



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA



Valencia, Septiembre 2014

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

FACULTAD DE ADMINISTRACIÓN Y DIRECCIÓN DE EMPRESAS

TRABAJO DE FIN DE GRADO

Grado en Administración y Dirección de Empresas

“PREVISIÓN DE LA DEMANDA TURÍSTICA MEDIANTE
INDICADORES DE INTERNET”

PRESENTADO POR:

Ana María Alexandra Bota

DIRIGIDO POR:

Josep Domènech i de Soria

ÍNDICE GENERAL

CAPÍTULO 1. INTRODUCCIÓN.

1.1. Resumen	4
1.2. Introducción	5
1.3. Objetivos del TFG	6
1.4. Justificación de asignaturas relacionadas	6

CAPÍTULO 2. SITUACIÓN ACTUAL: SISTEMAS DE NOWCAST Y GOOGLE TRENDS

2.1. Introducción	8
2.2. Tema elegido para hacer modelos nowcast	8
2.3. Sistemas de nowcast	
2.3.1. ¿QUÉ SON LOS SISTEMAS DE NOWCAST?	13
2.3.2. APLICACIONES DE LOS MODELOS DE NOWCAST	13
2.3.2.1. Modelos de Nowcast para la predicción del nivel de desempleo	13
2.3.2.2. Modelos de Nowcast para la predicción del nivel de inflación	14
2.3.2.3. Modelos de Nowcast para la predicción del PIB	14
2.3.2.4. Modelos de nowcast para la predicción de la entrada de turistas	15
2.4. Google Trends	
2.4.1. ¿QUÉ ES GOOGLE TRENDS?	15
2.4.2. APLICACIONES DE GOOGLE TRENDS	17
2.4.2.1. <i>Ámbito General</i>	17
2.4.2.2. <i>Ámbito económico</i>	17
2.4.2.3. <i>Ámbito del turismo</i>	18

CAPÍTULO 3. METODOLOGÍA

3.1. Variables utilizadas	
3.1.1. DATOS DEL INE	20
3.1.1.1. Descripción del INE	20
3.1.1.2. Funcionamiento	20
3.1.1.3. Registro de la pernoctación de turistas	21
3.1.2. DATOS DEL GOOGLE TRENDS	
3.2. Muestra	24
3.2.1. SERIES DEL INE	24

3.2.2 SERIES DEL GOOGLE TRENDS -----	27
3.3. Modelos propuestos	
3.3.1. MODELO DE REFERENCIA -----	30
3.3.2. MODELO CON DATOS DEL GOOGLE TRENDS -----	31
3.3.3. MODELOS 1 Y 2 CON DATOS DE UN PAÍS CONCRETO -----	32
<i>CAPÍTULO 4. RESULTADOS</i>	
4.1. Análisis del modelo de referencia -----	34
4.2. Análisis del modelo con Google Trends -----	35
4.3. Validación del modelo -----	37
4.3.1. CONTRASTE DE ERRORES DE ESPECIFICACIÓN -----	38
4.3.2. ANÁLISIS DE LA NORMALIDAD DE LAS PERTURBACIONES -----	40
4.3.3 CONTRASTES DE SIGNIFICATIVIDAD -----	41
<i>4.3.3.1 Contrastes de significatividad individual</i> -----	<i>41</i>
<i>4.3.3.2. Contrastes de significatividad conjunta</i> -----	<i>43</i>
4.3.4. ANÁLISIS DE LA HETEROCEDASTICIDAD -----	44
<i>4.3.4.1. Método gráfico</i> -----	<i>44</i>
<i>4.3.4.2. Método analítico</i> -----	<i>45</i>
4.3.5. ANÁLISIS DE LA AUTOCORRELACIÓN -----	47
4.4. Análisis y Validación de los restantes modelos -----	49
4.4.1. ANÁLISIS Y VALIDACIÓN DE LOS MODELOS CON DATOS DE FRANCIA -----	49
<i>4.4.1.2. Validación del modelo con datos del Google Trends</i> -----	<i>51</i>
4.4.2. VALIDACIÓN DEL MODELO DE GOOGLE TRENDS CON DATOS DEL REINO UNIDO -----	56
<i>4.4.2.1. Validación del modelo con datos del Google Trends</i> -----	<i>60</i>
4.4.3. VALIDACIÓN DEL MODELO DE GOOGLE TRENDS CON DATOS DE ALEMANIA -----	65
<i>4.4.3.1. Validación del modelo con datos del Google Trends</i> -----	<i>68</i>
4.5. Análisis de las hipótesis planteadas -----	74
<i>CAPÍTULO 5. CONCLUSIONES</i> -----	<i>77</i>
<i>CAPÍTULO 6. BIBLIOGRAFÍA</i>	
6.1. Bibliografía -----	80
6.2. Anexos -----	82

CAPÍTULO 1. INTRODUCCIÓN

1.1. Resumen

El uso de Internet a la hora de viajar ha aumentado exponencialmente en los últimos años. Google, el buscador más utilizado en Internet, ofrece una herramienta online (Google Trends) que proporciona la frecuencia de búsqueda en Internet de determinadas palabras desde el año 2004. Estas frecuencias se han usado en los últimos años para realizar predicciones a corto plazo sobre una gran variedad de series económicas y sociales en países donde hay un gran uso de Internet. Recientemente ha sido usada por Choi y Varian (2011) para el pronóstico de la situación presente de diversas series económicas (nowcast).

El presente Trabajo de Fin de Grado (TFG) tiene como objetivo validar el uso de Google Trends para predecir el volumen de pernoctaciones de turistas a la ciudad de Valencia en el corto plazo. Para ello, se han escogido los cuatro países de procedencia con mayor volumen de turistas a esta ciudad.

Dado que el turismo se ha convertido en el motor de la economía valenciana, a los establecimientos del sector les sería de utilidad un modelo que les proporcione una estimación lo más acertada posible de la pernoctación de turistas.

La metodología utilizada en este TFG se basa en el uso de modelos de predicción inmediata (nowcast). Numerosos artículos científicos tratan sobre el uso de estos modelos con datos del Google Trends e incluso el Banco de España publicó en 2012 un documento¹ en el cual se habla del posible uso de Google Trends para la previsión de variables de demanda y actividad en España. Los datos numéricos han sido obtenidos a través del Instituto Nacional de Estadística y de la herramienta Google Trends.

A lo largo de este trabajo se podrá apreciar que el término de búsqueda Valencia-City in Spain tiene una evolución estacional similar a la del número de pernoctaciones de turistas a Valencia y además que los turistas suelen realizar la búsqueda con antelación o el mismo mes en que van a ir de viaje, hecho que recalca la importancia de un modelo de predicción “nowcast”.

¹ Documento Ocasional Nº.1203 llamado “Las huellas del futuro están en la web: Construcción de indicadores adelantados a partir de las búsquedas en Internet”

1.2. Introducción

El uso de Internet para buscar información de todo tipo se ha convertido en una necesidad básica para las personas de todo el mundo. Tanto la información almacenada en Internet como las búsquedas de la misma quedan registradas. Estos registros de actividad online tienen un valor cada vez mayor para los investigadores, los cuales ven en ellos una fuente que pueden utilizar a la hora de hacer análisis económicos, sociales o de cualquier otro tipo. Es más, se trata de una fuente sin coste alguno que permite hacer análisis en tiempo real.

En los últimos años se ha escrito mucho sobre los indicadores económicos que ofrece internet y su utilidad para hacer análisis a corto plazo. Para la extracción de estos indicadores se ha hecho alto uso de la herramienta Google Trends, una herramienta online de Google que permite visualizar los términos de búsqueda más populares en un período de tiempo específico. Dado que esta herramienta permite generar indicadores en tiempo real, ya ha sido usada por varios investigadores. Entre otros, Ginsberg, J. et al. (2009) la ha usado para predecir la evolución de la gripe; Choi y Varian (2009) para predecir las ventas al por menor; Askitas y Zimmermann (2009) para mejorar la predicción del nivel de desempleo y Preis et al. (2010) para predecir el volumen de transacciones en bolsa.

Con respecto a la predicción de la entrada de turistas Choi y Varian (2009) y Gawlik et al. (2011) demostraron que la evolución de la popularidad de búsquedas del Google Trends ofrece un predictor útil de las tasas de turismo. El Banco de España publicó en 2012 un documento, redactado por Concha Artola y Enrique Galán, en que se proponía una aplicación muy específica del uso de Google Trends para la economía española: los flujos de turistas británicos entrados en España -los principales clientes de la industria turística española-. En el mismo se concluyó que los modelos que incluyen como indicadores los datos del Google Trends mejoraban la predicción dependiendo del modelo tomado como referencia, recalcando que esto implica que la información contenida en las búsquedas de Google ha de juzgarse en el contexto adecuado y que además esta herramienta permitía obtener indicadores ajustados del flujo de turistas británicos con casi un mes de adelanto.

La estructura de este trabajo se divide en cuatro partes, sin contar la introducción. El Capítulo 2 trata sobre el marco teórico referente a los modelos del nowcast en general y en particular a los que incluyen datos del Google Trends. El Capítulo 3 hace referencia a la metodología empleada, haciendo una exposición de las variables, la muestra y los modelos econométricos. En el Capítulo 4 se exponen las estimaciones de los modelos propuestos y se realiza su validación econométrica. Y por último, en el Capítulo 5 se presentan las conclusiones.

1.3. Objetivos del TFG

El objetivo principal de este TFG es el de validar el uso de Google Trends para predecir el volumen de pernoctaciones de turistas a la ciudad de Valencia en el corto plazo.

Objetivos derivados:

- Justificar que los datos obtenidos a través de Google Trends mejoran la predicción del número de pernoctaciones de turistas a Valencia que puede hacer un modelo de regresión básico (modelo de referencia con datos del INE).

-Proponer un modelo de predicción de turistas en Valencia desde los principales orígenes de estos.

1.4. Justificación de asignaturas relacionadas

Como justificación de asignaturas relacionadas se tiene que:

- Para el Capítulo 2 han sido útiles, en primer lugar, los conocimientos impartidos en la asignatura de Macroeconomía, en la cual se ha visto cómo hacer un estudio general de la economía de un país analizando las variables agregadas. En este caso, el ingreso por turismo representa una parte importante del PIB tanto a nivel nacional como a la ciudad de Valencia -objeto de nuestro estudio- y es por ello que resulta interesante el análisis del presente TFG. Por otra parte, los conocimientos adquiridos en las asignaturas de Economía Española y Mundial han sido útiles al ser el sector turístico un factor importante en la economía española y los países de procedencia de los turistas un reflejo de la economía de mercados internacionales.
- Para el Capítulo 3, han sido útiles los conocimientos de las asignaturas Métodos Estadísticos en Economía para la obtención de datos, e Introducción a la Estadística para la descripción estadística de los mismos. Asimismo, en la asignatura Econometría se adquieren los conocimientos para plantear los modelos econométricos regresión.
- Para el Capítulo 4 los conocimientos adquiridos en Métodos Estadísticos en Economía son útiles para entender las estimaciones obtenidas en los modelos y los adquiridos en Econometría para realizar la validación de los mismos.

CAPÍTULO 2. SITUACIÓN ACTUAL

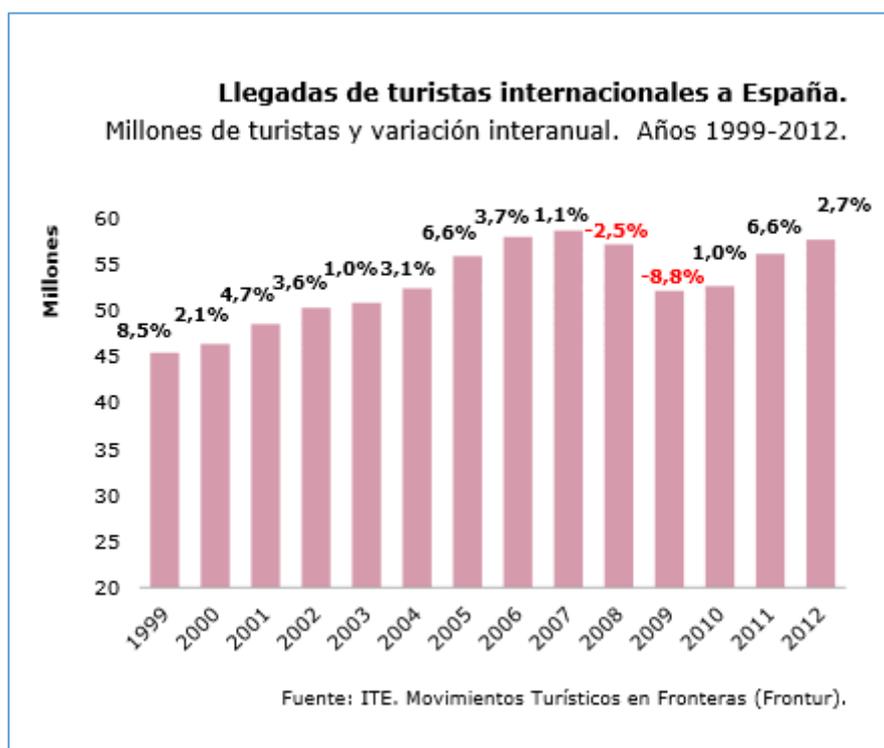
2.1. Introducción

En este capítulo se explican los modelos de nowcast en general y, en concreto, los modelos econométricos que incluyen datos del Google Trends. En primer lugar, se hace una explicación de la importancia del tema elegido para hacer modelos de predicción nowcast. A continuación, se expone el significado y funcionamiento de los modelos de nowcast, así como, de Google Trends. Por último, se muestran algunas investigaciones relevantes que han introducido en sus modelos datos del Google Trends.

2.2. Tema elegido para hacer modelos nowcast

En el Informe Anual del Instituto de Turismo de España (2012) se expone la importancia del sector turístico a España.

GRÁFICO 1: LLEGADA DE TURISTAS INTERNACIONALES A ESPAÑA (1999-2012)



FUENTE: INFORME ANUAL DEL TURISMO DE ESPAÑA (2012)

En el gráfico 1, se puede observar que la llegada de turistas tiene tendencia ascendente en el período mostrado, habiendo tenido descenso únicamente en 2008 y 2009, debido a la coyuntura económica de esos años.

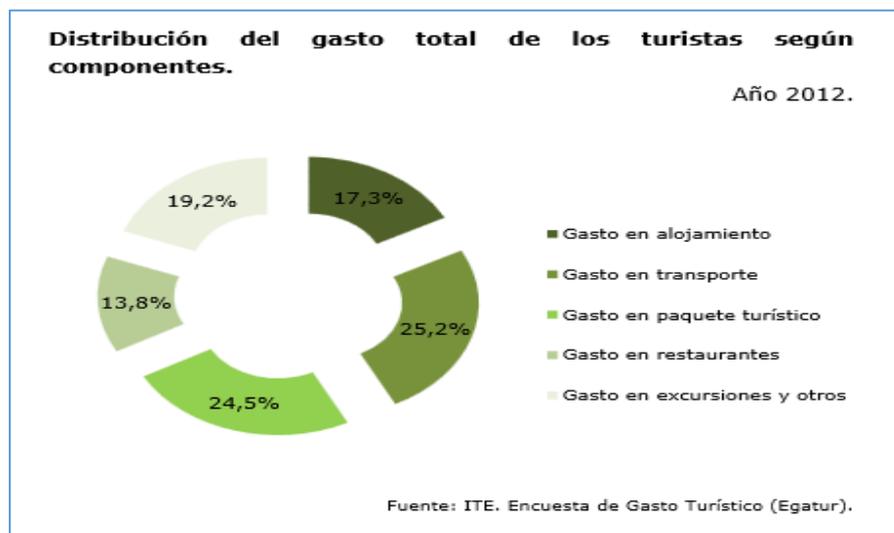
GRÁFICO 2. GASTO TOTAL DE TURISTAS INTERNACIONALES A ESPAÑA (2004-2012)



FUENTE: INFORME ANUAL DEL TURISMO DE ESPAÑA (2012)

En el Gráfico 2 se observa que el gasto total de dichos turistas también ha aumentado en el período de referencia, habiendo descendido únicamente en 2009, como consecuencia del descenso visto en el Gráfico 1.

GRÁFICO 3. DISTRIBUCIÓN DEL GASTO TOTAL DE LOS TURISTAS (2012)

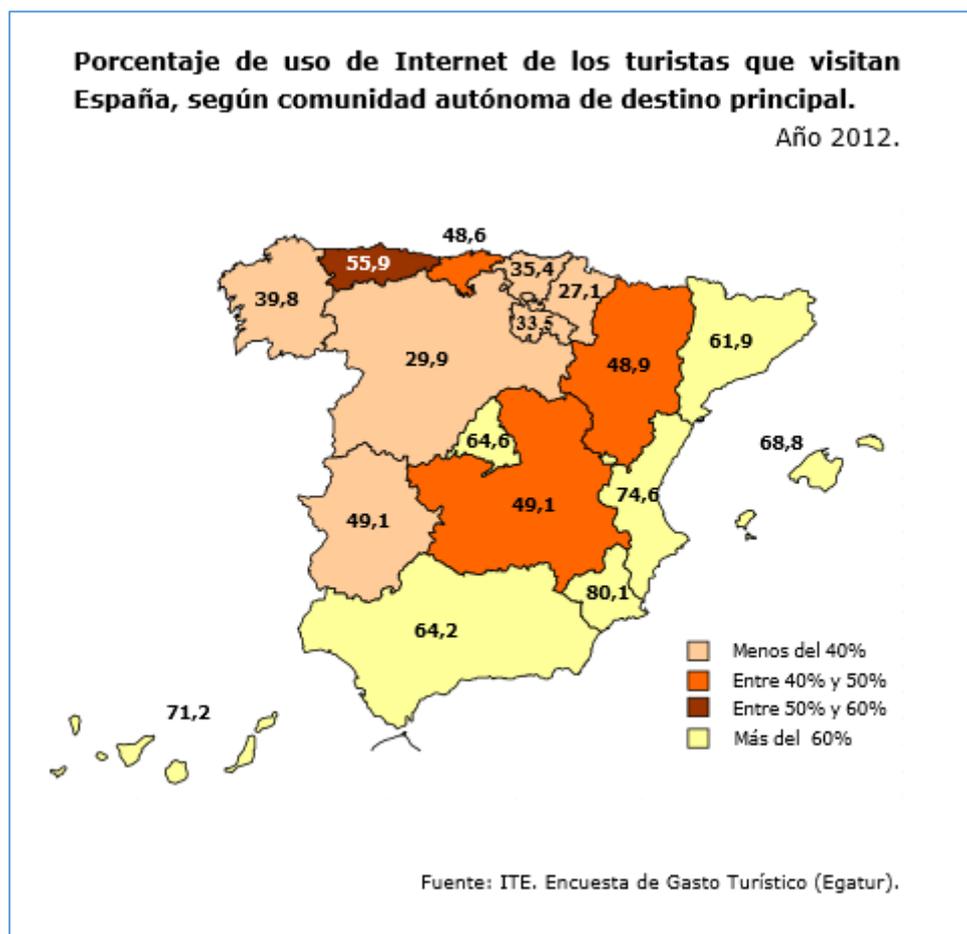


FUENTE: INFORME ANUAL DEL TURISMO DE ESPAÑA (2012)

El Gráfico 3 muestra que la mitad del gasto de estos turistas se da en el transporte y paquete turístico y la otra mitad en alojamiento, restaurantes, excursiones y otros.

Esto justifica el interés del presente trabajo para el sector turístico del país, ya que, a hoteles y restaurantes les sería de mucha ayuda poder saber a corto plazo una cifra aproximada de los turistas o, mejor aún, de su pernoctación (conllevando esta última más ingreso para los mismos).

GRÁFICO 4. PORCENTAJE DE USO DE INTERNET DE LOS TURISTAS QUE VISITAN ESPAÑA,
SEGÚN COMUNIDAD AUTÓNOMA DE DESTINO PRINCIPAL.

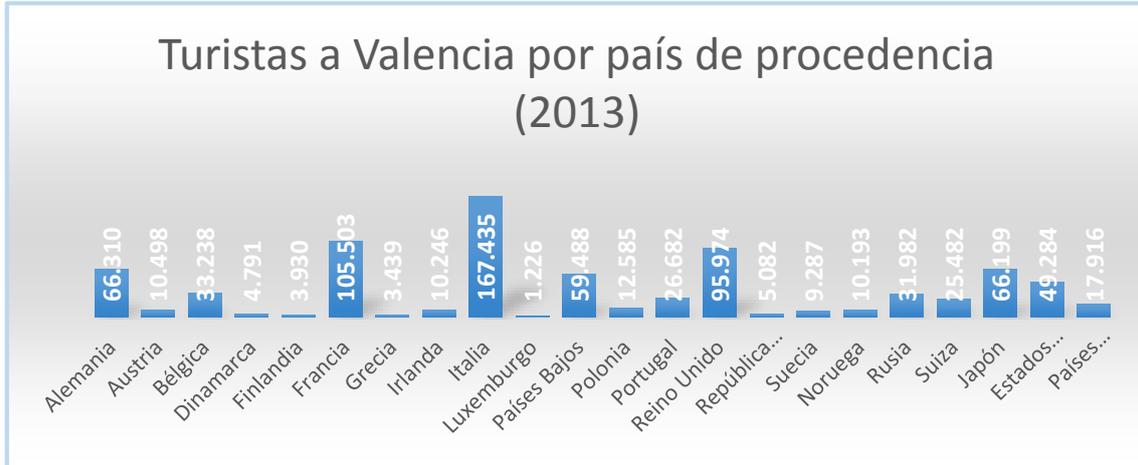


FUENTE: INFORME ANUAL DEL TURISMO DE ESPAÑA (2012)

En el Gráfico 4 se puede ver que el porcentaje de uso de Internet de los turistas que visitan la Comunidad Valenciana es de aproximadamente 75%, porcentaje superado únicamente por el uso de los turistas de la región de Murcia.

Se justifica así el uso de Google Trends para un estudio turístico a la Comunidad Valenciana.

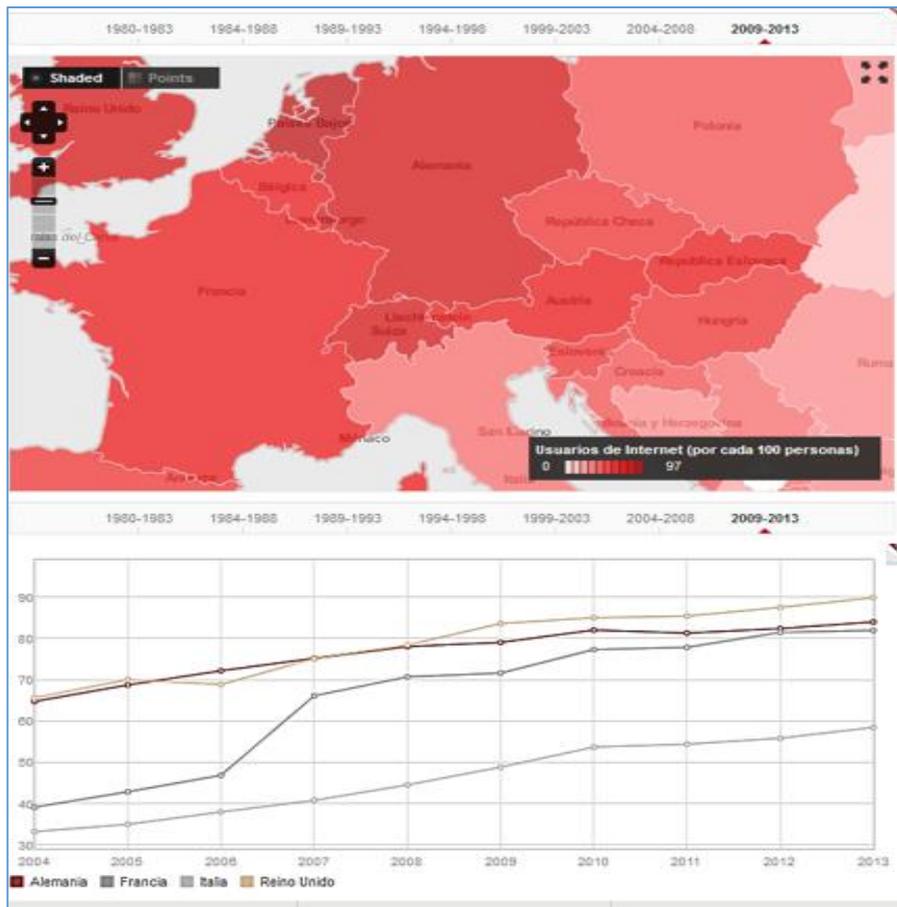
GRÁFICO 5. TURISTAS A VALENCIA POR PAÍS DE PROCEDENCIA (2013)



FUENTE: ELABORACIÓN PROPIA CON DATOS DEL INE

Centrando el presente trabajo a la provincia de Valencia, en el Gráfico 5 se puede observar que los principales países de procedencia de los turistas a esta provincia son Italia, Francia, Reino Unido y Alemania.

GRÁFICO 6. USUARIOS DE INTERNET (POR CADA 100 PERSONAS)

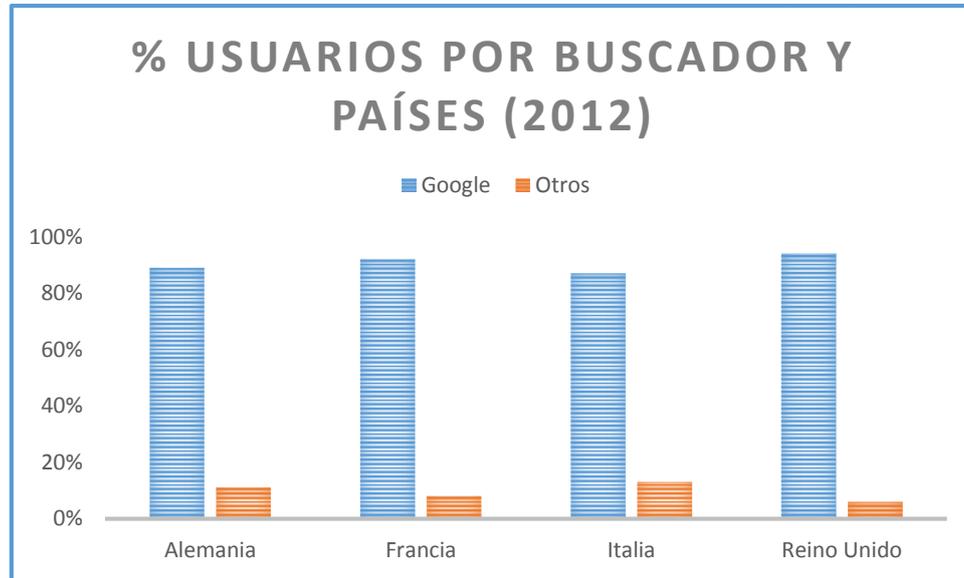


FUENTE: BANCO MUNDIAL ²

² <http://datos.bancomundial.org/indicador/IT.NET.USER.P2/countries/DE-FR-IT-GB?display=graph>

En el Gráfico 6 se puede ver que el uso de Internet está muy extendido a Alemania, Reino Unido y menos extendido a Francia e Italia. No obstante, dicho uso ha aumentado en los cuatro países durante el período analizado.

GRÁFICO 7. PORCENTAJE DE USUARIOS POR BUSCADOR Y PAÍSES (2012)



FUENTE: [HTTP://WWW.POSICIONAMIENTOINTERNACIONAL.COM/CUOTA-DE-MERCADO-DE-BUSCADORES-POR-PAISES/](http://www.posicionamientointernacional.com/cuota-de-mercado-de-buscadore-por-paises/)

Siendo Google el buscador más utilizado por los países de procedencia de turistas objeto de análisis en el presente trabajo, se justifica el uso de sistemas de nowcast con datos del Google Trends.

2.3. Sistemas de nowcast

2.3.1. ¿QUÉ SON LOS SISTEMAS DE NOWCAST?

Los sistemas de nowcast hacen referencia a los sistemas de predicción inmediata, es decir, predicciones del presente, del pasado reciente y del futuro cercano, siendo utilizados principalmente en ámbitos económicos y meteorológicos. El principio básico de este sistema se basa, por tanto, en explotar la información disponible publicada en busca de obtener una estimación con anterioridad a la siguiente publicación oficial de la variable en cuestión.

Tradicionalmente, la predicción inmediata se hacía mediante métodos heurísticos. Giannone et al. (2008) demostraron en un artículo³ que el proceso de predicción inmediata se puede formalizar en un modelo estadístico que produce predicciones sin necesidad de juicio informal. Sobre el ejemplo de predicción del PIB, expusieron que las señales sobre la dirección del cambio en el PIB se pueden extraer de un conjunto amplio y heterogéneo de fuentes de información (cifras de desempleo, pedidos industriales, balanza comercial, etc.) antes de que el PIB en sí sea publicado. Y que estos datos se pueden utilizar para calcular las secuencias de las estimaciones actuales trimestre del PIB en relación con el flujo en tiempo real de los datos publicados.

En cuanto al tipo de información, Bańbura et al. (2012) hace referencia a que tanto la información “hard” (dura), como son los informes de producción industrial, y la información “soft” (blanda), como son las encuestas, pueden proporcionar una estimación previa de los acontecimientos actuales de la economía.

2.3.2. APLICACIONES DE LOS MODELOS DE NOWCAST

Dado que para el presente trabajo nos interesan los modelos de nowcast en el ámbito económico, a continuación se exponen algunas de las aplicaciones más relevantes en esta materia.

2.3.2.1. Modelos de Nowcast para la predicción del nivel de desempleo

Los primeros que propusieron un modelo de nowcast para predecir el número de personas que acababan de ser desempleadas, basándose en los historiales de búsqueda de diversos buscadores fueron Ettredge et al. (2005). Su importancia radica en que en ellos se inspiraron los siguientes autores a partir de que Google ofreciera la herramienta Google Trends.

³ Giannone, Domenico y Reichlin, Lucrezia y Pequeños, David, 2008 "predicción inmediata: El contenido de la información en tiempo real de los datos macroeconómicos," *Journal of Monetary Economics*, Elsevier, vol. 55 (4), páginas 665-676, mayo.

2.3.2.2. Modelos de Nowcast para la predicción del nivel de inflación

En lo que respecta a modelos de nowcast para la predicción del nivel de inflación se puede destacar en la zona euro a Lenza y Qarmedinger (2010); Monteforte y Moretti (2010) y a Modugno (2011).

2.3.2.3. Modelos de Nowcast para la predicción del PIB

El uso de sistemas de nowcast para predecir el nivel del PIB ha sido el más extendido a nivel internacional.

- En Estados Unidos se puede nombrar a Giannone et al. (2008), Lahiri y Monokroussos (2011) y a Aastveit et al. (2011).
- En la zona euro a Angelini et al. (2010); Bańbura y Modugno (2010); Camacho y Pérez-Quirós (2010); Giannone et al. (2011) y Bańbura y Rünstler (2011).
- En Francia a Barhoumi et al. (2010).
- En Irlanda a Liebermann (2012).
- En Noruega a Aastveit y Trovik (2012).
- En España a Liedo y Fernandez (2010).

Conclusiones de los nombrados investigadores:

- La predicción del nivel de PIB se ve afectada, principalmente, por el trimestre actual (y el anterior).
- Las predicciones de modelos nowcasting son equiparable a las predicciones institucionales.
- Los modelos de nowcast mejoran sus predicciones conforme la información relevante se acumula, lo cual sugiere la necesidad de incorporar nuevos datos al modelo tan pronto como sean publicados (Bańbura, Giannone y Reichlin, 2011).

2.3.2.4. Modelos de nowcast para la predicción de la entrada de turistas

Gawlik et al. (2011) demostraron que la evolución de la popularidad de búsquedas del Google Trends ofrece un predictor útil de las tasas de turismo.

En España, Concha Artola y Enrique Galán (2012) publicaron en un documento del Banco de España² una aplicación específica del uso de Google Trends para la economía española -los flujos de turistas británicos entrados en España-. Como conclusión se obtuvo que los modelos nowcast con datos del Google Trends mejoraban la predicción y que estos datos permitían obtener indicadores ajustados del flujo de turistas británicos con casi un mes de adelanto. La mejora en la predicción dependía del modelo tomado como referencia, haciéndoles ver que la información contenida en las búsquedas de Google ha de juzgarse en el contexto adecuado.

2.4. Google Trends

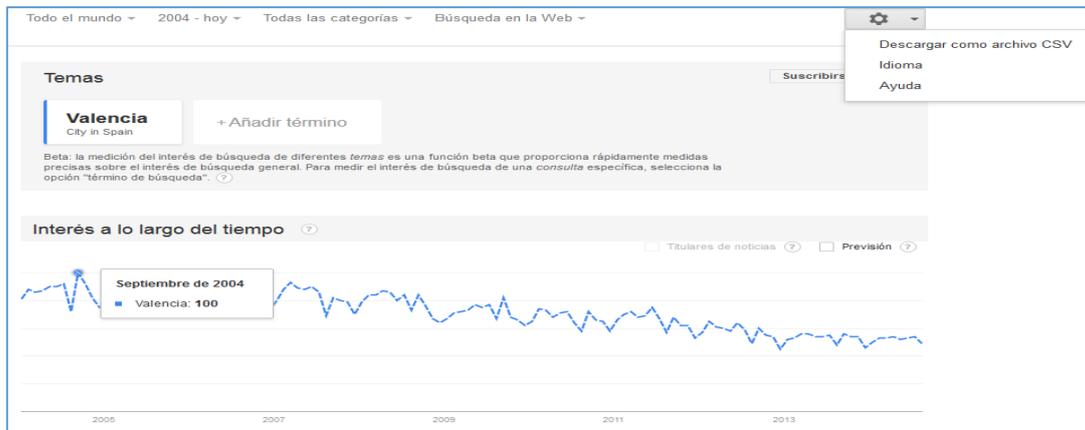
2.4.1. ¿QUÉ ES GOOGLE TRENDS?

Google Trends es una herramienta online, publicada por Google Labs el 10 de mayo de 2006 y cuya función es la de publicar series temporales del volumen de búsquedas que se han hecho de una determinada palabra o categoría en una región geográfica determinada (en todo el mundo, en un país, una ciudad...). Estos valores no son absolutos, como podría ser el número total de búsquedas de una determinada palabra, sino valores relativos, normalizados por Google Trends para que sean comparables.

El proceso de normalización que realiza consiste en dividir conjuntos de datos de búsqueda por variables comunes -como el número total de búsquedas-, para así, anular el efecto de las variables en los datos. El resultado de este proceso es la publicación de datos estandarizados mediante un índice de 0 a 100 (Google Index). Por tanto, los datos de dos regiones con volúmenes de búsqueda diferentes se pueden comparar de igual forma (ya que los datos se han normalizado en función del número total de búsquedas de cada región).

Las series temporales que ofrece se publican semanalmente desde el año 2004. En la Figura 1 se ve a modo de ejemplo la serie temporal del volumen de búsqueda de la palabra *Valencia* en la subcategoría de *City in Spain* en todo el mundo desde el 2004 hasta la actualidad, alcanzando dicho volumen el índice de popularidad máxima (100) en el mes de septiembre de 2004. Para ver qué semana exacta de ese mes habría descargar la serie como archivo CSV, obteniendo con ello el detalle semanal de todos los meses.

FIGURA 1. CAPTURA DE PANTALLA DE GOOGLE TRENDS



FUENTE: GOOGLE TRENDS.

La búsqueda puede hacerse especificando la palabra en cuestión cuya popularidad se desea conocer o escogiendo una categoría entre las que ofrece Google (Viajes, Arte y entretenimiento, Aficiones y tiempo libre...).

Por otro lado, se puede especificar en qué región nos interesa el volumen de búsqueda de la consulta en cuestión, pudiendo escoger 1 o más países o regiones a modo de comparación de sus volúmenes de búsqueda. Para ello, Google Trends normaliza los datos de manera que el índice 100 se alcance por el país que más volumen de búsqueda tenga en una semana/mes concreto. Es decir, la normalización se haría dividiendo los datos absolutos de cada país por el total de búsquedas de los cuatro países en el período determinado.

En la Figura 2 se puede apreciar la evolución comparada de la popularidad online del término *Valencia* (*City in Spain*) de cuatro países concretos, desde enero de 2004 hasta agosto de 2014, siendo la popularidad máxima alcanzada en Italia el mes de Abril de 2007.

FIGURA 2: CAPTURA DE PANTALLA DEL GOOGLE TRENDS



FUENTE: GOOGLE TRENDS

2.4.2. APLICACIONES DE GOOGLE TRENDS

2.4.2.1. *Ámbito General*

Dado que la descarga de datos estadísticos mediante la opción *Descargar como archivo CSV* solamente está disponible desde 2008, es a partir de este año que esta aplicación se comienza a utilizar estadísticamente.

El primer uso de Google Trends estadísticamente se hizo en el campo de la medicina por Ginsberg et al. (2009) y mediante él se demostró que la búsquedas online de 45 términos asociados con la gripe era útil para predecir brotes de gripe con anterioridad a la publicación de informes del CDC⁴. Este estudio sirvió a los siguientes investigadores para justificar la utilidad predictiva de Google Trends en la investigación.

La siguiente investigación con datos del Google Trends la hicieron Choi y Varian (2009) en el campo de las ventas, demostrando su utilidad para predecir ventas de inmuebles, automóviles y turismo (ver apartado 1.3.2.3).

En 2010, Goel investigó la utilización de las búsquedas de Google para predecir la venta de videojuegos el primer mes de su salida al mercado, los ingresos en traquilla de la industria del cine, así como, para predecir el ranking de canciones en la lista de Bollboard Hot 100. Como conclusión obtuvo que Google Trends sí era útil para predecir los días de ventas con antelación, pero no tenía tanta precisión para la predicción de canciones.

2.4.2.2. *Ámbito económico*

Uno de los primeros usos de mayor interés en el ámbito económico ha sido el estudio realizado por Da, Engelberg y Gao (2010). En él se demostró que el incremento en un conjunto de búsquedas online determinadas (como recesión, bancarrota...), el cual es signo de una negativa expectativa económica, conllevaba una volatilidad extrema en el flujo de los fondos de inversión. La alta correlación entre dichas búsquedas y el flujo de los fondos de inversión demostraba la utilidad predictiva de Google Trends para anticipar la situación económica próxima de inversión.

Otro estudio de interés en este ámbito es el de Guzman (2011), el cual demostró la utilidad de Google Trends para predecir la inflación con el menor error de predicción de entre todos los indicadores validados.

Por último, cabe destacar que ha habido numerosas publicaciones que demostraron el poder predictivo de Google Trends para el nivel de desempleo. Entre ellas se pueden nombrar Choi y Varian (2009), D'amuri y Marcucci (2009) y Askitas y Zimmermann (2009), destacando estos últimos por predecir el nivel agregado de desempleo en Alemania.

⁴ Centro para el Control Prevención de Enfermedades.

2.4.2.3 *Ámbito del turismo*

Choi y Varian (2009) realizaron uno de los primeros estudios de predicción, en el campo del sector turístico, con datos del Google Trends. Su objetivo era el de predecir las tasas de turismo en países específicos. Para ello realizaron un análisis de regresión de series de tiempo para predecir las estadísticas de la llegada de visitantes, incorporando los índices de volumen de búsqueda del Google Trends como datos de regresión.

Concretamente, se utilizaron como datos el término de búsqueda “Hong Kong” desde 9 países diferentes y los datos de los informes del Departamento de Turismo de Hong Kong. La idea era comprobar si un incremento en las búsquedas de este término desde los países analizados suponía un incremento en la cantidad de turistas, cosa que se demostró cierta al concluir que existía una alta correlación entre los dos parámetros.

Posteriormente, Gawlik et al. (2011) presentaron un enfoque más amplio al problema de la predicción de las tasas de turismo. Utilizaron consultas específicas de datos de búsqueda y aplicaron un algoritmo de selección de función para elegir la consulta más relevante. Dicho algoritmo lo evaluaron utilizando k-veces la validación cruzada y consiguieron no sólo un error bajo en el modelo –cosa que ya habían conseguido Choi y Varian- sino un error de ajuste bajo en la prueba.

En los dos estudios se demostró la utilidad del volumen de búsquedas que Google Trends ofrece para la predicción de la entrada de turistas. El objetivo del presente trabajo es validar el uso de datos del Google Trends para predecir el número de pernoctaciones de turistas a la ciudad de Valencia, con una precisión mayor que la ofrecida por un modelo de regresión básico -que no utilizan datos del Google Trends-, al que llamaremos modelo de referencia y en un periodo de tiempo anterior a publicaciones oficiales como las de INE. Esto último, se podrá conseguir una vez validado el modelo con datos del Google Trends, ya que la publicación de dichos datos es semanal –frente a la mensual del INE-.

Tras esta revisión bibliográfica, se pueden plantear las siguientes hipótesis:

H1: la popularidad del término de búsqueda *Valencia-City in Spain*⁵ ayuda a predecir el número de pernoctaciones de turistas a la ciudad de Valencia.

H2: la popularidad, en un país, del término de búsqueda *Valencia-City in Spain* mejora la predicción del número de pernoctaciones de turistas procedentes de ese país en la ciudad de Valencia.

⁵ Concepto de búsqueda elegido por ser el que más volumen de popularidad tiene (todas las semanas desde 2004 hasta la actualidad) en comparación con otras palabras relacionadas con las pernoctaciones de turistas a Valencia (Hotel in Valencia o Valencia dentro de la categoría viajes), las cuales presentan volúmenes de popularidad 0 en varias semanas de diferentes años.

CAPÍTULO 3. METODOLOGÍA

3.1. Variables utilizadas

3.1.1. DATOS DEL INE

3.1.1.1. Descripción del INE

El Instituto Nacional de Estadística se crea por la *Ley de 31 de diciembre de 1945* y es actualmente un organismo autónomo de carácter administrativo, con personalidad jurídica y patrimonio propio, adscrito al Ministerio de Economía y Competitividad a través de la Secretaría de Estado de Economía y Apoyo a la Empresa.

Su misión es la elaboración y perfeccionamiento de las estadísticas demográficas, económicas y sociales ya existentes, la creación de otras nuevas y la coordinación con los servicios estadísticos de las áreas provinciales y municipales.

Es referente nacional e internacional en la producción de estadísticas oficiales españolas y su visión está orientada a constituir un sistema estadístico basado en valores como la eficacia, eficiencia y colaboración.

Se rige por ley *Ley 12/1989, de 9 de mayo, de la Función Estadística Pública (LFEP)*, la cual le asigna un papel destacado en la actividad estadística pública. Entre otras funciones a destacar está la de ser responsable de la formación del Censo Electoral⁶.

3.1.1.2. Funcionamiento

El INE realiza encuestas y expone el resultado en estadísticas. Las estadísticas que realiza se pueden agrupar en tres grandes grupos temáticos, para los que cuenta con departamentos especializados:

-Cuentas nacionales y empleo (el mercado laboral en general). Se ocupa de la elaboración de estadísticas del mercado laboral y de las cuentas económicas con indicadores tan relevantes como la tasa de paro o el producto interior bruto.

-Estadísticas a empresas. En las que las principales fuentes de información son las empresas. Dentro de este grupo se encuentran los indicadores coyunturales, como el IPC, e información estructural sobre los distintos sectores de actividad, el medio ambiente, etc.

-Estadísticas de temática sociodemográfica. Entre ellas están los Censos de Población, que se realizan cada diez años debido a su enorme envergadura, o las estadísticas sanitarias o de condiciones de vida.

⁶ De acuerdo con lo establecido en la Ley Orgánica 5/1985, de 19 de junio, del Régimen Electoral General, y de la coordinación y gestión continua e informatizada del Padrón según lo establecido en la Ley 7/1985, de 2 de abril, Reguladora de las Bases del Régimen Local, y en el Real Decreto 2612/1996, de 20 de diciembre, por el que se modifica el Reglamento de Población y Demarcación Territorial de las Entidades Locales aprobado por el Real Decreto 1690/1986, de 11 de julio.

3.1.1.3. Registro de la pernoctación de turistas

Para entender mejor cómo trabaja se puede poner de ejemplo la Encuesta de ocupación hotelera -similar a las otras encuestas en alojamientos turísticos⁷- que realiza desde el año 1999, en el ámbito nacional, autonómico, provincial, así como en zonas y puntos turísticos y cuya publicación es mensual.

En la página oficial se puede ver que su objetivo es el de estudiar el comportamiento de una serie de variables⁸ que permitan describir la características fundamentales (oferta y demanda) del sector hotelero.

Los datos de la Encuesta reflejan la doble vertiente que se considera en el estudio del turismo: del lado de la demanda, se ofrece información sobre viajeros, pernoctaciones y estancia media distribuidos por país de residencia de los viajeros y categoría de los establecimientos que ocupan o por Comunidad o Ciudad Autónoma de procedencia en el caso de los viajeros residentes en España; del lado de la oferta se proporciona el número de establecimientos abiertos estimados, plazas estimadas, grados de ocupación e información sobre el empleo en el sector, según la categoría del establecimiento. Esta información se ofrece mensualmente, a nivel nacional, de comunidad autónoma, provincial, de zona turística y de punto turístico.

➤ Definición de variables

Establecimientos hoteleros abiertos estimados: se entiende por establecimientos hoteleros abiertos de temporada, aquel en el que el mes de referencia está comprendido dentro de su periodo de apertura.

Plazas estimadas: número de plazas estimadas por la encuesta de los establecimientos abiertos de temporada. El número de plazas equivale al número de camas fijas del establecimiento. No se incluyen, por tanto, las supletorias, y las de matrimonio dan lugar a dos plazas.

Viajeros entrados: todas aquellas personas que realizan una o más pernoctaciones seguidas en el mismo alojamiento. Los viajeros se clasifican por su país de residencia, y para los residentes en España, por la comunidad autónoma donde residen habitualmente.

Pernoctaciones o plazas ocupadas: se entiende por pernoctación cada noche que un viajero se aloja en el establecimiento. Al igual que en la entrada de viajeros, las plazas ocupadas se desglosan según el lugar de residencia.

Estancia media: esta variable es una aproximación al número de días que, por término medio, los viajeros permanecen en los establecimientos y se calcula como cociente entre las pernoctaciones y el número de viajeros.

⁷ Alojamientos turísticos considerados por el INE: Establecimientos Hoteleros, Acampamentos turísticos, Apartamentos Turísticos, Alojamientos de turismo rural.

⁸ Variables: viajeros, pernoctaciones, estancia media y grado de ocupación, por lugar de residencia, mes, tipo de alojamiento, categoría, ubicación. Personal ocupado por categoría, tipo de contratación y mes.

Grado de ocupación por habitaciones: relación, en porcentaje, entre la media diaria de habitaciones ocupadas en el mes y el total de habitaciones disponibles.

Grado de ocupación por plazas: relación, en porcentaje, entre el total de las pernoctaciones y el producto de las plazas, incluyendo las camas supletorias, por los días a que se refieren las pernoctaciones.

Grado de ocupación por plazas en fin de semana: relación, en porcentaje, entre las pernoctaciones del viernes y sábado que caen dentro de la semana de referencia y el producto de las plazas, incluyendo las camas supletorias de esos dos días, por los días a que se refieren las pernoctaciones, en este caso dos.

Personal ocupado: conjunto de personas, remuneradas y no remuneradas, que contribuyen mediante la aportación de su trabajo, a la producción de bienes y servicios en el establecimiento durante el mes que incluye el período de referencia de la encuesta, aunque trabajen fuera de los locales del mismo.

Punto turístico: municipio donde la concentración de la oferta turística es significativa.

Zona turística: conjunto de municipios en los que se localiza de forma específica la afluencia turística. Se ofrece información de las principales zonas de interés turístico.

➤ *Recogida de la información*

La consulta de datos básicos se refiere a siete días seguidos de cada mes, elegidos aleatoriamente, de tal manera que entre todos los establecimientos cubran el mes completo.

Se hace una segunda consulta sobre el total de viajeros alojados y de pernoctaciones causadas durante todo el mes al que se refieren los datos y sobre el periodo de apertura en el mes considerado; en los estratos de tres, cuatro y cinco estrellas de oro de todas las provincias.

La información es suministrada mensualmente por los establecimientos hoteleros, mediante cuestionario, al Instituto Nacional de Estadística. Es posible, asimismo, remitir la información mediante el envío telemático o carga de un fichero XML o a través de la conexión a internet mediante el sistema ARCE, cumplimentando directamente el cuestionario en pantalla.

➤ *Difusión de la información*

La información se presenta en diferentes niveles de desagregación geográfica: nacional, de comunidad autónoma, provincial, de zonas y de puntos turísticos. Se han considerado las zonas (conjunto de municipios), así como los municipios en los que se localiza de forma específica la afluencia turística.

Las estimaciones de los datos que se publican se realizan partiendo de los datos que contestan los establecimientos en la semana de referencia, excepto para los estratos de tres, cuatro y cinco estrellas de oro, en donde se calculan partiendo también de los totales mensuales.

➤ Procedimiento

En primer lugar, el INE elabora un directorio nacional de establecimientos hoteleros a partir de los registros enviados por las Consejerías de Turismo de las Comunidades Autónomas. Seguidamente, se utiliza un muestreo estratificado por provincias y categorías. Se recoge la información en cuestionarios (actualizados anualmente) con frecuencia mensual y se depura (de errores) de forma descentralizada. Se les aplica una serie de validaciones y se realiza una macrodepuración (de errores) a nivel central. Por último, los datos se unen a los de los ficheros XML y se obtienen los estimadores.

Así pues, se justifica que los datos del número de pernoctaciones de turistas a la ciudad de Valencia según país de procedencia que ofrece este instituto son los más fiables que podemos encontrar, y por ello se pueden tomar como referencia a la hora de intentar predecir dichas cifras.

➤ Período de publicación

El período de publicación de los resultados depende de la sección. En la Figura 1 podemos observar que los datos del sector turístico tienen publicación mensual, publicándose por ejemplo los datos de enero el 24 de febrero.

FIGURA 1. CALENDARIO DE DISPONIBILIDAD DE ESTADÍSTICAS COYUNTURALES

Calendario de disponibilidad de estadísticas coyunturales					
Estadísticas coyunturales	2014				
	Ene	Feb	Mar	Abr	May
CNTR <u>Contabilidad Nacional Trimestral</u>		27 (4ºT)			29 (1ºT)
PIB (Avance Trimestral) <u>Avance del PIB Trim.</u>	30 (4ºT)			30 (1ºT)	
CTH (Coyuntura Turística Hotelera) <u>Encuesta Ocupación Hotelera</u> <u>Índice de Precios Hoteleros</u> <u>Indicadores de Rentabilidad del Sector Hotelero</u>	24 (XII)	24 (I)	24 (II)	23 (III)	23 (IV)

FUENTE: INE

3.1.2. DATOS DEL GOOGLE TRENDS

En el Capítulo 1 ya se ha explicado en profundidad la herramienta online *Google Trends*. A continuación se exponen las características estadísticas más importantes⁹:

- La muestra analizada por Google Trends es aleatoria, por lo que no tiene en cuenta el total de búsquedas.
- El índice mediante el cual se estandarizan las series temporales de Google Trends se denomina Google Index y representa la popularidad de los términos de búsquedas en relación con el total de búsquedas realizadas en una determinada área geográfica y en un intervalo de tiempo especificado.

⁹ Extraído de : <https://support.google.com/trends/>.

- Las series temporales generadas no proporcionan valores absolutos del volumen de búsquedas, sino la frecuencia relativa de las mismas.
- Google Index está normalizado en una escala de 0 a 100, dividiendo la popularidad relativa en cada momento t por el máximo valor del periodo. Si el volumen de búsquedas se encuentra por debajo del umbral de tráfico mínimo, Google Trends asignará un valor de 0.
- El ámbito geográfico de búsqueda se puede delimitar en función del país o región.
- El periodo temporal, con origen en 2004, también se puede modificar (siendo la frecuencia siempre semanal).
- Las búsquedas se clasifican en 27 categorías de primer nivel y 241 subcategorías. Dicha clasificación se lleva a cabo de modo automático a través de un procesador de lenguaje natural.

3.2. Muestra

3.2.1. SERIES DEL INE

Las series temporales que ofrece el INE tienen periodicidad mensual y origen en 1999. Para el presente trabajo se tienen en cuenta las series temporales del número de pernoctaciones mensuales de turistas alemanes, franceses, ingleses e italianos a Valencia desde el 2004 hasta el 2013¹⁰.

Series utilizadas:

S1. PERNOCTACIONES MENSUALES DE TURISTAS ALEMANES A VALENCIA (2004-2013)



FUENTE: ELABORACIÓN PROPIA CON DATOS DEL INE.

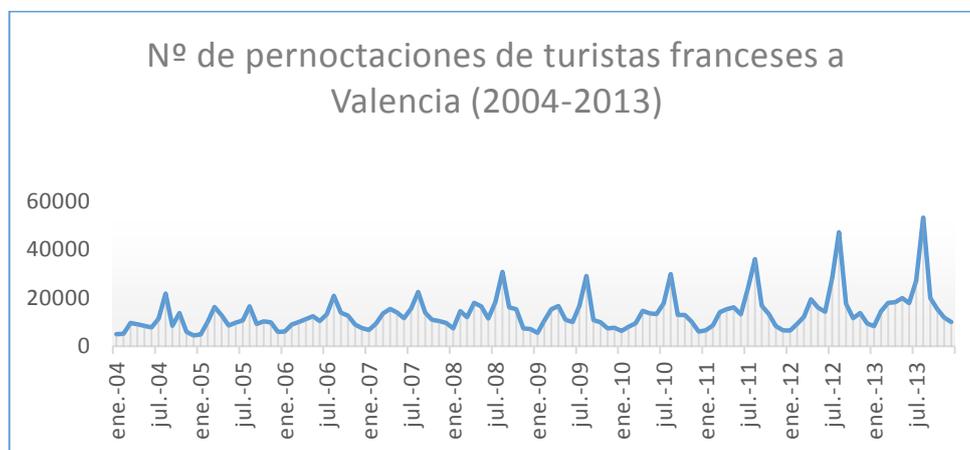
¹⁰ Se toman en cuenta estos años para ajustar con el período disponible en las series de Google Trends

Estadísticos más relevantes:

Media	11745,22
Desviación típica	5006,63
Máx.	30610
Min.	4944
n	120

La serie *Pernoctaciones mensuales de turistas alemanes a Valencia (2004-2013)* presenta estacionalidad. Concretamente, todos los años hay puntos de alta demanda turística en los meses de marzo y septiembre y puntos de baja demanda turística en el mes de enero.

S2. PERNOCTACIONES MENSUALES DE TURISTAS FRANCESES A VALENCIA (2004-2013)



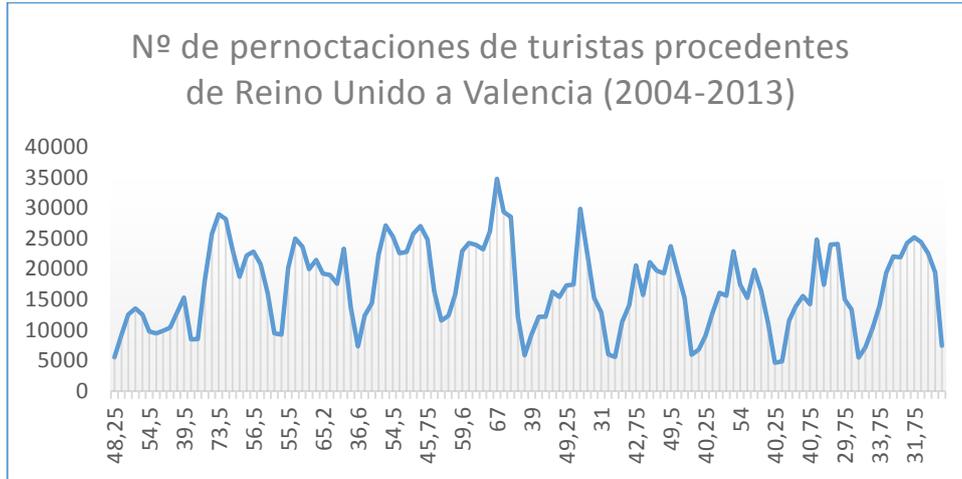
FUENTE: ELABORACIÓN PROPIA CON DATOS DEL INE.

Estadísticos más relevantes:

Media	12077,72
Desviación típica	7511,87
Máx.	53170
Min.	4556
n	120

La serie *Pernoctaciones mensuales de turistas franceses a Valencia (2004-2013)* presenta un componente estacional en la demanda, con puntos de alta demanda turística en el mes de agosto y puntos de baja demanda turística en enero.

S.3. PERNOCTACIONES MENSUALES DE TURISTAS INGLESES A VALENCIA (2004-2013)



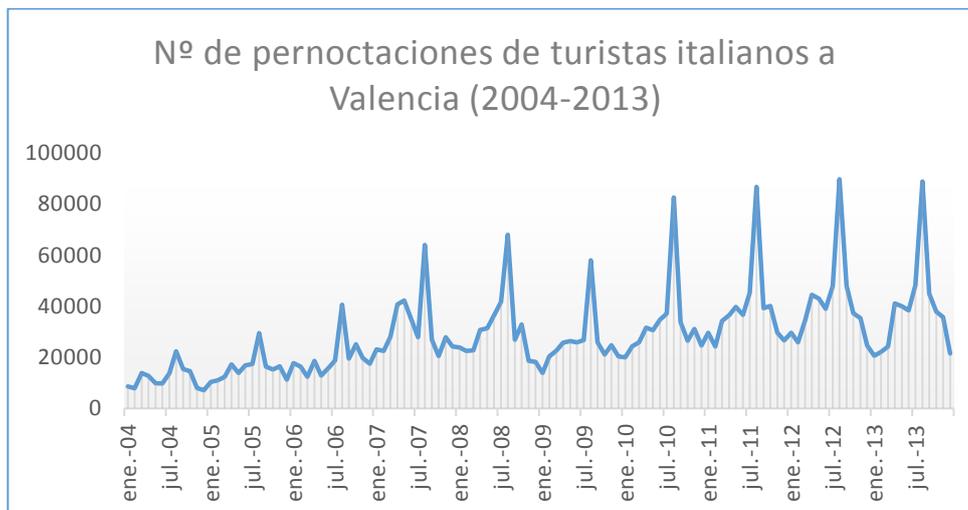
FUENTE: ELABORACIÓN PROPIA CON DATOS DEL INE

Estadísticos más relevantes:

Media	15613,30
Desviación típica	6704,90
Máx.	34766
Min.	4659
n	120

La serie *Pernoctaciones mensuales de turistas del reino unido a Valencia (2004-2013)* denota estacionalidad, con puntos de alta demanda turística en los meses de verano y puntos de baja demanda turística en el mes de enero.

S.4. PERNOCTACIONES MENSUALES DE TURISTAS ITALIANOS A VALENCIA (2004-2013)



FUENTE: ELABORACIÓN PROPIA CON DATOS DEL INE.

Estadísticos más relevantes:

Media	25367,95
Desviación típica	15941,73
Máx.	89749
Min.	7174
n	120

La serie *Pernoctaciones mensuales de turistas italianos a Valencia (2004-2013)* tiene estacionalidad, con puntos de alta demanda turística en agosto y puntos de baja demanda turística en enero.

3.2.2 SERIES DEL GOOGLE TRENDS

Las series temporales del volumen de búsquedas que ofrece Google Trends tienen periodicidad semanal y origen en 2004. Para que esta variable esté representada con la misma periodicidad que los datos del INE, se ha ajustado a periodicidad mensual realizando una media aritmética.

El término de búsqueda utilizado para el presente trabajo es *Valencia – City in Spain*, debido a que es el único que presenta volumen de búsqueda en todos los meses, desde 2004 hasta 2005.

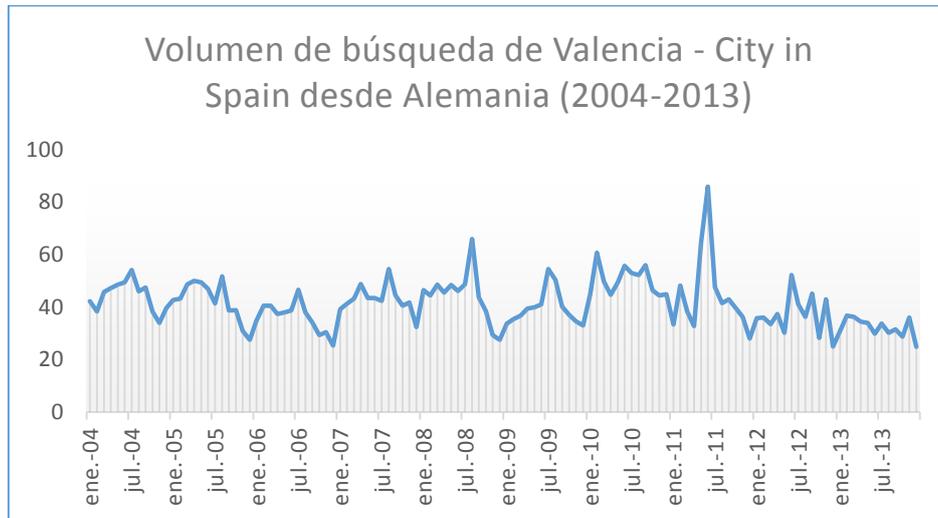
Otros términos de búsqueda como *-Hotel in Valencia-* no tienen volumen de búsqueda en todos los meses, hecho que no significa que no haya sido buscado todos los meses en el país en cuestión, sino que su volumen de búsqueda ese mes se encuentra por debajo del umbral de tráfico mínimo, y por tanto, como se veía en el punto 2.1.2., Google Trends le asigna valor 0.

Se espera que estas series ayuden a predecir el número de pernoctaciones de turistas procedentes de cada país, objeto de análisis en el presente trabajo, a la ciudad de Valencia porque cualquier turista que se quiera informar un poco acerca de la ciudad antes de venir -ya sea por ver dónde se alojará o que puestos turísticos hay dentro de la misma- lo primero que buscará será la ciudad por su nombre.

En las series que se muestran a continuación ningún mes tiene valor 100 debido a que los datos del Google Trends se han ajustado de semanales a mensuales.

Representación gráfica de variables:

V.1. VOLUMEN DE BÚSQUEDA DE VALENCIA - CITY IN SPAIN DESDE ALEMANIA (2004-2013)



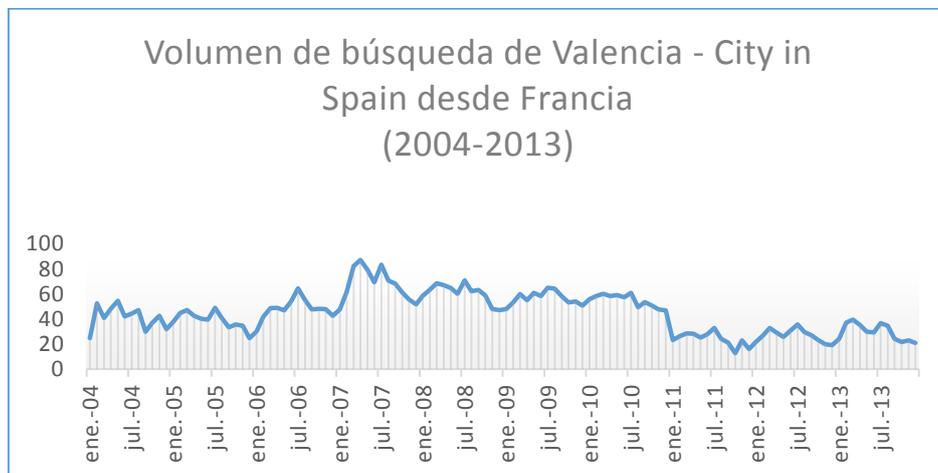
FUENTE: ELABORACIÓN PROPIA CON DATOS DEL GOOGLE TRENDS.

El Gráfico V.1. representa el volumen de búsqueda del término *Valencia -City in Spain*, desde el primer mes de 2004 hasta el último mes de 2013 en el ámbito geográfico Alemania.

El componente estacional en el volumen de búsquedas de dicho término guarda similitud con el componente estacional en la demanda de turistas de este país (Gráfico S.1.). Esto indica que, a priori, el número de pernoctaciones de alemanes a Valencia aumenta en el mes que aumenta el volumen de búsqueda del término (Marzo y Septiembre principalmente).

Más en adelante se podrá comprobar si los turistas buscan el destino turístico Valencia el mismo mes que van a realizar o con antelación.

V.2. VOLUMEN DE BÚSQUEDA DE VALENCIA - CITY IN SPAIN DESDE FRANCIA (2004-2013)



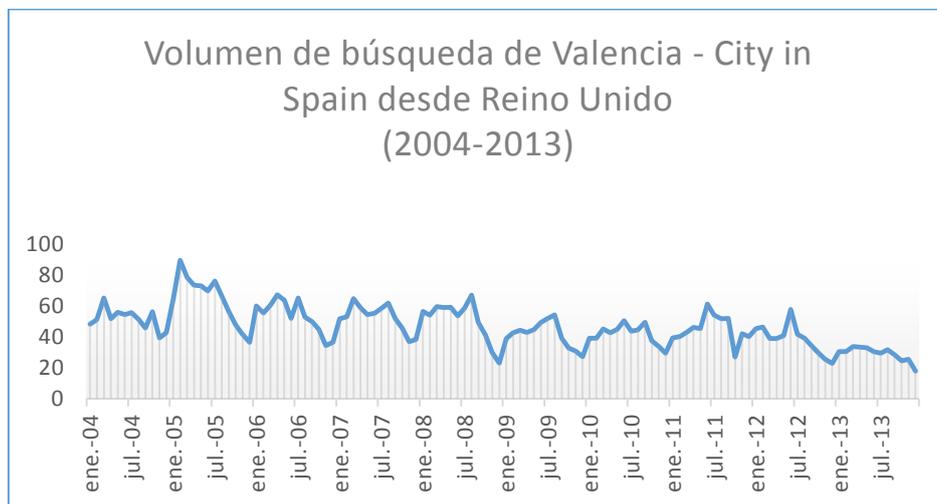
FUENTE: ELABORACIÓN PROPIA CON DATOS DEL GOOGLE TRENDS.

El Gráfico representa el término de búsqueda *Valencia – City in Spain*, desde el primer mes de 2004 hasta el último mes de 2013 en el ámbito geográfico Francia.

Viendo el gráfico, se observa estacionalidad, con puntos de alto volumen de búsqueda tanto el mes de alta demanda turística que se podían ver en el gráfico S.2 como 6 meses antes (en Marzo).

No obstante, hay falta de similitud en la tendencia de esta serie con respecto a la serie número de pernoctaciones mensuales de turistas franceses a Valencia S.2. En esta última se podía observar un incremento anual, en los últimos años, en el número de pernoctaciones, mientras que en el volumen de búsquedas de término se observa un decremento en los últimos años.

V.1. VOLUMEN DE BÚSQUEDA DE VALENCIA - CITY IN SPAIN DESDE REINO UNIDO (2004-2013)

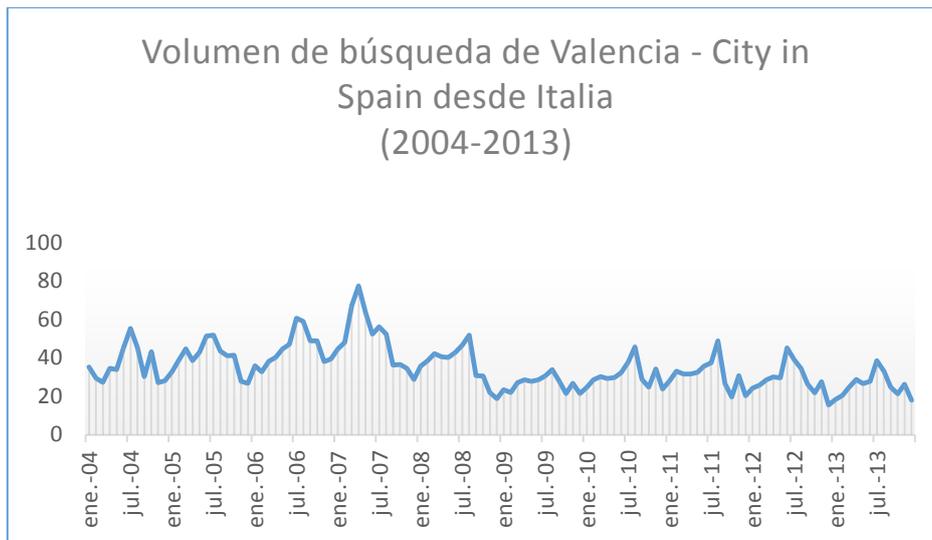


FUENTE: ELABORACIÓN PROPIA CON DATOS DEL GOOGLE TRENDS.

El Gráfico representa el término de búsqueda *Valencia – City in Spain*, desde el primer mes de 2004 hasta el último mes de 2013 en el ámbito geográfico Reino Unido.

Esta serie guarda similitud estacional con respecto a la serie S.3., en la que se representa el número de pernoctaciones mensuales de turistas ingleses a Valencia. Es más, los puntos de alto volumen de búsqueda se dan el mismo mes en que hay alta demanda turística de este país, o un mes antes.

V.1. VOLUMEN DE BÚSQUEDA DE VALENCIA - CITY IN SPAIN DESDE ITALIA (2004-2013)



FUENTE: ELABORACIÓN PROPIA CON DATOS DEL GOOGLE TRENDS.

El Gráfico representa el término de búsqueda Valencia – City in Spain, desde el primer mes de 2004 hasta el último mes de 2013 en el ámbito geográfico Italia.

Este gráfico una estacionalidad similar a la de la serie S.4., del número de pernoctaciones mensuales de turistas italianos a Valencia a partir del año 2008, con puntos de alto volumen de búsquedas el mismo mes en que se dan los puntos de alta demanda turística, o un mes antes.

3.3. Modelos propuestos

3.3.1. MODELO DE REFERENCIA

El primer modelo básico en que podríamos pensar a la hora de predecir el número de pernoctaciones de turistas a Valencia sería un modelo autorregresivo basado en los datos mensuales de pernoctaciones a Valencia según procedencia de los turistas que ofrece el INE.

En el Capítulo 2 se ha visto la estacionalidad que presentan las series del INE, con puntos de alta y baja demanda turística en los mismos meses de cada año. Por tanto, el modelo de referencia va a tener en cuenta este hecho.

Siendo el objetivo de este trabajo ver si los datos del Google Trends son útiles a la hora de predecir el número de pernoctaciones de turistas en una ciudad (Valencia por ejemplo), se puede utilizar el siguiente modelo como modelo de referencia.

Modelo1 ¹¹

$$\text{PERNOCT.INE}^i_t = \beta_0 + \beta_1 \text{PERNOCT.INE}^i_{t-12} + \varepsilon_t$$

donde,

PERNOCT.INE ^{IT} = NÚMERO DE PERNOCTACIONES, EN VALENCIA, DE TURISTAS
PROCEDENTES DEL PAÍS I EN EL MES T

ET = PERTURBACIÓN ALEATORIA

Se trata de un modelo autorregresivo de orden 12, es decir, un modelo que para predecir el número de turistas procedentes del país i, en el mes t, tiene en cuenta el número de turistas procedentes del país i en el mes t-12 (valor que había el mismo mes el año anterior y que se encuentra disponible en el INE).

3.3.2. MODELO CON DATOS DEL GOOGLE TRENDS

Partiendo del modelo de referencia, y teniendo en cuenta la estacionalidad de las series de Google Trends vista en el Capítulo 2, se puede plantear el siguiente modelo con datos del Google Trends. El poder predictivo de los datos del Google Trends se demostrará si mejora la predicción al incluirlos en el modelo de referencia.

Modelo2

$$\text{PERNOCT.INE}^i_t = \beta_0 + \beta_1 \text{PERNOCT.INE}^i_{t-12} + \text{BUSQUEDAS.GT}^i_t + \varepsilon_t$$

donde,

PERNOCT.INE ^{it} = número de pernoctaciones, en Valencia, de
turistas procedentes del país i, en el mes t

BUSQUEDAS.GT ^{it} = volumen de búsquedas del término Valencia-
City in Spain realizadas en el país i, en el mes t.

εt = perturbación aleatoria

¹¹ Todos los datos del modelo son datos mensuales debido a que los datos que se publican en el INE tienen frecuencia mensual.

3.3.3. MODELOS 1 Y 2 CON DATOS DE UN PAÍS CONCRETO

De los cuatro países que nombramos anteriormente como más predominantes dentro del turismo a Valencia (Italia, Francia, Alemania Y Reino Unido) se aplican estos modelos de manera detallada a los datos que ofrece el INE y el Google Trends para Italia¹². Posteriormente, se exponen los resultados y conclusiones de los demás países de procedencia predominantes de manera resumida.

Por tanto, como *modelo de referencia para Italia* se tiene:

$$\text{PERNOCT.INE}^{\text{ITAL}}_t = \beta_0 + \beta_1 \text{PERNOCT.ITAL.INE}^{\text{ITAL}}_{t-12} + \varepsilon_t$$

Y como *modelo con datos del Google Trends*:

$$\text{PERNOCT.INE}^{\text{ITAL}}_t = \beta_0 + \beta_1 \text{PERNOCT.INE}^{\text{ITAL}}_{t-12} + \text{BUSQUEDAS.GT}^{\text{ITAL}}_t + \varepsilon_t$$

donde,

PERNOCT.INE^{ITAL}_t = número de pernoctaciones, en Valencia, de turistas procedentes del Italia, en el mes t

BUSQUEDAS.GT^{ITAL}_t = volumen de búsquedas del término Valencia-City in Spain realizadas en Italia, en el mes t.

εt = perturbación aleatoria

¹² Dado que es el país de procedencia de los turistas predominantes a Valencia.

4.1. Análisis del modelo de referencia

FIGURA 1

Dependent Variable: PERNOCT_INE_ITAL				
Method: Least Squares				
Sample (adjusted): 2005M01 2013M12				
Included observations: 108 after adjustments				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	5785.253	1486.845	3.890960	0.0002
PERNOCT_INE_ITAL(-12)	0.898452	0.046818	19.19019	0.0000
R-squared	0.776496	Mean dependent var		30797.94
Adjusted R-squared	0.774387	S.D. dependent var		15652.83
S.E. of regression	7434.895	Akaike info criterion		20.68410
Sum squared resid	5.86E+09	Schwarz criterion		20.73377
Log likelihood	-1114.941	Hannan-Quinn criter.		20.70424
F-statistic	368.2635	Durbin-Watson stat		0.924169
Prob(F-statistic)	0.000000			

FUENTE: ELABORACIÓN PROPIA CON EVIEW

A priori, podemos observar que el modelo de referencia tiene un R^2 ajustado de aprox. 77%, es decir, con este modelo se explica en un 77% el comportamiento de la variable explicada (número de pernoctaciones en Valencia de turistas italianos en el mes t). El AIC es del 20,68 y el MAPE (Mean Absolut Percet Error) del 19.50%. Posteriormente se verá si estos datos mejoran al incluir como variable explicativa la serie del Google Trends.

FIGURA 2

Sample: 2005M01 2013M12						
Included observations: 108						
Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
		1	0.281	0.281	8.7578	0.003
		2	0.153	0.081	11.391	0.003
		3	0.172	0.120	14.747	0.002
		4	-0.001	-0.095	14.747	0.005
		5	-0.047	-0.058	15.004	0.010
		6	-0.064	-0.054	15.478	0.017
		7	-0.068	-0.015	16.020	0.025
		8	-0.053	-0.007	16.358	0.038
		9	0.007	0.048	16.364	0.060
		10	0.007	0.005	16.370	0.090
		11	0.048	0.043	16.647	0.119
		12	0.078	0.039	17.393	0.135
		13	0.055	0.011	17.767	0.167
		14	0.099	0.064	19.005	0.165
		15	0.014	-0.049	19.032	0.212
		16	-0.096	-0.115	20.222	0.210
		17	-0.017	0.028	20.260	0.261
		18	-0.038	0.001	20.453	0.308

FUENTE: ELABORACIÓN PROPIA CON EVIEW

En la Figura 2 se muestra el correlograma del modelo, en el cual podemos ver que el modelo presenta una leve correlación de orden 1. En la validación del modelo veremos si es significativa y, por tanto, si causa problemas en el mismo o no.

4.2. Análisis del modelo con Google Trends

Dada la estacionalidad de los datos del Google Trends, se puede plantear que los turistas italianos buscan el término Valencia-City in Spain un mes, o dos meses antes de viajar efectivamente a Valencia. Por tanto, hay que tener en cuenta mediante especificaciones este hecho, ya que implica que los datos de la variable explicativa del mes t pueden depender de los datos del mes $t-1$ o $t-2$.

Especificaciones que se han tenido en cuenta para los datos del Google Trends:

- *Especificación 1* : no se consideran términos autorregresivos.
- *Especificación 2* : se considera el término autorregresivo de primer orden.
- *Especificación 3* : se considera el término autorregresivo de segundo orden.
- *Especificación 4* : se considera tanto la popularidad del término Valencia-City in Spain en Italia en el mes actual, como en el mes anterior.

Dentro de estas especificaciones hay que elegir la especificación que mejor describe los datos y para ello podemos compararlas en base al coeficiente de determinación ajustado (R^2 ajustado)¹³, el Akaike Info Criterion (AIC)¹⁴ y el Mean Absolute Percent Error (MAPE)¹⁵, validando además que las variables explicativas sean significativas, es decir, que su P-valor sea inferior al 5% (descartando el término constante).

Una vez hecho esto, habrá que elegir un modelo y llevar a cabo una validación econométrica para comprobar su funcionalidad.

¹³ El R^2 ajustado es un coeficiente que indica el porcentaje del ajuste que se ha conseguido con el modelo lineal. A mayor porcentaje mejor es el modelo para predecir el comportamiento de la variable explicada.

¹⁴ El AIC mide de forma relativa la calidad de un modelo estadístico. Se basa en la complejidad del modelo y en la bondad de ajuste. Cuanto menor sea el nivel de AIC mejor será el modelo.

¹⁵ El MAPE es el error medio porcentual de un modelo. A menor porcentaje, más ajustada estará la predicción del modelo.

Variables	Coeficientes			
	Espec. 1	Espec. 2	Espec. 3	Espec. 4
Constante	-6179.84**	-7032.48***	-6330.71**	-8046.39***
<i>PERNOCT. INE ITAL t-12</i>	0,92***	0,93***	0,94***	0,93***
<i>BUSQUEDAS. GT ITAL t</i>	322.14***			144.77
<i>BUSQUEDAS. GT ITAL t-1</i>		341.67***		255.35**
<i>BUSQUEDAS. GT ITAL t-2</i>			311.49***	
<i>N</i>	108	108	108	108
<i>R² ajustado</i>	82.8%	83%	82.2%	83%
<i>AIC</i>	20,42	20,39	20,45	20,39
<i>MAPE</i>	18.92%	17.96%	18.34%	18.12%

Fuente: elaboración propia

La especificación que mejor describe los datos es la especificación 2, dado que a igual R^2 ajustado y AIC con la especificación 4, tiene el menor MAPE de todos. Esto quiere decir que el modelo de la especificación 2 es el que menos se equivoca, en promedio, al hacer la predicción de todos los analizados.

El R^2 ajustado es superior al de las especificaciones 1 y 3, queriendo decir esto que el modelo de la especificación 2 es mejor para predecir el comportamiento de la variable explicada. El AIC es inferior al de las especificaciones 1 y 3, por lo que el modelo de la especificación 2 es de mejor calidad y bondad de ajuste.

Como veíamos en la estacionalidad de las series del INE y del Google Trends para los datos de Italia, los puntos de alto volumen de búsqueda del término *Valencia-City in Spain* se solían dar el mismo mes o un mes antes que los puntos de alta demanda turística. Esto indica que los turistas Italianos suelen buscar dicho término aproximadamente el mismo mes o un mes antes de viajar, hecho favorable ya que implica que los datos del Google Trends pueden ser usados para predecir con mayor anterioridad a la publicación en el INE del número efectivo de turistas.

4.3. Validación del modelo

Se han considerado las siguientes hipótesis básicas para poder aplicar el método de estimación de Mínimos Cuadrados Ordinarios (MCO).

- I. El valor esperado de la perturbación es cero.

$$E(\varepsilon_t) = 0$$

- II. Homocedasticidad: La varianza de la perturbación es constante

$$V AR(\varepsilon_t) = E[\varepsilon_t - \varepsilon^*]^2 = \text{constante}$$

- III. Las perturbaciones no están relacionadas entre sí, es decir, no hay autocorrelación.

$$COV(\varepsilon_t, \varepsilon_j) = 0$$

- IV. Las variables explicativas son fijas, es decir, no les afecta la incertidumbre o la aleatoriedad.

$$COV(X_t, \varepsilon_t) = 0$$

- V. Ausencia de multicolinealidad perfecta, es decir, no hay una relación perfecta entre las variables explicativas

4.3.1. CONTRASTE DE ERRORES DE ESPECIFICACIÓN

Para averiguar si existen errores de especificación¹⁶, se puede aplicar el test RESET de Ramsey. Este test, estima un modelo transformado en el que se introduce como variable explicativa la variable estimada del modelo original, es decir, se introduce como variable explicativa $PERNOCT.INE^{ITAL}_t$ y por tanto se crea una β_3 .

Siendo el modelo original el elegido anteriormente (especificación 2):

$$PERNOCT.INE^{ITAL}_t = \beta_0 + \beta_1 PERNOCT.INE^{ITAL}_{t-12} + \beta_2 BUSQUEDAS.GT^{ITAL}_{t-1} + \epsilon_t$$

donde,

PERNOCT.INE^{ITAL}_t = número de pernoctaciones de turistas italianos a Valencia en el mes t.

PERNOCT.INE^{ITAL}_{t-12} = número de pernoctaciones de turistas italianos a Valencia registradas por el INE el mes t del año anterior

BUSQUEDAS.GT^{ITAL}_{t-1} = volumen de búsqueda del término Valencia-City in Spain en Italia en el mes anterior al mes t.

ϵ_t = perturbación aleatoria

FIGURA 3

Dependent Variable: PERNOCT_INE_ITAL				
Method: Least Squares				
Sample (adjusted): 2005M01 2013M12				
Included observations: 108 after adjustments				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-7032.483	2419.487	-2.906601	0.0045
PERNOCT_INE_ITAL(-12)	0.929010	0.040483	22.94838	0.0000
BUSQUEDAS_ITAL_GT(-1)	341.6706	54.79284	6.235680	0.0000
R-squared	0.836896	Mean dependent var		30797.94
Adjusted R-squared	0.833790	S.D. dependent var		15652.83
S.E. of regression	6381.488	Akaike info criterion		20.38757
Sum squared resid	4.28E+09	Schwarz criterion		20.46208
Log likelihood	-1097.929	Hannan-Quinn criter.		20.41778
F-statistic	269.3811	Durbin-Watson stat		1.287193
Prob(F-statistic)	0.000000			

FUENTE: ELABORACIÓN PROPIA CON EViews

¹⁶ Por ejemplo: utilizar una forma funcional incorrecta u omitir alguna variable explicativa relevante.

Viendo la estimación por MCO (Mínimos Cuadrados Ordinarios) del modelo en la FIGURA 3, para aplicar el test de RESET de Ramsey se define la variable FITTED como el valor estimado de $PERNOCT.INE^{ITAL}_t$:¹⁷

$$FITTED = -7032.48 + 0.93 \cdot PERNOCT.INE^{ITAL}_{t-12} + \\ + 341.67 \cdot BUSQUEDAS.GT^{ITAL}_{t-1} + \varepsilon_t$$

El modelo transformado queda como:

$$PERNOCT.INE^{ITAL}_t = \beta_0 + \beta_1 PERNOCT.INE^{ITAL}_{t-12} + \beta_2 BUSQUEDAS.GT^{ITAL}_{t-1} + \\ + \beta_3 FITTED^2 + \varepsilon_t$$

donde,

PERNOCT.INE ITAL_t = número de pernотaciones de italianos en Valencia en el mes t

PERNOCT.INE ITAL_{t-12} = número de pernотaciones de italianos en Valencia registradas por el INE doce meses antes del mes analizado t

BUSQUEDAS.GT ITAL_{t-1} = popularidad de la palabra Valencia-City in Spain siendo Italia la procedencia de la búsqueda en Google Trends en el mes anterior al mes t

ε_t = perturbación aleatoria

Por último, el test RESET consiste en efectuar un contraste de significatividad individual sobre β_3 . Para ello se utiliza la probabilidad exacta asociada al estadístico $t^*(\alpha^*(t^*\beta_3))$, ofrecida por EViews en la columna Prob. Las hipótesis de esta prueba son las siguientes:

$H_0 : \beta_i = 0 \rightarrow \beta_i$ no es significativo

$H_1 : \beta_i \neq 0 \rightarrow \beta_i$ sí es significativo

La regla de decisión de esta prueba es:

Si $\alpha^*(t^*\beta_3) > \alpha = 0,05 \rightarrow$ se acepta H_0

Si $\alpha^*(t^*\beta_3) < \alpha = 0,05 \rightarrow$ se rechaza H_0

En la Figura 4 podemos observar que el valor ofrecido por Eviews para β_3 es $0,27 > \alpha = 0,05$, hecho que lleva a aceptar H_0 , y ver que B_3 no es significativo y, por tanto, no existen errores de especificación.

¹⁷ Por razones de claridad, sólo se muestran dos cifras significativas de cada parámetro estimado.

FIGURA 4

Unrestricted Test Equation:				
Dependent Variable: PERNOCT_INE_ITAL				
Method: Least Squares				
Sample: 2005M01 2013M12				
Included observations: 108				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-3044.056	4336.672	-0.701934	0.4843
PERNOCT_INE_ITAL(-12)	0.774370	0.145348	5.327704	0.0000
BUSQUEDAS_ITAL_GT(-1)	289.3523	72.29619	4.002317	0.0001
FITTED^2	1.87E-06	1.68E-06	1.107664	0.2706
R-squared	0.838798	Mean dependent var		30797.94
Adjusted R-squared	0.834148	S.D. dependent var		15652.83
S.E. of regression	6374.603	Akaike info criterion		20.39437
Sum squared resid	4.23E+09	Schwarz criterion		20.49370
Log likelihood	-1097.296	Hannan-Quinn criter.		20.43464
F-statistic	180.3845	Durbin-Watson stat		1.265179
Prob(F-statistic)	0.000000			

FUENTE: ELABORACIÓN PROPIA CON EViews

4.3.2. ANÁLISIS DE LA NORMALIDAD DE LAS PERTURBACIONES

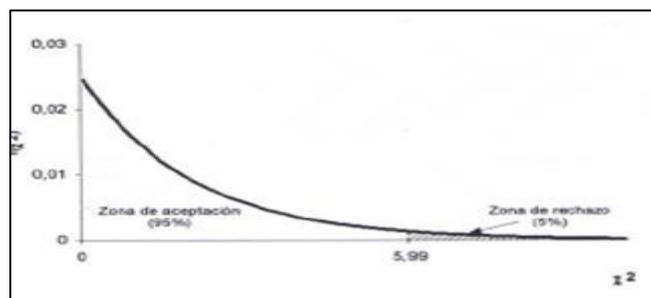
Para analizar la normalidad de las perturbaciones hay que llevar a cabo la prueba de Jarque-Bera. Las hipótesis de esta prueba son:

$H_0 : ut \sim N(0, \sigma^2) \rightarrow$ Las perturbaciones se distribuyen como una normal

$H_1 : ut \not\sim N(0, \sigma^2) \rightarrow$ No hay normalidad en las perturbaciones

Para llevar a cabo la prueba, se calcula el estadístico Jarque-Bera, que se distribuye como una χ^2 con 2 grados de libertad: $JB \sim \chi^2_2 = 5,99$, como muestra la siguiente figura:

DISTRIBUCIÓN DEL ESTADÍSTICO JARQUE-BERA



FUENTE: ELABORACIÓN PROPIA CON EViews

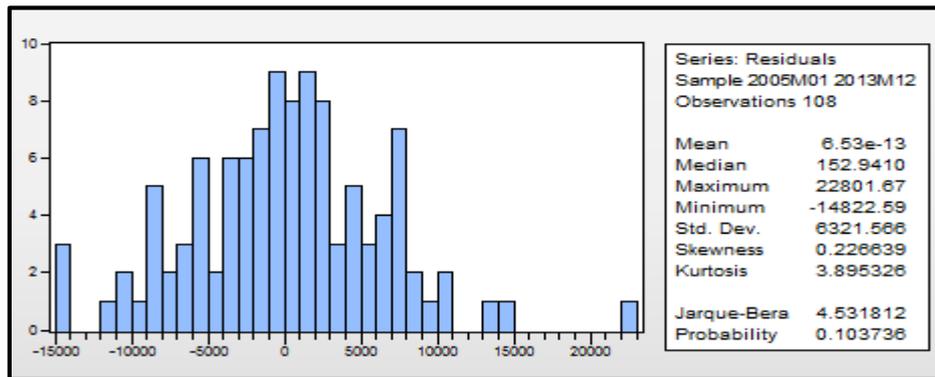
La regla de decisión a seguir es:

Si $JB^* < 5,99 \rightarrow$ se acepta H_0 , las perturbaciones se distribuyen como una normal

Si $JB^* > 5,99 \rightarrow$ se rechaza H_0 , las perturbaciones no son normales

EViews ofrece el cálculo del estadístico Jarque-Bera. Según muestra Figura 5, este estadístico presenta un valor de $4,5 < 5,9$, por lo que se acepta H_0 y se concluye que las perturbaciones se distribuyen conforme a la distribución normal. Este resultado también indica que el resto de contrastes que se van a efectuar son fiables, por lo que sí se puede realizar inferencia en el modelo.

FIGURA 5



FUENTE: ELABORACIÓN PROPIA CON EViews

4.3.3 CONTRASTES DE SIGNIFICATIVIDAD

En este epígrafe se realizan los contrastes de significatividad de los parámetros del modelo, primero de forma individual para cada parámetro y después, de forma conjunta.

4.3.3.1 Contrastes de significatividad individual

Se van a efectuar contrastes de significatividad sobre las pendientes parciales de las variables explicativas (β_1 y β_2).

El objetivo de estos contrastes es determinar si la variación de cada variable explicativa incide de forma significativa en la variable endógena.

Para ello, se empleará la prueba t y la probabilidad exacta asociada al estadístico t^* . Ambas pruebas comparten las mismas hipótesis, que son las que se enuncian a continuación:

$H_0 : \beta_i = 0 \rightarrow \beta_i$ no es significativo

$H_1 : \beta_i \neq 0 \rightarrow \beta_i$ sí es significativo

➤ *Contrastes de significatividad de β_1*

La prueba t para 1,66 y -1,66 del cálculo del estadístico $t^*\beta_1$, que se calcula de la siguiente forma:

$$t * \beta_1 = \frac{\beta_1}{ee(\beta_2)} = \frac{0.93}{0.04} = 22.95$$

Los valores de β_1 , su desviación típica así como el propio estadístico calculado, los proporciona EViews en la pantalla de resultados recogida en la Figura 3.

En este caso, el estadístico $t^*\beta_1$, se distribuye como una distribución t con 117 (=120-3 parámetros) grados de libertad. Por lo tanto, los valores críticos para cada una de las dos colas son 1,66 y -1,66.

De esta forma, la regla de decisión es la siguiente:

Si $t^*\beta_1 \in [1,66;-1,66] \rightarrow$ Se acepta H_0 , β_2 no es significativo

Si $t^*\beta_1 \notin [1,66;-1,66] \rightarrow$ Se rechaza H_0 , β_2 sí es significativo

Como $t^*\beta_1 = 22,95 \notin [1,66;-1,66]$, se rechaza H_0 y se concluye que β_1 sí que es significativo. Es decir, un cambio en el valor de la variable $PERNOCT.INE^{ITAL}_t -12$ afectará de forma estadísticamente significativa a la variable $PERNOCT.INE^{ITAL}_t$.

Otra manera de calcular la significatividad de β_1 es mediante la probabilidad exacta asociada al estadístico t^* , la cual se denota como $\alpha^*(t^*\beta_1)$ y viene ofrecida por EViews en la columna Prob. (Véase la Figura 3).

La regla de decisión es la siguiente:

Si $\alpha^*(t^*\beta_1) > \alpha = 0,05 \rightarrow$ se acepta H_0

Si $\alpha^*(t^*\beta_1) < \alpha = 0,05 \rightarrow$ se rechaza H_0

Como el valor proporcionado por EViews para $\alpha^*(t^*\beta_1)$ es $0,00 < \alpha = 0,05$, se rechaza H_0 , es decir, β_1 sí que es significativo.

➤ *Contrastes de significatividad de β_2*

La prueba t para 1,66 y -1,66 del cálculo del estadístico $t^*\beta_2$, que se calcula de la siguiente forma:

$$t * \beta_2 = \frac{\beta_2}{ee(\beta_2)} = \frac{341.67}{54.79} = 6.24$$

Los valores de β_2 , su desviación típica así como el propio estadístico calculado, los proporciona EViews en la pantalla de resultados recogida en la Figura 3.

En este caso, el estadístico $t^*\beta_2$, se distribuye como una distribución t con 117 (=120-3 parámetros) grados de libertad. Por lo tanto, los valores críticos para cada una de las dos colas son 1,66 y -1,66.

De esta forma, la regla de decisión es la siguiente:

Si $t^*\beta_2 \in [1,66;-1,66] \rightarrow$ Se acepta H_0 , β_2 no es significativo

Si $t^*\beta_2 \notin [1,66;-1,66] \rightarrow$ Se rechaza H_0 , β_2 sí es significativo

Como $t^*\beta_2 = 6.24 \notin [1,66;-1,66]$, se rechaza H_0 y se concluye que β_2 sí que es significativo.

De acuerdo con la probabilidad exacta asociada al estadístico t^* , la regla de decisión es la siguiente:

Si $\alpha^*(t^*\beta_2) > \alpha = 0,05 \rightarrow$ se acepta H_0

Si $\alpha^*(t^*\beta_2) < \alpha = 0,05 \rightarrow$ se rechaza H_0

Como el valor proporcionado por EViews para $\alpha^*(t^*\beta_2)$ es $0,00 < \alpha = 0,05$, se rechaza H_0 , es decir, β_2 sí que es significativo.

4.3.3.2. Contrastes de significatividad conjunta

Se van a efectuar contrastes de significatividad global sobre las pendientes parciales de las variables explicativas (β_1 y β_2).

El objetivo de estos contrastes es determinar si existe regresión de las variables explicativas sobre la variable endógena.

Para ello, se empleará la prueba F y la probabilidad exacta asociada al estadístico F^* . Ambas pruebas comparten las mismas hipótesis, que son las que se enuncian a continuación:

$H_0 : \beta_1 = \beta_2 = 0 \rightarrow$ no existe regresión

$H_1 : \text{Algún } \beta_1 \neq 0 \rightarrow$ sí existe regresión

La prueba F parte del cálculo del estadístico F^* , que se distribuye como una F con 2 grados de libertad en el numerador y 117 grados de libertad en el denominador (los datos numéricos están redondeados a la segunda unidad por claridad).

$$F^* = 1 + \frac{\frac{R^2}{k-1}}{\frac{(1-R^2)}{n-k}} + \frac{\frac{0.83}{3-1}}{\frac{(1-0.83)}{120-3}} = 269.38$$

El valor de R^2 , así como el propio estadístico calculado, es proporcionado por EViews en la pantalla de resultados recogida en la Figura 3. En este caso, el estadístico F^* se distribuye como una distribución F con 2(= 3 parámetros -1) grados de libertad en el

numerador y 117 (120 observaciones – 3 parámetros) grados de libertad en el denominador.

El valor crítico para $\alpha = 0,05$ según las tablas es de 3,09.

Si $F^* < F_{(2,117)} = 3,09 \rightarrow$ se acepta H_0 , no existe regresión

Si $F^* > F_{(2,117)} = 3,09 \rightarrow$ se rechaza H_0 , sí existe regresión.

Como $F^* = 269.38 > F_{(2,117)} = 3,09$, se rechaza H_0 y se concluye que algún $\beta_i \neq 0$ y, por tanto, sí que existe regresión.

Otra manera de verificar si existe regresión es mediante la probabilidad exacta al estadístico F^* , la cual viene denotada como $\alpha^*(F^*)$ y es ofrecida por EViews como Prob(F-Statistic). (Véase la Figura 3)

La regla de decisión de esta prueba es la siguiente:

Si $\alpha^*(F^*) > \alpha = 0,05 \rightarrow$ se acepta H_0

Si $\alpha^*(F^*) < \alpha = 0,05 \rightarrow$ se rechaza H_0

Como el valor estimado por EViews para la probabilidad asociada al estadístico F^* es $0,00 < \alpha = 0,05$, se rechaza H_0 , es decir, sí que existe regresión.

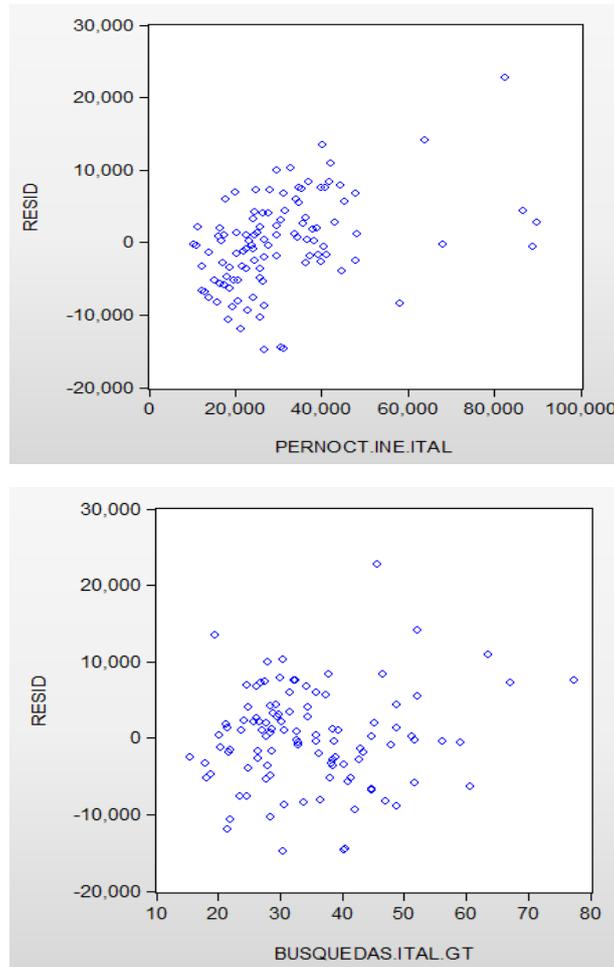
4.3.4. ANÁLISIS DE LA HETEROCEDASTICIDAD

En este apartado se comprobará la existencia de heterocedasticidad mediante dos métodos. En primer lugar se llevará a cabo el método gráfico y posteriormente se realizará el método analítico.

4.3.4.1. Método gráfico

El método gráfico de detección de heterocedasticidad consiste en analizar gráficamente en un diagrama de dispersión que compara el valor absoluto de los residuos con cada una de las variables explicativas. Estos diagramas de dispersión pueden verse en la Figura 6, de los que no se aprecia una clara relación funcional. En cualquier caso, será el método analítico el que confirme o rechace esta posible relación.

FIGURA 6



FUENTE: ELABORACIÓN PROPIA CON EViews

4.3.4.2. Método analítico

El método analítico de detección de heterocedasticidad consiste en aplicar la prueba de White. Esta prueba se basa en una regresión auxiliar en la que la variable endógena es el cuadrado de los residuos de la regresión original.

Como variables explicativas se incluyen las variables exógenas del modelo original además del producto de las variables explicativas del modelo original.

En este caso, la regresión auxiliar sería la siguiente:

$$\begin{aligned}
 e = & \alpha_0 + \alpha_1 \text{PERNOCT.INE}^{\text{ITAL}}_{t-12} + \alpha_2 \text{PERNOCT.INE}^{\text{ITAL}}_{t-12}{}^2 + \\
 & + \alpha_3 \text{PERNOCT.INE}^{\text{ITAL}}_{t-12} \cdot \text{BUSQUEDAS.GT}^{\text{ITAL}}_{t-1} + \alpha_4 \text{BUSQUEDAS.GT}^{\text{ITAL}}_{t-1} + \\
 & + \alpha_5 \text{BUSQUEDAS.GT}^{\text{ITAL}}_{t-1}{}^2
 \end{aligned}$$

Las hipótesis de esta prueba son:

$H_0: e^2_i \neq f(\chi_i) \rightarrow e^2_i = \alpha_i \rightarrow$ Homocedasticidad

$H_0: e^2_i = f(\chi_i) \rightarrow e^2_i \neq \alpha_i \rightarrow$ Homocedasticidad

FIGURA 7

Heteroskedasticity Test: White				
F-statistic	1.669634	Prob. F(2,105)	0.1933	
Obs*R-squared	3.328812	Prob. Chi-Square(2)	0.1893	
Scaled explained SS	4.554994	Prob. Chi-Square(2)	0.1025	
Test Equation:				
Dependent Variable: RESID^2				
Method: Least Squares				
Date: 09/07/14 Time: 03:46				
Sample: 2005M01 2013M12				
Included observations: 108				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	21108539	12624285	1.672058	0.0975
PERNOCT_INE_ITAL(-12)^2	0.006964	0.004860	1.432936	0.1548
BUSQUEDAS_ITAL_GT(-1)^2	8462.208	6905.251	1.225474	0.2231
R-squared	0.030822	Mean dependent var	39592181	
Adjusted R-squared	0.012362	S.D. dependent var	67682776	
S.E. of regression	67263135	Akaike info criterion	38.91351	
Sum squared resid	4.75E+17	Schwarz criterion	38.98801	
Log likelihood	-2098.329	Hannan-Quinn criter.	38.94372	
F-statistic	1.669634	Durbin-Watson stat	1.830692	
Prob(F-statistic)	0.193276			

FUENTE: ELABORACIÓN PROPIA CON EViews

Para llevar a cabo la prueba, se calcula el estadístico de White como $n \cdot R^2$ del modelo auxiliar.

Este estadístico se distribuye asintóticamente como una χ^2 con $k-1$ grados de libertad, donde k es el número de regresores de la regresión auxiliar ($4 = 5-1$ grados de libertad en este caso).

El valor crítico para $\alpha = 0,05$ es $\chi^2_4 = 9,49$. De esta forma, la regla de decisión es la siguiente.

Si $n \cdot R^2 < \chi^2_4 = 9,49 \rightarrow$ Se acepta H_0 , por tanto, existe homocedasticidad

Si $n \cdot R^2 > \chi^2_4 = 9,49 \rightarrow$ Se rechaza H_0 , por tanto, no existe heterocedasticidad

EViews ofrece el cálculo del estadístico de White. Según muestra la FIGURA 7, este estadístico presenta un valor de $3.33 < \chi^2_4 = 9.49$, por lo que se acepta H_0 y se concluye que existe homocedasticidad.

4.3.5. ANÁLISIS DE LA AUTOCORRELACIÓN

Para determinar la existencia de autocorrelación se empleará el contraste de Box-Ljung. Este contraste tiene como objetivo determinar la autocorrelación de cualquier orden. Dado que los datos que se dispone son mensuales, se incluirán 12 términos p en el contraste. Para llevar a cabo este contraste se utilizará la siguiente regresión auxiliar, donde p_p es el coeficiente de correlación de orden p :

$$e_t = p_1 \cdot e_{t-1} + p_2 \cdot e_{t-2} + p_3 \cdot e_{t-3} + p_4 \cdot e_{t-4} + p_5 \cdot e_{t-5} + p_6 \cdot e_{t-6} + p_7 \cdot e_{t-7} + p_8 \cdot e_{t-8} + p_9 \cdot e_{t-9} + p_{10} \cdot e_{t-10} + p_{11} \cdot e_{t-11} + p_{12} \cdot e_{t-12} + e_t$$

Las hipótesis de este contraste son las siguientes:

$H_0 : p_1 = p_2 = \dots = p_p = 0 \rightarrow$ no existe autocorrelación de orden p

$H_1 : \text{Algún } p_p \neq 0 \rightarrow$ existe autocorrelación de orden p

El contraste requiere del cálculo del estadístico Q para cada orden de autocorrelación que se quiera contrastar. Este estadístico se distribuye como una χ^2 con p grados de libertad. De esta forma, la regla de decisión es la siguiente para cada orden de autocorrelación.

Si $Q_p < \chi^2_p \rightarrow$ se acepta H_0 , por tanto, no existe autocorrelación de orden p

Si $Q_p > \chi^2_p \rightarrow$ se rechaza H_0 , por tanto, existe autocorrelación de orden p

En la Figura 8 se ve el cálculo del estadístico Q para cada orden de autocorrelación que hace EViews. Utilizando $\alpha = 0,05$ para calcular los valores críticos, se puede concluir que:

FIGURA 8

Autocorrelation		Partial Correlation		AC	PAC	Q-Stat	Prob	
				1	0.097	0.097	1.0404	0.308
				2	0.051	0.042	1.3295	0.514
				3	0.002	-0.007	1.3298	0.722
				4	-0.075	-0.077	1.9728	0.741
				5	0.022	0.037	2.0303	0.845
				6	0.024	0.026	2.0985	0.910
				7	0.043	0.036	2.3139	0.940
				8	-0.010	-0.027	2.3268	0.969
				9	0.138	0.145	4.6252	0.866
				10	0.003	-0.020	4.6263	0.915
				11	0.062	0.057	5.0923	0.927
				12	0.064	0.048	5.5928	0.935
				13	0.210	0.228	11.101	0.602
				14	0.299	0.269	22.391	0.071
				15	0.069	0.033	23.005	0.084
				16	-0.078	-0.119	23.794	0.094
				17	-0.068	-0.023	24.398	0.109
				18	-0.044	-0.016	24.651	0.135
				19	-0.044	-0.056	24.906	0.164

FUENTE: ELABORACIÓN PROPIA CON EViews

- Como $Q_1 = 1,04 < \chi^2_1 = 3,84 \rightarrow$ No existe autocorrelación de orden 1
- Como $Q_2 = 1,33 < \chi^2_2 = 5,99 \rightarrow$ No existe autocorrelación de orden 2
- Como $Q_3 = 1,33 < \chi^2_3 = 7,81 \rightarrow$ No existe autocorrelación de orden 3
- Como $Q_4 = 1,97 < \chi^2_4 = 9,49 \rightarrow$ No existe autocorrelación de orden 4
- Como $Q_5 = 2,03 < \chi^2_5 = 11,07 \rightarrow$ No existe autocorrelación de orden 5
- Como $Q_6 = 2,09 < \chi^2_6 = 12,59 \rightarrow$ No existe autocorrelación de orden 6
- Como $Q_7 = 2,31 < \chi^2_7 = 14,07 \rightarrow$ No existe autocorrelación de orden 7
- Como $Q_8 = 2,33 < \chi^2_8 = 15,51 \rightarrow$ No existe autocorrelación de orden 8
- Como $Q_9 = 4,62 < \chi^2_9 = 16,92 \rightarrow$ No existe autocorrelación de orden 9
- Como $Q_{10} = 4,63 < \chi^2_{10} = 18,31 \rightarrow$ No existe autocorrelación de orden 10
- Como $Q_{11} = 5,09 < \chi^2_{11} = 19,68 \rightarrow$ No existe autocorrelación de orden 11
- Como $Q_{12} = 5,59 < \chi^2_{12} = 21,03 \rightarrow$ No existe autocorrelación de orden 12

4.4. Análisis y Validación de los restantes modelos

4.4.1. ANÁLISIS Y VALIDACIÓN DE LOS MODELOS CON DATOS DE FRANCIA

- Modelo de referencia

FIGURA 9

Dependent Variable: PERNOCT_INE_FRAN Method: Least Squares				
Sample (adjusted): 2005M01 2013M12 Included observations: 108 after adjustments				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	499.9197	649.4486	0.769760	0.4432
PERNOCT_INE_FRAN(-12)	1.049706	0.044888	23.38526	0.0000
R-squared	0.837640	Mean dependent var		14006.57
Adjusted R-squared	0.836108	S.D. dependent var		7623.511
S.E. of regression	3086.267	Akaike info criterion		18.92566
Sum squared resid	1.01E+09	Schwarz criterion		18.97533
Log likelihood	-1019.985	Hannan-Quinn criter.		18.94580
F-statistic	546.8702	Durbin-Watson stat		1.427088
Prob(F-statistic)	0.000000			

FUENTE: ELABORACIÓN PROPIA CON EVIEW

En la Figura 9 podemos observar que el modelo de referencia tiene un R^2 ajustado del 83.61%, es decir, con este modelo se explica en un 83.61% el comportamiento de la variable explicada (número de pernoctaciones de turistas franceses en Valencia en el mes t). El AIC es del 18.92 y el MAPE (Mean Absolut Percet Error) del 18.9%.

FIGURA 10

Sample: 2005M01 2013M12 Included observations: 108						
Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
		1	0.038	0.038	0.1598	0.689
		2	-0.108	-0.110	1.4766	0.478
		3	-0.020	-0.012	1.5229	0.677
		4	0.013	0.003	1.5424	0.819
		5	0.186	0.184	5.5234	0.355
		6	0.136	0.129	7.6764	0.263
		7	0.118	0.160	9.3146	0.231
		8	-0.038	-0.009	9.4832	0.303
		9	-0.007	0.026	9.4898	0.393
		10	-0.060	-0.113	9.9310	0.447
		11	-0.056	-0.118	10.318	0.502
		12	0.081	-0.015	11.121	0.519
		13	0.088	0.045	12.092	0.520
		14	-0.092	-0.101	13.155	0.514
		15	-0.110	-0.055	14.690	0.474
		16	-0.103	-0.086	16.067	0.448
		17	-0.030	-0.027	16.187	0.511
		18	-0.090	-0.147	17.245	0.506

FUENTE: ELABORACIÓN PROPIA CON EVIEW

En la Figura 10 se puede observar que, a priori, este modelo no presenta problemas de autocorrelación.

- Modelo con datos del Google Trends

Especificaciones que se han tenido en cuenta para los datos del Google Trends:

- *Especificación 1* : no consideran términos autorregresivos.
- *Especificación 2* : se considera el término autorregresivo de primer orden.
- *Especificación 3* : se considera el término autorregresivo de sexto orden¹⁸.
- *Especificación 4* : se considera tanto la popularidad del término Valencia-City in Spain en Italia en el mes actual, como seis meses antes.

Tabla 2. Estimaciones del modelo con Google Trends

	Coeficientes			
	Espec. 1	Espec. 2	Espec. 3	Espec. 4
<i>Variables</i>				
<i>Constante</i>	1286.34	1308.73	2469.9**	2390.9**
<i>PERNOCT. INE FRAN t-12</i>	1.05***	1.05***	1.03***	1.03***
<i>BUSQUEDAS. GT FRAN t</i>	-18.19			9.39
<i>BUSQUEDAS. GT FRAN t-1</i>		-18.7		-44.59*
<i>BUSQUEDAS. GT FRAN t-6</i>			-38.2**	
<i>N</i>	108	108	108	108
<i>R² ajustado</i>	83.6%	83.6%	84.1%	83.9%
<i>AIC</i>	18.9	18.9	18.9	18.95
<i>MAPE</i>	19.2%	19.19%	18.8%	18.7%

Fuente: elaboración propia

Se elige la especificación 3, por ser la que mayor R² Ajustado, menor MAPE que la especificación 1 y 2 y la que tiene todas las variables con un P-Value < 0,05, es decir, todas las variables significativas.

¹⁸ En la serie gráfica de estos datos se podían observar picos de alto volumen de búsqueda tanto el mes de alto volumen de pernoctaciones de turistas como 6 meses antes, por lo que se plantea la posibilidad de que los franceses planeen en viaje de verano con 6 meses de antelación, concretamente en el mes de Marzo.

Por tanto, el modelo con datos del Google Trends elegido es:

$$\text{PERNOCT.INE}^{\text{FRAN}}_t = \beta_0 + \beta_1 \text{PERNOCT.INE}^{\text{FRAN}}_{t-12} + \text{BUSQUEDAS.GT}^{\text{FRAN}}_{t-6} +$$

ε_t donde,

$\text{PERNOCT.INE}^{\text{ITAL}}_t =$ número de pernотaciones, en Valencia, de turistas procedentes de Francia, en el mes t

$\text{BUSQUEDAS.GT}^{\text{ITAL}}_t =$ volumen de búsquedas del término Valencia-City in Spain realizadas en Francia, en el mes $t-6$.

$\varepsilon_t =$ perturbación aleatoria

4.4.1.2. Validación del modelo con datos del Google Trends

- *Contraste de errores de especificación*

FIGURA 11

Dependent Variable: PERNOCT_INE_FRAN				
Method: Least Squares				
Sample (adjusted): 2005M01 2013M12				
Included observations: 108 after adjustments				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	2469.923	1138.882	2.168726	0.0324
PERNOCT_INE_FRAN(-12)	1.033245	0.044887	23.01883	0.0000
BGVALF(-6)	-38.24056	18.29480	-2.090242	0.0390
R-squared	0.844126	Mean dependent var		14006.57
Adjusted R-squared	0.841157	S.D. dependent var		7623.511
S.E. of regression	3038.359	Akaike info criterion		18.90341
Sum squared resid	9.69E+08	Schwarz criterion		18.97791
Log likelihood	-1017.784	Hannan-Quinn criter.		18.93362
F-statistic	284.3105	Durbin-Watson stat		1.475719
Prob(F-statistic)	0.000000			

FUENTE: ELABORACIÓN PROPIA CON EViews

Viendo la Figura 11, para aplicar el test de RESET de Ramsey se tiene que que la variable FFITED se define como:

$$\text{FITTED} = 2469.92 + 1.03 \cdot \text{PERNOCT.INE}^{\text{FRAN}}_{t-12} - 38.24 \cdot \text{BUSQUEDAS.GT}^{\text{FRAN}}_{t-6} + \varepsilon_t$$

El modelo transformado queda:

$$\text{PERNOCT.INE}^{\text{FRAN}}_t = \beta_0 + \beta_1 \text{PERNOCT.INE}^{\text{FRAN}}_{t-12} + \beta_2 \text{BUSQUEDAS.GT}^{\text{FRAN}}_{t-6} + \beta_3 \text{FITTED}^2 + \varepsilon_t$$

FIGURA 12

Unrestricted Test Equation:				
Dependent Variable: PERNOCT_INE_FRAN				
Method: Least Squares				
Sample: 2005M01 2013M12				
Included observations: 108				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	5046.488	1367.974	3.689024	0.0004
PERNOCT_INE_FRAN(-12)	0.628785	0.135981	4.624069	0.0000
BGVALF(-6)	-29.08232	17.81149	-1.632784	0.1055
FITTED^2	9.01E-06	2.87E-06	3.136159	0.0022
R-squared	0.857594	Mean dependent var	14006.57	
Adjusted R-squared	0.853486	S.D. dependent var	7623.511	
S.E. of regression	2918.064	Akaike info criterion	18.83156	
Sum squared resid	8.86E+08	Schwarz criterion	18.93090	
Log likelihood	-1012.904	Hannan-Quinn criter.	18.87184	
F-statistic	208.7682	Durbin-Watson stat	1.336498	
Prob(F-statistic)	0.000000			

FUENTE: ELABORACIÓN PROPIA CON EViews

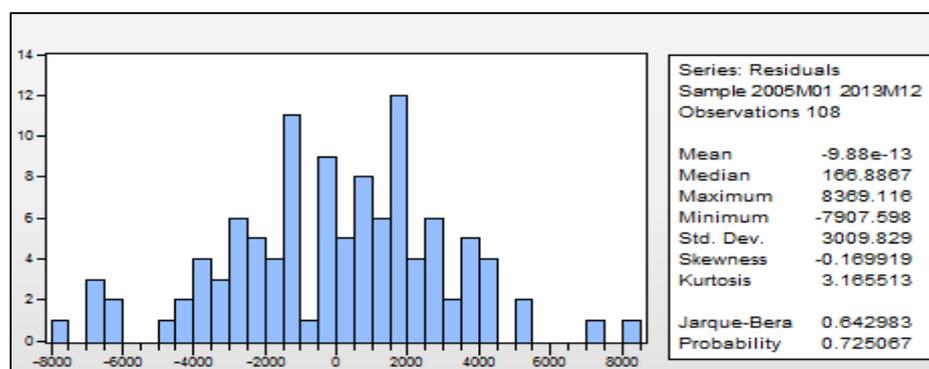
En la Figura 12 podemos observar que el valor ofrecido por Eviews para β_3 es $0,002 < \alpha = 0,05$, hecho que a nos lleva a rechazar la H_0 que veíamos en el apartado 4.3.1., y ver que B_3 es significativo y, por tanto, existen errores de especificación.

Esto no es preocupante, a priori, dado que es un modelo básico de predicción al que puede ser que le falte alguna explicativa relevante. Es decir, en el número de turistas franceses a la ciudad de Valencia influye no solamente el número de turistas que el INE registró el año anterior y las búsquedas del término *Valencia- City in Spain* que se hicieron en ese país y quedaron registradas en Google –siendo esto lo que hemos tenido en cuenta-, sino que puede ser que haya más variables explicativas que influyan en el mismo (factores económicos como la correlación entre el PIB, precios, renta disponible, gasto de los hogares o psicológicos como la preferencia de destino...).

No obstante, nuestro objetivo es validar el uso de Google Trends para predecir dicho nivel de turistas en la ciudad de Valencia y no plantear el mejor modelo de predicción de los mismos.

- *Análisis de la normalidad de las perturbaciones*

FIGURA 13



FUENTE: ELABORACIÓN PROPIA CON EViews

Vista la explicación del punto 4.3.2., en la Figura 13 se observa que el estadístico Jarque-Bera presenta un valor de $0.64 < 5.9$, por lo que se acepta H_0 y se concluye que las perturbaciones se distribuyen conforme a la distribución normal.

Este resultado también indica que el resto de contrastes que se van a efectuar son fiables, por lo que sí se puede realizar inferencia en el modelo.

- *Contraste de significatividad de β_1*

Vista la explicación en el apartado 4.3.3.1 y la Figura 11, se tiene que:

$$t * \beta_1 = \frac{\beta_1}{ee(\beta_2)} = \frac{1.033}{0.04} = 23.01$$

Como $t*\beta_1 = 23.01 \notin [1,66;-1,66]$, se rechaza H_0 y se concluye que β_1 sí que es significativo. Es decir, un cambio en el valor de la variable $PERNOCT.INE_{t-12}^{FRAN}$ afectará de forma estadísticamente significativa a la variable $PERNOCT.INE_{t..}^{FRAN}$.

Además, como el valor proporcionado por EViews para $\alpha^*(t*\beta_1)$ es $0,00 < \alpha = 0,05$, se rechaza H_0 , es decir, β_1 sí que es significativo.

Contraste de significatividad de β_2

$$t * \beta_1 = \frac{\beta_1}{ee(\beta_2)} = \frac{-38.24}{18.29} = -2.09$$

Como $t*\beta_2 = -2.09 \notin [1,66;-1,66]$, se rechaza H_0 y se concluye que β_2 sí que es significativo. Es decir, un cambio en el valor de la variable $BUSQUEDAS.GT_{t-6}^{FRAN}$ afectará de forma estadísticamente significativa a la variable $PERNOCT.INE_{t..}^{FRAN}$.

Además, como el valor proporcionado por EViews para $\alpha^*(t^*\beta_1)$ es $0,00 < \alpha = 0,05$, se rechaza H_0 , es decir, β_1 sí que es significativo.

- *Contraste de significatividad conjunta*

Teniendo la explicación en el apartado 4.3.3.2 se tiene que que :

$$F^* = 1 + \frac{\frac{R^2}{k-1}}{\frac{(1-R^2)}{n-k}} + \frac{\frac{0.84}{3-1}}{\frac{(1-0.84)}{120-3}} = 284.3$$

El valor de R^2 , así como el propio estadístico calculado, es proporcionado por EViews en la pantalla de resultados recogida en la Figura 11.

Como $F^* = 284.3 > F_{(2,117)} = 3.09$, se rechaza H_0 y se concluye que algún $\beta_i \neq 0$ y, por tanto, sí que existe regresión.

Además, como el valor estimado por EViews en la Figura 11 para la probabilidad asociada al estadístico F^* es $0,00 < \alpha = 0,05$, se rechaza H_0 , es decir, sí que existe regresión.

- *Análisis de la Heterocedasticidad*

Viendo la explicación del método analítico del punto 4.3.4 se tiene la siguiente regresión auxiliar:

$$e = \alpha_0 + \alpha_1 \text{PERNOCT.INE}^{\text{FRAN}}_{t-12} + \alpha_2 \text{PERNOCT.INE}^{\text{FRAN}}_{t-12}{}^2 + \\ + \alpha_3 \text{PERNOCT.INE}^{\text{FRAN}}_{t-12} \cdot \text{BUSQUEDAS.GT}^{\text{ITAL}}_{t-6} + \alpha_4 \text{BUSQUEDAS.GT}^{\text{FRAN}}_{t-6} \\ + \alpha_5 \text{BUSQUEDAS.GT}^{\text{FRAN}}_{t-6}{}^2$$

FIGURA 14

Heteroskedasticity Test: White				
F-statistic	0.615882	Prob. F(2, 105)	0.5421	
Obs*R-squared	1.252266	Prob. Chi-Square(2)	0.5347	
Scaled explained SS	0.858792	Prob. Chi-Square(2)	0.6509	
Test Equation:				
Dependent Variable: RESID^2				
Method: Least Squares				
Date: 09/07/14 Time: 12:40				
Sample: 2005M01 2013M12				
Included observations: 108				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.009992	0.070450	-0.141839	0.8875
(@LOG(PERNOCT_INE_FRAN(-12)))^2	0.000311	0.000678	0.459077	0.6471
(@LOG(BGVALF(-6)))^2	0.002119	0.001968	1.076738	0.2841
R-squared	0.011595	Mean dependent var	0.047558	
Adjusted R-squared	-0.007232	S.D. dependent var	0.057555	
S.E. of regression	0.057763	Akaike info criterion	-2.837552	
Sum squared resid	0.350339	Schwarz criterion	-2.763048	
Log likelihood	156.2278	Hannan-Quinn criter.	-2.807343	
F-statistic	0.615882	Durbin-Watson stat	1.708679	
Prob(F-statistic)	0.542104			

FUENTE: ELABORACIÓN PROPIA CON EViews

Según muestra la FIGURA 14, este estadístico de White presenta un valor de $1.25 < \chi^2_4 = 9,49$, por lo que se acepta H_0 y se concluye que existe homocedasticidad.

▪ *Análisis de la autocorrelación*

En el punto 4.3.5 ya se ha visto la explicación detallada de este análisis. Utilizando $\alpha = 0,05$ para calcular los valores críticos, se puede concluir que:

FIGURA 15

Sample: 2005M01 2013M12 Included observations: 108						
Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
		1	-0.024	-0.024	0.0625	0.803
		2	-0.083	-0.084	0.8341	0.659
		3	-0.065	-0.070	1.3132	0.726
		4	-0.048	-0.060	1.5817	0.812
		5	0.159	0.147	4.5092	0.479
		6	0.125	0.125	6.3295	0.387
		7	0.122	0.158	8.0886	0.325
		8	-0.012	0.042	8.1047	0.423
		9	0.004	0.064	8.1070	0.523
		10	-0.087	-0.086	9.0333	0.529
		11	-0.071	-0.115	9.6589	0.561
		12	0.112	0.025	11.207	0.511
		13	0.097	0.046	12.375	0.497
		14	-0.103	-0.139	13.714	0.471
		15	-0.087	-0.077	14.674	0.475
		16	-0.128	-0.116	16.798	0.399
		17	-0.003	-0.022	16.799	0.468
		18	-0.102	-0.182	18.163	0.445
		19	0.040	0.024	18.177	0.511

FUENTE: ELABORACIÓN PROPIA CON EViews

- Como $Q_1 = 0.06 < \chi^2_1 = 3,84 \rightarrow$ No existe autocorrelación de orden 1
- Como $Q_2 = 0.83 < \chi^2_2 = 5,99 \rightarrow$ No existe autocorrelación de orden 2
- Como $Q_3 = 1.31 < \chi^2_3 = 7,81 \rightarrow$ No existe autocorrelación de orden 3
- Como $Q_4 = 1.58 < \chi^2_4 = 9,49 \rightarrow$ No existe autocorrelación de orden 4
- Como $Q_5 = 4.51 < \chi^2_5 = 11,07 \rightarrow$ No existe autocorrelación de orden 5
- Como $Q_6 = 6.33 < \chi^2_6 = 12,59 \rightarrow$ No existe autocorrelación de orden 6
- Como $Q_7 = 8.08 < \chi^2_7 = 14,07 \rightarrow$ No existe autocorrelación de orden 7
- Como $Q_8 = 8.10 < \chi^2_8 = 15,51 \rightarrow$ No existe autocorrelación de orden 8
- Como $Q_9 = 4,62 < \chi^2_9 = 16,92 \rightarrow$ No existe autocorrelación de orden 9
- Como $Q_{10} = 4,63 < \chi^2_{10} = 18,31 \rightarrow$ No existe autocorrelación de orden 10
- Como $Q_{11} = 5,09 < \chi^2_{11} = 19,68 \rightarrow$ No existe autocorrelación de orden 11
- Como $Q_{12} = 5,59 < \chi^2_{12} = 21.03 \rightarrow$ No existe autocorrelación de orden 12

4.4.2. VALIDACIÓN DEL MODELO DE GOOGLE TRENDS CON DATOS DEL REINO UNIDO

➤ Modelo de referencia

FIGURA 16

Dependent Variable: PERNOCT_INE_RU Method: Least Squares				
Sample (adjusted): 2005M01 2013M12 Included observations: 108 after adjustments				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	6578.742	1307.411	5.031884	0.0000
PERNOCT_INE_RU(-12)	0.660986	0.071580	9.234282	0.0000
R-squared	0.445815	Mean dependent var		17811.50
Adjusted R-squared	0.440587	S.D. dependent var		6658.520
S.E. of regression	4980.168	Akaike info criterion		19.88266
Sum squared resid	2.63E+09	Schwarz criterion		19.93233
Log likelihood	-1071.664	Hannan-Quinn criter.		19.90280
F-statistic	85.27196	Durbin-Watson stat		0.679897
Prob(F-statistic)	0.000000			

FUENTE: ELABORACIÓN PROPIA CON EVIEW

En la Figura 16 podemos observar que el modelo de referencia tiene un R² ajustado del 44.5%, es decir, con este modelo se explica en un 44.5% el comportamiento de la variable explicada (número de pernoctaciones en Valencia, de turistas del Reino Unido en el mes t). El AIC es del 19.8 y el MAPE (Mean Absolut Percet Error) del 27%.

FIGURA 17

Sample: 2005M01 2013M12 Included observations: 108						
Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
		1	0.527	0.527	30.873	0.000
		2	0.306	0.038	41.341	0.000
		3	0.240	0.090	47.853	0.000
		4	0.210	0.062	52.890	0.000
		5	0.122	-0.043	54.621	0.000
		6	0.008	-0.097	54.627	0.000
		7	0.047	0.083	54.884	0.000
		8	-0.082	-0.188	55.681	0.000
		9	-0.062	0.066	56.145	0.000
		10	-0.053	-0.011	56.485	0.000
		11	-0.028	0.023	56.584	0.000
		12	-0.084	-0.083	57.457	0.000
		13	-0.090	0.016	58.461	0.000
		14	-0.039	-0.010	58.653	0.000
		15	-0.070	-0.027	59.283	0.000
		16	-0.054	-0.006	59.658	0.000
		17	-0.045	0.011	59.922	0.000

FUENTE: ELABORACIÓN PROPIA CON EVIEW

En la Figura 17 se puede observar que este modelo presenta problemas de autocorrelación, por lo que se procede a eliminarla introduciendo como variable explicativa $PERNOCT.INE^{RU}_{t-1}$, siendo el nuevo modelo de referencia para este país

$$PERNOCT.INE^{RU}_t = \beta_0 + \beta_1 PERNOCT.INE^{RU}_{t-1} + \beta_2 PERNOCT.INE^{RU}_{t-12} + \varepsilon_t$$

donde,

PERNOCT.INE^{RU}_t = número de pernoctaciones de turistas procedentes del Reino Unido en Valencia en el mes t.

PERNOCT.INE^{RU}_{t-1} = número de pernoctaciones de turistas procedentes del Reino Unido en Valencia en el mes anterior

PERNOCT.INE^{RU}_{t-12} = número de pernoctaciones de turistas procedentes del Reino Unido en Valencia doce meses antes

ε_t = perturbación aleatoria

FIGURA 18

Dependent Variable: PERNOCT_INE_RU				
Method: Least Squares				
Sample (adjusted): 2005M01 2013M12				
Included observations: 108 after adjustments				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1914.775	1261.690	1.517628	0.1321
PERNOCT_INE_RU(-1)	0.490090	0.068791	7.124300	0.0000
PERNOCT_INE_RU(-12)	0.421483	0.067949	6.202922	0.0000
R-squared	0.626406	Mean dependent var		17811.50
Adjusted R-squared	0.619290	S.D. dependent var		6658.520
S.E. of regression	4108.421	Akaike info criterion		19.50685
Sum squared resid	1.77E+09	Schwarz criterion		19.58135
Log likelihood	-1050.370	Hannan-Quinn criter.		19.53706
F-statistic	88.02683	Durbin-Watson stat		1.401431
Prob(F-statistic)	0.000000			

FUENTE: ELABORACIÓN PROPIA CON EVIEW

En la Figura 18 podemos observar que el nuevo modelo de referencia tiene un R² ajustado del 61.9%, es decir, con este modelo se explica en un 61.9% el comportamiento de la variable explicada (número de pernoctaciones en Valencia de turistas del Reino Unido en el mes t). El AIC es del 19.5 y el MAPE (Mean Absolut Percet Error) del 22.12%.

FIGURA 19

Sample: 2005M01 2013M12 Included observations: 108						
Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.231	0.231	5.9351	0.015
		2	0.015	-0.041	5.9596	0.051
		3	0.061	0.071	6.3781	0.095
		4	0.051	0.022	6.6719	0.154
		5	-0.042	-0.060	6.8718	0.230
		6	0.016	0.041	6.9001	0.330
		7	-0.057	-0.083	7.2882	0.399
		8	-0.051	-0.014	7.5984	0.474
		9	0.011	0.028	7.6131	0.574
		10	0.045	0.037	7.8590	0.643
		11	0.136	0.141	10.129	0.519
		12	-0.014	-0.093	10.154	0.602
		13	-0.086	-0.067	11.081	0.604
		14	0.111	0.143	12.648	0.554
		15	-0.090	-0.191	13.689	0.549
		16	-0.100	0.007	14.974	0.527
		17	-0.142	-0.155	17.615	0.414

FUENTE: ELABORACIÓN PROPIA CON EVIEW

En la Figura 19 se puede ver que la autocorrelación de orden 1 se ha corregido.

- Modelo con datos del Google Trends

Especificaciones que se han tenido en cuenta para los datos del Google Trends:

- *Especificación 1* : no consideran términos autorregresivos.
- *Especificación 2* : se considera el término autorregresivo de primer orden.
- *Especificación 3* : se considera el término autorregresivo de segundo orden.¹⁹
- *Especificación 4* : se considera tanto la popularidad del término *Valencia-City in Spain* en Italia en el mes actual, como el mes anterior.

¹⁹ En las series del volumen de búsquedas del término analizado en Reino Unido no habían pico claramente definidos, significando esto que los puntos de alto volumen se daban en varios meses seguidos, y por tanto, es posible que la búsqueda se realice los meses anteriores al viaje.

<i>Tabla 3. Estimaciones del modelo con Google Trends</i>				
<i>Variables</i>	<i>Coefficientes</i>			
	<i>Espec. 1</i>	<i>Espec. 2</i>	<i>Espec. 3</i>	<i>Espec. 4</i>
<i>Constante</i>	-4305.5***	-2464*	-1129.3	-4302.5**
<i>PERNOCT. INE RU t-1</i>	0.48***	0.36***	0.34***	0.47***
<i>PERNOCT. INE RU t-12</i>	0.35***	0.38***	0.42***	0.35***
<i>BUSQUEDAS. GT FRAN t</i>	163.23***			158.4***
<i>BUSQUEDAS. GT FRAN t-1</i>		157.11***		6.38
<i>BUSQUEDAS. GT FRAN t-2</i>			121.14***	
<i>N</i>	108	108	108	108
<i>R² ajustado</i>	72.6%	69.5%	65.4%	72.3%
<i>AIC</i>	19.18	19.2	19.4	19.2
<i>MAPE</i>	19.01%	19.8%	21.3%	19.09%

Fuente: elaboración propia

Se elige la especificación 1, por ser la que mayor R² Ajustado, menor AIC y MAPE, teniendo además todas las variables con un P-Value muy cercano al 0,000 < 0,05, es decir, todas las variables altamente significativas.

Por tanto, el modelo con datos del Google Trends elegido es:

$$\text{PERNOCT.INE}^{\text{RU}}_t = \beta_0 + \beta_1 \text{PERNOCT.INE}^{\text{RU}}_{t-1} + \beta_2 \text{PERNOCT.INE}^{\text{RU}}_{t-12} + \beta_3 \text{BUSQUEDAS.GT}^{\text{FRAN}}_t + \varepsilon_t$$

PERNOCT.INE^{RU}_t = número de pernoctaciones de turistas procedentes del Reino Unido en Valencia en el mes t.

PERNOCT.INE^{RU}_{t-1} = número de pernoctaciones de turistas procedentes del Reino Unido en Valencia en el mes anterior

PERNOCT.INE^{RU}_{t-12} = número de pernoctaciones de turistas procedentes del Reino Unido en Valencia doce meses antes

BUSQUEDAS.GT^{RU}_t = volumen de búsquedas del término Valencia-City in Spain realizadas en Reino Unido, en el mes t.

ε_t = perturbación aleatoria

4.4.2.1. Validación del modelo con datos del Google Trends

- *Contraste de errores de especificación*

FIGURA 20

Dependent Variable: PERNOCT_INE_RU				
Method: Least Squares				
Sample (adjusted): 2005M01 2013M12				
Included observations: 108 after adjustments				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-4305.504	1437.127	-2.995911	0.0034
PERNOCT_INE_RU(-1)	0.480246	0.058353	8.230036	0.0000
PERNOCT_INE_RU(-12)	0.351923	0.058609	6.004544	0.0000
BUSQUEDAS_RU_GT	163.2396	25.18104	6.482640	0.0000
R-squared	0.733923	Mean dependent var		17811.50
Adjusted R-squared	0.726248	S.D. dependent var		6658.520
S.E. of regression	3483.827	Akaike info criterion		19.18598
Sum squared resid	1.26E+09	Schwarz criterion		19.28532
Log likelihood	-1032.043	Hannan-Quinn criter.		19.22626
F-statistic	95.62144	Durbin-Watson stat		1.592736
Prob(F-statistic)	0.000000			

FUENTE: ELABORACIÓN PROPIA CON EViews

Viendo la Figura 20, para aplicar el test de RESET de Ramsey la variable FFITED se define como:

$$\text{FITTED} = -4305.5 + 0.48 \cdot \text{PERNOCT.INE}^{\text{RU}}_{t-1} + 0.35 \cdot \text{PERNOCT.INE}^{\text{RU}}_{t-12} + 163.2 \cdot \text{BUSQUEDAS.GT}^{\text{RU}}_t + \varepsilon_t$$

El modelo transformado queda:

$$\text{PERNOCT.INE}^{\text{RU}}_t = \beta_0 + \beta_1 \text{PERNOCT.INE}^{\text{RU}}_{t-1} + \beta_2 \text{PERNOCT.INE}^{\text{RU}}_{t-12} + \beta_3 \text{BUSQUEDAS.GT}^{\text{RU}}_t + \beta_4 \text{FITTED}^2 + \varepsilon_t$$

FIGURA 21

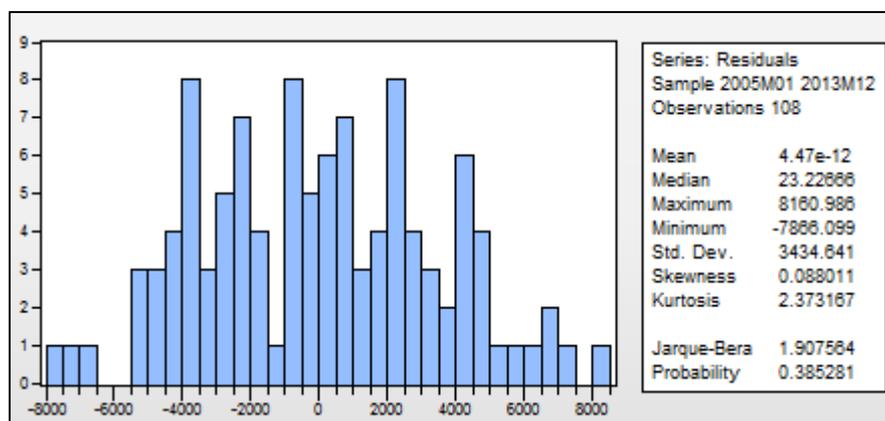
Unrestricted Test Equation:				
Dependent Variable: PERNOCT_INE_RU				
Method: Least Squares				
Sample: 2005M01 2013M12				
Included observations: 108				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-8991.127	4267.625	-2.106822	0.0376
PERNOCT_INE_RU(-1)	0.666163	0.169783	3.923621	0.0002
PERNOCT_INE_RU(-12)	0.484131	0.127609	3.793855	0.0003
BUSQUEDAS_RU_GT	228.1031	61.05394	3.736091	0.0003
FITTED^2	-1.11E-05	9.54E-06	-1.165795	0.2464
R-squared	0.737388	Mean dependent var	17811.50	
Adjusted R-squared	0.727190	S.D. dependent var	6658.520	
S.E. of regression	3477.828	Akaike info criterion	19.19139	
Sum squared resid	1.25E+09	Schwarz criterion	19.31557	
Log likelihood	-1031.335	Hannan-Quinn criter.	19.24174	
F-statistic	72.30346	Durbin-Watson stat	1.598613	
Prob(F-statistic)	0.000000			

FUENTE: ELABORACIÓN PROPIA CON EViews

En la Figura 21 podemos observar que el valor ofrecido por Eviews para β_4 es $0,24 > \alpha = 0,05$, hecho que a nos lleva a aceptar la H_0 que veíamos en el apartado 3.2.1., y ver que B_4 no es significativo y, por tanto, no existen errores de especificación.

- *Análisis de la normalidad de las perturbaciones*

FIGURA 22



FUENTE: ELABORACIÓN PROPIA CON EViews

Vista la explicación del punto 4.3.2., en la Figura 22 se observa que el estadístico Jarque-Bera presenta un valor de $1.9 < 5.9$, por lo que se acepta H_0 y se concluye que las perturbaciones se distribuyen conforme a la distribución normal.

Este resultado también indica que el resto de contrastes que se van a efectuar son fiables, por lo que sí se puede realizar inferencia en el modelo.

- *Contraste de significatividad de β_1*

Vista la explicación en el apartado 4.3.3.1 y la Figura 20, se tiene que :

$$t * \beta_1 = \frac{\beta_1}{ee(\beta_2)} = \frac{0.48}{0.05} = 8.23$$

Como $t*\beta_1 = 8.23 \notin [1,66;-1,66]$, se rechaza H_0 y se concluye que β_1 sí que es significativo. Es decir, un cambio en el valor de la variable $PERNOCT.INE^{RU}_{t-1}$ afectará de forma estadísticamente significativa a la variable $PERNOCT.INE^{RU}_t$.

Además, como el valor proporcionado por EViews para $\alpha^*(t*\beta_1)$ es $0,00 < \alpha = 0,05$, se rechaza H_0 , es decir, β_1 sí que es significativo.

- *Contraste de significatividad de β_2*

Vista la explicación en el apartado 4.3.3.1. y la Figura 20, se tiene que :

$$t * \beta_1 = \frac{\beta_1}{ee(\beta_2)} = \frac{0.35}{0.05} = 6$$

Como $t*\beta_1 = 6 \notin [1,66;-1,66]$, se rechaza H_0 y se concluye que β_2 sí que es significativo. Es decir, un cambio en el valor de la variable $PERNOCT.INE^{RU}_{t-12}$ afectará de forma estadísticamente significativa a la variable $PERNOCT.INE^{RU}_t$.

Además, como el valor proporcionado por EViews para $\alpha^*(t*\beta_1)$ es $0,00 < \alpha = 0,05$, se rechaza H_0 , es decir, β_2 sí que es significativo.

- *Contraste de significatividad de β_3*

$$t * \beta_1 = \frac{\beta_1}{ee(\beta_2)} = \frac{163.23}{25.18} = 6.48$$

Como $t*\beta_3 = 6,48 \notin [1,66;-1,66]$, se rechaza H_0 y se concluye que β_3 sí que es significativo. Es decir, un cambio en el valor de la variable $BUSQUEDAS.GT^{RU}_t$ afectará de forma estadísticamente significativa a la variable $PERNOCT.INE^{RU}_t$.

Además, como el valor proporcionado por EViews para $\alpha^*(t*\beta_1)$ es $0,00 < \alpha = 0,05$, se rechaza H_0 , es decir, β_1 sí que es significativo.

- *Contraste de significatividad conjunta*

Teniendo la explicación en el apartado 4.3.3.2 se tiene que :

$$F^* = 1 + \frac{\frac{R^2}{k-1}}{\frac{(1-R^2)}{n-k}} + \frac{\frac{0.73}{4-1}}{\frac{(1-0.73)}{120-3}} = 95.62$$

El valor de R², así como el propio estadístico calculado, es proporcionado por EViews en la pantalla de resultados recogida en la Figura 20.

Como F* = 95.62 > F_(3,117) = 2.70, se rechaza H₀ y se concluye que algún β_i ≠ 0 y, por tanto, sí que existe regresión.

Además, como el valor estimado por EViews en la Figura 19 para la probabilidad asociada al estadístico F* es 0,00 < α = 0,05, se rechaza H₀, es decir, sí que existe regresión.

- *Análisis de la Heterocedasticidad*

Viendo la explicación del método analítico del punto 4.3.4 se tiene la siguiente regresión auxiliar:

$$e = \alpha_0 + \alpha_1 \text{PERNOCT.INE}_{t-1}^{\text{RU}} + \alpha_2 \text{PERNOCT.INE}_{t-1}^{\text{FRAN}} + \alpha_3 \text{PERNOCT.INE}_{t-1}^{\text{RU}} \cdot \text{PERNOCT.INE}_{t-12}^{\text{RU}} + \alpha_4 \text{PERNOCT.INE}_{t-1}^{\text{RU}} \cdot \text{BUSQUEDAS.GT}_{t-12}^{\text{RU}} + \alpha_5 \text{PERNOCT.INE}_{t-12}^{\text{RU}} + \alpha_6 \text{PERNOCT.INE}_{t-12}^{\text{RU}^2} + \alpha_7 \text{PERNOCT.INE}_{t-12}^{\text{RU}} \cdot \text{BUSQUEDAS.GT}_{t-12}^{\text{RU}} + \alpha_8 \text{BUSQUEDAS.GT}_{t-12}^{\text{RU}} + \alpha_9 \text{BUSQUEDAS.GT}_{t-12}^{\text{FRAN}^2}$$

FIGURA 23

Heteroskedasticity Test: White				
F-statistic	1.163618	Prob. F(9,98)	0.3268	
Obs*R-squared	10.42694	Prob. Chi-Square(9)	0.3170	
Scaled explained SS	6.638492	Prob. Chi-Square(9)	0.6747	
Test Equation:				
Dependent Variable: RESID^2				
Method: Least Squares				
Date: 09/05/14 Time: 19:44				
Sample: 2005M01 2013M12				
Included observations: 108				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	10670100	23046342	0.462985	0.6444
PERNOCT_INE_RU(-1)	1646.800	1246.484	1.321156	0.1895
PERNOCT_INE_RU(-1)^2	-0.027789	0.039730	-0.699443	0.4859
PERNOCT_INE_RU(-1)*PERNOCT_INE_RU(-12)	0.011134	0.049733	0.223868	0.8233
PERNOCT_INE_RU(-1)*BUSQUEDAS_RU(-12)	-10.74807	18.16854	-0.591576	0.5555
PERNOCT_INE_RU(-12)	-368.1852	1227.684	-0.299902	0.7649
PERNOCT_INE_RU(-12)^2	0.012858	0.032872	0.391164	0.6955
PERNOCT_INE_RU(-12)*BUSQUEDAS_RU(-12)	-0.778317	17.99567	-0.043250	0.9656
BUSQUEDAS_RU_GT	-536134.0	773964.4	-0.692711	0.4901
BUSQUEDAS_RU_GT^2	6235.134	6288.270	0.991550	0.3239
R-squared	0.096546	Mean dependent var	11687531	
Adjusted R-squared	0.013575	S.D. dependent var	13759556	
S.E. of regression	13665841	Akaike info criterion	35.78672	
Sum squared resid	1.83E+16	Schwarz criterion	36.03506	
Log likelihood	-1922.483	Hannan-Quinn criter.	35.88741	
F-statistic	1.163618	Durbin-Watson stat	2.071509	
Prob(F-statistic)	0.326822			

FUENTE: ELABORACIÓN PROPIA CON EViews

Según muestra la FIGURA 23, este estadístico de White presenta un valor de $10.42 < \chi^2_8 = 15.51$, por lo que se acepta H_0 y se concluye que existe homocedasticidad.

- *Análisis de la autocorrelación*

En el punto 4.3.5 ya se ha visto la explicación detallada de este análisis. Utilizando $\alpha = 0,05$ para calcular los valores críticos, se puede concluir que:

FIGURA 24

Sample: 2005M01 2013M12 Included observations: 108						
Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
		1	0.085	0.085	0.8045	0.370
		2	0.025	0.018	0.8767	0.645
		3	0.130	0.127	2.7869	0.426
		4	-0.021	-0.043	2.8353	0.586
		5	-0.086	-0.087	3.6905	0.595
		6	-0.022	-0.024	3.7452	0.711
		7	0.022	0.038	3.8006	0.802
		8	0.065	0.086	4.3051	0.829
		9	0.104	0.096	5.6132	0.778
		10	-0.021	-0.059	5.6642	0.843
		11	0.183	0.169	9.7818	0.550
		12	-0.089	-0.147	10.765	0.549
		13	-0.009	0.042	10.775	0.630
		14	0.074	0.047	11.462	0.649
		15	-0.147	-0.138	14.210	0.510
		16	-0.091	-0.055	15.271	0.505
		17	-0.070	-0.103	15.902	0.531

FUENTE: ELABORACIÓN PROPIA CON EViews

- Como $Q_1 = 0.80 < \chi^2_1 = 3,84 \rightarrow$ No existe autocorrelación de orden 1
- Como $Q_2 = 0.87 < \chi^2_2 = 5,99 \rightarrow$ No existe autocorrelación de orden 2
- Como $Q_3 = 2.78 < \chi^2_3 = 7,81 \rightarrow$ No existe autocorrelación de orden 3
- Como $Q_4 = 2.83 < \chi^2_4 = 9,49 \rightarrow$ No existe autocorrelación de orden 4
- Como $Q_5 = 3.69 < \chi^2_5 = 11,07 \rightarrow$ No existe autocorrelación de orden 5
- Como $Q_6 = 3.74 < \chi^2_6 = 12,59 \rightarrow$ No existe autocorrelación de orden 6
- Como $Q_7 = 3.80 < \chi^2_7 = 14,07 \rightarrow$ No existe autocorrelación de orden 7
- Como $Q_8 = 4.30 < \chi^2_8 = 15,51 \rightarrow$ No existe autocorrelación de orden 8
- Como $Q_9 = 4,62 < \chi^2_9 = 16,92 \rightarrow$ No existe autocorrelación de orden 9
- Como $Q_{10} = 4,63 < \chi^2_{10} = 18,31 \rightarrow$ No existe autocorrelación de orden 10
- Como $Q_{11} = 5,09 < \chi^2_{11} = 19,68 \rightarrow$ No existe autocorrelación de orden 11
- Como $Q_{12} = 5,59 < \chi^2_{12} = 21,03 \rightarrow$ No existe autocorrelación de orden 12

4.4.3. VALIDACIÓN DEL MODELO DE GOOGLE TRENDS CON DATOS DE ALEMANIA

➤ Modelo de referencia

FIGURA 25

Dependent Variable: PERNOCT_INE_ALEM Method: Least Squares				
Sample (adjusted): 2005M01 2013M12 Included observations: 108 after adjustments				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	6658.670	1107.767	6.010895	0.0000
PERNOCT_INE_ALEM(-12)	0.523111	0.083229	6.285223	0.0000
R-squared	0.271498	Mean dependent var		13115.49
Adjusted R-squared	0.264625	S.D. dependent var		5023.051
S.E. of regression	4307.467	Akaike info criterion		19.59243
Sum squared resid	1.97E+09	Schwarz criterion		19.64210
Log likelihood	-1055.991	Hannan-Quinn criter.		19.61257
F-statistic	39.50403	Durbin-Watson stat		1.195429
Prob(F-statistic)	0.000000			

FUENTE: ELABORACIÓN PROPIA CON EVIEW

En la Figura 25 podemos observar que el modelo de referencia tiene un R² ajustado del 26,4%, es decir, con este modelo se explica en un 26,4% el comportamiento de la variable explicada (número de pernoctaciones de turistas franceses en el mes t). El AIC es del 19.59 y el MAPE (Mean Absolut Percet Error) del 25.9%.

FIGURA 26

Sample: 2005M01 2013M12 Included observations: 108						
Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
		1	0.394	0.394	17.242	0.000
		2	0.179	0.028	20.847	0.000
		3	0.018	-0.073	20.885	0.000
		4	0.108	0.142	22.219	0.000
		5	0.077	-0.000	22.898	0.000
		6	0.043	-0.019	23.112	0.001
		7	-0.036	-0.048	23.267	0.002
		8	-0.159	-0.162	26.261	0.001
		9	-0.074	0.056	26.910	0.001
		10	-0.084	-0.060	27.773	0.002
		11	-0.023	0.014	27.838	0.003
		12	-0.207	-0.197	33.159	0.001
		13	-0.010	0.184	33.171	0.002
		14	-0.181	-0.244	37.310	0.001
		15	-0.132	-0.030	39.520	0.001
		16	-0.173	-0.068	43.382	0.000
		17	-0.039	0.056	43.584	0.000
		18	0.031	0.066	43.713	0.001

FUENTE: ELABORACIÓN PROPIA CON EVIEW

En la Figura 26 se puede observar que este modelo presenta problemas de autocorrelación, por lo que se procede a eliminarla introduciendo como variable explicativa $PERNOCT.INE^{ALEM}_{t-1}$, siendo el modelo de referencia para este país

$$PERNOCT.INE^{ALEM}_t = \beta_0 + \beta_1 PERNOCT.INE^{ALEM}_{t-1} + \beta_2 PERNOCT.INE^{ALEM}_{t-12} + \varepsilon_t$$

donde,

$PERNOCT.INE^{ALEM}_t$ = número de pernoctaciones de turistas procedentes del Alemania en Valencia en el mes t.

$PERNOCT.INE^{ALEM}_{t-1}$ = número de pernoctaciones de turistas procedentes del Alemania en Valencia en el mes anterior

$PERNOCT.INE^{ALEM}_{t-12}$ = número de pernoctaciones de turistas procedentes del Alemania en Valencia doce meses antes

ε_t = perturbación aleatoria

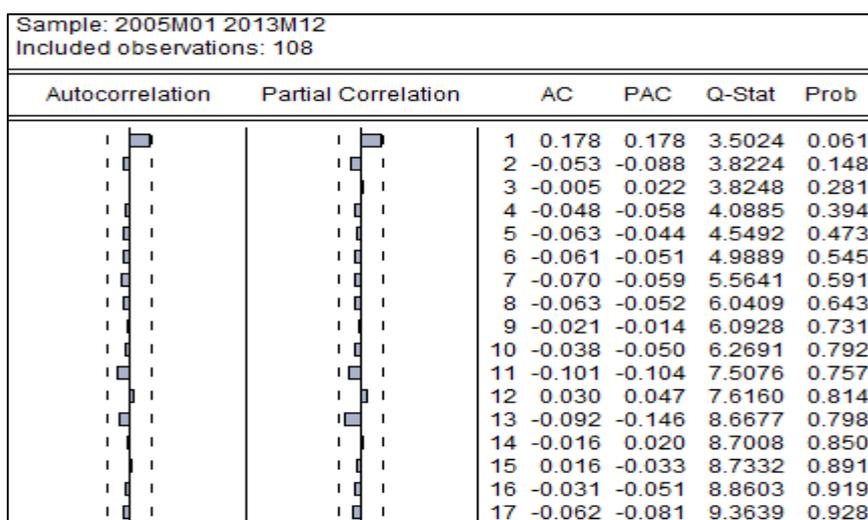
FIGURA 27

Dependent Variable: PERNOCT_INE_ALEM				
Method: Least Squares				
Sample (adjusted): 2005M01 2013M12				
Included observations: 108 after adjustments				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	3145.200	1252.635	2.510867	0.0136
PERNOCT_INE_ALEM(-1)	0.375821	0.079243	4.742653	0.0000
PERNOCT_INE_ALEM(-12)	0.408582	0.079639	5.130402	0.0000
R-squared	0.400023	Mean dependent var		13115.49
Adjusted R-squared	0.388595	S.D. dependent var		5023.051
S.E. of regression	3927.644	Akaike info criterion		19.41685
Sum squared resid	1.62E+09	Schwarz criterion		19.49136
Log likelihood	-1045.510	Hannan-Quinn criter.		19.44706
F-statistic	35.00335	Durbin-Watson stat		1.782327
Prob(F-statistic)	0.000000			

FUENTE: ELABORACIÓN PROPIA CON EVIEW

En la Figura 27 podemos observar que el nuevo modelo de referencia tiene un R^2 ajustado del 38.8%, es decir, con este modelo se explica en un 38.8% el comportamiento de la variable explicada (número de pernoctaciones de turistas del Alemania en el mes t). El AIC es del 19.4 y el MAPE (Mean Absolut Percet Error) del 23.3%.

FIGURA 28



FUENTE: ELABORACIÓN PROPIA CON EVIEW

En la Figura 28 se puede ver que la autocorrelación de orden 1 se ha corregido.

➤ Modelo con datos del Google Trends

Especificaciones que se han tenido en cuenta para los datos del Google Trends:

- *Especificación 1*: no consideran términos autorregresivos.
- *Especificación 2*: se considera el término autorregresivo de primer orden.
- *Especificación 3*: se considera el término autorregresivo de segundo orden.
- *Especificación 4*: se considera tanto la popularidad del término Valencia-City in Spain en Alemania en el mes actual, como en el mes anterior.

Variables	Coeficientes			
	Espec. 1	Espec. 2	Espec. 3	Espec. 4
Constante	-254.68	619.78	3502.5*	-769.69
PERNOCT. INE ALEM $t-1$	0.39***	0.35***	0.38***	0.37***
PERNOCT. INE ALEM $t-12$	0.38***	0.39***	0.40***	0.37***
BUSQUEDAS. GT ALEM t	86.01**			67.78
BUSQUEDAS. GT ALEM $t-1$		71.84*		35.14
BUSQUEDAS. GT ALEM $t-2$			-10.43	
<i>N</i>	108	108	108	108
<i>R</i> ² ajustado	40.8%	40%	38.3%	40.6%
<i>AIC</i>	19.39	19.4	19.4	19.4
<i>MAPE</i>	22.2%	22.7%	23%	22.2%

Fuente: elaboración propia

Se elige la especificación 1, por ser la que mayor R² Ajustado, menor AIC y MAPE y la que tiene todas las variables con un P-Value < 0.05, es decir, todas las variables significativas.

Por tanto, el modelo con datos del Google Trends elegido es:

$$\text{PERNOCT.INE}^{\text{ALEM}}_t = \beta_0 + \beta_1 \text{PERNOCT.INE}^{\text{ALEM}}_{t-1} + \beta_2 \text{PERNOCT.INE}^{\text{ALEM}}_{t-12} + \beta_3 \text{BUSQUEDAS.GT}^{\text{ALEM}}_t + \varepsilon_t$$

PERNOCT.INE^{ALEM}_t = número de pernoctaciones de turistas procedentes del Alemania en Valencia en el mes t.

PERNOCT.INE^{ALEM}_{t-1} = número de pernoctaciones de turistas procedentes del Alemania en Valencia en el mes anterior

PERNOCT.INE^{ALEM}_{t-12} = número de pernoctaciones de turistas procedentes del Alemania en Valencia doce meses antes

BUSQUEDAS.GT^{ALEM}_t = volumen de búsquedas del término Valencia-City in Spain realizadas en Alemania, en el mes t.

εt = perturbación aleatoria

4.4.3.1. Validación del modelo con datos del Google Trends

- *Contraste de errores de especificación*

FIGURA 29

Dependent Variable: PERNOCT_INE_ALEM				
Method: Least Squares				
Sample (adjusted): 2005M01 2013M12				
Included observations: 108 after adjustments				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-254.6835	2004.664	-0.127045	0.8991
PERNOCT_INE_ALEM(-1)	0.391917	0.078270	5.007241	0.0000
PERNOCT_INE_ALEM(-12)	0.380247	0.079403	4.788828	0.0000
BUSQUEDAS_ALEM_GT	86.01113	40.01511	2.149467	0.0339
R-squared	0.425543	Mean dependent var		13115.49
Adjusted R-squared	0.408972	S.D. dependent var		5023.051
S.E. of regression	3861.637	Akaike info criterion		19.39190
Sum squared resid	1.55E+09	Schwarz criterion		19.49124
Log likelihood	-1043.163	Hannan-Quinn criter.		19.43218
F-statistic	25.68020	Durbin-Watson stat		1.840378
Prob(F-statistic)	0.000000			

FUENTE: ELABORACIÓN PROPIA CON EViews

Viendo la Figura 29, para aplicar el test de RESET de Ramsey se tiene que que la variable FFITED se define como:

$$\text{FITTED} = -254.68 + 0.39 \cdot \text{PERNOCT.INE}^{\text{ALEM}}_{t-1} + 0.38 \cdot \text{PERNOCT.INE}^{\text{ALEM}}_{t-12} + 86.01 \cdot \text{BUSQUEDAS.GT}^{\text{ALEM}}_{t} + \varepsilon_t$$

El modelo transformado queda:

$$\text{PERNOCT.INE}^{\text{ALEM}}_{t} = \beta_0 + \beta_1 \text{PERNOCT.INE}^{\text{ALEM}}_{t-1} + \beta_2 \text{PERNOCT.INE}^{\text{ALEM}}_{t-12} + \beta_3 \text{BUSQUEDAS.GT}^{\text{ALEM}}_{t} + \beta_4 \text{FITTED}^2 + \varepsilon_t$$

FIGURA 30

Unrestricted Test Equation:				
Dependent Variable: PERNOCT_INE_ALEM				
Method: Least Squares				
Sample: 2005M01 2013M12				
Included observations: 108				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-12286.33	5188.612	-2.367941	0.0198
PERNOCT_INE_ALEM(-1)	1.131079	0.304964	3.708887	0.0003
PERNOCT_INE_ALEM(-12)	1.110071	0.301637	3.680154	0.0004
BUSQUEDAS_ALEM_GT	228.4695	69.00738	3.310799	0.0013
FITTED^2	-6.86E-05	2.74E-05	-2.503512	0.0139
R-squared	0.458494	Mean dependent var	13115.49	
Adjusted R-squared	0.437465	S.D. dependent var	5023.051	
S.E. of regression	3767.407	Akaike info criterion	19.35135	
Sum squared resid	1.46E+09	Schwarz criterion	19.47552	
Log likelihood	-1039.973	Hannan-Quinn criter.	19.40170	
F-statistic	21.80257	Durbin-Watson stat	1.733654	
Prob(F-statistic)	0.000000			

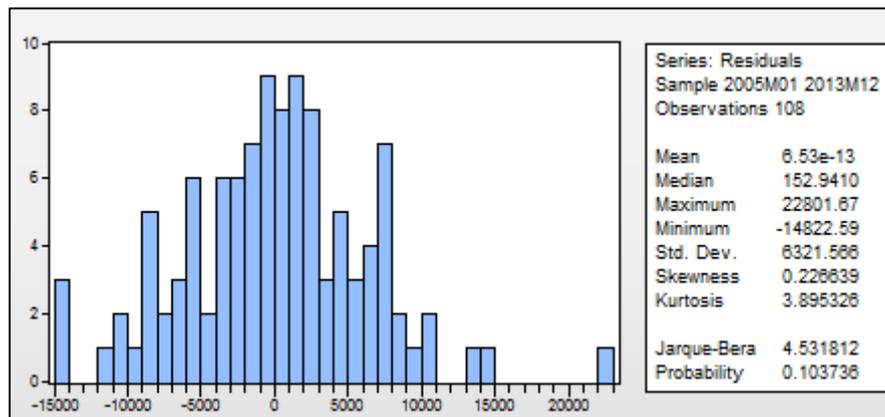
FUENTE: ELABORACIÓN PROPIA CON EViews

En la Figura 30 podemos observar que el valor ofrecido por Eviews para β_4 es $0,01 < \alpha = 0,05$, hecho que a nos lleva a rechazar la H_0 que veíamos en el apartado 4.3.1, y ver que B_4 es significativo y, por tanto, existen errores de especificación.

Esto no es preocupante, a priori, porque simplemente es indicio de que faltan variables explicativas en el modelo. Esto no es de extrañar, pues el modelo planteado es un modelo básico en base al objetivo del presente trabajo, el cual no es hacer un modelo que explique lo mejor posible la variable explicativa sino ver si los datos del Google Trends son útiles para explicarla y predecirla.

- *Análisis de la normalidad de las perturbaciones*

FIGURA 31



FUENTE: ELABORACIÓN PROPIA CON EViews

Vista la explicación del punto 4.3.2, en la Figura 31 se observa que el estadístico Jarque-Bera presenta un valor de $4.5 < 5.9$, por lo que se acepta H_0 y se concluye que las perturbaciones se distribuyen conforme a la distribución normal.

Este resultado también indica que el resto de contrastes que se van a efectuar son fiables, por lo que sí se puede realizar inferencia en el modelo.

- *Contraste de significatividad de β_1*

Vista la explicación en el apartado 4.3.3.1 y la Figura 29, se tiene que :

$$t * \beta_1 = \frac{\beta_1}{ee(\beta_2)} = \frac{0.39}{0.07} = 5$$

Como $t * \beta_1 = 5 \notin [1.66; -1.66]$, se rechaza H_0 y se concluye que β_1 sí que es significativo. Es decir, un cambio en el valor de la variable $PERNOCT.INE^{ALEM}_{t-1}$ afectará de forma estadísticamente significativa a la variable $PERNOCT.INE^{ALEM}_t$.

Además, como el valor proporcionado por EViews para $\alpha^*(t * \beta_1)$ es $0.00 < \alpha = 0.05$, se rechaza H_0 , es decir, β_1 sí que es significativo.

- *Contraste de significatividad de β_2*

Vista la explicación en el apartado 4.3.3.1 y la Figura 29, se tiene que :

$$t * \beta_1 = \frac{\beta_1}{ee(\beta_2)} = \frac{0.38}{0.07} = 4.7$$

Como $t_{\beta_1} = 4.7 \notin [1,66;-1,66]$, se rechaza H_0 y se concluye que β_2 sí que es significativo. Es decir, un cambio en el valor de la variable $PERNOCT.INE^{ALEM}_{t-12}$ afectará de forma estadísticamente significativa a la variable $PERNOCT.INE^{ALEM}_t$.

Además, como el valor proporcionado por EViews para $\alpha^*(t_{\beta_1})$ es $0,00 < \alpha = 0,05$, se rechaza H_0 , es decir, β_1 sí que es significativo.

- *Contraste de significatividad de β_3*

Vista la explicación en el apartado 4.3.3.1 y la Figura 29, se tiene que

$$t * \beta_1 = \frac{\beta_1}{ee(\beta_2)} = \frac{86.01}{40.01} = 2.15$$

Como $t_{\beta_3} = 2.15 \notin [1,66;-1,66]$, se rechaza H_0 y se concluye que β_3 sí que es significativo. Es decir, un cambio en el valor de la variable $BUSQUEDAS.GT^{ALEM}_t$ afectará de forma estadísticamente significativa a la variable $PERNOCT.INE^{ALEM}_t$.

Además, como el valor proporcionado por EViews para $\alpha^*(t_{\beta_1})$ es $0,00 < \alpha = 0,05$, se rechaza H_0 , es decir, β_1 sí que es significativo.

- *Contraste de significatividad conjunta*

Teniendo la explicación en el apartado 4.3.3.2 se tiene que:

$$F^* = 1 + \frac{\frac{R^2}{k-1}}{\frac{(1-R^2)}{n-k}} + \frac{\frac{0.41}{4-1}}{\frac{(1-0041)}{120-3}} = 25.68$$

El valor de R^2 , así como el propio estadístico calculado, es proporcionado por EViews en la pantalla de resultados recogida en la Figura 29.

Como $F^* = 25.68 > F_{(3,117)} = 2.70$, se rechaza H_0 y se concluye que algún $\beta_i \neq 0$ y, por tanto, sí que existe regresión.

Además, como el valor estimado por EViews en la Figura 29 para la probabilidad asociada al estadístico F^* es $0,00 < \alpha = 0,05$, se rechaza H_0 , es decir, sí que existe regresión.

▪ *Análisis de la Heterocedasticidad*

Viendo la explicación del método analítico del punto 4.3.4 se tiene la siguiente regresión auxiliar:

$$\begin{aligned}
 e = & \alpha_0 + \alpha_1 \text{PERNOCT.INE}^{\text{ALEM}}_{t-1} + \alpha_2 \text{PERNOCT.INE}^{\text{ALEM}}_{t-1}{}^2 + \\
 & + \alpha_3 \text{PERNOCT.INE}^{\text{ALEM}}_{t-1} \cdot \text{PERNOCT.INE}^{\text{ALEM}}_{t-12} + \alpha_4 \text{PERNOCT.INE}^{\text{ALEM}}_{t-1} \\
 & \text{BUSQUEDAS.GT}^{\text{ALEM}}_{t+} + \alpha_5 \text{PERNOCT.INE}^{\text{ALEM}}_{t-12} + \alpha_6 \text{PERNOCT.INE}^{\text{ALEM}}_{t-12}{}^2 \\
 & + \alpha_7 \text{PERNOCT.INE}^{\text{ALEM}}_{t-12} \cdot \text{BUSQUEDAS.GT}^{\text{ALEM}}_{t+} + \alpha_8 \text{BUSQUEDAS.GT}^{\text{ALEM}}_{t+} \\
 & + \alpha_9 \text{BUSQUEDAS.GT}^{\text{ALEM}}_{t+}{}^2
 \end{aligned}$$

FIGURA 32

Heteroskedasticity Test: White				
F-statistic	1.197788	Prob. F(9,98)	0.3051	
Obs*R-squared	10.70279	Prob. Chi-Square(9)	0.2966	
Scaled explained SS	27.21783	Prob. Chi-Square(9)	0.0013	
Test Equation:				
Dependent Variable: RESID^2				
Method: Least Squares				
Date: 09/05/14 Time: 22:12				
Sample: 2005M01 2013M12				
Included observations: 108				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-62455340	81497016	-0.766351	0.4453
PERNOCT_INE_ALEM(-1)	4612.211	5182.206	0.890009	0.3756
PERNOCT_INE_ALEM(-1)^2	0.002542	0.084504	0.030080	0.9761
PERNOCT_INE_ALEM(-1)*PERNOCT_IN...	0.225052	0.190327	1.182450	0.2399
PERNOCT_INE_ALEM(-1)*BUSQUEDAS_...	-144.0039	108.4629	-1.327678	0.1874
PERNOCT_INE_ALEM(-12)	-6581.416	5842.011	-1.126567	0.2627
PERNOCT_INE_ALEM(-12)^2	-0.066663	0.086972	-0.766487	0.4452
PERNOCT_INE_ALEM(-12)*BUSQUEDAS...	142.4992	100.4080	1.419202	0.1590
BUSQUEDAS_ALEM_GT	3216097.	2798314.	1.149298	0.2532
BUSQUEDAS_ALEM_GT^2	-32283.12	26702.87	-1.208976	0.2296
R-squared	0.099100	Mean dependent var	14359937	
Adjusted R-squared	0.016364	S.D. dependent var	33787515	
S.E. of regression	33509923	Akaike info criterion	37.58060	
Sum squared resid	1.10E+17	Schwarz criterion	37.82895	
Log likelihood	-2019.353	Hannan-Quinn criter.	37.68130	
F-statistic	1.197788	Durbin-Watson stat	1.887607	
Prob(F-statistic)	0.305145			

FUENTE: ELABORACIÓN PROPIA CON EViews

Según muestra la FIGURA 32, este estadístico de White presenta un valor de $10.7 < \chi^2_8 = 15.51$, por lo que se acepta H_0 y se concluye que existe homocedasticidad.

- *Análisis de la autorcorrelación*

En el punto 3.4. ya se ha visto la explicación detallada de este análisis. Utilizando $\alpha = 0,05$ para calcular los valores críticos, se puede concluir que:

FIGURA 33

Sample: 2005M01 2013M12 Included observations: 108						
Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
		1	0.074	0.074	0.6005	0.438
		2	0.053	0.047	0.9101	0.634
		3	-0.107	-0.115	2.2014	0.532
		4	0.122	0.138	3.8885	0.421
		5	0.122	0.117	5.6164	0.345
		6	0.147	0.107	8.1453	0.228
		7	0.072	0.075	8.7479	0.271
		8	-0.125	-0.145	10.599	0.226
		9	-0.045	-0.039	10.844	0.287
		10	-0.072	-0.089	11.473	0.322
		11	0.094	0.034	12.546	0.324
		12	-0.129	-0.147	14.593	0.264
		13	0.082	0.110	15.429	0.281
		14	-0.231	-0.184	22.191	0.075
		15	-0.075	-0.053	22.911	0.086
		16	-0.173	-0.120	26.778	0.044
		17	0.008	-0.015	26.786	0.061

FUENTE: ELABORACIÓN PROPIA CON EViews

- Como $Q_1 = 0,60 < \chi^2_1 = 3,84 \rightarrow$ No existe autocorrelación de orden 1
- Como $Q_2 = 0,91 < \chi^2_2 = 5,99 \rightarrow$ No existe autocorrelación de orden 2
- Como $Q_3 = 2,20 < \chi^2_3 = 7,81 \rightarrow$ No existe autocorrelación de orden 3
- Como $Q_4 = 3,88 < \chi^2_4 = 9,49 \rightarrow$ No existe autocorrelación de orden 4
- Como $Q_5 = 5,61 < \chi^2_5 = 11,07 \rightarrow$ No existe autocorrelación de orden 5
- Como $Q_6 = 8,14 < \chi^2_6 = 12,59 \rightarrow$ No existe autocorrelación de orden 6
- Como $Q_7 = 8,74 < \chi^2_7 = 14,07 \rightarrow$ No existe autocorrelación de orden 7
- Como $Q_8 = 10,6 < \chi^2_8 = 15,51 \rightarrow$ No existe autocorrelación de orden 8
- Como $Q_9 = 4,62 < \chi^2_9 = 16,92 \rightarrow$ No existe autocorrelación de orden 9
- Como $Q_{10} = 4,63 < \chi^2_{10} = 18,31 \rightarrow$ No existe autocorrelación de orden 10
- Como $Q_{11} = 5,09 < \chi^2_{11} = 19,68 \rightarrow$ No existe autocorrelación de orden 11
- Como $Q_{12} = 5,59 < \chi^2_{12} = 21,03 \rightarrow$ No existe autocorrelación de orden 12

4.5. Análisis de las hipótesis planteadas

En el Capítulo 2 se plantearon las siguientes hipótesis:

H1: la popularidad del término de búsqueda *Valencia-City in Spain* ayuda a predecir el número de pernoctaciones de turistas a la ciudad de Valencia.

H2: la popularidad, en un país, del término de búsqueda *Valencia-City in Spain* mejora la predicción del número de pernoctaciones de turistas procedentes de ese país en la ciudad de Valencia.

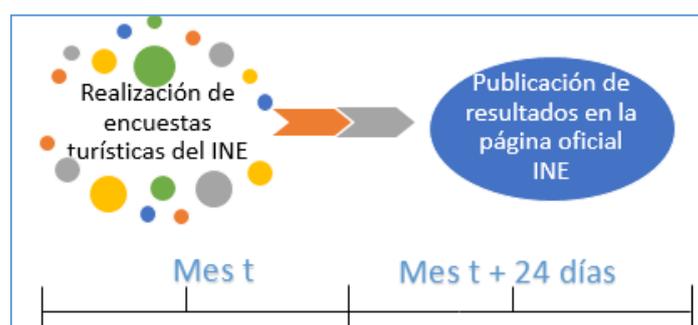
4.5.1. HIPÓTESIS 1

H1: la popularidad del término de búsqueda Valencia-City in Spain ayuda a predecir el número de pernoctaciones de turistas a la ciudad de Valencia.

La hipótesis 1 se puede aceptar, ya que, como se ha podido comprobar en los modelos con datos del Google Trends vistos en el apartado 4.3, el valor β de la variable $BUSQUEDAS.GT^i_t$ es distinto de cero, y por tanto, existe una relación entre el volumen de búsquedas del término analizado en un país determinado en el instante t y el número de turistas procedentes de ese país en el mismo instante.

Por otra parte, un modelo autorregresivo básico, como el modelo de referencia planteado, utiliza datos oficiales que tardan en publicarse. En la figura 34 podemos ver que los datos del INE para el sector turístico se publican 24 días posteriores a al mes t

FIGURA 34



FUENTE: ELABORACIÓN PROPIA CON EViews

Con el modelo con datos del Google Trends, se puede predecir el número de turistas procedentes del país i al final del mes t , con un error inferior al del modelo de referencia ($18,9 \% < 19,5\%$), por lo que se justifica que los datos del Google Trends ayudan a predecir el número de pernoctaciones de turistas en la ciudad Valencia.

4.5.2. HIPÓTESIS 2

H2: la popularidad, en un país, del término de búsqueda Valencia-City in Spain mejora la predicción del número de pernoctaciones de turistas procedentes de ese país en la ciudad de Valencia.

La hipótesis 2 se demuestra cierta al comprobar que modelos de predicción con datos del Google Trends como los vistos en el apartado 4.2 presentan mayor R^2 ajustado. En el caso de Italia se podía ver que mientras el modelo de referencia explicaba el comportamiento de la variable explicada en un 77%, el modelo con datos del Google Trends lo explicaba en 83%. Por tanto, los datos del Google Trends del país analizado ayudan en la predicción de turistas de ese país al aumentar el porcentaje de explicación del modelo predictivo.

Por otra parte, el AIC y el MAPE eran menores en los modelos con datos del Google Trends de los países objeto de estudio que en los de referencia. En el modelo de referencia de Italia el AIC era de un 20,7 mientras que en el modelo con datos del Google Trends era de un 20,4. Este $20,4 < 20,7$ dejar ver que el modelo con datos del Google Trends es de mayor calidad y bondad de ajuste. En cuanto al MAPE, en el modelo de referencia era de un 19,5% y en el modelo con datos del Google Trends de un 18,9%, denotando esto que el modelo con datos del Google Trends tiene un error de predicción medio inferior al de referencia y por tanto mejora la predicción.

El presente Trabajo de Fin de Grado ha validado el uso de Google Trends para predecir el nivel de pernoctaciones de turistas en Valencia. Además, en el mismo se han podido confirmar los resultados de investigaciones previas con respecto al uso de Google trends para predecir el nivel de turismo.

Las hipótesis planteadas se han demostrado ciertas, por lo que la popularidad del término de búsqueda *Valencia-City in Spain* ayuda a predecir el número de pernoctaciones de turistas a la ciudad de Valencia no solo de manera agregada (Hipótesis 1) sino para turistas de un origen determinado de búsqueda (Hipótesis 2). De ahí que el carácter innovador del presente trabajo está en no quedarse solo en validar el uso de Google Trends para la predicción agregada de turistas en una país/zona determinado/a, como hicieron Choi y Varian, sino en validarlo para predecir el número de pernoctaciones de turistas procedentes de un país en concreto, teniendo en cuenta el volumen de búsquedas del mismo.

En cuanto a los países tenidos en cuenta, es de interés ver que aunque a priori un país tenga un uso más extendido de internet que otro, esto no implica que los datos de su volumen de búsquedas sean más significativo para el análisis en el sector turístico que el que se de en un país de uso menos extendido de Internet. Así pues, se ha visto como en los modelos de predicción para los países de uso menos extendido de internet (Italia y Francia), la variable explicativa “Volumen de búsqueda del término analizado, en el país analizado y en el instante de tiempo t) aumentaba mucho más el porcentaje de explicación del modelo que en los países de uso más extendido de internet (Alemania y Reino Unido).

Otro hecho a destacar de los análisis hechos es que si se compara la estacionalidad de las series del Google Trends con la estacionalidad de series oficiales como la del INE, se pueden sacar datos importantes a la hora de plantear un buen modelo. Por ejemplo, para el caso de Francia en la series temporal del volumen de búsqueda se podía apreciar altos volúmenes de búsqueda que iban desde el mes de marzo hasta el mes de agosto, mientras que en la serie del número de pernoctaciones de estos turistas a Valencia se apreciaban puntos de alta demanda más claros en Agosto. Al plantear el modelo y tener en cuenta esto en la variable con datos de Google Trends, haciendo que se tengan en cuenta los datos 6 meses atrás al período t , se consiguió un modelo mejor y con mayor porcentaje de explicación de la variable endógena que si se hubieran tenido en cuenta datos del período t o $t-1$.

A modo de recalcar la importancia del presente trabajo para el sector turístico en general y concretamente, para el sector turístico de la provincia de Valencia, se puede decir que los modelos presentados ofrecen una oportunidad para estimar el número aproximado de pernoctaciones / turistas que habrá en el corto plazo y con anterioridad a la publicación del mismo por páginas oficiales como el INE. Esta oportunidad se traduce en un margen de maniobra, para los establecimientos del sector turístico, más amplio para tomar decisiones en cuanto a pedidos que tengan que hacer o al personal que necesitan tener para atender esa demanda turística (útiles tanto para la hostelería como para los establecimientos hoteleros).

Como limitaciones, este trabajo ha tenido, en primer lugar, la periodicidad mensual del indicador principal del número de pernoctaciones de turistas a Valencia por país de procedencia (el INE), la cual dificulta el uso de los datos semanales del Google Trends. Esto se explica en que se han tenido que ajustar los datos semanales de Google Trends con los datos mensuales del INE, y para ello se han realizado medias aritméticas.

Por otra parte, el uso de internet a la hora de viajar se ha visto incrementado de manera más significativa en los últimos años, por lo que, aunque dispongamos de volúmenes de búsqueda semanales desde el año 2004, es a partir del año 2008 cuando se observa una estacionalidad más marcada en las búsquedas del término analizado. Es decir, se observa que es en los últimos años cuando los turistas hacen más uso de internet a la hora de buscar información sobre la ciudad en cuestión a la que van a viajar, habiendo sido más utilizadas anteriormente otras vías para ello, como podrían ser las agencias de viajes. Por tanto, esta dificultad radica en que actualmente no se dispone de una muestra significativa de volúmenes de búsquedas, fruto de años de uso de internet predominante a la hora de hacer el viaje, con los cuales conseguir un modelo válido para toda la serie temporal.

A pesar de estas limitaciones, el uso de Google Trends para predecir el nivel de pernoctaciones de turistas procedentes de un país i , en un instante de tiempo t ha sido validado. Como trabajo futuro se podría proponer el comprobar si los porcentajes de explicación de los modelos planteados aumentan significativamente al introducir más años con volúmenes de búsqueda de términos turísticos altos, signo de una mayor uso de internet a la hora de realizar viajes turísticos. También se podría plantear un modelo econométrico que eligiera el término de búsqueda más explicativo de nivel de turistas para cada Comunidad Autónoma. Por último, la metodología empleada en este trabajo se podría aplicar a cualquier otro tipo de variable. Incluso en un enfoque básico, podría ser usado por las tiendas de ropa para anticipar que tipo de prendas se van a vender en el corto plazo y así evitar tener acumulaciones de stock.

CAPÍTULO 6. BIBLIOGRAFÍA

CAPÍTULO 6. BIBLIOGRAFÍA

6.1. Bibliografía

- ✚ Banco Mundial (2014): Información sobre los usuarios de internet (<http://datos.bancomundial.org/indicador/IT.NET.USER.P2/countries/DE-FR-IT-GB?display=graph>) (Consulta: 5 Agosto 2014)
- ✚ BARHOUMI, K.; DARN, O. K.; FERRARA, L. (2010) Are disaggregate data useful for factor analysis in forecasting French GDP? *Journal of Forecasting*, nº 29(1-2) pp. 132-144.
- ✚ CONCHA ARTOLA Y ENRIQUE GALÁN "LAS HUELLAS DEL FUTURO ESTÁN EN LA WEB: CONSTRUCCIÓN DE INDICADORES ADELANTADOS A PARTIR DE LAS BÚSQUEDAS EN INTERNET". *Documentos Ocasionales de Banco de España 2012*, N.º 1203.
- ✚ CHOI, H.; VARIAN, H. "Predicting the Present with Google Trends". *Googleusercontent*, 10 Abril,2009, p. 15-17. Disponible en : http://static.googleusercontent.com/media/www.google.com/es//googleblogs/pdfs/google_predicting_the_present.pdf
- ✚ CHOI, H.; VARIAN, H. (2009) Predicting the present with Google Trends. *Economic Record*, nº88 pp. 2-9.
- ✚ DZIELINSKI, M. (2012) Measuring economic uncertainty and its impact in the stock market. *Finance Research Letters*, nº 9 (3) pp. 167-175.
- ✚ EVAN GAWLIK, HARDIK KABARIA AND SHAGANDEEP KAUR, "Predicting tourism trends with Google Insights" *Stanford.edu*, 2011, Abstract. Disponible en: <http://cs229.stanford.edu/proj2011/GawlikKaurKabariaPredictingTourismTrendsWithGoogleInsights.pdf>
- ✚ GIANNONE, DOMENICO Y REICHLIN, LUCREZIA Y PEQUEÑOS, "predicción inmediata: El contenido de la información en tiempo real de los datos macroeconómicos," *Journal of Monetary Economics, Elsevier*, 2008, vol. 55 (4), páginas 665-676.
- ✚ GIANNONE, D.; REICHLIN, L.; SMALL, D. (2008) Nowcasting: The real-time informational content of macroeconomic data. *Journal of Monetary Economics*, nº55(4) pp. 665-676.
- ✚ GINSBERG, J.; MOHEBBI, M. H.; PATEL, R. S. ; BRAMMER, L.; SMOLINSKI, M. S.; BRILLIANT, L. (2009) Detecting influenza epidemics using search engine query data. *Nature*, nº457 pp. 1012-1014.
- ✚ GOEL, S; HOFMAN, J. M.; LAHAIE, S.; PENNOCK, D. M.; WATTS, D. J.(2010) Predicting consumer behaviour with web search. *Proceedings of the National Academy of Sciencis*, nº 7 (41) pp. 17846-17490.

🚩 GOOGLE (2014): Información sobre Google Trends

(https://support.google.com/trends/topic/13973?hl=es&ref_topic=13761) (Consulta: 10 Julio 2014)

🚩 GUZMAN, G. (2011) Internet search behaviour as an economic forecasting tool. *The Journal of Economic and Social Measurement*, nº 36 (1-3) pp. 337-386.

🚩 INE(2014): Información sobre las series temporales

(http://www.ine.es/inebmenu/mnu_hosteleria.htm) (Consulta: 10 Julio 2014)

🚩 INE(2014): Información sobre la institución

(http://www.ine.es/explica/docs/estaofi_quesine.pdf) (Consulta 10 Julio 2014)

🚩 Posicionamiento Internacional "CUOTA DE MERCADO DE BUSCADORES POR PAÍSES" 25 de Septiembre de 2012 (Consulta: 22 Julio de 2014). Disponible en: <http://www.posicionamientointernacional.com/cuota-de-mercado-de-buscadores-por-paises>

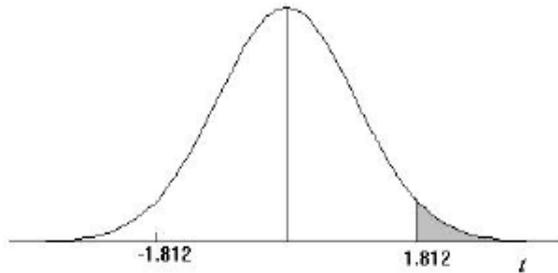
🚩 S.G. de Conocimiento y Estudios Turísticos "FRONTUR EGATUR". *Informe anual 2012*, <http://www.060.es>, NIPO: 072-13-002-6

6.2. Anexos

Anexo 1: Distribución t de Student

TABLA 2: DISTRIBUCIÓN t DE STUDENT

Puntos de porcentaje de la distribución t



Ejemplo

Para $\phi = 10$ grados de libertad:

$$P[t > 1.812] = 0.05$$

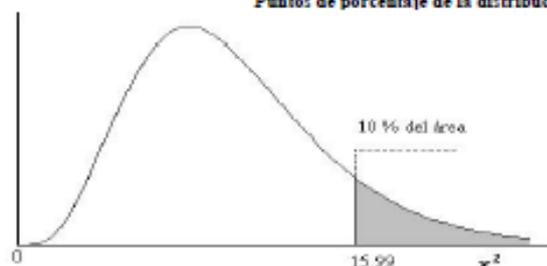
$$P[t < -1.812] = 0.05$$

α Γ	0,25	0,2	0,15	0,1	0,05	0,025	0,01	0,005	0,0005
1	1,000	1,376	1,963	3,078	6,314	12,706	31,821	63,656	636,578
2	0,816	1,061	1,386	1,886	2,920	4,303	6,965	9,925	31,600
3	0,765	0,978	1,250	1,638	2,353	3,182	4,541	5,841	12,924
4	0,741	0,941	1,190	1,533	2,132	2,776	3,747	4,604	8,610
5	0,727	0,920	1,156	1,476	2,015	2,571	3,365	4,032	6,869
6	0,718	0,906	1,134	1,440	1,943	2,447	3,143	3,707	5,959
7	0,711	0,896	1,119	1,415	1,895	2,365	2,998	3,499	5,408
8	0,706	0,889	1,108	1,397	1,860	2,306	2,896	3,355	5,041
9	0,703	0,883	1,100	1,383	1,833	2,262	2,821	3,250	4,781
10	0,700	0,879	1,093	1,372	1,812	2,228	2,764	3,169	4,587
11	0,697	0,876	1,088	1,363	1,796	2,201	2,718	3,106	4,437
12	0,695	0,873	1,083	1,356	1,782	2,179	2,681	3,055	4,318
13	0,694	0,870	1,079	1,350	1,771	2,160	2,650	3,012	4,221
14	0,692	0,868	1,076	1,345	1,761	2,145	2,624	2,977	4,140
15	0,691	0,866	1,074	1,341	1,753	2,131	2,602	2,947	4,073
16	0,690	0,865	1,071	1,337	1,746	2,120	2,583	2,921	4,015
17	0,689	0,863	1,069	1,333	1,740	2,110	2,567	2,898	3,965
18	0,688	0,862	1,067	1,330	1,734	2,101	2,552	2,878	3,922
19	0,688	0,861	1,066	1,328	1,729	2,093	2,539	2,861	3,883
20	0,687	0,860	1,064	1,325	1,725	2,086	2,528	2,845	3,850
21	0,686	0,859	1,063	1,323	1,721	2,080	2,518	2,831	3,819
22	0,686	0,858	1,061	1,321	1,717	2,074	2,508	2,819	3,792
23	0,685	0,858	1,060	1,319	1,714	2,069	2,500	2,807	3,768
24	0,685	0,857	1,059	1,318	1,711	2,064	2,492	2,797	3,745
25	0,684	0,856	1,058	1,316	1,708	2,060	2,485	2,787	3,725
26	0,684	0,856	1,058	1,315	1,706	2,056	2,479	2,779	3,707
27	0,684	0,855	1,057	1,314	1,703	2,052	2,473	2,771	3,689
28	0,683	0,855	1,056	1,313	1,701	2,048	2,467	2,763	3,674
29	0,683	0,854	1,055	1,311	1,699	2,045	2,462	2,756	3,660
30	0,683	0,854	1,055	1,310	1,697	2,042	2,457	2,750	3,646
40	0,681	0,851	1,050	1,303	1,684	2,021	2,423	2,704	3,551
60	0,679	0,848	1,045	1,296	1,671	2,000	2,390	2,660	3,460
120	0,677	0,845	1,041	1,289	1,658	1,980	2,358	2,617	3,373
∞	0,674	0,842	1,036	1,282	1,645	1,960	2,326	2,576	3,290

Anexo 2: Distribución X^2

TABLA 3: DISTRIBUCIÓN χ^2

Puntos de porcentaje de la distribución χ^2



Ejemplo:

Para $\phi = 10$ grados de libertad

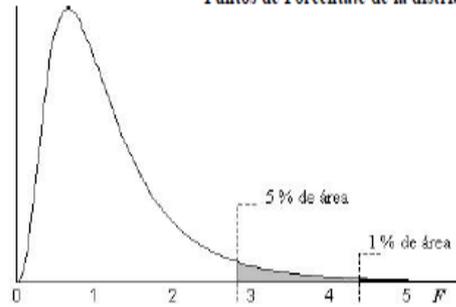
$$P[\chi^2 > 15.99] = 0.10$$

ϕ	0.995	0.99	0.975	0.95	0.9	0.75	0.5	0.25	0.1	0.05	0.025	0.01	0.005	Z_{α}
1	3.93E-05	1.57E-04	9.83E-04	3.93E-03	1.58E-02	0.102	0.455	1.323	2.71	3.84	5.02	6.63	7.88	1
2	1.00E-02	2.01E-02	5.06E-02	0.103	0.211	0.575	1.386	2.77	4.61	5.99	7.38	9.21	10.60	2
3	7.17E-02	0.115	0.216	0.352	0.584	1.213	2.37	4.11	6.25	7.81	9.35	11.34	12.84	3
4	0.207	0.297	0.484	0.711	1.064	1.923	3.36	5.39	7.78	9.49	11.14	13.28	14.86	4
5	0.412	0.554	0.831	1.145	1.610	2.67	4.35	6.63	9.24	11.07	12.83	15.09	16.75	5
6	0.676	0.872	1.237	1.635	2.20	3.45	5.35	7.84	10.64	12.59	14.45	16.81	18.55	6
7	0.989	1.239	1.690	2.17	2.83	4.25	6.35	9.04	12.02	14.07	16.01	18.48	20.3	7
8	1.344	1.647	2.18	2.73	3.49	5.07	7.34	10.22	13.36	15.51	17.53	20.1	22.0	8
9	1.735	2.09	2.70	3.33	4.17	5.90	8.34	11.39	14.68	16.92	19.02	21.7	23.6	9
10	2.16	2.56	3.25	3.94	4.87	6.74	9.34	12.55	15.99	18.31	20.5	23.2	25.2	10
11	2.60	3.05	3.82	4.57	5.58	7.58	10.34	13.70	17.28	19.68	21.9	24.7	26.8	11
12	3.07	3.57	4.40	5.23	6.30	8.44	11.34	14.85	18.55	21.0	23.3	26.2	28.3	12
13	3.57	4.11	5.01	5.89	7.04	9.30	12.34	15.98	19.81	22.4	24.7	27.7	29.8	13
14	4.07	4.66	5.63	6.57	7.79	10.17	13.34	17.12	21.1	23.7	26.1	29.1	31.3	14
15	4.60	5.23	6.26	7.26	8.55	11.04	14.34	18.25	22.3	25.0	27.5	30.6	32.8	15
16	5.14	5.81	6.91	7.96	9.31	11.91	15.34	19.37	23.5	26.3	28.8	32.0	34.3	16
17	5.70	6.41	7.56	8.67	10.09	12.79	16.34	20.5	24.8	27.6	30.2	33.4	35.7	17
18	6.26	7.01	8.23	9.39	10.86	13.68	17.34	21.6	26.0	28.9	31.5	34.8	37.2	18
19	6.84	7.63	8.91	10.12	11.65	14.56	18.34	22.7	27.2	30.1	32.9	36.2	38.6	19
20	7.43	8.26	9.59	10.85	12.44	15.45	19.34	23.8	28.4	31.4	34.2	37.6	40.0	20
21	8.03	8.90	10.28	11.59	13.24	16.34	20.3	24.9	29.6	32.7	35.5	38.9	41.4	21
22	8.64	9.54	10.98	12.34	14.04	17.24	21.3	26.0	30.8	33.9	36.8	40.3	42.8	22
23	9.26	10.20	11.69	13.09	14.85	18.14	22.3	27.1	32.0	35.2	38.1	41.6	44.2	23
24	9.89	10.86	12.40	13.85	15.66	19.04	23.3	28.2	33.2	36.4	39.4	43.0	45.6	24
25	10.52	11.52	13.12	14.61	16.47	19.94	24.3	29.3	34.4	37.7	40.6	44.3	46.9	25
26	11.16	12.20	13.84	15.38	17.29	20.8	25.3	30.4	35.6	38.9	41.9	45.6	48.3	26
27	11.81	12.88	14.57	16.15	18.11	21.7	26.3	31.5	36.7	40.1	43.2	47.0	49.6	27
28	12.46	13.56	15.31	16.93	18.94	22.7	27.3	32.6	37.9	41.3	44.5	48.3	51.0	28
29	13.12	14.26	16.05	17.71	19.77	23.6	28.3	33.7	39.1	42.6	45.7	49.6	52.3	29
30	13.79	14.95	16.79	18.49	20.6	24.5	29.3	34.8	40.3	43.8	47.0	50.9	53.7	30
40	20.7	22.2	24.4	26.5	29.1	33.7	39.3	45.6	51.8	55.8	59.3	63.7	66.8	40
50	28.0	29.7	32.4	34.8	37.7	42.9	49.3	56.3	63.2	67.5	71.4	76.2	79.5	50
60	35.5	37.5	40.5	43.2	46.5	52.3	59.3	67.0	74.4	79.1	83.3	88.4	92.0	60
70	43.3	45.4	48.8	51.7	55.3	61.7	69.3	77.6	85.5	90.5	95.0	100.4	104.2	70
80	51.2	53.5	57.2	60.4	64.3	71.1	79.3	88.1	96.6	101.9	106.6	112.3	116.3	80
90	59.2	61.8	65.6	69.1	73.3	80.6	89.3	98.6	107.6	113.1	118.1	124.1	128.3	90
100	67.3	70.1	74.2	77.9	82.4	90.1	99.3	109.1	118.5	124.3	129.6	135.8	140.2	100
Z_{α}	-2.58	-2.33	-1.96	-1.64	-1.28	-0.674	0.000	0.674	1.282	1.645	1.96	2.33	2.58	Z_{α}

Para $\phi > 100$ tómesese $\chi^2 = \frac{1}{2}(Z_{\alpha} + \sqrt{2\phi - 1})^2$. Z_{α} es la desviación normal estandarizada correspondiente al nivel de significancia y se muestra en la parte superior de la tabla.

TABLA 4: DISTRIBUCIÓN F DE FISHER

Puntos de Porcentaie de la distribución F



Ejemplo:

Para $n_1 = 9, n_2 = 12$ grados de libertad:
 $P[F > 2.80] = 0.05$
 $P[F > 4.39] = 0.01$

n_1		5% (normal) y 1% (negritas) puntos para la distribución de F																									n_2	
		n1 grados de libertad (para el mayor cuadrado medio)																										
		1	2	3	4	5	6	7	8	9	10	11	12	14	16	20	24	30	40	50	75	100	200	500	∞			
1	1	161	199	216	225	230	234	237	239	241	242	243	244	245	246	248	249	250	251	252	253	253	254	254	254	254	1	1
1	2	4052	4999	5404	5624	5764	5859	5928	5981	6022	6056	6083	6107	6143	6170	6209	6234	6260	6286	6302	6324	6334	6350	6360	6366	6366	2	2
1	3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.76	8.74	8.71	8.69	8.66	8.64	8.62	8.59	8.58	8.56	8.55	8.54	8.53	8.53	8.53	3	3
1	4	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.34	27.23	27.13	27.05	26.92	26.83	26.69	26.60	26.50	26.41	26.35	26.28	26.24	26.18	26.15	26.13	26.13	4	4
1	5	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.94	5.91	5.87	5.84	5.80	5.77	5.75	5.72	5.70	5.68	5.66	5.65	5.64	5.63	5.63	5	5
1	6	12.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66	14.55	14.45	14.37	14.25	14.15	14.02	13.93	13.84	13.75	13.69	13.61	13.58	13.52	13.49	13.46	13.46	6	6
1	7	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.70	4.68	4.64	4.60	4.56	4.53	4.50	4.46	4.44	4.42	4.41	4.39	4.37	4.37	4.37	7	7
1	8	15.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.54	6.47	6.36	6.28	6.16	6.07	5.99	5.91	5.86	5.79	5.75	5.70	5.67	5.65	5.65	8	8
1	9	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.94	2.91	2.86	2.83	2.77	2.74	2.70	2.66	2.64	2.60	2.59	2.56	2.55	2.54	2.54	9	9
1	10	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.77	4.71	4.60	4.52	4.41	4.33	4.25	4.17	4.12	4.05	4.01	3.96	3.93	3.91	3.91	10	10

n_1		5% (normal) y 1% (negritas) puntos para la distribución de F																									n_2	
		n1 grados de libertad (para el mayor cuadrado medio)																										
		1	2	3	4	5	6	7	8	9	10	11	12	14	16	20	24	30	40	50	75	100	200	500	∞			
11	11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.82	2.79	2.74	2.70	2.65	2.61	2.57	2.53	2.51	2.47	2.46	2.43	2.42	2.40	11	11	
11	12	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54	4.46	4.40	4.29	4.21	4.10	4.02	3.94	3.86	3.81	3.74	3.71	3.66	3.62	3.60	12	12	
11	13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.63	2.60	2.55	2.51	2.46	2.42	2.38	2.34	2.31	2.28	2.26	2.23	2.22	2.21	13	13	
11	14	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10	4.02	3.96	3.86	3.78	3.66	3.59	3.51	3.43	3.38	3.31	3.27	3.22	3.19	3.17	14	14	
11	15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.51	2.48	2.42	2.38	2.33	2.29	2.25	2.20	2.18	2.14	2.12	2.10	2.09	2.07	15	15	
11	16	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.73	3.67	3.56	3.49	3.37	3.29	3.21	3.13	3.08	3.01	2.98	2.92	2.89	2.87	16	16	
11	17	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.46	2.42	2.37	2.33	2.28	2.24	2.19	2.15	2.12	2.09	2.07	2.04	2.02	2.01	17	17	
11	18	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	3.62	3.55	3.45	3.37	3.26	3.18	3.10	3.02	2.97	2.90	2.86	2.81	2.78	2.75	18	18	
11	19	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.37	2.34	2.29	2.25	2.19	2.15	2.11	2.06	2.04	2.00	1.98	1.95	1.93	1.92	19	19	
11	20	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51	3.43	3.37	3.27	3.19	3.08	3.00	2.92	2.84	2.78	2.71	2.68	2.62	2.59	2.57	20	20	
11	21	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.46	2.42	2.38	2.34	2.31	2.26	2.21	2.16	2.11	2.07	2.03	2.00	1.96	1.94	1.91	1.89	1.88	21	21	
11	22	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43	3.36	3.30	3.19	3.12	3.00	2.92	2.84	2.76	2.71	2.64	2.60	2.55	2.51	2.49	22	22	
11	23	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.31	2.28	2.22	2.18	2.12	2.08	2.04	1.99	1.97	1.93	1.91	1.88	1.86	1.84	23	23	
11	24	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40	3.31	3.24	3.17	3.07	2.99	2.88	2.80	2.72	2.64	2.58	2.51	2.48	2.42	2.38	2.36	24	24	
11	25	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.26	2.23	2.17	2.13	2.07	2.03	1.98	1.94	1.91	1.87	1.85	1.82	1.80	1.78	25	25	
11	26	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26	3.18	3.12	3.02	2.94	2.83	2.75	2.67	2.58	2.53	2.46	2.42	2.36	2.33	2.31	26	26	
11	27	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.24	2.20	2.15	2.11	2.05	2.01	1.96	1.91	1.88	1.84	1.82	1.79	1.77	1.75	27	27	
11	28	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30	3.21	3.14	3.07	2.97	2.89	2.78	2.70	2.62	2.54	2.48	2.41	2.37	2.32	2.28	2.26	28	28	
11	29	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.22	2.16	2.13	2.09	2.03	1.98	1.94	1.89	1.86	1.82	1.80	1.77	1.75	1.73	29	29	
11	30	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17	3.09	3.03	2.93	2.85	2.74	2.66	2.58	2.49	2.44	2.37	2.33	2.27	2.24	2.21	30	30	
11	31	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.20	2.16	2.11	2.07	2.01	1.96	1.92	1.87	1.84	1.80	1.78	1.75	1.73	1.71	31	31	
11	32	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22	3.13	3.06	2.99	2.89	2.81	2.70	2.62	2.54	2.45	2.40	2.33	2.29	2.23	2.19	2.17	32	32	
11	33	4.23	3.37	2.97	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.18	2.15	2.09	2.05	1.99	1.95	1.90	1.85	1.82	1.78	1.76	1.73	1.71	1.69	33	33	
11	34	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18	3.09	3.02	2.96	2.86	2.78	2.66	2.58	2.50	2.42	2.36	2.29	2.25	2.19	2.16	2.13	34	34	
11	35	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20	2.17	2.13	2.08	2.04	1.97	1.93	1.88	1.84	1.81	1.76	1.74	1.71	1.69	1.67	35	35	
11	36	7.68	5.49	4.60	4.11	3.78	3.56	3.39	3.26	3.15	3.06	2.99	2.93	2.82	2.75	2.63	2.55	2.47	2.38	2.33	2.26	2.22	2.16	2.12	2.10	36	36	
11	37	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.15	2.12	2.06	2.02	1.96	1.91	1.87	1.82	1.79	1.75	1.73	1.69	1.67	1.65	37	37	
11	38	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12	3.03	2.96	2.90	2.79	2.72	2.60	2.52	2.44	2.35	2.30	2.23	2.19	2.13	2.09	2.06	38	38	
11	39	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18	2.14	2.10	2.05	2.01	1.94	1.90	1.85	1.81	1.77	1.73	1.71	1.67	1.65	1.64	39	39	
11	40	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.09	3.00	2.93	2.87	2.77	2.69	2.57	2.49	2.41	2.33	2.27	2.20	2.16	2.10	2.06	2.03	40	40	
11	41	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.13	2.09	2.04	1.99	1.93	1.89	1.84	1.79	1.76	1.72	1.70	1.66	1.64	1.62	41	41	
11	42	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.91	2.84	2.74	2.66	2.55	2.47	2.39	2.30	2.25	2.17	2.13	2.07	2.03	2.01	42	42	
11	43	4.15	3.29	2.90	2.67	2.51	2.40	2.31	2.24	2.19	2.14	2.10	2.07	2.01	1.97	1.91	1.86	1.82	1.77	1.74	1.69	1.67	1.63	1.61	1.59	43	43	
11	44	7.50	5.34	4.46	3.97	3.65	3.43	3.26	3.13	3.02	2.93	2.86	2.80	2.70	2.62	2.50	2.42	2.34	2.25	2.20	2.12	2.08	2.02	1.98	1.96	44	44	

		5% (normal) y 1% (negritas) puntos para la distribución de F																									
		n1 grados de libertad para el mayor cuadrado medio																									
D_1		1	2	3	4	5	6	7	8	9	10	11	12	15	20	24	30	40	50	75	100	200	500	∞	D_2		
34	4.13	3.28	2.88	2.65	2.49	2.38	2.29	2.23	2.17	2.12	2.08	2.05	1.99	1.95	1.89	1.84	1.80	1.75	1.71	1.67	1.65	1.61	1.59	1.57	34		
	7.44	5.29	4.42	3.93	3.61	3.39	3.22	3.09	2.98	2.89	2.82	2.76	2.66	2.58	2.46	2.38	2.30	2.21	2.16	2.08	2.04	1.96	1.94	1.91			
36	4.11	3.26	2.87	2.63	2.48	2.36	2.28	2.21	2.15	2.11	2.07	2.03	1.98	1.93	1.87	1.82	1.78	1.73	1.69	1.65	1.62	1.59	1.56	1.55	36		
	7.40	5.25	4.38	3.89	3.57	3.35	3.18	3.05	2.95	2.86	2.79	2.72	2.62	2.54	2.43	2.35	2.26	2.18	2.12	2.04	2.00	1.94	1.90	1.87			
38	4.10	3.24	2.85	2.62	2.46	2.35	2.26	2.19	2.14	2.09	2.05	2.02	1.96	1.92	1.85	1.81	1.76	1.71	1.68	1.63	1.61	1.57	1.54	1.53	38		
	7.35	5.21	4.34	3.86	3.54	3.32	3.15	3.02	2.92	2.83	2.75	2.68	2.59	2.51	2.40	2.32	2.23	2.14	2.09	2.01	1.97	1.90	1.86	1.84			
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.06	2.04	2.00	1.95	1.90	1.84	1.79	1.74	1.69	1.66	1.61	1.59	1.55	1.53	1.51	40		
	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80	2.73	2.66	2.56	2.48	2.37	2.29	2.20	2.11	2.06	1.98	1.94	1.87	1.83	1.81			
42	4.07	3.22	2.83	2.59	2.44	2.32	2.24	2.17	2.11	2.05	2.03	1.99	1.94	1.89	1.83	1.78	1.73	1.68	1.65	1.60	1.57	1.53	1.51	1.49	42		
	7.28	5.15	4.29	3.80	3.49	3.27	3.10	2.97	2.86	2.78	2.70	2.64	2.54	2.46	2.34	2.26	2.18	2.09	2.03	1.95	1.91	1.85	1.80	1.78			
44	4.05	3.21	2.82	2.58	2.43	2.31	2.23	2.16	2.10	2.05	2.01	1.98	1.92	1.88	1.81	1.77	1.72	1.67	1.63	1.59	1.56	1.52	1.49	1.48	44		
	7.25	5.12	4.26	3.78	3.47	3.24	3.08	2.95	2.84	2.75	2.68	2.62	2.52	2.44	2.32	2.24	2.15	2.07	2.01	1.93	1.89	1.82	1.78	1.75			
46	4.05	3.20	2.81	2.57	2.42	2.30	2.22	2.15	2.09	2.04	2.00	1.97	1.91	1.87	1.80	1.76	1.71	1.65	1.62	1.57	1.55	1.51	1.48	1.46	46		
	7.22	5.10	4.24	3.76	3.44	3.22	3.06	2.93	2.82	2.73	2.66	2.60	2.50	2.42	2.30	2.22	2.13	2.04	1.99	1.91	1.86	1.80	1.76	1.73			
48	4.04	3.19	2.80	2.57	2.41	2.29	2.21	2.14	2.08	2.03	1.99	1.95	1.90	1.86	1.79	1.75	1.70	1.64	1.61	1.56	1.54	1.49	1.47	1.45	48		
	7.19	5.08	4.22	3.74	3.43	3.20	3.04	2.91	2.80	2.71	2.64	2.58	2.48	2.40	2.28	2.20	2.12	2.02	1.97	1.89	1.84	1.78	1.73	1.70			
50	4.03	3.18	2.79	2.56	2.40	2.29	2.20	2.13	2.07	2.03	1.99	1.95	1.89	1.85	1.78	1.74	1.69	1.63	1.60	1.55	1.52	1.48	1.46	1.44	50		
	7.17	5.06	4.20	3.72	3.41	3.19	3.02	2.89	2.78	2.70	2.63	2.56	2.46	2.38	2.27	2.18	2.10	2.01	1.95	1.87	1.82	1.76	1.71	1.68			
55	4.02	3.16	2.77	2.54	2.38	2.27	2.18	2.11	2.06	2.01	1.97	1.93	1.88	1.83	1.76	1.72	1.67	1.61	1.58	1.53	1.50	1.46	1.43	1.41	55		
	7.12	5.01	4.16	3.68	3.37	3.15	2.98	2.85	2.75	2.66	2.59	2.53	2.42	2.34	2.23	2.15	2.06	1.97	1.91	1.83	1.78	1.71	1.67	1.64			
60	4.00	3.15	2.75	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.95	1.92	1.86	1.82	1.75	1.70	1.65	1.59	1.56	1.51	1.48	1.44	1.41	1.39	60		
	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.56	2.50	2.39	2.31	2.20	2.12	2.03	1.94	1.88	1.79	1.75	1.68	1.63	1.60			
65	3.99	3.14	2.75	2.51	2.36	2.24	2.15	2.08	2.03	1.98	1.94	1.90	1.85	1.80	1.73	1.69	1.63	1.58	1.54	1.49	1.46	1.42	1.39	1.37	65		
	7.04	4.95	4.10	3.62	3.31	3.09	2.93	2.80	2.69	2.61	2.53	2.47	2.37	2.29	2.17	2.09	2.00	1.91	1.85	1.77	1.72	1.65	1.60	1.57			
70	3.98	3.13	2.74	2.50	2.35	2.23	2.14	2.07	2.02	1.97	1.93	1.89	1.84	1.79	1.72	1.67	1.62	1.57	1.53	1.48	1.45	1.40	1.37	1.35	70		
	7.01	4.92	4.07	3.60	3.29	3.07	2.91	2.78	2.67	2.59	2.51	2.45	2.35	2.27	2.15	2.07	1.98	1.89	1.83	1.74	1.70	1.62	1.57	1.54			
80	3.96	3.11	2.72	2.49	2.33	2.21	2.13	2.06	2.00	1.95	1.91	1.88	1.82	1.77	1.70	1.65	1.60	1.54	1.51	1.45	1.43	1.38	1.35	1.33	80		
	6.96	4.88	4.04	3.56	3.25	3.04	2.87	2.74	2.64	2.55	2.48	2.42	2.31	2.23	2.12	2.03	1.94	1.85	1.79	1.70	1.65	1.58	1.53	1.50			
100	3.94	3.09	2.70	2.46	2.31	2.19	2.10	2.03	1.97	1.93	1.89	1.85	1.79	1.75	1.68	1.63	1.57	1.52	1.46	1.42	1.39	1.34	1.31	1.29	100		
	6.90	4.82	3.98	3.51	3.21	2.99	2.82	2.69	2.59	2.50	2.43	2.37	2.27	2.19	2.07	1.98	1.89	1.80	1.74	1.65	1.60	1.52	1.47	1.43			
125	3.92	3.07	2.68	2.44	2.29	2.17	2.08	2.01	1.95	1.91	1.87	1.83	1.77	1.73	1.66	1.60	1.55	1.49	1.45	1.40	1.36	1.31	1.27	1.25	125		
	6.84	4.78	3.94	3.47	3.17	2.95	2.79	2.66	2.55	2.47	2.39	2.33	2.23	2.15	2.03	1.94	1.85	1.76	1.69	1.60	1.55	1.47	1.41	1.37			
150	3.90	3.06	2.66	2.43	2.27	2.16	2.07	2.00	1.94	1.89	1.85	1.82	1.76	1.71	1.64	1.59	1.54	1.48	1.44	1.38	1.34	1.29	1.25	1.22	150		
	6.81	4.75	3.91	3.45	3.14	2.92	2.76	2.63	2.53	2.44	2.37	2.31	2.20	2.12	2.00	1.92	1.83	1.73	1.66	1.57	1.52	1.43	1.38	1.33			
200	3.89	3.04	2.65	2.42	2.26	2.14	2.06	1.99	1.93	1.88	1.84	1.80	1.74	1.69	1.62	1.57	1.52	1.46	1.41	1.35	1.32	1.26	1.22	1.19	200		
	6.76	4.71	3.88	3.41	3.11	2.89	2.73	2.60	2.50	2.41	2.34	2.27	2.17	2.09	1.97	1.89	1.79	1.69	1.63	1.53	1.48	1.39	1.33	1.28			
400	3.86	3.02	2.63	2.39	2.24	2.12	2.03	1.96	1.90	1.85	1.81	1.78	1.72	1.67	1.60	1.54	1.49	1.42	1.38	1.32	1.28	1.22	1.17	1.13	400		
	6.70	4.66	3.83	3.37	3.06	2.85	2.68	2.56	2.45	2.37	2.29	2.23	2.13	2.05	1.92	1.84	1.75	1.64	1.58	1.48	1.42	1.32	1.25	1.19			
1000	3.85	3.00	2.61	2.38	2.22	2.11	2.02	1.95	1.89	1.84	1.80	1.76	1.70	1.65	1.58	1.53	1.47	1.41	1.36	1.30	1.26	1.19	1.13	1.08	1000		
	6.66	4.63	3.80	3.34	3.04	2.82	2.66	2.53	2.43	2.34	2.27	2.20	2.10	2.02	1.90	1.81	1.72	1.61	1.54	1.44	1.38	1.28	1.19	1.12			
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.79	1.75	1.69	1.64	1.57	1.52	1.46	1.39	1.35	1.28	1.24	1.17	1.11	1.00	∞		
	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32	2.25	2.18	2.08	2.00	1.88	1.79	1.70	1.59	1.52	1.42	1.36	1.25	1.15	1.00			