INTERNATIONAL REAL ESTATE REVIEW

2021 Vol. 24 No. 4: pp. 613 – 631

Use of Google Trends to Predict the Real Estate Market: Evidence from the United Kingdom

Grzegorz Michal Bulczak
Gdynia Maritime University, Faculty of Management and Economics
Email: g.bulczak@wznj.umg.edu.pl

This paper demonstrates how Google Trends data can be used to improve real estate market predictions. Online searches produce valuable data that precede economic decisions. This study explores the usefulness of Google search engine data in predicting the real estate markets. The results indicate that Google data can be an additional source of information for investors and policymakers. This analysis adds to the existing literature that explores the role of behavioral factors in the decision-making process. Google Trends data are identified as an important predictor of real estate market prices and sales volume.

Keywords

Real Estate Prices, Google Trends, Forecasting

1. Introduction

Governments, economists, investors and the media are following publications of periodic economic data with interest. These publications are usually available with a significant delay. This is particularly true in the case of the real estate market. Information about the housing market, for a given month, is published a month later and usually updated as new data come in. However, observing real estate market movements earlier is associated with important benefits including better-informed decisions. Selected companies, examples of which include Google, MasterCard and UPS, provide data likely to be useful in predicting changes in economic activity (Jun, Yoo and Choi, 2018). This article will examine the usefulness of data from Google Trends (GT) in an analysis of changes in the real estate market in the United Kingdom (UK) during and shortly after the 2007/2008 global financial crisis. GT is based on the number of queries related to a selected search topic. For example, the number of queries 'sell/buy a house' in a given week of the month. Presumably, the frequency of a selected query may be correlated with the level of economic activity (specifically the real estate market) in a given period of time. Queries entered in search engines will likely be useful leading indicators of consumer decisions, especially in situations where consumers are beginning to plan purchases well before making a decision (Preis, Moat and Stanley, 2013; Yang, Santillana and Kou, 2015; Cervellin, Comelli and Lippi, 2017; Jun, Yoo and Choi, 2018; Silva *et al.*,2019).

In many settings, the activity of Internet users at a given point in time is assumed to reflect collective behavior, and shows the interests, concerns and intentions of the observed population (Yang, Santillana and Kou, 2015). From this point of view, one can speculate that the object of interest of Internet users today is correlated with their actions in the near future. Consumers who are considering buying a house use search engines to compare different locations, types of real estate, prices and financing. Similarly, prospective film viewers may check the release date of a film before heading out to the movie theatres and individuals who are planning for a holiday search for information on travel destinations or flight and hotel room prices.

The purpose of this article is to present possible applications of GT in an analysis of economic activity and in particular, the real estate market. The analysis here focuses on the real estate market of the UK particularly on the 2007/08 global financial crisis and the demand side. The selected period is chosen purposefully as it is characterised by dynamic changes and high levels of uncertainty. In such settings, valid predictions are arguably the most valuable (Krugman, 2009; Oust and Eidjord, 2020). The aim is to verify the usefulness of search engine data and show the possibilities and further applications of this type of analysis in future research work. Studying the real estate market is particularly important in the face of the recent financial crisis and the economic slowdown. This article will contribute to studies related to nano-economics and individuals who are making decisions about buying or selling (Arrow, 1987).

Previous studies suggest that Internet search engines are important for predicting economic activity and consumer behavior (Cooper *et al.*, 2005; Ettredge, Gerdes and Karuga, 2005; Choi and Varian, 2012). Goel *et al.* (2010) provide an overview of recent epidemiological research work, macroeconomic time series and consumer activity related to movies, music and video games. The benefits and deficiencies of data obtained through search engines are discussed. Other previous studies pay particular attention to the prediction of the current state of population health, which include the impacts of COVID-19, influenza and zika epidemics (Polgreen *et al.*, 2008; Teng *et al.*, 2017; Flanagan, Kuo and Staller, 2021; Strzelecki, 2020) or oil consumption, precious metals and macroeconomic indicators (Castle, Fawcett and Hendry, 2009; Yu *et al.*, 2019; Salisu, Ogbonna and Adewuyi, 2020). Askitas and Zimmermann (2009) further examine the practicality of this type of data in the context of unemployment in the United States (US). In addition, Guzmán (2011) examines the usefulness of search engine data in inflation forecasting.

Cooper *et al.* (2005) show that searches for information on cancer in 2001-2003 are correlated with the estimated occurrence of the disease. Eysenbach (2006) finds a high correlation between the number of the sponsored search results for keywords related to the flu and epidemiological data in Canada. Similar results are found in Sweden (Hulth, Rydevik and Linde, 2009). Choi and Varian (2012) emphasize that forecasting the status quo is particularly important for central banks and other government agencies. They use a number of search variables to improve the accuracy of forecasting economic activities, including car sales, tourism and unemployment benefits in the US. The potential value of using Google data to analyze the real estate market in the UK is also supported by related studies in the US (Jun, Yoo and Choi, 2018). In 2012, 90% of homes buyers in the US used the internet to verify the information related to the purchase (National Association of Realtors, 2012). Similarly, the Association of Realtors in California reports that in 2008, 63% of home buyers were searching for a real estate agent via a web search (Appleton-Young, 2008). Furthermore, a recent study in Taiwan finds that there is a link between the number of visits to an actual price registration system and real estate transactions, transaction volume and prices (Lin and Hsu, 2020).

Analyses that use data from search engines have a number of problems, particularly those in the field of econometrics, which include topics such as the selection of appropriate variables and data correction. The recent financial crisis in the US has shown that none of the models could anticipate problems in the housing market and the wider economy (Krugman, 2009). This, to some extent, could be the result of the variety of techniques used to extract information from the data affected by noise and a high degree of measurement

error (Simon, 1984). The social sciences however can overcome these problems with tools to observe a phenomenon analyzed at a higher resolution.
The literature presented above shows that data from web search engines have value for forecasting economic phenomena. The literature review here suggests that in the UK, which is a much smaller real estate market relative to the US, search data from the GT will improve the prediction accuracy of price and sales volume models. The main factors that may cause differences in the estimation results are: the difference in popularity of the Google search engine and English language differences between the US and the UK.

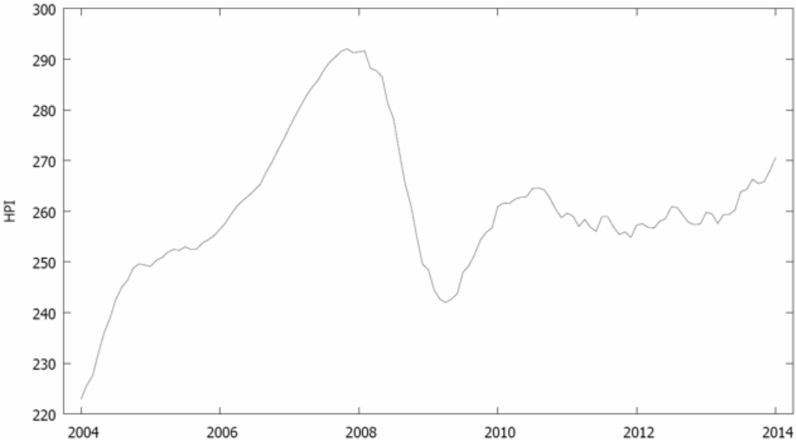
2. Methods

This paper uses GT and monthly data of the real estate market in the UK for 2004-2014. In 2004, Google provided data on indexes of queries typed in the search tool. These data allow the observation of changes in the popularity of queries in a given time period, globally, nationally and regionally. Query indexes take values from 0 to 100, and measure the popularity of a given word or phrase (index equal to 100 refers to maximum popularity). A given index measures the share of queries, which is counted as the share of the query in a given geographic location, divided by the total number of queries in this region at the moment. Google is a very popular search tool worldwide, however, significant differences between countries can be observed. Most online queries in the US (67%) and the UK (91%), are entered through Google (Seymour, Frantsvog and Kumar, 2011; ComScore, 2012).

In addition to the search data, this study will also use data from The Land Registry House Price Index (HPI). This index shows the changes in the value of residential property prices in the UK. The HPI is based on transactions in the housing market in England and Wales. The HPI informs about changes in both the national and regional markets. The HPI also includes the number of transactions in a given month (sales volume). Data from the HPI account for seasonal fluctuations in prices. Figure 1 shows the HPI and Figure 2 shows the average annual change in GT queries ‘house for sale’ ‘and mortgage’, for 2004-2014. The HPI peaked in 2008 but queries of 'house for sale' typed into the Google search bar box reached the local maximum in 2007. Figure 2 shows a trend of increase in popularity of the query 'house for sale' from 2008 to 2014.

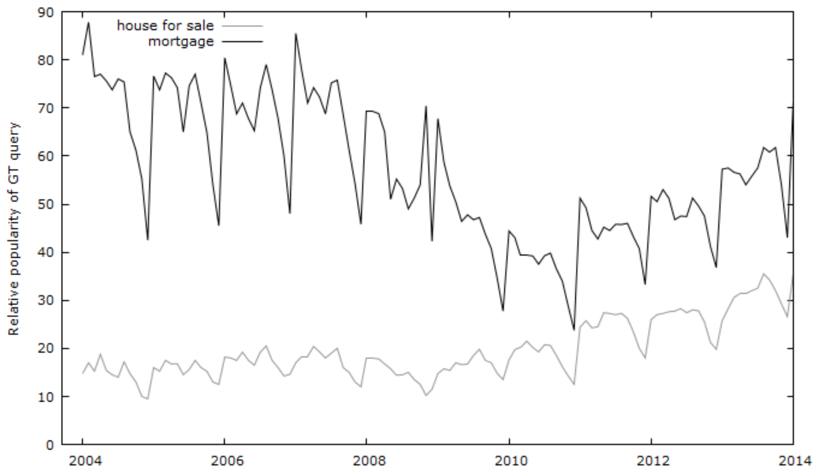
It seems plausible that in certain conditions there may be very little variation in the average house prices with only the number of sales providing valuable information on the movement of the real estate market. For this reason, the following models will not only focus on predicting the prices but also the volume of sales. The first point of the analysis is using a basic autoregressive model (AR) AR-1, in which home sales (sales vol) (y_t) will depend on sales in the past time period (y_{t-1}) and seasonal fluctuations of S_i. The analysis covers the period of January 2004 to February 2014.

Figure 1 HPI – United Kingdom



Source: The Land Registry House Price Index, 2014.

Figure 2 Google Trends Query '*House for Sale*'



Source: Google Trends, 2014.

$$y_t = b_1 y_{t-1} + \sum S_i + e_t \tag{1a}$$

Similarly, for the HPI, HPI (y_t) is dependent on past values of the index (y_{t-1}). The HPI is already adjusted for seasonality.

$$y_t = b_1 y_{t-1} + e_t \tag{1b}$$

and

$$y_t = b_1 y_{t-1} + b_2 SalesVol_{t-1} + e_t \tag{1.1b}$$

Models 1a and 1b will serve as a reference for the alternative models in which GT data are introduced:

$$y_t = b_1 y_{t-1} + b_2 GT_t + b_3 GT_{t-1} + \sum S_i + e_t \tag{2}$$

The extended models also include the changes in the price index in the previous time period (HPI_{t-1}), where y_t is the sales volume:

$$y_t = b_1 y_{t-1} + b_2 GT_t + b_3 GT_{t-1} + b_4 HPI_{t-1} + \sum S_i + e_t \tag{3a}$$

Similarly, changes in the number of sales in the previous period ($SalesVol_{t-1}$) are included, where y_t is the HPI_t:

$$y_t = b_1 y_{t-1} + b_2 GT_t + b_3 GT_{t-1} + b_4 SalesVol_{t-1} + e_t \tag{3b}$$

Based on the results from previous studies on the real estate market in the US, the expected impact of the GT, for specific searches, will be a significant one, particularly for short lags (Choi and Varian, 2012; Jun, Yoo and Choi, 2018; Oust and Eidjord, 2020). The addition of GT is likely to enhance the accuracy of alternative models. After a preliminary selection of the most suitable queries based on Wu and Brynjolfsson (2013) and Choi and Varian (2012), the following search words are considered in the models: 'rentals', 'house for sale' and 'mortgage'. The selected lag structure is directly based on Wu and Brynjolfsson (2013) and Choi and Varian (2012), where it is shown that the shortest lags are most relevant in predicting the housing market. To support this approach, the lag structure estimates are presented in Tables A3-A5 in the Appendix. The results of this analysis may deviate from those of previous studies due to the English language differences and different characteristics of the real estate markets in the US and the UK. Furthermore, the difference in popularity of Google between the US and the UK may also be a significant factor. In addition, in models where the HPI is the outcome of interest, it might be difficult to differentiate between the forces of supply and demand related to a given search query. All of the variables used in the following analysis are first differences to ensure stationarity. The results of the unit root test are provided in Table A1 in the Appendix.

3. Results

Table 1 shows the results of the estimation for Models 1a, 2a and 3a (base models). Model 1a explains for 64% of the variation in sales volume. This base model shows that the number of transactions in a given month is heavily dependent on transactions made in the previous month (t-1). Model 2a includes the GT query index for 'rentals'. The results suggest that the GT query index for 'rentals' is negatively correlated to the number of sales in the same period of time. In the analyzed period, a one-percent increase in inquiries about renting is associated with 389 fewer transactions (this is 0.5% of the average number of transactions in the analyzed period). In Model 3a, the HPI_{t-1} variable (the price index of the past month) is added in addition to the variables from Model 2a. The results remain largely unaffected and still suggest a negative relationship between the price index in the previous time period and the number of transactions. The size and significance level of the coefficient on 'rentals' are noticeably increased.

Table 1 Estimation of Sales Volume, 2004 - 2014

	Sales_vol[1]	**Sales_vol**	**Sales_vol**
Model	1a	2a	3a
$SalesVol_{t-1}$	0.941***	1.038***	1.0634***
Rentals		-389.088***	-449.816***
HPI t-1			-136.201**
Additional Controls	seasonal	seasonal	seasonal
Adjusted R-Squared	0.638	0.633	0.633
AIC	21.26	21.53	20.68

Notes: [1] *, **, *** indicates the levels of significance of 10%, 5%, and 1. Only significant coefficients are reported.

The next step in the analysis is to examine the usefulness of the GT data in predicting the HPI. As in Wu and Brynjolfsson (2013) and Choi and Varian (2012), the results from Model 1b indicate a strong correlation between HPI_t and HPI_{t-1} (Table 2). Model 1.1b is a modification of Model 1a which includes the variable $SalesVol_{t-1}$ (the number of homes sold in the previous period). The alternative model or Model 2b is enriched with data from the GT query of 'mortgage'. The results indicate a weak but statistically significant negative effect of the number of queries in the period t-2, and a stronger and positive effect for t = 0, on the HPI. After adding $SalesVol_{t-1}$ to the model, the 'mortgage' ceases to be statistically significant, and holds irrespective of the number of lags (Model 3b). This result suggests the limited usefulness of 'mortgage' in modeling the HPI. The consequences of this query could be too spread out in time to be a valuable source of information for the model. In addition, an increase in the number of 'mortgage' queries may indicate negating effects. Both negative and positive factors can be at play here including increasing problems with the repayment of loans and searches for sources of finance). Next, the analysis focuses on using a more explicit GT query. The 'mortgage' query in Model 3.1b is replaced by the ‘house for sale’ query. This query is associated with a large and significant increase in the HPI during the same period of time.

Table 2 Estimation of HPI, 2004 – 2014

	HPI[1]	HPI	HPI	HPI	HPI
Model	1b	1.1b	2b	3b	3.1b
HPI_{t-1}	0.961***	0.941***	0.959***	0.944***	0.946***
$SalesVol_{t-1}$		0.00001***		0.00001***	0.001***
Mortgage			0.042***		
$Mortgage_{t-1}$			-0.037**		
House for sale					0.047***
Adjusted R-Squared	0.28	0.15	0.33	0.34	0.22
AIC	1.4055	1.564	1.33	1.33	1.68

Notes: [1] *, **, *** indicates the levels of significance of 10%, 5%, and 1%. Only significant coefficients are reported.

The Granger causality test was performed for all of the explanatory variables. The results reported in Table A2 in the Appendix show the likelihood of causal relationships in most cases. As part of the auxiliary analysis, the HPI models are used to test if additional macroeconomic variables, in particular, monthly earnings and the unemployment rate, affect the main estimates. The results, reported in Tables A6 and A7, show that these variables do not play a major role in improving the models.

3.1 Predictions

The next step in the analysis is to compare the ability of the HPI models to predict prices. Monthly data for 2004-2007 are used to calibrate the base models. This will be followed by estimating the monthly predictions for 2008 - 2014. Based on the results of the previously discussed studies on the US real estate market, it is anticipated that the alternative models (Models 2b and 3b) will have a significantly higher accuracy thanks to the additional information from the GT data. The mean absolute error (MAE) is used to compare the accuracy of the base and alternative models. This measure will provide information about the mean deviation of the monthly forecasts from the actual values.

$$MAE = \frac{1}{N}\sum_{t=1}^{N}\left|\frac{y_t - \hat{y}_t}{y_t}\right| \quad (4)$$

Table 3 presents the MAEs calculated for the base and alternative HPI models.

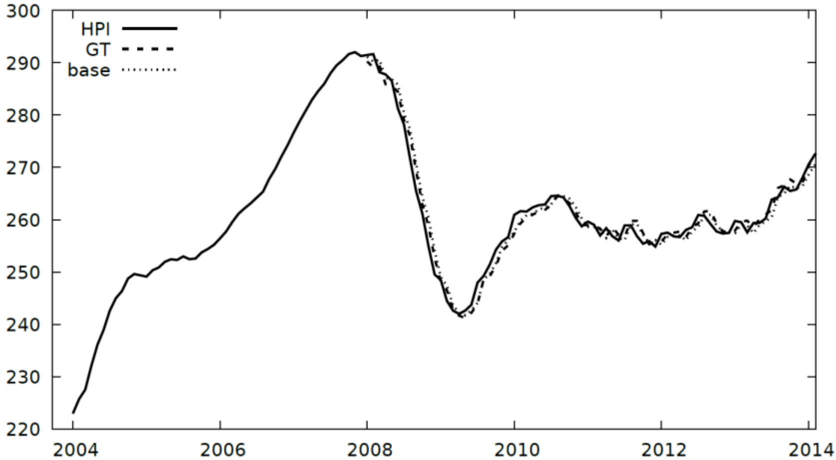
Table 3 MAE For Selected Models

Prediction from	MAE	MAE	%
Models	1.1b	3.1b GT: ‘House for sale’	ratio
2008	1.64	1.53	0.93
2007	1.54	1.42	0.92
Models	1.1b	3.b GT: ‘Mortgage’	
2008	1.64	1.61	0.97

Note: MAE based on monthly predictions for 2008 - 2014

As expected, the results presented in Table 3 show that the predictions based on the models with GT data are more accurate. The forecast error of the base model, that is, Model 1.1b, is 1.64% and 1.53% for the alternative model, that is, Model 3.1b. In the case of the latter, there is a reduction in the MAE of about 7.5%. The enhanced accuracy of the alternative models fluctuate between 2.5% - 8% depending on the query itself, and the start date of the prediction (2007 or 2008). The differences in the percentage of the MAE between the base and alternative models show that the GT data enhance the accuracy of predicting the HPI to a similar extent as the data on the volume of sales in the previous time period (month). Previous studies on the application of GT data for the real estate market in the US indicate an increase in the accuracy of estimation from 2.3% (Wu and Brynjolfsson, 2013) to 12% (Choi and Varian, 2012). Wu and Brynjolfsson (2013) obtain the highest improvement in accuracy or 7.1% for the 'real estate agencies' query, while Choi and Varian (2012) report 12% for their `Rental Listings & Referrals' query. Figure 3 shows the differences between the predicted and actual values for Models 1.1b and 3.1b. The differences for 2008-2009 are relatively small. However, these differences increase at points where the trend of the HPI changes, for example, during the first half of 2009 and the second half of 2010.

Figure 3 **Forecasting HPI with '*House for Sale*' GT Query**



Note: base = Model 1.1b and GT = Model 3.1b.

4. Discussion

The results of this study confirm that the GT data have a significant impact on enhancing the accuracy of estimations of the HPI for the UK during 2007-2014. Accurate predictions by the alternative models (GT) produce forecasts that are on average 8% more accurate relative to the base models without the use of GT data. This result is in line with previous studies on the real estate market in the US (Choi and Varian, 2012). It is likely that any differences in the results between this study and previous studies are due to the selection of different GT queries and, to some extent, the differences between the UK and US housing markets. This selection is crucial, and the query word or phrase which precedes investment decisions is likely to change with time and location as shown in a recent study by Oust and Eidjord (2020) done in the US. Enhancing the accuracy of predictions depends largely on the selection of appropriate queries. Moreover, it is possible based on evidence from Oust and Eidjord (2020) that a specific query will enhance forecasting only during a given phase of the market. This paper provides evidence that for the UK, as is the case for the US, search engine data can be used as a reliable means to improve forecasting for a much smaller real estate market. Furthermore, it is also shown that this type of data is robust enough to be used in times when the markets experience substantial variation. GT data are an interesting alternative to other data sources where their collection is often time-consuming and costly. The main advantages of using queries from web search engines are efficiency and availability. In addition, GT data provide timely information about changes in searches in countries, regions and cities, thus facilitating analysis in various regional settings. The richness of GT data allows options of queries associated with a given phenomenon, for example, the demand for a particular product. This analysis shows the need for further research into the possible application of GT data for other markets and countries. This type of data also provides valuable insight into studying the effectiveness of advertising campaigns. The high capacity of GT data allows weekly changes to be tracked in popular queries. Specific applications of the web search data in forecasting can apply to countries where data on the current state of the economy are published with a considerable delay or where data credibility is low.

It should be noted that this study has its limitations. The choice of queries may be changing in time and space. Particular queries could be less relevant in different regions even if the same language is spoken. Furthermore, new developments and tools may render some queries less relevant or even obsolete. Another aspect often ignored previously is that increased automation changes the way that internet users search for information. With the increased potential of personalization and automatic notices, more individuals may simply skip the search engine stage. This problem may be accounted for in future studies by using additional sources of information on automated or personalized internet activity. Furthermore, it should be of interest to scholars in this field to examine the usefulness of the search engine data at a less aggregate level, with particular focus on rural/urban differences. City or state-level analyses are however beyond the scope of this paper, but could be of interest to studies on emerging economies. It is likely that city, rural/urban level differences may provide new findings. The underdevelopment of internet resources might be an important issue for such studies which would require care in analysis as in some localities, the reduced availability in specific settings may result in unreliable search engine data. Finally, researchers in this field who conduct related future studies should pay particular attention to differences in the popularity of the Google search engine with time and space.

5. Conclusions

This investigation shows that data on online search activity can be reliably used to improve predictions of the real estate market. The usefulness of Google data in predicting the UK real estate market after the recent financial crisis is discussed here. The data may prove to be an additional source of information for individual buyers, investors and governments. This analysis is an important addition to the existing literature that explores the role of behavioral factors in decision making processes. The results of this study shed new light on the behavior of markets. GT data are identified as an important predictor of real estate market changes.

References

Appleton-Young, L. (2008). *State of the California housing market 2008–2009*. California: California Association of Realtors.

Arrow, K. J. (1987). Reflections on the Essays, *Arrow and the Foundations of the Theory of Economic Policy*, 727 - 734. doi: https://doi.org/10.1007/978-1-349-07357-3_36.

Askitas, N., and Zimmermann, K. F. (2009). Google Econometrics and Unemployment Forecasting, *DIW Berlin Discussion Paper No. 899*. doi: https://dx.doi.org/10.2139/ssrn.1465341.

Castle, J. L., Fawcett, N. W. P., and Hendry, D. F. (2009). Nowcasting is not just contemporaneous forecasting, *National Institute Economic Review*, 210, 71 - 89. doi: https://doi.org/10.1177/0027950109354412.

Cervellin, G., Comelli, I., and Lippi, G. (2017). Is Google Trends a reliable tool for digital epidemiology? Insights from different clinical settings, *Journal of Epidemiology and Global Health*, 7(3), 185 - 189. doi: https://dx.doi.org/10.1016/j.jegh.2017.06.001.

Choi, H., and Varian, H. (2012). Predicting the Present with Google Trends, *Economic Record*, 88, 2 - 9. doi: https://doi.org/10.1111/j.1475-4932.2012.00809.x.

ComScore (2012). *qSearch: A comprehensive view of the search landscape*, *Technical report*, *comScore*. https://www.comscore.com/Insights/Press-Releases/2012/11/comScore-Releases-October-2012-US-Search-Engine-Rankings

Cooper, C. P., Mallon, K.P., Leadbetter, S., Pollack, L.A., and Peipins, L.A.. (2005). Cancer internet search activity on a major search engine, United States 2001-2003, *Journal of Medical Internet Research*, 7(3). doi: https://doi.org/10.2196/jmir.7.3.e36.

Ettredge, M., Gerdes, J., and Karuga, G. (2005). Using web-based search data to predict macroeconomic statistics, *Communications of the ACM*, 48(11), 87 - 92. doi: https://doi.org/10.1145/1096000.1096010.

Eysenbach, G. (2006). Infodemiology: tracking flu-related searches on the web for syndromic surveillance, *AMIA Annual Symposium proceedings*. *AMIA Symposium*, 2006, 244 - 248

Flanagan, R., Kuo, B., and Staller, K. (2021). Utilizing Google Trends to Assess Worldwide Interest in Irritable Bowel Syndrome and Commonly

Associated Treatments, *Digestive Diseases and Sciences*, 66(3), 814 - 822. doi: https://doi.org/10.1007/s10620-020-06290-7.

Goel, S., Hofman, J.M., Lahaie, S., Pennock, D.M., and Watts, D.J. (2010). Predicting consumer behavior with web search, *Proceedings of the National Academy of Sciences of the United States of America*, 107(41), 17486 - 17490. doi: https://doi.org/10.1073/pnas.1005962107.

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. doi: https://doi.org/10.3233/JEM-2011-0342.

Hulth, A., Rydevik, G., and Linde, A. (2009). Web queries as a source for syndromic surveillance, *PLoS ONE*, 4(2), e4378. doi: https://doi.org/10.1371/journal.pone.0004378.

Jun, S. P., Yoo, H. S., and Choi, S. (2018). Ten years of research change using Google Trends: From the perspective of big data utilizations and applications, *Technological Forecasting and Social Change*, 130, 69 - 87. doi: https://doi.org/10.1016/j.techfore.2017.11.009.

Krugman, P. (2009). How did economists get it so wrong?, *New York Times*.

Lin, T. C., and Hsu, S. H. (2020). Forecasting Housing Markets from Number of Visits to Actual Price Registration System, *International Real Estate Review*, 23(4), 505–536. doi: https://doi.org/10.53383/100312.

National Association of Realtors (2012). *Profile of home Buyers and Sellers 2012*. https://www.nar.realtor/sites/default/files/2012-profile-international-home-buying-activity-2012-06.pdf

Oust, A., and Eidjord, O. M. (2020). Can google search data be used as a housing bubble indicator?, *International Real Estate Review*, 23(2), 267 - 308. doi: https://doi.org/10.53383/100302.

Polgreen, P. M., Chen, Y., Pennock, D.M., Nelson, F.D., and Weinstein, R.A. (2008). Using Internet Searches for Influenza Surveillance, *Clinical Infectious Diseases*, 47(11), 1443 - 1448. doi: https://doi.org/10.1086/593098.

Preis, T., Moat, H. S., and Stanley, H. E. (2013). Quantifying Trading Behavior in Financial Markets Using Google Trends, *Scientific Reports*, 3(1), 1 - 6. doi: https://doi.org/10.1038/srep01684.

Salisu, A. A., Ogbonna, A. E., and Adewuyi, A. (2020). Google trends and the predictability of precious metals, *Resources Policy*, 65, 101542. doi: https://doi.org/10.1016/j.resourpol.2019.101542.

Seymour, T., Frantsvog, D., and Kumar, S. (2011). History of Search Engines, *International Journal of Management & Information Systems (IJMIS)*, 15(4), 47 - 58. doi: https://doi.org/10.19030/ijmis.v15i4.5799.

Silva , E. S., Hassani, H., Madsen, D. Ø., and Gee, L. (2019). Googling Fashion: Forecasting Fashion Consumer Behaviour Using Google Trends, *Social Sciences*, 8(4), 111. doi: https://doi.org/10.3390/socsci8040111.

Simon, H. A. (1984). On the behavioral and rational foundations of economic dynamics, *Journal of Economic Behavior and Organization*, 5(1), 35 - 55. doi: https://doi.org/10.1016/0167-2681(84)90025-8.

Strzelecki, A. (2020). The second worldwide wave of interest in coronavirus since the COVID-19 outbreaks in South Korea, Italy and Iran: A Google Trends study, *Brain, behavior, and immunity*, 88, 950–951. doi: https://doi.org/10.1016/j.bbi.2020.04.042.

Teng, Y., Bi, D., Xie, G., Jin, Y., Huang, Y., Lin, B., An, X., Feng, D., and Tong, Y. (2017). Dynamic Forecasting of Zika Epidemics Using Google Trends, *PLoS ONE*, 12(1), e0165085. doi: https://doi.org/10.1371/journal.pone.0165085.

Wu, L., and Brynjolfsson, E. (2013). The Future of Prediction: How google Searches Foreshadow Housing prices and quantities. Available at SSRN: https://ssrn.com/abstract=2022293 or http://dx.doi.org/10.2139/ssrn.2022293

Wu, L., and Brynjolfsson, E. (2015). Chapter 3 - The Future of Prediction, in *Economic Analysis of the Digital Economy*. University of Chicago Press, 89–118. doi: https://doi.org/10.7208/chicago/9780226206981.003.0003.

Yang, S., Santillana, M., and Kou, S. C. (2015). Accurate estimation of influenza epidemics using Google search data via ARGO, *Proceedings of the National Academy of Sciences of the United States of America*, 112(47), 14473 - 14478. doi: https://doi.org/10.1073/pnas.1515373112.

Yu, L., Zhao, Y., Tang, L., and Yang, Z. (2019). Online big data-driven oil consumption forecasting with Google trends, *International Journal of Forecasting*, 35(1), 213 - 223. doi: https://doi.org/10.1016/j.ijforecast.2017.11.005.

Appendix

Table A1 Unit-root Test Results

Variable	ADF (n0= unit-root)
HPI	p-value =0.001
SalesVol	p-value =0.034
Mortage	p-value =0.005
House_for_Sale	p-value =0.001
Rentals	p-value =0.001
Unemployment	p-value =0.010
Earnings	p-value =0.001

Note: ADF unit-root tests for all variables considered in the estimation. All tests conducted with 12 lags as frequency of data is monthly.

Table A2 Granger Causality Test, HPI and Sales Volume

Outcome:	**SalesVol t-1**		**Mortgage t-1**		**House_for _sale t-1**		**Rentals**		**HPIt-1**	
	F-test	p-val	F-test	p-val	F-stat	p-val	F-test	p-val	F-test	p-val
HPI	12.082	0.0006	15.323	0.001	6.68	0.0109				
SalesVol							16.090	0.0001	15.221	0.0002

Table A3 Lag Length, HPI and Sales Volume

Equation 1: d_houseprices	***Coefficient***	***Std. Error***	***t-ratio***	***p-value***	
const	−0.0240596	0.0475412	−0.5061	0.6142	
d_houseprices_1	−0.480669	0.101409	−4.740	<0.0001	***
d_houseprices_2	−0.402761	0.113104	−3.561	0.0006	***
d_houseprices_3	−0.351691	0.122412	−2.873	0.0052	***
d_houseprices_4	−0.206559	0.125648	−1.644	0.1041	
d_houseprices_5	−0.0851553	0.125732	−0.6773	0.5002	
d_houseprices_6	0.0298479	0.122421	0.2438	0.8080	
d_houseprices_7	−0.0108329	0.122020	−0.08878	0.9295	
d_houseprices_8	0.106886	0.122891	0.8698	0.3870	
d_houseprices_9	0.0981398	0.123178	0.7967	0.4279	
d_houseprices_10	−0.00410521	0.118656	−0.03460	0.9725	
d_houseprices_11	0.0116696	0.110105	0.1060	0.9159	
d_houseprices_12	−0.152852	0.100133	−1.526	0.1308	

(Continued…)

(Table A3 Continued)

d_houseforsale	0.0360872	0.0952483	3.789	0.0003	***
d_houseforsale_1	0.00105559	0.106657	0.009897	0.9921	
d_houseforsale_2	0.0458748	0.105761	0.4338	0.6656	
d_houseforsale_3	0.0347034	0.113820	0.3049	0.7612	
d_houseforsale_4	−0.114731	0.118262	−0.9701	0.3349	
d_houseforsale_5	−0.162234	0.116881	−1.388	0.1689	
d_houseforsale_6	−0.0581775	0.119113	−0.4884	0.6266	
d_houseforsale_7	−0.123073	0.118055	−1.043	0.3003	
d_houseforsale_8	−0.170669	0.118128	−1.445	0.1524	
d_houseforsale_9	−0.130730	0.114809	−1.139	0.2582	
d_houseforsale_10	0.0208491	0.111356	0.1872	0.8519	
d_houseforsale_11	−0.157587	0.107047	−1.472	0.1449	
d_houseforsale_12	−0.00961479	0.0992182	−0.09691	0.9230	
d_sales_vol	−6.37002e-06	5.87205e-06	−1.085	0.2812	
d_sales_vol_1	1.29004e-05	6.05650e-06	2.130	0.0362	**
d_sales_vol_2	1.15032e-05	6.05166e-06	1.901	0.0609	*
d_sales_vol_3	9.96380e-06	6.02468e-06	1.654	0.1020	
d_sales_vol_4	6.51410e-06	6.07689e-06	1.072	0.2869	
d_sales_vol_5	9.51835e-07	6.03978e-06	0.1576	0.8752	
d_sales_vol_6	−4.82858e-06	5.76101e-06	−0.8381	0.4044	
d_sales_vol_7	7.23508e-06	5.75130e-06	1.258	0.2120	
d_sales_vol_8	4.62221e-06	5.74208e-06	0.8050	0.4232	
d_sales_vol_9	1.04617e-05	5.67642e-06	1.843	0.0690	*
d_sales_vol_10	7.33153e-06	5.44790e-06	1.346	0.1821	
d_sales_vol_11	6.69002e-06	5.49869e-06	1.217	0.2273	
d_sales_vol_12	−2.78959e-06	5.51458e-06	−0.5059	0.6143	

Mean dependent var	−0.025000	S.D. dependent var	0.511095
Sum squared resid	11.58017	S.E. of regression	0.378107
R-squared	0.627468	Adjusted R-squared	0.452699
F(38, 81)	3.590283	P-value(F)	7.03e-07
rho	−0.069473	Durbin-Watson	2.109908

Note: *, **, *** indicates the levels of significance of 10%, 5%, and 1%.

Table A4 Lag Length Selection, HPI, Sales Volume and Mortgage

Equation 1: d_houseprices	***Coefficient***	***Std. Error***	***t-ratio***	***p-value***	
const	−0.0419583	0.0449809	−0.9328	0.3537	
d_houseprices_1	−0.529988	0.108666	−4.877	<0.0001	***
d_houseprices_2	−0.408384	0.123917	−3.296	0.0015	***

(Continued…)

(Table A4 Continued)

d_houseprices_3	−0.313747	0.124446	−2.521	0.0137	**
d_houseprices_4	−0.173824	0.127644	−1.362	0.1770	
d_houseprices_5	−0.133325	0.124238	−1.073	0.2864	
d_houseprices_6	−0.0916922	0.121770	−0.7530	0.4536	
d_houseprices_7	−0.0908494	0.125383	−0.7246	0.4708	
d_houseprices_8	0.00879572	0.127426	0.06903	0.9451	
d_houseprices_9	0.137484	0.128084	1.073	0.2863	
d_houseprices_10	0.226122	0.129318	1.749	0.0842	*
d_houseprices_11	0.0502786	0.122494	0.4105	0.6826	
d_houseprices_12	−0.0623174	0.105892	−0.5885	0.5578	
d_sales_vol	7.54221e-06	5.86285e-06	1.286	0.2020	
d_sales_vol_1	7.31034e-06	5.54597e-06	1.318	0.1912	
d_sales_vol_2	1.23617e-05	5.62767e-06	2.197	0.0309	**
d_sales_vol_3	2.24051e-06	5.92339e-06	0.3782	0.7062	
d_sales_vol_4	−3.87002e-08	5.91382e-06	−0.006544	0.9948	
d_sales_vol_5	−4.63714e-06	5.72727e-06	−0.8097	0.4205	
d_sales_vol_6	−2.83861e-06	5.59744e-06	−0.5071	0.6134	
d_sales_vol_7	8.97620e-06	5.60173e-06	1.602	0.1130	
d_sales_vol_8	6.41332e-06	5.61839e-06	1.141	0.2570	
d_sales_vol_9	8.20424e-06	5.49657e-06	1.493	0.1394	
d_sales_vol_10	1.05854e-05	5.32246e-06	1.989	0.0501	*
d_sales_vol_11	−6.74978e-07	5.28330e-06	−0.1278	0.8987	
d_sales_vol_12	−3.14576e-06	5.32763e-06	−0.5905	0.5565	
d_mortgage	0.0891332	0.0292016	3.052	0.0031	***
d_mortgage_1	0.0300138	0.0368342	0.8148	0.4176	
d_mortgage_2	0.0457034	0.0398249	1.148	0.2545	
d_mortgage_3	−0.0233672	0.0402483	−0.5806	0.5631	
d_mortgage_4	−0.0351838	0.0413430	−0.8510	0.3973	
d_mortgage_5	−0.0138884	0.0399217	−0.3479	0.7288	
d_mortgage_6	0.0147483	0.0380067	0.3880	0.6990	
d_mortgage_7	0.0325056	0.0377137	0.8619	0.3913	
d_mortgage_8	−0.0124284	0.0375757	−0.3308	0.7417	
d_mortgage_9	−0.0522990	0.0366070	−1.429	0.1569	
d_mortgage_10	−0.0565436	0.0361878	−1.563	0.1221	
d_mortgage_11	−0.0224512	0.0352872	−0.6362	0.5264	
d_mortgage_12	0.0243820	0.0308436	0.7905	0.4315	

Mean dependent var	−0.025000	S.D. dependent var	0.511095
Sum squared resid	12.13577	S.E. of regression	0.387071
R-squared	0.609594	Adjusted R-squared	0.426441
F(38, 81)	3.328324	P-value(F)	2.86e-06
rho	−0.029662	Durbin-Watson	2.034660

Note: *, **, *** indicates the levels of significance of 10%, 5%, and 1%.

Table A5 Lag Length Selection Compact, HPI and Sales Volume

Equation 1: d_houseprices	*Coefficient*	*Std. Error*	*t-ratio*	*p-value*	
const	−0.0484929	0.0377249	−1.285	0.2011	
d_houseprices_1	−0.371564	0.0884049	−4.203	<0.0001	***
d_houseprices_2	−0.247426	0.0798621	−3.098	0.0024	***
d_houseprices_3	−0.144145	0.0798917	−1.804	0.0737	*
d_houseforsale	0.403507	0.0585202	6.895	<0.0001	***
d_houseforsale_1	0.122840	0.0706748	1.738	0.0848	*
d_sales_vol	−1.00585e-05	3.99827e-06	−2.516	0.0132	**
d_sales_vol_1	2.60233e-06	3.79225e-06	0.6862	0.4939	
d_sales_vol_2	−1.89710e-07	3.67019e-06	−0.05169	0.9589	
d_sales_vol_3	4.88407e-06	3.88992e-06	1.256	0.2117	

Mean dependent var	−0.010078	S.D. dependent var	0.551857
Sum squared resid	21.42054	S.E. of regression	0.424269
R-squared	0.450500	Adjusted R-squared	0.408942
F(9, 119)	10.84007	P-value(F)	3.54e-12
rho	−0.082383	Durbin-Watson	2.163740

Note: *, **, *** indicates the levels of significance of 10%, 5%, and 1%.

Table A6 HPI, Mortgage and Earnings and Unemployment

HPI	***Coefficient***	***p-value***	
d_houseprices_1	0.921645	0.0001	***
d_sales_vol	3.89127e-06	0.3416	
d_mortgage	0.003473	<0.0001	***
d_earnings	0.0389604	0.6163	
d_unempl. rate	0.0184662	0.9602	

Mean dependent var	−0.012500	S.D. dependent var	0.565529
Sum squared resid	27.60471	S.E. of regression	0.468065
R-squared	0.341127	Adjusted R-squared	0.314981
F(5, 126)	13.04714	P-value(F)	3.10e-10
rho	−0.146203	Durbin-Watson	2.281483

Note: *, **, *** indicates the levels of significance of 10%, 5%, and 1%.

Table A7 HPI, House for Sale, Earnings and Unemployment

	Coefficient	*p-value*	
d_houseprices_1	0.945605	0.0002	***
d_sales_vol	1.39009e-05	0.0002	***
d_houseforsale	0.04056	<0.0001	***
d_earnings	0.0592272	0.4411	
d_unempl. rate	0.0950721	0.7948	
Mean dependent var	−0.012500	S.D. dependent var	0.565529
Sum squared resid	26.97754	S.E. of regression	0.462717
R-squared	0.356097	Adjusted R-squared	0.330545
F(5, 126)	13.93630	P-value(F)	7.75e-11
rho	−0.123315	Durbin-Watson	2.229009

Note: *, **, *** indicates the levels of significance of 10%, 5%, and 1%.