neural networks vs time series models. *Economic Modelling*, 36(C):220-228.

Damuri, F. and Marcucci, J. (2017). The predictive power of google searches in forecasting us unemployment. *International Journal of Forecasting*, 33(4):801-816.

Dickey, D. A. and Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American statistical association*, 74(366a):427-431.

Dukic, V., Lopes, H. F., and Polson, N. G. (2012). Tracking Epidemics With Google Flu Trends Data and a State-Space SEIR Model. *Journal of the American Statistical Association*, 107(500):1410-1426.

Engle, R. F., Ghysels, E., and Sohn, B. (2013). Stock market volatility and macroeconomic fundamentals. *The Review of Economics and Statistics*, 95(3):776-797.

Fantazzini, D. and Fomichev, N. (2014). Forecasting the real price of oil using online search data. *International Journal of Computational Economics and Econometrics*, 4(1/2):4-31.

Fantazzini, D. and Toktamysova, Z. (2015). Forecasting german car sales using google data and multivariate models. *International Journal of Production Economics*, 170:97-135.

Fondeur, Y. and Karam, F. (2013). Can Google data help predict French youth unemployment? *Economic Modelling*, 30(C):117-125.

Ghysels, E., Santa-Clara, P., and Valkanov, R. (2006). Predicting volatility: getting the most out of return data sampled at different frequencies. *Journal of Econometrics*, 131(1):59-95.



**6to Simposio Internacional de Investigación en Ciencias Económicas,
Administrativas y Contables - Sociedad y Desarrollo**
*2do Encuentro Internacional de Estudiantes de Ciencias Económicas,
Administrativas y Contables*



Bogotá, 12, 13 y 14 de septiembre de 2019

Ghysels, E., Sinko, A., and Valkanov, R. (2007). Midas regressions: Further results and new directions. *Econometric Reviews*, 26(1):53-90.

Ghysels, E. and Wright, J. H. (2009). Forecasting Professional Forecasters. *Journal of Business & Economic Statistics*, 27(4):504-516.

Girardin, E. and Joyeux, R. (2013). Macro fundamentals as a source of stock market volatility in china: A garch-midas approach. *Economic Modelling*, 34(Supplement C):59-68.

Goel, S., Hofman, J., Lehaie, S., Pennock, D. M., and Watts, D. J. (2010). Predicting consumer behavior with Web search. *Proceedings of the National Academy of Sciences of the United States of America*, 107(41):4864-90.

Gonzalez, P. and Moral, P. (1995). An analysis of the international tourism demand in Spain. *International Journal of Forecasting*, 11(2):233-251.

Hadavandi, E., Ghanbari, A., Shahanaghi, K., and Abbasian-Naghneh, S. (2011). Tourist arrival forecasting by evolutionary fuzzy systems. *Tourism Management*, 32(5):1196-1203.

Hong, W.-C. (2011). Electric load forecasting by seasonal recurrent SVR (support vector regression) with chaotic artificial bee colony algorithm. *Energy*, 36(9):5568-5578.

Kuzin, V., Marcellino, M., and Schumacher, C. (2011). Midas vs. mixed-frequency var: Nowcasting gdp in the euro area. *International Journal of Forecasting*, 27(2):529-542.

Kwiatkowski, D., Phillips, P. C., Schmidt, P., and Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of Econometrics*, 54(1):159-178.



**6to Simposio Internacional de Investigación en Ciencias Económicas,
Administrativas y Contables - Sociedad y Desarrollo**
*2do Encuentro Internacional de Estudiantes de Ciencias Económicas,
Administrativas y Contables*



Bogotá, 12, 13 y 14 de septiembre de 2019

Park, S., Lee, J., and Song, W. (2017). Short-term forecasting of Japanese tourist inflow to South Korea using Google Trends data. *Journal of Travel & Tourism Marketing*, 34(3):357-368.

Pavlicek, J. and Kristoufek, L. (2015). Nowcasting unemployment rates with Google searches: Evidence from the Visegrad group countries. *PloS one*, 10(5):e0127084.

Peng, B., Song, H., and Crouch, G. I. (2014). A meta-analysis of international tourism demand forecasting and implications for practice. *Tourism Management*, 45(Supplement C):181-193.

Phillips, P. C. B. and Perron, P. (1988). Testing for a unit root in time series regression. *Biometrika*, 75(2):335-346.

Rivera, R. (2016). A dynamic linear model to forecast hotel registrations in Puerto Rico using Google Trends data. *Tourism Management*, 57(Supplement C):12-20.

Siliverstovs, B. A. and Wochner, D. S. (2017). Google Trends and reality: Do the proportions match?: Appraising the informational value of online search behavior: Evidence from Swiss tourism regions. *Journal of Economic Behavior & Organization*.

Song, H., Li, G., Witt, S. F., and Athanasopoulos, G. (2011). Forecasting tourist arrivals using time-varying parameter structural time series models. *International Journal of Forecasting*, 27(3):855-869.

Yu, L., Zhao, Y., Tang, L., and Yang, Z. (2018). Online big data-driven oil consumption forecasting with Google Trends. *International Journal of Forecasting*.