

# **Prediciendo la llegada de turistas a Colombia a partir de los criterios de Google Trends**

---

**Alexander Correa**

Lecturas de Economía - No. 95. Medellín, julio-diciembre 2021



Alexander Correa

## **Prediciendo la llegada de turistas a Colombia a partir de los criterios de Google Trends**

**Resumen:** *Este artículo examina si los criterios de búsqueda de Google Trends son útiles para predecir la llegada mensual de turistas a Colombia. Para este fin, se compara un modelo base que utiliza como predictor los rezagos propios de la llegada de turistas con dos especificaciones alternativas: (i) el modelo base aumentado con la inclusión de datos mensuales de Google Trends; y (ii) el modelo base, pero modificado con la inclusión de datos semanales de Google Trends. Los resultados obtenidos presentan evidencia estadísticamente significativa de que los datos de Google Trends aportan beneficios a la evaluación y predicción de llegadas de turistas a Colombia. En particular, se encuentra que datos de alta frecuencia (semanales) agregan alto valor predictivo en comparación con los modelos que usan datos de la misma frecuencia (mensuales). De este modo, la industria del turismo y los encargados de la política pública de turismo pueden apoyarse de la capacidad predictiva de los datos de Google Trends para mejorar sus procesos de planeación en el corto y mediano plazo.*

**Palabras clave:** *demanda de turismo, Google Trends, proyecciones, mixed data sampling, llegada de turistas.*

**Clasificación JEL:** C53, I83, Z32.

## **Forecasting Tourist Arrivals to Colombia from Google Trends Search Criteria**

**Abstract:** *This study examines whether the Google Trends search criteria are useful in forecasting the monthly arrival of tourists to Colombia. To this end, a baseline model that employs as a predictor the lags values of tourist arrivals is compared with two alternative specifications: (i) the baseline model augmented with monthly data from Google Trends; and (ii) the baseline model but modified with the inclusion of weekly data from Google Trends. The results show statistically significant evidence that Google Trends data provide benefits for the evaluation and prediction of tourist arrivals to Colombia. High-frequency (weekly) data adds high predictive value compared to models that use data of the same frequency (monthly). In this way, the tourism industry and those in charge of tourism public policy can rely on the predictive capacity of Google Trends data to improve their planning processes in the short and medium run.*

**Keywords:** *tourism demand, Google Trends, forecasting, Mixed Data Sampling, tourist arrivals*

<https://doi.org/10.17533/udea.le.n95a343462>



Este artículo y sus anexos se distribuyen por la revista *Lecturas de Economía* bajo los términos de la Licencia Creative Commons Atribución-NoComercial-CompartirIgual 4.0. <https://creativecommons.org/licenses/by-nc-sa/4.0/>

## **Prévision des arrivées touristiques en Colombie selon les critères de Google Trends**

**Résumé:** *Cet article cherche à savoir si les critères de recherche de Google Trends sont utiles pour prévoir les arrivées touristiques mensuelles en Colombie. Pour ce faire, nous proposons un modèle de base qui utilise comme prédicteur les décalages inhérents à l'arrivée des touristes. Ce modèle est ensuite comparé avec deux spécifications alternatives : (i) le modèle de base augmenté par l'inclusion des données mensuelles issues de Google Trends ; et (ii) le modèle de base augmenté par l'inclusion des données hebdomadaires issues également de Google Trends. Nous montrons que les données de Google Trends apportent des avantages statistiquement significatifs à l'évaluation et la prévision des arrivées touristiques en Colombie. Tout particulièrement, les modèles qui utilisent des données à haute fréquence (hebdomadaire), ajoutent une valeur prédictive plus élevée par rapport aux modèles utilisant des données de la même fréquence (mensuelle). Ainsi, l'industrie du tourisme et les responsables de sa politique publique peuvent s'appuyer sur la capacité prédictive des données Google Trends, afin d'améliorer leurs processus de planification à court et moyen terme.*

**Mots clés:** *demande touristique, Google Trends, prévisions dans le tourisme, échantillonnage de données mixtes, arrivées touristes.*

### **Cómo citar / How to cite this item:**

Correa, A. (2021). Prediciendo la llegada de turistas a Colombia a partir de los criterios de Google Trends. *Lecturas de Economía*, 95, 105-134.

<https://doi.org/10.17533/udea.le.n95a343462>

# Prediciendo la llegada de turistas a Colombia a partir de los criterios de Google trends

Alexander Correa <sup>a</sup>

– Introducción. –I. Revisión de literatura. –II. Modelo econométrico. –III. Discusión de resultados. –Conclusiones. –Referencias

*Primera versión recibida el 13 de agosto de 2020; versión final aceptada el 21 de noviembre de 2020*

## Introducción

La globalización ha hecho que viajar sea una actividad cada vez más común. Mientras que en 1985 el 75 % de la población mundial necesitaba una visa para viajar a destinos extranjeros, para 2018 este porcentaje se había reducido a 53 %. De esta mayor frecuencia de viajes, los de turismo representaban el 50 % en el año 2000, y ya para el año 2018 esta participación había subido al 56 % (UNWTO, 2019).

Es ampliamente aceptado que el desarrollo del turismo conduce a mayor crecimiento económico, aunque en la literatura académica se encuentran diferencias empíricas con respecto a los canales de transmisión (Adnan & Ali, 2013; Kadir & Karim, 2012; Katircioglu, 2009). Los gobiernos, en particular los de países en desarrollo, apoyan y promueven el turismo debido a su alto efecto multiplicador en términos de generación de empleo, aumento de los beneficios en tasa de cambio, efecto positivo sobre la balanza de pagos, y estímulo de la oferta en los sectores relacionados con esta actividad (Kessy et al., 2018).

El turismo también impacta el desarrollo de la comunidad a través de la reducción de la pobreza que se da producto del mayor crecimiento económico (Croes & Vanegas, 2008; Croes & Rivera, 2018). De hecho, los numerales 8 y 10 de los Objetivos de Desarrollo Sostenible (ODS) abordan directamente

---

<sup>a</sup> *Alexander Correa*: profesor asociado de la Universidad EAN. Dirección electrónica: [acorrea@universidadean.edu.co](mailto:acorrea@universidadean.edu.co) <https://orcid.org/0000-0002-4025-6612>

el papel del turismo como una fuente de crecimiento económico sostenible (UNWTO & UNDP, 2017).

Para aquellos países receptores, estos flujos de turismo han significado que este sector sea un generador importante de ingresos y empleos, contribuyendo de manera importante al crecimiento económico. Colombia no es ajena a esta realidad y, de acuerdo con el Departamento Nacional de Estadísticas (DANE, 2020), entre 2010 y 2019 el turismo (medido desde la rama de hoteles y restaurantes) creció en promedio al 3,67% anual y contribuyó al crecimiento del PIB total en 11,86%. El Banco Mundial estima que los ingresos del turismo internacional representaron en 2018 el 12,23% de las exportaciones totales (World Bank, 2020).

Por esta razón, el país ha aumentado sus inversiones en la industria turística, incluyendo campañas y estrategias que resaltan el atractivo de Colombia como destino turístico. Para que la industria turística pueda mantener una oferta continua de productos y servicios que satisfagan el creciente flujo de turistas se hace necesario contar con proyecciones sobre la demanda futura de turistas al país.

Un elemento crucial en la proyección de la demanda turistas es la información oportuna. No obstante, en el mejor de los casos, la mayoría de los datos sobre turismo se publican con un rezago de dos meses. De igual forma, dado que la información extrae de los participantes de la industria, el acceso a la misma se vuelve costoso y dispendioso. Tener información en tiempo real sobre la demanda de turistas puede ayudar a mejorar aspectos claves tales como la asignación de presupuesto de mercadeo al permitir estimar el comportamiento futuro de viajeros de mercados entrantes al mercado de destino.

Tanto para la industria turística como para los encargados de la política pública del sector, es crucial tener la información más actualizada posible sobre la demanda de turistas. Esta necesidad de “pronosticar el presente” —o *nowcasting* como se conoce en la literatura económica— ha llevado a incorporar en los modelos estadísticos variables que estén disponibles lo más pronto posible, siendo en su mayoría de casos encuestas cualitativas de tendencias sobre el sector. Sin embargo, debido al sesgo evidente que se

presenta en la medición de variables de este tipo, es necesario buscar fuentes alternativas de información que se produzcan en tiempo real.

Diversos estudios han mostrado que la información producida por motores de búsqueda —en especial la de Google Trends— mejora los pronósticos de modelos tradicionales, al incorporar en los mismos información útil y diversa. En particular, el creciente uso del internet hace que cada vez más personas “revelen” su intención de visitar un destino a través de sus búsquedas en la web. En este sentido, los datos que producen los criterios de búsqueda sobre destinos turísticos podrían estar altamente correlacionados con la llegada de turistas al país y, por lo tanto, ser usados como un indicador en tiempo real de la demanda de turismo por la industria y los encargados de la política pública.

A través de este estudio se indaga si los datos producidos por el criterio de búsqueda “Colombia” en la categoría “travel” en Google Trends permiten mejorar la precisión de pronóstico de la de la demanda de turistas con respecto a un modelo base. En particular, se asume que las personas realicen búsquedas previas sobre sus destinos turísticos en internet mientras planean sus vacaciones y dado que Google Trends presenta información sobre el interés relativo de diferentes términos de búsqueda, esto se puede interpretar como un indicador de viaje hacia el destino turístico.

Para este fin propuesto en este artículo, se analiza el grado en que las series de tiempo generadas en Google Trends mejoran la capacidad de predicción de un modelo simple que utiliza como predictor los rezagos propios de la llegada de turistas en comparación con dos especificaciones alternativas: (i) el modelo base aumentado con la inclusión de datos mensuales de Google Trends); y (ii) el modelo base, pero modificado con la inclusión de datos semanales de Google Trends.

Para los próximos años se espera que el sector del turismo sea el motor de la economía colombiana y —dada la poca literatura sobre proyección de flujos de turismo que existe para el caso de Colombia— este artículo pretende ser una contribución importante al desarrollo de métodos cuantitativos de predicción que puedan ser usados por esta industria y las autoridades encargadas de la política de turismo en el diseño y planeación de programas

que permitan que Colombia se consolide como uno de los principales destinos turísticos de la región.

Este artículo está organizado de la siguiente manera: en esta introducción se discuten el problema y objetivo de investigación; en la primera sección se revisan los estudios más relevantes relacionados con la predicción de la demanda de turistas; la segunda presenta los datos a utilizar, las transformaciones realizadas a los mismos, y la metodología a emplear; la sección tres discute los resultados; finalmente, se presentan las conclusiones más relevantes y las principales implicaciones para la industria turística y para los encargados de la política pública de turismo.

## **I. Revisión de literatura**

Los estudios sobre proyección de flujos de turismo emplean métodos cuantitativos causales y no causales con el propósito de caracterizar y analizar el comportamiento de los turistas, identificar variables relevantes que incidan en la demanda de turistas al igual que patrones estacionales y de tendencia (Guizzardi & Mazzocchi, 2010). Gonzáles y Moral (1995) usan índices de precios de los países de origen y destino para evaluar el poder predictivo que las preferencias de los turistas tienen en la elección de España como destino turístico. Song et al. (2011) estudian los flujos de turismo hacia Hong Kong provenientes de China, Corea del Sur, Reino Unido, y los Estados Unidos usando índices de precios y diferentes de métodos de series de tiempo.

Claveria y Torra (2014) comparan el poder predictivo de diferentes métodos de series de tiempo y encuentran que modelos del tipo ARIMA (Autoregressive Integrated Moving Average) presentan un mejor desempeño que métodos más sofisticados como SETAR (Self Exciting Threshold Autoregressive) y ANN (Artificial Neural Network). Valencia et al. (2017) comparan la capacidad de predecir la llegada de turistas a Medellín (Colombia) de modelos de regresión clásica y modelos bayesianos, encontrando que los segundos presentan una mayor capacidad de predicción que los primeros.

La última década se ha caracterizado por un aumento sin precedentes en la cantidad de información, recursos, servicios, y datos, a los cuales

se puede acceder a través de internet. Este crecimiento en información y datos se denomina como *big data* y ha implicado una transformación social profunda por las diferentes aplicaciones que se le da a este mayor volumen de información. De acuerdo con Khoury e Ioannidis (2014), *big data* es aquella información interconectada, constituida por altos volúmenes de datos que presentan estructuras complejas. Como ejemplos representativos de *big data* es posible mencionar datos de redes sociales, registros de llamadas, datos de sitios web de carácter comercial, información geográfica, datos sobre trayectorias de vehículos, y datos de motores de búsqueda.

Entre la información producida en sitios web se destaca la de Google Trends, herramienta producida por Google, y la cual analiza la popularidad de las consultas de búsqueda realizadas en el motor de búsqueda en diferentes regiones e idiomas. Desde su lanzamiento en 2006, Google Trends ha atraído el interés tanto de la industria como de la academia como instrumento para predecir diferentes fenómenos. Es posible encontrar aplicaciones en epidemiología (Wilson & Brownstein, 2009; Ginsberg et al., 2009), sector automotor (Carriere-Swallow & Labbé, 2013), desempleo (Askitas & Zimmerman, 2009), comportamiento del consumidor (Silva et al., 2019).

En el área de turismo, los datos de Google Trends han sido utilizados por diferentes autores tales como Rivera (2016) quien encuentra que existe una correlación alta entre las búsquedas en Google y el número de visitantes en los hoteles de Puerto Rico. De igual forma, Yang et al (2015) comparan el poder de Google Trends y Baidu en la proyección de visitantes a China y como principal resultado encuentran que Baidu tiene un mejor desempeño que Google Trends ya que Baidu es más popular en China que Google.

Önder (2017) utiliza Google Trends para proyectar la demanda de turismo en dos ciudades (Viena y Barcelona) y dos países (Austria y Bélgica), pero con la novedad que no solo utiliza criterios de búsqueda sino también búsqueda de imágenes de los destinos estudiados. En general, este autor encuentra que la calidad de las proyecciones para Viena aumenta con la inclusión de imágenes y en menor medida para los otros tres destinos.



Finalmente, Bangwayo-Skeete y Skeete (2015) y Havranek y Zeynalov (2021) examinan la utilidad de Google Trends para proyectar los flujos de turismo hacia cinco destinos del Caribe y a la ciudad de Praga (República Checa), respectivamente. Estos autores sugieren que incluir datos semanales de Google Trends mejora la capacidad explicativa y predictora con relación a modelos simples autorregresivos.

## II. Modelo econométrico

### A. Datos

De acuerdo con el portal StatCounter (2020), el motor de búsqueda de Google cuenta con una participación de mercado del 92,71 % a nivel mundial, seguido por Bing (2,73 %) y Yahoo (1,47 %). De hecho, “googlear” se ha vuelto un verbo común cuando las personas se refieren a buscar algo en internet, lo que evidencia la importancia de Google como motor de búsqueda de cualquier tipo información. Dado que el turismo es una industria intensiva en información, Google contiene una cantidad de información relacionada con este sector. De manera específica, la herramienta Google Trends, permite analizar la popularidad que han tenido diferentes criterios de búsqueda a través del tiempo.

Los datos de Google Trends siempre son relativos ya que son índices del volumen de criterio de búsqueda (por ejemplo, “Colombia”) en un área geográfica dada (por ejemplo, “worldwide”) y, de manera opcional, para una categoría de interés (por ejemplo, “travel”). Los índices se construyen tomando el porcentaje de consultas —es decir, el número de consultas para “Colombia” en proporción a todas las demás consultas en el área geográfica especificada en la categoría de interés (si se especificó)— y el volumen diario más alto durante un periodo determinado de tiempo es normalizado a 100 (Choi & Varian, 2012).

Las consultas de búsqueda incluyen descripciones cortas o breves de los términos que están relacionados a las necesidades individuales del individuo tal como “vuelos a Colombia”. Una persona puede usar diversos criterios para obtener la información deseada del motor de búsqueda, no obstante, Google Trends agrupa todos los criterios de búsqueda relacionados con viajes en la

categoría “travel”, incluyendo atracciones, hospedaje, transporte, entre otros (Google Trends, 2020). En ese sentido, al usar Google Trends se pueden obtener datos que capturan el criterio de búsqueda usado en este estudio (“Colombia”) en la categoría “travel”.

Las series sobre llegada mensual de turistas procedentes de diferentes partes del mundo para el período comprendido entre abril del 2011 y junio del 2019 se obtuvieron de los sitios web Centro de Información Turística de Colombia (CITUR) al igual que del sitio web de Migración Colombia. Las llegadas de turistas corresponden a las hechas por vía aérea para tener concordancia con los criterios utilizados en la obtención de las series de Google Trends<sup>1</sup>.

La Figura 1 compara la llegada mensual de turistas con el índice de intensidad de búsqueda mensual de Google Trends mensual. Las dos variables presentan una correlación fuerte con tendencia creciente, lo que indica que tanto el atractivo de Colombia como destino turístico al igual que el interés de las personas de otros países por viajar a Colombia han ido aumentando con el tiempo. Adicionalmente, las series exhiben patrones estacionales, algo común en los datos de turismo. La Figura 2 compara la llegada mensual de turistas con el índice de intensidad de búsqueda semanal de Google Trends mensual exhibiendo los mismos patrones que para los datos mensuales.

Para garantizar relaciones funcionales de carácter lineal entre las variables de estudio, los datos correspondientes a llegada mensual de turistas y los índices de Google Trends son transformados a logaritmos naturales. De igual forma, los datos son ajustados estacionalmente usando una descomposición del tipo STL<sup>2</sup>. De esta manera, las diferentes proyecciones de la llegada

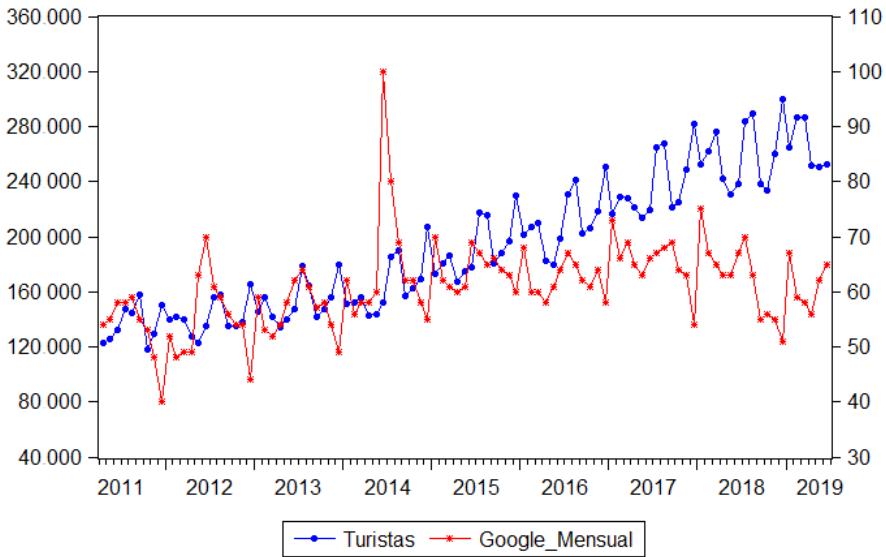
---

<sup>1</sup> De acuerdo a la página de CITUR: “La llegada de extranjeros al país por vía aérea corresponde a corrientes de turistas que se desplazan a nivel mundial con el objetivo de tener nuevas experiencias en diferentes ámbitos de la vida como ocio, convenciones, salud, cultura, entre otras.”

<sup>2</sup> La descomposición STL es un método de ajuste estacional que descompone una serie en sus componentes estacional, de tendencia, y remanente usando un algoritmo que usa regresiones del tipo LOESS. La principal ventaja de STL sobre otros métodos de ajuste estacional es que se puede utilizar en datos de cualquier frecuencia, y que se puede calcular en series de tiempo con patrones irregulares y valores perdidos (Cleveland et al., 1990).

de turistas se pueden ajustar por su componente estacional para poder compararlas con los datos originales de llegada de turistas<sup>3</sup>.

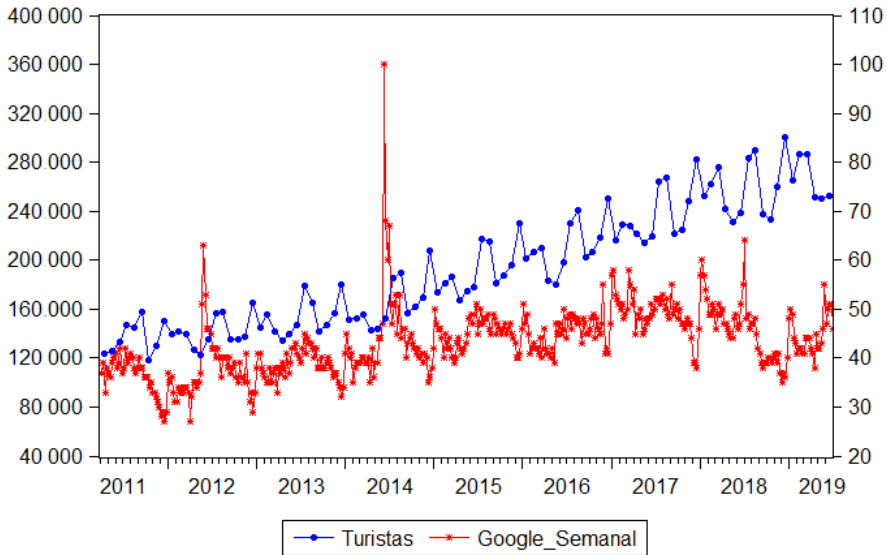
**Figura 1.** Llegada mensual de turistas a Colombia e intensidad de búsqueda mensual del criterio “Colombia” en la categoría “Travel” de Google Trends



Fuente: elaboración propia con datos de Google Trends y CITUR (2020).

<sup>3</sup> Es posible argumentar que utilizar datos estacionalizados produce proyecciones más confiables. Sin embargo, Önder y Gunter (2016) prueban esta hipótesis en la proyección de llegada de turistas a Viena usando Google Trends y encuentran que usar datos estacionalizados en contraste a datos desestacionalizados no produce proyecciones más precisas. En el trabajo de estos autores se muestra que los valores RSME y MAE para los datos ajustados estacionalmente son prácticamente idénticos a los datos originales. De todas formas, siempre es posible volver a la serie original (sin desestacionalizar) incluyendo los factores estacionales en la serie proyectada.

**Figura 2.** Llegada mensual de turistas a Colombia e intensidad de búsqueda semanal del criterio “Colombia” en la categoría “Travel” de Google Trends



*Fuente:* elaboración propia con datos de Google Trends y CITUR (2020).

La Tabla 1 presenta los resultados de tres pruebas diferentes utilizadas para determinar la existencia de raíces unitarias en las diferentes series. Los resultados de las pruebas indican que se rechaza la existencia de una raíz unitaria para el logaritmo natural de la llegada de turistas cuando se incluye constante y tendencia. De igual forma, se rechazar la hipótesis nula de una raíz unitaria para el logaritmo natural de los datos mensuales y semanales de Google Trends.

**Tabla 1. Pruebas de raíz unitaria**

Variable	ADF (Augmented Dickey-Fuller)	PP (Phillips Perron)	KPSS (Kwiatkowski Phillips Schmidt Shin)
Turistas	-4,7222*** (0,0012)	-4,5651*** (0,0020)	0,1814 (0,1460)
Google Mensual	-4,0650*** (0,0097)	-3,9965** (0,0118)	0,2321 (0,1460)
Google Semanal	-5,6597*** (0,0000)	-9,5601*** (0,0000)	0,2940 (0,1460)

*Nota:* para las pruebas ADF y PP la hipótesis nula es que la variable tiene una raíz unitaria y los valores entre paréntesis son los p-valores. En la prueba KPSS la hipótesis nula es que la variable es estacionaria y los valores entre paréntesis corresponden a los valores críticos al 5 %.

*Fuente:* elaboración propia con herramienta Eviews 10.

## ***B. Metodología***

Para determinar la utilidad de Google Trends como predictor de la llegada mensual de turistas a Colombia, se parte de un modelo base el cual se compara con dos modelos alternativos que usan información de Google Trends. En el primer modelo alternativo esta información es mensual, en tanto que en el segundo esta información es semanal.

El modelo base asume que es posible predecir la llegada mensual de turistas a partir de sus propios rezagos. El segundo modelo de comparación establece que es posible explicar la llegada mensual de turistas utilizando no solo rezagos propios sino también los criterios de búsqueda mensual de Google Trends. Este modelo es el modelo base, pero asumiendo que la llegada de turistas depende no solo de su propio pasado (parte autorregresiva) sino de valores actuales y pasados de variables explicativas adicionales (parte de rezagos distribuidos).

La novedad de este trabajo es el tercer modelo, que aumenta el modelo base con datos de Google Trends cuya frecuencia es semanal. Los modelos de regresión estándar requieren que las regresiones tengan la misma frecuencia que la variable dependiente. Esta restricción no siempre se cumple en la

práctica —en economía la mayoría de series estadísticas se producen de manera anual, trimestral, mensual e inclusive diaria—.

Por lo general, existen dos aproximaciones para estimar regresiones donde las variables dependiente e independiente exhiben frecuencias mixtas. La primera aproximación consiste en sumar o promediar los datos de alta frecuencia para que queden en la misma frecuencia que los datos de baja frecuencia. Para este fin, se incluye un coeficiente por cada variable de alta frecuencia para que de esta forma cada valor en el promedio tenga la misma ponderación.

En la segunda aproximación, los componentes individuales de los datos de alta frecuencia son incluidos en la regresión de manera tal que se tenga un coeficiente separado para cada componente de alta frecuencia. Por ejemplo, cuando se estima una regresión anual con regresores de frecuencia mensual, es posible agregar cada uno de los componentes mensuales como regresores. Esta aproximación lleva a que se tenga que estimar un número alto de coeficientes de regresión.

Debido a la pérdida de eficiencia en estimación que exhiben estas dos aproximaciones, Ghysels et al. (2006); Andreou et al. (2010); Ghysels et al. (2020), proponen una alternativa en la cual se permite que la variable dependiente en la regresión tenga una frecuencia menor a las variables explicativas. Esta alternativa es conocida en la literatura de series de tiempo como MIDAS (Mixed Data Sampling) y permite de manera simple y flexible usar datos de alta frecuencia para explicar datos de baja frecuencia.

El método MIDAS es el punto medio de las dos aproximaciones mencionadas anteriormente, ya que permite funciones diferentes de ponderación simultáneamente reduciendo el número de coeficientes a estimar. De esta manera, MIDAS ofrece una aproximación a la estimación de frecuencias mixtas caracterizada por la flexibilidad y una parametrización simple y flexible de la respuesta de la variable dependiente de baja frecuencia a cambios en la variable independiente de alta frecuencia.

Específicamente, el modelo bajo consideración es:

$$Y_t = f\left(\left\{X_{\frac{t}{S}}^H\right\}, \theta, \lambda\right) + \varepsilon_t, \quad (1)$$

donde  $Y_t$  es la variable dependiente de baja frecuencia,  $\left\{ X_{\frac{t}{S}}^H \right\}$  es un regresor (regresores) de alta frecuencia con  $S$  valores para cada valor de baja frecuencia,  $f$  es una función que describe el efecto de la variable de alta frecuencia sobre la variable dependiente de baja frecuencia, y  $\theta, \lambda$  son parámetros para estimar.

El método de estimación MIDAS ofrece diferentes funciones de ponderación que se ubican en un punto medio entre las aproximaciones de coeficientes individuales y la agregación con ponderaciones iguales. El uso de funciones de ponderación en el método MIDAS permite que se reduzca el número de parámetros en el modelo ya que estas funciones fijan restricciones sobre los efectos rezagados de la variable de alta frecuencia. La parametrización de la función de ponderación puede tomar diferentes formas funcionales para mantener la flexibilidad y parsimonia del modelo. No obstante, las especificaciones más utilizadas en trabajo empírico son las ponderaciones Almon y Beta.

La ponderación Almon es una extensión del polinomio de rezagos distribuidos (Almon, 1965) que fija restricciones sobre los coeficientes de los rezagos en modelos autorregresivos, lo que a su vez la convierte en candidato natural para la ponderación de frecuencias mixtas. Por cada rezago de alta frecuencia hasta  $k$ , los coeficientes de la regresión se estiman utilizando rezagos polinomiales de dimensión  $p$  en los parámetros  $\theta$  de la regresión MIDAS. En este caso, el modelo de regresión es:

$$Y_t = \sum_{i=0}^n Y_{t-i} + \sum_{\tau=0}^{k-1} X_{(t-\tau)/S}^H \left( \sum_{j=0}^p \tau^j \theta_j \right) + \varepsilon_t, \quad (2)$$

donde  $p$  es el orden del polinomio Almon, y el número de rezagos seleccionados  $k$  pueden ser menores que o mayores que  $S$ .

Es importante señalar que el número de coeficientes a ser estimados depende del orden del polinomio y no del número de rezagos de alta frecuencia. Esto puede ser visto de manera más clara escribiendo (2) de la siguiente manera:

$$Y_t = \sum_{i=0}^n Y_{t-i} + \sum_{l=0}^p Z_{l,t} \theta_l + \varepsilon_t \quad (2a)$$

$$Z_{l,t} = \sum_{\tau=0}^{k-1} \tau^l X_{(t-\tau)/S}^H \quad (2b)$$

De esta manera, es más fácil ver la asociación de los coeficientes  $\theta_l$  con cada uno de los ordenes  $p$  de la variable  $Z_{l,t}$ .

La ponderación Beta (Ghysels et al., 2006) hace uso de la función beta normalizada. El modelo de regresión correspondiente esta dado por:

$$Y_t = \sum_{i=0}^n Y_{t-i} + \sum_{\tau=0}^{k-1} X_{t-\tau}^H \left( \frac{\omega_\tau^{\theta_1-1} (1-\omega_\tau)^{\theta_2-1}}{\sum_{j=0}^k \omega_j^{\theta_1-1} (1-\omega_j)^{\theta_2-1}} + \theta_3 \right) \lambda + \varepsilon_t, \quad (3)$$

donde  $k$  es un numero de rezagos,  $\lambda$  es un coeficiente de pendiente que es común para todos los rezagos, y

$$\omega_i = \begin{cases} \delta & i = 0 \\ i/(k-1) & i = 1, \dots, k-2 \\ 1-\delta & i = k, \end{cases} \quad (3a)$$

donde  $\delta$  es un número muy pequeño que en la práctica es aproximadamente igual a  $2,22\exp^{-16}$ .

La función beta es extremadamente flexible y puede tomar diferentes formas, las cuales pueden ser creciente, decreciente, plana, jorobada, o en forma de U dependiendo de los tres parámetros AR-MIDAS ( $\theta_1, \theta_2, \theta_3$ ).

### ***C. Modelos***

Para verificar la hipótesis de que los datos de Google Trends mejoran el poder predictivo de la llegada de turistas se plantean los siguientes modelos:



- Un modelo base con valores rezagados de llegada de turistas y una tendencia determinística lineal como las únicas variables explicativas.

$$\ln(llegadas_t) = \alpha + \sum_{i=1}^n \beta_i L^i \ln(llegadas_t) + \vartheta tendencia_t + \epsilon_t. \quad (4a)$$

- Un modelo alternativo I que incluye valores rezagados de la llegada de turistas al igual que valores contemporáneos y rezagados de los criterios de búsqueda Google Trends mensuales más una tendencia determinística lineal (denominado GTM).

$$\begin{aligned} \ln(llegadas_t) = \alpha + \sum_{i=1}^n \beta_i L^i \ln(llegadas_t) \\ + \sum_{i=0}^n \gamma_i L^i \ln(googlet^m) + \vartheta tendencia_t + \epsilon_t. \end{aligned} \quad (4b)$$

- Un modelo alternativo II que incluye valores rezagados de la llegada de turistas y valores contemporáneos y rezagados de los criterios de búsqueda Google Trends semanales más una tendencia determinística lineal (denominado GTS).

$$\begin{aligned} \ln(llegadas_t) = \alpha + \sum_{i=1}^n \beta_i L^i \ln(llegadas_t) \\ + \delta \sum_{k=0}^{w-1} B(k; \theta) L^{k/w} \ln(googlet^w) + \vartheta tendencia_t + \epsilon_t^w. \end{aligned} \quad (4c)$$

La variable dependiente  $\ln(llegadas_t)$  es una variable de baja frecuencia y representa el logaritmo natural desestacionalizado de la llegada mensual de turistas.  $\ln(googlet^m)$  es el logaritmo natural desestacionalizado de la intensidad de búsqueda mensual del criterio “Colombia” en la categoría “Travel” de Google Trends.  $\ln(googlet^w)$  es una variable de alta frecuencia y equivale al logaritmo natural desestacionalizado de la intensidad de búsqueda semanal del criterio

“Colombia” en la categoría “Travel” de Google Trends. La función  $B(k; \theta)$  es una especificación polinomial que determina las ponderaciones para agregación temporal tales como Beta o Almon.  $L^i$  es el operador de rezagos de la variable dependiente  $llegadas_t$ , y  $L^{k/w}$  representa el operador de rezagos para la variable independiente de alta frecuencia  $google_t^w$ .

Dado que la variable dependiente está en logaritmo,  $\beta_i$  representa la elasticidad de los valores rezagados de  $llegadas_t$ ,  $\gamma_i$  es la elasticidad de  $google_t^m$  sobre  $llegadas_t$ ,  $\delta$  es la elasticidad de la variable de alta frecuencia  $google_t^w$  sobre  $llegadas_t$ . Finalmente,  $\alpha$  es una constante,  $\vartheta$  el efecto del tiempo sobre la tasa de crecimiento mensual de  $llegadas_t$ , y  $\epsilon_t$  y  $\epsilon_t^w$  son errores independiente e idénticamente distribuidos  $N(0, \sigma_\epsilon^2)$ .

### III. Discusión de resultados

Los resultados de los modelos base, Google Trends mensual, y dos especificaciones de Google Trends semanal, se presentan en la Tabla 2. Todos los modelos exhiben  $R^2$ s altos con coeficientes individuales significativos. La tendencia desempeña un papel importante en explicar la llegada de turistas para todos los modelos.<sup>4</sup>

El modelo base muestra que aumentos en la llegada de turistas de uno y cinco meses atrás ayuda a explicar estadísticamente la llegada actual de turistas. El modelo alternativo I, el cual es el modelo base aumentando por datos mensuales Google Trends, muestra que cuando el índice de búsquedas aumenta su intensidad en 10 % doce meses después se evidencia un aumento estadísticamente significativo del 10,36 % en la llegada de turistas.

El modelo alternativo II toma el modelo base y lo extiende incluyendo datos semanales de Google Trends. Este modelo se estima utilizando regresiones AR-MIDAS<sup>5</sup> con dos especificaciones diferentes de la ponderación

<sup>4</sup> Las pruebas post-estimación indican errores normalmente distribuidos, libres de correlación serial y heterocedasticidad.

<sup>5</sup> Dado que los modelos autoregresivos por lo general proporcionan proyecciones con mejor desempeño que las obtenidas con modelos estáticos que incluyen variables explicativas, incluir un término autoregresivo en el modelo MIDAS es una extensión deseable (Ghysels & Marcellino, 2016)

semanal y la llegada del mes actual se explica utilizando un rezago de 52 semanas, incluyendo la primera semana del mes a explicar. Los resultados muestran que la inclusión de los datos semanales de Google Trends se justifica dada la significancia estadística del primer rezago (los demás rezagos no se presentan).

**Tabla 2.** Resultados estimación modelos base y alternativos. Abril 2011 – Junio 2019

Variable explicativa	Modelo base	Modelo GTM	Modelo GTS	
			Almon	Beta
$\beta_1$	0,5213*** (0,0016)	0,4289*** (0,0000)	0,3891*** (0,0001)	0,3963*** (0,0001)
$\beta_5$	0,1419* (0,0666)	0,2265** (0,0102)	0,2350*** (0,0081)	0,2262** (0,0131)
$\gamma_{12}$		0,1700*** (0,0005)		
$\delta$			0,0185	0,0718
$\alpha$	3,9500*** (0,0016)	3,3718*** (0,0071)	3,6243*** (0,0035)	3,2605** (0,0040)
$\vartheta$	0,0029*** (0,0015)	0,0025*** (0,0073)	0,0025*** (0,0070)	0,0025*** (0,0071)
$\theta_1$			0,0202*** (0,0006)	-8,4041*** (0,0000)
$\theta_2$			-0,0018*** (0,0008)	-8,4079*** (0,0000)
$\theta_3$			0,0000*** (0,0000)	0,0070 (0,5041)
$R^2$ ajustado	0,9872	0,9877	0,9886	0,9881

*Nota:* \*\*\*, \*\* y \* indican significancia a los niveles de 1 %, 5 % y 10 % respectivamente. P-valores entre paréntesis.  $\beta$ 's muestran los coeficientes rezagados de la llegada mensual de turistas.  $\gamma$ 's denotan los coeficientes de Google Trends mensual en la regresión ARDL.  $\delta$ 's son los coeficientes de Google Trends semanal en las regresiones MIDAS.  $\alpha$  es la constante en las diferentes regresiones.  $\vartheta$  es la tendencia en las diferentes regresiones.  $\theta$ 's son hiperparámetros que determinan la forma de la función de ponderación en las regresiones MIDAS.

*Fuente:* elaboración propia con herramienta Eviews 10.

Estos resultados son consistentes con otros estudios, tales como los de Bangwayo-Skeete y Skeete (2015), Oender (2017), y Havranek y Zeynalov (2019), en los cuales se evidencia que utilizar los datos de Google Trends produce modelos que explican la llegada de turistas con mayor bondad de ajuste y mejor capacidad de predicción que un modelo base de valores rezagados de la variable dependiente.

### ***A. Evaluación proyecciones***

Ya que el interés principal de este artículo es determinar la utilidad de Google Trends para predecir la llegada mensual de turistas, se requiere evaluar cada uno de los tres modelos estimados, no solo en términos de su bondad de ajuste en el periodo muestral sino también en el desempeño de cada uno para predecir por fuera del periodo muestral.

Para determinar qué modelo se desempeña de mejor manera en términos de capacidad predictiva se usan los estadísticos MAPE (Mean Absolute Percentage Error) y RMSE (Root Mean Squared Error). Adicionalmente, un modelo de predicción de promedio simple se usa como contraste. Las proyecciones de cada modelo se evalúan para periodos de 1, 2, 3, 6, y 12 meses. Los resultados de las evaluaciones de predicción se presentan en la Tabla 3 y en la misma se indica la clasificación de cada modelo, donde 1 indica el mejor modelo para predecir por horizonte de tiempo.

Según la información presentada en la Tabla 3, los modelos que incluyen datos semanales de Google Trends exhiben la mejor calidad de predicción para todos los horizontes excepto para un horizonte de 6 meses. También es de destacar que el modelo de pronóstico simple supera al modelo al modelo base que solo usa valores rezagados de la variable dependiente. Vale la pena anotar que en un horizonte de seis meses el modelo que usa datos mensuales de Google Trends es el mejor seguido por el modelo de promedio simple que también exhibe mejor desempeño que los modelos de Google Trends que usan datos semanales.

**Tabla 3. Evaluación proyecciones. Horizonte: 1, 2, 3, 6 y 12 meses**

Modelo	1 mes			2 meses			3 meses			6 meses			12 meses		
	MAPE	C	RMSE	MAPE	C	RMSE	MAPE	C	RMSE	MAPE	C	RMSE	MAPE	C	RMSE
Base	0,3941	5	0,0492	0,2567	5	0,0390	0,2821	5	0,0395	0,1838	3	0,0291	0,2124	5	0,0327
GTM	0,3053	3	0,0381	0,1577	1	0,0263	0,1609	4	0,0250	0,1418	1	0,0231	0,1665	4	0,0281
GTS-Almon	0,2774	2	0,0346	0,1597	2	0,0229	0,1170	2	0,0191	0,1874	4	0,0283	0,1589	3	0,0230
GTS-Beta	0,2756	1	0,0344	0,1636	3	0,0226	0,1137	1	0,0185	0,1993	5	0,0295	0,1542	2	0,0239
Promedio simple	0,3131	4	0,0391	0,1648	4	0,0271	0,1465	3	0,0245	0,1623	2	0,0260	0,1460	1	0,0247

*Nota:* Valores bajos de los estadísticos MAPE y RMSE indican que un modelo en particular ofrece una mejor calidad de predicción que los otros modelos de comparación.

*Fuente:* elaboración propia con herramienta Eviews 10.

Las ponderaciones Almon y Beta exhiben un desempeño bastante consistente (siempre en primer o segundo lugar, a excepción de seis meses), aunque el polinomio Almon es un poco irregular (en términos de MAPE y RMSE relativo al polinomio Beta). Esto, en cierta medida, es esperado ya que el polinomio Almon es sensible a la selección del rezago. Es probable que el número óptimo de rezagos cambie dependiendo del horizonte de tiempo. Sin embargo, en este artículo se utiliza un número fijo de rezagos para todas las estimaciones y por lo tanto no es sorprendente que el desempeño del polinomio Almon cambie.

Estos resultados muestran que los modelos AR-MIDAS producen reducciones significativas en los errores medios al cuadrado para horizontes cortos de tiempo. En particular, el modelo GTS-Beta proporciona valores significativos que reducen los errores de predicción en comparación con los otros modelos. En particular, se puede observar que para todos los horizontes de tiempo (a excepción de seis meses), al menos uno de los modelos que incluyen datos semanales de Google Trends tiene mejor poder de pronóstico que los base, de promedio simple, o que usa datos mensuales de Google Trends. De hecho, se resalta que para un horizonte de tiempo de 12 meses el modelo de promedio simple se ubica según el criterio MAPE, aunque el modelo GTS-Beta quedó de segundo.

Aunque en la práctica, los indicadores tradicionales, tales como MAPE y RMSE, son ampliamente utilizados para evaluar y comparar pronósticos debido a su facilidad y comprensión, los mismos presentan limitaciones en algunos casos. Por un lado, cuando la diferencia estocástica tiene un impacto significativo en los pronósticos producidos por los diferentes modelos, los indicadores tradicionales pueden incluso dar resultados confusos en los casos más desfavorables (Fildes, 2008). Por otra parte, como lo muestran Armstrong y Collopy (1993) estos indicadores tradicionales se ven afectados por la existencia de datos atípicos y no son independientes de la unidad de medida, a menos que estén expresados como porcentaje.

Para solventar estos problemas, se utiliza la prueba Diebold-Mariano (Diebold & Mariano, 1995) para evaluar el desempeño de cada uno de los modelos que incluyen datos de Google Trends en referencia al modelo base.

Para este fin, y teniendo en cuenta que la industria y los encargados de política planifican en horizontes de corto y mediano plazo, se selecciona un periodo de evaluación de dos años. Por lo tanto, para los tres modelos, la muestra de entrenamiento va de julio de 2011 a junio de 2017 y la muestra de proyección de julio de 2017 a junio de 2019.

De manera formal, la prueba Diebold-Mariano (DM) compara la calidad de las predicciones generadas por un modelo alternativo con las de un modelo base utilizando el siguiente estadístico:

$$DM = \frac{\tilde{d}}{s_d}, \quad (5)$$

donde  $\tilde{d}$  y  $s_d$  representan el promedio y la desviación estándar muestral de  $d$ . Donde  $d = \epsilon_1 - \epsilon_2$ .  $\epsilon_i$  puede ser la diferencia absoluta o cuadrada de los valores proyectados y los valores reales de los dos modelos ( $i = 1, 2$ ). Bajo la hipótesis nula de esta prueba las predicciones de los dos modelos tienen la misma exactitud y bajo la hipótesis alternativa las proyecciones del modelo alternativo superan a las del modelo base.

La Tabla 4 presenta los resultados de esta prueba y para cada modelo se rechaza la hipótesis nula. Por lo tanto, es posible concluir que los modelos alternativos I y II son mejores que el modelo base y, en ese sentido, la inclusión de los datos de Google Trends puede ayudar a tener mejores predicciones sobre flujos turísticos hacia Colombia.

Es posible afirmar que índices cuidadosamente identificados sobre intensidad de búsqueda en sitios web, tales como los producidos por Google Trends, incluyen señales tempranas que pueden ayudar de manera considerable a predecir la llegada de turistas a Colombia. Los resultados presentados indican que existe una ventaja informativa en usar datos de alta frecuencia (semanales) para pronosticar una variable de baja frecuencia (mensuales) tal y como lo es la llegada de turistas al país. En ese sentido, para investigaciones futuras, se sugiere examinar si el uso de múltiples variables explicativas de alta frecuencia (diarias o semanales) pueden mejorar aún más los pronósticos de llegada de turistas.

**Tabla 4.** Evaluación proyecciones. Prueba Diebold-Mariano

Modelo base	GTM	GTS	
		Almon	Beta
	2,2581**	2,1489**	2,2187**
	(0,0169)	(0,0212)	(0,0183)

*Nota:* Un valor positivo y significativo de la prueba DM significa que las proyecciones producidas por los modelos alternativos 1 y 2 (ARDL y AR-MIDAS) son estadísticamente superiores a las producidas por el modelo base (AR). En contraste, un valor no significativo implica que las proyecciones del modelo alternativo son estadísticamente iguales a las del modelo base. El estadístico utilizado es la diferencia cuadrada,  $\epsilon_i = (\hat{y}_i - y_i)^2$ .

*Fuente:* elaboración propia con herramienta Eviews 10.

## Conclusiones

Contar con proyecciones adecuadas sobre la demanda de turistas es importante no solo para la industria turística, sino también para la economía como un todo, dada la contribución que este sector hace al crecimiento económico de países cuyo principal motor es el turismo. Este tema cobra más valor en un entorno global donde la incertidumbre es la regla, lo que a su vez modifica el comportamiento de turistas internacionales haciendo más difícil los procesos de planeación tanto de la industria turística como de las autoridades encargadas del diseño de la política de turismo. A esto también se suma que las proyecciones en la industria del turismo por lo general se realizan usando datos provenientes de fuentes secundarias cuya disponibilidad pública se da con meses de retraso.

Los motores de búsqueda son una fuente importante para obtener datos sobre turismo y evaluar dinámicas de su consumo. De hecho, búsquedas en la categoría “Travel” de Google sobre países destino pueden indicar llegadas futuras o potenciales de turistas. Esto ha motivado que en este estudio se determine el impacto que los datos de Google Trends sobre intensidad de búsqueda del criterio “Colombia” tienen sobre la llegada de turistas al país.

Los resultados evidencian que la industria del turismo y las autoridades encargadas de la política pública de turismo se pueden beneficiar en utilizar los



datos de Google Trends —como complemento a la información tradicional— para predecir la llegada de turistas en horizontes de un año y mejorar de esta manera sus procesos de planeación. Simultáneamente, otras industrias también pueden utilizar esta metodología basándose en sus propios criterios de búsqueda para predecir tendencias útiles.

No obstante, es importante anotar que usar Google Trends como la única variable para predecir no está exenta de sesgos, puesto que la motivación para buscar información en motores de búsqueda depende de factores internos y externos difíciles de determinar. De igual forma, tampoco se puede asumir que todos los que buscan información sobre un destino tienen la intención de visitar ese lugar; lo que realmente indican estas búsquedas es un interés de las personas. Por lo tanto, es importante contar con información adicional que complemente los pronósticos en tiempo real que se producen utilizando Google Trends.

Por ejemplo, el criterio de búsqueda “Colombia” en la categoría “Travel” se puede mejorar si se incluyen las búsquedas en el idioma nativo del turista (por ejemplo, mandarín o ruso); también si la información de Google Trends no se limita a utilizar términos de búsqueda, sino que se expande para incluir imágenes sobre los destinos turísticos de interés. Así, se pueden asignar recursos de manera mucho más eficiente y reducir costos.

Poder reaccionar de manera adecuada y rápida a picos previsibles de corto plazo de llegada de turistas de ciertos mercados permitiría —por ejemplo— que haya suficientes folletos en chino, un número adecuado de guías turísticos que hablen italiano, o los buses necesarios para el traslado de turistas del aeropuerto a los hoteles respectivos. De igual forma, a diferencia de muchos de los datos que tradicionalmente se utilizan para pronosticar la llegada de turistas, los datos de Google Trends están disponibles de manera gratuita y oportuna (semanales) y con el uso de los modelos adecuados, pueden ser fácilmente incorporados por la industria.

Es importante señalar que, incluso si los datos de Google Trends indican intención de visitar, puede existir un sesgo de selección con Google Trends ya que esta herramienta no recoge información de otros grupos demográficos que no usan esta herramienta y que pueden ser bastante heterogéneos en

términos de edad, ingreso, y nivel educativo. Adicionalmente, en el mercado también existen un variado número de aplicaciones que ofrecen información y recomiendan diferentes sitios turísticos que eventualmente pueden limitar el uso de motores de búsqueda para buscar información sobre destinos turísticos. En ese sentido, se podrían utilizar algoritmos de machine learning y data mining que permitiesen recopilar información demográfica de los potenciales turistas para complementar los datos de Google Trends.

### Referencias

- Adnan Hye, Q. M., & Ali Khan, R. E. (2013). Tourism-Led Growth Hypothesis: A Case Study of Pakistan. *Asia Pacific Journal of Tourism Research*, 18(4), 303-313. <https://doi.org/10.1080/10941665.2012.658412>
- Almon, S. (1965). The Distributed Lag Between Capital Appropriations and Expenditures. *Econometrica*, 33(1), 178-196. <https://doi.org/10.2307/1911894>
- Andreou, E., Ghysels, E., & Kourtellos, A. (2010). Regression models with mixed sampling frequencies. *Journal of Econometrics*, 158(2), 246-261. <https://doi.org/10.1016/j.jeconom.2010.01.004>
- Armstrong, J. S., & Collopy, F. (1993). Error measures for generalizing about forecasting methods: Empirical comparisons. *Long Range Planning*, 26(1), 150. [https://doi.org/10.1016/0024-6301\(93\)90280-s](https://doi.org/10.1016/0024-6301(93)90280-s)
- Askitas, N., & Zimmermann, K. F. (2009). Google Econometrics and Unemployment Forecasting. *Applied Economics Quarterly*, 55(2), 107-120. <https://doi.org/10.3790/aeq.55.2.107>
- Bangwayo-Skeete, P. F., & Skeete, R. W. (2015). Can Google data improve the forecasting performance of tourist arrivals? Mixed-data sampling approach. *Tourism Management*, 46, 454-464. <https://doi.org/10.1016/j.tourman.2014.07.014>

Correa: Prediciendo la llegada de turistas a Colombia a partir de los criterios de Google Trends

- Carrière-Swallow, Y., & Labbé, F. (2013). Nowcasting with Google Trends in an Emerging Market. *Journal of Forecasting*, 32(4), 289-298. <https://doi.org/10.1002/for.1252>
- Centro de Información Turística de Colombia (CITUR). (2020). Flujos turísticos. <http://www.citur.gov.co/estadisticas#gsc.tab=0>
- Choi, H., & Varian, H. (2012). Predicting the Present with Google Trends. *Economic Record*, 88(1), 2-9. <https://doi.org/10.1111/j.1475-4932.2012.00809.x>
- Claveria, O., & Torra, S. (2014). Forecasting tourism demand to Catalonia: Neural networks vs. time series models. *Economic Modelling*, 36, 220-228. <https://doi.org/10.1016/j.econmod.2013.09.024>
- Cleveland, R., Cleveland, W., McRae, J., & Terpening, I. (1990). STL: A Seasonal- Trend Decomposition Procedure Based on Loess. *Journal of Official Statistics*, 6(1), 3-73. <https://www.wessa.net/download/stl.pdf>
- Croes, R., & Rivera, M. (2018). Tourism and poverty alleviation: a reply to Sharpley. *Journal of Policy Research in Tourism, Leisure and Events*, 11(1), 170-178. <https://doi.org/10.1080/19407963.2018.1482702>
- Croes, R., & Vanegas, M. (2008). Cointegration and Causality between Tourism and Poverty Reduction. *Journal of Travel Research*, 47(1), 94-103. <https://doi.org/10.1177/0047287507312429>
- Departamento Nacional de Estadísticas (DANE). (2020). Estadísticas por tema. <https://www.dane.gov.co/index.php/estadisticas-por-tema>
- Diebold, F. X., & Mariano, R. S. (1995). Comparing Predictive Accuracy. *Journal of Business & Economic Statistics*, 13(3), 253-263. <https://doi.org/10.2307/1392185>
- Fildes, R. (2008). Forecasting. *Strategic Direction*, 24(5). <https://doi.org/10.1108/sd.2008.05624ead.005>
- Ghysels, E., Kvedaras, V., & Zemlys-Balevičius, V. (2020). Chapter 4. Mixed data sampling (MIDAS) regression models. En Hrishikesh D., &

- Vinod, C. R. Rao (eds.), *Handbook of Statistics* (Vol. 42, pp. 117-153). <https://doi.org/10.1016/bs.host.2019.01.005>
- Ghysels, E., & Marcellino, M. (2016). The econometric analysis of mixed frequency data sampling. *Journal of Econometrics*, 193(2), 291-293. <https://doi.org/10.1016/j.jeconom.2016.04.007>
- Ghysels, E., Santa-Clara, P., & Valkanov, R. (2006). Predicting volatility: getting the most out of return data sampled at different frequencies. *Journal of Econometrics*, 131(1-2), 59-95. <https://doi.org/10.1016/j.jeconom.2005.01.004>
- Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S., & Brilliant, L. (2009). Detecting influenza epidemics using search engine query data. *Nature*, 457(7232), 1012-1014. <https://doi.org/10.1038/nature07634>
- González, P., & Moral, P. (1995). An analysis of the international tourism demand in Spain. *International Journal of Forecasting*, 11(2), 233-251. [https://doi.org/10.1016/0169-2070\(94\)00570-3](https://doi.org/10.1016/0169-2070(94)00570-3)
- Google Trends. (2020). Trends Help. <https://support.google.com/trends/?hl=en-GB#topic=6248052>
- Guizzardi, A., & Mazzocchi, M. (2010). Tourism demand for Italy and the business cycle. *Tourism Management*, 31(3), 367-377. <https://doi.org/10.1016/j.tourman.2009.03.017>
- Havranek, T., & Zeynalov, A. (2021). Forecasting tourist arrivals: Google Trends meets mixed-frequency data. *Tourism Economics*, 27(1). <https://doi.org/10.1177/1354816619879584>
- Kadir, N., & Karim, M. Z. A. (2012). Tourism and Economic Growth in Malaysia: Evidence from Tourist Arrivals from Asean-S Countries. *Economic Research-Ekonomska Istraživanja*, 25(4), 1089-1100. <https://doi.org/10.1080/1331677X.2012.11517550>
- Katircioglu, S. T. (2009). Revisiting the tourism-led-growth hypothesis for Turkey using the bounds test and Johansen approach for cointegration.

- Tourism Management*, 30(1), 17-20. <https://doi.org/10.1016/j.tourman.2008.04.004>
- Kessy, D., Kiage, O., & Kiprutto, N. (2018). Multiplier effects of tourism in selected areas of Arusha, Tanzania. *African Journal of Hospitality, Tourism and Leisure*, 7(3), 1-17. [https://www.ajhtl.com/uploads/7/1/6/3/7163688/article\\_17\\_vol\\_7\\_\\_3\\_\\_2018.pdf](https://www.ajhtl.com/uploads/7/1/6/3/7163688/article_17_vol_7__3__2018.pdf)
- Khoury, M. J., & Ioannidis, J. P. A. (2014). Big data meets public health. *Science*, 346(6213), 1054-1055. <https://doi.org/10.1126/science.aaa2709>
- Önder, I. (2017). Forecasting tourism demand with Google trends: Accuracy comparison of countries versus cities. *International Journal of Tourism Research*, 19(6), 648-660. <https://doi.org/10.1002/jtr.2137>
- Önder, I., & Gunter, U. (2016). Forecasting Tourism Demand with Google Trends For a Major European City Destination. *Tourism Analysis*, 21(2), 203-220. <https://doi.org/10.3727/108354216X14559233984773>
- Rivera, R. (2016). A dynamic linear model to forecast hotel registrations in Puerto Rico using Google Trends data. *Tourism Management*, 57, 12-20. <https://doi.org/10.1016/j.tourman.2016.04.008>
- Silva, E., Hassani, H., Madsen, D., & Gee, L. (2019). Googling Fashion: Forecasting Fashion Consumer Behaviour Using Google Trends. *Social Sciences*, 8(4), 111. <https://doi.org/10.3390/socsci8040111>
- Song, H., Li, G., Witt, S. F., & Athanasopoulos, G. (2011). Forecasting tourist arrivals using time-varying parameter structural time series models. *International Journal of Forecasting*, 27(3), 855-869. <https://doi.org/10.1016/j.ijforecast.2010.06.001>
- StatCounter. (2020). Search Engine Market Share Worldwide. <https://gs.statcounter.com/search-engine-market-share>
- World Tourism Organization (UNWTO). (2019). International Tourism Highlights. <https://www.e-unwto.org/doi/pdf/10.18111/9789284421152>

- Valencia, M., Vanegas, J., Correa, J., & Restrepo, J. (2017). Comparación de pronósticos para la dinámica del turismo en Medellín, Colombia. *Lecturas de Economía*, 86, 199-230. <https://doi.org/10.17533/udea.le.n86a08>
- Wilson, K., & Brownstein, J. S. (2009). Early detection of disease outbreaks using the Internet. *Canadian Medical Association Journal*, 180(8), 829-831. <https://doi.org/10.1503/cmaj.1090215>
- World Bank. (2020). International Tourism, Receipts (% of total exports) - Colombia. <https://data.worldbank.org/indicator/ST.INT.RCPT.XP.ZS?locations=CO>
- World Tourism Organization (UNWTO) & United Nations Development Programme (UNDP). (2017), *Tourism and the Sustainable Development Goals - Journey to 2030, Highlights*. UNWTO. <https://doi.org/10.18111/9789284419340>
- Yang, X., Pan, B., Evans, J. A., & Lv, B. (2015). Forecasting Chinese tourist volume with search engine data. *Tourism Management*, 46, 386-397. <https://doi.org/10.1016/j.tourman.2014.07.019>