



The Google trends effect on the behavior of the exchange rate Mexican peso - US dollar

Efecto de las tendencias de Google en el comportamiento del tipo de cambio peso mexicano- dólar estadounidense

Mario Durán Bustamante*, Adrián Hernández del Valle,
Ambrosio Ortiz Ramírez

Instituto Politécnico Nacional, México

Received October 3, 2017; accepted February 20, 2018

Available online November 5, 2018

Abstract

We show the advantage of using Google search engine trends to forecast the volatility of the short-term (weekly) exchange rate between the Mexican peso and United States dollar. We perform a comparison of models in the literature that have used Google Trends to examine explanatory variables. Some of the models are based on time series, whereas others are based on the similarity function, which captures the cognitive form of human reasoning. For example, an investor who needs to know the value that a variable will take in the future will take into account relevant, known, and available information, and weigh it to calculate the forecast. We conclude that taking into account the Google Trends variable helps explain partially the behaviour of volatility; and it is necessary to incorporate more aggregation levels. Moreover, to the best of our knowledge, literature on the subject of using Google Trends to explain relevant economic variables is relatively scarce.

JEL Codes: C01, C13, F31.

Keywords: Google trends; Empirical similarity; Aggregation levels; Volatility, Exchange rate.

*Corresponding author.

E-mail address: mario.d.bustamante@gmail.com (M. Durán Bustamante)

Peer Review under the responsibility of Universidad Nacional Autónoma de México.

<http://dx.doi.org/10.22201/fca.24488410e.2018.1710>

0186- 1042/© 2019 Universidad Nacional Autónoma de México, Facultad de Contaduría y Administración. This is an open access article under the CC BY-NC-SA (<https://creativecommons.org/licenses/by-nc-sa/4.0/>)

Resumen

En este trabajo se muestra la ventaja de usar tendencias del motor de búsqueda de Google para pronosticar la volatilidad de corto plazo (semanal) del tipo de cambio peso mexicano - dólar estadounidense. Se realizó un comparativo de modelos que en la literatura ya han incorporado las tendencias de Google como variable explicativa. Algunos de estos modelos utilizan series de tiempo, mientras que otros utilizan la función de similitud, que captura la forma cognitiva del razonamiento humano. Por ejemplo, un inversor que desea conocer el valor que tomará una variable en el futuro tomará en consideración información relevante, conocida y disponible, y la ponderará para calcular su pronóstico. Se concluye que el tomar en consideración la variable de tendencia de Google no es suficiente para explicar el comportamiento volátil semanal del tipo de cambio y que es necesario incorporar más niveles de agregación. Además, de que de acuerdo con nuestro saber existe escasa literatura que utilice las tendencias de Google para explicar variables económicas relevantes.

Códigos JEL: C01, C13, F31.

Palabras clave: Tendencias de Google; Similitud empírica; Niveles de agregación; Volatilidad; Tipo de cambio.

Introduction

Predicting the volatility of the exchange rate in the short term is undoubtedly a challenge for heavily traded emerging markets, which are subject to natural speculation against less volatile currencies in the short term. The Mexican peso (MXN) is the tenth most traded currency in the world. The daily turnover of the peso oscillates around USD 112 000 million and 97% of these transactions are made against the US dollar. The latter currency had a total trading volume of all currencies of 87.6% in 2016.¹

In the Mexican current account in the first quarter of 2016 imports reached USD 89 133 million and exports were USD 85 148 million.² The amount of Mexican pesos that moves over two days in the foreign exchange market is much greater than the exports and imports combined in a single quarter; this is a key factor in the exchange rate being subject to speculation within the market currency.

Moreover, purchasing power parity (PPP) considers the exchange rate as a price; (Taylor & Taylor, 2004) demonstrate that PPP shows deviations from the expected value in the short term. The price or exchange ratio can be considered a demand variable, which is influenced by the external market of goods, the financial market (derivatives), and its own market: the exchange rate (the variable itself generates its own dynamics) and the speculation of this same market. It is necessary to take into consideration that in the short term not all variables play an important role in the supply–demand relationship for foreign exchange.

The financial system has become very complex and this has led to different social science fields, such as psychology and neuroscience, wondering how financial decisions are generated. They consider that financial decisions happen at different levels of aggregation as opposed to economic theories. (Frydman & Camerer, 2016) mentions the example of the efficient market

¹ See the Triennial Bank Survey April by the Bank for International Settlements (BIS, 2016).

² See the (BANXICO, 2016) website for information on the balance of payments in the second quarter.

hypothesis, which only takes into account data from market levels, leaving aside information from individuals.³

(Frydman & Camerer, 2016) find four levels of aggregation: (i) household, (ii) individual trading patterns, (iii) decisions that determine the price of assets in the market, and (iv) corporate investment funds. Data from search engines may belong to more than one level of economic aggregation, although it is difficult to distinguish to what level they belong; nonetheless, part of the results obtained herein provide us with an idea, as mentioned below. Google Trends should belong to (i) or (ii) above, because (iii) and (iv) may well use other specialized sources of information, such as Bloomberg.

Empirical studies have proposed a number of simple models to explain economic variables using Google Trends. For example, (Choi & Varian, 2012) use autoregressive (AR) models to forecast auto parts sales, tourism, and unemployment. They find that Google Trends are often correlated with economic indicators and are useful for improving short-term prediction as they attain the most successful outcomes at breakpoints or inflection points. In addition, (Carrière-Swallow & Labbé, 2013) find that the AR(1) model has a very good fit and produces more successful results outside the sample than considering other higher order specifications. They also analyze consumer behavior, i.e., interest in buying a car, by building an indicator that involves Google Trends.

Other models are based on the *similarity function*, which most closely resembles the natural form of human reasoning. According to (Lieberman, 2010), this function can be used to predict the variable in question, taking into account similar information and the same historical variable, as proposed by (Gilboa, Lieberman, & Schmeidler, 2011). One example is an economic analyst who needs to forecast next year's inflation. Some current applications of the similarity function have been to explain the volatility of the stock exchange (Golosnoy, Hamid, & Okhrin, 2014), and of even greater relevance to this work, it has been used to explain volatility using Google's search engine trends (Hamid & Heiden, 2015). The latter authors find that in short periods of high market volatility, prediction improves when investors' attention to the market is taken into account if Google Trends are used.

The Google search engine is the one most used worldwide, embracing 67.78% of the desktop market, and on mobile devices and tablets 94.4% of the market see (Market share, 2017).

The objective of this paper is to develop different methodologies to examine the MXN–USD exchange rate variable. On the one hand, there are models based on AR, using time series that directly incorporate the trend; on the other hand, there are models based on the similarity, which performs a transformation, as shown in section Empirical volatility models.

Theory and literature review

The use of big data indicators in economics is growing. (Guzmán, 2011) uses Google Trends to forecast inflation. (McLaren & Shanbhogue, 2011) propose that the data generated by a web search can be used by central banks in the management of economic nowcasting. (Carrière-Swallow & Labbé, 2013) use this means for auto sales nowcasting in emerging markets (Smith, 2012) investigates if the activity of the search engines is related to the volatility of the foreign

³ The article by Cary Frydman refers to different examples employing different theories of finance, noting that there are cognitive data that should be considered to obtain a more realistic approach, and that it is a challenge in finance to explain all the different levels of behaviors observed based on just a few principles.

exchange market. Specifically, he takes into consideration the search terms: economic crisis, financial crisis, inflation, and recession. With these terms in Google Trends, he constructs two estimators, one over the short term and one over the long term, for which he considers a four-week moving average. He employs these short-term and long-term estimators in three different ordinary least squares models, incorporating volatility for seven currencies. However, the R^2 statistics are poor considering the use of three variables to explain volatility in all cases and one intercept. It is thought that this may be due to the fact that the search terms chosen are somewhat technical and it is not known whether there is an empirical correlation of these terms with the volatility variable to be explained. Moreover, there is a question concerning whether these particular Google Trends capture the real behavior of investors. Nonetheless, the author concludes that these keywords can predict volatility in the foreign exchange market, albeit they only capture part of the volatility of the entire market of financial decisions that spin around the currency market.

(Hamid & Heiden, 2015) forecast volatility by incorporating Google Trends to forecast the variance of the Dow Jones index based on five-minute returns, focusing on weekly forecasts.

On the other hand, GARCH family models are widely used to analyze volatility. (Kumari & Mahakud, 2015) built a sentiment index to analyze the return on assets of the Indian market. They incorporated the index in the mean and variance of the GARCH model. They find that their sentiment index helps to anticipate negative and positive return volatility.

(Afkhami, Cormack, & Ghoddusi, 2017) looked for terms in the Google engine that best represented market investors in order to forecast energy price volatility. They used ninety search terms related to the energy market and they contrasted them with the usual GARCH.

(Morimoto & Kawasaki, 2017) modelled market volatility taking into account on line intraday news using HAR models, big data techniques and text mining techniques.

Other authors suggest that investor attention is a driver of volatility over short periods of time (Vlastakis & Markellos, 2012), and that there is a close relationship in phases of high volatility (Andrei & Hasler, 2015).

In this paper, we use Google Trends data as volatility variables in relation to the exchange rate because people's interest when looking for relevant information concerning a financial economic variable can be translated as a precursor or an indicator of whether there is volatility in that variable and considering this as one direct variable in the model would imply a lower final impact of this variable. (Alvarez, Atkeson, & Kehoe, 2007) consider that in the short run, the exchange rate is not affected in the first statistical moment; they propose a theoretical model in which they analyze the effects of the second statistical moment on the currency.

Empirical volatility models

According to (Gilboa, Lieberman, & Schmeidler, 2006), there are sometimes issues to be solved that are unique, such as calculating the sale price of a work of art, knowing if a person's body will react successfully to a certain operation, or an analyst determining the rate of inflation for the following year. Thus, sometimes the information is not sufficiently relevant to explain a particular case and it is sometimes necessary to weigh the information, counting it and assigning it a weight according to its importance for the case under study. This seems a more accurate approach than simply taking all available information directly and expecting it to suffice to answer the issue that the analyst wishes to address; in this context considering "similar" variables adopted in other studies can improve the prediction process. Economic

agents prefer to obtain good results based on similar cases that have previously been well adjusted.

Mathematical models that incorporate cognitive processes, observed behaviors, and novel data sources generated by individuals or the economy are considered models of economic behavior. These types of predictive models are formulated using the similarity function see (Gilboa *et al.*, 2011). (Lieberman, 2010) argues that the similarity function can be considered a natural model of human reasoning and its statistical validity has been shown through the axioms of (Gilboa *et al.*, 2006).

We define the process of exhibiting interest y_{t+1} as follows:

$$y_{t+1} = \sum_{i=1}^p \phi[y_t, x_{i,t-1}] y_{i,t} + \varepsilon_{t+1} \quad (1)$$

with $\varepsilon_t \sim (0, \sigma^2)$.

(Golosnoy *et al.*, 2014) interpret $x_{i,t-1}$ as previous forecasts similar to y_t . The variable $x_{i,t-1}$ must naturally be similar to y_t in the past. For this reason, the distance between the two must be small, so as to give:

$$\phi[y_t, x_{i,t-1}] = \frac{\theta[y_t, x_{i,t-1}]}{\sum_{j=1}^p \theta[y_t, x_{j,t-1}]} \quad (2)$$

The forecast is given by:

$$\hat{y}_{t+1} = \sum_{i=1}^p \phi[y_t, x_{i,t-1}] y_{i,t} \quad (3)$$

The similarity function is represented by a Euclidean distance $d(y_t, x_{i,t-1}) = \omega_i (y_t - x_{i,t-1})^2$, the main function of which is to measure the distance between the study problem and similar situations with the objective of transmitting these weighted measurements to the similarity function, and this in turn to the model of “empirical similarity.”

The concept of “similarity” can be extended to analyze volatilities. This would imply that y_t must be a known volatility variable to be forecast, and the similarity function must be constructed with variables that can anticipate future volatility behavior based on past behavior; in line with this, the economic agent will take the information and consider it in accordance with the function of similarity to construct the estimate of the volatility that will arise in the future.

The similarity function is specified as:

$$v_{t+1} = \sum_{i=1}^p \phi[v_t, G_{i,t-1}] v_{i,t} + \varepsilon_{t+1} \quad (5)$$

where $G_{i,t-1}$ is the trend behavior variable generated by the individuals.

Temporal models

The random walk can be represented by an AR(1) process. (Charles R. Nelson & Charles R. Plosser, 1982) analyze different macroeconomic series, concluding that the best prognosis for these is very close to the random walk. In different works, it has been established that for the main currencies a random walk can be a good approximation of the nominal exchange rate.

The simpler model proposed by (Choi & Varian, 2012), based on time series, follows the following dynamic:

$$y_t = b_1 y_{t-1} + b_{12} y_{t-12} + b_0 g_t + e_t \quad (6)$$

With $e_t \sim (0, \sigma^2)$. Where y_t is the process to be explained, g_t the variable corresponding to Google trends, and $b_{12} y_{t-12}$ the seasonal component, if this exists.

Data

Google trends data

Google Trends data are collected through the company's search engine; this collates relevant searches for any topic browsed on one's page and that can be downloaded from the (Google, 2017b) web page. The data are standardized by the keywords that generate the highest number of searches on a specific date. This value represents 100% and the other values for data must be lower than this amount.

The Google Trends website allows one to view and download up to five joint variables, but one must be very careful because when they are downloaded together, they are standardized based on the most searched variable. If the query is downloaded in isolation, it is standardized based on the most searched data. It is possible to select the data by region of the world and choose the temporal frequency. It is also possible to select where one looks for the topic of interest: in the web browser, in the image finder, or in the news. Moreover, it also allows one to visualize geographically the countries represented in the search area, shading the most important regions in navy blue and the less important in light blue in terms of the total number of searches.

We use historical data for the trend of the "precio dolar," counted weekly beginning in the week 04/01/2004 to 10/01/2004 and ending in the week 03/20/2016 to 03/26/2016 (639 observations). The observations begin on Sunday and end on Saturday. The process of choosing the "precio dolar" variable is as described in Annex A. In terms of the selection of trend variables, the geographical area was not delimited, the intention being to visualize which countries are interested in acquiring USD and what might start speculation in demand for USD, i.e., what would motivate more foreign investors to purchase this reserve currency, or simply to reflect the behavior of an entire area that might acquire USD, directly affecting the MXN-USD exchange rate.

Exchange rate data

The historical data for the exchange rate were downloaded from the Banco de México (BANXICO, 2017) portal; the information is daily from January 1, 2001 to March 29, 2016. The data for Fridays are taken as the closing values of the exchange rate price in each week, so that the data are weekly and are homogeneous with the weekly trend data.

Realized volatility of the exchange rate

Realized volatility refers to volatility that occurred in the past and usually concerns derivatives. For example, if we wish to examine monthly volatility, it can be calculated by taking the standard deviations of the daily returns in the desired month according to the (Nasdaq, 2017) website.

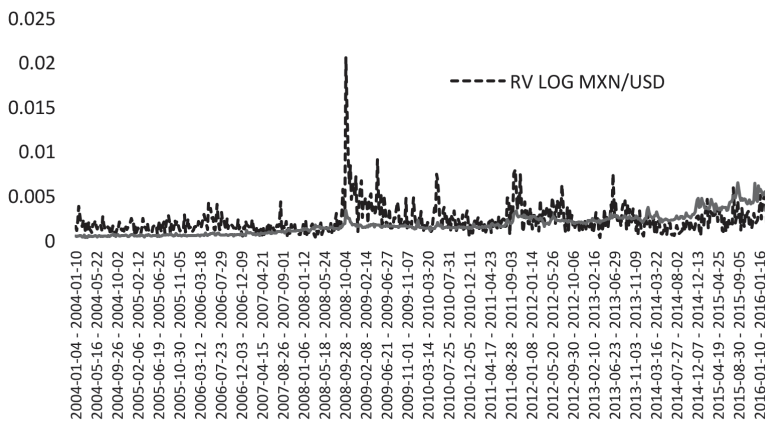


Figure 1. Realized volatility of the log exchange rate, MXN–USD, against Google Trend data. (2004–2016).
Source: Own estimation with data for the MXN–USD exchange rate and the variable “precio dolar” in Google Trends; both variables are on a similar scale. (2004–2016)

For the second time (the first being in the selection of trend variables), we use Granger’s causality test on the realized volatility and the Google Trends variable, finding that in its first and sixth lags in the test, the Google Trends variable Granger causes realized volatility with a significance level of 5%.

It should be noted that the data, even for the past, are updated according to the maximum value recorded in the specified period of time. For example, in a future period when a higher query volume for a search term is registered than in the previous data, this maximum modifies all the series, adjusting them with the new value. In considering a specific range in the past that does not consider the present period, clearly in that time range the maximum search value will be different.⁴ Thus, when considering a fixed data series, it is incorrect to consider dividing the series within the sample. It would only be valid if the maximum value of the whole series were considered in the same range within the sample, or another series were downloaded for

⁴ Interest over time: “Numbers represent the search interest in relation to the maximum value of the chart for the specified region and time. The value 100 indicates the maximum popularity for the term, 50 implies half of popularity and 0 is equivalent to a lesser popularity of 1% compared to the value 100.” (Google, 2017b).

the selected period; this might change the specification of the model that takes the present into consideration. We consider that the best option is to perform contrasts outside and not within the sample.

Methodology and results

According to (Choi & Varian, 2012), trend variables can be incorporated directly into the model. Applying the same idea to explain realized volatility, we have the following:

$$RV_t = intercept + b_1RV_{t-1} + b_0g_t + e_t \quad (7)$$

In this specification, all coefficients are found to be significant and the standard error has a level consistent with the coefficients see Table 1.

Table 1
Parameters of the time series of volatility.

	Coefficient	Standard Error	t value	Pr (> t)
Intercept	0.00130	0.00042	3.09019	0.00224
b_0	0.39596	0.12528	3.16068	0.00178
b_1	0.42051	0.05292	7.94531	7.51E-14

Adj R-squared: 0.2093, AIC: -10.5437

Source: Own elaboration.

On the other hand, the empirical similarity (ES) method is not interested in directly analyzing the variables over time. The methodology tries to weigh similar cases to improve the forecast. This feature is an advantage when the historical information does not exhibit regular and consistent periodicity in the entire sample.

Here, we adopt a variation of the ES methodology, as described by (Hamid & Heiden, 2015):

$$RV_t = intercept + b_1 \exp(-\omega_1(RV_{t-1} - G_{t-1}))RV_{t-1} + \varepsilon_t \quad (8)$$

RV_t and $RW_t = \exp(-\omega_1(RV_t - G_t))RV_t$ are weakly stationary in both models. Table 2 reports our results of unit root tests (Dickey & Fuller, 1979, 1981) and (Phillips & Perron, 1988).

Table 2
Unit root tests in levels.

Variable	Unit root tests(Level)	Test statistic	1% Critical value	5% Critical value	10% Critical value
RV_t	Dickey-Fuller(DF)	-3.65289	-3.457061	-2.87319	-2.573054
	Phillips-Perron(PP)	-10.8164	-3.457061	-2.87319	-2.573054
RW_t	Dickey-Fuller(DF)	-3.89707	-3.457061	-2.87319	-2.573054
	Phillips-Perron(PP)	-11.19332	-3.457061	-2.87319	-2.573054

Source: Own elaboration.

Table 3 shows the values of the parameters obtained for the ES model; the Akaike information criterion (AIC) is quite similar in both models. Interestingly, the significance level of the second model is preferable based on the values of the t-statistic.

Table 3
 Parameters of the empirical volatility model.

	Coefficient	Standard Error	t value	Pr(> t)
Intercept	0.00091	8.86E-05	10.29908	4.16E-23
b_1	0.58949	0.03244	18.17408	9.61E-60

Adj R-squared: 0.3421, AIC: -10.4203

Source: Own elaboration.

Figure 2 shows the out-of-sample auto one-step-ahead predictions with fixed coefficients. The period visualized is from April 4, 2016 to October 29, 2016; as we can see, both methodologies fail to capture the full magnitude of the volatility, leading us to believe that the financial decisions that are reflected only represent a part of the total behavioral decisions within the foreign exchange market.

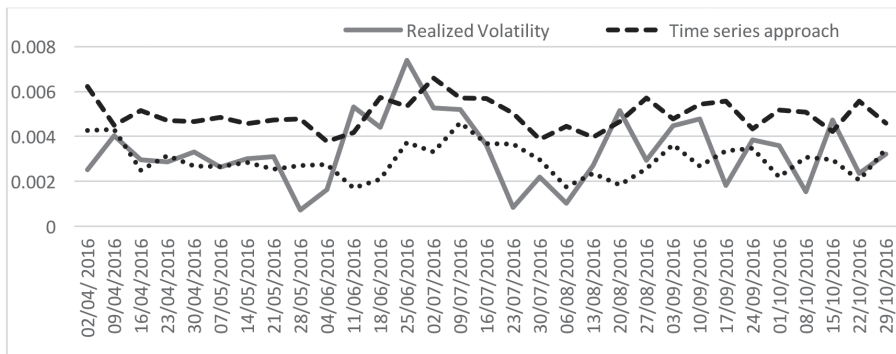


Figure 2. Out-of-sample; realized volatility; time series approach and ES approach. (2016).
 Source: Own elaboration with Banxico and Google Trends data. (2016)

The ES approach follows the real volatility more closely, but it does not capture the full breadth of the volatility or its shape.

In Figure 3, the behavior shown by investors before the historical maximum of the MXN–USD exchange rate shows no considerable variation. If we undertake a visualization using a window of more years, it is clear that there is no change in tendency in the series over the period. In Table 4, we can confront this with the real values of the exchange rate. Over time, interest only increases highly from September 19, 2016 to September 20, 2016 (see Figure 3) and from September 21, 2016 to September 26, 2016 the interest rate is down. In this last period, higher values of the MXN–USD exchange rate are achieved.

Table 4

Variation in the MXN–USD exchange rate in the month of September.

Daily Date	USD/MXN
07/09/2016	18.3689
08/09/2016	18.5427
09/09/2016	18.8451
12/09/2016	19.0646
13/09/2016	19.152
14/09/2016	19.2275
15/09/2016	19.2514
19/09/2016	19.6097
20/09/2016	19.777
21/09/2016	19.8394
22/09/2016	19.5965
23/09/2016	19.7211
26/09/2016	19.8322

Source: Own elaboration with data of BANXICO.

Note: In this period, there is no marked prior interest for different keywords; only on 09/19/2016 and 09/20/2016 does it show a sudden interest, but this disappears quickly.

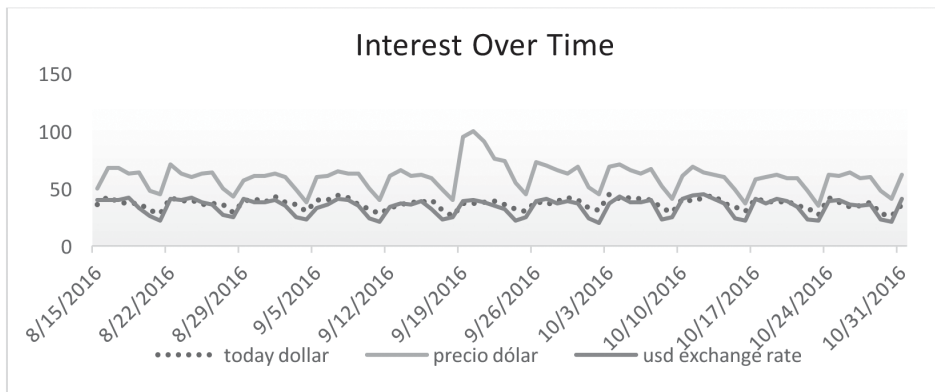


Figure 3. The MXN achieves its maximum historical value of USD 19.83 (MXN/USD) on 09/21/16. (2016)

Source: Elaborated using the Google Trends portal for the keywords “today dollar”, “precio dolar”, and “USD exchange rate”. (2016)

Conclusions

The way in which the volatility of the MXN–USD rate is modeled in this work is quite novel because it incorporates the behavior of individuals in different regions of the world whose financial decisions are focused on acquiring USD, possibly to avoid a devaluation in capital, or simply to try to generate profits by anticipating the depreciation of an emerging country’s currency. All this directly affects the Mexican peso in the short term. The similarity model represents a clear way of capturing such behavior. Even so, a lot of work remains to be done; in particular, it is necessary to include more relevant variables for short-term impact.

For example, in the results obtained, even in the most volatile periods it is not possible to capture the behavior of all the investors. Examining different variables in Google Trends,

they do not reflect drastic upward movements in the month of September 2016 (see Fig. 3), which makes us think that there are different levels of economic aggregation at which different financial decisions are made. Moreover, the Google Trends variable only represents a part of the financial decisions that are made in the currency market.

An extension of this study would be to determine what sources of information allow us to anticipate the behavior of big capital, which surely comes from large corporate investment funds and cannot be captured by the Google Trends volatility proxy chosen here.

References

- Afkhami, M., Cormack, L., & Ghodusi, H. (2017). Google search keywords that best predict energy price volatility. *Energy Economics*, 67, 17–27. <https://doi.org/10.1016/J.ENERCO.2017.07.014>
- Alvarez, F., Atkeson, A., & Kehoe, P. J. (2007). If Exchange Rates Are Random Walks, Then Almost Everything We Say About Monetary Policy Is Wrong. *American Economic Review*, 97(2), 339–345. <https://doi.org/10.1257/aer.97.2.339>
- Andrei, D., & Hasler, M. (2015). Investor Attention and Stock Market Volatility. *Review of Financial Studies*, 28(1), 33–72. <https://doi.org/10.1093/rfs/hhu059>
- BANXICO. (2016). Balanza de pagos, comercio e inversión, Banco de México. Available in: <http://www.banxico.org.mx/informacion-para-la-prensa/comunicados/sector-externo/balanza-de-pagos/indexpage.html> ; Retrieved: 7/11/2017
- BANXICO. (2017). Consulta de Estructura de Información. Available in: <http://www.banxico.org.mx/SieInternet/consultarDirectorioInternetAction.do?accion=consultarCuadro&idCuadro=CF373> ; Retrieved: 7/11/2017
- Billot, A., Gilboa, I., & Schmeidler, D. (2008). Axiomatization of an exponential similarity function. *Mathematical Social Sciences*, 55(2), 107–115. <https://doi.org/10.1016/J.MATHSOCSCI.2007.08.002>
- BIS. (2016). *Triennial Central Bank Survey Foreign exchange turnover in April 2016*. Available in: <https://www.bis.org/publ/rpfx16fx.pdf>
- 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>
- Charles R. Nelson, & Charles R. Plosser. (1982). Trends and random walks in macroeconomic time series: Some evidence and implications. *Journal of Monetary Economics*, 10(2), 139–162. [https://doi.org/10.1016/0304-3932\(82\)90012-5](https://doi.org/10.1016/0304-3932(82)90012-5)
- Choi, H., & Varian, H. (2012). Predicting the Present with Google Trends. *Economic Record*, 88(s1), 2–9. <https://doi.org/10.1111/j.1475-4932.2012.00809.x>
- Dickey, D. A., & 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. <https://doi.org/10.1080/01621459.1979.10482531>
- Dickey, D. A., & Fuller, W. A. (1981). Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root. *Econometrica*, 49(4), 1057. <https://doi.org/10.2307/1912517>
- Frydman, C., & Camerer, C. F. (2016). The Psychology and Neuroscience of Financial Decision Making. *Trends in Cognitive Sciences*, 20(9), 661–675. <https://doi.org/10.1016/J.TICS.2016.07.003>
- Gilboa, I., Lieberman, O., & Schmeidler, D. (2006). Empirical Similarity. *Review of Economics and Statistics*, 88(3), 433–444. <https://doi.org/10.1162/rest.88.3.433>
- Gilboa, I., Lieberman, O., & Schmeidler, D. (2011). A similarity-based approach to prediction. *Journal of Econometrics*, 162(1), 124–131. <https://doi.org/10.1016/J.JECONOM.2009.10.015>
- Golosnoy, V., Hamid, A., & Okhrin, Y. (2014). The empirical similarity approach for volatility prediction. *Journal of Banking & Finance*, 40, 321–329. <https://doi.org/10.1016/J.JBANKFIN.2013.12.009>
- Google. (2017a). Google Correlate. Available in: <https://www.google.com/trends/correlate> ; Retrieved: 7/11/2017
- Google. (2017b). Google Trends. Available in: <https://trends.google.com.mx/trends/> ; Retrieved: 7/11/2017

- Guzmán, G. (2011). Internet search behavior as an economic forecasting tool: The case of inflation expectations. *Journal of Economic and Social Measurement*, 36(3), 119–167. <https://doi.org/10.3233/JEM-2011-0342>
- Hamid, A., & Heiden, M. (2015). Forecasting volatility with empirical similarity and Google Trends. *Journal of Economic Behavior & Organization*, 117, 62–81. <https://doi.org/10.1016/J.JEBO.2015.06.005>
- Kumari, J., & Mahakud, J. (2015). Does investor sentiment predict the asset volatility? Evidence from emerging stock market India. *Journal of Behavioral and Experimental Finance*, 8, 25–39. <https://doi.org/10.1016/J.JBEF.2015.10.001>
- Lieberman, O. (2010). ASYMPTOTIC THEORY FOR EMPIRICAL SIMILARITY MODELS. *Econometric Theory*, 26(4), 1032–1059. <https://doi.org/10.1017/S0266466609990454>
- Market share. (2017). Report. Available in: <https://netmarketshare.com/>
- McLaren, N., & Shanbhogue, R. (2011). Using Internet Search Data as Economic Indicators. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.1865276>
- Morimoto, T., & Kawasaki, Y. (2017). Forecasting Financial Market Volatility Using a Dynamic Topic Model. *Asia-Pacific Financial Markets*, 24(3), 149–167. <https://doi.org/10.1007/s10690-017-9228-z>
- Nasdaq. (2017). Realized volatility Definition. Available in: <http://www.nasdaq.com/investing/glossary/r/realized-volatility> ; Retrieved: 7/11/2017
- Phillips, P. C. B., & Perron, P. (1988). Testing for a unit root in time series regression. *Biometrika*, 75(2), 335–346. <https://doi.org/10.1093/biomet/75.2.335>
- Smith, G. P. (2012). Google Internet search activity and volatility prediction in the market for foreign currency. *Finance Research Letters*, 9(2), 103–110. <https://doi.org/10.1016/J.FRL.2012.03.003>
- Taylor, A. M., & Taylor, M. P. (2004). The Purchasing Power Parity Debate. *The Journal of Economic Perspectives*, 18, 135–158. <https://doi.org/10.1257/0895330042632744>
- Vlastakis, N., & Markellos, R. N. (2012). Information demand and stock market volatility. *Journal of Banking & Finance*, 36(6), 1808–1821. <https://doi.org/10.1016/J.JBANKFIN.2012.02.007>

Annex A

Selection of trend variables

In the selection of the main variable, we have taken into account that the geographical area reflecting the search term should not repeat countries. We consider whether the region can represent the behavior of an entire economic zone that is able to acquire large amounts of the currency, or anticipate a previous demand for USD as a refuge in the face of the possibility of a devaluation of the exchange rate, motivating the investors in other parts of the world to generate profits on the currencies of different emerging markets that are affected by the demand for USD.

We have also ensured that the correlations calculated by the Google Correlate tool see (Google, 2017a) are consistent with keywords related to the exchange rate and not with other Internet search terms.

Graphically, we have corroborated that the behavioral patterns of Internet searches hold a similar trend to the historical data for the currency in question.

The following table shows the searches made for different terms related to the MXN–USD exchange rate.

Table A1

*The country (countries) generating the highest level of search for each term.

Search Term	Relevant Countries	Important Correlations in Google correlate
peso dólar	Chile *, Argentina, Ecuador, Uruguay, Mexico	0.9457 with cambio peso dólar
today dollar rate	Pakistan *, Nigeria *, India *, Uganda, Philippines	0.9844 with today dollar
usd to mxn	Mexico *, United States, Philippines	0.9985 with 1 usd to mxn
usd dollar to	Philippines *, Mauritius, Sri Lanka	without results
cotizacion del dólar	Uruguay *, Peru	0.9414 with cotización dólar
precio dolar	Venezuela *, Colombia, Argentina, Chile	0.9593 with precio del. 0.9424 with dólar Americano
price dollar	Pakistan *, Nepal *, India, Bangladesh	without results
tipo de cambio dólar	Venezuela *, Colombia, Argentina, Chile	0.964 with cambio dólar
Usd	Singapore *, United Arab Emirates,	0.9868 with to usd
cambio dólar	Costa Rica, Nicaragua, Honduras	0.9649 with tipo de cambio dólar, 0.9564 with valor del dólar
mexican peso	Mexico *, Canada, United States	0.8964 with dollar vs peso
1 usd to mxn	Mexico, Philippines, United States	0.9985 with usd to mxn
tipo de cambio dólar	Costa Rica, Peru, Nicaragua	0.9649 with cambio dólar, 0.9520 with peso mexicano
mxn to usd	Mexico*	0.9909 with mxn
dólar peso mexicano	El Salvador *, Mexico *, Guatemala	0.9658 with el dólar hoy
dollar price	Pakistan *, Nepal *, India *, Burma, Bangladesh	0.9409 with today dollar
cambio peso dólar	Chile *, Ecuador, Panama, Mexico, Argentina	0.9928 with tipo de cambio peso dólar
dollar to	Ghana, Nigeria, Philippines, Lebanon	0.9938 with usd dollar to
usd exchange rate	Zimbabwe, Singapore	0.9941 with usd exchange
dollar to mexican peso	Mexico *, Canada, United States	0.9864 with us dollar to mexican peso
1 usd dollar	Philippines *, Syria	without results
usd exchange	Maldives *, Zimbabwe	0.9941 with usd exchange rate
rate dollar	Ghana *, Zambia, Nigeria, Pakistan, Nepal	0.9925 with exchange rate dollar
dolar hoy	Argentina *, Colombia	0.9612 with tipo de cambio
dollar rate	Ghana *, Zambia, Nigeria, Pakistan	0.9238 with today dollar
tipo de cambio peso dólar	Mexico*	0.9928 with cambio peso dólar
exchange rate dollar	Zambia *, Nigeria *, South Africa, Zimbabwe	0.9925 with rate dollar
dólar americano	Ecuador *, Venezuela	0.9952 with dolares
peso mexicano	Mexico *, Guatemala, Honduras	0.9589 with dólar peso mexicano
dollar exchange rate	Zambia *, Nigeria, South Africa, Zimbabwe, Ghana	0.9082 with dollar exchange

Source: Own elaboration with Google Trends and Google Correlate.

Taking into account that the keyword must maintain a similar trend with the exchange rate (i.e., one can introduce the keywords in the portal and see a similar trend), Google Correlate was consistent in terms of keywords related to the exchange rate. The geographical area can be interested in acquiring USD or it can start speculation in demand for USD. The 30 search terms are reduced to only three: “precio dolar,” “USD exchange rate,” and “today dollar.”

From the calculation of the correlation between these variables, which is very high (almost close to 1), we obtain the values of the R^2 statistic for the MXN–USD exchange rate vs. the three Google trends, as shown in Table A2.

Table A2

Correlations between the three keywords.

	Correlations			R^2
	precio dólar	usd exchange rate	today dollar	usd/mxn
precio dólar	1	0.9416	0.9060	0.8073
usd exchange rate	0.9416	1	0.95726958	0.7329
today dollar	0.9060	0.9572	1	0.6790

Source: Own elaboration.

Based on Table A2, the concern is to identify which of the three variables should be chosen as the proxy variable; adopting the three variables would not be appropriate because of the high correlation between them. In addition, simply considering the

R^2 statistic as a selection method would not be appropriate.

We have used the Granger causality test to determine whether one variable could be a causal precursor of another, and to consider the variable showing the largest source of previous speculation in the analysis. The results show bidirectional causation for the USD and the today dollar rate with a level of significance of 1%. In terms of the second trend variable, the USD exchange rate causes the USD price with a level of significance of 1%, and in the opposite direction, the USD price causes the USD exchange rate with a level of significance of 10%.

Moreover, when the number of lags increases, we have another consistent result, namely that the USD exchange rate causes the today dollar rate at different levels of lags with a level of significance of 1% for a lag one to a lag ten which suggests that the relationship between the two variables is very strong even with many lags; thus, we discard the today dollar variable.