MACROECONOMIC ANALYSIS

Peru | Internet searches and sales projections of departments in Lima

Vanessa Belapatiño and Yalina Crispin

This report presents a purchasing interest index regarding the buying of apartments for the city of Lima (IICDG), which has been developed using information provided by Google Trends. There is evidence that the IICDG anticipates movements (by five quarters) and that this improves its ability to predict transactions in the real estate sector. Finally, using the IICDG in a projection model for sales of apartments, it is expected that these sales will remain stable during 2017.

Web searches, Google Trends and the analysis of the behaviour of economic agents

Internet searches contain relevant information about the preferences and intentions of millions of buyers and businesses, and its digitization is enabling its systematic use. Google, one of the main information search engines on the Internet at a global level, has played a key role in this process by channelling and consolidating the diverse data available on to various web pages to facilitate access of all users.

A free tool from Google is Google Trends, which provides information about the number of searches for a word or phrase. This information is provided by a weekly frequency index and it is possible to limit the search to a particular geographical region. The index is constructed on the basis of two processes. The first is standardization: the number of searches for the word or phrase is divided by the total searches carried out on Google, controlling for the geographical area and time interval represented. The aim of this process is to isolate the data from the effects of the growth in the number of internet users and the increase in Google's popularity as a search engine[1], so it becomes an indicator of interest. The second process corresponds to scaling: the highest number of searches throughout the period of analysis is given a value of 100, so the index is limited to values between 0 and 100.

Two important features of the information provided by Google Trends: it is real-time, high frequency, allowing for the updating of query data, unlike other sources that provide monthly or quarterly information which is subject to a certain period of delay.

Recent studies show that Internet search data are useful for carrying out projections in various disciplines

In recent years, several studies have shown the effectiveness of using data from Google to improve projections in the fields of epidemiology and the economy. Regarding the first discipline, among the best-known works are those of Polgreen et al (2008) and Ginsberg et al (2009), which showed that internet search data help predict the incidence of diseases such as influenza. These works were widely disseminated and stimulated additional findings in epidemiology.

[1] Carrie - Swallow and Labbé (2010).

In the field of economy, Choi and Varian (2009 and 2011) describe how to use Google searches to predict economic variables such as retail sales, sales of cars, homes and travel[2]. The authors found that models incorporating the Google Trends index improve their predictive capability.

Other authors have examined the use of internet information for labour market analysis. For example, Askitas and Zimmerman (2010) use internet information to predict unemployment in Germany. Along the same lines, mention may also be made of the work of Baker and Fradkim (2011) on the dynamics of employment in the United States and of Chang and Rio (2013), who estimated a prediction model for the level of employment for Peru.

On the other hand, Pena and Huang (2009) built a consumer confidence index for the United States on the basis of changes in search patterns in Google. They found that the index they built highly correlated with the consumer confidence index of the University of Michigan (ICS), and also with the consumer confidence index of the Conference Board (CCI). The authors also find that the Google index foresees the changes in the two reference indices. It also found that the Google index tends to perform better than both the ICS and the CCI at predicting growth in personal consumer spending and retail sales.

Other works that use Internet search information are related to the housing market. Wu and Brynjolfsson (2009) constructed a search index of US housing that allows them to predict sales in the real estate market and prices. In Askitas and Zimmermann (2011), there is evidence of the relationship between mortgage delinquencies and Internet search requests for loan modifications, (a borrower writes to his or her bank when they have problems and they externalise their economic decline).

As can be seen, the use of information searches on the internet has several applications and extensive development potential for various disciplines such as economics.

Development of an index measuring Apartment Purchase Interest for Lima with information from Google Trends (IICDG)

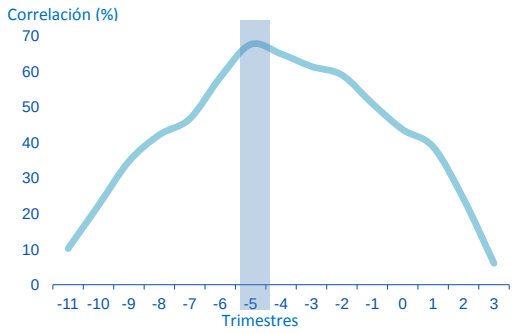
The construction of the IICDG for Lima has proceeded as follows:

1. For Internet searches, what was defined was a list of phrases related to being interested in purchasing a department. Specifically, the five phrases that showed the greatest representativeness were selected[3]: (i) "apartments sale Lima", (ii) "apartments for sale in Lima", (iii) "sale of apartments in Lima", (iv) "real estate in Lima" and (v) "estate agencies in Lima".
2. The selected series change - through simple averages - on a monthly or quarterly basis[4]. These changes are carried out to match the IICDG's frequency of apartment sales in Lima (which is only available on a quarterly basis)[5], to then analyse if there is a relationship between these two variables.
3. The five series were weighted by the inverse of their standard deviations to reduce the importance of their volatility and uncertainty in the IICDG (for more detail see Appendix 1).

[2] Hal Varian is Google Inc.'s Chief Economist.
[3] Because Google Trends shows the exact search words or phrases, some of the selected phrases are similar.
[4] The average for July and August was considered to be data corresponding to the third quarter of 2016.
[5] The series of apartment sales in Lima prepared by TINSA was considered (information available from the first quarter of 2008).

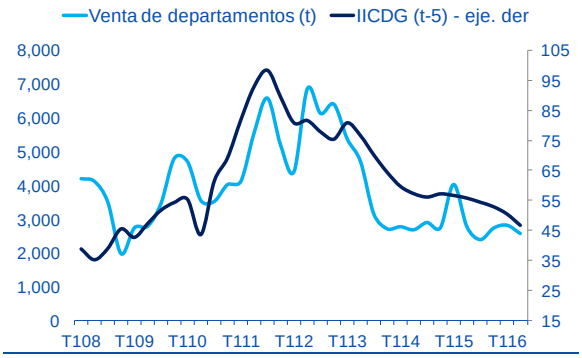
As can be seen in Figure 1, the dynamic correlation analysis reveals that there is a positive relationship between the IICDG and the sale of apartments[6], reaching its highest association when the IICDG is under way over five quarters (see Figure 1). Thus, by plotting the sale of apartments and with the IICDG having a lag of five quarters[7], a similar behaviour in both series (see Figure 2) is shown. In addition, when testing statistical precedence[8], it was found that the IICDG anticipates (cause) the behaviour of apartment sales in Lima[9] (see Appendix 2).

Figure 1
Correlation coefficient between Sales of apartments in Lima and the IICDG with different lags



Source: TINSA, Google
Preparation: BBVA Research

Figure 2
Sale of apartments in Lima and IICDG (number of units and index*)



*In the first quarter of 2010, the highest level of searches was recorded, so the index takes the value of 100.
Source: TINSA, Google
Preparation: BBVA Research

With the above, we found evidence that apartment purchasing interest, measured by the IICDG, anticipates the completion of the sale. This result is consistent in practice due to the fact that purchasing an apartment requires time to evaluate factor such as: location, price, features to suit the buyer, which estate agent to go to, the time needed to save for the initial deposit and which bank to go to for a loan, etc.

As well as the time needed to search for an apartment, you must add the time it takes you to get the bank to approve a mortgage. The assessment does not take a long time, if the buyer is a dependent employee of average income, but this is not always the case. There is another longer evaluation process which is used when the worker is self-employed. In this case, because the buyer cannot guarantee a permanent income, the banking system gives them the possibility of accessing a loan, if they maintain a constant deposit for at least 6 months.

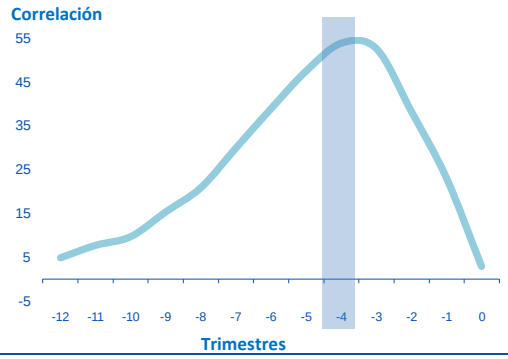
In this context, the relationship between the index of interest regarding purchasing an apartment (IICDG) and mortgage lending growth is analysed. It is evident that the IICDG also anticipates the dynamics of mortgage lending. The lagging of mortgage loans over 4 quarters, is the highest correlation found (see Figure 3 and 4)**.** Also, when performing tests of statistical precedence it was found that the IICDG anticipates (cause) the growth of mortgage loans (see Appendix 2).

Figure 3
Correlation coefficient between the growth in

Figure 4
Mortgage lending in the banking system and the

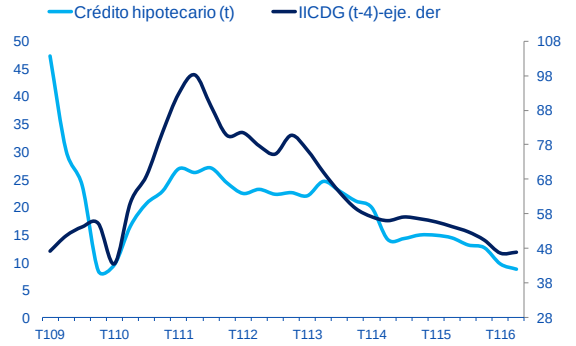
[6] The two series are seasonally adjusted.
[7] For example, the data for the sale of apartments from the second quarter of 2016 is associated with the value of the IICDG from the first quarter of 2015; i.e. the value of the IICDG anticipates the behaviour of apartment sales.
[8] Testing Granger causality. See details in Appendix 2.
[9] However, it is important to note that the statement that the "IICDG causes the sale of apartments in Lima" does not imply that the selling of apartments in Lima is the effect or the result of the IICDG, therefore, other factors also intervene.

mortgage lending in the banking system and the IICDG



Source: Google, ASBANC
Preparation: BBVA Research

IGCD (var.% YoY and Index*)

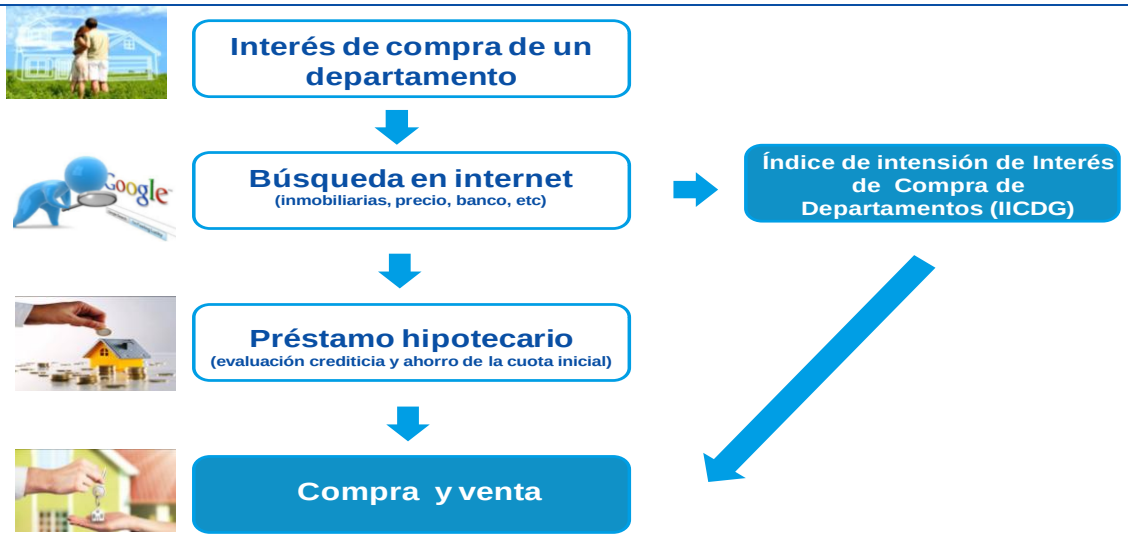


*In the first quarter of 2010, the highest level of searches was recorded, so the index takes the value of 100. Source: ASBANC & Google
Preparation: BBVA Research

In summary, the results show evidence of the relationship between Internet searches, which reflects the interest families have in buying an apartment, and in the sale of this one (the completion of the transaction).

Figure 5

Relationship between the interest of families in buying an apartment (measured by the IGCD) and purchase and sale



Preparation: BBVA Research

The index of purchasing interest in apartments in Lima (IICDG) improves the ability to forecast sales of apartments

To assess whether the use of the IICDG improves the predictive capacity of apartment sales, two econometric models were estimated (See Appendix 3): one with an ARMA specification (Reference Model) and another that is included in the IICDG (Augmented Model). It was found that the predictive ability of the second model is better than the reference one. As can be seen in Figure 1, the model that is included in the IICDG presents a lower mean square error (ECM) and a lower mean absolute error (EMA) than the reference model. In addition, the R^2 increases slightly in the Augmented Model. Also, when the estimate is carried out outside the sample in the Augmented Model, there are lower ECM and EMA values, 46% and 52% respectively, with respect to the Reference Model.

Table 1
Predictive capacity indicators of the two projection models

Indicadores	Modelo de referencia	Modelo aumentado
R-squared	81	83
Dentro de la muestra		
ECM 1/	575	**539**
EMA	464	**385**
Fuera de de la muestra		
ECM 1/	1,485	**804**
EMA	1,276	**612**

*1/ This value is considered to be the root of the mean quadratic error (ECM).
Preparation: BBVA Research

The level the IICDG has shown so far this year is consistent with a stable sales level during 2017

Interest in purchasing departments, measured by the IICDG, stopped its slowdown in the third quarter of 2015 and remained stable until August this year. Considering the time lapse between showing interest and purchasing is five quarters (as evidenced in the previous sections), the evolution of the IICDG over the year suggests that the level of sales of apartments will remain relatively stable in the remainder of 2016 and 2017.

The Model of projection that is included in the IICDG (described in the previous section) also points to the fact that the sale of apartments in Lima will stop its deceleration and maintain levels similar to current ones in the coming months (see Figure 6).

Figure 6
Sale of apartments in Metropolitan Lima: results of the augmented model * (in number of units)

*Projection outside the sample.
Source: TINSA
Preparation: BBVA Research

It should be added that some recent events, which are still not being captured by the IICDG, could give further impetus to the real estate market in the following months. For example, regulatory changes to pension funds (an amount being withdrawn for the purchase of new housing[10]).

Conclusions

Information searches on the internet enable you to forecast the real estate market over the short-term. For this, a purchase interest indicator was constructed (IICDG), via Google Trends, which anticipates, by five quarters, the behaviour of future sales of the latter.

The index of interest in purchasing apartments in Lima (IICDG) improves the ability to forecast the sales of apartments An estimate of two econometric models has been put forward: one with an ARMA specification (Reference Model) and another that includes information from the IICDG (Augmented Model). It was found that the predictive ability of the Augmented Model is better than the reference one.

The level that the IICDG has shown so far this year is consistent with a stable level of sales during the 2017. The interest in the purchase of apartments stopped its slowdown in the third quarter of 2015 and remained stable until August of this year. Considering the time lapse between the interest in and purchase of an apartment is five quarters, the level of sales of apartments will remain stable for the rest of 2016 and 2017. It should be added that some events could boost the real estate market, as is the case of the liberalization of the funds from pension funds for the acquisition of housing.

[10] In June 2016, the Law on Pension Funds was promulgated and this will drive the purchase of homes. Law No. 30478 allows for the withdrawal of up to 25% from the fund of affiliates of an AFP, with the condition that this money is allocated to the initial down payment for the purchase of a home or to the advance payment of a mortgage loan, both for a first home. It is estimated that around 1 million people will benefit from the law.

Appendix 1: Construction of the Google Index for Apartment Purchasing

For the construction of the Google index for Apartment Purchasing (IICDG), the five series obtained from the searches in Google Trends were indexed with the following phrases: "apartments sale Lima", "apartments for sale in Lima", "sale of apartments in Lima", "real estate in Lima" and "estate agents in Lima"

The first step for indexing was to calculate the standard deviation of each of the series that is represented by each one of the five phrases. See equation 1 and 2, where σ_n represents the deviation of σ_n representa la desviación de n, being σ_n representa la desviación de n, siendo n one of the 5 series and σ_n representa la desviación de n, siendo n una de las 5 series y t one of the quarters.

$$\sigma_n = \sqrt{\frac{\sum_{t=1}^{T}(x_{tn} - \bar{x}_n)^2}{T-1}} \qquad \forall \quad t = 1\ a\ T\ , \qquad n = 1\ a\ N \tag{1}$$

$$donde \qquad \bar{x}_n = \frac{\sum_{t=1}^{T} x_{tn}}{T} \tag{2}$$

Then, each weighting factor is calculated as the division of the inverse of the standard deviation of each series between the sum of the reciprocals of the standard deviations (equation 3).

$$\rho_n = \frac{\frac{1}{\sigma_i}}{\sum_{n=1}^{N}\frac{1}{\sigma_i}} \tag{3}$$

From this, you get the sum product of the five initial series and their respective weights (equation 4).

$$\acute{I}ndice\_ponderado_t = \sum_{n=1}^{N} \rho_n x_{tn} \tag{4}$$

Finally, for the construction of the IGCD, the scale changes to the calculated series in equation 4. To do this, the quarter that shows the greatest interest in purchasing an apartment is valued at 100.

Appendix 2: Causality test like Granger [11]

Causalidad a lo Granger: IICDG y Venta de departamentos en Lima*

Hipótesis nula	Probabilidad
a. IICDG no causa a las ventas de departamento	0.06
b. Ventas de departamento no causa IICDG	0.83

*With a probability of 10% the null hypothesis is rejected. That is to say the IICDG is a variable which lets us predict sales of apartments in Lima (the test was carried out with a lag and five lags and the null hypothesis "a" is rejected.)

Causalidad a lo Granger: IICDG y el crecimiento del crédito hipotecario *

Hipótesis nula	Probabilidad
a. IICDG no causa a créditos	0.00
b. créditos no causa IICDG	0.11

*With a probability of 5% the null hypothesis is rejected. That is to say, the IICDG is a variable that allows us to predict the growth of mortgage lending (the test was carried out with a lag and four lags and the null hypothesis "a" is rejected.)

[11] The Granger test analyses the relationship of causality between two variables, a different assessment to calculate only correlations. This test allows us to say whether a variable "y" is caused by another "x", i.e. if "x" helps to estimate and "y" predict. To that end, it evaluates if the coefficients of the lagged variable "x" are statistically significant when estimating "y". However, statement "x" causing "y" does not imply that "y" is the result of "x", because, in addition other factors outside of "x" intervene and affect "y". Sometimes, you can find a double Granger-causality: "x" causes "y" and "y" causes "x".

Appendix 3: Projection Models

Estimated parameters, endogenous variable: Sale of apartments in Lima (y_t)*

Variables exógenas	Modelo de referencia	Modelo aumentado
Yt(-1)	0.73	0.47
t- st.	5.7	2.89
MA(12)	-0.91	0.90
t- st.	29.4	29.40
Constante	4,016	-2,533
t- st.	8.5	-2.5
IGCD**		94.64
t- st.		6.6

* The ARMA model specification considered Box-Jenkins methodology (the variables are stationary and the models have no autocorrelation or heteroscedasticity). ** The variable is lagging behind in five quarters. Preparation: BBVA Research

Unit Root Test for IGCD series and selling apartments

Null Hypothesis: IGCD_SA has a unit root
Exogenous: Constant
Lag Length: 7 (Automatic - based on SIC, maxlag=8)

		t-Statistic
Elliott-Rothenberg-Stock DF-GLS test statistic		-4,185157
Test critical values:	1% level	-2,653401
	5% level	-1,953858
	10% level	-1,609571

*MacKinnon (1996)

Null Hypothesis: VENTA_SA has a unit root
Exogenous: Constant
Lag Length: 0 (Automatic - based on SIC, maxlag=8)

	T-Statistic
Elliott-Rothenberg-Stock DF-GLS test statistic	-1,947061

Test critical values:	1% level	-2,636901
	5% level	-1,951332
	10% level	-1,610747

*MacKinnon (1996)

Test: Breusch-Godfrey autocorrelation

Reference Model

Breusch-Godfrey Serial Correlation LM Test:

F-statistic	0,380924	Prob. F(2.28)	0,6867
Obs*R-squared	0,420578	Prob. Chi-Square(2)	0,8103

Augmented Model

Breusch-Godfrey Serial Correlation LM Test:

F-statistic	0,293380	Prob. F(2.27)	0,7481
Obs*R-squared	0,241060	Prob. Chi-Square(2)	0,8865

Test: White heteroscedaticity

Base Model

Heteroskedasticity Test: White

F-statistic	1,749012	Prob. F(9.23)	0,1344
Obs*R-squared	13,40841	Prob. Chi-Square(9)	0,1450
Scaled explained SS	11,08604	Prob. Chi-Square(9)	0,2699

Augmented Model

Heteroskedasticity Test: White

F-statistic	2.770427	Prob. F(4,28)	0.05
Obs*R-squar	9.357225	Prob. Chi-Square(4)	0.05
Scaled expla	14.81991	Prob. Chi-Square(4)	0.01

References

1. Carrière-Swallow y. and F. Labbé (2010). Nowcasting with Google Trends in an emerging market, Working Document, 2010-588, Central Bank of Chile.

2. Choi Hyunyoung and Hal Varian. Predicting the Present with Google Trends. Google Inc. 2009

3. Choi Hyunyoung and Hal Varian. Predicting the Present with Google Trends. Google Inc. 2011

4. Della Penna, N. and H. Huang (2009). Constructing Consumer Sentiment Index for U.S. Using Google Searches. University of Alberta, Canada. 2009.

5. Jeremy Ginsberg, Matthew H. Mohebbi, Rajan S. Patel, Lynnette Brammer, Mark S.Smolinski, and Larry Brilliant. Detecting influenza epidemics using search engine query data. 2009

6. Jillie Chang and Andrea Del Río. Google Trends: Predicting the level of aggregate employment in Peru using data in real time, 2005-2011. BCRP Working Paper. 2013

7. Lynn Wu and Erik Brynjolfsson. How Google Searches Foreshadow Housing Prices and Sales. 2009

8. Nikos Askitas and Klaus F. Zimmermann. Detecting Mortgage Delinquencies with Google Trends. 2010.

9. Nikos Askitas and Klaus F. Zimmermann. Google econometrics and unemployment forecasting. 2010.

10. Scott Baker and Andrey Fradkin. What Drives Job Search? Evidence from Google Search Data. The Stanford Institute for Economic Policy Research. 2011

11. P. M. Polgreen, y. Chen, D. M. Pennock, and F. D. Nelson. Using internet searches for influenza surveillance. Clinical Infectious Diseases. 2008.

LEGAL NOTICE

This document, prepared by the Department of BBVA Research, is informative in nature and contains data, opinions or estimates relating to the date of the same, they are from its own research or based on sources believed to be reliable, and have not been independently verified by BBVA. BBVA therefore makes no guarantee, express or implied, as to the document's accuracy, completeness or correctness.

The estimates that this document contains have been carried out according to generally accepted methodologies and should be taken as such, i.e. as estimates or projections. The historical evolution of economic variables (positive or negative) does not guarantee their equivalent evolution in the future.

The content of this document is subject to change without notice for example, depending on the economic context or market fluctuations. BBVA does not make any commitment to update any of the content or communicate such changes.

BBVA assumes no responsibility for any loss, direct or indirect, that may result from the use of this document or its contents.

Neither this document nor its content, constitutes an offer, invitation or solicitation to acquire, divest or obtain any interest in assets or financial instruments, nor can it form the basis of any contract, commitment or decision of any kind.

Particularly as regards investment in financial assets that may be related to the economic variables that this document develops, readers should be aware that in no case should they take this document as the basis for their investment decisions; and that persons or entities that can potentially offer them investment products are legally obliged to provide all the information they need to take these decisions.