

www.uoc.edu/idp

ARTÍCULO

L'activitat de cerques en Google anticipa els resultats electorals

Raúl Gómez Martínez

Departament d'Economia de l'Empresa
 Facultat de Ciències Jurídiques i Socials
 Universitat Rey Juan Carlos de Madrid

Camilo Prado Román

Departament d'Economia de l'Empresa
 Facultat de Ciències Jurídiques i Socials
 Universitat Rey Juan Carlos de Madrid

Data de publicació: juny de 2014

Resum

Aquest treball de recerca pretén analitzar la utilitat d'ús de les estadístiques de cerca a Google a l'hora d'explicar i/o predir diferents esdeveniments. En aquest estudi proposem un model economètric de dades de panell que demostra que les estadístiques de cerca de Google tenen capacitat explicativa i predictiva sobre l'evolució de la intenció de vot en les eleccions espanyoles. La relació entre ambdues variables és significativa, positiva i amb un coeficient superior a la unitat; per tant, si les estadístiques de cerca de Google augmenten d'un punt, interpretarem que la intenció de vot augmentarà de més d'un 1% del cens electoral.

Paraules clau

Google Insights, votants, intenció de vot, predicció eleccions, Espanya

Tema

estadístiques electorals i Google

Google search activities for predicting election results

Abstract

This research aims to analyse the usefulness of Google searches for explaining or predicting events. A panel-data econometric model is proposed that demonstrates that Google search statistics have explanatory and predictive capacity for the trend in voting intention in the Spanish elections. The relationship between votes and searches is significant, positive and greater than unity, which means that if Google search statistics increase by 1 point, voting intention increases by more than 1% of the electorate.

Keywords

Google Insights, voters, voting intentions, election prediction, Spain

Topic

electoral statistics and Google

Introducció

Aquest treball de recerca analitza la utilitat de l'ús de les estadístiques de cerca a Google a l'hora d'explicar i/o predir diferents esdeveniments. En el nostre cas, pretenem analitzar si la intenció de vot als diferents partits polítics participants en unes eleccions pot ser explicada a través de l'activitat de cerca registrada a Google dels noms i sigles de cada partit, és a dir, si aquest pot ser un indicador de la intenció de vot i pot explicar els resultats electorals.

El desenvolupament tecnològic i d'internet, així com els hàbits dels usuaris, estan provocant que sorgeixi una eina amb més «potència»; alguns estudis, com el de Lemmon i Portniaguina (2006), van demostrar que les enquestes de confiança del consumidor tenien capacitat predictiva sobre l'evolució dels mercats borsaris, i recentment s'ha observat que el desenvolupament tecnològic i internet estan confirmant aquesta apreciació, sent aquestes eines complementàries de les ja utilitzades tradicionalment. Les estadístiques de cerques de Google s'ofereixen de manera gratuïta¹ i presenten certs avantatges sobre les enquestes tradicionals: són gratuïtes, encara que hi ha eines d'enquestes que també ho són; és una manera ràpida d'obtenir informació per a l'usuari; la informació es pot obtenir en format electrònic (Excel, Word, CSV –*comma-separated*

values–, etc.), cosa que en facilita el tractament; es poden personalitzar per categoria, període, zona geogràfica, etc.; i pràcticament són immediates en el temps, perquè estan disponibles en línia.

A la tesi doctoral *El préstamo de valores en España: Relevancia de la venta en corto y el estado de ánimo de los inversores en la rentabilidad de la Bolsa Española* (Gómez-Martínez, 2012) s'analitza la utilitat de les estadístiques de cerques a Google a l'hora d'explicar o predir diversos esdeveniments econòmics o financers, en aquest cas l'evolució dels mercats borsaris en funció de l'estat d'ànim dels inversors i el seu nivell d'aversion al risc. Gómez-Martínez (2013) utilitza les estadístiques de cerques a Google per a quantificar el grau d'aversion al risc dels inversors. Com més cerques fan els inversors en Google sobre termes com *crisi* o *crash*, més aversion hi ha al risc dels inversors, i s'observen caigudes en les cotitzacions dels principals índexs borsaris. A Gómez-Martínez i Medrano-García (2012) es destaca la capacitat predictiva de les estadístiques de cerques a Google que s'estan utilitzant com a indicadors de qui guanyarà les eleccions amb almenys la mateixa precisió que les enquestes però amb els avantatges anteriorment descrits, i a manera il·lustrativa d'aquesta capacitat predictiva s'esmentava el cas de les passades eleccions generals espanyoles del 20 de novembre de 2011, sobre les quals es va fer un exercici

1. Les estadístiques de cerques de Google estan disponibles a: www.google.com/insights/search/

descarregant les estadístiques de cerques de les sigles dels partits polítics amb representació al Congrés restringint la zona al període de la campanya electoral i a la zona geogrà-

fica d'Espanya. Les següents mitjanes dels índexs de cerques realitzades anticipaven, amb una certa precisió, els resultats en percentatge de vot reals observats amb posterioritat.

Taula 1: Informació d'eleccions espanyoles, 20.11.2012

	PP	PSOE	IU	UPyD	CiU	PNV	ERC	BNG	TOTAL
MITJANA	56	35	11	18	4	3	3	2	133
% DE VOTS	42%	26%	8%	14%	3%	2%	2%	2%	100%

Font: Elaboració pròpia

Taula 2: Resultats d'eleccions espanyoles, 20.11.2012.

	Previsió	Real
PP	42%	45%
PSOE	26%	29%
IU	8%	7%
UPyD	14%	5%
CiU	3%	4%
PNV	2%	1%
ERC	2%	1%
BNG	2%	1%
Amaiur		1%

Font: Elaboració pròpia

Aquest exercici il·lustratiu va despertar un cert interès en l'audiència del Congrés, i es va repetir per a les eleccions autonòmiques catalanes i gallegues del 2012,² amb resultats novament interessants. Tot això ens ha portat a aprofundir-hi i a elaborar una anàlisi més profunda i rigorosa sobre aquesta matèria. Altres estudis, com ara el de Kessler (2012) per a Mashable.com, mostren la capacitat predictiva de les estadístiques de cerca de Google com a indicador de qui guanyarà les properes eleccions. L'estudi de Kessler es va fer per a les últimes eleccions generals dels Estats Units d'Amèrica i va obtenir resultats similars a les enquestes (mateixa precisió). La utilització d'aquestes estadístiques està obtenint rellevància en l'àmbit acadèmic, com ho demostren els treballs de Graefe i Armstrong (2010) o el de

Lui *et al.* (2011), que estudien les estadístiques de cerca de Google com a eina/indicador de predicció de les eleccions als Estats Units. Assumim que, cada vegada que sorgeix interès per un tema concret, el primer que acostuma a fer l'usuari és una cerca relacionada amb aquest interès. Google és el cercador majoritari d'internet, amb més d'un 80% de les cerques totals fetes (Banc Mundial, 2012). Google registra les cerques fetes pels usuaris d'internet, les monitoritza i en posa les estadístiques a la disposició dels mateixos usuaris perquè les consultin o en facin ús.

Ja en el terreny de la recerca acadèmica trobem un creixent interès en l'estudi de l'activitat registrada a internet, tant per a identificar què està passant en aquests moments (Ginsberg *et al.*, 2009; Choi i Varian, 2009; Metaxas i Mustafaraj, 2010) com per a predir el futur (Asur i Huberman, 2010; Tumasjan *et al.*, 2010; O'Connor *et al.*, 2010). Tots aquests treballs tenen com a punt en comú el supòsit que el volum de paraules clau buscades a Google o contingudes en converses de xarxes socials com Twitter estan revelant el pensament actual d'una secció molt representativa de la població (Khrabrov i Cybenko, 2010; Romero *et al.*, 2010). Els resultats observats en aquests estudis de recerca revelen que es poden utilitzar les estadístiques de cerques a Google o el volum de paraules presents en els xats de Twitter per a predir tota mena de dades socials i de consum, entre les quals destaquem l'èxit de les noves pel·lícules abans de la seva estrena (Asur i Huberman, 2010), el futur èxit de la comercialització de béns de consum (Choi i Varian, 2009) o, i això és més rellevant per al nostre estudi, la predicció de les eleccions alemanyes (Tumasjan *et al.*, 2010) i enquestes (O'Connor *et al.*, 2010).

2. Comentaris publicats a la secció «Actualidad financiera» (<<https://sites.google.com/site/culturillafinanciera/>>).

La principal aportació d'aquest estudi és que apliquem aquesta nova font d'informació al cas espanyol, sobre el qual no hem identificat treballs anteriors sobre el tema, de manera que això queda emmarcat en un model econòmic consistent que aporta una valoració quantitativa del fenomen estudiat.

L'estructura del treball és la següent: en un primer moment es farà una revisió de la literatura i es plantejarà la hipòtesi que s'ha de contrastar; després, es donarà una explicació tant de la mostra com de la metodologia emprada per a l'anàlisi; es prosseguirà amb els resultats obtinguts, i s'acabarà amb la discussió i les conclusions.

Per tant, sobre aquesta base, en aquest treball es planteja verificar la hipòtesi següent:

H1 Les estadístiques de cerques d'informació en internet sobre els diferents partits polítics participants en unes eleccions poden ser un indicador de la intenció de vot i poden explicar els resultats electorals. A més cerques a Google sobre el nom i les sigles dels partits participants en unes eleccions, correspon un nombre més alt de vots? La validació d'aquesta hipòtesi ens proveiria d'una eina gratuïta i precisa a l'hora d'anticipar el resultat d'unes eleccions.

Si assumim que el votant tria un partit polític perquè «compra» el seu programa polític, ideologia o valors, po-

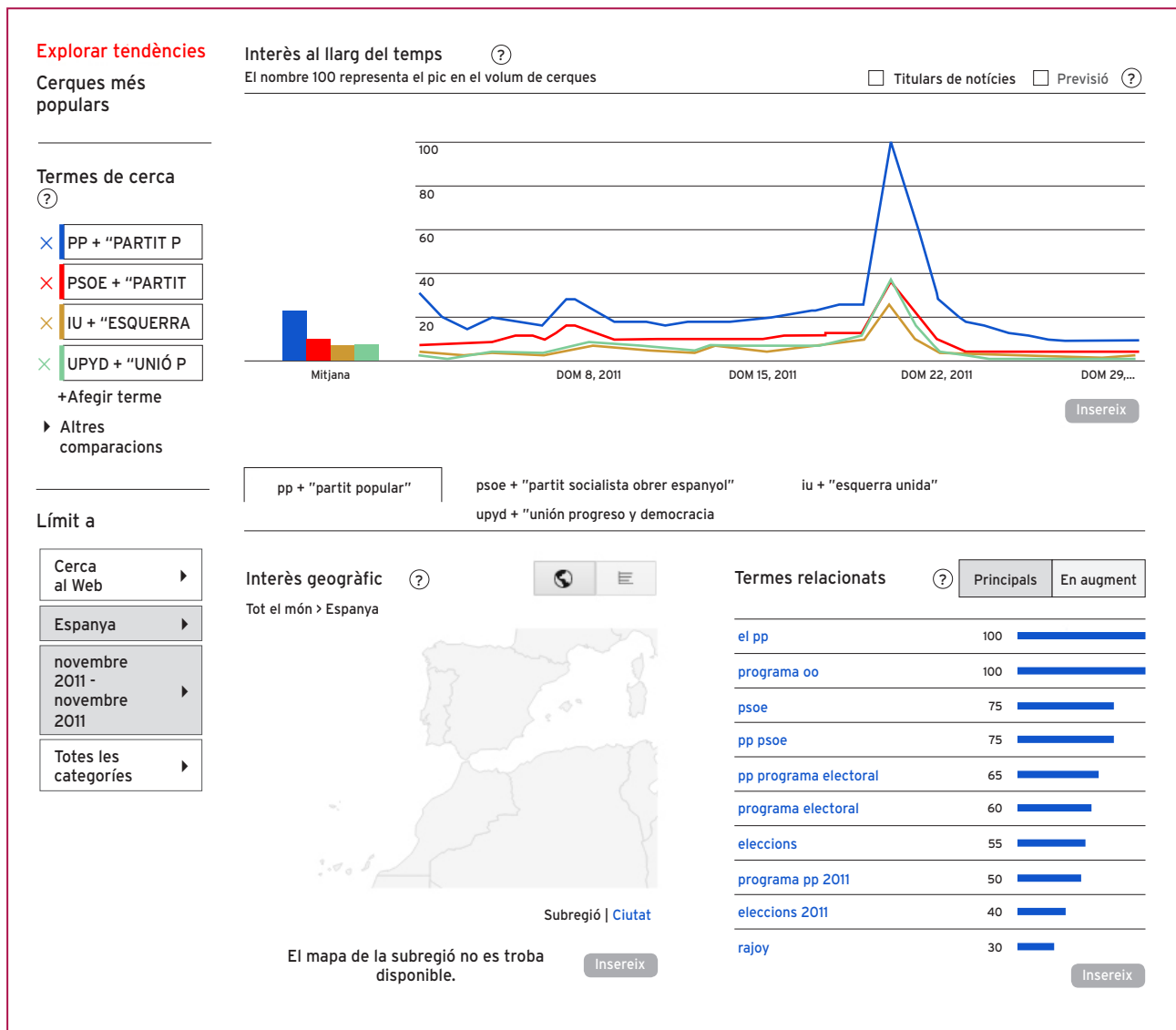
dem assimilar la decisió de vot a la decisió de compra i al comportament del consumidor. El model de decisió del consumidor (Blackwell *et al.*, 2005) proposa que els individus prenen decisions lògiques sobre la base d'uns fets realitzats amb anterioritat, és a dir, el model es fonamenta en una seqüència d'etapes:

1. Reconeixement de la necessitat
2. Cerca d'informació
3. Avaluació d'alternatives
4. Decisió
5. Conducta després de la decisió

Entenem que aquest model és aplicable a l'elecció del vot davant d'una sèrie d'alternatives polítiques seguint aquest flux:

1. Convocatòria d'eleccions.
2. Cerca d'informació sobre els programes polítics. Aquest seria el punt en què els votants entren en internet buscant informació sobre els partits polítics, variable explicativa en la nostra anàlisi.
3. Avaluació d'alternatives.
4. Decisió de vot, sobre la base de la informació recollida en el pas 2. La validació de l'H1 implicaria que com més informació es busca en el pas 2, més vots són realitzats en el pas 4.
5. Conducta després de la decisió.

Imatge 1: Cerca de l'estudi en Google Insights.



Font: <https://www.google.es/trends/>

Mostra i metodologia

Mostra

Les dades dels resultats electorals han estat descarregades directament del web del Ministeri de l'Interior.³ De les dades disponibles de vots, representants (diputats o regidors), percentatge de vàlids, percentatge de candidatura i percentatge de cens, hem pres aquest últim per a les dades de la

variable endògena. La xifra de nombre de vots pot presentar problemes d'escalabilitat si passem d'eleccions generals a eleccions locals, mentre que el percentatge de vots sobre el cens cridat a urnes serà una variable homogènia amb la qual treballar amb dades de panell.

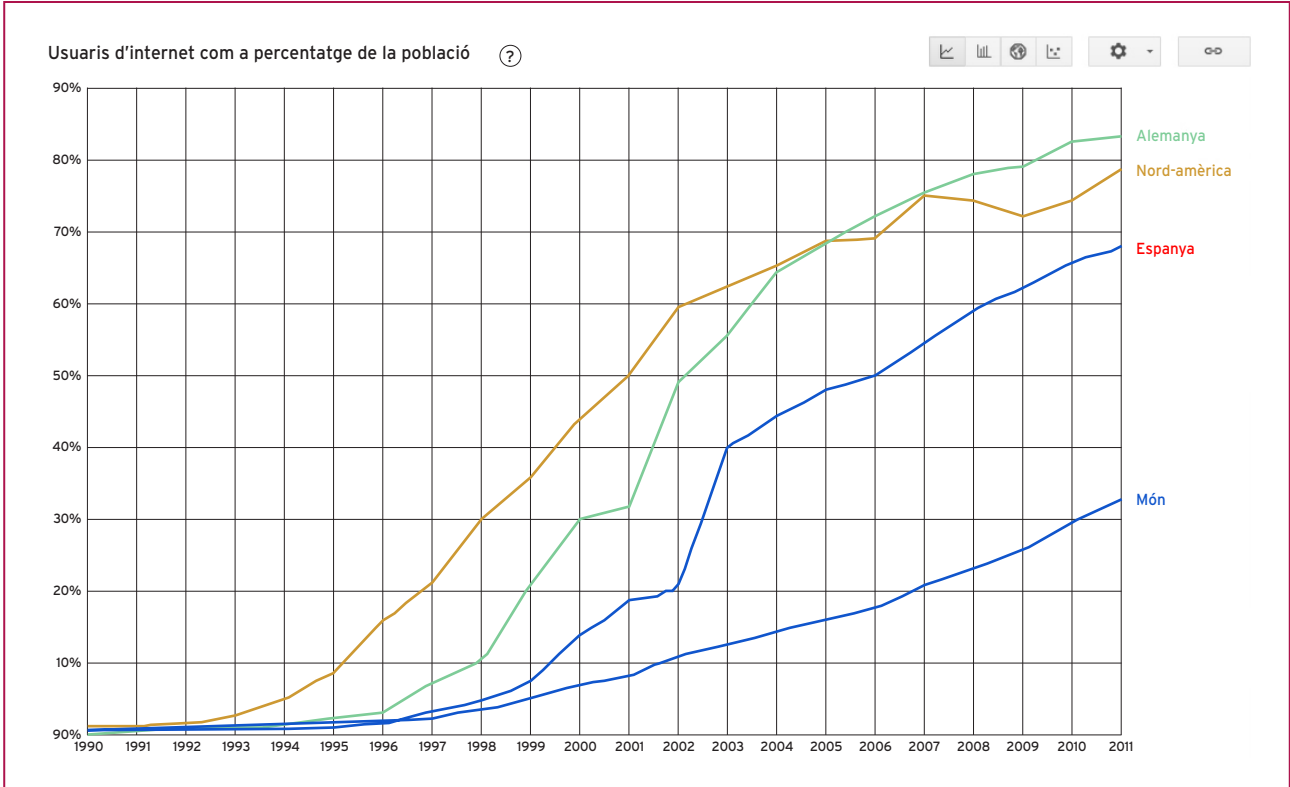
A partir dels resultats electorals anteriors hem consultat a Google Trends les estadístiques de cerques a Google fetes amb les sigles dels partits polítics o el nom complet oficial del partit. La

3. Els resultats electorals estan disponibles a l'URL <http://www.infoelectoral.mir.es/min/>

cerca s'ha restringit al mes en què s'han celebrat les eleccions (Google no permet discriminar per períodes més breus) i a la zona geogràfica d'influència. Per exemple, per a les eleccions

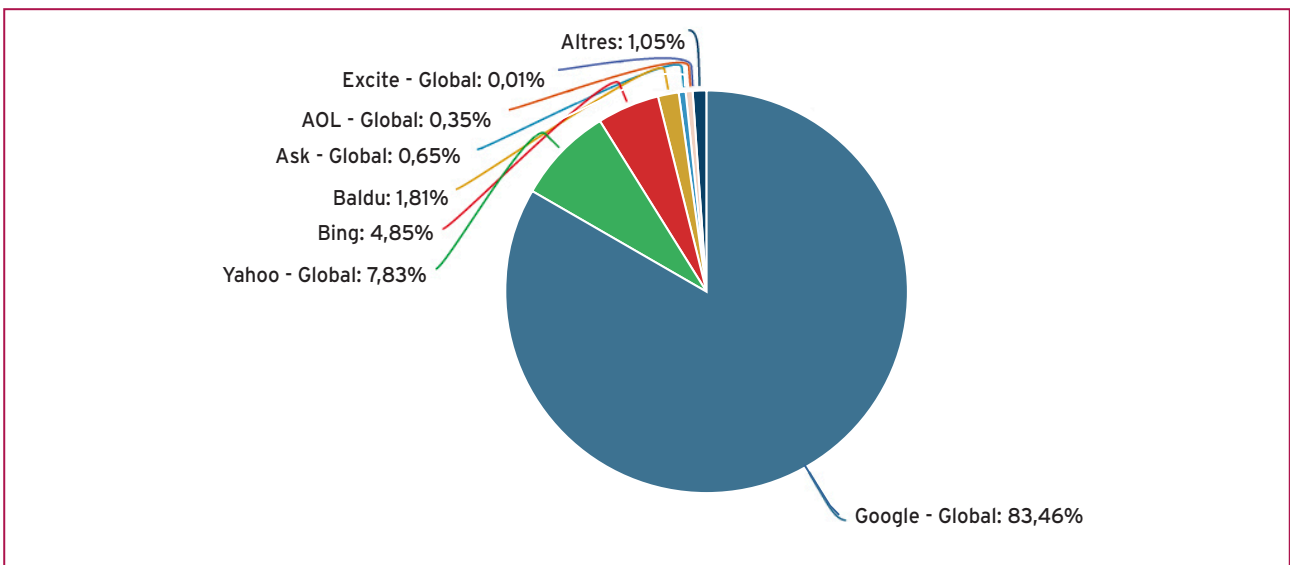
al Congrés celebrades al novembre del 2011 hem obtingut les estadístiques de cerques limitades a la zona «Espanya» i limitades a l'interval temporal que abraça el novembre del 2011.

Imatge 2: Usuaris d'internet com a percentatge de la població.



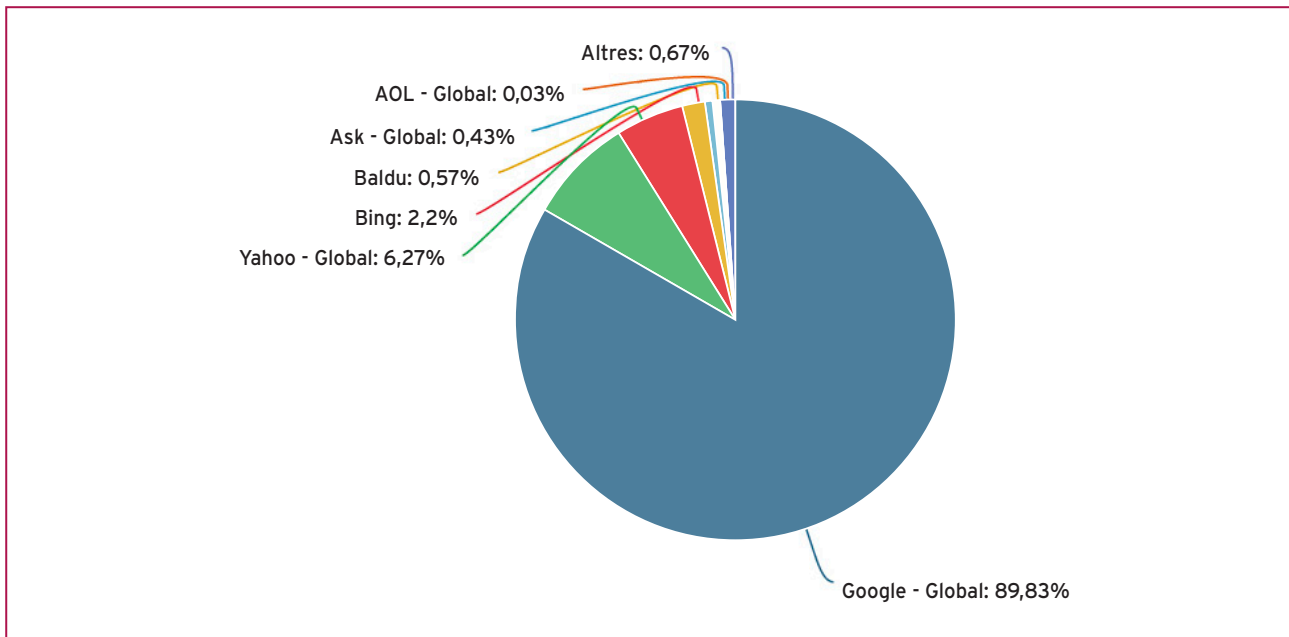
Font: Banc Mundial, 2012.

Imatge 3: Cercadors d'internet més utilitzats.



Font: <http://karmasnack.com/>.

Imatge 4: Cercadors d'internet mòbil més utilitzats.



Font: <<http://karmasnack.com/>>

Google ofereix dades des de l'any 2004, per la qual cosa la nostra anàlisi la centrem en les eleccions celebrades des del 2004 fins a l'actualitat. Pensem que és una mostra rellevant, ja que l'accés a internet a Espanya, segons dades del Banc Mundial, des de l'any 2004 és creixent i va des d'un 44,05% el 2004 fins a gairebé un 68% el 2011, xifra de població que considerem rellevant.

Metodologia

Per a fer aquest estudi proposem utilitzar un model economètric de dades de panell. Un model economètric de dades de panell és aquell que inclou una mostra d'agents o individus (en el nostre cas seran els diferents partits polítics participants en les eleccions) per a un període determinat de temps (la data en la qual se celebren les eleccions), per la qual cosa combinem la dimensió temporal i l'estructural (Noves, 2000). L'especificació general d'un model economètric amb dades de panell és la següent:

$$Y_{it} = \alpha_{it} + \beta X_{it} + u_{it} \text{ (Model I)}$$

On, en el nostre cas:

- Y_{it} : És el nombre de vots rebuts pel partit i en les eleccions celebrades en t . ($i=1, \dots, n$; $t=1, \dots, t$)

- X_{it} : És la mitjana de cerques a Google del partit i en el període t corresponent a les eleccions celebrades. ($i=1, \dots, n$; $t=1, \dots, t$)
- v_{it} : És el terme d'error que per a dades de panell sol descompondre's en:

- μ_i : Efectes no observables que difereixen entre les unitats d'estudi però no en el temps.
- δ_t : Efectes no quantificables que varien en el temps però no entre les unitats d'estudi.
- ε_{it} : Terme d'error purament aleatori.

Si δ_t i μ_i fossin iguals a 0, el model es refereix al cas en què no hi ha heterogeneïtat en el sistema de dades de panell i, per tant, s'empra el mètode de mínims quadrats ordinaris (MQO). Si això no és així, en presència d'heterocedasticitat, l'estimador de mínims quadrats ordinaris és lineal, sense biaix i consistent, però deixa de ser eficient, per la qual cosa hauríem de fer l'estimació seguint dues alternatives:

- Model d'efectes aleatoris: $Y_{it} = \alpha_{it} + \beta X_{it} + v_{it}$. El mètode de mínims quadrats ordinaris (MQO) no és aplicable, atès que no es compleixen els supòsits que permeten que l'estimador sigui consistent. Per això és preferible, en aquest cas, utilitzar el mètode de mínims quadrats generalitzats (MQG), les estimacions del qual són superiors a les del

MQO en cas que no es compleixin els supòsits tradicionals i són similars en cas contrari.

- Model d'efectes fixos: aquest model assumeix que hi ha un terme constant diferent per a cada individu, i suposa que els efectes individuals són independents els uns dels altres i per a la i -èsima unitat de tall transversal, la relació és la següent: $Y_i = \alpha_i + \beta X_i + v_i$.

En el nostre treball explorarem les tres alternatives: MQO, MQG i model d'efectes fixos

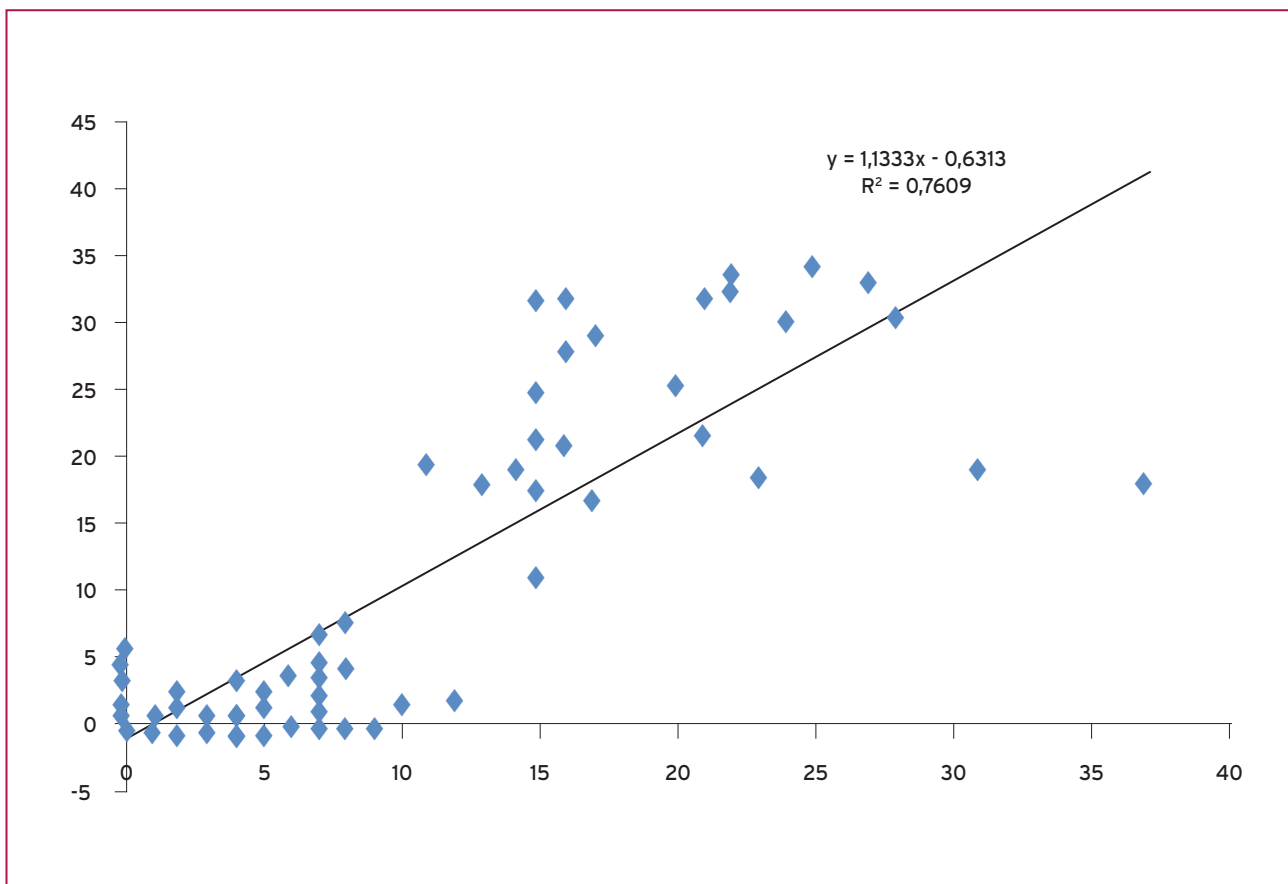
Resultats

Apliquem les dades descrites en l'epígraf anterior al model I, utilitzant com a variable endògena (Y) el percentatge de vots obtinguts sobre el cens electoral i, com a variable explicativa (X), la mitjana de cerques a Google dels partits

polítics realitzades en el mes de les eleccions a la zona geogràfica d'influència.

Les regressions s'han calculat utilitzant el paquet economètric GRTL 19.5cvs, els llistats del qual queden reflectits a l'annex. De les regressions dutes a terme, observem que el paràmetre b , calculat mitjançant el mètode de mínims quadrats ordinaris (MQO), de la variable explicativa (X), representada per la mitjana de cerques a Google, és significatiu segons l'estadístic t i diferent de 0 en un interval de confiança del 99%, ja que presenta un paràmetre positiu superior a la unitat. Podem interpretar aquest paràmetre assumint que un major (menor) interès en el partit polític observat per més (menys) cerques a Google d'aquest partit implica una major (menor) intenció de vot i un percentatge major (menor) de vots sobre el cens. A més, el coeficient de determinació R^2 és molt elevat, per la qual cosa es pot explicar el 76% de la intenció de vot solament tenint en compte l'interès despertat per un partit mesurat per les seves cerques a Google.

Gràfic 1: Model 1 - MQO combinats.



Font: Elaboració pròpia.

Encara que no s'aprecia autocorrelació en els residus, d'acord amb el test de Durbin-Watson, no se supera el test de White, per la qual cosa no tenim homocedasticitat, i els residus no són normals. Això implica que el paràmetre obtingut és lineal, sense biaix i consistent, però no és eficient, per la qual cosa procedim a calcular la regressió mitjançant mínims quadrats generalitzats (MQG) utilitzant el mètode de mínims quadrats ponderats. En aquesta regressió el coeficient de determinació R^2 s'eleva fins al 84% i, de nou, el paràmetre b és significatiu al 99%, positiu i amb un paràmetre superior a la unitat, però continuem sense superar el test de normalitat dels residus, per la qual cosa b segueix no sent eficient.

Fem la regressió alternativa del model d'efectes fixos. En aquesta regressió el paràmetre b continua sent significatiu al 99% i superior a la unitat, i, a més, en aquest cas no podem rebutjar en un interval de confiança del 95% les hipòtesis nul·les de normalitat en els residus i de l'existència d'un intercepte comú, per la qual cosa podem concloure que, a més, és eficient. El coeficient R^2 segueix sent molt elevat, ja que se situa per sobre del 80%.

Per a corregir l'heterocedasticitat observada en la regressió realitzada per MQO, procedim a fer una transformació logarítmica de les variables participants. Aquesta transformació linealitzaria el model davant d'una possible relació quadràtica entre les seves variables, fonamentada en el fet que partits nous, rellevants tot i que amb un menor volum d'electors, desperten més curiositat que els partits tradicionals, per la qual cosa susciten més cerques a través d'internet de les que serien proporcionals a la seva intenció de vot. El resultat del paràmetre b torna a ser positiu, superior a la unitat i significatiu al 99%, i es corregeix l'heterocedasticitat del model, encara que els residus no arriben a ser normals.

A partir dels resultats anteriors, si les estadístiques de cerca

de Google augmenten, això ho hem d'interpretar com un increment en la intenció de vot.

Els resultats obtinguts en les regressions dutes a terme es mostren a la taula 3

Quantitativament, d'acord amb les dades anteriors, hem d'interpretar que una campanya electoral que desperti un increment en l'interès del partit polític mesurable per un punt en la mitjana de cerques a Google significarà un increment en la intenció de vot de més d'un 1% sobre la del cens electoral que participa en aquestes eleccions.

Discussió i conclusió

L'objectiu del treball és analitzar la utilitat de l'ús de les estadístiques de cerca a Google a l'hora d'explicar i/o predir diferents esdeveniments. Partint del cas analitzat de les passades eleccions generals espanyoles del 20 de novembre de 2011, desenvolupem un model economètric que modelitzi i quantifiqui aquesta relació. L'estudi demostra que les estadístiques de cerca de Google tenen capacitat explicativa i predictiva sobre l'evolució de la intenció de vot en les eleccions espanyoles. Amb aquest objectiu apliquem el model I, descrit anteriorment, el qual utilitza com a variable endògena (Y) el percentatge de vots obtinguts sobre el cens electoral, i, com a variable explicativa (X), la mitjana de cerques a Google dels partits polítics realitzades en el mes de les eleccions a la zona geogràfica d'influència.

Dels resultats analitzats podem concloure que, aplicant la regressió per MQO, podem assumir que un major (menor) interès en el partit polític observat per més (menys) cerques a Google d'aquest partit implica una major (menor) intenció de vot i un percentatge major (menor) de vots sobre el cens. Aquesta interpretació va obtenir un coeficient de determinació R^2 elevat, i el 76% de la intenció de vot es pot explicar

Taula 3: Resultats dels models aplicats.

	α	β	t	Valor p	Significatiu al		R^2
MCO	-0,63	1,13	17,20	0,00	99%	Consistent	76%
MCG	-1,45	1,23	22,64	0,00	99%	Consistent	85%
Efectes fixos	-0,55	1,12	16,20	0,00	99%	Eficient	80%
MCO (log)	-1,96	1,55	10,28	0,00	99%	Consistent	59%

Font: Elaboració pròpia

solament tenint en compte l'interès despertat per un partit mesurat per les seves cerques a Google. El problema rau en el fet que no se supera el test de normalitat dels residus, per la qual cosa b és no eficient.

Aplicant la regressió del model de dades de panell mitjançant mínims quadrats generalitzats (MQG) utilitzant el mètode de mínims quadrats ponderats, s'obtenen resultats similars. El coeficient de determinació R^2 s'eleva fins al 84%, encara que es continua sense superar el test de normalitat dels residus, per la qual cosa b segueix no sent eficient.

Fem la regressió alternativa del model d'efectes fixos. El paràmetre b segueix sent significatiu al 99% i superior a la unitat, i, a més, per aquest mètode sí que és eficient. El coeficient R^2 segueix sent molt elevat, ja que se situa per sobre del 80%.

Els resultats obtinguts pels tres mètodes són similars, per la qual cosa es pot interpretar, en els tres, que un major (menor) interès en el partit polític observat per més (menys)

cerques a Google d'aquest partit implica una major (menor) intenció de vot i un percentatge major (menor) de vots sobre el cens.

Quan validem la hipòtesi sobre si les estadístiques de cerques d'informació en internet dels diferents partits polítics participants en unes eleccions poden ser un indicador de la intenció de vot i poden explicar els resultats electorals, podem indicar que aquestes estadístiques proveeixen la societat d'una eina gratuïta i precisa a l'hora d'anticipar el resultat d'unes eleccions.

Per a valorar la capacitat predictiva de les estadístiques obtingudes a Google Insights hem fet un estudi profund mitjançant els tres models utilitzats anteriorment. A partir dels resultats obtinguts podem concloure que, si les estadístiques de cerca de Google augmenten, això ho hem d'interpretar com un increment en la intenció de vot. Aquests resultats els hem d'utilitzar amb cautela, esperant que es confirmin amb estudis posteriors en altres països i en altres moments de temps passats i/o futurs.

Referències bibliogràfiques

- ASUR, S.; HUBERMAN, B. A. (2010). *Predicting the Future with Social Media*. <<http://arxiv.org/abs/1003.5699>>
- BANC MUNDIAL (2012). *Informe Usuarios de Internet como porcentaje de la población*. Banc Mundial. <<http://datos.bancomundial.org/indicador/it.net.user.p2/countries/>>
- BLACKWELL, R. D.; MINIARD, P. W.; ENGEL, J. F. (2005). *Consumer Behavior*. South-Western College Pub.
- CHAFFEE, S. (1981). «Mass media in political campaigns: an expanding role». A: D. RICE & W. PAISLEY (ed.). *Public Communication Campaigns*. Beverly Hills: Sage Pub.
- CHOI, H.; VARIAN, H. (2009). *Predicting the Present with Google Trends*. Google Research Blog. [Data de consulta: abril del 2009]. <<http://googleresearch.blogspot.com/2009/04/predicting-present-with-google-trends.html>>
- GINSBERG, J.; MOHEBBI, M.; PATEL, R.; BRAMMER, L.; SMOLINSKI, M.; BRILLIANT, L. (2009). «Detecting influenza epidemics using search engine query data». *Nature*. Vol. 457. [Data de consulta: febrer del 2009]. <<http://dx.doi.org/10.1038/nature07634>>
- GÓMEZ-MARTÍNEZ, R. (2012). *El préstamo de valores en España: Relevancia de la venta en corto y el estado de ánimo de los inversores en la rentabilidad de la Bolsa Española*. Alemania: EAE Publishing.
- GÓMEZ-MARTÍNEZ, R. (2013). «Señales de inversión basadas en un índice de aversión al riesgo». *Investigaciones Europeas de Dirección y Economía de la Empresa*. <<http://dx.doi.org/10.1016/j.iedee.2012.12.001>>
- GÓMEZ-MARTÍNEZ, R.; MEDRANO-GARCÍA, M. L. (2012). «Señales de inversión basadas en estadísticas de búsquedas en Internet». *AEDEM Proceedings del XXVI Congreso Anual*. Barcelona.
- GRABER, D. (1980). *Mass media; Social aspects; Political aspects; United States*. Washington: Congressional Quarterly Press

- GRAEFE, A.; ARMSTRONG, J. S. (2010). «Predicting elections from the most important issue: A test of the take-the-best heuristic». *Journal of Behavioral Decision Making*. Vol. 25, núm. 1, pàg. 41-48. <<http://dx.doi.org/10.1002/bdm.710>>
- KESSLER, S. (2012). «Are Google Search Trends Better Election Predictors than Polls?». Mashable.com. <<http://mashable.com/2012/02/02/google-search-trends-election/>>
- KHRABROV, A.; CYBENKO, G. (2010). «Discovering Influence in Communication Networks using Dynamic Graph Analysis». Article inclòs a *In SocialCom 2010*. Minneapolis, 20-22 d'agost. Pàg. 288-294.
- LEMMON, M.; PORTNIAGUINA, E. (2006). «Consumer Confidence and Asset Prices: Some Empirical Evidence». *The Review of Financial Studies*. Vol. 19, núm. 4, pàg. 1499-1529. <<http://dx.doi.org/10.1093/rfs/hhj038>>
- LUI, C.; METAXAS, P. T.; MUSTAFARAJ, E. (2011). *On the Predictability of the U.S. Elections through Search Volume Activity*. Wellesley (Massachusetts): Wellesley College Press.
- McCLURE, R. D.; PATTERSON, T. E. (1974). «Television news and political advertising». *Communication Research*. Núm. 1, pàg. 3-31. <<http://dx.doi.org/10.1177/009365027400100101>>
- METAXAS, P.; MUSTAFARAJ, E. (2010). «From Obscurity to Prominence: Political Speech and Real Time Search». A: *Web Science 2010*. Raleigh (Carolina del Nord).
- NOVALES, A. (2000). *Econometría*. Madrid: McGraw Hill.
- ROMERO, D.; GALUBA, W.; ASUR, S. [et al.] (2010). «Influence and passivity in social media». *Lecture Notes in Computer Science*. Volume 6913, págs. 18-33.
- ROSE, R.; MOSSAWIR, H. (1980). «Voting and elections: A functional analysis». *Political Studies*. Núm. 15, pàg. 173-201. <<http://dx.doi.org/10.1111/j.1467-9248.1967.tb01843.x>>
- TUMASJAN, A.; SPRENGER, T. O.; SANDNER, P. G.; WELPE, I. M. (2010). «Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment». Publicat a *4th International AAAI Conference on Weblogs and Social Media (ICWSM)*. Pàg. 178-185.
- O'CONNOR, B.; BALASUBRAMANYAN, R.; ROUTLEDGE, B.; SMITH, A. (2010). «From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series». Publicat a *1st International AAAI Conference on Weblogs and Social Media*. Washington. Maig. Pàg. 122-129.

Annex

Tabla 4: Model 1 - MCO combinat

Model 1: MQO combinats, utilitzant 95 observacions
S'han inclòs 19 unitats de secció creuada
Llargada de la sèrie temporal = 5
Variable dependent: p_censo

	Coefficient	Desv. típica	Estadístic t	Valor p
const	-0,631299	0,760087	-0,8306	0,4083
Google	1,13332	0,0658714	17,20	1,19e-030 ***

	Mitjana de la vble. dep.	D.T. de la vble. dep.	Mitjana de la vble. dep.	D.T. de la regressió
Suma de quad. residus	2871,570	5,556717	R-quadrat	0,758361
R-quadrat	0,760932	R-quadrat corregit	0,758361	F(1, 93)
F(1, 93)	296,0105	Valor p (d'F)	1,19e-30	Log-versemblança
Log-versemblança	-296,7142	Criteri d'Akaike	597,4284	Criteri de Schwarz
Criteri de Schwarz	602,5361	Crit. d'Hannan-Quinn	599,4923	rho
rho	-0,212386	Durbin-Watson	1,827462	

Contrast d'heterocedasticitat de White -
Hipòtesi nul·la: No hi ha heterocedasticitat
Estadístic de contrast: LM = 41,9468
amb valor p = P(Ji quadrat(2) > 41,9468) = 7,7868e-010

Contrast de normalitat dels residus -
Hipòtesi nul·la: l'error es distribueix normalment
Estadístic de contrast: Ji quadrat(2) = 28,0887
amb valor p = 7,95463e-007

Font: Elaboració pròpia

Taula 6: Model 3 - Efectes fixos

Model 3: Efectes fixos, utilitzant 95 observacions
S'han inclòs 19 unitats de secció creuada
Llargada de la sèrie temporal = 5
Variable dependent: p_censo

	Coefficient	Desv. típica	Estadístic t	Valor p
const	-0,547641	0,781460	-0,7008	0,4856
Google	1,12235	0,0692874	16,20	3,36e-026 ***

	Mitjana de la vble. dep.	D.T. de la vble. dep.	Mitjana de la vble. dep.	D.T. de la regressió
Suma de quad. residus	2358,944	5,608260	R-quadrat	0,753858
R-quadrat	0,803610	R-quadrat corregit	0,753858	F(19, 75)
F(19, 75)	16,15226	Valor p (d'F)	2,79e-19	Log-versemblança
Log-versemblança	-287,3735	Criteri d'Akaike	614,7471	Criteri de Schwarz
Criteri de Schwarz	665,8246	Crit. d'Hannan-Quinn	635,3863	rho
rho	-0,371620	Durbin-Watson	2,236325	

Contrast de diferents interceptes per grups -
Hipòtesi nul·la: Els grups tenen un intercepte comú
Estadístic de contrast: F(18, 75) = 0,905467
amb valor p = P(F(18, 75) > 0,905467) = 0,574121

Contrast de normalitat dels residus -
Hipòtesi nul·la: l'error es distribueix normalment
Estadístic de contrast: Ji quadrat(2) = 5,9018
amb valor p = 0,05229266

Font: Elaboració pròpia.

Taula 5: Model 2 - MC ponderats

Model 2: MQ ponderats, utilitzant 95 observacions
S'han inclòs 19 unitats de secció creuada
Variable dependent: p_censo
Ponderacions basades en variàncies dels errors per unitat

	Coefficient	Desv. típica	Estadístic t	Valor p
const	-1,44589	0,564623	-2,561	0,0121 **
Google	1,23130	0,0543881	22,64	1,31e-039 ***

Estadístics basats en les dades ponderades:

	Suma de quad. residus	D.T. de la regressió	R-quadrat	F(1, 93)
Suma de quad. residus	91,67335	0,992842	R-quadrat	0,844764
R-quadrat	0,846416	R-quadrat corregit	0,844764	F(1, 93)
F(1, 93)	512,5302	Valor p (d'F)	1,31e-39	Log-versemblança
Log-versemblança	-133,1060	Criteri d'Akaike	270,2120	Criteri de Schwarz
Criteri de Schwarz	275,3198	Crit. d'Hannan-Quinn	272,2759	

Estadístics basats en les dades originals:

	Mitjana de la vble. dep.	D.T. de la vble. dep.	Mitjana de la vble. dep.	D.T. de la regressió
Suma de quad. residus	2940,315	5,622837		

Contrast de normalitat dels residus -
Hipòtesi nul·la: l'error es distribueix normalment
Estadístic de contrast: Ji quadrat(2) = 28,7423
amb valor p = 5,73715e-007

Font: Elaboració pròpia.

Taula 7: Model 4 - MCO combinats (log)

Model 4: MQO combinats, utilitzant 75 observacions
S'han inclòs 19 unitats de secció creuada
Llargada de la sèrie temporal: mínim 3, màxim 5
Variable dependent: l_p_censo

	Coefficient	Desv. típica	Estadístic t	Valor p
const	-1,95888	0,313214	-6,254	2,43e-08 ***
l_Google	1,54962	0,150676	10,28	7,61e-016 ***

	Mitjana de la vble. dep.	D.T. de la vble. dep.	Mitjana de la vble. dep.	D.T. de la regressió
Suma de quad. residus	153,3847	1,449538	R-quadrat	0,586058
R-quadrat	0,591651	R-quadrat corregit	0,586058	F(1, 73)
F(1, 73)	105,7689	Valor p (d'F)	7,61e-16	Log-versemblança
Log-versemblança	-133,2502	Criteri d'Akaike	270,5003	Criteri de Schwarz
Criteri de Schwarz	275,1353	Crit. d'Hannan-Quinn	272,3510	rho
rho	-0,383599	Durbin-Watson	2,031240	

Contrast d'heterocedasticitat de White -
Hipòtesi nul·la: No hi ha heterocedasticitat
Estadístic de contrast: LM = 5,64767
amb valor p = P(Ji quadrat(2) > 5,64767) = 0,0593778

Contrast de normalitat dels residus -
Hipòtesi nul·la: l'error es distribueix normalment
Estadístic de contrast: Ji quadrat(2) = 36,8451
amb valor p = 9,98131e-009

Font: Elaboració pròpia.

Citació recomenada

GÓMEZ MARTÍNEZ, Raúl; PRADO ROMÁN, Camilo (2014). «L'activitat de cerques a Google anticipa els resultats electorals». *IDP. Revista d'Internet, Dret i Política*. Núm. 18, pàg. 2-15. UOC. [Data de consulta: dd/mm/aa].

<<http://journals.uoc.edu/index.php/idp/article/view/n18-gomez-prado/n18-gomez-prado-ca>>

<<http://dx.doi.org/10.7238/idp.v0i18.1941>>



Els textos publicats en aquesta revista estan subjectes -llevat que s'indiqui el contrari- a una llicència de Reconeixement-Sense obres derivades 3.0 Espanya de Creative Commons. Podeu copiar-los, distribuir-los i transmetre'ls públicament sempre que citeu l'autor, la revista i la institució que els publica (*IDP. Revista d'Internet, Dret i Política*; UOC); no en feu obres derivades. La llicència completa es pot consultar a <http://creativecommons.org/licenses/by-nd/3.0/es/deed.ca>.

Sobre els autors

Raúl Gómez Martínez

raul.gomez.martinez@urjc.es

Departament d'Economia de l'Empresa
 Facultat de Ciències Jurídiques i Socials
 Universitat Rey Juan Carlos de Madrid

Despatx 283

Edifici Departamental I (Vicálvaro)

Passeig d'Artilleros, s/n

28032 Madrid

Llicenciat en Ciències Econòmiques i Empresariales, màster d'Anàlisi econòmica i economia financera, estudis del màster oficial de Direcció d'empreses i doctor en Economia de l'Empresa i Finances per la Universitat Rey Juan Carlos de Madrid.

Des de l'any 2007 imparteix i coordina assignatures emmarcades en els plans docents de la Universitat Rey Juan Carlos de diplomatures, llicenciatures, graus i postgraus, com ara Valoració i adquisició d'empreses, Mercats financers nacionals, Gestió bancària, Productes financers, Assessorament i planificació financera, Direcció financera, Mercats de valors, Finances corporatives, etc. La seva activitat docent es pot seguir a: <<http://www.culturillafinanciera.com/>>. Ha participat en projectes de recerca com el del diccionari econòmic i financer de la Reial Acadèmia de Ciències Econòmiques i Financeres (RACEF). Ha publicat diversos articles en premsa i revistes de recerca, així com participacions en congressos i seminaris. Les seves publicacions es poden seguir a: <<http://elblogderaulgomez.blogspot.com/>>. Té més de quinze anys d'experiència col·laborant en institucions financeres de primer nivell com ara Grup Santander, BBVA, Bankia, Banesto, Caixa Rural o ING Direct, principalment en projectes de consultoria relacionats amb tresoreria, valors, productes d'inversió i assegurances de qualitat de programari. La seva trajectòria professional es pot seguir a: <<http://www.linkedin.com/pub/raul-gomez-martinez/27/a69/3b2>>.

Camilo Prado Román
camilo.prado.roman@urjc.es
Departament d'Economia de l'Empresa
Facultat de Ciències Jurídiques i Socials
Universitat Rey Juan Carlos de Madrid

Despatx 271
Edifici Departamental I (Vicálvaro)
Passeig d'Artilleros, s/n
28032 Madrid

Doctor en Ciències Econòmiques i Empresariales (Àrea de Finances) per la Universitat Rey Juan Carlos (URJC), és especialista en metodologia de la recerca quantitativa i tècniques estadístiques per la Universitat Politècnica de Madrid. És llicenciat en Administració i Direcció d'Empreses, especialitat Auditoria i direcció d'empreses, pel Colegio Universitario de Estudios Financieros (CUNEF). Va exercir la docència durant deu anys a la Universitat Rey Juan Carlos de Madrid, dins l'Àrea d'Economia Financera i Comptabilitat, impartint assignatures com ara Introducció a l'empresa, Mercats financers nacionals, Gestió de carteres, Planificació financera, Valoració d'empreses, Direcció financera, Inversions alternatives o Finances aplicades al màrqueting. Va ser analista financer a la Central de Balanços del Banc d'Espanya l'any 2002. Forma part del personal docent i investigador de l'URJC des del novembre del 2003, desenvolupant activitats de recerca en el camp de la inversió en actius no financers, específicament en economia del col·leccionisme, legitimitat organitzativa, gestió de la qualitat i gestió de carteres.