

www.uoc.edu/idp

ARTÍCULO

La actividad de búsquedas en Google anticipa los resultados electorales

Raúl Gómez Martínez

Departamento de Economía de la Empresa
Facultad de Ciencias Jurídicas y Sociales
Universidad Rey Juan Carlos de Madrid

Camilo Prado Román

Departamento de Economía de la Empresa
Facultad de Ciencias Jurídicas y Sociales
Universidad Rey Juan Carlos de Madrid

Fecha de publicación: junio de 2014

Resumen

Este trabajo de investigación pretende analizar la utilidad de uso de las estadísticas de búsqueda en Google a la hora de explicar y/o predecir diferentes acontecimientos. En este estudio proponemos un modelo econométrico de datos de panel que demuestra que las estadísticas de búsqueda de Google tienen capacidad explicativa y predictiva sobre la evolución de la intención de voto en las elecciones españolas. La relación entre ambas variables es significativa, positiva y con un coeficiente superior a la unidad; por lo que si las estadísticas de búsqueda de Google aumentan en un punto interpretaremos que la intención de voto aumentará en más de un 1% del censo electoral.

Palabras clave

Google Insights, votantes, intención de voto, predicción elecciones, España

Tema

estadísticas electorales y 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

Introducción

El presente trabajo de investigación analiza la utilidad del uso de las estadísticas de búsqueda en Google a la hora de explicar y/o predecir diferentes acontecimientos. En nuestro caso, pretendemos analizar si la intención de voto a los diferentes partidos políticos participantes en unas elecciones puede ser explicada a través de la actividad de búsqueda registrada en Google de los nombres y siglas de cada partido, es decir, si este puede ser un indicador de la intención de voto y explicar los resultados electorales.

El desarrollo tecnológico y de Internet, así como los hábitos de los usuarios están provocando que surja una herramienta con mayor «potencia»; algunos estudios, como el de Lemmon y Portniaguina (2006), demostraron que las encuestas de confianza del consumidor tenían capacidad predictiva sobre la evolución de los mercados bursátiles, y recientemente se ha observado que el desarrollo tecnológico e Internet están confirmando esta apreciación; siendo estas herramientas complementarias a las ya utilizadas tradicionalmente. Las estadísticas de búsquedas de Google se ofrecen de manera gratuita¹ y presentan ciertas ventajas sobre las encuestas tradicionales: son gratuitas, aunque existen herramientas de encuestas que también lo son; es una manera rápida de obtención de información para el usuario; se puede obtener

la información en formato electrónico (Excel, Word, CSV –*comma-separated values*–, etc.) lo que facilita su tratamiento; se pueden personalizar por categoría, período, zona geográfica, etc.; y prácticamente son inmediatas en tiempo, al estar disponibles en línea.

En la tesis 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) se analiza la utilidad de las estadísticas de búsquedas en Google a la hora de explicar o predecir diversos acontecimientos económicos o financieros, en este caso la evolución de los mercados bursátiles en función del estado de ánimo de los inversores y su nivel de aversión al riesgo. Gómez-Martínez (2013) utiliza las estadísticas de búsquedas en Google para cuantificar el grado de aversión al riesgo de los inversores. Cuantas más búsquedas realizan los inversores en Google sobre términos como crisis o *crash*, mayor es la aversión al riesgo de los inversores, y se observan caídas en las cotizaciones de los principales índices bursátiles. En Gómez-Martínez y Medrano-García (2012) se destaca la capacidad predictiva de las estadísticas de búsquedas en Google que están siendo usadas como indicadores de quién ganará las elecciones con al menos la misma precisión que las encuestas pero con las ventajas anteriormente descritas, y a modo ilustrativo de esta ca-

1. Las estadísticas de búsquedas de Google están disponibles en: <www.google.com/insights/search/>.

pacidad predictiva se mencionaba el caso de las pasadas elecciones generales españolas del 20 de noviembre de 2011, sobre las que se hizo un ejercicio descargando las estadísticas de búsquedas de las siglas de los partidos políticos con representación en el Congreso restringiendo

la zona al período de la campaña electoral y a la zona geográfica de España. Los siguientes promedios de los índices de búsquedas realizadas anticipaban con cierta precisión los resultados en porcentaje de voto reales observados con posterioridad.

Tabla 1: Información de elecciones españolas, 20-11-2012

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

Fuente: Elaboración propia

Tabla 2: Resultados de elecciones españolas, 20-11-2012

	Previsión	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%

Fuente: Elaboración propia

Este ejercicio ilustrativo despertó cierto interés en la audiencia del Congreso, y el ejercicio se repitió para las elecciones autonómicas catalanas y gallegas de 2012,² con resultados de nuevo interesantes. Todo ello nos ha llevado a profundizar y elaborar un análisis más en profundidad y riguroso sobre esta materia. Otros estudios, como el de Kessler (2012) para Mashable.com, muestran la capacidad predictiva de las estadísticas de búsqueda de Google como indicador de quién ganará las próximas elecciones. El estudio de Kessler fue realizado para las últimas elecciones generales de los Estados Unidos de América, y obtuvo resultados similares a las encuestas (misma precisión). La utilización de estas estadísticas está obteniendo relevancia en el ámbito académ-

mico, como demuestran los trabajos de Graefe y Armstrong (2010) o el de Lui *et al.* (2011) que estudian las estadísticas de búsqueda de Google como herramienta/indicador de predicción de las elecciones en los Estados Unidos. Asumimos que, cada vez que aparece el interés por un tema concreto, lo primero que suele hacer el usuario es una búsqueda de ese interés. Google es el buscador mayoritario de Internet, con más de un 80% de las búsquedas totales realizadas en Internet (Banco Mundial, 2012). Google registra las búsquedas realizadas por los usuarios de Internet, las monitoriza y pone las estadísticas a disposición de los usuarios para su consulta y/o uso.

Las investigaciones sobre campañas electorales y el efecto de los medios de difusión en los resultados electorales fueron llevadas a cabo por un grupo muy heterogéneo de académicos, que actuaban en áreas de periodismo (Chaffee, 1981) o ciencia política (Graber, 1980; Patterson y McClure, 1974; Rose *et al.*, 1980). Entre las conclusiones más importantes de estos trabajos podemos destacar la cuantificación de estas campañas en la intención de voto, de las que se concluye que, en los Estados Unidos, entre el 7% y el 11% de los electores cambia de intención de voto, de un partido a otro. Entre los que no tienen una identificación partidaria arraigada, este porcentaje sube entre el 10% y el 28%. Las investigaciones enmarcadas en esta perspectiva comunicacional se centraron en el análisis de los medios masivos de difusión, sin embargo, en varios estudios se señalaba la fuerte influencia de los contactos personales en el cambio de intención de voto. Disponer de una herramienta en línea

2. Comentarios publicados en la sección «Actualidad financiera» (<<https://sites.google.com/site/culturillafinanciera/>>).

que sirva a los gestores de estas campañas para monitorizar el resultado de sus esfuerzos puede ser de gran utilidad.

Ya en el plano de la investigación académica, encontramos un creciente interés en el estudio de la actividad registrada en Internet, tanto para identificar qué está pasando en estos momentos (Ginsberg *et al.*, 2009; Choi y Varian, 2009; o Metaxas y Mustafaraj, 2010) como para predecir el futuro (Asur y Huberman, 2010; Tumasjan *et al.*, 2010; O'Connor *et al.*, 2010). Todos estos trabajos tienen como punto en común el supuesto de que el volumen de palabras clave buscadas en Google o contenidas en charlas de redes sociales como Twitter están revelando el pensamiento actual de una sección muy representativa de la población (Khrabrov y Cybenko, 2010; Romero *et al.*, 2010). Los resultados observados en los estudios de investigación citados son que se pueden utilizar las estadísticas de búsquedas en Google o el volumen de palabras presentes en los chats de Twitter para predecir todo tipo de datos sociales y de consumo, entre los que destacamos el éxito de las nuevas películas antes de su estreno (Asur y Huberman, 2010), el futuro éxito de la comercialización de bienes de consumo (Choi y Varian, 2009) o, lo que es más relevante para nuestro estudio, la predicción de las elecciones alemanas (Tumasjan *et al.*, 2010) y encuestas (O'Connor *et al.*, 2010).

La principal aportación de este estudio es que aplicamos esta nueva fuente de información al caso español, sobre el que no hemos identificado trabajos anteriores al respecto, lo cual queda enmarcado en un modelo econométrico consistente que aporta una valoración cuantitativa del fenómeno estudiado.

La estructura del trabajo es la siguiente, en un primer momento se realizará una revisión de la literatura y se planteará la hipótesis que se va a contrastar, después se realizará una explicación tanto de la muestra como de la metodología empleada para el análisis, para proseguir con los resultados obtenidos y terminar con la discusión y las conclusiones.

Por lo tanto, sobre esta base, en el presente trabajo se plantea verificar la siguiente hipótesis:

H1 Las estadísticas de búsquedas de información en Internet sobre los diferentes partidos políticos participantes en unas elecciones pueden ser un indicador de la intención de voto y explicar los resultados electorales. ¿A mayores búsquedas en Google sobre el nombre y las siglas de los partidos participantes en unas elecciones le corresponde un mayor número de votos? La validación de esta hipótesis nos proveería de una herramienta gratuita y precisa a la hora de anticipar el resultado de unas elecciones.

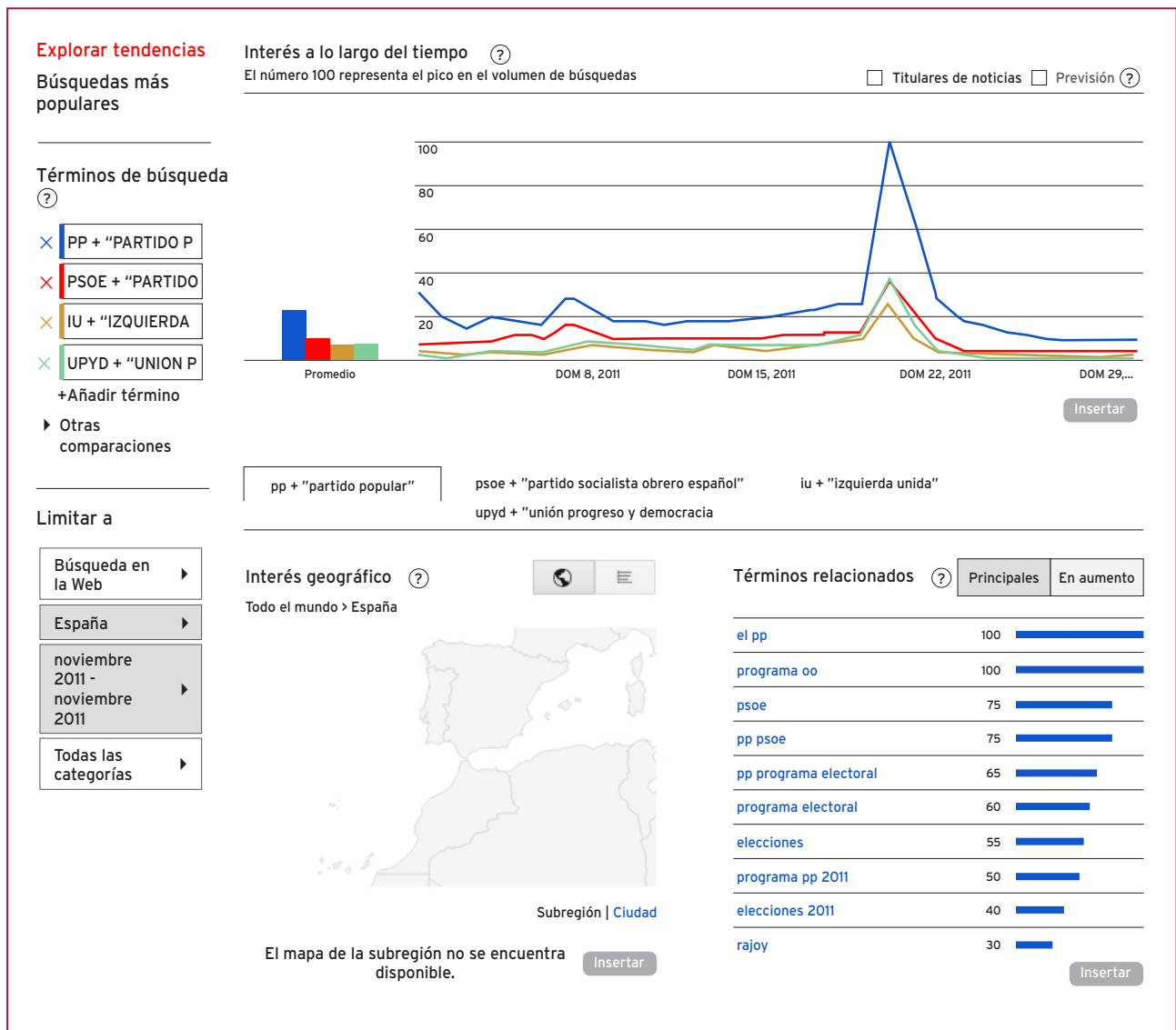
Si asumimos que el votante elige un partido político porque «compra» su programa político, ideología o valores, podemos asimilar la decisión de voto a la decisión de compra y al comportamiento del consumidor. El modelo de decisión del consumidor (Blackwell *et al.*, 2005) propone que los individuos toman decisiones lógicas sobre la base de unos hechos realizados con anterioridad, es decir, el modelo se fundamenta en una secuencia de etapas:

1. Reconocimiento de la necesidad
2. Búsqueda de información
3. Evaluación de alternativas
4. Decisión
5. Conducta tras la decisión

Entendemos que este modelo es aplicable a la elección del voto ante una serie de alternativas políticas siguiendo este flujo:

1. Convocatoria de elecciones
2. Búsqueda de información sobre los programas políticos. Este sería el punto en el que los votantes entran en Internet buscando información sobre los partidos políticos, variable explicativa en nuestro análisis.
3. Evaluación de alternativas
4. Decisión de voto, sobre la base de la información recogida en el paso 2. La validación de la H1 implicaría que cuanto más información se busca en el paso 2 más votos son realizados en el paso 4.
5. Conducta tras la decisión

Imagen 1: Búsqueda del estudio en Google Insights



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

Muestra y metodología

Muestra

Los datos de los resultados electorales han sido descargados directamente de la web del Ministerio del Interior.³ De los datos disponibles de votos, representantes (diputados o concejales), porcentaje de válidos, porcentaje de candidatura y porcentaje de censo, hemos tomado este último para los datos de la variable endógena. La cifra de número de votos

puede presentar problemas de escalabilidad si pasamos de elecciones generales a elecciones locales, mientras que el porcentaje de votos sobre el censo llamado a urnas va a ser una variable homogénea con la que trabajar con datos de panel.

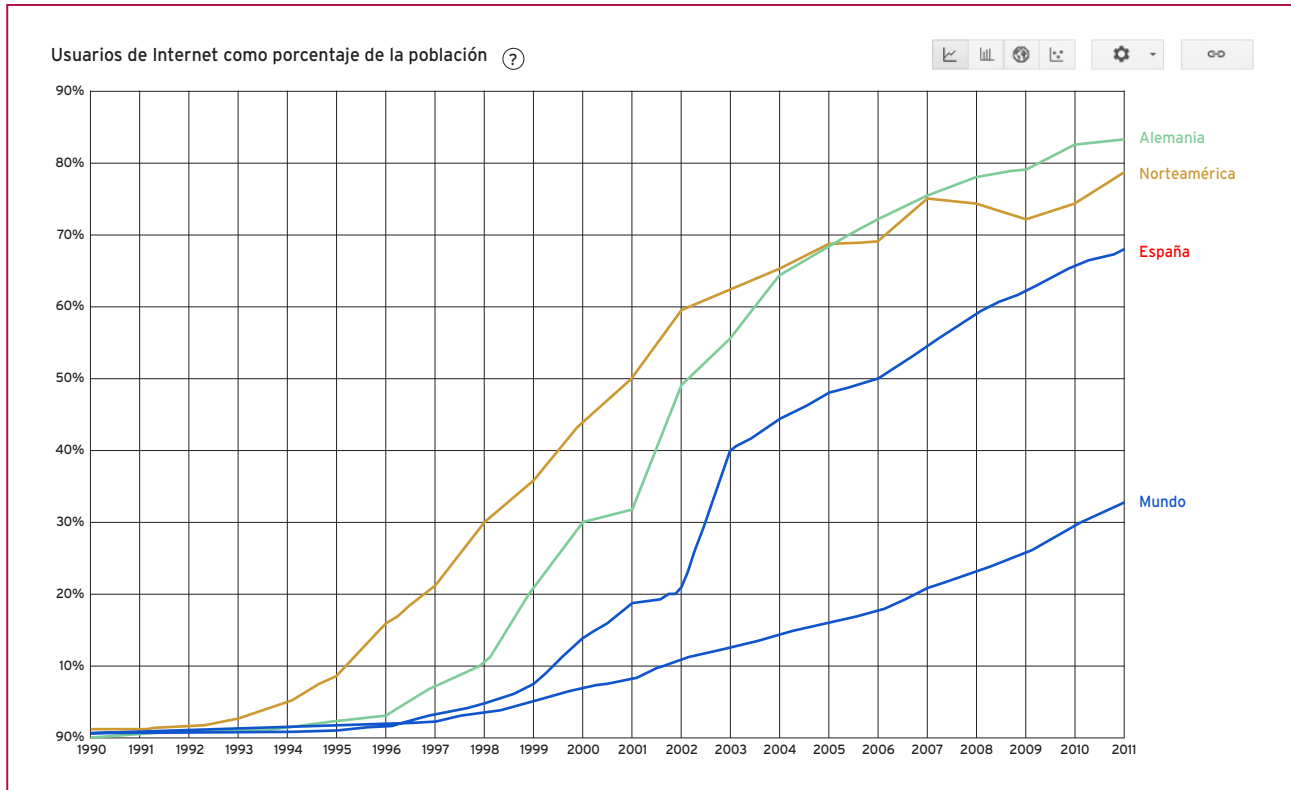
A partir de los resultados electorales anteriores hemos consultado en Google Trends las estadísticas de búsquedas en Google realizadas con las siglas de los partidos políticos o el nombre completo oficial del partido. La búsqueda se ha restringido al mes en el que se han celebrado las elecciones

3. Los resultados electorales están disponibles en la URL <http://www.infoelectoral.mir.es/min/>.

(Google no permite discriminar por períodos más breves) y a la zona geográfica de influencia. Por ejemplo, para las elecciones al Congreso celebradas en noviembre de 2011,

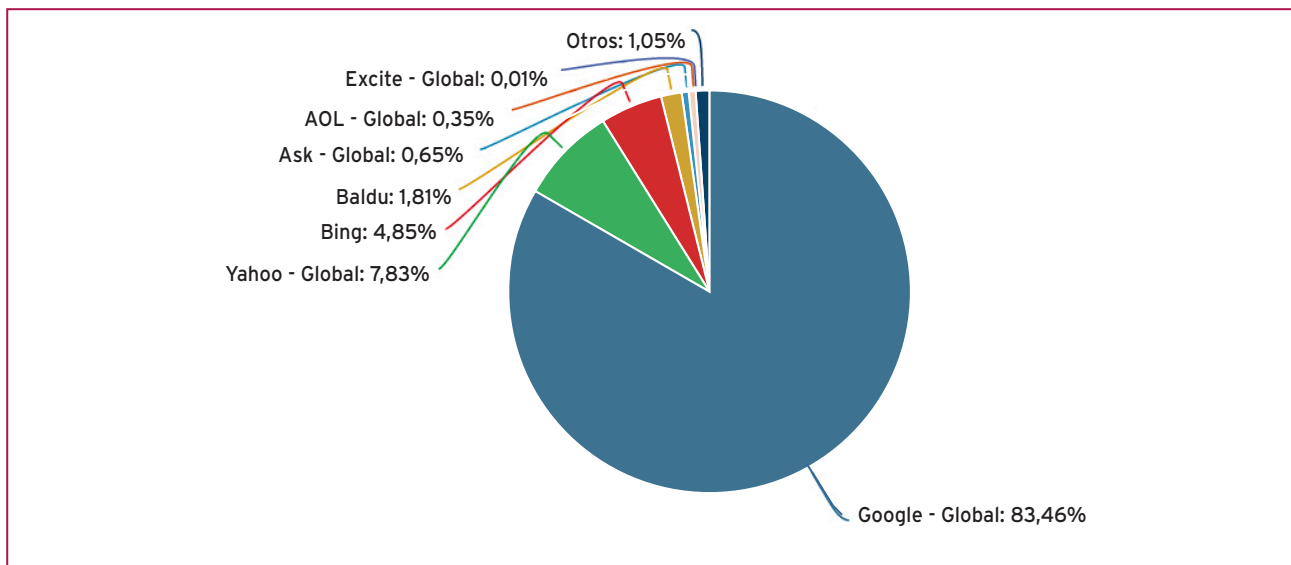
hemos obtenido las estadísticas de búsquedas limitadas a la zona «España» y limitadas al intervalo temporal que abarca noviembre de 2011.

Imagen 2: Usuarios de Internet como porcentaje de la población



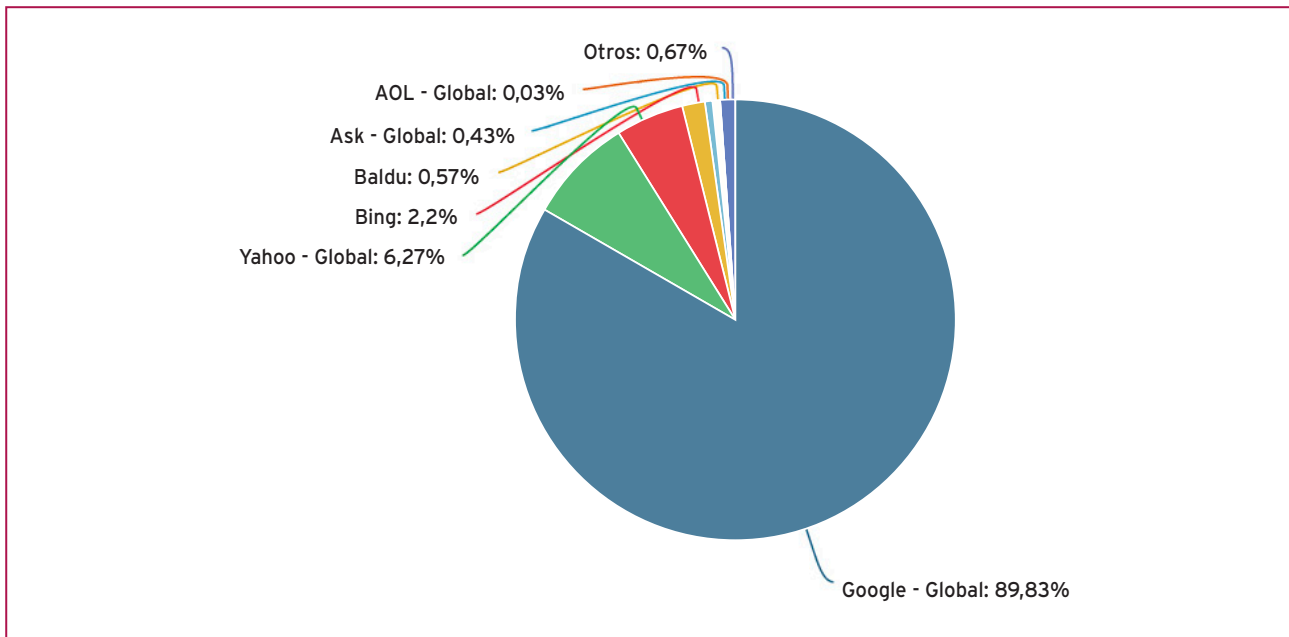
Fuente: Banco Mundial, 2012

Imagen 3: Buscadores de Internet más utilizados



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

Imagen 4: Buscadores de Internet móvil más utilizados



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

Google ofrece datos desde el año 2004, por lo que centramos nuestro análisis en las elecciones celebradas desde 2004 hasta la actualidad. Pensamos que es una muestra relevante ya que el acceso a Internet en España, según datos del Banco Mundial, desde el año 2004 es creciente y va desde un 44,05 % en 2004 hasta casi un 68 % en 2011, lo que consideramos una población relevante.

Metodología

Para realizar este estudio proponemos utilizar un modelo econométrico de datos de panel. Un modelo econométrico de datos de panel es aquel que incluye una muestra de agentes o individuos (en nuestro caso serán los diferentes partidos políticos participantes en las elecciones) para un período determinado de tiempo (la fecha en la que se realizan las elecciones), por lo que combinamos la dimensión temporal y estructural (Novales, 2000). La especificación general de un modelo econométrico con datos de panel es la siguiente:

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

Donde, en nuestro caso:

- Y_{it} : Es el número de votos recibidos por el partido «i» en las elecciones celebradas en «t». (i=1, ..., n; t=1, ..., t)

- X_{it} : Es el promedio de búsquedas en Google del partido «i» en el período «t» correspondiente a las elecciones celebradas. (i=1, ..., n; t=1, ..., t)
- u_{it} : Es el término de error que para datos de panel suele descomponerse en:

- μ_{it} : Efectos no observables que difieren entre las unidades de estudio pero no en el tiempo.
- δ_{it} : Efectos no cuantificables que varían en el tiempo pero no entre las unidades de estudio.
- ε_{it} : Término de error puramente aleatorio.

Si δ_t y μ_t fuesen igual a 0, el modelo se refiere al caso en que no existe heterogeneidad en el sistema de datos de panel y por tanto se emplea el método de mínimos cuadrados ordinarios (MCO). Si esto no es así, en presencia de heterocedasticidad, el estimador de mínimos cuadrados ordinarios es lineal, insesgado y consistente, pero deja de ser eficiente, por lo que deberíamos realizar la estimación siguiendo dos alternativas:

- Modelo de efectos aleatorios: $Y_{it} = \alpha_{it} + \beta X_{it} + v_{it}$. El método de mínimos cuadrados ordinarios (MCO) no es aplicable dado que no se cumplen los supuestos que permiten que el estimador sea consistente. Por lo que es preferible en este caso utilizar el método de mínimos cuadrados gene-

realizados (MCG), cuyas estimaciones son superiores al de MCO en caso de no cumplirse los supuestos tradicionales y son similares en caso contrario.

- Modelo de efectos fijos: Este modelo asume que existe un término constante diferente para cada individuo, y supone que los efectos individuales son independientes entre sí y para la i -ésima unidad de corte transversal, la relación es la siguiente: $Y_i = i\alpha_t + \beta X_i + v_i$

En nuestro trabajo exploraremos las tres alternativas, MCO, MCG y modelo de efectos fijos.

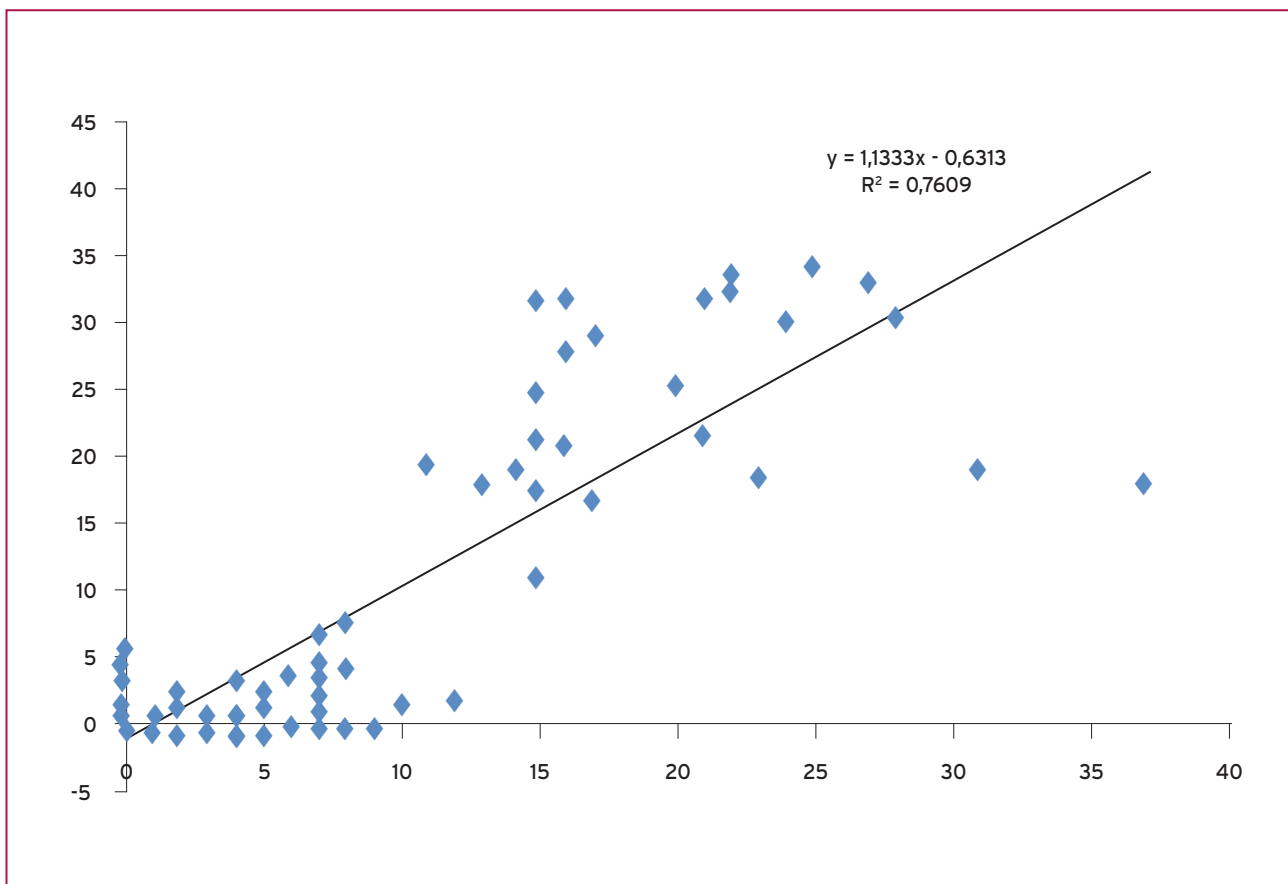
Resultados

Aplicamos los datos descritos en el epígrafe anterior al modelo 1, utilizando como variable endógena (Y) el porcentaje de votos obtenidos sobre el censo electoral y como variable explicativa (X) el promedio de búsquedas en Google de los

partidos políticos realizadas en el mes de las elecciones en la zona geográfica de influencia.

Las regresiones se han calculado utilizando el paquete econométrico GRTL 1.9.5cvs, cuyos listados quedan reflejados en el anexo. De las regresiones realizadas, observamos que el parámetro b , calculado mediante el método de mínimos cuadrados ordinarios (MCO), de la variable explicativa (X), representada por el promedio de búsquedas en Google, es significativo según el estadístico «t» y distinto de 0 en un intervalo de confianza del 99%, pues presenta un parámetro positivo superior a la unidad. Podemos interpretar este parámetro asumiendo que un mayor (menor) interés en el partido político observado por mayores (menores) búsquedas en Google de este implica una mayor (menor) intención de voto y un porcentaje mayor (menor) de votos sobre el censo. Además, el coeficiente de determinación R^2 es muy elevado, por lo que puede explicarse el 76 % de la intención de voto solamente atendiendo al interés despertado por un partido medido por sus búsquedas en Google.

Gráfico 1: Modelo 1 - MCO combinados



Fuente: Elaboración propia.

Aunque no se aprecia autocorrelación en los residuos, a tenor del test de Durbin-Watson, no se supera el test de White, por lo que no tenemos homocedasticidad, y los residuos no son normales. Esto implica que el parámetro obtenido es lineal, insesgado y consistente, pero no es eficiente, por lo que procedemos a calcular la regresión mediante mínimos cuadrados generalizados (MCG) usando el método de mínimos cuadrados ponderados. En esta regresión el coeficiente de determinación R^2 se eleva hasta el 84 % y de nuevo el parámetro b es significativo al 99%, positivo y con un parámetro superior a la unidad, pero seguimos sin superar el test de normalidad de los residuos por lo que b sigue no siendo eficiente.

Realizamos la regresión alternativa del modelo de efectos fijos. En esta regresión el parámetro b sigue siendo significativo al 99 % y superior a la unidad, y además, en este caso no podemos rechazar en un intervalo de confianza del 95 % las hipótesis nulas de normalidad en los residuos y de la existencia de un intercepto común por lo que podemos concluir que además es eficiente. El coeficiente R^2 sigue siendo muy elevado pues se sitúa por encima del 80 %.

Para corregir la heterocedasticidad observada en la regresión realizada por MCO procedemos a realizar una transformación logarítmica de las variables participantes. Esta transformación linealizaría el modelo ante una posible relación cuadrática entre sus variables, fundamentada en que partidos nuevos, relevantes aunque con un menor volumen de electores, despiertan más curiosidad que los partidos tradicionales, por lo que suscitan más búsquedas a través de Internet de las que serían proporcionales a su intención de voto. El resultado del parámetro b vuelve a ser positivo, superior a la unidad y significativo al 99 %, y se corrige la heterocedasticidad del modelo aunque los residuos no llegan a ser normales.

A partir de los resultados anteriores, si las estadísticas de búsqueda de Google aumentan, debemos interpretarlo como un incremento en la intención de voto.

Los resultados obtenidos en las regresiones realizadas se muestran en la tabla 3.

Cuantitativamente, a tenor de los datos anteriores, debemos interpretar que una campaña electoral que despierte un incremento en el interés del partido político medible por un punto en el promedio de búsquedas en Google significará un incremento en la intención de voto de más de un 1 % sobre la del censo electoral que participa en esas elecciones.

Discusión y conclusión

El objetivo del trabajo es analizar la utilidad del uso de las estadísticas de búsqueda en Google a la hora de explicar y/o predecir diferentes acontecimientos, partiendo del caso analizado de las pasadas elecciones generales españolas del 20 de noviembre de 2011 desarrollamos un modelo econométrico que modelice y cuantifique esta relación. El estudio demuestra que las estadísticas de búsqueda de Google tienen capacidad explicativa y predictiva sobre la evolución de la intención de voto en las elecciones españolas. Para ello aplicamos el modelo I, descrito anteriormente, el cual utiliza como variable endógena (Y) el porcentaje de votos obtenidos sobre el censo electoral y como variable explicativa (X) el promedio de búsquedas en Google de los partidos políticos realizadas en el mes de las elecciones en la zona geográfica de influencia.

De los resultados analizados podemos concluir que, aplicando la regresión por MCO, podemos asumir que un mayor (menor) interés en el partido político observado por mayores (menores) búsquedas en Google de este implica una mayor (menor) intención de voto y un porcentaje mayor (menor) de

Tabla 3: Resultados de los modelos aplicados

	α	β	t	Valor p	Significativo al		R^2
MCO	-0,63	1,13	17,20	0,00	99%	Consistente	76%
MCG	-1,45	1,23	22,64	0,00	99%	Consistente	85%
Efectos fijos	-0,55	1,12	16,20	0,00	99%	Eficiente	80%
MCO (log)	-1,96	1,55	10,28	0,00	99%	Consistente	59%

Fuente: Elaboración propia

votos sobre el censo. Esta interpretación obtuvo un coeficiente de determinación R^2 elevado, y puede explicarse el 76% de la intención de voto solamente atendiendo al interés despertado por un partido medido por sus búsquedas en Google. El problema se encuentra en que no se supera el test de normalidad de los residuos por lo que b es no eficiente.

Aplicando la regresión del modelo de datos de panel mediante mínimos cuadrados generalizados (MCG) usando el método de mínimos cuadrados ponderados, se obtienen resultados similares. El coeficiente de determinación R^2 se eleva hasta el 84%, aunque se sigue sin superar el test de normalidad de los residuos, por lo que b sigue no siendo eficiente.

Realizamos la regresión alternativa del modelo de efectos fijos. El parámetro b sigue siendo significativo al 99% y superior a la unidad, y, además, por este método, sí es eficiente. El coeficiente R^2 sigue siendo muy elevado pues se sitúa por encima del 80%.

Los resultados obtenidos por los tres métodos son similares, por lo que puede interpretarse en los tres que un mayor

(menor) interés en el partido político observado por mayores (menores) búsquedas en Google de este implica una mayor (menor) intención de voto y un porcentaje mayor (menor) de votos sobre el censo.

Al validar la hipótesis sobre si las estadísticas de búsquedas de información en Internet de los diferentes partidos políticos participantes en unas elecciones pueden ser un indicador de la intención de voto y explicar los resultados electorales, podemos indicar que estas estadísticas proveen a la sociedad de una herramienta gratuita y precisa a la hora de anticipar el resultado de unas elecciones.

Para valorar la capacidad predictiva de las estadísticas obtenidas en Google Insights hemos realizado un estudio profundo mediante los tres modelos utilizados anteriormente, a partir de los resultados obtenidos podemos concluir que, si las estadísticas de búsqueda de Google aumentan, debemos interpretarlo como un incremento en la intención de voto. Estos resultados los debemos utilizar con cautela esperando que se confirmen con estudios posteriores en otros países y en otros momentos de tiempo pasados y/o futuros.

Referencias bibliográficas

- ASUR, S.; HUBERMAN, B. A. (2010). *Predicting the Future with Social Media*. <<http://arxiv.org/abs/1003.5699>>
- BANCO MUNDIAL (2012). *Informe Usuarios de Internet como porcentaje de la población*. Banco 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». En: RICE, D.; PAISLEY, W. (eds). *Public Communication Campaigns*. Beverly Hills: Sage Pub.
- CHOI, H.; VARIAN, H. (2009). *Predicting the Present with Google Trends*. Google Research Blog. [Fecha de consulta: 2 de abril de 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. [Fecha de consulta: 19 de febrero de 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, DC: 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, n.o 1, 41-48. <<http://dx.doi.org/10.1002/bdm.710>>
- KESSLER, S. (2012). «Are Google Search Trends Better Election Predictors than Polls?». Mashable.com. USA. <<http://mashable.com/2012/02/02/google-search-trends-election/>>
- KHRABROV, A.; CYBENKO, G. (2010). «Discovering Influence in Communication Networks using Dynamic Graph Analysis». En: *In SocialCom 2010*.
- LEMMON, M.; PORTNIAGUINA, E. (2006). «Consumer Confidence and Asset Prices: Some Empirical Evidence». *The Review of Financial Studies*. Vol. 19, n.o 4, 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*. Department of Computer Science. EE. UU.: Wellesley College Press.
- MCCLURE, R. D.; PATTERSON, T. E. (1974). «Television news and political advertising». *Communication Research*. N.o 1, pág. 3-31. <<http://dx.doi.org/10.1177/009365027400100101>>
- METAXAS, P.; MUSTAFARAJ, E. (2010). «From Obscurity to Prominence: Political Speech and Real Time Search». En: *Web Science 2010*. Raleigh, NC.
- NOVALES CINCA, A. (2000). *Econometría*. 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*. N.o 15, pág. 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». En: *4th International AAAI Conference on Weblogs and Social Media (ICWSM)*.
- O'CONNOR, B.; BALASUBRAMANYAN, R.; ROUTLEDGE, B.; SMITH, A. (2010). «From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series». En: *1st International AAAI Conference on Weblogs and Social Media*. Washington, DC. Mayo

Anexo

Tabla 4: Modelo 1 - MCO combinados

Modelo 1: MCO combinados, utilizando 95 observaciones
Se han incluido 19 unidades de sección cruzada
Largura de la serie temporal = 5
Variable dependiente: p_censo

	Coefficiente	Desv. típica	Estadístico t	Valor p
const	-0,631299	0,760087	-0,8306	0,4083
Google	1,13332	0,0658714	17,20	1,19e-030 ***
Media de la vble. dep.	8,017684	D.T. de la vble. dep.	11,30407	
Suma de cuad. residuos	2871,570	D.T. de la regresión	5,556717	
R-cuadrado	0,760932	R-cuadrado corregido	0,758361	
F(1, 93)	296,0105	Valor p (de F)	1,19e-30	
Log-verosimilitud	-296,7142	Criterio de Akaike	597,4284	
Criterio de Schwarz	602,5361	Crit. de Hannan-Quinn	599,4923	
rho	-0,212386	Durbin-Watson	1,827462	

Contraste de heterocedasticidad de White -
Hipótesis nula: No hay heterocedasticidad
Estadístico de contraste: LM = 41,9468
con valor p = P(Ji cuadrado(2) > 41,9468) = 7,7868e-010

Contraste de normalidad de los residuos -
Hipótesis nula: el error se distribuye normalmente
Estadístico de contraste: Ji cuadrado(2) = 28,0887
con valor p = 7,95463e-007

Fuente: Elaboración propia

Tabla 6: Modelo 3 - Efectos fijos

Modelo 3: Efectos fijos, utilizando 95 observaciones
Se han incluido 19 unidades de sección cruzada
Largura de la serie temporal = 5
Variable dependiente: p_censo

	Coefficiente	Desv. típica	Estadístico t	Valor p
const	-0,547641	0,781460	-0,7008	0,4856
Google	1,12235	0,0692874	16,20	3,36e-026 ***
Media de la vble. dep.	8,017684	D.T. de la vble. dep.	11,30407	
Suma de cuad. residuos	2358,944	D.T. de la regresión	5,608260	
R-cuadrado	0,803610	R-cuadrado corregido	0,753858	
F(19, 75)	16,15226	Valor p (de F)	2,79e-19	
Log-verosimilitud	-287,3735	Criterio de Akaike	614,7471	
Criterio de Schwarz	665,8246	Crit. de Hannan-Quinn	635,3863	
rho	-0,371620	Durbin-Watson	2,236325	

Contraste de diferentes interceptos por grupos -
Hipótesis nula: Los grupos tienen un intercepto común
Estadístico de contraste: F(18, 75) = 0,905467
con valor p = P(F(18, 75) > 0,905467) = 0,574121

Contraste de normalidad de los residuos -
Hipótesis nula: el error se distribuye normalmente
Estadístico de contraste: Ji cuadrado(2) = 5,9018
con valor p = 0,0522926

Fuente: Elaboración propia

Tabla 5: Modelo 2 - MC ponderados

Modelo 2: MC ponderados, utilizando 95 observaciones
Se han incluido 19 unidades de sección cruzada
Variable dependiente: p_censo
Ponderaciones basadas en varianzas de los errores por unidad

	Coefficiente	Desv. típica	Estadístico t	Valor p
const	-1,44589	0,564623	-2,561	0,0121 **
Google	1,23130	0,0543881	22,64	1,31e-039 ***
Media de la vble. dep.	8,017684	D.T. de la vble. dep.	11,30407	
Suma de cuad. residuos	91,67335	D.T. de la regresión	0,992842	
R-cuadrado	0,846416	R-cuadrado corregido	0,844764	
F(1, 93)	512,5302	Valor p (de F)	1,31e-39	
Log-verosimilitud	-133,1060	Criterio de Akaike	270,2120	
Criterio de Schwarz	275,3198	Crit. de Hannan-Quinn	272,2759	

Estadísticos basados en los datos ponderados:

Estadísticos basados en los datos originales:

Media de la vble. dep. 8,017684 D.T. de la vble. dep. 11,30407
Suma de cuad. residuos 2940,315 D.T. de la regresión 5,622837

Contraste de normalidad de los residuos -
Hipótesis nula: el error se distribuye normalmente
Estadístico de contraste: Ji cuadrado(2) = 28,7423
con valor p = 5,73715e-007

Fuente: Elaboración propia

Tabla 7: Modelo 4 - MCO combinados (log)

Modelo 4: MCO combinados, utilizando 75 observaciones
Se han incluido 19 unidades de sección cruzada
Largura de la serie temporal: mínimo 3, máximo 5
Variable dependiente: l_p_censo

	Coefficiente	Desv. típica	Estadístico t	Valor p
const	-1,95888	0,313214	-6,254	2,43e-08 ***
l_Google	1,54962	0,150676	10,28	7,61e-016 ***
Media de la vble. dep.	0,763821	D.T. de la vble. dep.	2,252992	
Suma de cuad. residuos	153,3847	D.T. de la regresión	1,449538	
R-cuadrado	0,591651	R-cuadrado corregido	0,586058	
F(1, 73)	105,7689	Valor p (de F)	7,61e-16	
Log-verosimilitud	-133,2502	Criterio de Akaike	270,5003	
Criterio de Schwarz	275,1353	Crit. de Hannan-Quinn	272,3510	
rho	-0,383599	Durbin-Watson	2,031240	

Contraste de heterocedasticidad de White -
Hipótesis nula: No hay heterocedasticidad
Estadístico de contraste: LM = 5,64767
con valor p = P(Ji cuadrado(2) > 5,64767) = 0,0593778

Contraste de normalidad de los residuos -
Hipótesis nula: el error se distribuye normalmente
Estadístico de contraste: Ji cuadrado(2) = 36,8451
con valor p = 9,98131e-009

Fuente: Elaboración propia

Cita recomendada

GÓMEZ MARTÍNEZ, Raúl; PRADO ROMÁN, Camilo (2014). «La actividad de búsquedas en Google anticipa los resultados electorales». *IDP. Revista de Internet, Derecho y Política*. N.º 18, pág. 2-15. UOC. [Fecha de consulta: dd/mm/aa].

<<http://journals.uoc.edu/index.php/idp/article/view/n18-gomez-prado/n18-gomez-prado-es>>
 <<http://dx.doi.org/10.7238/idp.v0i18.1941>>



Los textos publicados en esta revista están -si no se indica lo contrario- bajo una licencia Reconocimiento-Sin obras derivadas 3.0 España de Creative Commons. Puede copiarlos, distribuirlos y comunicarlos públicamente siempre que cite a su autor y la revista y la institución que los publica (IDP. Revista de Internet, Derecho y Política; UOC); no haga con ellos obras derivadas. La licencia completa se puede consultar en <http://creativecommons.org/licenses/by-nd/3.0/es/deed.es>.

Sobre los autores

Raúl Gómez Martínez
 raul.gomez.martinez@urjc.es

Departamento de Economía de la Empresa
 Facultad de Ciencias Jurídicas y Sociales
 Universidad Rey Juan Carlos de Madrid

Despacho 283
 Edificio Departamental I (Vicálvaro)
 Paseo de Artilleros, s/n
 28032 - Madrid - España

Licenciado en Ciencias Económicas y Empresariales, máster en Análisis económico y economía financiera, estudios del máster oficial de Dirección de empresas y doctor en Economía de la Empresa y Finanzas por la Universidad Rey Juan Carlos de Madrid.

Desde el año 2007 viene impartiendo y coordinando asignaturas enmarcadas en los planes docentes de la Universidad Rey Juan Carlos de diplomaturas, licenciaturas, grados y posgrados, como Valoración y Adquisición de Empresas, Mercados Financieros Nacionales, Gestión Bancaria, Productos Financieros, Asesoramiento y Planificación Financiera, Dirección Financiera, Mercados de Valores, Finanzas Corporativas, etc. Puedes seguir su actividad docente en: <<http://www.culturillafinanciera.com/>>. Participación en proyectos de investigación como el del diccionario económico y financiero de la Real Academia de Ciencias Económicas y Financieras (RACEF). Varios artículos publicados en prensa, revistas de investigación y participaciones en congresos y seminarios. Puedes seguir sus publicaciones en: <<http://elblogderaulgomez.blogspot.com/>>. Más de quince años de experiencia colaborando en instituciones financieras de primer nivel como Grupo Santander, BBVA, Bankia, Banesto, Caja Rural, ING Direct, etc., principalmente en proyectos de consultoría relacionados con tesorería, valores, productos de inversión y aseguramiento de calidad de software. Puedes seguir su trayectoria profesional en: <<http://www.linkedin.com/pub/raul-gomez-martinez/27/a69/3b2>>.

Camilo Prado Román
camilo.prado.roman@urjc.es
Departamento de Economía de la Empresa
Facultad de Ciencias Jurídicas y Sociales
Universidad Rey Juan Carlos de Madrid

Despacho 271
Edificio Departamental I (Vicálvaro)
Paseo de Artilleros, s/n
28032 - Madrid - España

Doctor en Ciencias Económicas y Empresariales (Área de Finanzas) por la Universidad Rey Juan Carlos (URJC). Especialista en Metodología de la investigación cuantitativa y técnicas estadísticas por la Universidad Politécnica de Madrid. Licenciado en Administración y Dirección de Empresas, especialidad Auditoría y dirección de empresas por el Colegio Universitario de Estudios Financieros (CUNEF). Diez años de docente en la Universidad Rey Juan Carlos de Madrid, dentro del Área de Economía Financiera y Contabilidad, impartiendo asignaturas como Introducción a la empresa, Mercados financieros nacionales, Gestión de carteras, Planificación financiera, Valoración de empresas, Dirección financiera, Inversiones alternativas o Finanzas aplicadas al marketing. Analista financiero en la Central de Balances del Banco de España en 2002. Personal docente e investigador de la URJC desde noviembre 2003, en la que desarrolla actividades de investigación en el campo de la inversión en activos no financieros, específicamente en la economía del coleccionismo, legitimidad organizativa, gestión de la calidad y gestión de carteras.