



---

# **Universidad de Valladolid**

**Facultad de Ciencias  
Económicas y Empresariales**

**Grado en A.D.E.**

## **Modelo econométrico de demanda turística para España**

Presentado por:

***Alfredo de Paz Sastre***

*Valladolid, 30 de Noviembre de 2015*

## INDICE

<b>1. INTRODUCCIÓN</b> .....	3
<b>2. ESTUDIOS CUANTITATIVOS SOBRE LA DEMANDA TURÍSTICA EN ESPAÑA</b> .....	5
<b>3. ELABORACIÓN DEL MODELO ECONOMÉTRICO</b> .....	9
<b>3.1. DESCRIPCIÓN DE LOS DATOS</b> .....	9
3.1.1. Fuentes de datos.....	9
3.1.2. Análisis descriptivo de las variables consideradas .....	10
<b>3.2. METODOLOGÍA</b> .....	12
3.2.1. Estadísticos descriptivos.....	12
3.2.2. Análisis de correlación.....	14
<b>3.3. MODELOS CONSIDERADOS</b> .....	16
3.3.1. Elección del modelo de partida.....	17
3.3.2. Análisis de los posibles problemas del modelo elegido .....	20
3.3.3. Solución a los problemas existentes.....	24
3.3.4. Interpretación de los resultados en el modelo final.....	27
<b>4. CONCLUSIONES</b> .....	28
<b>5. BIBLIOGRAFÍA</b> .....	29

# 1. INTRODUCCIÓN

Según la Organización Mundial del Turismo (OMT) el turismo es “un fenómeno social, cultural y económico relacionado con el movimiento de las personas a lugares que se encuentran fuera de su lugar de residencia habitual por motivos personales o de negocios/profesionales. Estas personas se denominan visitantes (que pueden ser turistas o excursionistas; residentes o no residentes) y el turismo tiene que ver con sus actividades, de las cuales algunas implican un gasto turístico”.

A la vista de los datos actuales del Ministerio de Industria, Energía y Turismo, el turismo representa en España más del 11 % del PIB y se trata de uno de los sectores más dinámicos de la economía, ya que a su vez representa el 12 % del empleo total en España, pero no sólo es importante por el valor añadido y el empleo que genera, sino por el dinamismo y efectos de arrastre que genera en otros sectores. La llegada de turistas está alcanzando unas cifras de récord, posicionando a España como el tercer destino turístico mundial en número de turistas y el segundo en nivel de ingresos por turismo.

El sector turístico ha sido clave en la recuperación de la economía. En el pasado año 2014 el sector creció más del doble que el resto de la economía (crecimiento del 2,9% del PIB turístico frente al 1,4 % del conjunto de la economía española) y éste año 2015 crecerá también a mayor ritmo, en concreto un 3,4 % según las últimas proyecciones de la Alianza para la Excelencia Turística (Exceltur).

Además, según el Informe de Competitividad en Viajes y Turismo 2015, España tiene por primera vez el sector turístico más competitivo del mundo, gracias, entre otros factores, a nuestras infraestructuras, recursos culturales, la adaptación a las tendencias del nuevo consumidor digital y la importancia de la política turística dentro de las políticas públicas.

En relación con la definición de turismo, se entiende como demanda turística el número de personas que viajan (demanda real) o que desean viajar (demanda potencial), y que consumen productos o servicios turísticos en lugares distintos a los de trabajo y residencia habituales.

El objetivo principal de este trabajo es analizar los factores que influyen en la demanda turística en España. Para ello elaboraremos un modelo econométrico que reúna las variables de mayor relevancia para el análisis, cuya regresión nos de un buen ajuste y los estimadores con las mejores propiedades.

El trabajo se estructura en tres partes. En la primera vamos a realizar una revisión de los estudios de estimación y previsión de la demanda turística elaborados en España. En estos estudios veremos las distintas técnicas aplicadas, variables usadas, previsiones (evolución de la demanda turística), factores de influencia, tipos de datos, tipo de turismo, predicciones y objetivos fijados por los autores, de modo que se puede observar como se ha ido modelizando hasta ahora la demanda turística.

La segunda parte del trabajo es la más importante, ya que en ella realizaremos nuestro modelo econométrico. En primer lugar haremos una descripción de los distintos tipos de datos que vamos a usar así como de las variables dependientes y explicativas de nuestro interés. Seguidamente, antes de comenzar con las pruebas para la elección del modelo realizamos un análisis de correlación, y mediante los resultados del coeficiente de correlación simple vemos las variables que pueden ser más relevantes en las regresiones.

Para llegar al modelo econométrico óptimo vamos a realizar varias pruebas de modelos hasta conseguir el que tenga mayor  $\bar{R}^2$ . Una vez elegido el modelo analizamos los posibles problemas (multicolinealidad, normalidad, linealidad, autocorrelación y heteroscedasticidad) y las soluciones a estos.

Por último, vemos los resultados obtenidos en el modelo econométrico finalmente elegido (corregidos los problemas).

La tercera parte contiene las principales conclusiones obtenidas.

## 2. ESTUDIOS CUANTITATIVOS SOBRE LA DEMANDA TURÍSTICA EN ESPAÑA

La mayoría de los estudios sobre la demanda turística elaborados en España se han realizado en los últimos treinta años. Básicamente, según Sánchez, M. (1998), podrían agruparse en tres bloques:

- Estudios de estimación y previsión de la demanda turística.
- Estudios acerca del grado de satisfacción del turista.
- Estudios del comportamiento de los españoles ante las vacaciones.

Si nos centramos en los primeros, que son los que nos interesan para nuestro trabajo, tanto en la tesis doctoral de *Sánchez, M. (1998)* como en *Esteban, A. y Reinares, E (1996)* se citan y analizan los estudios de previsión de la demanda turística (realizados con el uso de técnicas econométricas) más relevantes realizados hasta esas fechas.

El primer trabajo en el que se realizó un análisis econométrico de la demanda turística española que tuvo gran influencia en las investigaciones futuras fue el realizado por *Pulido San Román (1966)* "*Introducción a un análisis econométrico del turismo*".

El trabajo comienza realizando un planteamiento económico del fenómeno del turismo, para después, centrándose en la demanda de turismo de la unidad familiar y habiendo analizado las variables explicativas más relevantes, proponer algunos modelos econométricos de predicción de la demanda para el periodo 1965-1966, en las áreas de emisión y recepción de turistas. Este autor se encontró con el problema de disponer de información estadística limitada a la hora de analizar el consumo turístico.

Su principal aportación fue un cuadro completo de movimientos turísticos entre una serie de países de origen y destino, con la evolución de las cuotas de participación y la distribución del mercado turístico internacional.

Destacan también los estudios realizados por distintas instituciones especializadas en el turismo como el *Instituto de Estudios turísticos, la*

*Secretaría General de Turismo o la Dirección General de Política turística (años 80 y 90).*

Por un lado, lo que suelen hacer es un análisis de la evolución de los modelos existentes de demanda turística para observar las relaciones estructurales y causales entre las variables (a través de sus elasticidades) que forman parte de los procesos relativos al fenómeno turístico, buscar los motivos de los desajustes en las tendencias de esos modelos respecto a la realidad, revisar los factores de influencia, etc.

Por otra parte, se dedican también a la elaboración de modelos econométricos para conocer la estructura de comportamiento de la demanda turística. Estos modelos se elaboran con base en las series de evolución de las variables turísticas en España durante un periodo (mes, año), y con ellos se realizan previsiones de evolución de las series para el año/s siguientes. Entre las variables de interés destacan las siguientes:

- Volumen de visitantes entrados en España (flujos turísticos), sus nacionalidades y su gasto medio (relacionado con la renta)
- Medios de transporte utilizados
- Ingresos por turismo
- Índice de precios turísticos (estimaciones, precios relativos de países).

No podemos dejar de mencionar los estudios realizados por una de las investigadoras que ha abordado con más profundidad esta temática, *Águeda Esteban Talaya*.

Dentro de ellos, destaca su tesis doctoral *“Análisis de la demanda: aplicación a la actividad turística de las técnicas de predicción” (1987)*.

En este trabajo se estudian los factores determinantes del comportamiento de la demanda turística (socioeconómicos, demográficos, psicográficos y técnicos, además de los turísticos) y se describen las técnicas de predicción (subjettivas, causales y de series temporales) aplicadas en las investigaciones sobre la demanda turística para conocer las principales estimaciones efectuadas, los modelos, las variables (turísticas y explicativas) y los resultados.

La principal aportación de este trabajo es la estimación de modelos econométricos de la demanda turística para generar predicciones propias diferenciadas por el ámbito temporal:

- largo plazo: panel de expertos con el horizonte del año 2000
- medio plazo: modelos econométricos entre 1986 y 1990
- corto plazo: técnicas Box-Jenkins para un año desagregado mensualmente

Se proponen modelos para las siguientes variables:

- Visitantes extranjeros entrados por fronteras (terrestre, marítima, aérea)
- Pernoctaciones totales en establecimientos
- Ingresos por turismo

También realizó años más tarde un estudio en la misma línea: modelos de evolución temporal de la demanda turística española pero con una segmentación por países de origen. El objetivo es apreciar diferencias en la evolución de los visitantes extranjeros entrados a España por fronteras (variable dependiente), en función de las variables económicas de influencia, que en este caso son la renta y los precios (variables explicativas). Lo relevante es que realiza un análisis segmentado con los países de procedencia de los visitantes, formando 2 grupos según las cuotas de mercado considerando los visitantes entrados por fronteras. Este estudio se puede realizar de 2 formas:

- Análisis de las elasticidades de demanda, mediante los que se observan los resultados de la sensibilidad de la demanda frente a las variaciones de los factores de influencia.
- Construcción de modelos econométricos, los cuales incluyen la renta con el PIB del país y los precios con el IPC en varias formas, tanto para representar la evolución total de la demanda como para diferenciar los segmentos de procedencia de la demanda turística hacia España.

Por último mencionamos estudios más recientes (años 2000) realizados por las autoras *Guisán, Aguayo y Neira*, las cuáles abordaron sobre todo el tema del turismo hotelero y extrahotelero en las regiones y provincias españolas.

Así, en *Guisán y Aguayo (2008)* se estiman varios modelos econométricos con una combinación de series temporales y atemporales en los años 1990-95 para explicar la distribución regional y provincial del turismo hotelero en España. En cuanto a las variables del modelo:

La variable endógena (Y) es la proporción de pernoctaciones hoteleras anuales de la región en el total nacional, y las variables explicativas son:

- Proporción del número de plazas hoteleras de la región en el total nacional
- Proporción de la media ponderada del precio por plaza de cada región respecto a la media nacional

Algunas técnicas usadas son introducir las variables en forma logarítmica (el modelo incluye el efecto de los precios) o incluir las variables explicativas con retardos (endógena) e incrementos (exógenas), es decir, un modelo dinámico mixto. Una vez analizado los resultados del modelo (datos de coeficientes de los estimadores, bondad del ajuste, coeficiente de determinación,...) se puede observar la importancia de varios factores como el clima, costes del transporte o la proximidad a grandes núcleos de población. Otras variables significativas para el turismo hotelero son el precio y el número de plazas hoteleras (esta última tiene una relación dinámica con el número de pernoctaciones, ya que el incremento en el número de plazas depende del incremento retardado de las pernoctaciones).

También tiene importancia para estas autoras otro tipo de turismo menos estudiado, el turismo extrahotelero en España. Lo analizaron en *Guisán y Neira (2001)* y observaron que en su conjunto tiene un peso económico tan importante o más que el hotelero por su impacto sobre la producción y el empleo (sectores como la construcción o el comercio).

En este caso para comparar el turismo extrahotelero entre distintas regiones el indicador que se usa es el número total de viviendas secundarias, y para ver la intensidad del turismo extrahotelero en relación con la población utilizaron las viviendas secundarias por habitante. Presentan modelos econométricos explicativos de la distribución regional del número de viviendas secundarias (variable endógena), incluyendo unas variables ficticias climatológicas (según



el tipo de clima) y de proximidad (distancia respecto a grandes núcleos de población) y una variable relativa al Valor añadido. Como resultado, según sean significativos los valores de los coeficientes vemos el impacto que tienen sobre el turismo extrahotelero factores o variables como la renta, población, clima y proximidad.

Además de estos estudios del turismo hotelero y extrahotelero, también citar que realizaron otros modelos econométricos del estilo basados en el impacto del turismo sobre el desarrollo y empleo regional, junto con factores relativos a la inversión industrial y el capital humano. Estos modelos ponen de manifiesto el impacto positivo que el incremento del Valor Añadido real por habitante tiene sobre las tasas de empleo en todos los sectores (*Guisán, Aguayo y Carballas (2004) y Guisán y Aguayo (2005) y (2007)*).

### **3. ELABORACIÓN DEL MODELO ECONOMÉTRICO**

Una vez vistos los trabajos y estudios de los autores sobre la materia en cuestión, voy a comenzar a realizar mi modelo econométrico de demanda turística.

#### **3.1. DESCRIPCIÓN DE LOS DATOS**

##### **3.1.1. Fuentes de datos**

Las fuentes de datos que he usado han sido el INE (para la mayoría de los datos) y el Instituto de Turismo de España a modo complementario. Ambos organismos elaboran estadísticas que permiten conocer los flujos turísticos registrados durante un período de tiempo determinado en una zona concreta.

Están disponibles datos de corte transversal (nivel nacional, CCAA, provincias) y temporales (trimestre, cuatrimestre, mes, año). Para mi trabajo los datos que utilizaré son datos de corte transversal referidos a las 19 comunidades autónomas para el año 2014.

### 3.1.2. Análisis descriptivo de las variables consideradas

De la multitud de variables que ofrecen he seleccionado las que he creído más acordes a la demanda turística, y son las siguientes:

#### VARIABLE DEPENDIENTE

-Los viajeros son las personas que realizan una o más pernoctaciones (cada noche que un viajero se aloja en un establecimiento) seguidas en el mismo alojamiento. Además cabe citar que todos los tipos de viajeros relacionados con el turismo se denominan visitantes, ya sean turistas (visitantes que pernoctan) o excursionistas (visitantes del día).

Lo mediremos en número total de viajeros según la Encuesta de Ocupación Hotelera (EOH).

#### VARIABLES EXPLICATIVAS

-Índice de precios de consumo (IPC): Su objetivo es medir la evolución del nivel de precios de los bienes y servicios de consumo adquiridos por los hogares residentes en España.

Si el IPC aumenta en una CCAA o provincia, el poder adquisitivo del dinero (por el aumento de la inflación) en ella será menor y el “coste de vida” mayor, lo cuál puede perjudicar a la llegada de turistas eligiendo éstos destinos vacacionales alternativos.

La mediremos en porcentaje con base 2011.

-Índice de precios hoteleros (IPH): Este índice mide la evolución del conjunto de precios aplicados por los empresarios a los distintos clientes que se alojan en los hoteles de España.

Debido a ello si se incrementan los precios de la oferta hotelera de una zona, será perjudicial para la llegada de visitantes.

La mediremos en porcentaje con base 2008.

-Producto interior bruto (PIB): Es la suma del valor de todos los bienes y servicios finales producidos en un país/región/provincia en un año.

Este indicador sirve para comparar la riqueza generada o el nivel de crecimiento de una economía, de modo que el incremento del PIB en una zona puede generar un efecto atracción por el bienestar económico.

Utilizaremos el PIB a precios de mercado (precios corrientes), en Miles de €.

-Población: Se incluye la medida de la población total residente en CCAA/provincia ya que nos interesa ver si las zonas con gran cantidad de población atraen un mayor número de visitantes que las de menos habitantes.

La mediremos en número de habitantes.

-Tasa de paro = parados / activos, expresando el nivel de desempleo entre la población económicamente activa (no tiene empleo y lo está buscando).

-Tasa actividad = activos / población en edad de trabajar, siendo a su vez la población activa la suma de los ocupados (empleados) y los parados (desempleados).

Estas variables pueden influir de modo que si en una zona la tasa de paro es elevada o la tasa de actividad reducida, podría perjudicar a sus habitantes a la hora de relizar viajes turísticos.

Ambas variables se miden en tasas (cocientes en %).

-Un establecimiento hotelero es aquel que presta servicios de alojamiento colectivo a personas mediante precio de forma habitual y profesional, con o sin otros servicios complementarios, mediante la denominación de hotel, hostel, apartahotel, motel, pensión o similares. A efectos de la Encuesta de Ocupación Hotelera no formarían parte de este grupo el resto de establecimientos de alojamiento turístico sujetos a una normativa específica, como viviendas turísticas vacacionales, agroturismos, campings, etc.

-Las plazas hoteleras es el espacio que ocupa una persona en un establecimiento hotelero y equivale al número de camas fijas (las de matrimonio dan lugar a 2 plazas, y no se incluyen las supletorias).

-El personal ocupado es el conjunto de personas, fijas o eventuales, que se encuentran ejerciendo una labor (remunerada o no) para el establecimiento. Si la oferta hotelera (nº establecimientos y/o plazas) es mayor en una zona podría influir en la llegada de un mayor número de viajeros.

Las mediremos en número de establecimientos abiertos y plazas estimadas según la EOH (media anual), y el personal ocupado en número de personas.

## 3.2. METODOLOGÍA

### 3.2.1. Estadísticos descriptivos

Distribución de la variable dependiente N° VIAJEROS en las diferentes CCAA

CCAA	Población	VIAJEROS (Y)	% en el total CCAA
Andalucía	8.392.537	15.569.414	17,77%
Aragón	1.328.759	2.227.147	2,54%
Asturias	1.054.408	1.500.982	1,71%
Balears (Illes)	1.121.739	8.446.299	9,64%
Canarias	2.120.534	8.473.585	9,67%
Cantabria	586.395	1.037.949	1,18%
Castilla y León	2.484.932	4.169.642	4,76%
Castilla-La Mancha	2.067.191	1.903.896	2,17%
Cataluña	7.400.684	16.754.341	19,13%
Comunitat Valenciana	4.949.219	7.284.705	8,32%
Extremadura	1.093.587	1.201.656	1,37%

<b>Galicia</b>	2.739.229	3.615.588	4,13%
<b>Madrid</b>	6.376.610	10.265.196	11,72%
<b>Murcia</b>	1.463.797	1.111.188	1,27%
<b>Navarra</b>	636.024	901.160	1,03%
<b>País Vasco</b>	2.166.043	2.469.884	2,82%
<b>La Rioja</b>	314.062	542.395	0,62%
<b>Ceuta</b>	84.618	65.994	0,08%
<b>Melilla</b>	83.685	58.018	0,07%
<b>Total Nacional</b>	46.464.053	87.599.030	

*Fuente: Elaboración propia (datos del INE a 2014)*

Como podemos observar en la tabla las CCAA que reciben un mayor nº de viajeros son Cataluña con 16.754.341 y un 19,3% en el total regional, Andalucía con 15.569.414 y un 17,77% y Madrid con 10.265.196 y un 11,72%.

Cabe destacar que en relación a la población, las regiones con mayor nº de visitantes serían las Islas (Canarias tiene aproximadamente 2 millones de habitantes y reciben alrededor de 8,5 millones de viajeros; Baleares tiene aproximadamente 1 millón de habitantes y también recibe alrededor de 8,5 millones). También la Comunidad Valenciana sería relevante en cuanto a población y viajeros.

Por otro lado entre las CCAA que reciben un menor nº de viajeros, sin contar con Ceuta y Melilla, destacan las uniprovinciales como La Rioja con 542.395 y un 0,62% del total, Navarra con 901.160 y un 1,03% y Cantabria con 1.037.949 y un 1,18%.

### 3.2.2. Análisis de correlación

Una herramienta para la selección de las variables a incluir en el modelo es la matriz de correlaciones entre las variables, analizando los coeficientes de correlación simple.

Mediante el análisis de los coeficientes de correlación simple entre la variable endógena y cada una de las exógenas podemos ver si están correlacionadas, es decir, si a priori tienen una relación. Si estas correlaciones son altas ( $>0.85$ ) supone una cierta garantía de acierto en la inclusión de la variable en el modelo para su posterior regresión múltiple. En cambio, un bajo coeficiente de correlación indica que esa variable es posible que no sea importante al indicar una ausencia de causalidad.

A pesar de lo dicho, un elevado coeficiente de correlación no implica que la relación sea segura, ni que una variable no sea relevante a pesar de ausencia de correlación ya que esa variable explicativa puede tener importancia en la regresión múltiple por la relación con el conjunto de variables.

También nos permite ver si las variables explicativas están correlacionadas entre ellas, de modo que un coeficiente de correlación elevado entre 2 o más variables nos avisan de un posible problema de multicolinealidad entre ellas en el posterior análisis de regresión múltiple. Esto se debe a que si están muy relacionadas entre sí estarán explicando lo mismo del comportamiento de la endógena y al incluirlas en el modelo juntas, su aportación a la explicación de la endógena se compensará.

#### MATRIZ DE CORRELACIONES

He realizado mediante Excel la matriz de correlaciones de Y (nº de viajeros) con todas las X consideradas obteniendo los siguientes resultados:

	<i>viajeros (Y)</i>	<i>IPC</i>	<i>IPH</i>	<i>IPC h,c,r.</i>	<i>Población</i>	<i>PIB</i>	<i>T. actividad</i>	<i>T. paro</i>	<i>establec.</i>	<i>nº plazas</i>	<i>personal</i>
viajeros (Y)	1										
IPC	0,2799	1,0000									
IPH	0,1836	-0,2851	1,0000								
IPC h,c,r.	0,0561	0,6550	0,0518	1,0000							
Población	0,9019	0,2697	-0,0830	-0,0444	1,0000						
PIB	0,8612	0,3359	-0,1452	0,0425	0,9473	1,0000					
T. actividad	0,4987	0,1903	0,1189	0,2838	0,3169	0,4089	1,0000				
T. paro	0,1428	-0,6170	0,4118	-0,3466	0,1401	-0,0629	-0,0585	1,0000			
establec.	0,8350	0,3986	-0,0754	0,0799	0,8831	0,7882	0,1058	0,0736	1,0000		
nº plazas	0,9374	0,1659	0,3851	0,0632	0,7370	0,6562	0,5072	0,2645	0,7110	1,0000	
personal	0,8661	0,0794	0,4616	0,0539	0,6250	0,5569	0,5192	0,2916	0,5819	0,9797	1

Como podemos observar las variables mas relacionadas con el número de viajeros (Y) son la Población, el PIB, el nº de establecimientos, el nº de plazas y el personal.

#### -Análisis de correlación por grupos de variables

- Variables que recogen la oferta hotelera (establecimientos, plazas, personal): PLAZAS y PERSONAL están muy correlacionadas (0,979) y entre las dos elijo PLAZAS al tener a su vez una correlación alta con la Y (0,937). En teoría además parece más lógico considerar el nº de plazas hoteleras que el personal empleado. Entre PLAZAS y ESTABLECIMIENTOS la correlación es de 0,71 y, aunque establecimientos tiene una correlación con la Y algo más baja que plazas (0,83), probaré modelos de regresión con una y con otra.
- Variables que recogen el nivel de precios (IPC, IPH, IPC (hoteles, cafés, restaurantes)): En este grupo de variables a diferencia del anterior ninguna de las 3 tiene una correlación alta con la Y, pero al tener el IPC de hotel, café, restaurante una correlación casi nula (0,05) con la Y, no la usaré en las regresiones. Entre IPC y IPH no existe apenas correlación (-0,28) con lo que probaré las dos aunque con indicios de que no serán relevantes para el modelo debido a sus bajas correlaciones con la Y.

- VARIABLES QUE RECOGEN EL TAMAÑO (ECONÓMICO O DE HABITANTES) DE LA CCAA (PIB y Población): Ambas variables tienen correlaciones altas con la Y (0,86 y 0,90 respectivamente), pero a su vez están muy correlacionadas entre ellas (0,947). Como ambas tienen correlaciones con la Y similares, probaré modelos de regresión con una y con otra.
- VARIABLES QUE RECOGEN EL NIVEL DE ACTIVIDAD (T.actividad y T.paro): Entre ellas no existe apenas correlación (-0,05), pero al expresar ambas lo mismo (empleo) elegiremos una de las dos. Ninguna tiene correlación alta con la Y pero como la de la T.paro es casi nula (0,14) no la usaré en los modelos de regresión.

En principio por tanto, no usaré las variables PERSONAL, IPC hotel, café, restaurante y T.PARO en las regresiones.

### 3.3. MODELOS CONSIDERADOS

Una vez realizado el análisis de correlación he seleccionado 4 modelos distintos para hacer el análisis de regresión múltiple, según incluya las variables PLAZAS/ESTABLECIMIENTOS y POBLACIÓN/PIB.

MODELO 1: Viajeros → Plazas - T.Actividad - Población - IPC - IPH

MODELO 2: Viajeros → Plazas - T.Actividad - PIB - IPC - IPH

MODELO 3: Viajeros → Establecimientos - T.Actividad - Población - IPC - IPH

MODELO 4: Viajeros → Establecimientos - T.Actividad - PIB - IPC - IPH

Mediante las regresiones iré comprobando la significación de las variables en cada modelo, comenzando con las 5 variables explicativas y quitando o añadiendo variables. Pero la referencia para elegir el mejor modelo más que la significación es el  $R^2$  ajustado, de modo que aunque una variable sea no significativa individualmente (mirando el estadístico "t" y su p-valor) si al quitarla del modelo el  $R^2$  ajustado se reduce habría que dejarla (estaríamos cometiendo



sino un error por omisión de variable relevante), pero si el  $R^2$  ajustado aumenta o se queda igual se debe quitar esa variable (estaríamos cometiendo sino un error por inclusión de variable irrelevante).

En el caso de que dos o más modelos tuvieran el mismo  $R^2$  ajustado cualquiera sería válido, pero para elegir uno nos interesa la simplicidad y por ello entre 2 modelos con igual  $R^2$  ajustado elegiríamos el más sencillo, es decir, el que tiene menos regresores.

### 3.3.1. Elección del modelo de partida

MODELO 1:

Viajeros  $\rightarrow$  Plazas - T.Actividad - Población - IPC – IPH

La regresión múltiple con las 5 variables explicativas nos da como resultado un  $\bar{R}^2 = 0.97546$  y son significativas individualmente las variables PLAZAS y POBLACION, además de ser conjuntamente significativas.

Las pruebas de regresión con 4 variables (eliminando del modelo una distinta cada vez) nos dan en general  $\bar{R}^2$  más bajos. Sólo mejora el modelo quitando la variable IPH al aumentar el  $\bar{R}^2 = 0.97687$ , por lo tanto estábamos cometiendo un error por inclusión de variable irrelevante (IPH).

En las pruebas con 3 variables (eliminando de 2 en 2) quitando junto al IPH otra (T.ACTIVIDAD o IPC, ya que vimos que si quitamos PLAZAS o POBLACION el modelo empeora) los resultados no mejoran al anterior modelo, sólo se acerca al eliminar las variables T.ACTIVIDAD e IPH con un  $\bar{R}^2 = 0.97548$ .

En conclusión, en el modelo 1 la mejor regresión sería con las variables PLAZAS, POBLACIÓN e IPC.

MODELO 2:

Viajeros  $\rightarrow$  Plazas - T.Actividad - PIB - IPC – IPH

En este modelo se mejoran los resultados del modelo 1 en el análisis de regresión múltiple con 5 variables, con un  $\bar{R}^2=0.98060$  siendo significativas individualmente las variables PLAZAS y ahora el PIB, además de ser conjuntamente significativas.

En las pruebas con 4 variables se obtienen peores resultados al eliminar las variables significativas, pero mejoran al quitar cualquiera de las otras 3 (T.ACTIVIDAD, IPC o IPH):

- a) PLAZAS, PIB, T.ACTIVIDAD, IPC  $\rightarrow \bar{R}^2= 0.981993$
- b) PLAZAS, PIB, T.ACTIVIDAD, IPH  $\rightarrow \bar{R}^2= 0.980974$
- c) PLAZAS, PIB, IPC, IPH  $\rightarrow \bar{R}^2= 0.981743$

Como vemos el mejor modelo es el “a” eliminando la variable IPH (estábamos cometiendo error por inclusión de variable irrelevante, como en el modelo 1). Las variables plazas y PIB siguen siendo significativas y tanto tasa de actividad como IPC son no significativas.

En cuanto a las pruebas con 3 variables, quitando además del IPH la tasa de actividad o el IPC, también mejora el modelo en ambos casos:

- a) Plazas, PIB, tasa actividad  $\rightarrow \bar{R}^2= 0.982173$
- b) Plazas, PIB, IPC  $\rightarrow \bar{R}^2= 0.982960$

El mejor modelo sería el b) eliminando además del IPH la variable tasa de actividad. Las variables plazas y PIB siguen siendo significativas y el IPC es no significativa.

Por último, si eliminamos la variable IPC la cuál es no significativa obtenemos el mejor resultado posible:

$$\text{PLAZAS, PIB} \rightarrow \bar{R}^2= 0.983145$$

MODELO 3: Viajeros  $\rightarrow$  Establecimientos - T.Actividad - Población - IPC

En este modelo cambiamos en el análisis la variable plazas por establecimientos, con la que vimos que estaba correlacionada. En la regresión múltiple con 5 variables a diferencia de los modelos 1 y 2, además de ser significativas las variables establecimientos y población también lo son la tasa

actividad y el IPH, siendo únicamente no significativa el IPC. Pero el  $\bar{R}^2=0.92351$  es menor a los obtenidos en los anteriores modelos.

Este modelo queda descartado al ser inferior el  $\bar{R}^2$ , ya que a pesar de mejorar al eliminar la variable no significativa (IPC), los resultados son peores a los de los modelos 1 y 2.

MODELO 4: Viajeros → Establecimientos - T.Actividad - PIB - IPC - IPH

En este último modelo los resultados son similares a los del modelo 3. Las variables significativas son las mismas (cambiando la POBLACION por el PIB) y el IPC no es significativa, siendo el  $\bar{R}^2=0.91960$  incluso menor que el anterior.

También descartamos este modelo por el mismo motivo que el modelo 3.

En resumen, después de realizar estas pruebas he elegido el MODELO 2 que tiene como endógena el nº de VIAJEROS (Y) y como variables explicativas el nº de PLAZAS y el PIB, por tener el mayor  $\bar{R}^2$ .

$$\text{VIAJEROS} = \beta_0 + \beta_1 \cdot \text{PIB} + \beta_2 \cdot \text{PLAZAS} + \varepsilon_i$$

Los resultados obtenidos con el programa Eviews mediante la estimación por MCO de la regresión múltiple han sido los siguientes:

Dependent Variable: VIAJEROS				
Method: Least Squares				
Date: 11/16/15 Time: 20:47				
Sample: 1 19				
Included observations: 19				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
PIB	0.036284	0.003405	10.65700	0.0000
PLAZAS	39.17594	2.430160	16.12073	0.0000
C	-364424.1	217349.5	-1.676673	0.1130
R-squared	0.985018	Mean dependent var		4610476.
Adjusted R-squared	0.983145	S.D. dependent var		5126678.
S.E. of regression	665583.8	Akaike info criterion		29.79866
Sum squared resid	7.09E+12	Schwarz criterion		29.94778
Log likelihood	-280.0872	Hannan-Quinn criter.		29.82389
F-statistic	525.9604	Durbin-Watson stat		2.370294
Prob(F-statistic)	0.000000			

### 3.3.2. Análisis de los posibles problemas del modelo

#### 1. Multicolinealidad

Ya realizamos la matriz de correlaciones anteriormente donde vimos que el coeficiente de correlación lineal simple entre la variable PIB y PLAZAS era de  $r_{\text{PIB,PLAZAS}} = 0,6562$ . A pesar de estar algo correlacionadas, ambas variables eran significativas y necesarias en la regresión ya que si quitamos cualquiera de ellas el  $R^2$  disminuye.

También podríamos calcular el factor de inflación de la varianza (FIV) del estimador que acompaña a cualquiera de los 2 regresores. El coeficiente de correlación múltiple entre los dos regresores es  $R^2_{\text{PIB,PLAZAS}} = 0.430635$ , con lo que  $FIV(\beta_i) = 1/1-R^2_{x_j} = 1,756$ .

Se considera que existen problemas cuando el  $FIV > 10$ , es decir, cuando el  $R^2_{x_j} > 0,90$  de modo que en nuestro caso a pesar de estar algo correlacionadas, no habría problema y la varianza del estimador no estaría “inflada”.

#### 2. Normalidad

El incumplimiento de la hipótesis clásica de normalidad de las perturbaciones hace que los estadísticos usados para los contrastes de hipótesis no sigan distribuciones conocidas.

Se analiza con el Test de Normalidad Jarque-Bera a partir de la forma de la distribución de los residuos, examinando sus discrepancias respecto a la curva campaniforme del modelo normal.

$H_0 = \text{Normalidad}$

$H_1 = \text{No normalidad}$

Vemos que el coeficiente de simetría es  $= -0,68$  (0 sería simétrica, hay asimetría a la izquierda) y el coeficiente de kurtosis  $= 3,85$  (3 sería mesocúrtica, es una distribución leptocúrtica con apuntamiento algo mayor que la normal). A pesar de ello el valor del estadístico J-B es pequeño  $= 2,077$  y su p-valor  $= 0,3538$  con lo que no rechazamos  $H_0$  y asumimos que hay normalidad.

### 3. Linealidad

Otra de las hipótesis clásicas del modelo de regresión es la de la forma funcional lineal, de modo que una especificación incorrecta en la forma funcional genera estimadores sesgados e inconsistentes.

El contraste Reset de Ramsey sirve para la detección de no linealidades y consiste en especificar un modelo alternativo al inicial ( $Y=X\beta+\varepsilon$ ) de la forma:  $Y=X\beta+Z\gamma+u$  introduciéndose en "Z" potencias de los valores estimados de la variable dependiente.

$H_0: \gamma=0$  (linealidad)

$H_1: \gamma\neq 0$  (No linealidad)

---

#### Ramsey RESET Test:

---

F-statistic	9.639231	Prob. F(1,15)	0.0072
Log likelihood ratio	9.429505	Prob. Chi-Square(1)	0.0021

---

Si introducimos una sola potencia, el modelo sería del tipo  $Y_i = \beta_0 + \beta_1 \cdot X_1 + \beta_2 \cdot X_2 + \beta_3 \cdot \hat{Y}^2 + \varepsilon_i$

El estadístico  $F=9,63$  y su p-valor  $=0,0072$  con lo que rechazo  $H_0$ . El modelo planteado no es lineal, de modo que habrá que solucionar el problema.

### 4. Heteroscedasticidad y autocorrelación

En el Modelo de Regresión Lineal Clásico ( $Y=X\beta+\varepsilon$ ) sabemos que la matriz de varianzas-covarianzas es  $E(\varepsilon\varepsilon') = \sigma^2 \cdot I$  y por lo tanto las varianzas de las perturbaciones son iguales (hipótesis de homoscedasticidad) y las covarianzas entre ellas son nulas (hipótesis de incorrelación).

En caso de que no se verifiquen cualquiera de las dos hipótesis tendremos un Modelo de Regresión Lineal Generalizado (MRLG) con los casos particulares de heteroscedasticidad o autocorrelación de las perturbaciones, y en consecuencia la matriz es  $E(\varepsilon\varepsilon') = \sigma^2 \cdot \Omega$ .

Aunque en datos de corte transversal no suele darse autocorrelación vamos a comprobar si existe o no.

La autocorrelación se produce cuando las covarianzas de las perturbaciones no son iguales a cero, es decir, se rompe con la hipótesis de incorrelación entre las perturbaciones del MRLC:

$$\text{Cov}(\varepsilon_t \varepsilon_s) = E(\varepsilon_t \varepsilon_s) \neq 0$$

La relación de correlación entre las perturbaciones puede ser diversa, siendo la forma habitual el proceso autorregresivo de orden 1, AR(1):

$\varepsilon_t = \rho \varepsilon_{t-1} + u_t$ , donde  $u_t$  es un ruido blanco (variable aleatoria que cumple las hipótesis clásicas) y “ $\rho$ ” es un parámetro a estimar.

Otros posibles esquemas de autocorrelación son:

$$\text{AR}(p): \quad \varepsilon_t = \phi_1 \cdot \varepsilon_{t-1} + \phi_2 \cdot \varepsilon_{t-2} + \dots + \phi_p \cdot \varepsilon_{t-p} + u_t$$

$$\text{MA}(q): \quad \varepsilon_t = u_t + \theta_1 \cdot u_{t-1} + \theta_2 \cdot u_{t-2} + \dots + \theta_q \cdot u_{t-q}$$

Para analizar la posible existencia de autocorrelación podemos utilizar por ejemplo el contraste de Breusch-Godfrey

$H_0$ : Incorrelación ( $\rho_1 = \rho_2 = \dots = \rho_m = 0$ )

$H_1$ : AR(1) o MA(1)

---

Breusch-Godfrey Serial Correlation LM Test:

F-statistic	1.539278	Prob. F(1,15)	0.2338
Obs*R-squared	1.768292	Prob. Chi-Square(1)	0.1836

---

El estadístico de contraste es  $T \cdot R_{aux}^2 \rightarrow \chi_m^2$ , siendo “m” el número de retardos. Con lo cual aquí no nos interesa el valor del estadístico F, sino el de “Obs\*R-squared” = 1,768 y su “Prob. Chi-Square” (1 retardo) o p-valor asociado al estadístico = 0,1836. Con este p-valor no rechazamos la hipótesis nula y suponemos que no hay autocorrelación.

La heteroscedasticidad se da cuando las varianzas de las perturbaciones son desiguales

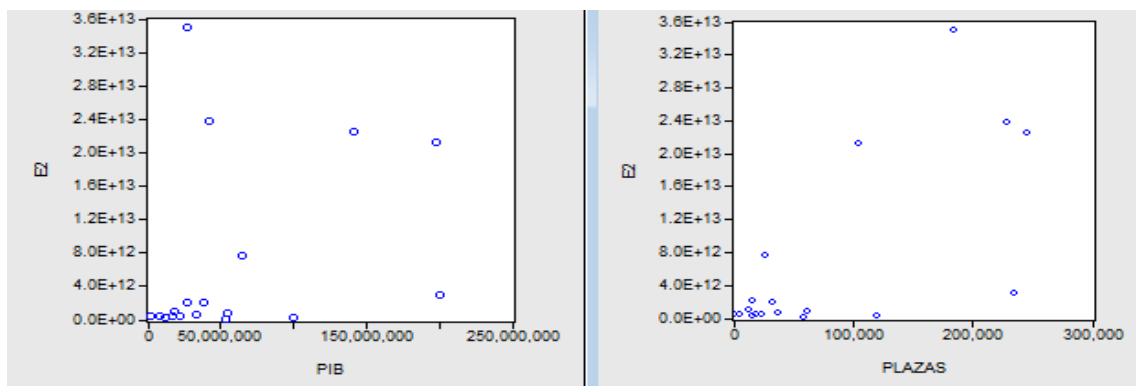
$$(\sigma_i^2 \neq \text{cte}) \rightarrow \text{Var}(\varepsilon_i) = \sigma^2 \cdot w_i$$

Este problema afecta más a datos de corte transversal y puede deberse también a que en el modelo existan errores de especificación, como omisión de variable, cambio estructural o en la forma funcional.

Analizamos a continuación si existe este problema en nuestro modelo mediante diferentes métodos.

a) Método gráfico de los residuos

Consiste en realizar un diagrama de dispersión de los residuos al cuadrado (ordenadas) en función de cada uno de los regresores (abscisas). Si los residuos tienen algún tipo de tendencia dependiendo de los regresores implicaría que la dispersión o varianza de las perturbaciones depende de los valores de X (varianzas no ctes) y habría heteroscedasticidad.



Como vemos en ambas variables la mayoría de los valores pequeños se acumulan juntos pero los residuos son mayores y más dispersos para valores grandes del regresor. Esto nos indica la posible existencia de heteroscedasticidad.

b) Método analítico (contraste de hipótesis)

En las pruebas estadísticas al igual que en los gráficos, como no es observable  $\varepsilon_i$  usamos  $e_i$  y utilizamos  $e_i^2$  como una aproximación a la varianza de las perturbaciones, de modo que queremos ver si los residuos MCO al cuadrado ( $e_i^2$ ) son función de los regresores.

Contraste de White:

Este test analiza la significación de una regresión auxiliar que trata de explicar los  $e_i^2$  (residuos al cuadrado de la regresión inicial), a partir de los regresores, sus cuadrados y sus productos cruzados dos a dos.

H<sub>0</sub>: homoscedasticidad

H<sub>1</sub>: heteroscedasticidad

Heteroskedasticity Test: White

F-statistic	16.85635	Prob. F(5,13)	0.0000
Obs*R-squared	16.46098	Prob. Chi-Square(5)	0.0056
Scaled explained SS	16.65891	Prob. Chi-Square(5)	0.0052

El estadístico de contraste es  $N \cdot R_{aux}^2 \rightarrow \chi^2$ , siendo el  $R_{aux}^2$  el coeficiente de determinación de la regresión auxiliar realizada antes. Con lo cual aquí no nos interesa el valor del estadístico F, sino el de “Obs\*R-squared” ( $N \cdot R_{aux}^2$ )= 16,460 y su “Prob. Chi-Square” o p-valor asociado al estadístico de White = 0,0056. Con este p-valor rechazo la hipótesis nula para cualquier nivel de significación de los habituales, lo que significa que hay heteroscedasticidad y habrá que tener en cuenta este problema.

En resumen, nos hemos encontrado con falta de linealidad y con heteroscedasticidad.

### 3.3.3. Solución a los problemas existentes

Intentaremos solucionar en primer lugar el problema de la heteroscedasticidad puesto que en ocasiones al hacerlo también se soluciona la no linealidad.

Las consecuencias de aplicar MCO a un MRLG son:

- $\hat{\beta}_{MCO}$  sigue siendo insesgado, lineal y consistente.
- La matriz de varianzas-covarianzas ya no es la misma con lo que el estimador no es óptimo ni eficiente, y además el  $S^2$  deja de ser insesgado y consistente al ser el estimador de las varianzas de las perturbaciones. También debido a que ahora  $E(\varepsilon\varepsilon') \neq \sigma^2 \cdot I$  los estadísticos “t” y “F” no son válidos (distribución exacta no válida) y  $\hat{\beta}_{MCO} \neq \hat{\beta}_{MV}$ .

-Cuando en un modelo hay heteroscedasticidad o no linealidad lo primero que vamos a hacer es ver si se debe a un error de omisión, introduciendo algún factor que creamos que es relevante y no esté en el modelo.



He tratado de ver si faltaban variables que recogieran por ejemplo el hecho de que la CCAA tenga o no costa, o que tenga o no algún centro estratégico, por ejemplo, de negocios como Madrid o Barcelona, etc, introduciendo ficticias, pero no fueron significativas ni solucionaban el problema.

-Una vez analizado lo anterior, una posible solución es estimar por Mínimos Cuadrados Ponderados (MCP), que es el estimador que minimiza la suma de los cuadrados de los residuos ponderados. Este método en la teoría también se llama Mínimos Cuadrados Generalizado Factible (MCGF).

Lo que hace este estimador es dar más ponderación a las observaciones provenientes de poblaciones con menor varianza que a las de mayor, lo que da más precisión al estimar (el estimador MCO en el MRLG no es eficiente ya que cada error recibe la misma ponderación).

Para obtener este estimador tenemos que transformar el modelo para conseguir que las varianzas de las perturbaciones sean constantes y no tenga heteroscedasticidad. De este modo cumpliría las hipótesis clásicas y tendrá las propiedades de la estimación MCO en el MRLC (los mejores estimadores).

Partiendo de un modelo con heteroscedasticidad:

$\text{Var}(\varepsilon_i) = \sigma^2 \cdot w_i$  se transforma el modelo ponderando cada observación por  $1/\sqrt{w_i}$  de modo que las distribuciones de mayor varianza tienen menos peso.

Pero como no conocemos “ $w_i$ ” lo que hacemos es estimarla ( $\hat{w}_i$ ) transformando el modelo de regresión original, de modo que se plantean hipótesis suponiendo que las varianzas de las perturbaciones son proporcionales a los valores de un regresor o de Y estimado.

He probado a estimar la  $w_i$  de varias formas, solucionando el problema con una de ellas:

Si considero  $\text{Var}(\varepsilon_i) = \sigma^2 \cdot \text{PIB}^2 \rightarrow$  no se soluciona el problema (p-valor en White = 0,0094).

Si consideramos  $\text{Var}(\varepsilon_i) = \sigma^2 \cdot \text{PLAZAS}^2 \rightarrow$  se arregla el problema de heteroscedasticidad (p-valor = 0,1251) y además el problema de la no linealidad (p-valor = 0,2434).

Como vemos, al estimar por MCGF suponiendo que  $\text{Var}(\varepsilon_i) = \sigma^2 \cdot \text{PLAZAS}^2$  se arreglan los problemas de nuestro modelo, y los estimadores no tendrían propiedades finitas pero sí asintóticas.

-Otra forma para corregir la heteroscedasticidad no consiste en cambiar el método de estimación, sino en realizar una transformación de los datos tomando logaritmos y estimando el modelo "log-lineal".

En EVIEWS tras generar las variables LVIAJEROS, LPIB y LPLAZAS y estimar por MCO comprobamos con el test de White que no hay heteroscedasticidad (p-valor=0,9251) ya que al tomar logaritmos se reduce la variabilidad del modelo. Además el problema de la no linealidad también se ha solucionado (p-valor=0,4549).

En este caso, puesto que se trata de estimadores de MCO serían ELIO y consistentes.

-Otra opción cuando existe heteroscedasticidad es estimar el modelo por MCO pero con la matriz de varianzas y covarianzas de White. De este modo aunque no se corrige la heteroscedasticidad los estimadores son insesgados y consistentes aunque no sean eficientes, pero en este caso seguiríamos teniendo el problema de la no linealidad.

En resumen, aunque cualquiera de las dos primeras opciones (estimación por MCP o transformando las variables en logaritmos) sería válida, elegimos el modelo estimado en logaritmos porque los estimadores conservan además las propiedades finitas, y dado que nuestra muestra no es muy grande es preferible.

El modelo final una vez corregidos los problemas es el siguiente:

$$\text{LVIAJEROS} = \beta_0 + \beta_1 \cdot \text{LPIB} + \beta_2 \cdot \text{LPLAZAS} + \varepsilon_i$$

Dependent Variable: LVIAJEROS				
Method: Least Squares				
Date: 11/16/15 Time: 20:34				
Sample: 1 19				
Included observations: 19				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
LPIB	0.302251	0.052684	5.737013	0.0000
LPLAZAS	0.714948	0.042802	16.70367	0.0000
C	1.948585	0.536898	3.629342	0.0023
R-squared	0.994264	Mean dependent var		14.53459
Adjusted R-squared	0.993548	S.D. dependent var		1.595079
S.E. of regression	0.128128	Akaike info criterion		-1.127629
Sum squared resid	0.262670	Schwarz criterion		-0.978507
Log likelihood	13.71247	Hannan-Quinn criter.		-1.102391
F-statistic	1386.814	Durbin-Watson stat		2.488597
Prob(F-statistic)	0.000000			

### 3.3.4. Interpretación de los resultados en el modelo final

Estimadores de los coeficientes de las variables: Estos coeficientes (términos  $\beta$ ) representan en general el incremento de la variable dependiente cuando las variables independientes aumentan en 1 unidad (ceteris paribus).

En este caso, como las variables están en logaritmos los incrementos se miden en porcentaje.

La  $\hat{\beta}_1$  correspondiente a la variable LPIB = 0,302251 significa que si el PIB aumenta en un 1% el nº de viajeros aumentaría un 0,30 % si lo demás no cambia.

La  $\hat{\beta}_2$  correspondiente a la variable LPLAZAS = 0,714948 significa que si aumenta un 1% el nº de plazas, el nº de viajeros aumentará en 0,71 % si lo demás no cambia.

Bondad de ajuste:

$R^2 = 0.994264 \rightarrow$  indica que el 99,4 % de la variabilidad de Y viene explicado por las 2 variables independientes, y los residuos sólo explican el 0,6 % restante, con lo que es muy buen ajuste.

### Significación:

Para contrastar la significación individual y colectiva de las variables, Eviews proporciona en las salidas de la estimación tanto el valor muestral del estadístico “t” como del estadístico “F”.

#### *Regresor aislado:*

LPIB →  $H_0: \beta_1=0$  (no significativa)

$$H_1: \beta_1 \neq 0$$

El estadístico  $t= 5,737$  y  $p\text{-valor}=0,00$  (la probabilidad de equivocarnos si rechazamos  $H_0$  es nula) con lo que rechazamos  $H_0$  y es significativa

LPLAZAS →  $H_0: \beta_2=0$  (no significativa)

$$H_1: \beta_2 \neq 0$$

El estadístico  $t= 16,703$  y  $p\text{-valor}=0,00$  con lo que rechazamos  $H_0$  y es significativa

#### *Contraste de significación de la regresión:*

$H_0: \beta_1=\beta_2=0$  (no significativa)

$$H_1: \beta_1=\beta_2 \neq 0$$

El estadístico  $F=1386,81$  y  $p\text{-valor}=0,00$  con lo que rechazamos  $H_0$  y son conjuntamente significativas (la regresión es totalmente significativa).

## **4. CONCLUSIONES**

Por último veremos a modo resumen una serie de conclusiones acerca del trabajo:

- En España el sector del turismo es una de las actividades económicas más relevantes al generar una gran cantidad de riqueza y empleo. Además, tiene una importante cuota de participación en la economía nacional, siendo un elemento clave para la recuperación económica.
- Desde que se iniciaron en los años 60 los estudios de previsión sobre la demanda turística española mediante un análisis econométrico, se han

ido mejorando las técnicas aplicadas y aumentando y actualizando las variables y los factores de influencia de los modelos econométricos. Los distintos autores han ido modelizando la demanda turística, ampliando los periodos de predicción, encontrando los desajustes en las tendencias y conociendo la estructura de comportamiento y su evolución según las distintas circunstancias.

- Las CCAA que reciben un mayor número de viajeros en España son las regiones costeras de Cataluña y Andalucía, las Islas (Baleares y Canarias) y la capital, Madrid.
- Gracias a los indicios de la matriz de correlaciones nos hacemos una idea de las variables explicativas que están más relacionadas con la variable endógena (número de viajeros), y mediante los análisis de regresión siguientes observamos que las variables más relevantes para nuestro modelo econométrico son el PIB y las PLAZAS.
- Nuestro modelo, al ser datos de corte transversal, presentaba heteroscedasticidad, pero se solucionó trabajando con las variables en logaritmos.

## 5. BIBLIOGRAFÍA

### -Páginas web:

Organización Mundial del Turismo (OMT) <http://www2.unwto.org/es>

Ministerio de Industria, Energía y Turismo <http://www.minetur.gob.es/>

Turespaña <http://www.tourspain.es/>

Alianza para la Excelencia Turística (Exceltur) <http://www.exceltur.org/>

Instituto Nacional de Estadística (INE) <http://www.ine.es/>

Institutos de Estudios Turísticos (IET) <http://www.iet.tourspain.es/>

### -Noticias de prensa (datos del turismo):

<http://www.minetur.gob.es/es->

[ES/GabinetePrensa/NotasPrensa/2015/Paginas/20151112-camara-comercio-sevilla.aspx](http://www.minetur.gob.es/es-ES/GabinetePrensa/NotasPrensa/2015/Paginas/20151112-camara-comercio-sevilla.aspx)

<http://www.minetur.gob.es/es-ES/GabinetePrensa/NotasPrensa/2015/Paginas/20151104-desayuno-operadores.aspx>

<http://sabemosdigital.com/trip/366-cuanto-aporta-el-turismo-a-espana-despues-de-tres-anos-de-record>

-Artículos y libros:

Chang, S. y Aguayo, E. (2002): “Modelo econométrico de la demanda turística internacional en China”, *Estudios Económicos Regionales y Sectoriales*, Vol. 3-2, AEEADE, pp. 93-106.

Chica, J. (2014): “Elaboración de un modelo econométrico”, *Proyecto de innovación docente Guime 2.0*, Universidad de Granada.

Esteban, A. (2004): “Modelos de la demanda turística en España: segmentación por países de procedencia”, *Las nuevas formas del turismo*, Cajamar, pp. 81-101.

Esteban, A. y Reinares, E. (1996): “La investigación de la demanda turística en España: recopilación y análisis”, *Estudios turísticos*, nº 129, Ed. Instituto de Estudios Turísticos, pp. 97-100.

Guisán, M.C. y Aguayo, E. (2008): “Evolución del turismo hotelero en las regiones españolas (2001-2008)”, *Regional and Sectorial Economic Studies*, Vol. 8-1, pp. 127-136.

Guisán, M.C. y Neira, I. (2001): “Un análisis econométrico del turismo hotelero y extrahotelero en las regiones y provincias españolas”, *Estudios Económicos Regionales y Sectoriales*, Vol. 1-2, AEEADE, pp. 1-13.

Sánchez, M. (1998): “Modelización estadística de tablas de contingencia: aplicación al análisis de la demanda turística española”, *Tesis doctoral de la Universidad de Extremadura*, pp. 25-27.

Wooldrige, J.M. (2001): *Introducción a la econometría: Un enfoque moderno*, Capítulo 19, ED. Thomson Learning.