



Diciembre 2019 - ISSN: 1696-8352

EVALUACIÓN DE LA PERFORMANCE DEL *GOOGLE TREND INDEX* PARA ESTIMAR LAS VENTAS DE AUTOMÓVILES EN ARGENTINA. UN CASO PRÁCTICO.

Jonathan Alejandro Collison¹

Analista de Proyectos - Accenture (Buenos Aires)

E-mail: jcollison@live.com.ar

Para citar este artículo puede utilizar el siguiente formato:

Jonathan Alejandro Collison (2019): "Evaluación de la performance del google trend index para estimar las ventas de automóviles en Argentina. Un caso práctico", Revista Observatorio de la Economía Latinoamericana (diciembre 2019). En línea:

<https://www.eumed.net/rev/oel/2019/12/ventas-automoviles-argentina.html>

Resumen:

En los últimos años, la literatura económica ha hecho uso de las nuevas tecnologías para intentar mejorar los *forecasts* de diversas variables de interés. Con la masividad de *Big Data* y la enorme cantidad de información disponible sobre hábitos de consumo, muchos modelos intentan explotar estas herramientas para estimar ventas. En este contexto, las búsquedas en *Google* pueden proporcionar valiosa información sobre el interés de los consumidores sobre un producto en particular.

El propósito de este artículo es evaluar si las búsquedas en Google pueden contribuir a una mejor estimación de las ventas de automóviles marca Ford en Argentina. Para lograr este objetivo, construye un modelo incorporando el *Google Trends Index* para capturar la frecuencia de búsqueda en internet y se lo compara con un modelo sin dicha variable explicativa.

Palabras clave: estimación, modelado, automóvil, Google, búsqueda, ventas

Abstract:

Over the last few years, economic literature has used technology to try to improve forecasting. With Big Data and the vast amount of information regarding consumer habits, many sales forecasting models tried to exploit this feature. In this context, Google searches can provide valuable information regarding the consumer preferences for a product.

¹ Candidato a Máster en Economía (Universidad de Buenos Aires). Analista de Proyectos en Accenture, Argentina.

The purpose of this paper is to assess the contribution of Google searches in forecasting Ford car sales in Argentina. For this objective, we construct a model incorporating the Google Trend Index to capture search frequency in internet, and we compare this model with a standard model without said Index as an explanatory variable, and asses the difference in performance between the two.

Keywords: forecasting, modeling, car, google, search, sales

1. INTRODUCCIÓN

Recientemente, la cantidad de datos de consumidores que se capturan y almacenan ha crecido exponencialmente. Con el auge de Big Data, las empresas e investigadores poseen una herramienta poderosa para analizar diferentes fenómenos.

Una de las principales aplicaciones de Big Data, es el de analizar las tendencias en las búsquedas por internet. Con relación a ello, diversos autores han abordado la investigación empírica de búsquedas online para estimar diversas variables, desde desempleo (Pavlicek y Kristoufek, 2014) hasta epidemias (Ginsberg et al., 2009), (Young y Zhang, 2018).

Google Trends es un sitio web público que pertenece a Google Inc. y ofrece datos de búsquedas en Google, mostrando la frecuencia de un término en particular en comparación con todos los otros términos, en diferentes regiones y lenguajes (Google, 2017). *Google Trends* fue lanzado en 2006, y en 2016 las búsquedas anuales en *Google Trends* alcanzaron 3 trillones (Sullivan, 2016).

En lo que respecta a los trabajos sobre ventas de automóviles, (Graevenitz, G. et al, 2016) plantea un análisis de datos de panel para estudiar la demanda para el mercado automotriz de Alemania y el Reino Unido, utilizando *Google Trends*. Asimismo, en (Choi y Varian 2012) también se utiliza *Google Trends* para estimar las ventas de diversos bienes, entre ellos los automóviles. Dicho trabajo se ocupa de “predecir el presente”, es decir, estimaciones a muy corto plazo (mensuales). Un trabajo similar que también realiza estimaciones mensuales y utiliza *Google Trends* para analizar el mercado automotriz es (Tomczyk, E yT. Doligalski (2015). Los autores utilizan el término “*newcasting*” para definir la modalidad de estimar mes a mes la variable de interés. Este enfoque se utilizará en el presente trabajo.

2. METODOLOGÍA

Google Trends publica un índice de búsquedas desagregado por producto y país en un período determinado de tiempo. Los datos publicados no están expresados en valores absolutos, sino que el índice se publica expresado de 0 a 100, donde el punto 100 representa la máxima cantidad de búsquedas de ese producto en esa región durante el período de tiempo especificado.

Los reportes están disponibles públicamente en el sitio web: <http://www.google.com/insights/search> y los datos se pueden bajar en un archivo con formato .csv.

El período de análisis seleccionado es de febrero 2011 hasta diciembre 2018 para la estimación de coeficientes del modelo, y el período enero-junio 2019 para testear la capacidad predictiva (*out of sample prediction*).

El gráfico a continuación muestra el índice publicado por *Google Trends*, para las búsquedas en Argentina de la palabra clave “Ford” durante el período enero 2011 a diciembre 2018. Como se puede apreciar, el

mes de mayor cantidad de búsquedas realizadas es febrero de 2016 (equivalente a un *Google Index* de 100).

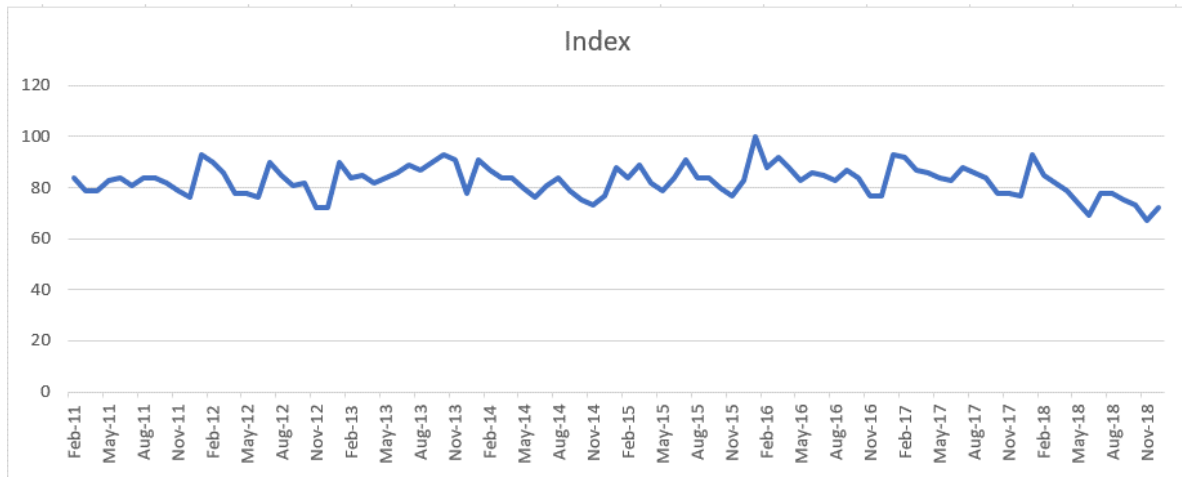


Figura 1: Google Index para la búsqueda Ford Argentina.

Fuente: Elaboración propia

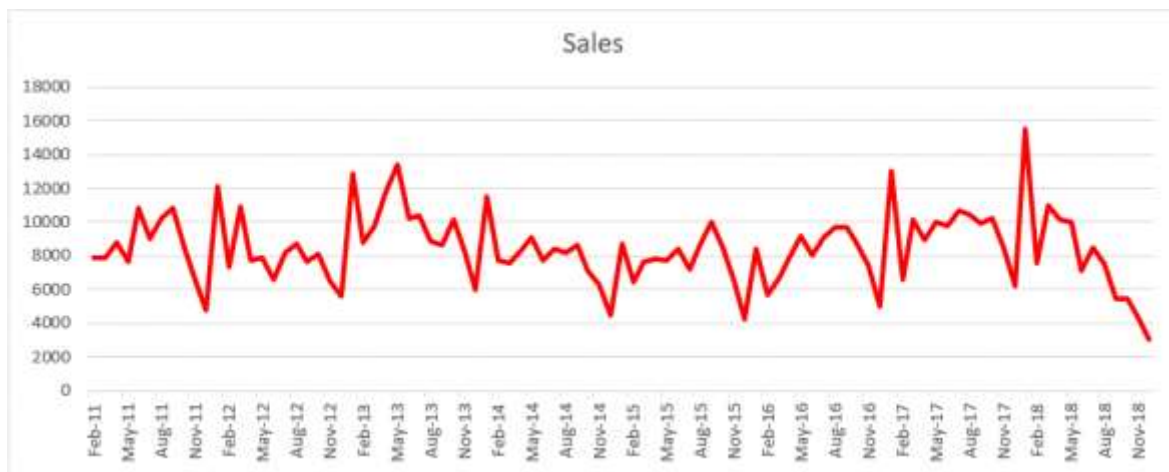


Figura 2: Google Index para la búsqueda Ford Argentina durante el período seleccionado

Fuente: Elaboración propia

3. CONSTRUCCIÓN Y COMPARACIÓN DE MODELOS

Al examinar ambos gráficos, se observa que tanto las ventas como el *Google Index* presentan estacionalidad. Es por ello que el primer modelo se plantea de la siguiente manera:

$$\text{Modelo 1: } \log(\text{Ventas}_t) = \beta_0 + \beta_1 * \log(\text{Ventas}_{t-1}) + \beta_2 * \log(\text{Ventas}_{t-12}) + e_t$$

Como es usual, la variable e es un término de error.

Desde el punto de vista de la estructura, el modelo aquí presentado es un AR (1) con estacionalidad anual. Con el objetivo de incorporar la información de búsquedas online, en el segundo modelo se incorporó el *Google Trend Index* como término regresor:

$$\text{Modelo 2: } \log(\text{Ventas}_t) = \beta_0 + \beta_1 * \log(\text{Ventas}_{t-1}) + \beta_2 * \log(\text{Ventas}_{t-12}) + X_t + e_t$$

Donde X_t es el *Google Trend Index*.

Al realizar una estimación de mínimos cuadrados para ambas ecuaciones, se obtienen los siguientes resultados:

Modelo 1:

$$\log(\text{Ventas}_t) = 0,09 + 0,34 * \log(\text{Ventas}_{t-1}) + 0,54 \log(\text{Ventas}_{t-12}) + e_t$$

Coefficients:	Estimate	Std Error	t value	Pr(> t)
(Intercept)	1,1038	1,3104	0,842	0,401806
log(Sales.Lag1)	0,3371	0,0947	3,559	0,000591 ***
log(Sales.Lag12)	0,5409	0,0952	5,682	1,55E-07 ***

Signif. codes: 0 '***' 0,001 '**' 0,01 '*' 0,05 '.' 0,1 ' ' 1

Residual standard error: 0,2256 on 92 degrees of freedom

Multiple R-squared: 0,2983, **Adjusted R-squared:** 0,283

F-statistic: 19,55 on 2 and 92 DF, **p-value:** 8,384e-08

Modelo 2:

$$\log(\text{Ventas}_t) = 1,1 + 0,34 * \log(\text{Ventas}_{t-1}) + 0,46 * \log(\text{Ventas}_{t-12}) + 0,02 * X_t + e_t$$

Coefficients:	Estimate	Std Error	t value	Pr(> t)
(Intercept)	0,08744	1,10868	0,079	0,937
log(Sales.Lag1)	0,33891	0,07928	4,275	4,71E-05 ***
log(Sales.Lag12)	0,46046	0,0807	5,706	1,43E-07 ***
Google Trend Index	0,02081	0,00328	6,345	8,49E-09 ***

Signif. codes: 0 '***' 0,001 '**' 0,01 '*' 0,05 '.' 0,1 ' ' 1

Residual standard error: 0,1889 on 91 degrees of freedom

Multiple R-squared: 0,5135, Adjusted R-squared: 0,4975

F-statistic: 32,02 on 3 and 91 DF, p-value: 3,206e-14

Como se evidencia en las salidas de R anteriores, ambos modelos muestran resultados consistentes y coeficientes similares, los cuales tienen significatividad estadística al 99% de confianza.

Es fundamental aclarar que este trabajo no pretende obtener la mejor estimación posible para las ventas mensuales de vehículos marca Ford, sino que el objetivo es destacar la mejora en la estimación que resulta de incorporar el *Google Trend Index* como regresor en la especificación del modelo.

Para comprobar esta hipótesis se realizaron una serie de estimaciones mensuales (out of sample forecast) utilizando ambos modelos y se los comparó computando los errores de predicción. Para realizar las estimaciones mensuales se hace una regresión para cada mes, agregando el valor real observado del mes anterior. Para obtener el *Google Trend Index* de cada nuevo mes, se utiliza el índice que toma en cuenta la primera semana de dicho mes.

Como se mencionó anteriormente, el período comprendido entre febrero de 2011 hasta diciembre de 2018 se utilizó para la estimación de coeficientes del modelo, y el período enero-junio 2019 se utilizó para testear la capacidad predictiva de los dos modelos propuestos.

Para calcular los errores de estimación se utiliza el error medio absoluto (Choi y Varian 2012):

$$EMA = \frac{1}{T} \sum_{t=1}^T |EP_t|$$

Donde se define EP (Error de la predicción):

$$EP = \log(\text{Ventas estimadas}_t) - \log(\text{Ventas Reales}_t)$$

El gráfico y la tabla a continuación resumen los resultados:

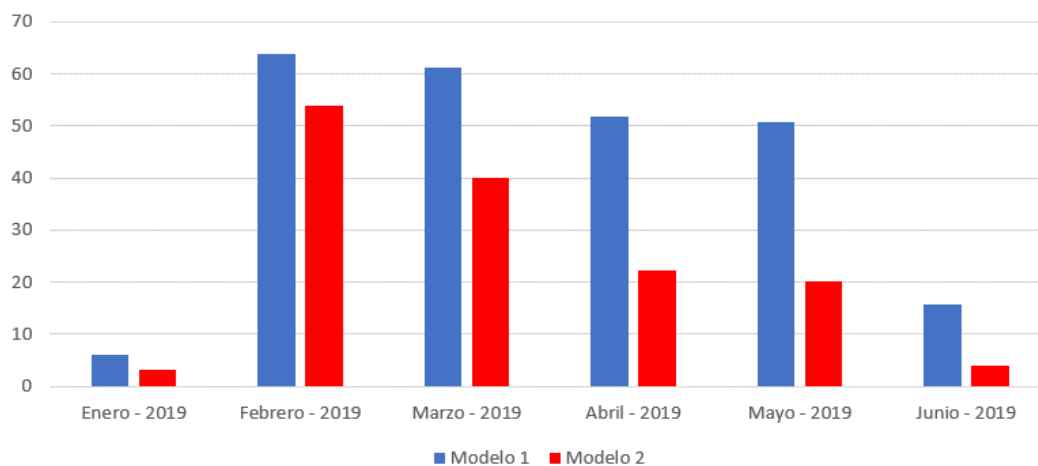


Figura 3: Error en la estimación mensual

Fuente: Elaboración propia

	Promedio del Error de Estimación
Modelo 1:	41,5%
Modelo 2:	23,9%

Figura 4: Promedio del Error de Estimación para todo el período

Fuente: Elaboración propia

Nótese que el modelo que incluye el *Google Trend Index* posee un error absoluto menor tanto en el promedio de todo el período de estimación como en cada mes en particular. Ambos modelos tienden a sobreestimar las ventas, pero en el modelo 2 la sobreestimación es significativamente menor. El *Google Trend Index* captura el hecho de que la demanda por automóviles Ford ha declinado durante el período de estimación.

Sin duda alguna que la contracción de la economía argentina en general (INDEC 2019), y la del mercado automotriz en particular, contribuyen a provocar un aumento del error de estimación. Una posible respuesta sería controlar por dicho fenómeno y tratarlo como un cambio estructural, pero ello excede el propósito de este trabajo.

Correlación Serial - Durbin-Watson test

A menudo cuando se analizan series de tiempo se observa correlación serial entre los errores de la estimación (Graevenitz, G. et al, 2016). En este trabajo, se utilizó el test de Durbin-Watson (Diebold, 1997) para evaluar los errores. La salida de R para el test es la siguiente:

DW = 1,7111, p-value = 0,06198

Siguiendo a (Diebold,1997), se toma 1,5 como valor de corte para determinar la existencia de autocorrelación. Como se observa, el estadístico es de 1,7111, por lo que se concluye que no existe autocorrelación serial entre los errores de la estimación.

4. CONCLUSIONES

A través del análisis se observa que el modelo que incorpora al *Google Trend Index* tiene una mejor performance al estimar las ventas de automóviles Ford en Argentina durante el período bajo estudio. Si bien el análisis se realizó tomando en cuenta un producto en particular, la contribución de *Google Trend Index* en la estimación de ventas puede ser significativa para diferentes bienes. Queda como una futura línea de investigación el análisis de diferentes productos y mercados, pero la aplicación de los motores de búsqueda para ayudar a predecir el comportamiento de diversas variables seguramente se incrementará con el correr del tiempo.

Bibliografía

Choi, H. y H. Varian (2012): “Predicting the Present with Google Trends,” *Economic Record*, 88, 2–9.

De Los Santos, B., A. Hortac, Su, y M. R. Wildenbeest (2012): “Testing Models of Consumer Search Using Data on Web Browsing and Purchasing Behavior,” *The American Economic Review*, 102, 2955–2980.

Diebold, F. (1997): “Elements of Forecasting”, South-Western College Publishing, p. 27-28.

Ginsberg, J., M. H. Mohebbi, R. S. Patel, L. Brammer, M. S. Smolinsky, y L. Brilliant (2009): “Detecting Influenza Epidemics Using Search Engine Query Data,” *Nature*, 457, 1012–1014.

Google (2017): “Google Trends Help”, Google, Inc.

Graevenitz, G. v., C. Helmers, Millot, V. y Turnbull O. (2016): “Does Online Search Predict Sales? Evidence from Big Data for Car Markets in Germany and the UK”, *CCP Working Paper* 16-7, University of East Anglia, 2016.

Granger, C. W. y P. Newbold (1974): “Spurious Regressions in Econometrics,” *Journal of Econometrics*, 2, 111–120.

INDEC (2019): Estimador mensual de actividad económica – Junio de 2019 (indec.gob.ar)

Jun, S., H. Yoo, y S. Choi (2017): “Ten years of research change using Google Trends: From the perspective of big data utilizations and applications”, *Technological Forecasting and Social Change* Volume 130, May 2018, Pages 69-87

Moraga-González, J. L., Z. Sandor, y M. R. Wildenbeest (2015): “Consumer Search and Prices in the Automobile Market,” *CEPR DP* 10487.

Pavlicek, J. y L. Kristoufek (2014): “Can Google Searches Help Nowcast and Forecast Unemployment Rates in the Visegrad Group Countries?” *arXiv preprint* arXiv:1408.6639.

Preis, T., H. S. Moat, y H. E. Stanley (2013): “Quantifying Trading Behavior in Financial Markets

Using Google Trends,” *Scientific Reports*, 3.

Sullivan, D. (2016): “Google now handles at least 2 trillion searches per year”, *Search Engine Land*, Danny Sullivan (2016)

Tomczyk, E yT. Doligalski (2015): “Predicting New Car Registrations: Nowcasting with Google Search and Macroeconomic Data”, *Sł. Partycki (ed.), E-społeczeństwo w Europie Środkowej i Wschodniej. Terażniejszość i perspektywy rozwoju (e-Society in Middle and Eastern Europe. Present and Development Perspectives)*, Wydawnictwo KUL, Lublin 2015, p. 228-236.

Wooldridge, J. (2012): *Introductory Econometrics: A Modern Approach*, *Cengage Learning*.

Young SD, Zhang Q (2018): Using search engine big data for predicting new HIV diagnoses. *PLoS ONE* 13(7)