



E.T.S. de Ingenieros de Caminos, Canales y Puertos

Tesis Doctoral

**MODELIZACIÓN DE LA DEMANDA DE
TRANSPORTE PÚBLICO ANTE FASES
CONTRACTIVAS DEL CICLO ECONÓMICO**

Autor

César Canales García

Director:

Borja Alonso Oreña

Santander, Enero de 2016





RESUMEN

Avanzar en el conocimiento relativo a los factores que llevan a que se produzcan cambios en la demanda de viajes en transporte público puede resultar de gran interés de cara a una mayor eficacia en las labores de gestión y planificación del transporte público y así contribuir al mejor desarrollo de un sistema de transporte más racional y eficiente que suponga una reducción de costes sin necesidad de que empeore la calidad del servicio prestado.

Poder modelizar la demanda de transporte público con el fin de prever su comportamiento en función de determinados cambios en la dinámica macroeconómica, como pueden ser un periodo de recesión o un periodo de recuperación económica, supone una ayuda de interés para gestores y planificadores del Transporte Público y así tener la opción de poder ajustar la oferta de transporte a una demanda que se va a ver influenciada por las distintas dinámicas socioeconómicas y que en muchas ocasiones pueden hacer variar su comportamiento de cara a utilizar el transporte público. Las dinámicas socioeconómicas van a estar relacionadas en gran medida con variaciones en el nivel de renta por un lado, o por el aumento de la tasa de paro por otro. Variables cuyas variaciones se han visto influenciadas con el avance de la crisis económica en los últimos años que ha generado de forma general un descenso del número de viajeros, excepto en el tramo final del periodo analizado y en consecuencia una caída de la recaudación.

En diversas ciudades los servicios de transporte se planifican de una manera “rígida” y con unos horarios determinados. Dicha planificación en muchas ocasiones se mantiene a lo largo de varios años y sin importantes modificaciones de tal forma que se convierte en una especie de “foto fija” que apenas tiene en cuenta los cambios económicos responsables a su vez de los cambios de conducta de los usuarios. Una estrategia que puede suponer, de modo general, una consecuente caída de viajeros y por tanto de la recaudación. Para resolver este tipo de problemas se han tomado diferentes medidas como pueden ser el mantenimiento de las tarifas durante varios años sin incrementar el



precio, poner en marcha mejoras en la calidad del servicio ofrecido (frecuencias, limpieza, trato del personal), la redefinición de la red de transporte urbano para dar servicio a las nuevas áreas residenciales que se consolidan en zonas periurbanas o la prestación de ayudas a colectivos sociales desfavorecidos como pueden ser los parados además de ventajas para familias numerosas, tercera edad y estudiantes. Este tipo de medidas no siempre han contribuido necesariamente a incrementar el número de viajeros por lo que el análisis de diferentes variables socio-económicas como el coste del combustible, el IPC (Índice de Precios al consumo) o la tasa de motorización resultan de interés para tratar de prever de la forma más ajustada posible cual puede ser la respuesta de la demanda considerando también la dinámica macroeconómica.

Un modelo de previsión de la demanda de viajeros en transporte público es capaz de estimar el número de usuarios en equilibrio y a corto y largo plazo considerando distintas variables de ámbito socioeconómico, macroeconómico y demográfico. En este estudio se establecen cuales de esas variables son las que tienen una mayor incidencia sobre la demanda de viajeros llegándose a calcular las elasticidades de cada una de ellas. Para la construcción del mismo se realiza el cálculo basado en regresiones y en series temporales. En el caso de los modelos de regresión, se utilizan como variables explicativas aquellas que han obtenido un alto coeficiente de correlación con la variable a estimar (independiente) como es la demanda de viajeros (Número de viajeros mensuales). Para llegar a construir este modelo de previsión, ha sido necesario llevar a cabo un profundo análisis no sólo de la demanda de viajeros a lo largo de un periodo importante de tiempo (12 años) sino que también han sido objeto de un exhaustivo estudio toda una serie de variables socioeconómicas, macroeconómicas y demográficas para el mismo periodo de tiempo de manera que se pudieran extraer conclusiones que pudieran explicar los distintos cambios experimentados en la demanda de viajeros en función de la variación de los datos de esas variables macroeconómicas y socioeconómicas. La razón por la que se ha centrado el interés en esas variables económicas está relacionada con la importante recesión experimentada por la economía española y de cómo esa recesión y otros periodos más expansivos de la



economía pueden llegar a incidir en las demandas de viajeros en los servicios de transporte público.

La metodología utilizada para la obtención del modelo se divide en dos partes. En la primera de ellas se procede a la construcción de una base de datos en la que se ha recogido información relativa al número de viajeros mensuales en un servicio municipal de transportes a lo largo de una serie de doce años (2001 – 2012). El número de viajeros mensuales se convierte así en la variable a prever o estimar convirtiéndose en la variable dependiente. Junto a esa variable se han tenido en cuenta otras relacionadas con aspectos sociales y aquellas relacionadas con los ciclos económicos para ese mismo periodo de tiempo y con la misma escala temporal. Estas últimas variables se convertirán en las variables independientes. Una vez construida la base de datos, se procedió a un riguroso análisis de las variables y se estableció una comparativa entre estas variables y la variable a estimar para así poder obtener algunas conclusiones que ayudaran a entender qué factores podían influir más en la demanda de viajeros.

En la segunda parte se procede a un tratamiento de la base de datos que consiste en la aplicación de distintas técnicas estadísticas (análisis de tendencias, estacionalidad, estacionariedad, correlación entre las variables, cálculo de la recta de regresión lineal, deflactación de variables monetarias...etc). Una vez la base de datos es tratada se desarrolla la transformación de todas las variables obteniendo su logaritmo natural para que los parámetros obtenidos puedan interpretarse directamente como elasticidades. El cálculo de elasticidades se realiza mediante un modelo estático de regresión de tipo Log-Log, es decir como el cambio porcentual de equilibrio de la variable dependiente ante los cambios porcentuales de las variables independientes. Además se estiman las elasticidades de equilibrio, a corto y a largo plazo de las variables de interés mediante un modelo de regresión log – log considerando la presencia de autocorrelación en los residuos y la endogeneidad de la oferta de transporte.

Una vez se ha obtenido el modelo de previsión en el siguiente paso se lleva a cabo la validación del modelo para lo cual se cuentan con datos reales pertenecientes al año



2013. De esa forma se comprueba el porcentaje de error existente entre los datos arrojados por el modelo con los datos reales. El objetivo principal es evaluar mediante un modelo econométrico la influencia de variables macroeconómicas como Tasa de Desempleo y el nivel de ingresos “per cápita”, sobre la demanda de transporte público con datos de la ciudad de Santander (TUS), calculando las elasticidades de las variables dependientes de interés estimando los modelos a partir de una serie temporal obtenida a partir de los servicios de transporte público. La finalidad es aportar datos de interés a planificadores y gestores de los transportes públicos de cara al ajuste en la gestión de sus servicios en función de la coyuntura económica existente.

La investigación llevada a cabo en este trabajo ha sido publicada en:

- Canales C., Cordera R., Dell’Olio L., Ibeas., (2015). “Public transport demand elasticities during the recessionary phases of economic cycles”. *Transport Policy* 42 (2015) 173-179. Elsevier.



ABSTRACT

Advance knowledge concerning the causes or reasons that lead to different trends occurring in the demand for public transport trips can be of great interest with regard to greater efficiency in the work of management and planning of public transport and contribute the better development of a system more rational and efficient transport would mean a optimization and cost reduction without worsening the quality of service provided.

To model the demand for public transport in order to predict its behavior in line with certain changes in economic flows, as they can be a recession or a period of economic recovery, is an aid appeal to managers and planners of Public Transport and to adjust the supply of transport to a demand that is going to be influenced by the different socio-economic dynamics and many times can vary their behavior in the face of public transport use. The socio-economic dynamics will be largely related to variations in the level of income on the one hand, or by increasing the unemployment rate on the other. Variables whose variations have been influenced by the progress of the economic crisis in recent years has generated generally decreased the number of travelers and consequently a drop in revenues, so the service costs have become more expensive.

A model for forecasting the demand for public transport travelers is able to estimate the number of users in the short and medium term considering other variables of socioeconomic, demographic and macroeconomic level. In this model also which of these variables are set are those that have a greater impact on the demand of travelers arriving to calculate the elasticities of each. For its construction it based on time-series regression and calculation. In the case of regression models they are used as explanatory variables those who have obtained a high correlation coefficient with the variable estimate (independent) as passenger demand (number of monthly passengers). To get to build forecasting model has been necessary to conduct a thorough analysis not only of passenger demand over a significant period of time (12 years) but have also been the



subject of exhaustive study a whole series of socio-economic, macroeconomic and demographic for the same period of time so that they could draw conclusions that could explain the various changes experienced in passenger demand depending on the variation in the data of these macroeconomic and socioeconomic variables. The reason has focused interest in these economic variables is related to the major recession experienced by the Spanish economy and how the recession and more expansive periods of the economy may come to influence the demands of travelers services public transport.

The serious recession suffered by the Spanish economy has had an important impact on the working of transport systems. This thesis provides evidence of how the state of the economy, measured using per capita income levels and unemployment rates, can influence the demand for public transport by bus with data from the city of Santander (In the north of the Spain area) for the period 2001-2012. The methodology used has involved the short and long run equilibrium demand elasticities estimation using a log-log regression model considering the presence of autocorrelation in the residuals and the endogeneity of the transport supply. The results show that the demand for transport is sensitive to changes in unemployment rates with an elasticity estimated at 0.133 in the static equilibrium model and of 0.210 in the long run dynamic model. The resulting elasticity for income levels was not significant in all the models, with a parameter estimated in the static model of -0.505 and of -0.861 in the long run dynamic model. The model estimated by two stage last squares validated using data from 2013 with unemployment rate as an independent variable gave the lowest average square error in validation.



AGRADECIMIENTOS

Quisiera dedicar estas primeras palabras de agradecimiento y, especialmente, de afecto, a todos los que, en mayor o menor medida han contribuido a mi desarrollo como investigador para poder alcanzar el objetivo de desarrollar esta tesis. Entre ellos a mi director de tesis D. Borja Alonso Oreña por su dedicación y paciencia, así como del resto de profesores del departamento de transportes D. Ángel Ibeas, D. Luigi dell’Olio, y D. José Luis Moura. Igualmente quiero agradecer, muy especialmente, a mi colega geógrafo D. Rubén Cordera Piñera, su inestimable ayuda y no menor paciencia, para que este estudio pudiera llegar a buen puerto.

Mi agradecimiento también al TUS (Transportes Urbanos de Santander) por su valiosa aportación ya que gracias a los datos que me han facilitado ha sido posible la realización de esta tesis.

No quisiera olvidarme de todos mis compañeros del Grupo de Investigación de Sistemas de Transporte de la Universidad de Cantabria, tanto los antiguos como los actuales y de todos ellos una mención especial a Juan Pablo, Rosa, Álex y María, así como los miembros del Grupo Ferrocarriles de la Universidad de Cantabria Roberto y Raúl.

He querido dejar para el final a toda mi familia, a todos y cada uno de los miembros que la componen por su interés, por su preocupación y por su apoyo incondicional. Un recuerdo muy entrañable para mis dos abuelos “Pepe” y Gonzalo, que no han podido llegar a ver este momento pero que estoy seguro de que se alegran de que haya llegado hasta aquí.

La familia es una piedra fundamental sobre la que apoyarse tanto en los buenos como en los malos momentos al igual que es importante también apoyarse sobre los amigos. Para ellos van dedicadas también estas palabras y esta cita:

“Sin amigos nadie querría vivir, aunque poseyera los demás bienes, porque la prosperidad no sirve de nada si se está privado de la posibilidad de hacer el bien,



la cual se ejercita, sobre todo, respecto de los amigos. Asimismo, en los infortunios se considera a los amigos como único refugio”.

(Aristóteles, libros VIII y IX de la “Ética a Nicómano”).

Por último y en un lugar destacado quiero dedicarle estas letras a mi mujer, María, a quien tengo que agradecer todo su apoyo y comprensión en estos años y toda la paciencia que ha demostrado tener. Su apoyo, especialmente en el último tramo de este trabajo, ha sido de vital importancia para no permitir que decayera en el esfuerzo. Cada página de esta tesis lleva también un aporte de esfuerzo y ánimo suyo.



ÍNDICE

RESUMEN	3
ABSTRACT	7
AGRADECIMIENTOS.....	9
1. INTRODUCCIÓN, OBJETIVOS Y APORTACIONES	20
1.1. Motivación	20
1.2. Objetivos	25
1.3. Estructura.....	27
1.4. Aportaciones	30
2. ESTADO DEL ARTE	34
3. METODOLOGÍA.....	52
3.1. Las Series Temporales	53
3.1.1. Tipos de variables y datos de la modelización econométrica.....	60
3.1.2. Análisis de una Serie Temporal	61
3.1.3. Componentes de una Serie Temporal.....	63
3.2. Los Modelos de Regresión	72
3.2.1. Los Modelos de Regresión Simple (MRLS)	75
3.2.2. Los Modelos de Regresión Múltiple (MRLM).....	77
3.2.3. Hipótesis del modelo de regresión lineal múltiple (MRLM).....	80
3.2.4. Hipótesis sobre el término de perturbación	82
3.2.5. Estimación del Modelo de Regresión Lineal Múltiple (MRLM).....	84
3.2.6. Medidas de Bondad del ajuste del Modelo (MRLM)	92
3.2.7. Significación de los parámetros del modelo	94
3.3. Datos de Panel.....	105
3.3.1. Tipos de análisis con datos de panel	109
3.3.2. Especificación general de un modelo de datos de panel. Estimación con datos de panel	110
3.3.3. Cuándo utilizar Datos de Panel	112
3.3.4. Supuestos necesarios para usar Datos de Panel.....	114
3.3.5. Elección entre los distintos tipos de datos de panel. Datos de Panel: Estáticos o dinámicos y posibles correcciones. Comprobación de la Endogeneidad.	115
3.3.6. Estimador con variables instrumentales.....	116



3.4.	Validación de los modelos.....	118
3.4.1.	El Error cuadrático Medio	118
3.4.2.	Otros Estadísticos:.....	119
3.5.	Deflactación de variables monetarias.....	120
3.6.	Cálculo de elasticidades	123
4.	DESCRIPCIÓN DEL AREA DE ESTUDIO Y ANÁLISIS DE DATOS.....	130
4.1.	Área de Estudio	130
4.2.	Datos utilizados	137
4.2.1.	Origen de los datos.....	138
4.2.2.	EL TUS (Transportes Urbanos de Santander).....	139
4.2.3.	Evolución de las líneas del TUS (2001-2012).....	140
4.3.	Construcción y tratamiento de los datos	144
4.3.1.	Homogeneidad de los datos.....	145
4.3.2.	Deflactación de variables monetarias.....	146
4.3.3.	Tendencia y Estacionalidad	148
4.3.4.	La Estacionalidad.....	150
4.3.5.	Desestacionalización de los datos y análisis de la correlación.....	151
4.4.	Análisis de datos.....	152
4.4.1.	Evolución de la demanda de viajeros y el desempleo	152
4.4.2.	Evolución de la demanda de viajeros y nivel de ingresos	157
4.4.3.	Demanda de viajeros y evolución del coste del combustible.	159
4.4.4.	Demanda de viajeros y nivel de Motorización	166
4.4.5.	Evolución de la Demanda en el periodo de Crisis Económica (Fase Contractiva) (2009-2012).....	167
4.5.	Análisis de datos incluyendo variable demográfica. Per cápita. Datos generales. ...	174
4.6.	Análisis de datos incluyendo variable demográfica. Per cápita. Datos por línea. Estadísticos Descriptivos.....	177
4.6.1.	LINEA 1 - Parque Tecnológico (PCTCAN) – Adarzo – Arsenio Odriozola.	177
4.6.2.	LINEA 3 - Ojáiz – Paseo Pereda – Rectorado (Universidad)	180
4.6.3.	LINEA 5C1 – Miranda/Plaza Italia – Plaza de los Remedios – General Dávila... ..	183
4.6.4.	LINEA 12 – Carrefour - Canalejas.....	186
5.	ANÁLISIS Y DISCUSIÓN DE RESULTADOS.....	190
5.1.	Metodología y aplicación (Modelo teórico).....	190



5.2.	Resultados y discusión	192
5.2.1.	Resultados del modelo agregado de líneas.....	192
URATE.....		197
5.2.2.	Resultados del modelo desagregado por líneas	198
5.2.2.1.	LINEA 1. Parque Tecnológico (PCTCAN) – Adarzo –Arsenio Odriozola	199
5.2.2.2.	LINEA 3. Ojáiz – Paseo Pereda –Rectorado (Universidad).....	202
5.2.2.3.	LINEA 5C1. Miranda/Plaza de Italia – Plaza de los Remedios- General Dávila. 204	
5.2.2.4.	LINEA 12. Carrefour – Canalejas.....	206
5.3.	Validación del modelo.....	208
5.3.1.	Validación del modelo con datos agregados.....	208
5.3.2.	Validación del modelo con datos desagregados. Por líneas. Las más significativas.	209
5.3.2.1.	Validación del modelo. Línea 1.	209
5.3.2.2.	Validación del modelo. Línea 3.	210
5.3.2.3.	Validación del modelo. Línea 5C1.	210
5.3.2.4.	Validación del modelo. Línea 12.	211
6.	CONCLUSIONES Y LINEAS FUTURAS DE INVESTIGACIÓN	214



ÍNDICE DE ILUSTRACIONES

Ilustración 1. Fuente: Notas de Prensa del INE.	22
Ilustración 2 Esquema general del análisis de una serie temporal (Medina Moral 2002).....	57
Ilustración 3. Representación gráfica de una serie de datos a lo largo de una serie de tiempo. Sánchez Fernández Jesús. “Introducción a la Estadística Empresarial. Cap.4. Series Temporales.	63
Ilustración 4. Representación gráfica de la línea de tendencia sobre la serie de datos temporales expresada gráficamente.	66
Ilustración 5. Expresión gráfica de la captación de la tendencia de una serie temporal aplicando distintos procesos de descomposición.	67
Ilustración 6. Ejemplo gráfico de la desestacionalización de una serie.	69
Ilustración 7 Representación gráfica del ajuste de datos a partir de la recta de regresión. Fuente: (Molina Ibáñez, Giménez et al. 2009)	72
Ilustración 8. Esquema para el estudio de un análisis de regresión.	74
Ilustración 9. Función de regresión poblacional y diagrama de dispersión.....	77
Ilustración 10. Esquema de análisis de un Modelo de Regresión Lineal Múltiple.	79
Ilustración 11. Ejemplo de introducción de observaciones en forma de datos de panel.....	106
Ilustración 12. . Esquema gráfico de los efectos individuales α_i	113
Ilustración 13. Variaciones de la demanda. (Frank and Piras 2003).	124
Ilustración 14. Tipos de elasticidades. Elasticidades Punto y Arco.	127
Ilustración 15. Situación del término municipal de Santander. Fuente: PGOU de Santander.	133
Ilustración 16. Densidad de población de los municipios de Cantabria. Fuente: ICANE.	134
Ilustración 17. Estructura del término municipal de Santander. Entidades de población. Fuente: PGOU	135
Ilustración 18. Término municipal de Santander y su área de influencia. Fuente: ICANE (Instituto Cántabro de Estadística) a partir de Cartografía Regional del Gobierno de Cantabria y datos del INE.	137
Ilustración 19. Líneas de Autobús en el área de estudio.	139
Ilustración 20. Recorrido de la línea 1. Fuente: Aplicación cartográfica del (S.A.E). Sistema de ayuda a la explotación del TUS de Santander.....	178
Ilustración 21. Recorrido de la línea 3. Fuente: Aplicación cartográfica del (S.A.E). Sistema de ayuda a la explotación del TUS de Santander.....	181
Ilustración 22. Recorrido de la línea 5C1. Fuente: Aplicación cartográfica del (S.A.E). Sistema de ayuda a la explotación del TUS de Santander.....	183
Ilustración 23. Recorrido de la línea 12. Fuente: Aplicación cartográfica del (S.A.E). Sistema de ayuda a la explotación del TUS de Santander.....	186



ÍNDICE DE TABLAS

Tabla 1. Viajeros transportados y tasas de variación del transporte urbano en España.Fuente: Notas de Prensa del INE.	22
Tabla 2. Transporte urbano (autonbús) por comunidades autónomas en España.Fuente: Notas de Prensa del INE.	23
Tabla 3. Cambio de valores en x_i con respecto a y (Causelo 2012)	124
Tabla 4. Fuentes principales consultadas para la obtención de la información.	138
Tabla 5. Elaboración propia a partir de los datos obtenidos del Ministerio de Industria, Turismo y Comercio. Comparativa entre el coste real del litro de gasolina y el coste deflactado de forma semestral.....	165
Tabla 6. Elaboración propia a partir de los datos del TUS de Santander. Porcentajes de tipos de viaje según perfil.	171
Tabla 7. Estadísticas descriptivas de las variables contenidas en la base de datos mensual de las líneas objeto de estudio agregadas (N=144)	176
Tabla 8. Estadísticas descriptivas de las variables contenidas en la base de datos correspondiente a la línea 1. (N=144).....	180
Tabla 9. Estadísticas descriptivas de las variables contenidas en la base de datos correspondiente a la línea 3.(N=144).....	183
Tabla 10. Estadísticas descriptivas de las variables contenidas en la base de datos correspondiente a la línea 5C1. (N=144).....	185
Tabla 11. Estadísticas descriptivas de las variables contenidas en la base de datos correspondiente a la línea 12.(N=144).....	188
Tabla 12. Parámetros estimados de los modelos de demanda. Modelos teniendo en cuenta los datos agregados de las líneas objeto de estudio (p valor del test t aparece entre paréntesis, N=144).....	194
Tabla 13. Cálculo de las elasticidades de equilibrio para las variables estructurales del estudio. Estimaciones a corto plazo (SR) y a largo plazo (LR).	197
Tabla 14. Parámetros estimados de los modelos de demanda. Modelos teniendo en cuenta los datos de la línea 1 (p valor del test t aparece entre paréntesis, N=144).....	200
Tabla 15. Parámetros estimados de los modelos de demanda. Modelos teniendo en cuenta los datos de la línea 3 (p valor del test t aparece entre paréntesis, N=144).....	203
Tabla 16. Parámetros estimados de los modelos de demanda. Modelos teniendo en cuenta los datos de la línea 5C1 (p valor del test t aparece entre paréntesis, N=144).....	205
Tabla 17. Parámetros estimados de los modelos de demanda. Modelos teniendo en cuenta los datos de la línea 12 (p valor del test t aparece entre paréntesis, N=144).	207
Tabla 18. MSE de los resultados de los modelos frente a las observaciones de 2013.	208
Tabla 19. MSE de los resultados de los modelos frente a las observaciones de 2013 de la línea 1.....	209
Tabla 20. MSE de los resultados de los modelos frente a las observaciones de 2013 de la línea 3.....	210
Tabla 21. MSE de los resultados de los modelos frente a las observaciones de 2013 de la línea 5C1.....	211



Tabla 22. MSE de los resultados de los modelos frente a las observaciones de 2013 de la línea
12..... 211



CAPÍTULO 1

Introducción, objetivos y aportaciones



1. INTRODUCCIÓN, OBJETIVOS Y APORTACIONES

1.1. Motivación

La importante recesión experimentada por las economías del sur de Europa, entre ellas la española, ha tenido un fuerte impacto sobre múltiples dinámicas sociales, incluidas las relacionadas con el sistema de transporte. Variables macroeconómicas como el nivel de ingresos per cápita o la tasa de paro, relacionadas con el ciclo económico, pueden tener un fuerte impacto sobre la demanda de transporte. Una relación de gran interés para operadores y gestores de los sistemas de transporte público, ya que los cambios en la demanda pueden impactar a su vez sobre la recaudación, el grado de ocupación de los vehículos y en general sobre la calidad y la sostenibilidad del sistema en su conjunto (Litman 2014). El origen de este estudio parte del análisis del número de viajeros mensuales habidos en el Servicio de Transportes Urbanos de Santander entre los años 2001 y 2012, entendiendo como “Transporte Urbano” aquel que discurre íntegramente por suelo urbano o urbanizable o se dedica a comunicar entre sí núcleos urbanos diferentes situados dentro del mismo municipio. En el caso de los Transportes Urbanos de Santander (T.U.S) estos transportes se refieren única y exclusivamente al sistema de transporte municipal de autobuses, aquel destinado a transportar a todo tipo de pasajeros en autobuses o autocares que tienen un itinerario preestablecido con sujeción a calendarios y horarios prefijados, tomando a los pasajeros en paradas fijas dentro de un mismo término municipal (Estadística de Transporte de viajeros del INE 2010).

En ese periodo de tiempo se pudo comprobar cómo a partir de los primeros años de la serie el número total de viajeros experimentaba una caída progresiva que se iba acentuando con los años iniciales de la crisis económica y en los años finales de esta misma serie se denotaba un leve ascenso del número de viajeros con respecto a los años anteriores coincidiendo a su vez con una serie de medidas adoptadas en materia de tarifas y que beneficiaban a sectores de población como pensionistas, jóvenes, y miembros de familias numerosas.



A tenor de dichas coincidencias surgió el interés y la necesidad de intentar explicar las razones por las cuales el número de viajeros había experimentado dichos cambios y puesto que coincidían con momentos en los que la economía entraba en una fase más o menos recesiva se optó por iniciar una investigación basada en el estudio de diferentes variables tanto macroeconómicas como socioeconómicas y demográficas para el mismo periodo 2001-2012 y poder así obtener conclusiones que pudieran justificar qué variables eran las que podían influir en mayor o menor medida en la demanda de viajeros en el caso de Santander. Estas variables y su mayor o menor grado de incidencia sobre el número de viajeros es lo que denominaremos elasticidades y el objetivo de la investigación sería poder determinar cómo se comportan dichas elasticidades en la demanda de transporte público ante fases contractivas del ciclo económico. A su vez lo que se pretende es establecer un modelo econométrico que sea capaz de prever la demanda futura en función de los distintos valores que presenten las diferentes variables tenidas en cuenta.

Por otro lado el impacto de la crisis económica a partir de los años 2008-2009 en España provocó el descenso generalizado de los niveles de demanda en transporte público urbano, un descenso que en el caso de muchas ciudades españolas siguió en años sucesivos (véase Ilustración 1 y Tabla 1. Según el INE, en el mes de Noviembre de 2009 más de 263 millones de viajeros utilizaron el transporte urbano lo que suponía una caída de un 2,3% con respecto al mismo mes del año anterior. El transporte por autobús registró una bajada del 2,1% y el metropolitano de un 2,6%. Esas caídas pueden apreciarse también de forma generalizada en la mayor parte de comunidades autónomas (véase Tabla 2).

En el caso de la ciudad de Santander el impacto de la crisis económica medida en términos económicos como la tasa de paro, los precios de los combustibles, entre otros, también tuvo su inmediata repercusión en el número de viajeros/usuarios dando como resultado un descenso tanto del número de viajeros como de la recaudación en el servicio de transporte urbano. Sin embargo en el caso de Santander, a diferencia de la tendencia generalizada en muchas ciudades españolas, tras el primer año de impacto de la crisis el número de usuarios en un primer momento se estabilizó para después



recuperarse levemente en los años comprendidos entre 2010 y 2012 (véase Gráfico 1). Esa diferencia de tendencia con respecto a las cifras registradas en otras ciudades suscitó aún mayor interés por la investigación recopilada en esta tesis ya que había que profundizar aún más en las causas responsables de que el número de viajeros se recuperase a pesar de que los indicadores macroeconómicos principales iban empeorando paulatinamente.

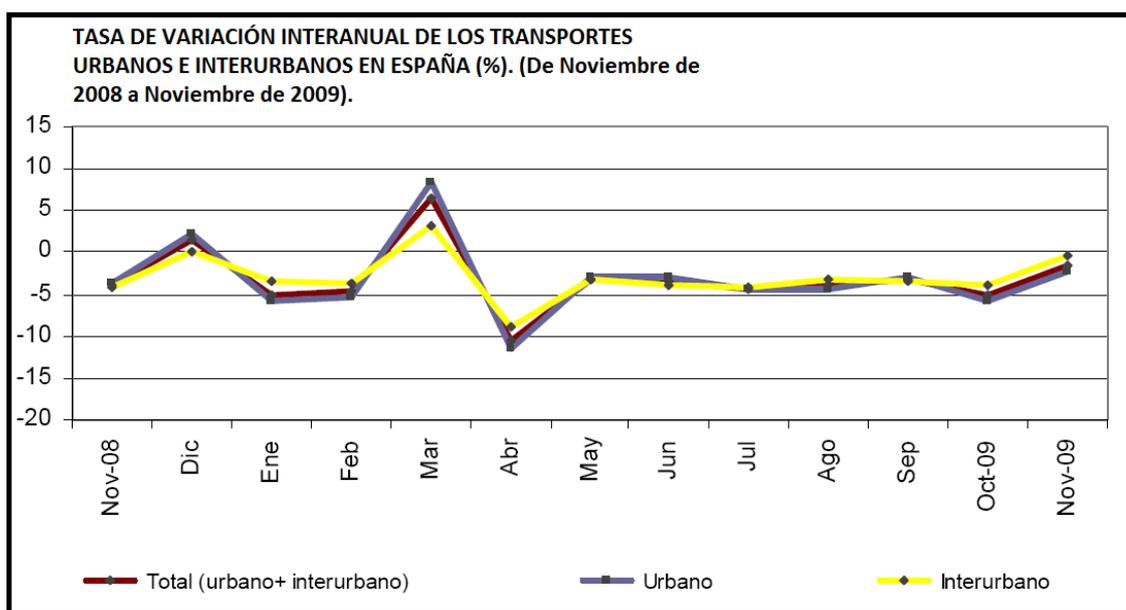


Ilustración 1. Fuente: Notas de Prensa del INE.

	Viajeros transportados (Millones)	Tasa de variación %	
		Interanual	Interanual acumulada
Transporte urbano			
TOTAL	263.343	-2,3	-3,8
Autobuses	161.069	-2,1	-2,8
- Regular	159.256	-2,1	-2,8
· Regular general	150.220	-2,3	-2,9
· Regular especial	9.036	1,0	-1,5
- Discrecional	1.813	-0,5	-2,4
Metropolitano	102.274	-2,6	-5,3

Tabla 1. Viajeros transportados y tasas de variación del transporte urbano en España. Fuente: Notas de Prensa del INE.



	Viajeros transportados Millones	Tasa de variación %	
		Interanual	Interanual acumulada
TOTAL	150.220	-2,3	-2,9
Andalucía	21.638	0,5	-2,3
Aragón	10.873	-3,3	-7,7
Asturias (Principado de)	3.540	-2,0	-3,1
Canarias	3.474	-4,2	-6,5
Castilla y León	6.485	-0,9	-1,9
Castilla-La Mancha	2.067	-1,7	-1,8
Cataluña	26.480	-3,4	-3,0
Comunitat Valenciana	11.746	-5,8	-6,6
Extremadura	1.218	0,8	-5,5
Galicia	5.374	2,7	-1,3
Madrid (Comunidad de)	40.754	-3,5	-0,8
Murcia (Región de)	1.604	2,8	-5,7
País Vasco	6.000	-1,3	-2,9
Otras ¹	8.967	1,5	-1,7

1 A fin de mantener el secreto estadístico, se agrupan las comunidades autónomas en las que se pudiera identificar el dato a nivel de empresa.

Tabla 2. Transporte urbano (autobús) por comunidades autónomas en España. Fuente: Notas de Prensa del INE.

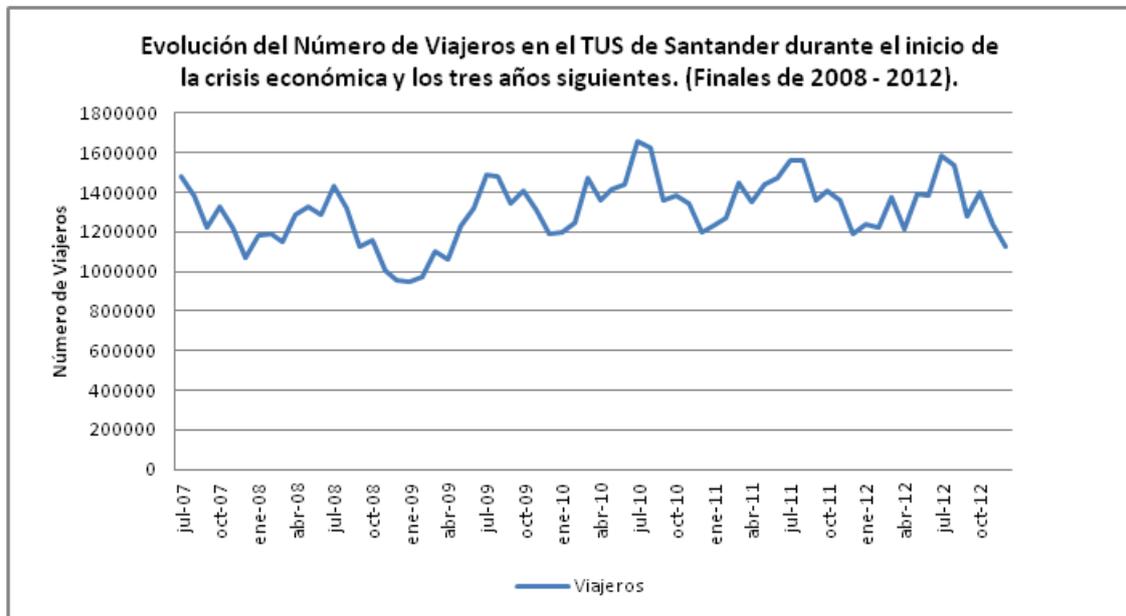


Gráfico 1. Elaboración propia a partir de los datos obtenidos del TUS (Transportes Urbanos de Santander)



En la literatura disponible se han realizado diversas estimaciones empíricas sobre la variación en dirección y magnitud de la demanda de transporte en bus ante distintas variables de cara a la ayuda a la toma de decisiones. Elasticidades bajas o no significativas implican variables que tienen una escasa incidencia sobre la demanda, mientras que elasticidades con magnitud elevada, positivas o negativas, señalan variables sobre las que se puede incidir de forma efectiva para incrementar la demanda de transporte público. A su vez en la literatura internacional los modelos de transportes propuestos son cada vez más eficaces y precisos en su capacidad de previsión. Las investigaciones se centran generalmente en zonas urbanas donde se incluyen principalmente los transportes públicos.

Las variables más estudiadas en la literatura han sido aquellas relacionadas con las características internas al sistema de transporte público como el precio y el nivel de servicio. Otras variables relacionadas con el entorno del sistema como la tasa de motorización de la población, el nivel de ingresos o el precio del combustible también han sido tenidas en cuenta en las investigaciones disponibles aunque en menor medida que las anteriores (Balcombe, Mackett et al. 2004). Finalmente variables macroeconómicas como la tasa de paro generalmente no se han considerado relevantes a pesar de que a priori pueden tener un impacto directo en la demanda de transporte público por efecto de las oscilaciones del ciclo económico.

Los ciclos económicos en periodización más simple se pueden descomponer en dos grandes fases. Una fase expansiva con crecimientos del Gross Domestic Product (GDP), y una fase recesiva con contracciones del GDP (Filardo 1994). La economía española tras experimentar una fase expansiva del ciclo en el periodo 1996 - 2007, entró en una fase recesiva desde 2008 sin haber alcanzado aún en 2013 los valores de GDP obtenidos en el punto álgido del ciclo (GDP de 100 en 2008 versus 94 en 2013). Esta dinámica ha venido asociada a su vez con un aumento acusado de la tasa de paro que desde el 8% de 2007 se situó en 2013 alrededor del 25% (INE 2013). Los ciclos económicos y sus cambios así como el estudio de los efectos económicos de los movimientos de personas y mercancías son también algunos de los principales objetivos de la denominada “Economía del Transporte”.



Por tanto en esta tesis se presenta una metodología que, partiendo de una serie de datos temporales entre los que se incluyen la demanda de transporte como de otras de tipo socio-económico, permite establecer, por un lado, qué variables son aquellas que influyen sobre la demanda de transporte público urbano, y por otro, en qué grado esas variables lo hacen, de forma que las fluctuaciones en las variables socio-económicas nos indiquen si eso supone un incremento o descenso de la demanda y cuanto.

1.2. Objetivos

En el presente estudio se propone evaluar mediante un modelo econométrico la influencia de variables macroeconómicas, del nivel de servicio y socioeconómicas en la demanda de transporte público. El objetivo principal es poder realizar estimaciones de la demanda futura en los servicios de transporte público a partir del diseño de un modelo de predicción de la demanda de viajes que permitan estimar cual puede ser el comportamiento de la misma y qué variables de distinto ámbito socioeconómico y de oferta del sistema, pueden influir más y por lo tanto ante cuales la demanda posee una mayor elasticidad. Para la consecución del principal objetivo de este estudio será necesario el desarrollo de los siguientes objetivos específicos:

- Construcción y agregación de datos históricos representados en forma de series temporales. Esos datos reflejan, por un lado, cual ha sido el número de viajeros mensuales y por otro lado distintas variables independientes de tipo socioeconómico.
- Ligado a la obtención del modelo que cuantifique con precisión la demanda de transporte público en función de los distintos cambios que puedan darse en el ámbito social y económico, se define otro de los objetivos a tener en cuenta como es el cálculo de las elasticidades de las variables de interés, para poder saber cuánto aumenta o disminuye la demanda en función de los cambios en esas variables de interés.



- Otro objetivo es el de analizar la demanda de transporte público diferenciándolo por líneas y comprobar cuales son aquellas que registran un mayor número de viajeros y cuales tienen un menor peso cuantitativo analizando sus series temporales de datos y tendencias.
- Realizar un análisis desagregado por líneas para después aplicar el modelo de predicción obtenido anteriormente y determinar los cambios de las elasticidades en las distintas líneas de transporte y así comprobar cómo influyen los cambios en las variables sociales y económicas en las distintas líneas ya que unas líneas pueden verse más afectadas que otras a determinados cambios en los ciclos económicos. Se trata de obtener las elasticidades para todas las líneas a partir del modelo general y de esa forma poder ver qué líneas se ajustan más a dicho modelo
- Comprobar la variabilidad de los distintos tipos de modelos utilizados (regresión lineal (RL) considerando correlación en los residuos, Regresión lineal en dos etapas y Regresión lineal en dos etapas considerando la correlación en los residuos.
- Atender a los diferentes comportamientos que puede adoptar la demanda en función de los cambios en los indicadores socio-económicos y que estos puedan ser medidos a través de un modelo de previsión que pueda servir de cara a los gestores del transporte público y así poder actuar sobre la planificación del mismo logrando una mayor eficiencia.
- Aplicar le metodología seguida al caso de la ciudad de Santander con datos socioeconómicos provenientes tanto de su área urbana como de datos originarios del propio sistema público de transportes urbano.



1.3. Estructura

Esta tesis se divide en seis capítulos a lo largo de los cuales se desarrollará una metodología cuyo resultado es el de obtener un modelo que sirva para prever la demanda total anual del transporte público y estimar también el modelo con la demanda desagregada por líneas que conforman la red de transporte público. A su vez se desarrollarán los cálculos de elasticidades de todas las variables. El presente capítulo (1) describe los objetivos y las aportaciones que esta tesis proporciona al estado del conocimiento. En el capítulo (2) se expone el “estado del arte” realizando un estudio sobre las diferentes metodologías existentes para poder hacer predicciones de demanda y hacer un seguimiento de las distintas variables que se han tenido en cuenta a la hora de desarrollar distintos modelos de previsión.

Asimismo se recopilan las bases sobre las cuales se asienta la metodología seguida en esta tesis como pueden ser el uso de los modelos de regresión, la aplicación de técnicas estadísticas y diferentes tipos de análisis tanto a través de series temporales como de los llamados “Datos de Panel”. Por otro lado en este capítulo (2) se hace mención también a las fases expansivas y recesivas que caracterizan a los ciclos económicos ya que enmarcan el contexto socioeconómico en diferentes momentos a lo largo del tiempo y en él se pueden medir el peso de variables económicas directamente relacionadas con la demanda de transporte público, como pueden ser por ejemplo las tarifas y el nivel de servicio. Tanto la tarifa como el nivel de servicio han sido los principales objetos de estudio en la gran mayoría de trabajos realizados que se han revisado de cara a iniciar la presente investigación.

En el capítulo (3) se muestra una exposición de una metodología general en la cual se ha desarrollado conceptualmente distintas técnicas matemáticas y estadísticas entre cuyos objetivos se encuentra también el de poder establecer predicciones de una determinada variable a partir de la relación existente con otras variables diferentes a lo largo de un periodo de tiempo determinado. Las herramientas que aparecen citadas en



este capítulo son las Series Temporales, los modelos de regresión (simple y múltiple) y los Datos de Panel.

En primer lugar se hace alusión al concepto de series temporales, su expresión matemática y objetivo de las mismas, para después adentrarse en los componentes que configuran una serie temporal y cómo deben de analizarse dichos componentes como pueden ser la estacionariedad, estacionalidad o la aleatoriedad de los residuos o variaciones accidentales y que procedimientos se ha de seguir en caso de que exista presencia de una u otra en nuestros datos de partida.

La segunda de las técnicas matemáticas citadas es la de modelos predictivos o modelos de regresión en los que a partir de una expresión lógico-matemática se pueden establecer una serie de predicciones para la variable que sea objeto de estudio o análisis, también llamada dependiente, y de cómo los valores de la misma fluctúan en función de los cambios que puedan experimentar el resto de variables a considerar, denominadas variables independientes o explicativas.

Asimismo dentro de los modelos de regresión se diferencia entre los modelos de Regresión Simple (MRLS) y los Modelos de Regresión Múltiple (MRLM). En el caso de los modelos de regresión tanto simple como múltiple se expresan tanto la formulación matemática como los requisitos e hipótesis que estos deben cumplir. Al igual que en los modelos de regresión simple, se deben de tener en cuenta ciertos requisitos e hipótesis que se deben de cumplir en cuanto a los datos para que los modelos obtenidos posean un alto grado de fiabilidad. Tras la fijación de requisitos e hipótesis se aborda la cuestión sobre la estimación de los modelos de regresión múltiple y de alguno de los métodos de estimación existentes como pueden ser, el “Método de Mínimos Cuadrados Ordinarios” (MCO), el “Método de Mínimos Cuadrados Ordinarios en dos etapas” (MC2E), el “Método de Mínimos Cuadrados Ordinarios considerando correlación en los residuos” o la “Estimación por Máxima Verosimilitud”.

Una vez enumerados los diferentes métodos de estimación se pasa a explicar cuál es el ajuste del modelo a partir de la estimación de mínimos cuadrados ordinarios, teniéndose en cuenta una serie de medidas de bondad de ajuste y sus respectivas expresiones matemáticas. Por último y en relación a los modelos de regresión se abordan las cuestiones referidas a la significación de los parámetros del modelo y al análisis de los residuos en el análisis de regresión.



La tercera y última herramienta matemática a la que se alude en este extenso capítulo es la de Datos de Panel. Se hace hincapié en la necesidad de hacer uso de este instrumento metodológico y se enumeran las ventajas así como las desventajas frente a otras herramientas. Al igual que en los casos de herramientas anteriores se procede a presentar la especificación general de un modelo de regresión con datos de panel así como las necesidades específicas que requieren su uso y los supuestos necesarios que los datos con los que contemos han de cumplir para que se pueda emplear esta técnica. Tras mencionar la posibilidad de elección de distintos tipos de datos de panel se define un método de estimación basado en el uso de variables instrumentales y que pueden resultar de gran ayuda de cara a conseguir unos estimadores más sólidos de los parámetros cuando estos son inconsistentes debido a la falta de variables significativas.

A continuación se explica la necesidad de seguir un protocolo de evaluación o de validación del modelo o modelos que se obtengan. Para tal validación se aboga por el uso de estadísticos simples como el Error Cuadrático Medio, el Error Absoluto Medio o el Sesgo.

Los dos últimos puntos del capítulo (3) serán los que hagan mención por un lado al tratamiento de variables monetarias que se contengan en una base de datos como la “Deflactación”, y por otro al cálculo de la “Elasticidad” de distintas variables, una elasticidad que se define como la capacidad de reacción de una variable en relación con cambios en otra.

El capítulo (4) consta de una descripción tanto del área de estudio como del origen y análisis de los datos manejados, todos ellos referidos tanto a la ciudad de Santander desde el punto de vista demográfico y territorial, como del servicio municipal de transportes, el TUS (Transportes Urbanos de Santander) del que se han obtenido datos relacionados con la demanda, la recaudación y demás variables relacionadas con el nivel de servicio. En este mismo capítulo se procede a explicar qué pasos se establecieron para construir la base de datos y los distintos tratamientos aplicados a la misma. Igualmente se procederá a hacer un análisis de las principales variables objeto de estudio y que se erigen como principales variables explicativas y que a la vez se



comparan con la demanda de viajeros que a su vez conforma la variable dependiente y de la cual se obtienen finalmente una serie de predicciones. Mención especial se hace al análisis de un momento concreto de la serie (2001-2012) como es el del periodo comprendido entre finales del año 2008 y principios del año 2009 en donde se produce el impacto de la crisis económica en el área de estudio. Finalmente se amplía el análisis de los datos incluyendo la variable demográfica y teniendo en cuenta las líneas del TUS más significativas.

En el capítulo (5) se atiende al análisis y discusión de los resultados con la metodología establecida y la aplicación del modelo teórico generado. Los resultados mostrarán, en un primer lugar, los referidos al modelo general o agregado en el que se tienen en cuenta la totalidad de las líneas objeto de estudio y por otro los resultados del modelo pero desagregado por líneas, tomándose aquellas que resultaron más significativas. A continuación y en último lugar dentro de este capítulo se acomete la validación del modelo y los resultados obtenidos de la misma.

En el capítulo (6), y último, se recogen las conclusiones finales sobre el estudio realizado y se proponen las futuras líneas de investigación.

1.4. Aportaciones

Las principales aportaciones de esta tesis son:

- El desarrollo de modelos de previsión de la demanda en transporte público teniendo en cuenta variables socio-económicas como el Paro y la Renta per Cápita, variables íntimamente ligadas a los procesos de crecimiento o decrecimiento económico y que complementan a las ya tenidas en cuenta en otros modelos de previsión como la tarifa y el nivel de servicio. De esta manera se contribuye a los escasos trabajos hallados en la literatura internacional.
- El presente estudio aporta como novedad estimar la elasticidad de las variables más relevantes a la hora de considerar el impacto del ciclo económico en la demanda de transporte como es la tasa de paro. Se ha profundizado en la



comparación de las elasticidades a corto y largo plazo para tener en cuenta el efecto temporal del cambio en las variables.

- La inclusión de otras variables socio-económicas como objeto de análisis y que también tienen relación con los ciclos expansivos o recesivos de la economía como pueden ser los datos turísticos (número de visitantes y pernoctaciones) y el precio medio mensual de los combustibles muy ligado a las fluctuaciones del precio del petróleo en el mercado (Gasoil y gasolina).
- Considerar la presencia de endogeneidad entre variables (entre demanda y oferta de transporte), que generalmente no se tiene en cuenta en estudios de previsión de la demanda de transportes. Para ello se han empleado modelos de Regresión lineal en dos etapas.
- Se ha ampliado el análisis anterior considerando la posible existencia de correlación en los residuos mediante el uso de modelos de Regresión lineal en dos etapas considerando la correlación en los residuos. La presentación de la variabilidad de los distintos tipos de modelos utilizados (regresión lineal (RL) considerando
- Se han aplicado los modelos resultantes a un caso real como es el de la ciudad de Santander, logrando validar modelos, contrastar los resultados de los modelos con los datos reales y obteniendo unos resultados que pueden servir para conseguir una gestión más eficiente del sistema municipal de transportes en Santander.





CAPÍTULO 2

Estado del arte



2. ESTADO DEL ARTE

El término “Econometría” se cree que fue acuñado por primera vez por el economista noruego Ragnar Frisch, fundador junto al economista Jon Tinbergen de la conocida como “Econometric Society” en 1933, por lo que podemos establecer el origen de la econometría como disciplina a comienzos de la década de los 30 del pasado siglo XX. Esta disciplina “... surge por la necesidad de cuantificar los modelos teóricos, para tratar de contrastar si las diversas teorías económicas se cumplen cuando se ven enfrentados a los datos”. (Reyes, Morcillo et al. 1997)

En la actualidad se define la econometría como la disciplina responsable del estudio combinado de modelos económicos, estadística matemática y datos de naturaleza económica. A su vez dentro de lo que es la econometría se distinguen la “Teoría Econométrica” por un lado y la “Econometría Aplicada” por otro. La primera de ellas hace referencia tanto al desarrollo de herramientas y métodos, como del estudio de las propiedades que poseen los métodos econométricos. Podría afirmarse que la teoría econométrica concierne al ámbito de la estadística.

La econometría aplicada es un concepto que explica la confección de modelos económicos cuantitativos y el empleo de métodos econométricos a estos modelos recurriendo a datos económicos. El uso de modelos econométricos resulta de gran interés de cara a poder cuantificar las relaciones existentes entre las distintas variables de interés que tengamos y de cómo estas influyen en otras que pretendamos predecir. Por ello cabe destacar las principales utilidades que los modelos econométricos poseen, análisis estructural, Predicción y Simulación o evaluación de políticas (Medina Moral 2002):

- Análisis Estructural: cuantificación de las relaciones que entre la serie de tiempo analizada ha tenido lugar entre las variables implicadas, a través del conocimiento del signo y valor de los parámetros estimados. Es decir, sirve para conocer como inciden en la variable dependiente alteraciones de las variables explicativas.



- Predicción: Teniendo unos valores a futuro para las variables independientes, y sabiendo la expresión matemática que relaciona las variables independientes y la variable dependiente o a estimar, es posible predecir los valores que tomará a futuro la variable objeto de estudio.
- Simulación o evaluación de políticas: Los efectos que experimenta la variable dependiente cuando se actúa de una u otra forma aplicando distintas estrategias sobre las variables explicativas. Un ejemplo sería el de poder comprobar los efectos en la demanda de transporte público una vez se decide incrementar en cierto porcentaje la tarifa del mismo.

El uso generalizado de estos modelos econométricos puede aplicarse a distintas formas de recolección de datos con las que poder calcular estimaciones. Entre las distintas formas de recolección de datos se pueden señalar como las principales tenidas en cuenta para la elaboración de esta tesis, las series temporales, los modelos de regresión lineal (simple y múltiple) y los datos de panel. Sería a comienzo de los años 70 cuando, G.E.P. Box, profesor de Estadística de la Universidad de Wisconsin, y G.M. Jenkins, profesor de Ingeniería de Sistemas de la Universidad de Lancaster, introdujeron una pequeña revolución en el enfoque del análisis de series temporales, en sus trabajos sobre el comportamiento de la contaminación en la bahía de San Francisco, con el propósito de establecer mejores mecanismos de pronóstico y control, Box and Jenkins (1976), en el que describen la metodología, se convirtió rápidamente en un clásico, y sus procedimientos se utilizan ampliamente desde entonces en diferentes ramas de la ciencia, conociéndose como modelos ARIMA y también como modelos Box–Jenkins.

La metodología Box-Jenkins para la obtención de predicciones es diferente de otros muchos métodos debido a que no supone ningún patrón particular en los datos históricos de las series que se pretenden pronosticar. Este método se apoya en un enfoque iterativo para reconocer un modelo posible a partir de una clase general de modelos. Posteriormente, el modelo seleccionado se compara con los datos históricos que tenemos para ver si describe la serie con exactitud. El modelo está bien ajustado si



los residuos son por lo general de poca entidad, están distribuidos aleatoriamente y no contienen información útil. Si el modelo especificado no es satisfactorio, el proceso se repite usando un nuevo modelo trazado para optimizar el original. Este procedimiento iterativo continúa hasta que se encuentra un modelo satisfactorio. En ese momento, el modelo se considera útil para poder establecer previsiones.

Tomando como base la econometría y los modelos que de ella pueden resultar se ha acometido una revisión de aquellos en los que se han utilizado variables relacionadas con el ámbito del transporte, principalmente del transporte urbano. El número de estudios que han aportado evidencia sobre la influencia de distintas variables en la demanda de transporte público son numerosos. La gran mayoría de los trabajos se han centrado en la influencia de dos variables: la tarifa y el nivel de servicio. En el ámbito español De Rus (1990) estimó un modelo de regresión con datos de panel como instrumento para varias ciudades españolas en el periodo 1980-1988. Las elasticidades estimadas para la tarifa y el nivel de servicio mostraron como la demanda de transporte, es decir, el número de viajeros, era sensible a estos factores. El autor calculó una elasticidad en el modelo estático de -0.3% para la tarifa y de 0.71% para el nivel de servicio, sugiriendo que se podría aumentar el número de usuarios de los sistemas de transporte público incrementando a la vez las tarifas y los niveles de servicio.

De Rus (1990) además de estimar modelos a partir del uso de datos de panel estableció también una clasificación en tres categorías de los distintos métodos en lo que agrupación de datos en forma de series temporales se refiere. Esas tres categorías las dividen en una metodología basada en proximidad temporal, otra metodología basada en la representación y una tercera metodología basada en modelos. La idea de los dos últimos métodos es convertir las series de tiempo en datos estáticos o parámetros de modelos, para luego aplicar directamente métodos de agrupamiento desarrollados para manipular los datos estáticos para completar las tareas de agrupamiento. A su vez en su trabajo introducen las bases de agrupamiento de series temporales y suministran una revisión de trabajos realizados en años posteriores.



En otro artículo (Waters et al. 1992) examinan las elasticidades del precio en la demanda de transporte. Además de revisar las estimaciones de elasticidad empíricas, tanto para la demanda de carga y pasajeros, su documento también se ocupa de cuestiones teóricas y empíricas para estimar elasticidades de la demanda de transporte. En una segunda parte se refieren principalmente a las estimaciones empíricas de las elasticidades de la demanda de sistema de transporte público y el uso del automóvil. En su estudio recogen una bibliografía y hacen una revisión de la literatura en la que se recogen las características que afectan a la demanda de transporte. Posteriormente pasan a definir el concepto de “elasticidad” y establecen diferencias en cuanto a su cálculo atendiendo a los diferentes periodos de tiempo, tanto a corto como a largo plazo. Finalmente muestran un resumen de estimaciones de elasticidades, a corto y largo plazo, del uso del automóvil en países como Estados Unidos, Australia y Reino Unido.

Mención aparte supone el cálculo de elasticidades de demanda en transporte urbano a partir del uso de diferentes tipos de datos en forma de series temporales, datos de sección transversal y datos agrupados. En el caso de elasticidades vinculadas al uso del automóvil se estimaron valores comprendidos entre -0.09 a -0.52. Aunque las estimaciones de la elasticidad a largo plazo son en general más altas, la diferencia no parece ser significativa, aunque esto puede reflejar el hecho de que pocos estudios desarrollan verdaderos modelos a largo plazo y que a su vez tengan en cuenta los cambios en la propiedad de vehículos y la elección de ubicación. En su análisis obtienen también elasticidades derivadas de la demanda de transporte teniendo en cuenta el motivo de viaje, manifestando ser estos por negocios o trabajo, por ocio o placer o por otros motivos desconocidos. Para esos tres motivos de viaje se especifican las elasticidades en función de la forma en la que se han tratado los datos bien a partir de series temporales, bien a partir de datos de sección transversal u otros.

Para ese mismo año Goodwin (1992) realizó una revisión de más de 50 estudios existentes sobre la influencia de la tarifa como variable en la demanda de transporte en bus, llegando a obtener una elasticidad media de -0.41%. Esta elasticidad fue por lo tanto superior al valor de -0.3% que hasta entonces se consideraba como de referencia en el



ámbito británico derivado de los estudios de Webster and Bly (1981). Adicionalmente el autor estimó que a largo plazo las elasticidades son entre dos y tres veces más elevadas que las estimadas a corto plazo.

En otra serie de estudios se estimaron también las elasticidades de variables relacionadas con determinantes procedentes del entorno del sistema de transporte público como los ingresos de la población, el hecho de tener automóvil y el precio de la gasolina. Precisamente Goodwin (1992) fue quien recopilara tres estudios con datos sobre la elasticidad de la demanda en transporte público respecto a los precios de la gasolina con un resultado medio de 0.34. En relación al cálculo de elasticidades teniendo en cuenta las mismas variables de Goodwin, un estudio adicional de Paulley, Balcombe et al. (2006) recopiló varios resultados de las elasticidades del ingreso y la posesión de coche. En el caso del nivel de ingresos, si el transporte público en bus se considera un bien normal, su elasticidad debería ser positiva. Sin embargo diversos estudios han encontrado elasticidades negativas que señalan que los usuarios lo consideran más bien como un bien inferior. A esta conclusión llegaría Paulley, Balcombe et al. (2006) para su estudio desarrollado en Gran Bretaña con estimaciones de la elasticidad cuyos valores oscilaban entre -0.5 y -1 siendo estas calculadas a largo plazo. Uno de los problemas que se plantearon al respecto fue que si se introducían de forma simultánea las variables de ingreso y de posesión de coche en el modelo, podían aparecer problemas de colinealidad, de ahí que en algunos estudios se utilizaran sólo una de ellas.

Hasta donde llega el conocimiento del autor, los estudios que han utilizado únicamente la posesión de coche en la función de demanda son más escasos que los que lo han hecho con el precio de la gasolina, pero han encontrado en general elasticidades también negativas. Este problema es común en los estudios ya que además no sólo pueden darse problemas de colinealidad entre los ingresos y la posesión de coche sino también de los precios de la gasolina con otras variables incluidas en el estudio.

Previo a los trabajos desarrollados por Goodwin, Matas (1991) calculó estimaciones sobre elasticidades para poder evaluar cuales eran las variables determinantes, no de variables exógenas a los sistemas de transporte como son el coste del combustible o el



nivel de ingresos, sino de la demanda en transporte urbano. La estimación de su modelo permite calcular las elasticidades de la demanda respecto a los principales factores que la determinan como son: la tarifa y el tiempo de viaje dividido en distintas categorías (de viaje, de espera y de segunda espera). Para abordar el análisis de los determinantes de la demanda de transporte urbano, estima una función de demanda de tipo probabilístico en el que los usuarios se enfrentan a dos alternativas de transporte, el público y el privado. En estos modelos no se tienen en cuenta el contexto socio-económico en el que se encuentran inmersos los usuarios sino que se centra más en los atributos de los medios de transporte y en las características socio-económicas de los individuos sin analizar el periodo económico existente bien sea expansivo (periodos de bonanza económica) o contractivo (periodos de crisis económica).

En su estudio se capta el efecto de la renta a partir de una relación funcional simplificada consistente en segmentar la muestra. Es decir, dividiendo el conjunto de observaciones en varias sub-muestras según su nivel de renta, de forma que los coeficientes a estimar toman distintos valores según la categoría a la que pertenece cada individuo. En su cálculo de la elasticidad para la variables “tarifa” obtiene como resultado el valor (-0,15) demostrando así que la variable tiempo de viaje (-0,35), tiempo de espera (-0,25) y tiempo de segunda espera (-0,69), es mucho más sensible y presenta una mayor elasticidad que la “tarifa”.

Volviendo a los estudios de comienzos del siglo XXI, que ahondaban en el análisis de la demanda de transporte, como los (Dargay and Hanly 2002), hicieron un análisis de la demanda de los servicios de transporte de autobuses locales en Inglaterra. Su estudio se basaba en el uso de un modelo dinámico en el que se tenían en cuenta la población como variable demográfica interpretándose así la demanda per cápita o volumen de pasajeros por unidad de población. De igual forma la dimensión demográfica se combinó con otras variables como la tarifaria, el nivel de renta y variables del nivel de servicio.

Para calcular las estimaciones a partir de un modelo de regresión, recurrieron a la elaboración de una serie de datos históricos a la que aplicaron el uso de distintas técnicas e instrumentos como eran las series temporales y los datos de sección



transversal. Los datos sobre los cuales aplicaron dichas técnicas se referían a datos de demanda de una serie de condados ingleses. La ventaja en el uso de este tipo de técnicas, como pueden ser las series temporales, radica en que sus resultados se ajustan mejor a la búsqueda de soluciones de gestión con fines políticos y en ese sentido el método no sólo es capaz de generar estimaciones para un corto periodo de tiempo sino que también es capaz de hacer predicciones a más largo plazo aportando así una información relevante de cara a los políticos responsables de gestionar el transporte.

Los resultados obtenidos en ese estudio realizado para varios condados ingleses, y obtenidos partiendo de una base de datos compuesta de (414) observaciones, apuntaron a que la demanda era relativamente sensible a las tarifas y con una amplia variación de las elasticidades. Su objetivo se centraba en la obtención de estimaciones de las elasticidades relativas a la tarifa y que sirvieran a los gestores políticos para hacer determinados cálculos que pudieran estar orientados al cambio en la demanda de transportes en autobús a nivel nacional. Para llegar a obtener estimaciones de la elasticidad de la tarifa, se apoyaron en datos obtenidos de diferentes fuentes, por un lado datos reales sobre la demanda o volumen de pasajeros, y por otra datos provenientes de la realización de encuestas de preferencia declarada. La estimación de las elasticidades en las tarifas de autobús se basó en la información aportada por los operadores de transporte, que proporcionaron datos para un periodo comprendido entre los años 1986 y 1996. Esos datos de demanda los agregaron de forma que disponían de la demanda por condado y contaban también con información referida a los ingresos de la población de dichos condados.

Las elasticidades de tarifas son calculadas sobre la base de la econometría dinámica con modelos relativos a la demanda de autobús bus per cápita (todos los viajes), los ingresos per cápita, las tarifas reales de autobús (ingreso promedio por viaje), de nivel de servicio (kilómetros por vehículo), los costes derivados del uso del automóvil, y las variables demográficas. En su metodología distinguen a su vez entre los impactos en la variación de la elasticidad de las tarifas tanto a corto plazo como a largo plazo.



En sus resultados señalan que los valores más probables de la elasticidad de precios en Inglaterra en su conjunto se sitúan alrededor de -0,4 a corto plazo y a -0,9 a largo plazo por lo que se sugiere que las elasticidades a largo plazo son aproximadamente el doble. Por otro lado concluyeron que modelos con elasticidades de tarifas separadas para cada uno de los condados objeto de estudio son estadísticamente más relevantes que las especificaciones en la que la elasticidad de tarifas se toma de la misma manera por igual en todos los condados. Los resultados de los modelos más flexibles muestran una variación considerable en la elasticidad de tarifas si se tienen en cuenta los condados de forma desagregada con unos valores comprendidos entre 0 y -3.0 a largo plazo.

Dos años más tarde Bresson, Dargay et al. (2004) estimaron mediante *generalized method of moments (GMM)* modelos para una muestra con datos de 62 áreas de transporte público francesas en el periodo 1975 – 1995. Los autores calcularon una elasticidad de la tarifa de -0.65 a largo plazo. Estos autores también realizaron estimaciones sobre los kilómetros recorridos por los autobuses con elasticidad positiva calculada de 0.68 a largo plazo.

Dos años más tarde Balcombe, Mackett et al. (2004) elaboraron una guía para el Transport Research Laboratory (TRL) sobre los factores que inciden en la demanda de transporte público en superficie. Esta guía actualizaba los resultados presentados en los años ochenta por Webster and Bly (1981). Los autores recopilaron los estudios existentes y establecieron que la elasticidad de la tarifa en bus estaba a corto plazo en el entorno del -0.4% y a largo plazo en torno al -1%. Bresson, Dargay et al. (2004) igualmente realizaron estimaciones sobre las elasticidades del nivel de ingreso calculándolas en -0.23 a largo plazo. Estos resultados pueden consultarse también en un artículo posterior (Paulley, Balcombe et al. 2006).

Investigaciones más recientes como la de Holmgren (2007) han realizado meta-análisis a partir de los resultados de diversos estudios para explicar la variación en las elasticidades obtenidas en las distintas investigaciones. El autor obtuvo elasticidades medias en un conjunto de 81 estudios de -0.38 para la tarifa. El valor medio para el nivel de servicio en un total de 58 estudios fue de 0.72. El autor recomendó el tratamiento



endógeno de la variable del nivel de servicio, ya que en caso de no tratar esta endogeneidad los parámetros estimados en esa y otras variables podrían estar sesgados. Teniendo en cuenta esta endogeneidad el autor calculó las elasticidades en -0.75 para la tarifa y 1.05 para el nivel de servicio, notablemente superiores por lo tanto a las calculadas sin tener en cuenta la dependencia del nivel de servicio de la demanda de transporte. Por otra parte Holmgren (2007) en su meta – análisis recomendó también utilizar este tipo de variables en las funciones de demanda para evitar parámetros sesgados en las variables relacionadas con la tarifa o el nivel de servicio. Según Holmgren las estimaciones realizadas con los estudios disponibles a corto plazo para Europa presentaron valores de las elasticidades de -0.62 para el ingreso, 0.4 para el precio de la gasolina y -1.48 para la posesión de coche. Estas cifras según el autor señalan que la demanda de transporte público en Europa funciona como un bien inferior y por lo tanto puede seguir experimentando caídas en el futuro si continúan aumentando los ingresos de la población y los niveles de posesión de coche.

Tras lo anteriormente citado queda demostrado que el número de estudios que han aportado evidencia sobre la importancia en el tratamiento del agrupamiento de datos de series de tiempo también denominado series temporales son también numerosos. Como consecuencia del rápido desarrollo de las tecnologías de la información y la comunicación (TIC) la obtención y recopilación de todo tipo de datos han hecho que aparezcan problemas en distintos campos como pueden ser los de la economía, el medio ambiente, la salud o los transportes por lo que sus tendencias, aclaración y sobre todo su predicción se han convertido en objeto de estudio de vital importancia. Hasta tal punto adquiere importancia el uso de las nuevas técnicas econométricas que hasta se tiende a desarrollar métodos de evaluación de las diferentes técnicas.

Rico (2006) obtiene una serie de resultados de una investigación cuya meta es evaluar las técnicas de análisis de series temporales, como una posible herramienta en la elaboración de pronósticos de la demanda de transporte de carga en aeropuertos mexicanos y que bien podría servir para aplicarlo en demandas de otros modos de transporte. Para ello utiliza una técnica conocida como descomposición factorial, consistente en la desagregación y tratamiento individual de los tres puntos



característicos en una serie temporal como son, la tendencia, la variación estacional y la variación aleatoria. La técnica de descomposición factorial fue aplicada a series temporales de la carga transportada a través de los aeropuertos de México, Guadalajara y Monterrey separando la información en dos categorías: doméstica e internacional.

El análisis practicado confirmó la validez de los supuestos iniciales ya que en todos los casos tenidos en cuenta se hallaron valores característicos de tendencia, variación estacional y variación aleatoria. Finalmente se concluye que la técnica de análisis de series temporales empleada en el estudio es la adecuada para poder obtener un modelo que sea capaz de prever el comportamiento de la demanda de transporte de carga aérea en aeropuertos siempre y cuando se tengan en cuenta los supuestos básicos de los modelos tendenciales, y que los horizontes de pronóstico sean lo suficientemente cortos, con el fin de que se pueda esperar que no existan cambios significativos en los factores sociales, económicos y tecnológicos que condicionan a los sistemas de transporte.

Tras las investigaciones de Rico cabe resaltar a Sánchez .P (2008) quien centró su análisis de las series temporales teniendo en cuenta los llamados cambios estructurales. Estos cambios estructurales están caracterizados por su capacidad de perturbación permanente sobre la componente determinística de la serie temporal lo cual incide en la pérdida de peso en la identificación del modelo y en las pruebas. El estudio abordado al respecto surge desde las siguientes perspectivas: primero, muchos autores coinciden en afirmar que no tener en cuenta un cambio estructural en la representación del modelo de la serie da lugar a una especificación incorrecta ya que no explica de forma apropiada su comportamiento. Segundo: los anteriormente citados cambios estructurales influyen permanentemente sobre la componente determinística llegando a afectar a la estructura de autocorrelación de la serie temporal y por lo tanto dificulta la identificación del modelo a través de sus patrones de autocorrelación, lo cual implica la obtención de parámetros inconsistentes. Tercero: un cambio estructural al final de un periodo de medición afecta de manera importante tanto a la especificación del modelo, como de su capacidad predictiva, debido a que su presencia conlleva efectos en la



función de autocorrelación y los intervalos de predicción ante cambios estructurales pueden llegar a engañar pues dichos sucesos aumentan la variación estimada de la serie.

Además del cálculo de elasticidades haciendo uso de series temporales como herramienta para la obtención de estimaciones, existen más herramientas o técnicas que pueden resultar válidas también de cara a obtener estimaciones y calcular las elasticidades. Otra herramienta que sirve para poder hacer un óptimo análisis de los datos de las variables con las que partimos sería un análisis con datos de panel. De cara a poder establecer toda una serie de estimaciones a partir del uso de modelos de datos de panel, señalar que desde años atrás y hasta la fecha el uso de la metodología a partir de datos de panel ha sido una técnica econométrica ampliamente utilizada.

(Nijkamp and Pepping 1998) evaluaron los factores clave que influyen en los usuarios de los medios de transporte mediante el análisis comparativo de los diferentes valores de la elasticidad de la demanda para el transporte, que son los utilizados en algunos de los diferentes Estados miembros de la Unión Europea. El punto de partida sobre el cual apoyaron su investigación se focalizó en el cálculo de elasticidades en países como Noruega, Finlandia, los Países Bajos y el Reino Unido. En un documento identifican un conjunto de factores potenciales que causan varianzas entre los resultados de diferentes estudios. Por otro lado presentan un pormenorizado análisis en conjunto de las causas de las variaciones entre los valores de elasticidad en los cuatro países estudiados. El análisis que llevan a cabo está basado en la literatura existente hasta entonces, lo que indica que la diferencia entre información agregada, y la investigación basada en métodos empíricos mediante el uso de modelos de elección desglosados, así como supuestos del modelo, explican la varianza de la elasticidad de los valores obtenidos en los estudios.

De las distintas técnicas econométricas mencionadas, en todas ellas el cálculo de estimaciones se realiza a partir de los conocidos modelos de regresión lineal. Los modelos de regresión lineal son los más utilizados de cara a poder predecir los valores de una variable cuantitativa o dependiente a partir de los valores de otra variable explicativa también cuantitativa (modelo de regresión lineal simple). Cuando se cuenta



con más de una variable explicativa o independiente que puede ayudar también a predecir el valor de la variable independiente, entonces hablamos de modelos de regresión lineal múltiple.

El análisis de regresión es una técnica estadística fundamental a la hora de modelizar y sirve para explicar el comportamiento de dos medias encontrando una fórmula, que permita llegar a una conclusión dada una cantidad, cual sea, de una variable y cómo se comportaría la otra variable. Por ello el planteamiento del modelo debe ser fundamental y tener una base teórica. En materia de investigación el análisis de regresión es utilizado para hacer previsiones de un amplio espectro de fenómenos. A su vez el análisis de regresión lleva asociados una serie de procedimientos de diagnóstico como el análisis de los residuos, el análisis de la correlación entre variables, puntos de influencia,...etc, que proporcionan pistas acerca de cuestiones a tener en cuenta como son la estabilidad e idoneidad del análisis.

(Batarce, M. et al., 2005) explican el tratamiento de los efectos espaciales en la calibración de modelos lineales de demanda de transporte y especifican como para modelos lineales calibrados existe correlación en las variables independientes, dilación de la variable a predecir o dependiente y auto correlación en los residuos. Dado que un mismo conjunto de datos puede ser resultado de la presencia de uno, otro o ambos fenómenos, es necesario que el análisis de la presencia de auto correlación se haga cuidadosamente.

Afirman que cuando se detecta que la variable dependiente está autocorrelacionada entonces es necesario recalibrar el modelo utilizando como alternativa el método de Máxima Verosimilitud. En su trabajo presentan un marco de análisis econométrico además de diferentes test desarrollados con el objetivo de identificar la presencia de efectos espaciales. Aplicarán sus cálculos sobre una serie de datos de viajes a nivel interurbano en la macrozona del sur de Chile y a escala urbana en la ciudad de Santiago. En su análisis resaltan que para el uso de modelos lineales calibrados aparece correlación en las variables independientes tenidas en cuenta y rezago de la variable dependiente así como autocorrelación en los residuos.



Velásquez, J. M. (2009) examinó la relación existente entre la demanda de Transmilenio (Sistema de transporte público en Bogotá) y las tarifas y costes de los distintos modos de transporte que compiten con el servicio municipal de transportes. A su vez amplía la relación existente a distintas variables económicas y demográficas mediante el análisis de un modelo de regresión lineal múltiple. En sus resultados concluye como la demanda crece principalmente como resultado del incremento en materia económica y el progresivo aumento demográfico. Los datos con los que contó fueron analizados utilizando varios modelos de regresión lineal múltiple y se probaron diferentes combinaciones de las variables independientes que ayudaran a explicar la demanda (variable dependiente). Un apunte del estudio indica que el modelo fue elaborado con la finalidad de comprobar cuáles y cómo estas variables exógenas (independientes o explicativas) incidían en la demanda de TransMilenio durante el periodo de análisis y, por lo tanto, no se trataba de una propuesta de modelación encaminada a lograr pronósticos de la demanda futura.

(Palomares, G. et al., 2010) estudiaron la conexión existente entre variables de tipo urbanístico y la demanda de transporte público. Para analizar la relación entre dichas variables y la demanda de transporte trabajaron con las estaciones de metro de Madrid, tratando de explicar el número de viajeros que accede a pie a dichas estaciones en función de las características urbanísticas de su entorno próximo. Para poder definir esta relación desarrollan un modelo de estimación apoyado sobre un sistema de Información Geográfica (GIS) necesario para delimitar el área de influencia de las estaciones y obtener sus características urbanísticas y sobre un modelo de regresión lineal múltiple con el objetivo de poder explicar el número de entradas por estación en función de su entorno próximo. En su modelo resultan como variables de interés la densidad del empleo, la diversidad de usos del suelo o la densidad viaria de cara a explicar el número de entradas generadas en las distintas estaciones por lo que su utilidad es doble. Por un lado permite prever la demanda de futuras estaciones a construir a diferentes modos de transporte público y por otro valorar el peso que la morfología urbana tiene sobre el uso del transporte público.



Cuando surge la necesidad de ampliar o actuar sobre las redes de transporte público, por parte de la parte encargada de la planificación existe la necesidad de concluir una primera evaluación rápida y suficientemente ajustada de la demanda de viajes que podría producirse. De esta necesidad se impulsó el desarrollo de una metodología simple y de respuesta rápida cuyo propósito se encaminaba a la estimación de las entradas de viajeros en las estaciones que componen la red de metro. El estudio se realizó con datos obtenidos a partir de una encuesta domiciliaria realizada en Noviembre del año 2004 por el Consorcio Regional de Transportes de Madrid (CRTM).

El número total de viajeros que accedieron en cada una de las estaciones de la red se erige como la variable dependiente y por tanto la variable a predecir (variable dependiente). La consecución de los datos de acceso a dichas estaciones se obtuvo a partir del aforo de accesos a la red de metro realizado por el CRTM durante noviembre de 2004, mismo mes en el que se realizó la encuesta domiciliaria. Posteriormente y una vez obtenidas el resto de variables independientes, se obtuvo un modelo de regresión lineal múltiple a partir del paquete estadístico SPSS 15. Una vez calcularon las variables urbanísticas referidas al entorno de la estación y teniendo en cuenta las variables internas que caracterizan cada una de las estaciones (variables independientes), obtuvieron un modelo de regresión lineal múltiple con el objetivo de responder a las entradas totales en cada una de las estaciones (variable dependiente). Los modelos de regresión lineal múltiple son uno de los métodos más utilizados a la hora de hacer análisis explicativos y/o predictivos.

En su análisis de resultados mostraron los coeficientes de correlación bivariada (Pearson) obtenidos entre cada una de las variables independientes consideradas y las entradas de viajeros. La bondad de ajuste de la ecuación de regresión obtenida incluyendo variables propias a las estaciones y variables urbanísticas de su entorno alcanzaron unos coeficientes de determinación R^2 y R^2 ajustado de 0,68 y 0,67, respectivamente, con un valor del estadístico F de 44,95, con una muy elevada significación estadística.



Las distintas técnicas y métodos explicados cuyo objetivo se centra en la obtención de modelos que sirvan para explicar el comportamiento de distintas variables a partir de otras y de cómo las fluctuaciones de unas repercuten sobre las demás, sería interesante que pudieran tener en cuenta también los contexto socio-económicos de los lugares o áreas en donde esas técnicas y métodos se apliquen. Por ejemplo, si queremos obtener modelos de previsión de la demanda en transporte público para un determinado entorno urbano o ciudad sería de gran interés poder incluir variables de ámbito socioeconómico como puedan ser los datos de desempleo, el precio de los carburantes, el nivel de ingresos o el número y tipología de las distintas ramas de actividad presentes en al área de estudio. Además de esas variables sería importante también añadir otras también de tipo económico pero de mayor escala como pueden ser variables macroeconómicas a distintas escalas (Nacional, autonómico, regional, provincial...) como valores del PIB (Producto Interior Bruto) o el IPC. De esa forma se podrían reconocer en los resultados de los modelos y por tanto en las estimaciones cómo pueden afectar las distintas fases de crecimiento (periodos de recuperación económica) y decrecimiento económico (periodos de crisis económica).

En cuanto al estudio de las distintas etapas de crecimiento y decrecimiento económico y por tanto el análisis de las distintas coyunturas económicas y sus distintos ciclos, éstas fueron señaladas como punto de interés, en los años 30 del Siglo XX como así lo recogen (Kydland and Prescott 1991), donde hacen un repaso de la teoría econométrica del ciclo económico desde sus orígenes y haciendo especial hincapié en el llamado “Enfoque de Equilibrio General” para más adelante introducirse en el “Enfoque de la Teoría del Ciclo Económico”. En su análisis centran su atención en la importancia de una elaboración sistemática de las cuentas nacionales de ingreso y producto, junto con series temporales de “inputs” y “outputs” añadidos desde el sector empresarial. Junto a esas mediciones de “inputs” y “outputs” agregados desde el sector empresarial, esas cuentas se conforman como series temporales agregadas que de alguna manera definen el campo de la macroeconomía, la cual está relacionada con las fluctuaciones del crecimiento y de los ciclos económicos.



En su artículo describen los modelos de equilibrio general definiéndolos como un “marco teórico en el que existe una situación estable y consistente del sector de las economías domésticas y del sector empresarial”. Incluyen también al llamado sector público, sujeto a su propia restricción presupuestaria. Otro de sus puntos de interés es el relativo a los años 80 del siglo XX haciendo alusión a los denominados “Shocks tecnológicos como fuente de fluctuaciones económicas”. En la década de los 80 la fuente de “shocks” se convirtió en objeto de interés como fuente principal de fluctuaciones. Desde esa perspectiva se llega a plantear en cuánto hubiera fluctuado la economía norteamericana de la posguerra en el caso de que los shocks tecnológicos hubieran sido la única fuente de fluctuaciones agregadas.

Para poder llegar a abordar esa cuestión, se centró el estudio en la selección de un cierto modelo de economía ampliando el modelo de crecimiento neoclásico e introduciendo el ocio como argumento de la función de utilidad de las economías domésticas. Puesto que en una gran medida las fluctuaciones del ciclo económico se explican por variaciones en el factor trabajo, es importante, añadir ese elemento.

En el artículo de (Figuroa 2005) se explica como la globalización y la adopción de políticas liberales han repercutido en el funcionamiento de la ciudad y por ende también en el comportamiento de los transportes urbanos. En el caso de las ciudades la influencia ha derivado en reconocidos procesos de expansión urbana así como de diferentes aspectos económicos han sido adoptados por parte de los sistemas de transporte. En su artículo Figuroa hace referencia también a como los cambios en las economías latinoamericanas tuvieron como consecuencia cambios en los comportamientos de la demanda tanto de transportes como de pasajeros.

En su artículo se resalta como la práctica del control tarifario, en los contextos de economías inflacionarias que fueron la principal característica durante muchos años de la “región latinoamericana”, produjo un comportamiento cíclico del transporte colectivo urbano, expresado en distintos periodos períodos de alta rentabilidad y expansión de la oferta, por lo general unidos a los momentos de reajuste tarifario, y momentos de baja



rentabilidad y contracción de la economía, vinculados a los períodos de reducción del valor de la tarifa en términos reales.

Se afirma que la crisis económica tuvo como consecuencia, en un primer momento, una caída en la movilidad y, por tanto, en la demanda de viajes, que se expresó en una reducción de las tasas de ocupación de los buses, lo que fue agravado por el crecimiento de la oferta de transporte al ingresar nuevos operadores al servicio. Como respuesta a este hecho la situación fue compensada por parte de los operadores bajo dos modalidades: la primera fue la reducción de los costes de operación y la segunda fue la de reajustes tarifarios. Esta segunda posibilidad, que inicialmente estaba más controlada por las autoridades, se pudo llevar a cabo, en especial en los casos en los que la mayor liberalización del sector permitía manejar más flexiblemente el precio del servicio. Un ejemplo de la caída en la demanda de transporte fueron Bogotá entre los años 1985 y 1995, Santiago entre los años 1978 y 1989 y Buenos Aires los años comprendidos entre 1992 y 1996. En los tres casos, la caída en las tasas de ocupación se debió a la caída de la demanda y al aumento de la flota de autobuses. Por último Figueroa introduce también el uso creciente del coche privado y su asociación con la degradación del sistema de transporte público como razón de impacto importante que explica la reducción de la demanda en transporte público

Otras variables relacionadas con el entorno económico como la tasa de paro han sido escasamente tenidas en cuenta por los estudios anteriores. Hasta donde llega el conocimiento del autor, el presente estudio aporta como novedad estimar la elasticidad de una variable relevante a la hora de considerar el impacto del ciclo económico en la demanda de transporte como es la tasa de paro, así como la de tener en cuenta distintas variables del contexto socio económico del área de estudio y variables macroeconómicas a escala autonómica. El conocimiento de este efecto puede ser importante para los gestores y los operadores de transporte público dado que las oscilaciones en la demanda pueden impactar notablemente en la sostenibilidad del sistema a través de cambios en la recaudación, la ocupación medida de los vehículos y otras.



CAPÍTULO 3

Metodología



3. METODOLOGÍA

En este capítulo se describe la metodología desarrollada para, bajo un mismo marco y de forma integral, modelizar la demanda de viajeros en transporte público de un área urbana incorporando variables macroeconómicas y así poder cuantificar el impacto de algunos factores económicos como la Tasa de Paro/Desempleo y la Renta Per Cápita. En esta sección se presenta la formulación teórica de los modelos de demanda de transporte público en bus incorporando variables macroeconómicas para cuantificar el impacto del ciclo económico como la tasa de paro y el ingreso per cápita.

La metodología utilizada ha consistido en la estimación de las elasticidades de equilibrio, a corto y a largo plazo de las variables de interés mediante un modelo de regresión log – log considerando la presencia de autocorrelación en los residuos y la endogeneidad de la oferta de transporte. Por otro lado para calcular las elasticidades de las variables de interés se ha estimado en primer lugar un modelo estático de regresión de tipo log – log. Dada esta forma funcional, los parámetros estimados pueden interpretarse directamente como elasticidades, es decir, como el cambio porcentual de equilibrio de la variable dependiente ante los cambios porcentuales en las variables independientes.

Los modelos han sido estimados mediante Mínimos Cuadrados Ordinarios (OLS) y Mínimos Cuadrados Ordinarios en dos etapas (2SLS) de cara a tener en cuenta en estos últimos la endogeneidad del nivel de servicio. Los modelos han considerado además la existencia de autocorrelación en los residuos aplicando la estimación propuesta por Prais – Winsten (Prais and Winsten 1954).

Las diferentes etapas seguidas para llegar a crear este modelo de previsión de la demanda de viajeros en transporte público urbano ha consistido en una serie de pasos secuenciales, los cuales se resumen en: la obtención de un modelo econométrico de previsión a partir de la construcción de una base de datos en la que para un periodo de tiempo determinado de doce años y a escala mensual. Como paso previo a la obtención de la expresión matemática del modelo se procedió, tras recabar toda la información necesaria, a fijar cual debía de ser la variable dependiente y por tanto objeto de



predicción que en el caso de este estudio se corresponde con el número de viajeros. Posteriormente se recopilaron los datos relativos al resto de variables independientes o explicativas. Una vez construida la base de datos, se hizo un tratamiento de la misma.

Dicho tratamiento se resume en la deflactación de aquellas variables de ámbito monetario y en la desestacionalización de los datos tras detectarse la presencia de estacionalidad en la serie de datos. Dicha desestacionalización se llevó a cabo introduciendo una serie de variables denominadas “Dummy” cuyo objetivo es el de corregir los efectos que los ciclos estacionales tienen (Ejemplo entre verano e invierno).

Igualmente se procedió a hacer un análisis de los datos con el objetivo de poder verificar la estacionariedad o no de la serie así como el análisis de la tendencia, estacionalidad y su componente aleatoria. Una vez dirimidos esos puntos se pasó a hacer una especificación del modelo y su posterior validación cuyos resultados se compararon con datos reales.

3.1. Las Series Temporales

En primer lugar se comienza por la “construcción” de una Serie Temporal. En esta primera tarea el objetivo es el de configurar una base de datos en la que se agrupe toda una serie de variables relacionadas con el nivel de servicio del transporte público y otras de tipo socioeconómico. El motivo es el poder recoger una serie de observaciones ordenadas y ligadas a un periodo de tiempo determinado. A la sucesión de observaciones de una o varias variables ordenadas en el tiempo se le conoce como “Serie Temporal”, “Serie Cronológica”, “Serie Histórica” o “Serie de Tiempo”. Irgoin (2011).

En este sentido podemos afirmar que una serie temporal es una sucesión de observaciones cuantitativas de uno o varios fenómenos ordenadas en el tiempo. El objetivo principal con el diseño de esta serie temporal es el de poder explicar el valor



que toma, en un momento determinado del tiempo, un fenómeno determinado (económico, socio-económico...etc) que muestra dependencia temporal.

Para poder estimar ese valor se precisa de la recogida de datos de una serie de variables y su evolución a lo largo del tiempo y poder determinar y explotar el patrón de regularidad que muestran dichos datos. En cuanto a la elección de las variables si todas las variables independientes son discretas, variables cualitativas o cuantitativas que sólo pueden adoptar un número finito de valores distintos, el modelo logit es el más idóneo debido a que permite una interpretación más clara de los coeficientes.

Para poder crear un modelo a partir de una serie temporal , lo único que se necesita es la información muestral de la variable a estudiar y analizar. Si se desea pronosticar el comportamiento de una variable temporal Y_t , un modelo de series temporales puede plantearse a partir de la siguiente ecuación o función:

$$Y_t = f(y_{t-1}, y_{t-2}, \dots) \quad (1)$$

El primer objetivo del análisis de una serie temporal consiste en obtener un modelo estadístico que describa correctamente el origen de dicha serie, de manera que las implicaciones teóricas del modelo resulten compatibles con las pautas muestrales observadas en la serie temporal (Chatfield 2013). Posteriormente, el modelo elaborado a partir de la serie temporal considerada puede utilizarse para:

- Prever la evolución futura de dicha serie.
- Describir la evolución observada de dicha serie así como sus relaciones y dinámicas.

Desde el momento en que los valores de una serie temporal van unidos a instantes del tiempo, entonces, podemos afirmar que el análisis de una serie implica el tratamiento conjunto de dos variables, siendo una de ellas la serie temporal y la otra los intervalos o



instantes del tiempo sobre los cuales se han realizado las observaciones, días, semanas, meses,...etc.

Se debe indicar que esa observación sincronizada de la variable en el tiempo supone que los valores de la misma han de permanecer perfectamente ordenados, de igual manera que los intervalos del tiempo también lo están. Otro aspecto a tener en cuenta es que los intervalos de tiempo utilizados siempre deben de tener la misma amplitud. Es decir, si se tienen en cuenta datos sobre el número de viajeros en un servicio de transporte público, estos deberán ser siempre o bien diarios o mensuales, trimestrales, anuales..etc, pero lo que nunca deberá procederse es a intentar trabajar con una serie que mezcle datos semanales con mensuales o referidos a cualquier otro periodo de tiempo. Esta condición lleva asociada la idea de “homogeneidad”. Para que el estudio y análisis de una serie temporal nos conduzca a conclusiones verídicas no basta con utilizar las técnicas idóneas, sino que será imprescindible que esos datos sean comparables y no lo serán nunca si no son homogéneos. En el análisis de series temporales existe la posibilidad de que la serie no siga un comportamiento regular a lo largo de todo el periodo objeto de estudio y ejemplos típicos de falta de homogeneidad incluyen datos atípicos, cambios de nivel, cambios en la variabilidad (dispersión) y cambios en la estructura del modelo (Peña Sánchez de Rivera 1990).

Con el objeto de predecir los valores de una variable cuantitativa y que denominaremos como variable dependiente o a estimar, existen otras formas de hacerlo, como por ejemplo a través de los modelos de regresión lineal múltiple (MRLM) muy utilizados de cara a poder predecir los valores de este tipo de variables a partir de los valores de otra variable explicativa también cuantitativa.

En el presente estudio se opta por el uso de modelos de regresión múltiple con datos de series temporales puesto que plantea una serie de ventajas frente a otros modelos que utilizan datos de sección transversal o cruzada. En comparación a dichos modelos, las ventajas de los modelos de regresión lineal múltiple:

- En una muestra con datos de sección transversal o cruzada, distintos agentes económicos de una naturaleza similar proporcionan información solicitada para



un mismo periodo de tiempo mientras que los modelos de regresión múltiple a partir del uso de datos de series temporales proporciona información acerca de una unidad económica, que bien podría ser un país o una empresa, a lo largo de un periodo de tiempo, algo de gran importancia puesto que de lo que se trata es de poder predecir, a partir de los datos generados a lo largo de una secuencia temporal, de cara al futuro y no para una misma franja temporal.

- Los modelos econométricos permiten conocer la forma en que la variable de interés se relaciona con las variables exógenas; éste puede ser el objetivo principal del análisis (Ejemplo, estimación de una elasticidad).

Las Series Temporales por tanto son un tipo de modelo econométrico donde los datos pueden corresponder a los valores de una variable en el tiempo pudiendo tener estos una frecuencia, diaria, semanal, mensual o anual. (Medina Moral 2002)

El esquema que se sigue para abordar con fiabilidad un análisis de series temporales será como se muestra a continuación (véase Ilustración 2).

La “Especificación” supone la primera etapa a la hora de elaborar un modelo econométrico. En esta etapa se considerarán cuatro elementos importantes como son: el modelo económico, el modelo econométrico, los supuestos estadísticos del modelo y los datos. El primero de esos elementos se especifica mediante el uso de la teoría económica. En otros supuestos, la teoría económica se utiliza menos formalmente en la construcción de un modelo económico. Tras la obtención del modelo económico es necesario transformar dicho modelo en un modelo econométrico aplicando distintas funciones de consumo. Para poder convertirlo en un modelo econométrico antes se deben de cumplir algunos requisitos como el de especificar la forma matemática de la función y la inclusión de una variable denominada de perturbación aleatoria o error. Por lo tanto el segundo de los requisitos sería la introducción del término de error en la ecuación.

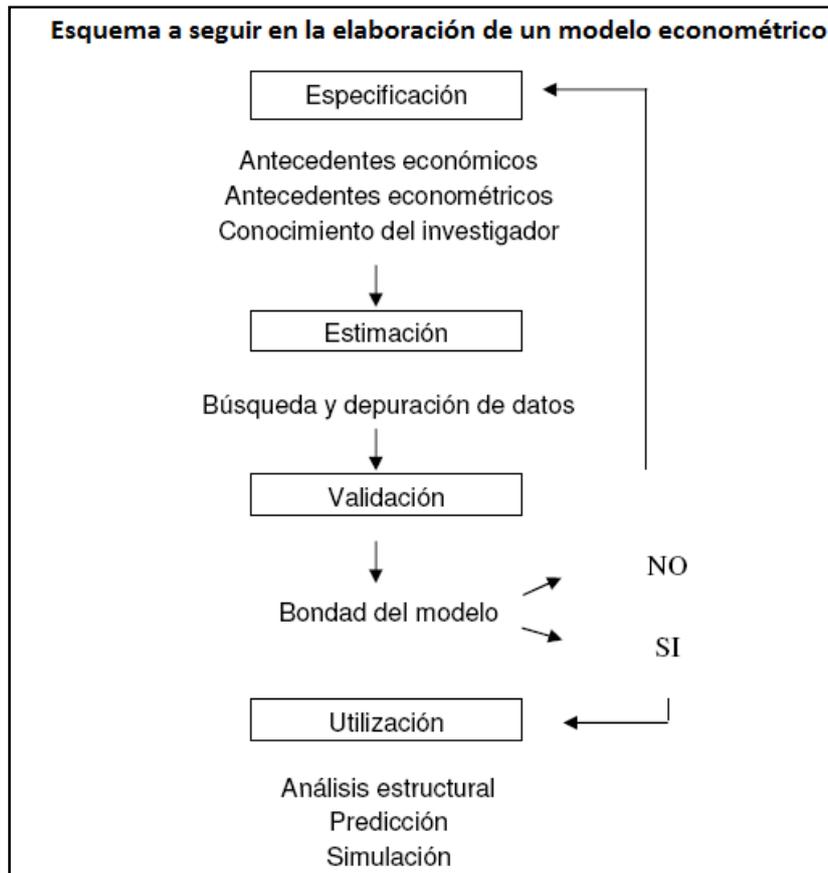


Ilustración 2 Esquema general del análisis de una serie temporal (Medina Moral 2002)

La “Estimación” es la segunda de las etapas a seguir de cara a la consecución de un modelo econométrico. En la estimación se obtienen los valores numéricos de los coeficientes de un modelo econométrico. Para completar esta etapa es necesario disponer de un conjunto de observaciones de todas las variables observables que aparecen en el modelo econométrico especificado, y, por otro lado, es casi obligado el elegir el método de estimación apropiado, teniendo en cuenta las implicaciones de esta elección en las propiedades estadísticas de los estimadores de los coeficientes. La diferencia entre un estimador y una estimación debe quedar meridianamente clara. Un estimador es el resultado de aplicar un método de estimación a una especificación econométrica. Por otra parte, una estimación consiste en la obtención de un valor numérico de un estimador para una muestra dada.



Tras la Estimación, el tercero de los pasos a seguir es el de la “Validación”. En esta tercera etapa se procede a realizar una evaluación de los resultados de manera que se pueda comprobar si las estimaciones obtenidas son aceptables, tanto por la teoría económica como desde el punto de vista estadístico. Se revisa, por un lado, si las estimaciones de los parámetros del modelo tienen los signos y magnitudes esperados, es decir, si satisfacen las limitaciones establecidas por la teoría económica. Desde el punto de vista estadístico, por otro lado, se desarrollan una serie de contrastes estadísticos sobre la “significatividad” de los parámetros del modelo en los que se aplican los supuestos estadísticos formulados en la etapa de especificación.

Resulta de vital importancia verificar si los supuestos estadísticos del modelo econométrico se cumplen, aunque se debe tener en cuenta que no todos los supuestos son contrastables. El no cumplimiento de alguno de estos supuestos conlleva, en general, la aplicación de otros métodos de estimación, que permitan obtener estimadores que cuenten con las mejores propiedades estadísticas posibles. Una forma de poder confirmar si el modelo es adecuado para hacer predicciones es utilizar el modelo fuera del período muestral, y después comparar los valores predichos de la variable endógena con los valores realmente observados. Por ejemplo, si el periodo muestral con el que se cuenta es de una serie anual de doce años en la que se recogen los datos de demanda en transporte público (2001-2012) de un determinado sistema de transporte, se procede a estimar los datos para el año 2013 y una vez obtenidos los resultados se comparan estos con los datos reales de ese mismo año y si el porcentaje de error no es demasiado elevado puede tenerse en cuenta como válido (Uriel and Rosat 1997).

Los tres tipos de datos que pueden utilizarse en la estimación de un modelo econométrico son fundamentalmente tres:

- Las Series Temporales.
- Los Datos de Corte Transversal.
- Los Datos de Panel.



En las series temporales, los datos constituyen observaciones de una variable a lo largo del tiempo. Por ejemplo: magnitudes de las cuentas nacionales, como el consumo, las importaciones, ingresos, etc. El orden cronológico de las observaciones arroja información potencialmente importante. En consecuencia, en una serie temporal la ordenación de las observaciones es por tanto relevante.

En los datos de corte transversal se dispone de una observación por sujeto y hacen alusión a un punto determinado en el tiempo. En la mayor parte de los estudios, los individuos encuestados son personas (por ejemplo, en la Encuesta de Población Activa (EPA), más de 100000 personas son entrevistadas cada trimestre), hogares (por ejemplo, la Encuesta de Presupuestos Familiares), empresas (por ejemplo, la Encuesta de Empresas Industriales) u otros agentes económicos. Las encuestas son una fuente típica para datos de corte transversal.

En un número importante de estudios econométricos basados en datos de corte transversal el tamaño muestral es bastante elevado. En los datos de corte transversal, las observaciones deben de ser obtenidas mediante un muestreo aleatorio, lo que implica que las observaciones sean independientes entre sí. El orden de las observaciones en los datos de corte transversal, a diferencia que en las series temporales, no supone tanta importancia de cara al análisis econométrico. El problema de los datos de corte transversal radica en que si los datos no se obtienen con una muestra aleatoria, entonces se manifiesta un problema de selección muestral (Uriel and Rosat 1997).

Los datos de panel, denominados también datos longitudinales, consisten en observaciones de corte transversal repetidas a lo largo del tiempo. Así pues, los datos de panel combinan elementos de datos de corte transversal y de series temporales. Estos conjuntos de datos consisten en un conjunto de individuos encuestados repetidamente a lo largo del tiempo. En la modelización se adopta generalmente el supuesto de que los individuos son independientes entre sí, pero que, para un individuo dado, las observaciones a lo largo del tiempo son mutuamente dependientes. Por lo tanto, el orden dentro de un corte transversal de un conjunto de datos panel no es tan



importante, pero el orden en la dimensión temporal es francamente importante. Cuando no se tiene en cuenta el tiempo en datos de panel, se entiende que estamos utilizando datos de corte transversal agrupados (pooled) (Uriel and Rosat 1997).

3.1.1. Tipos de variables y datos de la modelización econométrica

Se define como “variable” al concepto generalmente económico que se pretende analizar. Habitualmente se emplean variables cuantitativas, es decir, aquellas cuyos valores se expresan de forma numérica.

No obstante, también existe la posibilidad de añadir en el modelo econométrico información cualitativa, siempre que esta información pueda representarse de forma cuantitativa. Con respecto a este tipo de variables es aconsejable conocer las siguientes posibilidades que se le pueden presentar al investigador:

- Variables proxies (proxy): son variables cercanas a la variables objeto de análisis. Por ejemplo, si se requiere el uso de una variable que tenga en cuenta el nivel cultural de un país (variable cualitativa) se puede recurrir como variable proxy el número de bibliotecas existentes en un país, que si bien no contiene el concepto exacto que se pretende medir, si se aproxima al mismo pudiéndonos dar una referencia.
- Variables ficticias (dummy): Estas representan la incidencia que tiene sobre la variable endógena o dependiente y que es objeto de análisis, un determinado fenómeno cualitativo. Generalmente a dicha variable dummy se le atribuyen dos valores arbitrarios en función de si se da o no esa cualidad en un fenómeno. Así, se le puede asignar el valor 1 si tiene lugar un determinado fenómeno y 0 en caso contrario. Por ejemplo para una serie diaria en la que se quiere analizar los días de lluvia. Se opta por elegir el valor 1 para los días en los que llovió y un 0 para aquellos con ausencia de la misma. Estas variables “Dummy” resultan de gran ayuda



de cara a eliminar efectos estacionales de una serie de datos que esta pueda tener. (Diferencias cíclicas entre meses de verano e invierno).

Las variables ficticias se pueden incluir tanto en modelos de series temporales como en modelos de corte transversal. Por ejemplo, si se requiere un análisis de la venta de motocicletas a nivel de un centro urbano, se puede añadir una variable ficticia que valga 0 cuando la provincia no tenga metro como medio alternativo de transporte y 1 cuando si cuente con él.

Por otro lado se define como “datos” los diferentes valores que adquiere una variable. Esos datos pueden corresponderse con valores de una determinada variable a lo largo de un periodo de tiempo (Serie Temporal) o a valores para distintos sujetos en un momento concreto (datos de corte transversal).

A su vez cabe señalar, las distintas formas en las que un modelo puede representarse, siendo objeto necesariamente de transformaciones de los datos originales (tasas de crecimiento, diferencias, logaritmos) o incluso a un cierto tratamiento previo de los mismos (eliminación de tendencia, eliminación de estacionalidad, etc...) (Medina Moral 2002).

3.1.2. Análisis de una Serie Temporal

El objetivo del análisis de una serie temporal es doble. Por un lado se busca explicar las variaciones observadas en la serie en el pasado, tratando de determinar si responden a un determinado patrón de comportamiento. Y por otro, si se consigue definir ese patrón o modelo, se intentará pronosticar el comportamiento futuro de la misma. Para lograr este doble objetivo se utiliza una metodología bastante consolidada, según la cual se admite que la serie temporal es una función del tiempo: $Y_t = f(t)$.



Bajo esta premisa, la serie sería una variable dependiente (por ejemplo número de viajeros en un sistema de transporte público a lo largo de un periodo de tiempo determinado), y el tiempo una independiente o explicativa. Sin embargo, es necesario dejar bien claro que el tiempo, en sí, no es una variable explicativa, es simplemente el “soporte” o escenario en el que se realiza o tiene lugar la serie temporal. El tiempo no sirve para explicar el comportamiento de la serie. A esta forma de abordar el estudio de una serie temporal se le conoce como enfoque clásico, frente al causal, según el cual, cualquier serie, como variable que es, puede ser explicada por otra u otras series.

El primer paso de cara a iniciar el análisis de una serie temporal es hacer una representación gráfica de los datos. Para ello se utilizará un sistema cartesiano en el que los valores o periodos de tiempo puedan ser trasladados al eje de abscisas y los valores de la serie Y_t , sean ubicados en el eje de ordenadas. Mediante la representación gráfica se pueden detectar las características más llamativas de una serie, tales como el movimiento a largo plazo, la amplitud de las oscilaciones, la posible existencia de ciclos, los puntos de ruptura, la presencia de valores atípicos o anómalos, etc. El resultado es un diagrama de dispersión, con la característica de que el eje de abscisas se reserva siempre a la misma variable: El tiempo. (véase Ilustración 3).

El análisis de Series Temporales se desarrolla habitualmente usando dos tipos de técnicas:

- Los métodos tradicionales, como los de regresión, alisado exponencial, medias móviles, empleados si el número de observaciones es pequeño (hasta unos 40).
- Los métodos de Box y Jenkins o utilización de modelos ARIMA, si se dispone de suficiente material estadístico.



Ilustración 3. Representación gráfica de una serie de datos a lo largo de una serie de tiempo. Sánchez Fernández Jesús. “Introducción a la Estadística Empresarial. Cap.4. Series Temporales.

3.1.3. Componentes de una Serie Temporal

Tras la representación gráfica de los datos, no es posible tratar cualquier tipo de serie temporal, ya que en cada instante tenemos una variable con una distribución diferente, de la que sólo observamos un dato, por lo que es necesario imponer una serie de condiciones a la serie temporal. Una serie temporal es el resultado de cuatro componentes: tendencia, variaciones estacionales o estacionalidad, variaciones cíclicas y variaciones residuales o accidentales.

- La Tendencia: Entendida como el comportamiento o movimiento suave de la serie a largo plazo. Para poder detectarla es necesario que la serie conste de un número de observaciones elevado, a lo largo de muchos años, para que se puede determinar si la serie muestra un movimiento a largo plazo que responda a una determinada ley de crecimiento, decrecimiento o estabilidad.



En relación a la tendencia si la serie se comporta de una forma estable y tanto la media como la variabilidad son constantes a lo largo del tiempo entonces la serie es estacionaria. Esto significa que no se aprecian aumentos o disminuciones sistemáticos de sus valores. Esta característica, la de la estacionariedad significa que la serie posee un grado importante de validez de cara a poder hacer posteriores previsiones. Si los coeficientes de la Autocorrelación Simple (AC) no decaen o decrecen rápidamente entonces hay indicios que la serie es no estacionaria.

Por otro lado la identificación y estimación de los modelos de series temporales se desarrollaron precisamente para procesos estacionarios pues con series estacionarias podemos obtener predicciones de una forma menos complicada (Jiménez and Giménez 2000).

- Las Variaciones estacionales o Estacionalidad: Son movimientos de la serie que se repiten de forma periódica. La razón de estas variaciones pueden tener su explicación en causas de tipo climatológico o de ordenación del tiempo (los días de la semana condicionan el comportamiento de ciertas series). Estos movimientos o variaciones que se producen con cierta periodicidad, en general menor o igual a un año, es lo que se denomina estacionalidad.
- Las Variaciones Cíclicas: Son variaciones que se producen sin un periodo fijo y se caracterizan por ser ciclos superiores a un año ya que dichas variaciones suelen ir asociadas a fluctuaciones en los ciclos económicos y sociales de ahí que el estudio y análisis de estas variaciones sean fundamentales si se pretende analizar el impacto de coyunturas económicas. Un marcado carácter económico, suele ser el resultado de la sucesión de las fases expansivas y recesivas de la economía. Estas variaciones suponen movimientos a medio plazo, superiores a un año y que se suelen repetir de forma periódica aunque poseen una menor regularidad si las comparamos con las variaciones estacionales.



Es una componente que presenta ciertas dificultades si lo que se quiere es aislarla, pues con frecuencia es posible que lleguen a superponerse ciclos con periodos o amplitudes diferentes. Denominaremos “amplitud” al número de años que dura un ciclo completo. En ocasiones resulta complicado recoger esos ciclos, bien porque el periodo de tiempo que pretendemos analizar es demasiado corto o porque realmente no existen ciclos. De ahí que de cara a la predicción de determinadas variables en las que los ciclos económicos pueden llegar a ser importantes se necesita que la serie temporal contenga un periodo de tiempo lo suficientemente largo ya que los ciclos de la economía suelen producirse a lo largo de lustros o incluso décadas.

- Las Variaciones Accidentales: Esta componente no sigue ningún patrón reconocible de comportamiento, sino que se producen debido a hechos aleatorios o fortuitos que influyen de forma aislada y no permanente en la serie temporal. Estos factores suelen ir asociados a eventos extraordinarios.

La interacción de estas cuatro componentes será la que genere y caracterice la serie temporal.

3.1.3.1. Análisis de la Tendencia

De cara a aislar la “Tendencia” de una serie, se pueden utilizar distintos métodos, pero con independencia del que se utilice, el periodo de información necesario debe de ser lo suficientemente amplio para evitar identificar como “tendencia” otros movimientos distintos de la serie. (Véase Ilustración 4).

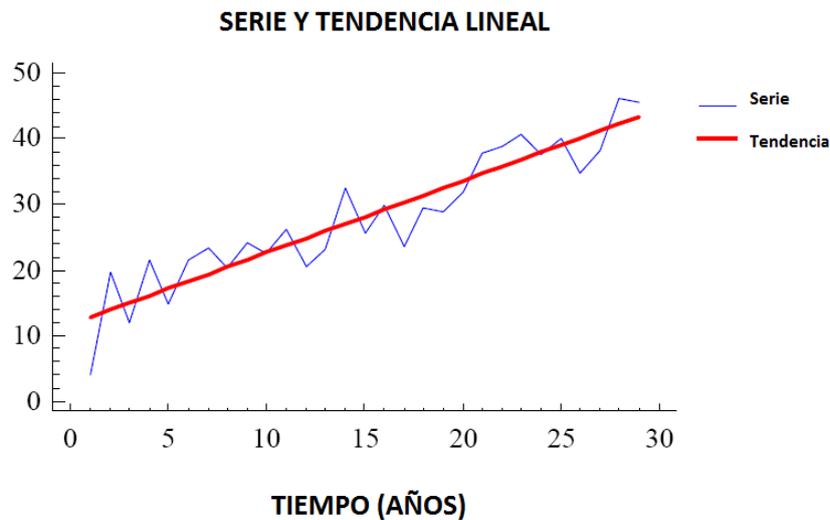


Ilustración 4. Representación gráfica de la línea de tendencia sobre la serie de datos temporales expresada gráficamente.

Los métodos más recurrentes para detectar y anular la tendencia de una serie se basan en la aplicación de filtros a los datos. Un filtro no es más que una función matemática que aplicada a los valores de la serie genera una nueva serie con unas características determinadas (Aguirre Jaime 1994). Entre esos filtros encontramos las “medias móviles”. El método de “medias móviles” consiste en promediar observaciones consecutivas de una serie temporal para “suavizar” el patrón que siguen los datos y así poder hacer predicciones a corto plazo (de Rivera 2005). Una media móvil se calcula, para cada punto, como un promedio del mismo número de valores a cada lado de ese punto. Así una media móvil de tres puntos se calcula como:

$$m(x_t) = \frac{x_{t-1} + x_t + x_{t+1}}{3} \quad (2)$$

Existen otros procedimientos para extraer la tendencia, como el método del ajuste analítico, método analítico de mínimos cuadrados, etc. Una clase de filtro, que resulta de bastante utilidad para eliminar la tendencia, se basa en aplicar diferencias a la serie hasta transformarla en estacionaria. Una diferencia de primer orden se obtiene restando dos valores contiguos:



$$\nabla x_{t+1} = x_{t+1} - x_t \quad (3)$$

Volviendo a diferenciar esa serie, restando los nuevos valores consecutivos obtenemos una nueva serie más suavizada.

$$\nabla^2 x_{t+2} = \nabla x_{t+2} - \nabla x_{t+1} \quad (4)$$

Una vez que se aplica un proceso clásico de descomposición mediante, por ejemplo, un procedimiento de medias móviles a los datos se obtiene las siguientes series (véase Ilustración 5).

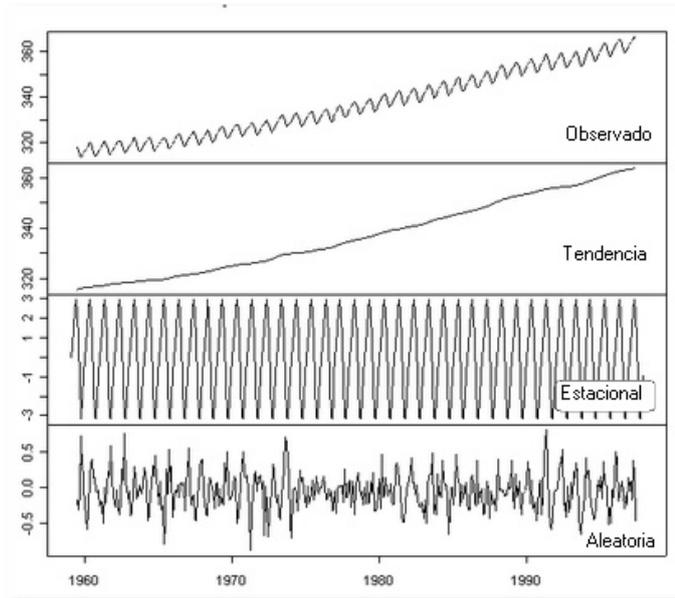


Ilustración 5. Expresión gráfica de la captación de la tendencia de una serie temporal aplicando distintos procesos de descomposición.

3.1.3.2. Análisis de la Estacionalidad

El objetivo de determinar las componentes estacionales puede ser: en primer lugar para tener conocimiento de ellas y posteriormente eliminando posibles influencias estacionales. A ese proceso de eliminación de efectos estacionales en una serie temporal se le denomina “Desestacionalización”.



De acuerdo con (Dagum 1988) las tres características más significativas del fenómeno estacional son:

- Se repite cada año con cierta regularidad, pero puede evolucionar.
- Es posible medirlo y separarlo de las otras fuerzas que influyen en el movimiento de la serie.
- Es posible medirlo y separarlo de las otras fuerzas que influyen en el movimiento de la serie. Es causado principalmente por fuerzas no económicas, exógenas al sistema económico, que los tomadores de decisiones no pueden controlar o modificar en el corto plazo. (Por ejemplo la climatología).

De ahí que sea necesario recurrir a la “desestacionalización” de los datos. La explicación principal es que las causas que contribuyen a la manifestación de la estacionalidad de una serie, se consideran factores exógenos, de naturaleza no económica y que influyen en la variable a estudiar, “oscureciendo” las características de la serie relacionadas con aspectos meramente económicos (Hylleberg, Engle et al. 1990) y (Zellner 1978).

La Desestacionalización de la serie temporal permite analizar una serie cronológica de datos sin tener en cuenta la componente estacional. La Desestacionalización o corrección estacional evita las distorsiones en los valores medios de los datos. Una serie se estima que está desestacionalizada cuando se ha eliminado el efecto de cada periodo mensual y que se obtiene restando al valor de cada mes el coeficiente estacional de dicho mes. (Véase Ilustración 6).

Para llevar a cabo esta operación resulta indispensable aislar, en primer lugar la componente estacional, lo cual permitirá su posterior eliminación. En los distintos métodos que pueden utilizarse para obtener la componente estacional, existe un paso previo que consiste en eliminar la tendencia, obtenida mediante alguno de los



procedimientos señalados con anterioridad, medias móviles o regresión, principalmente.

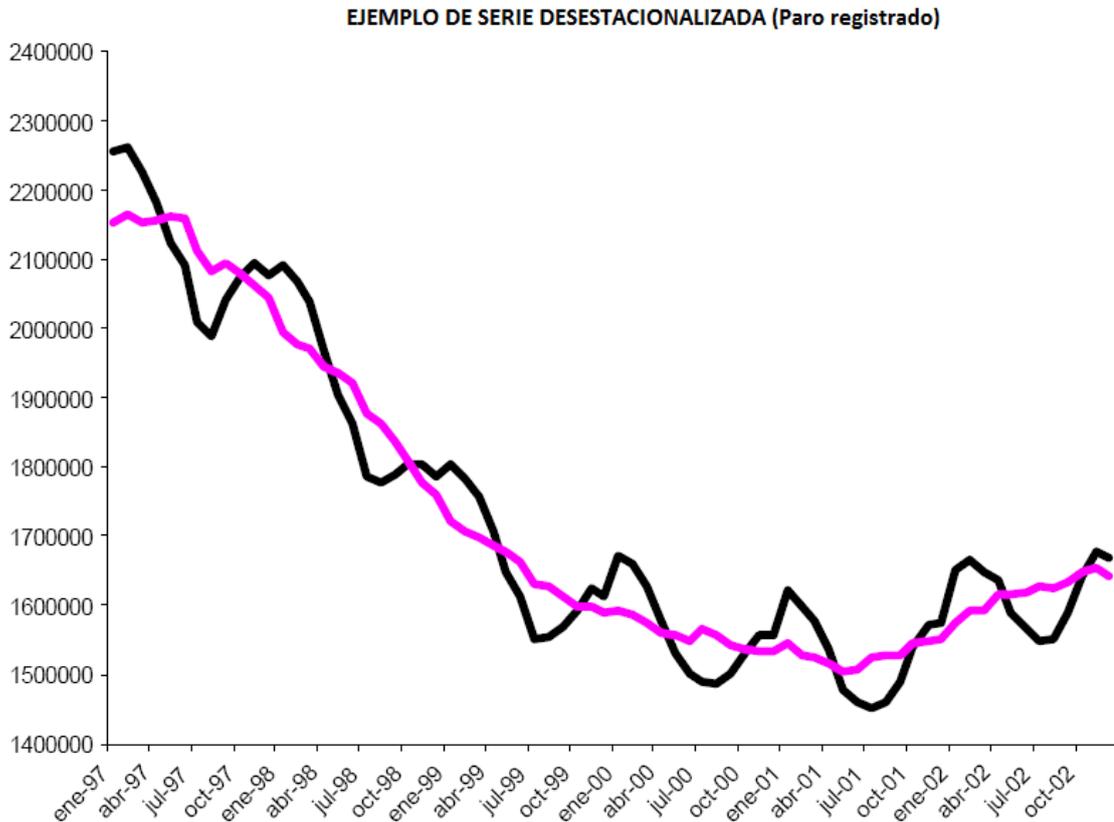


Ilustración 6. Ejemplo gráfico de la deseasonalización de una serie.

Otro método de deseasonalización es el de incluir a una serie temporal las llamadas variables “Dummy” para realizar su ajuste estacional. A estas variables artificiales se les otorga unos valores discretos generalmente de 0 y 1. La inclusión de este tipo de variables fue originalmente aplicada por Lovell (Kikut and Torres 1998), en los inicios de los años 60 del Siglo XX y su objetivo principal es el de “explicar” la estacionalidad en las series temporales. Cuando se trabaja con datos mensuales, la idea que surge es la de utilizar dichas variables artificiales, una para cada mes del año si es que los datos de la serie están agrupados mensualmente, que pueden definirse como:

$$q_1(\text{Enero}), q_2(\text{Febrero}), q_3(\text{Marzo}) \dots q_{12}(\text{Diciembre}) \quad (5)$$



No obstante aunque la inclusión de doce variables “dummy” pueda parecer natural, esto plantea un problema ya que las columnas correspondientes a las variables estacionales tendrían como resultado una combinación lineal exacta con la constante dándose así un caso de multicolinealidad perfecta. En caso de multicolinealidad el determinante de la matriz sería igual a cero y por tanto, singular no invertible, lo cual no permitiría estimar los coeficientes del modelo de regresión.

Para evitar este problema que se plantea deberían de utilizarse once de las doce variables “dummy” y por supuesto la constante. Al desestimar una de las doce variables estacionales Dummy (junto con la constante y las restantes variables que sean de interés) en la matriz, se evitaría la colinealidad perfecta con la columna de la constante asociada con el intercepto.

Por otra parte, para analizar la estacionalidad de una serie incluiremos un concepto interesante en el análisis de series temporales: la función de autocorrelación. La función de autocorrelación mide la correlación entre los valores de la serie distanciados un lapso de tiempo k . Recordemos la fórmula del coeficiente de correlación simple, dados n pares de observaciones x, y

$$r = \frac{\sum (y_i - \bar{y})(x_i - \bar{x})}{\sqrt{\sum (y_i - \bar{y})^2 \sum (x_i - \bar{x})^2}} \quad (6)$$

La función de autocorrelación es el conjunto de coeficientes de autocorrelación r, k desde 1 hasta un máximo que no puede superar la mitad de los valores observados, y es de vital importancia para detectar y estudiar la estacionalidad de la serie, ya que si ésta se manifiesta, los valores separados entre sí por intervalos iguales al periodo estacional deben estar correlacionados de alguna forma. Es decir que el coeficiente de



autocorrelación para un retardo igual al periodo estacional tiene que ser significativamente distinto de 0.

3.1.3.3. Análisis de la componente aleatoria o variaciones residuales

Se conoce como “residuos” a las diferencias entre los valores observados y los pronósticos. Los residuos son un componente de especial interés en lo que al análisis de regresión lineal se refiere. Estos residuos revelan sobre el grado de fidelidad de las predicciones: cuanto más bajo es el error típico de los residuos, mejores son las predicciones lo cual significa que mejor se ajusta la recta de regresión a la nube de puntos. Por otro lado, el estudio de las características de los casos con residuos altos (bien positivos o negativos; altos en valor absoluto) ayudan a captar casos infrecuentes, y por tanto, a refinar la ecuación de regresión a través de un estudio detallado de los mismos.

Una vez que tanto la tendencia como la estacionalidad han sido analizados y reconocidos y eliminados sus efectos, se mantienen una serie de residuos que pueden ser aleatorios o no serlo. El objetivo es estudiar qué tipo de comportamiento aleatorio presentan estos residuos, utilizando algún tipo de modelo probabilístico que los describa. La utilidad de esta última componente se basa en poder verificar si satisface ciertos supuestos o hipótesis, como el que sea realmente aleatoria. Si lo que se persigue es estudiar qué tipo de comportamiento aleatorio presentan estos residuos, es necesario un análisis de la correlación entre las distintas variables independientes tenidas en cuenta. Algunas de las fluctuaciones aparentemente irregulares, pueden explicarse en término de modelos probabilísticos, como por ejemplo los modelos ARIMA.

Los “Residuos Tipificados” (aquellos divididos por su error típico), presentan una media de 0 y una desviación típica de 1



3.2. Los Modelos de Regresión

En determinadas ocasiones en muchos problemas se da una relación característica entre dos o más variables y se necesita investigar la esencia de esa relación. El análisis de regresión es una técnica estadística utilizada para lograr el modelado e investigación de la relación existente entre dos o más variables. Según Molina Ibáñez, Giménez et al. (2009) los modelos predictivos o de regresión conforman la representación de una relación dada entre dos (o más) variables a partir de un modelo teórico o formal que se plasma en forma de expresión lógico-matemática que, aparte de resumir cómo es esa relación, va a permitir establecer predicciones de los valores que tomará una de las dos variables (la que se tenga en cuenta como variable de respuesta o variable a predecir) a partir de los valores que tome la otra (la independiente). A continuación se detalla el ejemplo de aplicación de un modelo de regresión lineal simple con el objetivo de modelar la distribución conjunta de las variables “Estrategias de Afrontamiento” y “Estrés”. En este ejemplo el modelo de regresión se concreta en el ajuste a los datos de la siguiente ecuación de regresión o “recta de regresión” véase Ilustración 7.

$$\hat{Y} = 75,4 + (-0,76) \cdot X$$

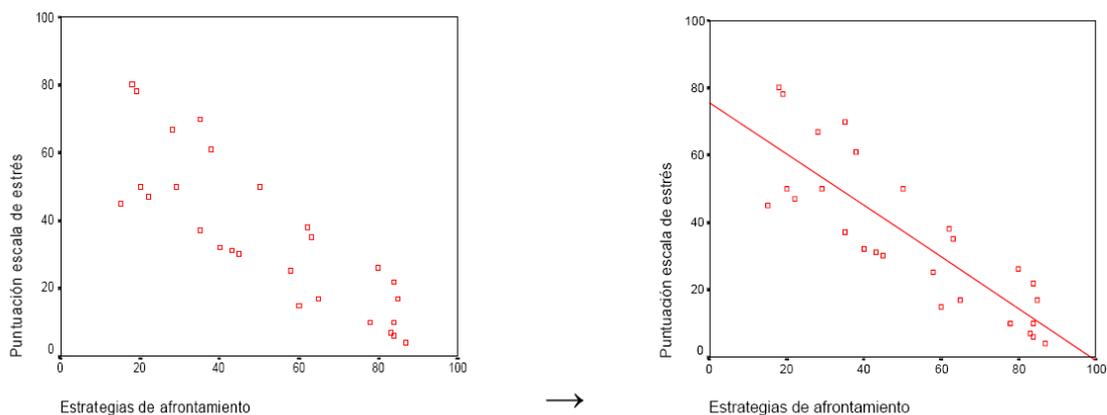


Ilustración 7 Representación gráfica del ajuste de datos a partir de la recta de regresión. Fuente: (Molina Ibáñez, Giménez et al. 2009)



Los modelos de regresión lineal (MRL) son los más utilizados de cara a poder predecir los valores de una variable cuantitativa o dependiente a partir de los valores de otra variable explicativa también cuantitativa (MRLS, modelos de regresión lineal simple). Si existe más de una variable explicativa o independiente que pueda también predecir el valor de la variable dependiente, entonces hablamos de modelos de regresión lineal múltiple (MRLM).

El análisis de regresión es una técnica estadística fundamental a la hora de modelizar y se encarga de explicar el comportamiento de dos medias encontrando una fórmula, que permita extraer una conclusión dada una cantidad de una variable y cómo se comportaría en la otra variable. Por ello el planteamiento del modelo debe ser fundamental y tener una base teórica. En materia de investigación el análisis de regresión es utilizado para predecir un amplio espectro de fenómenos. A su vez el análisis de regresión lleva asociados una serie de procedimientos de diagnóstico como análisis de los residuos, puntos de influencia,...etc que proporcionan pistas acerca de cuestiones a tener en cuenta como son la estabilidad e idoneidad del análisis.

Se recurre al análisis de regresión lineal cuando es necesario encontrar variables que sean explicativas y proceder así a crear un modelo donde tener en cuenta sólo aquellas variables que estén vinculadas con la respuesta, que suele ser la variable dependiente y por otro lado nos ayuda a descartar aquellas otras variables que no aportan información suficientemente explicativa. Además el análisis de regresión lineal, permite detectar interacciones entre las variables independientes que afectan a la variable dependiente o predicha. El modelo de regresión predice el valor de una variable dependiente (variable respuesta y , predicha o explicada, variable criterio) basándose en el valor de al menos una variable independiente (variable explicativa x o variable predictora). La situación de análisis idónea sería aquella en la que se pudiera disponer de una serie de variables independientes muy correlacionadas con la variable dependiente pero con poca correlación entre sí.

El análisis de regresión conlleva que se establezca una relación entre una variable dependiente y otra variable independiente. En el siguiente esquema se plantea el



,entonces puede considerarse que no hay presencia de variables explicativas o independientes que sean endógenas.

Si por alguna razón como, la falta de variables relevantes, errores de medida, simultaneidad...etc X_j está correlacionada con ε , se puede afirmar que X_j se trata de una variable independiente endógena.

En definitiva, un vínculo de causalidad entre las variables independientes y dependientes de un modelo conlleva endogeneidad. Si no se presta atención al modelo por endogeneidad, las estimaciones resultantes son sesgadas e inconsistentes y por tanto pueden arrojar conclusiones equivocadas. En relación a ello, existen test para probar la endogeneidad en un modelo, tal como el test de Hausman o el test de Wu.

El test de Hausman (1978) es un test chi cuadrado que dictamina si las diferencias entre dos estimaciones son sistemáticas y significativas. El empleo de dicho test tiene como objetivos, en primer lugar, saber si un estimador es lo suficientemente consistente y en segundo lugar tener el conocimiento de la relevancia o no de una determinada variable (Hausman 1978).

3.2.1. Los Modelos de Regresión Simple (MRLS)

En el caso de los modelos de regresión lineal simple se considera sólo un regresor, también llamado predictor x y una variable dependiente también conocida como respuesta Y . Se parte de la base de que la verdadera relación entre Y y x es una línea recta y que la observación Y en cada nivel x es una variable aleatoria. El modelo de regresión simple se expresa de la siguiente forma:

$$y = \beta_1 + \beta_2 x + u \tag{9}$$

En este modelo hay tres tipos de variables: y , x y u . En el modelo el único factor que puede explicar y es x mientras que el resto de los factores que afectan a y están



recogidos en u . y por tanto se trata de la variable endógena o dependiente. En la regresión lineal simple de y sobre x , a la variable x se le denomina variable exógena, explicativa o independiente. La variable u es la que recoge todos aquellos factores distintos de x que afectan a y . Esta variable es conocida como “error” o perturbación aleatoria. El término de perturbación puede captar también el error de medición de la variable dependiente. A su vez se entiende u como una variable “no observable”.

Los parámetros β_1 y β_2 son fijos y desconocidos. En la fórmula inicial se pueden distinguir dos componentes: un componente sistemático $\beta_1 + \beta_2 x$ y la perturbación aleatoria u . Si denominamos u como u_y al componente sistemático, podemos partir de la siguiente ecuación:

$$u_y = \beta_1 + \beta_2 x \quad (10)$$

A esta ecuación se la conoce como función de regresión poblacional o recta poblacional. Por lo que, como puede verse en la **¡Error! No se encuentra el origen de la referencia.**, u_i es una función lineal de x con término independiente igual a β_1 y pendiente igual a β_2 . La existencia de linealidad significa que un aumento de una unidad en x implica que el valor esperado de $y - u_i = E(y) -$ varíe en β_1 unidades. Suponiendo que se dispone de una muestra aleatoria de tamaño $n((y_i, x_i) : i = 1, \dots, n)$ extraída de la población estudiada. En el diagrama de dispersión se muestran los hipotéticos valores de la muestra.

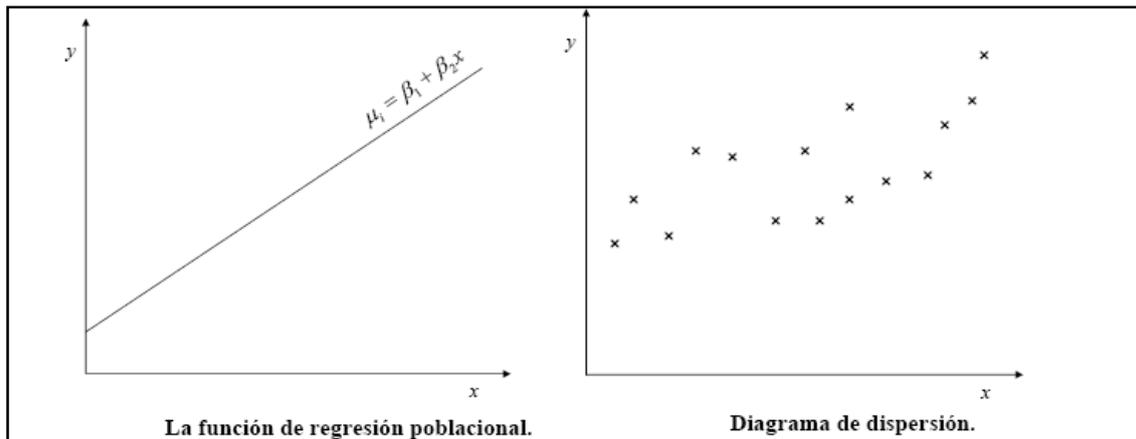


Ilustración 9. Función de regresión poblacional y diagrama de dispersión

El modelo poblacional para cada observación de la muestra se puede expresar de la siguiente forma:

$$y_i = \beta_1 + \beta_2 x_i + u_i \quad i = 1, 2, \dots, n \quad (11)$$

En la Ilustración 9 se muestran de forma conjunta la función de regresión poblacional y el diagrama de dispersión, pero es importante no olvidar que β_1 y β_2 son fijos, pero desconocidos. De acuerdo con este modelo es posible, desde un punto de vista teórico, hacer la siguiente descomposición:

$$y_i = \mu_{y_i} + u_i \quad i = 1, 2, \dots, n \quad (12)$$

que ha sido representada anteriormente para la observación i -ésima. Sin embargo, desde un punto de vista empírico, no es posible hacerlo debido a que β_1 y β_2 son desconocidos y, consecuentemente, u_i es no observable. En el análisis de regresión las estimaciones de β_1 y β_2 pueden obtenerse por medio del conocido como método de mínimos cuadrados.

3.2.2. Los Modelos de Regresión Múltiple (MRLM)

Cuando existen varias variables independientes a estimar o predecir, entonces el análisis pasa a denominarse “Análisis de Regresión Múltiple”. El análisis de regresión lineal múltiple, a diferencia del simple, se acerca más a escenarios de análisis real puesto que



los fenómenos, hechos o sucesos sociales, por definición, son complejos y, por tanto deben ser debidamente explicados y justificados en la medida de lo posible por la serie de variables que directa o indirectamente inciden en su concreción. Por tanto el Análisis de Regresión Lineal Múltiple nos permite establecer la relación que se produce entre una variable dependiente Y y un conjunto de variables independientes (Rodríguez Jaume and Mora Catalá 2001).

Cuando hay más de una variable explicativa (modelo de regresión lineal múltiple), se utiliza un subíndice para cada una de ellas, por ejemplo, para el caso de dos variables explicativas. El modelo de regresión lineal múltiple se expresa de la siguiente forma:

$$Y = \beta_1 + \beta_2 \cdot X_2 + \dots + \beta_K \cdot X_K + U \quad (13)$$

Al igual que en los modelos de regresión simple, se deben de tener en cuenta otros requisitos que deben de cumplir los datos de los modelos de regresión lineal múltiple (Abuín 2007):

- Linealidad: Esto significa que la ecuación de regresión toma una forma particular. Esto es, la variable predictiva o dependiente se corresponde con la suma de un conjunto de elementos: el origen de la recta, una combinación lineal de variables independientes o predictoras y los residuos. El no cumplimiento del supuesto de linealidad suele denominarse como error de especificación.
- Homocedasticidad: Se refiere a que la distribución de los errores respecto al plano de regresión es constante, es decir, homogénea alrededor del mismo. Para cada valor de la variable explicativa o independiente (o combinación de valores de las variables independientes), la varianza de los residuos es constante.
- Normalidad: Dicha distribución de los errores sigue una ley normal. Esto es, para cada valor de la variable independiente (o combinación de valores de las



variables independientes), los residuos se distribuyen normalmente con media cero. Si el valor del estadístico Durbin Watson está próximo a 2 entonces los residuos no están autocorrelacionados. Si su valor es 0 hay autocorrelación perfecta positiva. Si vale 4 existe autocorrelación perfecta negativa.

- Independencia de errores: Los errores o residuos son independientes entre sí (que no existe autocorrelación), no están relacionados tampoco con las variables predictivas (independientes) ni con la criterio o dependiente (Endogeneidad). Los residuos constituyen una variable aleatoria y es frecuente encontrar residuos autocorrelacionados cuando se trabaja con datos de series temporales.
-
- No-colinealidad. No existe relación lineal precisa entre ninguna de las variables independientes. El no cumplimiento de este supuesto tiene como resultado la colinealidad o multicolinealidad.

Esquema de análisis a partir de un (MRLM) (Véase Ilustración 10):

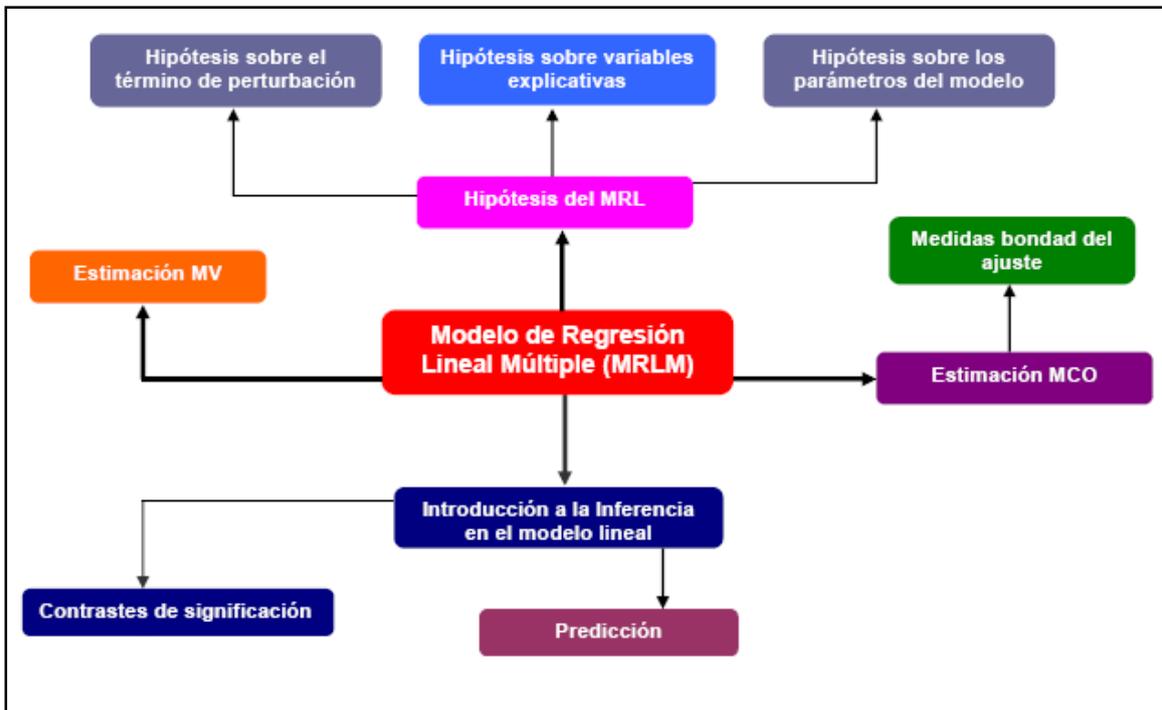


Ilustración 10. Esquema de análisis de un Modelo de Regresión Lineal Múltiple.



Además de los requisitos comentados, para poder llegar a disponer de unas estimaciones adecuadas de los coeficientes de la ecuación, el modelo de regresión lineal múltiple requiere que las variables predictoras no presenten un nivel alto de correlación entre sí.

Si aparecieran altos grados de correlación entre ellas, estos coeficientes ($\beta_1, \beta_2 \dots$) pueden experimentar cambios notables debido a que cada uno de ellos manifiesta el efecto específico de cada predictor con la variable a predecir eliminando en dicho cálculo la correlación que dicho predictor mantiene con el resto de predictores así como la que tienen estos predictores con la variable dependiente o criterio.

En este sentido se muestra que los coeficientes de la ecuación de regresión múltiple son coeficientes de “correlación parcial” por lo que la aparición de grandes porcentajes de variabilidad compartida entre predictores, también llamada colinealidad, provoca que las estimaciones calculadas experimenten muchos cambios respecto a las estimadas independientemente entre cada regresor particular con la variable dependiente. La estimación de los coeficientes de una regresión múltiple es un cálculo bastante complicado y laborioso, por lo que se requiere del empleo de programas de computación especializados. Sin embargo, la interpretación de los coeficientes es similar al caso de la regresión simple: el coeficiente de cada variable independiente mide el efecto separado que esta variable tiene sobre la variable dependiente.

3.2.3. Hipótesis del modelo de regresión lineal múltiple (MRLM)

Mediante un modelo de regresión lineal múltiple (MRLM) tratamos de explicar el comportamiento de una determinada variable que denominaremos variable a explicar, variable endógena o variable dependiente, (y representaremos con la letra Y) en función de un conjunto de k variables explicativas x_1, x_2, \dots, x_k mediante una relación de dependencia lineal (suponiendo $x_1 = 1$):



$$Y = \beta_1 + \beta_2 \cdot X_2 + \dots + \beta_k \cdot X_k + U \quad (14)$$

siendo U el término de perturbación o error

Para determinar el modelo anterior, es de obligado cumplimiento el poder hallar o estimar el valor de los coeficientes $\beta_1, \beta_2, \dots, \beta_k$. La propiedad de linealidad en parámetros hace posible la interpretación correcta de los parámetros del modelo. Esos parámetros miden la intensidad media de los efectos de las variables explicativas o independientes sobre nuestra variable dependiente o a estimar y se obtienen al tomar las derivadas parciales de la variable a explicar respecto de cada una de las variables independientes.

$$\beta_j = \frac{\partial Y}{\partial X_j}; \quad j = 1, \dots, k \quad (15)$$

El objetivo se centra en la labor de asignar valores numéricos a esos parámetros $\beta_1, \beta_2, \dots, \beta_k$. Por lo tanto se buscará el poder estimar el modelo de forma que, los valores ajustados de la variable endógena (aquella que viene explicada por el funcionamiento del modelo) resulten tan próximos a los valores realmente observados como sea posible.

A fin de poder establecer las propiedades de los estimadores obtenidos al aplicar distintos métodos de estimación y realizar diferentes contrastes, se ha de especificar un conjunto de hipótesis sobre el Modelo de Regresión Lineal múltiple que hemos formulado.

Existen tres grupos de hipótesis:

- Las hipótesis sobre el término de perturbación.
- Las hipótesis sobre las variables independientes/explicativas.
- Las hipótesis sobre los parámetros del modelo.



3.2.4. Hipótesis sobre el término de perturbación

Para una muestra de n observaciones (cada observación estará formada por una matriz con los valores de X_2, X_3, \dots, X_k y el valor de Y asociado), obtenemos el siguiente sistema de n ecuaciones lineales:

$$\left. \begin{array}{l} Y_1 = \beta_1 + \beta_2 \cdot X_{21} + \dots + \beta_k \cdot X_{k1} + u_1 \\ Y_2 = \beta_1 + \beta_2 \cdot X_{22} + \dots + \beta_k \cdot X_{k2} + u_2 \\ \dots \\ Y_n = \beta_1 + \beta_2 \cdot X_{2n} + \dots + \beta_k \cdot X_{kn} + u_n \end{array} \right\} \quad (16)$$

O en forma matricial: $Y = X \cdot B + U$, donde:

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \dots \\ Y_n \end{bmatrix}, X = \begin{bmatrix} 1 & X_{21} & \dots & X_{k1} \\ 1 & X_{22} & \dots & X_{k2} \\ \dots & \dots & \dots & \dots \\ 1 & X_{2n} & \dots & X_{kn} \end{bmatrix}, B = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \dots \\ \beta_k \end{bmatrix}, U = \begin{bmatrix} u_1 \\ u_2 \\ \dots \\ u_n \end{bmatrix} \quad (17)$$

En estas condiciones, las hipótesis del Modelo de Regresión Lineal múltiple se resumen en la esfericidad del término de perturbación, *i.e.* :

- El valor esperado de la perturbación es cero: $E[u_i] = 0 \quad \forall i = 1, \dots, n$
- Homoscedasticidad: Todos los términos de perturbación tienen la misma varianza (varianza constante):

$$Var[u_i] = Var[u_j] = \sigma^2 \quad \forall i \neq j \quad (18)$$



Por lo que, todos los términos de la diagonal principal de la matriz de varianzas y covarianzas serán iguales:

$$Var[U] = \begin{bmatrix} \sigma^2 & & & \\ & \sigma^2 & & \\ & & \dots & \\ & & & \sigma^2 \end{bmatrix} \quad (19)$$

c.) No Autocorrelación: Los errores son independientes unos de otros, *i.e.*: la matriz de varianzas y covarianzas es una matriz diagonal (fuera de la diagonal principal todo son ceros):

$$Var[U] = \begin{bmatrix} \sigma^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \sigma_n^2 \end{bmatrix} \quad (20)$$

Observar que, bajo las hipótesis de homoscedasticidad y no autocorrelación, la matriz de varianzas y covarianzas tendrá la forma siguiente:

$$Var[U] = \begin{bmatrix} \sigma^2 & 0 & \dots & 0 \\ 0 & \sigma^2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \sigma^2 \end{bmatrix} = \sigma^2 \cdot I_n \quad (21)$$

(I_n es la matriz identidad de orden n)

d.) El error de perturbación sigue una distribución normal *i.e.*:



$$U \approx N(0_n, \sigma^2 \cdot I_n) \quad (22)$$

3.2.4.1. Hipótesis sobre las variables explicativas o independientes

- a.) Las variables explicativas son fijas o deterministas.
- b.) Las variables explicativas están no correlacionadas con la perturbación aleatoria.
- c.) Las variables explicativas no presentan relación lineal exacta entre sí.
- d.) Además, supondremos que las variables explicativas son medidas sin error.
- e.) En el modelo no se excluyen las variables relevantes y que tampoco no se incluyen las variables irrelevantes, a la hora de explicar el comportamiento de la variable endógena.

3.2.4.2. Hipótesis sobre los parámetros del modelo

- a.) La única hipótesis que presentaremos sobre de los parámetros del modelo es la hipótesis de permanencia estructural, lo cual significa que los parámetros poblacionales β_j , se mantienen constantes a lo largo de toda la muestra.

3.2.5. Estimación del Modelo de Regresión Lineal Múltiple (MRLM)

Estimar el modelo es lo mismo que asignar valores numéricos a los parámetros desconocidos $\beta_1, \beta_2, \dots, \beta_k$, a partir de la información muestral disponible de las variables observables del modelo. Para ello se considerarán tres métodos de estimación:

- El Método de Mínimos Cuadrados Ordinarios (MCO)
- El Método de Mínimos Cuadrados en dos etapas (MC2E)
- El Método de Máxima Verosimilitud (MV)

3.2.5.1. Método de mínimos Cuadrados Ordinarios (MCO)

El método de estimación de mínimos cuadrados (MCO o “Ordinary Least Squares Regresión”) se fundamenta en la minimización de la suma cuadrada de los desvíos entre los valores observados y el modelo de prueba estimado para un conjunto de parámetros. Este método sirve a su vez, para encontrar la recta que mejor ajuste a los datos. (Johnston and Dinardo 2001).

La aplicación de Mínimos Cuadrados Ordinarios (MCO) en un sistema de ecuaciones no simultáneas genera (en ausencia de otros problemas de especificación individual de cada ecuación) estimaciones insesgadas, consistentes y eficientes en tanto que su utilización en modelos con simultaneidad (y, por tanto, con riesgo de regresores estocásticos correlacionados con las perturbaciones aleatorias) no garantiza la insesgadez (riesgo de estimaciones sesgadas) ni la consistencia (el sesgo no sólo se presenta en muestras pequeñas sino que se mantiene para muestras grandes) (Uriel 2013).

Con esta premisa, lo cierto es que el estimador de MCO sólo es analíticamente recomendable para modelos carentes de simultaneidad o recursivos (también llamados triangulares). Tal denominación de “triangulares” se refiere a la forma triangular de la matriz de coeficientes “gamma” de este tipo de modelos. En este tipo de modelos las variables endógenas que actúan como explicativas en las ecuaciones, no estarán relacionadas con las perturbaciones de las mismas, lo cual anula la posibilidad de que aparezcan problemas de sesgo en la estimación.

Este método presenta muchas ventajas en cuanto a la facilidad de uso que plantea y por lo adecuado del planteamiento estadístico matemático que permite adaptarse a los supuestos para los modelos econométricos. El término de MCO (Mínimos Cuadrados Ordinarios) está vinculado con la regresión y la correlación, ambas determinan la existencia de relación entre dos o más variables (siempre una dependiente y una o varias independientes). La existencia de variables independientes que sean endógenas anula



los estimadores (MCO) de los parámetros del modelo y entonces serán inconsistentes. Para resolver ese problema y así poder obtener estimadores consistentes de los parámetros del modelo en presencia de variables independientes endógenas se puede recurrir a métodos como a través del uso de variables instrumentales o aplicando el método de mínimos cuadrados bietápicos o mínimos cuadrados en dos etapas, que se explicará más adelante.

Partiendo de un modelo en forma matricial $Y = X \cdot B + U$. Partamos de la base de que el modelo se estimó obteniéndose \hat{Y} , vector de valores de la variable dependiente implicado por el modelo. La diferencia entre los valores observados y los valores estimados, $e = Y - \hat{Y} = Y - X \cdot \hat{B}$, la denominaremos vector de residuos. Ahora bien, el problema que se plantea solventar es el de minimizar la suma de los cuadrados $e'e$ con respecto del vector de parámetros estimados, B . De este problema de optimización se deduce la siguiente expresión de mínimos cuadrados ordinarios del modelo de regresión lineal múltiple.

$$\hat{B} = (X' \cdot X)^{-1} \cdot X' \cdot Y \quad (23)$$

Cuya varianza viene dada por:

$$\text{Var} \left[\hat{B} \right] = \sigma^2 (X' \cdot X)^{-1} \quad (24)$$

Por otro lado, el estimador MCO de la varianza del término de perturbación es:

$$\hat{\sigma}_u^2 = \frac{e'e}{n - k} \quad (25)$$



donde n es el número de observaciones y k es el número de elementos del vector B .

Bajo la hipótesis de perturbaciones esféricas, el estimador MCO del vector B cumple una serie de propiedades que le convierten en un insesgado (el valor esperado del estimador coincide con el valor real del parámetro), eficiente (de varianza mínima), y consistente.

Por otro lado, bajo la hipótesis de esfericidad, el estimador MCO de la varianza del término de error, $\hat{\sigma}_u^2$, es también insesgado.

3.2.5.2. Método de mínimos Cuadrados Ordinarios en dos etapas (MC2E)

En presencia de simultaneidad, una segunda estrategia para resolver los no deseados efectos derivados de la aplicación directa de Mínimos Cuadrados Ordinarios (MCO) (sesgo e inconsistencia) es la utilización de la estrategia de estimación conocida como MC2E (Juan, Kizys et al.).

El método consiste en utilizar Mínimos Cuadrados Ordinarios (MCO) sobre la forma estructural pero, antes de ello, reemplazar los valores reales originales de las variables explicativas de cada ecuación (es decir, las endógenas que aparecen en el lado derecho de cada ecuación) por sus valores MCO estimados en la forma reducida (de otro modo, no podríamos plantear la estimación de la forma reducida). Para representar el itinerario operativo de Mínimos Cuadrados en dos Etapas (MC2E), supongamos el siguiente modelo simultáneo con dos ecuaciones:

$$\begin{aligned} Y_{1i} &= \beta_{11}X_{1i} + \beta_{12}X_{2i} + \gamma_{12}Y_{2i} + U_{1i} \\ Y_{2i} &= \beta_{21}X_{1i} + \beta_{23}X_{3i} + \gamma_{21}Y_{1i} + U_{2i} \end{aligned} \tag{26}$$



Para el caso de la primera ecuación, antes de proceder a la estimación directa con MCO, reemplazamos los valores originales de la variable Y_{2i} (un regresor estocástico potencialmente relacionado con U_{1i}) por una estimación obtenida aplicando MCO sobre su forma reducida, o lo que es lo mismo:

$$\begin{aligned} Y_{2i} &= \pi_{21}X_{1i} + \pi_{22}X_{2i} + \pi_{23}X_{3i} + V_{2i} \rightarrow \\ \rightarrow \hat{Y}_{2i} &= \hat{\pi}_{21}X_{1i} + \hat{\pi}_{22}X_{2i} + \hat{\pi}_{23}X_{3i} \rightarrow \\ \rightarrow Y_{2i} &= \hat{\pi}_{21}X_{1i} + \hat{\pi}_{22}X_{2i} + \hat{\pi}_{23}X_{3i} + \hat{V}_{2i} \end{aligned} \quad (27)$$

De esta forma, la ecuación a estimar vendría dada de la siguiente manera:

$$Y_{1i} = \beta_{11}X_{1i} + \beta_{12}X_{2i} + \gamma_{12}(\hat{Y}_{2i} + \hat{V}_{2i}) + U_{1i} \quad (28)$$

o lo que es igual,

$$Y_{1i} = \beta_{11}X_{1i} + \beta_{12}X_{2i} + \gamma_{12}\hat{Y}_{2i} + (U_{1i} + \gamma_{12}\hat{V}_{2i}) \quad (29)$$

Como puede observarse, nos encontramos de nuevo ante una estimación con información limitada ya que, otra vez, no necesitamos conocer la especificación concreta de cada ecuación pero sí la lista de regresores (X) y endógenas (Y) del modelo.

El método de Mínimo Cuadrado en dos Etapas posee varias ventajas (Juan, Kizys et al.):

- Se aborda la estimación aislada de cada ecuación lo que, operativamente, supone una ventaja y evita el contagio a todo el modelo de los errores presentes en una ecuación.



- El uso de los valores estimados de las variables explicativas evita la presencia de regresores estocásticos relacionados con la perturbación aleatoria; las variables explicativas originales son aleatorias pero sus valores estimados procedentes de la forma reducida no lo son.
- En principio cabe pensar que el uso de estimadores de Mínimos Cuadrados Ordinarios en dos Etapas (MC2E) ante la presencia de simultaneidad, produce estimaciones consistentes (es decir, anula el problema de los regresores estocásticos). No obstante, la insesgadez y la eficiencia sólo se conseguirán para muestras grandes, sin que pueda garantizarse para estimaciones con conjuntos de datos reducidos.

Otras ventajas son:

- La sencillez en su aplicación ya que este método sólo requiere dos sencillas estimaciones sucesivas por Mínimos Cuadrados Ordinarios MCO.
- No requiere una identificación exacta de la ecuación; puede utilizarse también para ecuaciones superidentificables.
- Es más robusto que otros métodos ante la aparición de problemas de especificación o multicolinealidad en las ecuaciones.

A diferencia de otros métodos, la aplicación de Mínimos cuadrados Ordinarios en Dos Etapas (MC2E) sí permite disponer de una estimación de las varianzas de los parámetros. En esa segunda etapa se lleva a cabo una estimación de los parámetros estructurales " β " y " γ ", por tanto, contamos con unos residuos derivados de esta estimación que nos ofrecen la posibilidad de calcular las desviaciones típicas de los parámetros estimados.



Sin embargo, además de contar con esas ventajas, cabe destacar que presenta una limitación importante:

El procedimiento de Mínimos Cuadrados en Dos etapas (MC2E) exige la estimación de la forma reducida de cada ecuación lo cual sólo es posible si $n > k$.

3.2.5.3. Estimación por Máxima Verosimilitud

El método de estimación por mínimos cuadrados (MCO) se basa en otorgar valores numéricos a los parámetros desconocidos de forma que la suma cuadrática de errores sea mínima y sólo requiere que la matriz $X'X$ sea invertible. A continuación se mostrará un método de estimación alternativo como es el método de máxima verosimilitud (MV). Este método propone como un estimador el valor que maximiza la probabilidad de obtener la muestra ya disponible.

El método se apoya, prácticamente, en la distribución que sigue el término de error. A tales efectos, se suele suponer que las perturbaciones aleatorias se distribuyen con una distribución Normal que, además de cumplir las propiedades de una muestra grande, es una aproximación cómoda y fácil de tratar.

El modelo que se usará es $Y = X \cdot B + U$, y se supondrá que el término aleatorio sigue la distribución Normal con la siguiente función de densidad:

$$f(u_i) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left\{-\frac{u_i^2}{2\sigma^2}\right\}, i = 1, \dots, N. \quad (30)$$

Maximizar la probabilidad de obtener la muestra ya disponible equivale a maximizar la función de densidad conjunta del valor aleatorio, u . Para ello, se ha de suponer homoscedasticidad y ausencia de autocorrelación. Así, la expresión de la función de densidad conjunta es la siguiente:



$$f(U) = \prod_{i=1}^n f(u_i) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{-\frac{\sum u_i^2}{2\sigma^2}\right\} \quad (31)$$

Como U sigue una distribución Normal Multivariante de orden k , la variable Y , al ser una combinación lineal de las perturbaciones aleatorias, también se distribuirá con una distribución Normal Multivariante. Así pues, para que la función de densidad conjunta sea una función de verosimilitud, el vector aleatorio U ha de expresarse en función del vector Y , es decir:

$$L(Y; \beta, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{-\frac{(Y - X\beta)'(Y - X\beta)}{2\sigma^2}\right\} \quad (32)$$

Se trata, por tanto, de maximizar la función de verosimilitud. Como la expresión anterior resulta complicada, aplicaremos una transformación monótona; en concreto, una función logarítmica:

$$\ln L(Y; \beta, \sigma^2) = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln(2\sigma^2) - \frac{(Y - X\beta)'(Y - X\beta)}{2\sigma^2} \quad (33)$$

Derivando la función de verosimilitud con respecto de B y σ^2 , e igualando las derivadas a cero, obtenemos los resultados:

$$\hat{B}_{MV} = (X' \cdot X)^{-1} \cdot X' \cdot Y \quad (34)$$

Y cuya varianza es la siguiente:

$$\text{Var}\left[\hat{B}_{MV}\right] = \sigma^2 (X' \cdot X)^{-1} \quad (35)$$



Añadir que, el estimador MCO de la varianza del término de perturbación es:

$$\sigma_{MV}^2 = \frac{e' \cdot e}{n} \quad (36)$$

donde n es el número de observaciones y k es el número de elementos del vector B .

Se puede observar que el estimador de MV de B coincide con el MCO y no es insesgado aunque sí asintóticamente insesgado.

3.2.6. Medidas de Bondad del ajuste del Modelo (MRLM)

En la estimación por Mínimos Cuadrados Ordinarios (MCO) aún no es posible evaluar la calidad de ajuste del modelo. Por ello es necesario tener en cuenta una serie de medidas de bondad de ajuste. Inicialmente se tendrán en cuenta la Suma de los Cuadrados de Errores (SCE), que puede expresarse de distintas formas:

$$e' \cdot e = \sum_{i=1}^n e_i^2 = Y' \cdot Y - \hat{B}' \cdot X' \cdot Y = Y' \cdot Y - \hat{Y}' \cdot Y = \sum_{i=1}^n Y_i^2 - \sum_{i=1}^n \hat{Y}_i^2 \quad (37)$$

De esta fórmula, despejando la suma de cuadrados la variable endógena, toma esta otra expresión:

$$Y' \cdot Y = \hat{Y}' \cdot \hat{Y} + e' \cdot e \quad (38)$$

, o bien



$$\sum_{i=1}^n Y_i^2 = \sum_{i=1}^n \hat{Y}_i^2 + \sum_{i=1}^n e_i^2 \quad (39)$$

Si restamos a ambos lados la cantidad $n \cdot \bar{Y}^2$, tenemos como resultado:

$$Y' \cdot Y - n \cdot \bar{Y}^2 = \hat{Y}' \cdot \hat{Y} - n \cdot \bar{Y}^2 + e' \cdot e \quad (40)$$

,o bien

$$\sum_{i=1}^n (Y_i - \bar{Y})^2 = \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2 + \sum_{i=1}^n e_i^2 \quad (41)$$

La parte representada en el lado izquierdo representa la “Suma de Cuadrados Totales” (SCT) y esta no es sino la suma de los cuadrados de las desviaciones respecto a su media aritmética.

Por otra parte, si el modelo posee término independiente, a la cantidad;

$$\hat{Y}' \cdot \hat{Y} - n \cdot \bar{Y}^2 = \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2 \text{ se le denomina “Suma de Cuadrados de Regresión” (SCR).}$$

En definitiva, la variabilidad total de la variable endógena (SCT) puede descomponerse en dos partes: una la que es posible explicar a través del modelo especificado (SCR) y la parte que no es posible explicar, la suma de cuadrados de los errores (SCE).

A partir de la descomposición mostrada anteriormente de la SCT, se procede a definir el Coeficiente de Determinación R^2 , el cual se convierte en la primera medida de bondad del ajuste:

$$R^2 = 1 - \frac{SCE}{SCT} \quad (42)$$



Cuando el modelo tiene término independiente, entonces se cumple la igualdad $SCT=SCR+SCE$, y el Coeficiente de Determinación puede expresarse de otra forma alternativa:

$$R^2 = \frac{SCR}{SCT} \quad (43)$$

El Coeficiente de Determinación indica qué proporción de variabilidad total es explicada por la regresión. Si el modelo tiene término independiente, entonces R^2 pasa a tomar valores comprendidos entre 0 y 1.

En la práctica, el uso de R^2 cuenta con algunas limitaciones si lo que se pretende es comparar varios modelos desde la perspectiva de bondad del ajuste. Cuantas más variables explicativas son incorporadas al modelo, mayor es el Coeficiente de Determinación, pues la SCR disminuye a medida que aumenta el número de variables explicativas. De tal forma que si lo que queremos es llevar a cabo un análisis comparativo entre varios modelos, entonces recurrimos al R^2 corregido:

$$\bar{R}^2 = 1 - \frac{n-1}{n-k}(1-R^2) \quad (44)$$

Este estadístico es invulnerable ante la incorporación de extra variables vía interacción de dos efectos: el que permite aumentar R^2 , y el efecto contrario que aparece al descontar un mayor número de las variables explicativas, $\frac{n-1}{n-k}$.

3.2.7. Significación de los parámetros del modelo



Se distinguirá entre dos distintas dimensiones de significación: Por un lado la económica y por otro la estadística.

3.2.7.1. Significación Económica

La significación económica permite verificar si las estimaciones obtenidas son coherentes con la teoría económica. Según especificación del modelo, la interpretación y significación de los parámetros puede variar. Si el modelo está especificado en niveles, el parámetro refleja el efecto medio que tiene una variación unitaria de la variable explicativa sobre la variable endógena:

$$\beta_j = \frac{\partial Y}{\partial X_j} \quad (45)$$

En cambio, si el modelo está especificado en logaritmos neperianos, los parámetros pueden interpretarse como una elasticidad, como es el caso de la función de producción de Cobb-Douglas (Redondo 2011). El modelo de Cobb-Douglas, para la determinación de la función de producción, fue definido por Cobb y Douglas en 1928.

$$\beta_j = \frac{\partial \ln Y}{\partial \ln X_j} \quad (46)$$

3.2.7.2. Significación estadística



El análisis econométrico tiene como objetivo analizar por medio de una serie de contrastes, la significación (o significatividad) estadística individual y conjunta de los parámetros del modelo. Para contrastar las hipótesis de significatividad individual, tenemos:

$$\begin{aligned} H_0 : \beta_j &= 0 \\ H_A : \beta_j &\neq 0 \end{aligned} \tag{47}$$

El estadístico t-Student al que se recurre para realizar el test es el siguiente:

$$t_j = \frac{\hat{\beta}_j}{\sqrt{\hat{\sigma}_u^2 \cdot a_{jj}}} \sim t_{n-k} \tag{48}$$

Donde $\sqrt{\hat{\sigma}_u^2 \cdot a_{jj}}$ es el error estándar estimado de $\hat{\beta}_j$ y a_{jj} es el j-ésimo elemento de la diagonal principal de la matriz $(X'X)^{-1}$.

Dado un nivel de significación α , las tablas de distribuciones nos proporcionan la cantidad $t_{n-k, \alpha/2}$ que es el valor asociado a una t-Student con $n-k$ grados de libertad que deja a su derecha un área de $\alpha/2$ (o, equivalentemente, deja a su izquierda un área de $1 - \alpha/2$). La regla de decisión que utilizaremos para determinar si el parámetro asociado a la variable X_j es individualmente significativo o no es la siguiente:

- Si $|t_j| \geq t_{n-k, \alpha/2}$, el estadístico cae fuera de la región de aceptación, por lo que rechazamos la hipótesis nula. Concluimos, por tanto, que el parámetro es significativamente diferente de cero.
- Si $|t_j| < t_{n-k, \alpha/2}$, el estadístico cae dentro de la región de aceptación, por lo que no podemos rechazar la hipótesis nula. Por tanto, el parámetro no es individualmente significativo.



3.2.7.3. Análisis de los Residuos en el análisis de regresión.

Uno de los procedimientos de diagnóstico que van ligados al análisis de regresión es el del análisis de los residuos considerando su correlación. Los residuos son las diferencias entre los valores observados y los valores pronosticados por lo que son de vital importancia en el análisis de regresión ya que nos arrojan información del grado de exactitud de los pronósticos, de manera que cuanto más pequeño es el error típico de los residuos mejores son los pronósticos, lo cual significa que mejor se ajusta la recta de regresión a la nube de puntos.

El análisis de las características de los casos con residuos grandes (bien positivos o bien negativos en valor absoluto), puede servir para detectar casos atípicos, y por tanto ayudar al perfeccionamiento de la ecuación de regresión a través de un estudio detallado de los mismos. El análisis de los residuos es básico para chequear si se verifican las hipótesis del modelo de regresión.

Se denominan residuos a las diferencias entre los valores observados y los pronosticados: $(Y_i - \hat{Y}_i)$. Los residuos son de gran importancia dentro del análisis de regresión ya que proporciona información sobre el grado de exactitud de los pronósticos: cuanto más pequeño es el error típico de los residuos, mejores resultarán los pronósticos, o lo que es lo mismo, mejor se ajusta la recta de regresión a la nube de puntos. Por otro lado, el análisis de las características de los casos con residuos grandes (bien positivos o negativos; es decir, grandes en valor absoluto), puede ayudar a captar casos atípicos y, consecuentemente, a perfeccionar la ecuación de regresión a través de un estudio detallado de los mismos.

El verdadero interés de los residuos reside en el hecho de que el análisis de los mismos nos proporciona información crucial sobre el cumplimiento de varios supuestos del modelo de regresión lineal, por lo que se debe de contrastar que se dan las hipótesis básicas o supuestos del modelo:



$$\varepsilon_{it} = Y_{it} - (\mu + \alpha_i) \quad (49)$$

$$t = 1, \dots, n_i, \quad i = 1, \dots, I.$$

Esos supuestos son: independencia, homocedasticidad, normalidad y linealidad.

Puesto que los errores del modelo son desconocidos, los supuestos anteriores deben verificarse a partir de los residuos:

$$\varepsilon_{it} = y_{it} - \bar{y}_i \quad (50)$$

$$t = 1, \dots, n_i, \quad i = 1, \dots, I.$$

Y, en general, es aconsejable trabajar con residuos estandarizados, definidos por:

$$r_{it} = \frac{e_{it}}{\hat{S}_R}, \quad (51)$$

$$t = 1, \dots, n_i, \quad i = 1, \dots, I.$$

En cuanto que las suposiciones del modelo son correctas, se verifica que $\frac{\varepsilon_{it}}{\sigma}$ es una muestra aleatoria simple de una distribución $N(0,1)$, con lo cual, es razonable suponer un comportamiento similar para r_{it} .

El estudio descriptivo, analítico y gráfico de la muestra y de los residuos posibilita tener un cierto conocimiento o una aproximación acerca del cumplimiento de las hipótesis básicas anteriormente citadas. Para ello se recomienda lo siguiente:



- I. Anterior al cálculo o estimación del modelo se han de obtener los estadísticos básicos de la variable respuesta Y según el factor.
- II. Representaciones gráficas interesantes para un análisis previo son: gráficas de puntos de Y según el factor y la gráfica de cajas múltiple de Y según el factor.
- III. Tras el ajuste del modelo y captados los residuos, se deben de obtener los estadísticos básicos de los residuos según el factor.
- IV. Analizar el gráfico de puntos de los residuos, o gráfico de cajas múltiples, histograma de los residuos, el gráfico de los residuos frente a las predicciones o el gráfico de los residuos frente al índice.

Hipótesis o supuestos:

- Independencia:

Uno de los supuestos básicos del modelo de regresión lineal es el de independencia entre los residuos (supuesto éste particularmente relevante cuando los datos se han recogido siguiendo una secuencia temporal). El supuesto de que las observaciones muestrales son independientes es fundamental en el estudio de los modelos de regresión lineal. Con ello se parte de la base de que los errores $\{\varepsilon_i\}_{i=1}^n$ son variables aleatorias independientes. La ausencia de independencia tiene lugar en el momento en que se trabaja con variables aleatorias que se observan a lo largo del tiempo, esto significa, cuando se trabaja con series temporales o con datos de series temporales. Por ello, una primera medida, para tratar de evitar la dependencia de las observaciones consiste en aleatorizar la recogida muestral.

En el caso de ausencia de independencia en los residuos, se daría una situación complicada ya que dicha circunstancia afectaría enormemente a los resultados del modelo de regresión ya que se obtendrían estimadores de los parámetros y predicciones ineficientes y los intervalos de confianza y contrastes igualmente serían nulos.



En caso de no existir independencia, la ausencia de aleatoriedad es muy complicada de corregir, algo que puede resultar especialmente grave, ya que es posible que invalide por completo las conclusiones del análisis estadístico (generándose conclusiones erróneas). Todas las expresiones utilizadas para las varianzas son incorrectas y, por tanto, los intervalos de confianza y las pruebas de hipótesis deducidos a partir de ellas, poseerán una confianza o una potencia distinta a la supuesta.

Para poder comprobar si existe o no dependencia entre los residuos se puede recurrir al llamado contraste de Durbin-Watson, en adelante DW (Durbin and Watson 1951). El Test de Durbin-Watson consigue calificar si existe autocorrelación en una Regresión lineal, sea simple o múltiple. De esta se pretende ver si los valores presentan algún tipo de dependencia en cuanto al orden de obtención. Si fuera así se estaría incumpliendo una de las condiciones del modelo y cuando se incumplen las condiciones de un modelo de Regresión lineal (normalidad, homogeneidad de varianzas, independencia de los datos) las estimaciones de los parámetros del modelo (los coeficientes del modelo) no tienen los criterios de calidad que se suponen.

El contraste desarrollado por Durbin y Watson es la prueba más frecuentemente utilizada para detectar la presencia de autocorrelación en los modelos de regresión. Este contraste permite verificar la hipótesis de no-autocorrelación frente a la alternativa de correlación de primer orden bajo un esquema autorregresivo AR(1), es decir:

$$x_t = c + \phi_1 x_{t-1} + \varepsilon_t \quad (52)$$

Analíticamente el contraste se representa del siguiente modo:

$$\left\{ \begin{array}{l} H_0 : \phi_1 = 0 \\ H_1 : 0 < |\phi_1| < 1 \end{array} \right. \quad (53)$$



La forma específica de la hipótesis alternativa marca unas cotas para el coeficiente de correlación; estas se necesitan para garantizar algunas características del modelo, en concreto que la varianza sea finita. El estadístico de Durbin-Watson proporciona información sobre el grado de independencia existente entre los residuos a partir de la siguiente expresión:

$$DW = \frac{\sum_{i=2}^n (e_i - e_{i-1})^2}{\sum_{i=1}^n e_i^2} \quad 0 \leq DW \leq 4 \quad (54)$$

(e_i se refiere a los residuos $e_i = Y_i - \hat{Y}_i$). El estadístico DW varía entre 0 y 4, y toma el valor 2 cuando los residuos son independientes. Los valores inferiores a 2 significan autocorrelación positiva y los mayores que 2 indican autocorrelación negativa. Se puede entender independencia entre los residuos cuando DW admite valores entre 1,5 y 2,5.

- Normalidad:

Uno de los supuestos básicos del modelo es que los errores del modelo, los residuos, siguen una distribución normal. La falta de normalidad significa la existencia de un problema. Ese problema afecta de forma significativa a la estimación de la varianza del modelo, por lo que no se conseguirán intervalos de confianza correctos. La normalidad de los residuos puede contrastarse gráficamente mediante el gráfico probabilístico normal. En ese gráfico la diagonal representa la ubicación teórica de los residuos en el supuesto de que estos sigan una distribución normal. Las desviaciones de dichos puntos con respecto a esa diagonal indican alteración de la normalidad de los residuos. El supuesto de que los errores siguen una distribución normal puede verificarse bien de forma gráfica o bien de forma analítica. Los gráficos que sirven para captar y estudiar la “normalidad” de los residuos son:

- El histograma de residuos.



- El gráfico de normalidad para los datos $i = 1^n$, en el que se representan los pares $i = 1^n$ y $i = 1^n$, donde F_n es la frecuencia relativa acumulada de la muestra y F es la distribución teórica (en este caso la distribución normal). Las curvas se muestran en unos ejes escalados de manera que los puntos $i = 1^n$ están sobre la recta $y = x$. Por lo que si los puntos $i = 1^n$ se muestran cercanos a esa recta, se aceptará el supuesto de normalidad.

- Homocedasticidad:

El supuesto de Homocedasticidad supone que la varianza de los residuos es constante y no varía en los diferentes niveles del factor. A la falta de homocedasticidad se la conoce como heterocedasticidad. Igualdad de varianzas de los residuos y los pronósticos.

La falta de homocedasticidad demuestra que si $(n_i = m, \forall_i = 1, \dots, I)$ los niveles de significación de las pruebas de hipótesis y los niveles de confianza de los intervalos que apenas se ven afectados por la existencia de heterocedasticidad, excepto que la varianza de la respuesta para algún tratamiento particular sea considerablemente mayor que para otros. En el caso de tamaños muestrales de grupos similares, la heterocedasticidad no afecta al $F - test$ ni a las diferentes técnicas de comparaciones múltiples siempre que:

$$\frac{\hat{S}_{Max}^2}{\hat{S}_{Min}^2} < 3 \quad (55)$$

Cuando el tamaño de la muestra es muy distinto, se analiza que:



- Si los tratamientos con tamaños muestrales (n_i) pequeños tienen mayor varianza la probabilidad de cometer un error de tipo I en las pruebas de hipótesis será menor de lo que se obtiene y los niveles de confianza de los intervalos serán inferiores a lo que se presupone;
- Si los tratamientos con tamaños muestrales (n_i) grandes poseen una varianza mayor, entonces resultará el efecto contrario.

Para verificar si se cumple el supuesto de homocedasticidad del modelo pueden llevarse a cabo distintos análisis descriptivos y gráficos como:

- El cálculo de la varianza (o desviación típica) de los residuos.
- El gráfico de cajas múltiple que da una idea de la distribución de los residuos.
- El gráfico de los residuos (e_{it}) frente a las predicciones (\bar{y}) resulta de gran interés porque en múltiples ocasiones, la varianza de los residuos por niveles, aumenta con las predicciones.

Por tanto, el análisis gráfico puede revelar posibles violaciones de la hipótesis de homocedasticidad, como por ejemplo, si se detecta que el tamaño de los residuos se incrementa o decrece de forma sistemática a medida que aumenta (\bar{Y}). Si dicho gráfico no muestra patrón ninguno entonces es posible aceptar que se cumple la hipótesis de varianza de los residuos constante.

Además existen distintos contrastes que sirven para captar la heterocedasticidad, entre los que podemos citar los de Cochran, Bartlett, Hartley, o Levene entre otros (Glass 1966).

Un modelo usual de heterocedasticidad es el que sigue:

$$Y_{ij} = \mu_i \varepsilon_{ij} \tag{56}$$



$$Y_{ij} = \mu_i \varepsilon_{ij} , j = 1, \dots, n_i , i = 1, \dots, I$$

, donde los ε_u son variables aleatorias de media 1 y varianza constante. En este modelo los grupos con una media mayor poseen una mayor variabilidad. De cara a la corrección de este problema se escogen logaritmos en el modelo y a partir de ahí el siguiente modelo con el que se trabaje si será homocedástico:

$$\bar{Y}_{ij} = \lg Y_{ij} = \lg \mu_i + \lg \varepsilon_{ij} = \bar{\mu}_i + \bar{\varepsilon}_{ij} \quad (57)$$

$$, j = 1, \dots, n_i \quad i = 1, \dots, I.$$

- Linealidad:

Este supuesto afirma que si no existe linealidad se advierte un error llamado de especificación. En el caso de que sean varias variables independientes, se ha de generar el gráfico para visualizar y analizar la regresión de forma que se pueda obtener los diagramas de dispersión parcial para cada una de las variables independientes. En ellos se ha anulado el efecto proveniente del resto de variables y de esta forma la relación que muestran es la relación neta entre las variables representadas.

- Colinealidad:

La no-colinealidad aparece por la inexistencia de colinealidad. Esta colinealidad puede ser “perfecta” si una de las variables independientes posee una relación lineal con otra/as independientes, o puede ser colinealidad “parcial” si entre las variables independientes se producen altas correlaciones. En relación al concepto de colinealidad podría hablarse también de Multi-colinealidad, algo que tiene lugar cuando entre las variables independientes o explicativas existe algún tipo de dependencia lineal entre ellas, o lo que es lo mismo, si se produce una fuerte correlación entre las mismas. La correlación no sólo se refiera a las distintas variables dos a dos, sino a cualquiera de ellas con cualquier grupo de los restantes.



3.3. Datos de Panel

El uso de los Datos de Panel como instrumento metodológico en el ámbito de la economía y otros campos es uno de los más extendidos. Los Datos de Panel se entienden como un conjunto de datos que alterna una dimensión temporal (serie de tiempo) y otra transversal (individuos). Según Cameron and Trivedi (2005), “Los datos de panel son observaciones de un mismo corte seccional para varios períodos de tiempo. Otros términos que se utilizan para ese tipo de base de datos son observaciones longitudinales o repetidas”

Un modelo econométrico de datos de panel es aquel que incluye una muestra de observaciones para un determinado periodo de tiempo. Este método combina ambos tipos de datos (dimensión temporal y estructural). El principal objetivo de aplicar y estudiar los datos en panel, es capturar la heterogeneidad no observable, ya sea entre variables económicas o de estudio así como también en el tiempo, dado que esta heterogeneidad no se puede detectar ni con estudios de series temporales ni tampoco con los de corte transversal. (Mayorga and Muñoz 2000).

Los datos de panel no significan sólo una suma de unidades individuales y series, sino que conforman una nueva visión de la exposición de la información, que ha generado el desarrollo de herramientas estadísticas que a su vez contribuyen a la estimación de parámetros con una serie de que propiedades ideales. En el campo de la Econometría el concepto de datos de panel se refiere a datos en dos dimensiones, de este modo, son generadas múltiples observaciones puntuales para cada unidad económica o cualquier otro tipo de unidad a estimar. Por lo general, los paneles de datos se diferencian unos de otros en función de su amplitud transversal y temporal. Así, los paneles con un número muy amplio de observaciones transversales y un número de períodos reducido se denominan Paneles Micro. En el caso contrario, número de periodos elevado e información transversal reducida se conoce con el nombre de Paneles Macro.(Rodríguez, Estive et al. 2010). Ejemplo de observaciones recogidas en forma de datos de panel (véase Ilustración 11).



	i=1	i=2	i=3	i = N
1980	3.5	2.5	4.2		1.8
1981	2.8	2.1	4.6		1.9
1982	2.6	1.9	4.1		1.7
1983	3.2	2.2	3.9		2.0

Ilustración 11. Ejemplo de introducción de observaciones en forma de datos de panel.

Con los datos de panel se obtiene información para cada uno de los individuos $i = 1, 2, 3 \dots N$ y para cada momento del tiempo $t = 1, 2, 3 \dots T$ tratándose de $N \times T$ observaciones. Las variables observadas se identifican para cada individuo, i , y momento del tiempo $t : Y_{it}$.

La técnica de datos de panel posee una serie de ventajas y también desventajas si los comparamos con modelos de series de tiempo y de corte transversal. Evitar problemas de agregación y facilitar el seguimiento del comportamiento individual en el tiempo son dos grandes ventajas de los datos de panel sobre las series temporales y los datos de corte transversal, el trabajo con este tipo de datos permite comparar observaciones de un individuo particular con su propio pasado. (Rodríguez, Estive et al. 2010)

Algunas de esas ventajas son las siguientes, (Mayorga and Muñoz 2000):

- ✓ Permite trabajar de forma simultánea con varios periodos de tiempo y los efectos individuales, y a su vez, abordar la cuestión que plantea el problema de la presencia de la “Endogeneidad”. La “Endogeneidad” se define como la presencia de correlación entre la variable a estimar o variable dependiente y el término de error y hace referencia a la relación causal entre las variables.
- ✓ Una ventaja de usar datos de panel de cara a establecer estimaciones es que se obtiene una mayor precisión de los parámetros estimados, debido



al gran número de observaciones utilizadas al combinar datos de corte seccional con las de series de tiempo. No obstante, esa ventaja sólo resulta cuando se corrige por cualquier tipo de correlación serial en las observaciones relacionadas con cualquier individuo. En ese tipo de estructura de datos, además, se puede controlar por efectos individuales no observables, que pudieran causar sesgos en estimaciones con otro tipo de datos.

- ✓ Permite realizar un análisis más dinámico al incorporar la dimensión temporal de los datos, lo que enriquece el estudio, particularmente en períodos de grandes cambios. El uso de esta metodología ayuda a analizar dos aspectos de suma importancia cuando se trabaja con este tipo de información y que forman parte de la heterogeneidad no observable:
 - Los Efectos Individuales específicos (Aquellos que afectan de manera distinta a cada uno de los agentes de estudio integrados en la muestra, los cuales son invariables en el tiempo y que afectan de manera directa las decisiones que tomen dichas unidades.
 - Los Efectos Temporales (Aquellos que afectan por igual a todas las unidades individuales del estudio).
- ✓ Esta técnica permite al investigador disponer de un mayor número de observaciones incrementando los grados de libertad, reduciendo la colinealidad entre las variables independientes o explicativas y mejorando la eficiencia de las estimaciones econométricas.
- ✓ Esta técnica ofrece la oportunidad de capturar la heterogeneidad no observable ya sea entre unidades individuales de estudio como en el tiempo. Basándonos en lo citado, la técnica ofrece la oportunidad de



poder aplicar una serie de pruebas de hipótesis para confirmar o rechazar dicha heterogeneidad y cómo capturarla.

- ✓ Los datos de panel suponen y agregan, en el análisis, el hecho de que los individuos a considerar son heterogéneos. Sin embargo los análisis basados en series temporales no tratan de controlar esta heterogeneidad pudiendo arriesgarse a obtener resultados sesgados.
- ✓ Permite estudiar mejor la dinámica de los procesos de ajuste. Esto ocurre en estudios sobre el grado de duración y permanencia de ciertos niveles de condición económica como por ejemplo el desempleo.
- ✓ Ofrece la capacidad de elaborar y testar modelos relativamente complejos de comportamiento en comparación con los análisis de series temporales.

En cuanto a sus desventajas:

En términos generales, las desventajas están asociadas a la técnica de datos de panel, se relacionan con los procesos cuyo objetivo es la obtención y el procesamiento de la información estadística sobre las unidades individuales de estudio, cuando esta se obtiene mediante encuestas, entrevistas o utilizando algún otro medio de recepción de datos. Como desventajas principales cabe señalar:

- ✓ Dificultades de implementación, tanto metodológicos como operativos. En trabajos desarrollados muchos años atrás se recurría al método conocido como de Mínimos Cuadrados Ordinarios (MCO u OLS sus siglas en inglés), sin embargo esta metodología presenta algunas críticas: Por un lado no permite el estudio de los efectos individuales (Castellacci 2008) y por otro, los estimadores son inconsistentes y pueden ser



insesgados cuando tratemos de analizar varios periodos de tiempo y efectos individuales.

3.3.1. Tipos de análisis con datos de panel

Existen dos tipos de análisis de cara a trabajar con datos de panel, los estáticos y los dinámicos (Arellano and de España 1992): Ambos permiten evaluar un conjunto de variables como explicativas de algún suceso o fenómeno en estudio y establecer así el conjunto de datos que presentan efectos individuales fijos o variables. No obstante este tipo de método presenta una serie de carencias, dentro de las cuales se encuentra la incapacidad de tratar de una forma adecuada el problema de la endogeneidad, por lo que no es posible estudiar desde un punto de vista evolucionista la dependencia del pasado o el proceso acumulativo.

A pesar de la presencia de este problema, los paneles dinámicos son igualmente recomendados ya que permiten incorporar en el modelo una estructura endógena, mediante la integración de efectos pasados a través de variables instrumentales. La endogeneidad causada por la relación bidireccional entre el fenómeno que queremos predecir y sus variables independientes o explicativas fue el principal problema que se intentó solventar haciendo uso de los modelos de regresión a través de paneles dinámicos.

Una primera fase tenía en cuenta la inclusión de la variable dependiente como explicativa. Debido a que no es posible incorporarla directamente por problemas de correlación, algunos autores lograron identificar instrumentos adecuados para poder abordar este tipo de análisis recurriendo los retardos de la variable dependiente (y) como regresor (es) (Y_{it-n}) .

Entonces podemos expresar, el segundo término de la ecuación como la variable dependiente retardada (Y_{it-n}) más las variables independientes (x_{it}) . Dada que la



casualidad está temporalmente afectada, el regresor se expresa como retardo de (y_t)

$$Y_t = \alpha Y_{t-n} + \beta X_{it} + \mu_{it} \quad (58)$$

Donde:

Y_t : Variable dependiente del individuo i en el tiempo t

Y_{t-n} : Variable dependiente del individuo i en el tiempo $t - 1$

α : Constante del modelo

β_i : Coeficiente de la variable i

X_{it} : Variable dependiente i en el tiempo t

3.3.2. Especificación general de un modelo de datos de panel.

Estimación con datos de panel

La especificación general de un modelo de regresión con datos de panel se expresa de la siguiente forma:

$$Y_{it} = a_{it} + b_1 x_{1it} + b_2 x_{2it} + \dots + b_k x_{kit} + U_{it}; \quad \text{con } i = 1, \dots, n \text{ y } t = 1, \dots, T \quad (59)$$

Donde i hace referencia al individuo o a la unidad de estudio (corte transversal), t a la dimensión temporal, a es un vector de interceptos que puede tener entre 1 y $n + t$ parámetros, b es un vector de k parámetros y X_{it} es la i -ésima observación al momento t para las k variables explicativas o independientes X_1, X_2, \dots, X_k .



En este caso, la muestra total de las observaciones en el modelo viene dado por $n \times T$

A partir de este modelo general, y partiendo de ciertos supuestos y restricciones acerca del valor de algunos de los parámetros, se pueden derivar algunas otras variantes de modelos de datos de panel.

a) Componentes del error

Es habitual interpretar los modelos de datos de panel a través de sus componentes de errores. El término de error U_{it} incluido en la ecuación (59), puede descomponerse de la siguiente forma:

$$U_{it} = u_i + v_t + w_{it} \quad (60)$$

Donde:

- u_i representa los efectos no observables que difieren entre las unidades de estudio pero no en el tiempo.
- v_t es identificado con efectos no cuantificables que fluctúan en el tiempo pero no entre las unidades de estudio.
- w_{it} hace alusión al término de error puramente aleatorio.

La mayoría de las aplicaciones con datos de panel utilizan el modelo de componente de error conocido como “one way” (Rafael Bustamante (DE) para el cual $v_t = 0$. Este tipo de análisis supone que no existen efectos no cuantificables que varíen en el tiempo pero no entre las unidades individuales de estudio. Existe además otro modelo como es el denominado “two way” en el que el componente de error $v_t \neq 0$, a través del cual se pretende capturar efectos temporales específicos que no están incluidos en la regresión.

3.3.3. Cuándo utilizar Datos de Panel

Una ingente cantidad de estudios de investigación realizados en los últimos años han recurrido a este tipo de metodología de Datos de Panel. La razón se ha debido al gran avance experimentado en las bases de datos, las cuales se han construido recogiendo cada vez más volumen de información de individuos a lo largo del tiempo. Otros trabajos anteriores que utilizaron como método las regresiones lineales, habían sido analizados a partir de técnicas de series de tiempo y de sección cruzada. Una extensión a las técnicas anteriores es lo que se denomina “Pool” de datos, según la cual cada individuo en un momento concreto de tiempo constituye por sí mismo una observación (Labra and Torrecillas 2014). El desarrollo de técnicas de datos de panel, sin embargo, puede tratar de forma independiente el conjunto de datos de un individuo en el tiempo, lo que se conoce como efectos individuales (α_i)

Como se puede ver en la Ilustración 12, un conjunto de datos pueden ser analizados de diferente manera. La figura 12A muestra una serie de observaciones analizadas, mientras que la figura 12B, presenta un análisis de regresión lineal, sin diferenciar individuos, es decir como un Pool de datos.

Por otro lado, la figura 12C muestra como el conjunto de datos está compuesto en la realidad por tres subgrupos (individuos observados a través del tiempo) y que por tanto, podría llevarse a cabo otro tipo de análisis que considere esta condición.

La figura 12D representa las regresiones para cada subgrupo o individuo (individuo 1, 2 y 3), de lo que se entiende que cada uno posee un comportamiento diferente y que debe ser analizado teniendo en cuenta esta particularidad.

Por último, en la figura 12E se presenta una regresión lineal teniendo en cuenta los efectos individuales. La función final para el conjunto de individuos (línea continua de color negro) es totalmente distinta a la que obtendríamos si el análisis no se hubiese hecho a través de las técnicas de panel.



Concluyendo, este ejemplo gráfico, se puede deducir que es importante considerar los efectos individuales cuando estos existen, ya que el análisis y sus resultados pueden variar al usar uno u otro método.

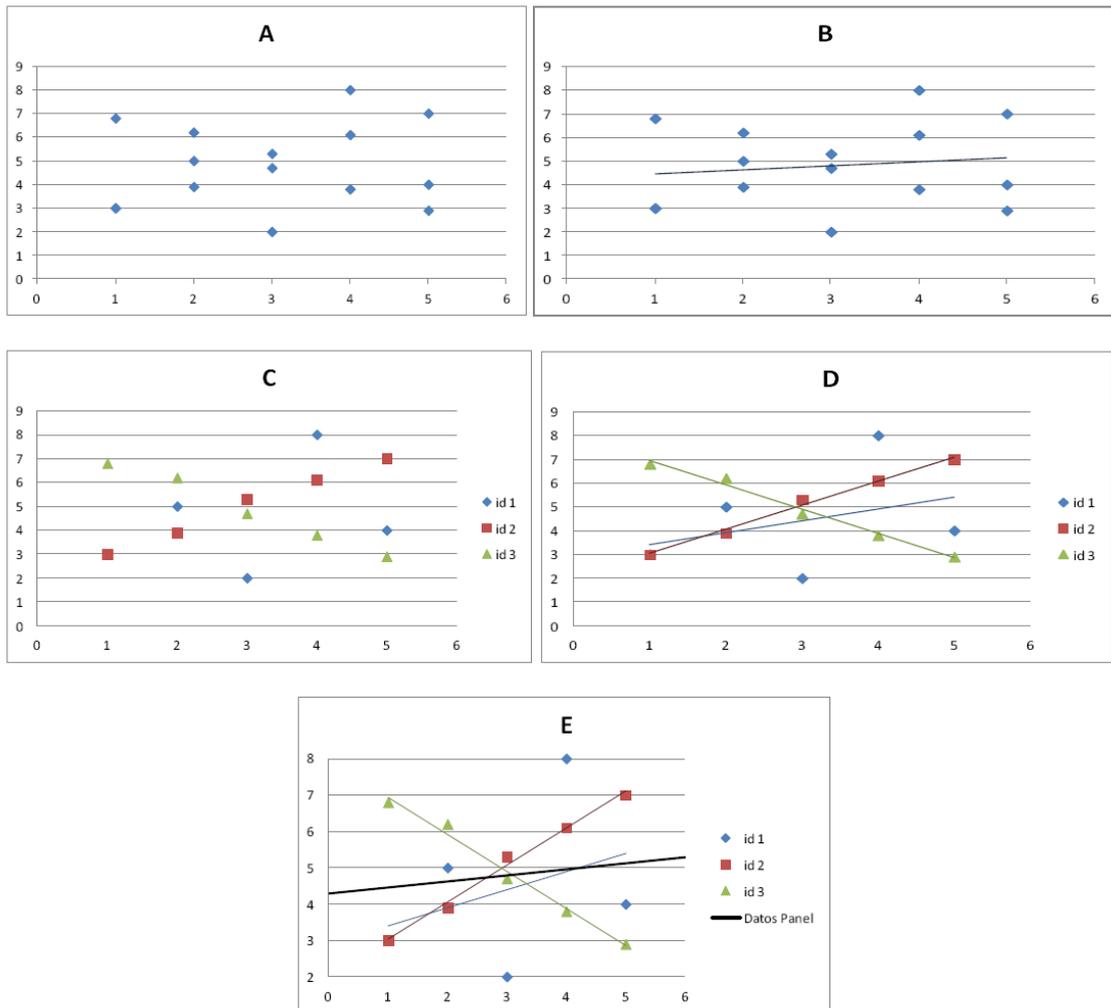


Ilustración 12. . Esquema gráfico de los efectos individuales α_i
(Labra and Torrecillas 2014)

Cabe señalar que dentro de esta última técnica, datos de panel, existen dos grandes métodos: los basados en Paneles Estáticos y los basados en Paneles Dinámicos, cuya principal diferencia radica en la capacidad y forma de tratar la endogeneidad de las variables.



3.3.4. Supuestos necesarios para usar Datos de Panel.

Para poder hacer uso de este procedimiento es necesario que se cumplan dos requisitos:

- I. Tener un conjunto de individuos pudiendo ser estos, por ejemplo, empresas, líneas de transporte de un servicio de transportes, o diferentes modos de transporte.
- II. La existencia de observaciones de los mismos individuos (al menos parcialmente) durante un determinado periodo de tiempo (un conjunto de años).

Además de los citados requisitos existen también una serie de limitaciones. En general, para ambos métodos (datos de panel estáticos y dinámicos) se recomienda el uso de bases de datos con un número de individuos (n) elevado y un período de tiempo (t) pequeño. La literatura no entrega un valor exacto de (n) y (t) sin embargo, algunos autores han indicado que un adecuado n podría ser mayor a 100, mientras que el (t) no debiera sobrepasar 15, e idealmente ser inferior a 10 en el caso de paneles dinámicos (Roodman 2009).

Esta metodología se ha aplicado con más frecuencia en los estudios de carácter micro debido a que estas bases de datos tienen, normalmente, un volumen grande de observaciones para franjas de tiempo más pequeñas. Un caso distinto observamos para los trabajos macro, en cuyas bases de datos es frecuente encontrar una relación entre número de individuos y el periodo de tiempo es menor, lo que reviste serias dificultades para estudios con variables endógenas.

Sí la relación n/t es pequeña, la literatura muestra diversas alternativas para solucionar este problema, las que van desde dividir el periodo de tiempo de análisis, agrupar los años en bienios, trienios, o el periodo que sea más razonable para el tipo de estudio que



se esté realizando. Sin embargo, esto tiene la limitación de reducir la información empleada para el análisis, afectando la varianza de muestra.

Los análisis de datos de panel también se pueden aplicar cuando el panel está incompleto, y no se encuentra completa toda la serie de datos para un individuo, o faltan individuos para ciertos años de una misma variable. Cuando los datos con los que no se cuenta información son elevados, pueden surgir limitaciones en el análisis, generando inconsistencia en los resultados, o simplemente impidiendo realizar la estimación del modelo o función.

3.3.5. Elección entre los distintos tipos de datos de panel. Datos de Panel: Estáticos o dinámicos y posibles correcciones. Comprobación de la Endogeneidad.

El primer paso de cara a seleccionar el modelo es la identificación de la presencia de endogeneidad. Existen varias formas para poder identificarla, dentro de las cuales se pueden citar:

1. Mediante la detección de variables endógenas.
2. Analizando la correlación entre los errores de X_i y los errores del modelo.

$$Cov(x_i, \varepsilon) \neq 0 \tag{61}$$

3. Desarrollando el test de endogeneidad denominado Durwin Wu Hausman, el cual compara estimaciones efectuadas mediante diferentes métodos: Mínimos Cuadrados Ordinarios (MCO) y modelo endógeno utilizando variables instrumentales.



4. En el caso de análisis con paneles dinámicos el test de Sagan/Hansen analiza la validez de los instrumentos, por lo que es posible si nos encontramos ante un modelo endógeno.

En el supuesto de que se consideren las consideraciones anteriores y no se detecte endogeneidad y no se tengan indicios de sospecha de ella, entonces el procedimiento más recomendado es el de la utilización de paneles estáticos o MCO (Mínimos Cuadrados Ordinarios). Si por el contrario, la literatura o los test dan indicios o confirman la presencia de endogeneidad, necesariamente deberemos utilizar un método que permita tratarla. Como por ejemplo el Test de Hausman (Hausman 1978) o el Test de Wu.

3.3.6. Estimador con variables instrumentales.

El método de Variables Instrumentales permite la consecución de estimadores robustos de los parámetros en casos en los que el estimador MCO es inconsistente (por falta de variables significativas, errores de medida o simultaneidad)(Arellano and Bover 1990).

En términos generales, tenemos que a partir del modelo:

$$Y = \beta_0 + \beta_1 X + \varepsilon \tag{62}$$

que $C(X, \varepsilon) \neq 0$ los parámetros que resultan con mayor interés β_0 y β_1 no coinciden con los parámetros de la proyección lineal $L(Y / X)$ y los estimadores MCO (Mínimos Cuadrados Ordinarios) $\hat{\beta}_0$ y $\hat{\beta}_1$:



$$\begin{aligned}
 p \lim \hat{\beta}_1 &= \frac{p \lim \left(\frac{1}{n} \sum_i X_i Y_i \right)}{p \lim \left(\frac{1}{n} \sum_i X_i^2 \right)} = \frac{p \lim \left[\frac{1}{n} \sum_i X_i (\beta_1 X_i + \varepsilon_i) \right]}{p \lim \left(\frac{1}{n} \sum_i X_i^2 \right)} = \beta_1 + \frac{p \lim \left(\frac{1}{n} \sum_i X_i \varepsilon_i \right)}{p \lim \left(\frac{1}{n} \sum_i X_i^2 \right)} \\
 &= \beta_1 + \frac{C(X, \varepsilon)}{V(X)} \neq \beta_1
 \end{aligned} \tag{63}$$

Donde tenemos, $y_i = Y_i - \bar{Y}$, $x_i = X_i - \bar{X}$

En el modelo:

$$Y = \beta_0 + \beta_1 X + \varepsilon \tag{64}$$

Donde, $C(X, \varepsilon) \neq 0$,

Se precisa de “información adicional” (en forma de variables a incluir nuevas o adicionales) y así poder obtener estimaciones de una mayor consistencia tanto de β_0 como de β_1 .

Partiendo del supuesto que tenemos una variable Z a la que denominaremos como variable instrumental que:

- No presente correlación con el error del modelo **(a)** $C(Z, \varepsilon) = 0$
- Exista correlación con la variable endógena X : **(b)** $C(Z, X) \neq 0$

Con la ayuda de Z como instrumento, es posible obtener estimadores robustos y fiables de β_0 y β_1 .



3.4. Validación de los modelos

Llegado el punto el cual el modelo ha sido ya identificado y se ha programado, el paso siguiente consistiría en aplicarlo al problema concreto que nos ocupe. En primer lugar es necesario seguir un protocolo de evaluación para ese modelo. Una parte importante de los estudios científicos se decantan por una evaluación de los resultados cualitativa y en aquellos donde se puede cuantificar el análisis eligen el uso de estadísticos simples como el Error Cuadrático Medio, el Error Absoluto Medio o el Sesgo.

Para la validación del modelo, por tanto, se pueden adoptar los siguientes métodos estadísticos:

- El de Error Cuadrático Medio (RMSE).
- El de Error Absoluto Medio (MAE)
- El de Error Absoluto Medio Normalizado (NMAE)
- El de Sesgo (BLAS).

Estos estadísticos permiten medir la precisión de la simulación o de aquello que se busque prever. El término de “precisión” lo entendemos como el promedio del grado de correspondencia entre pares individuales de valores pronosticados y valores observados. Los valores observados serán aquellos obtenidos a partir de una fuente determinada para convertirla en una variable dependiente o variable a predecir y que sirvieron para construir una serie temporal.

3.4.1. El Error cuadrático Medio

Para calcular dicha precisión se recurre al Error Medio Cuadrático (RMSE), definido como:

$$RMSE = \sqrt{\sum_{i=1}^N \frac{(\phi - \phi_{iobs})^2}{N}} \quad (65)$$



en donde:

ϕ_i es el valor pronosticado para la celda i

ϕ_{iobs} es el valor observado para la celda i

N es el número de valores analizados

El Error cuadrático Medio nos ofrece la medida de las diferencias en promedio entre los valores pronosticados y los observados.

3.4.2. Otros Estadísticos:

Existen además del citado estadístico de Error Medio Cuadrático otros estadísticos que nos describen una información similar como son:

El Error Absoluto Medio expresado como:

$$MAE = \sum_{i=1}^N \frac{|\phi_i - \phi_{iobs}|}{N} \quad (66)$$

Para tener en cuenta el peso del error respecto al valor de la variable medida se normaliza el error absoluto, teniendo el error absoluto medio normalizado:

$$NMAE = \sum_{i=1}^N \frac{|\phi_i - \phi_{iobs}|}{N \phi_{iobs}} \quad (67)$$



Por último tenemos el Sesgo (BIAS) que proporciona información sobre la tendencia del modelo a sobreestimar o subestimar una variable y cuantifica el error sistemático del modelo:

$$BIAS = \sum_{i=1}^N \frac{(\phi_1 - \phi_{iobs})}{N} \quad (68)$$

3.5. Deflactación de variables monetarias

Cuando trabajamos con una serie estadística de datos sobre la valoración de alguna magnitud económica (renta per cápita, recaudación o ingresos, precios de tarifas, entre otros), lo usual es que la valoración monetaria de estos datos se realice a precios corrientes de cada período, o también denominados precios nominales, estos son los precios establecidos; los que marcan la etiqueta. El problema radica en que si pretendemos comparar precios en el tiempo, en la medida en que los precios experimentan fluctuaciones de unos periodos a otros, la serie así representada no permite llevar a cabo comparaciones.

Para solucionar este problema hay que expresar la serie en términos de precios constantes de un determinado periodo (año base). Los precios reales son los precios al que se le han descontado ya el efecto de la inflación. De esta forma podemos comparar los precios reales de un determinado producto en el año 20011 y de uno similar en el año 2012, descontando el “encarecimiento de la vida en esos años”. Esta sería la forma conveniente y correcta de establecer comparaciones de precios a lo largo de una serie de tiempo. La transformación de la serie original a la serie valorada en precios constantes se llama deflactación, y el índice a través del cual se puede pasar de una serie a la otra se llama deflactor. “Un deflactor es un índice de precios que se obtiene por comparación entre una magnitud en corrientes y la misma magnitud en pesetas constantes. La base de este índice de precios será, evidentemente, la base en la que venga expresada la magnitud en constantes” (Mahía 2001).

Para deflactar las variables que contienen datos económicos/monetarios hay que utilizar un índice de precios (IPC) y elegir un año como “año base” cuyo IPC será igual a 100. A los años



superiores habrá que ir sumándoles la inflación al año base (102, 104.5...) y por detrás del año base habrá que ir restando la inflación (99.5, 97...). Así se hace una columna con los años que queremos comparar. Será nuestro índice de precios.

Una vez obtenido el índice de precios, se aplica la siguiente fórmula:

$$VPC = \frac{Vpcorr}{IP} * 100 \quad (69)$$

Donde:

- VPC es el valor en precios constantes.
- $Vpcorr$ es el valor en precio corriente.
- IP es el índice de precios.

El Deflactor se construye de la siguiente manera:

$$DeflactorPIB = \frac{PIB_{Nominal}}{PIB_{Real}} * 100 \quad (70)$$

El deflactor del PIB es el cociente entre el PIB nominal y el PIB real (el PIB real se obtiene dividiendo el PIB nominal por su deflactor). En el año base toma siempre el valor 1 (ó 100) y, por tanto, se trata de un número índice. El deflactor del PIB es un indicador de la evolución de los precios de una economía que incluye todos los bienes producidos (no sólo los de consumo como el IPC). El deflactor del PIB en el año t , P_t , es el cociente entre el PIB nominal y el PIB real en el año t :



$$P_t = \frac{Y_t \text{€}}{Y_t} \quad (71)$$

El PIB nominal es igual al deflactor del PIB multiplicado por el PIB real:

$$\$Y_t = P_t Y_t \quad (72)$$

El deflactor del PIB es lo que se denomina número-índice: se fija en 100 en el año base.

La tasa de variación del deflactor del PIB es igual a la tasa de inflación:

$$\frac{(P - P_{t-1})}{P_{t-1}} \quad (73)$$

La construcción de un deflactor tiene como objetivo crear un instrumento que permita aislar el efecto de la variación general de los precios en una serie que originalmente se encuentra expresada en precios corrientes (recaudación mensual, coste medio del litro de combustible, tarifa del billete en servicio público de transportes, valor del Pib y valor del pib per cápita), de modo que éste pueda centrar la atención de su análisis en la evolución de las cantidades físicas (volumen), expresadas a precios constantes (Mahía 2001).

Para la construcción del deflactor de las variables monetarias existentes en una serie temporal se toma el año inicial de la serie como base y se calcula a partir de los índices que proporciona para tal efecto el INE (Instituto Nacional de Estadística).



3.6. Cálculo de elasticidades

En ocasiones, no basta conocer sólo si una cantidad de cualquier producto demandada aumenta o disminuye en respuesta a un cambio de precio sino que es más interesante saber en "cuanto" aumenta o en "cuanto" disminuye (Frank and Piras 2003).

La elasticidad puede definirse como la capacidad de reacción de una variable en relación con cambios en otra. Este término mide la variación porcentual que experimenta una al cambiar otra. Estas variables son cuantitativas. Se recurre a este concepto, para examinar la relación entre precios y demanda (Causelo 2012).

La elasticidad, por tanto, constituye un instrumento mediante el cual medimos el grado en que la cantidad demandada responde (cambia) a las variaciones en el precio del propio bien. Dicho de otra manera, la elasticidad es la sensibilidad de la demanda al precio.

A continuación puede observarse en la siguiente ilustración (véase Ilustración 13) que una misma variación de un precio tiene como consecuencia una variación diferente de la cantidad demandada, dependiendo de la forma concreta que tenga la curva de demanda. Se entiende que las unidades de medida empleadas son las mismas para una y otra curva. Pues en caso contrario curvas de demanda iguales serían visualizadas de forma diferente:

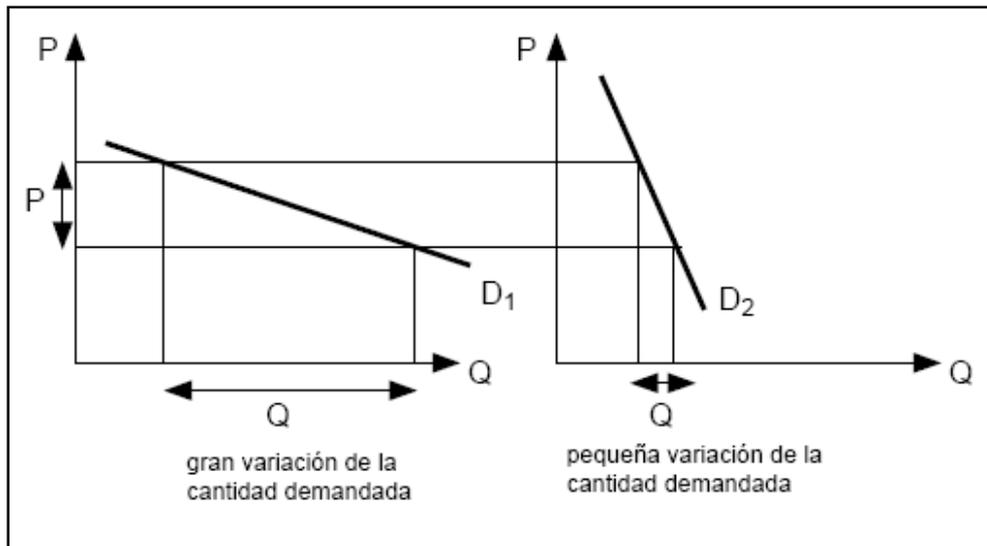


Ilustración 13. Variaciones de la demanda. (Frank and Piras 2003).

Si contamos con una función de n variables expresada como $y = f(x_1, x_2, \dots, x_n)$, el concepto de elasticidad se emplea para cuantificar la sensibilidad de la variable dependiente y ante una fluctuación en el valor de una de las variables independientes, x_i , cuando se mantiene constante el valor que adquieren las demás variables independientes o explicativas. Sirva como ejemplo el cambio en el valor de y asociado al cambio en el valor de x_i que se puede observar en la tabla siguiente:

y	x_i
100	10
140	12

Tabla 3. Cambio de valores en x_i con respecto a y (Causelo 2012)

Un incremento de dos unidades en la variable x_i ha supuesto un incremento de cuarenta unidades en la variable y . Para poder comprobar cómo ha respondido la variable y ante esa variación en x_i se podría recurrir como primera medida a la tasa media de variación (TMeV):



$$TMeV_{x_i}^y(10,12) = \frac{\Delta y}{\Delta x_i} = \frac{40}{2} = 20 \quad (74)$$

Lo cual indicaría que por cada unidad de aumento en x_i la y aumenta, en promedio, veinte unidades.

De otra forma, se podría calcular la tasa a la que ha variado la y para después poder compararla con la tasa a la que ha variado x_i . En ese caso el incremento de la y ha representado un 40% ($\frac{\Delta y}{y^0} = 0,4$), variación atribuible a un aumento del 20% ($\frac{\Delta x_i}{x_i^0} = 0,2$) de la x_i . Si se hace uso del cociente de ambas tasas para medir la sensibilidad de la y ante el cambio en la x , entonces se está usando lo que se denomina como “Elasticidad Arco”:

$$\varepsilon_{y/x_i}(10,12) = \frac{\frac{\Delta y}{y^0}}{\frac{\Delta x_i}{x_i^0}} = \frac{40\%}{20\%} = 2. \quad (75)$$

Para este ejemplo, la elasticidad tiene como valor 2, indicando que por cada 1% que se ha incrementado la variable x_i la y ha crecido un promedio de un 2%. El valor de la elasticidad no está supeditado a las unidades de medidas utilizadas, siendo esta una de sus principales ventajas frente a la tasa media de variación, cuyo valor si depende de las unidades en que sean medidas y y x_i .

En definitiva, la “elasticidad arco” nos informa acerca de la variación en porcentaje (%) de la variable dependiente, y , por cada 1% de fluctuación de la variable independiente o explicativa x_i cuando se comportan de forma constante el resto de variables independientes. Cabe señalar que el valor de la elasticidad definida de esta forma dependería, además de del punto de la función que consideremos, (x, y) , del valor de Δx_i .



$$\varepsilon_{y/x_i}^A (X, \Delta x_i) = \frac{\Delta y}{\Delta x_i} \frac{x_i}{y} \quad (76)$$

En resumen, en una curva de demanda, el coeficiente de elasticidad precio de la demanda entre dos puntos se denomina elasticidad arco (véase Ilustración 14). Como el coeficiente de elasticidad precio de la demanda tiende a ser diferente en todos los puntos a lo largo de la curva, la elasticidad arco es solamente una estimación.

En los modelos teóricos es habitual servirse de funciones que son continuas y diferenciables, en cuyo caso en lugar de la “elasticidad arco” se usa la “elasticidad punto” (véase Ilustración 14) . La “elasticidad punto” es a la “elasticidad arco”, lo que el cociente de incrementos (tasa media de variación) es a la derivada parcial (tasa marginal de variación). Por lo que, si tenemos la expresión $y = f(x_1, x_2, \dots, x_n)$ y otro valor de la variable independiente x_i^0 se define la elasticidad en dicho punto como:

$$\varepsilon_{y/x_i} (X^0) = \lim_{\Delta x_i \rightarrow 0} \frac{\frac{\Delta y}{y^0}}{\frac{\Delta x_i}{x_i^0}} = \lim_{\Delta x_i \rightarrow 0} \frac{\Delta y}{\Delta x_i} \frac{x_i^0}{y^0} = \frac{\partial y}{\partial x_i} (X^0) \frac{x_i^0}{y^0} \quad (77)$$

En resumen, para un punto cualquiera $X = (x_1, \dots, x_n)$ resulta:

$$\varepsilon_{y/x_i} (X) = \frac{\partial y}{\partial x_i} (X) \frac{x_i}{y} \quad (78)$$

Este concepto se explica como la variación en porcentaje (%) de la variable dependiente por cada 1% de fluctuación de la variable explicativa ante un incremento infinitesimal de esta última. Como sucede con el caso del concepto de derivada y de derivada parcial, el valor de la elasticidad así definida es igualmente una función cuyo valor se subordina al punto x que se considere. En relación a la “elasticidad de punto”, a medida que el arco va siendo más pequeño, la aproximación mejora y se aproxima a un valor puntual en el límite, cuando el cambio en el precio tiende a cero.

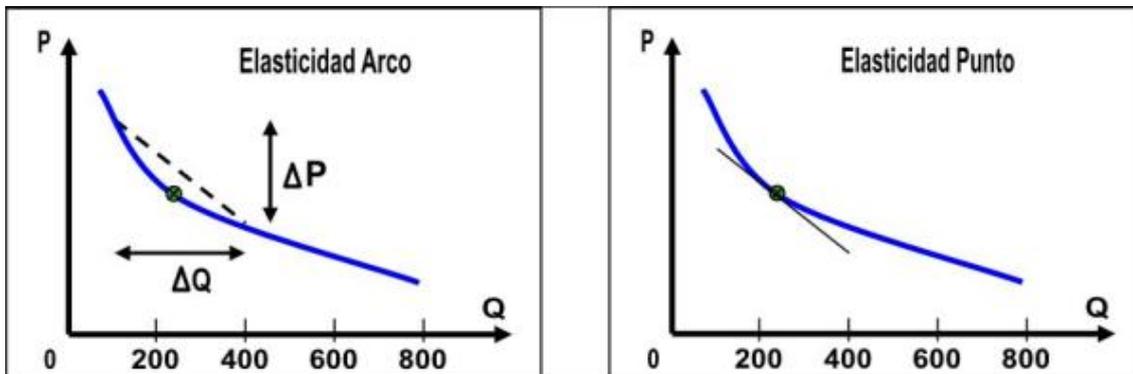


Ilustración 14. Tipos de elasticidades. Elasticidades Punto y Arco.

Una curva de demanda rectilínea que contacte en ambos ejes se considera elástica por encima de su punto medio, es unitaria en el punto medio y es inelástica por debajo de su punto medio. Siguiendo ese criterio la demanda puede ser:

- Elástica: cuando ($\varepsilon > 1$)
- Inelástica: cuando ($\varepsilon < 1$)
- Unitaria: cuando ($\varepsilon = 1$)

Otro método que sirva para poder calcular las elasticidades de las distintas variables independientes que puedan ser de interés es a través de la estimación en un primer paso, de un modelo estático de regresión de tipo logarítmico log-log. A partir de su expresión funcional, los parámetros estimados pueden interpretarse directamente como elasticidades, es decir, como el cambio porcentual de equilibrio de la variable dependiente ante los cambios porcentuales en las variables independientes.





CAPÍTULO 4

Descripción del área de estudio y análisis de datos



4. DESCRIPCIÓN DEL AREA DE ESTUDIO Y ANÁLISIS DE DATOS

4.1. Área de Estudio

Los modelos de demanda de transporte en bus se han estimado con datos provenientes de la ciudad de Santander (España) y obtenidos a partir de la información proporcionada por parte del servicio municipal de transportes (TUS, Transportes Urbanos de Santander) y de diversas fuentes como el INE (Instituto Nacional de Estadística), el ICANE (Instituto cántabro de estadística) y otras fuentes de la administración tanto de Santander como de la comunidad de Cantabria entre las que cabe señalar al Servicio Cántabro de empleo.

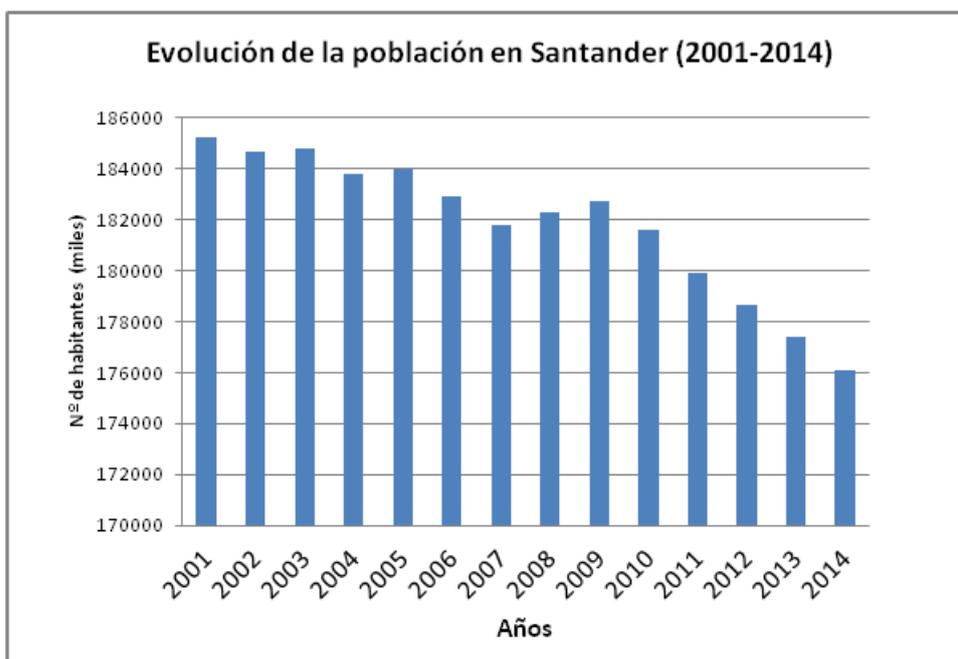


Gráfico 2. Evolución demográfica de Santander (2001-2014). Elaboración propia a partir de los datos obtenidos del ICANE (Instituto Cántabro de Estadística). Octubre de 2015.



Santander es una ciudad de tipo medio, que cuenta actualmente con algo más de 176.000 habitantes en su núcleo urbano y más de 280,000 en su área de influencia (Coppola, Ibeas et al. 2013). En lo que a población se refiere, la ciudad de Santander está experimentando una lenta pero progresiva pérdida de población algo que queda constatado a través de su censo en el que en 2001 se registraba una cifra de 185.231 habitantes y en el año 2014 esa cifra apenas superaba los 176.000, lo que representa una pérdida de población aproximada del 5% (aprox. 9.000 habitantes) con respecto al año 2001 (véase Gráfico 2).

La pérdida de población de Santander se ha producido de la misma manera que los municipios limítrofes o próximos, situados en el arco de la bahía de Santander, han ido incrementando su número de efectivos demográficos (véase Gráfico 3).

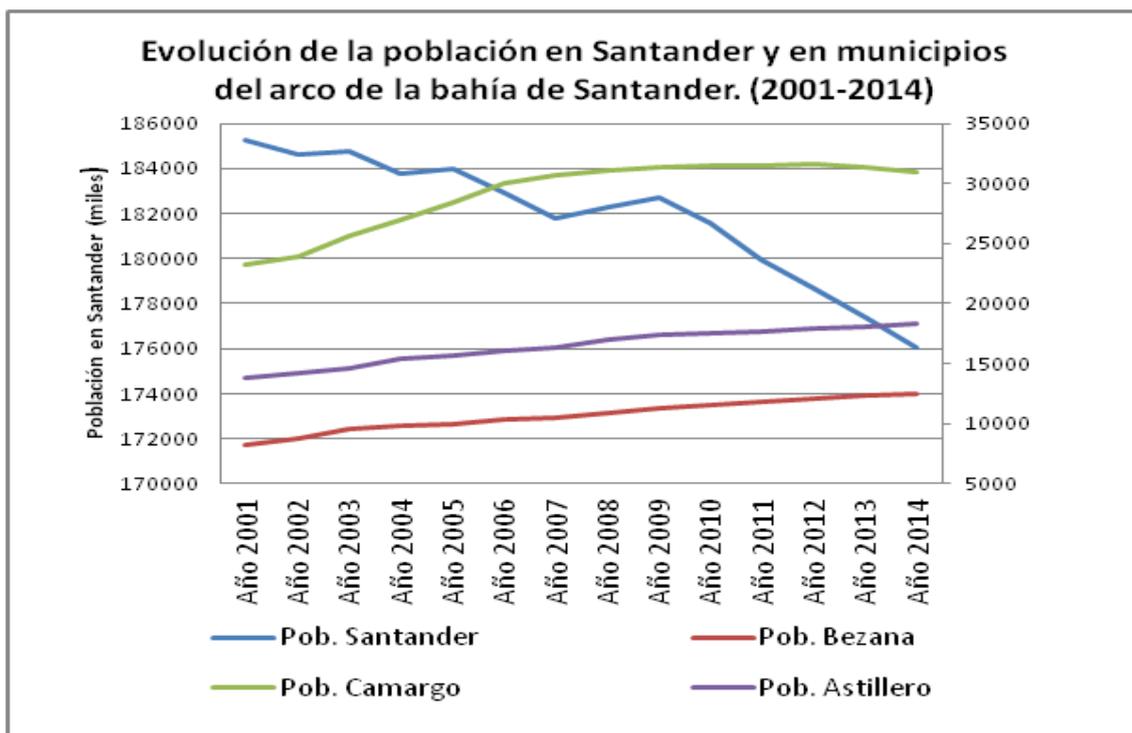


Gráfico 3. Evolución demográfica de la ciudad de Santander con respecto a municipios próximos (Arco de la Bahía) (2001-2014).). Elaboración propia a partir de los datos obtenidos del ICANE (Instituto Cántabro de Estadística). Octubre de 2015.

En cuanto a la estructura de población, en Santander (véase Gráfico 4), el mayor grupo de edad es el que engloba al de adultos (de 31 a 65 años) que prácticamente suponen el 50% de la población total en el año 2014. El siguiente grupo de edad con mayor número de individuos es el de los jubilados (mayores de 65 años) un 22,3 % del total. En el grupo de edades comprendidas entre los 0 y los 31 años se concentra el resto de la población (algo menos del 28%).

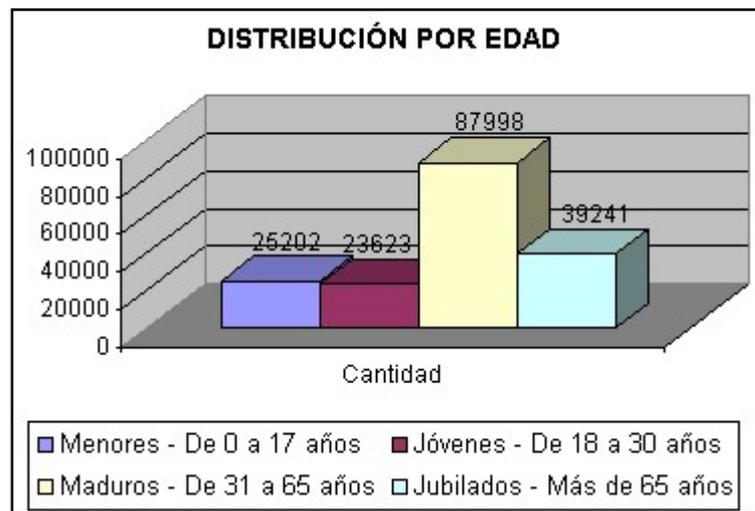


Gráfico 4. Estructura de la población de Santander por edades (2014). Fuente: Ayuntamiento de Santander.

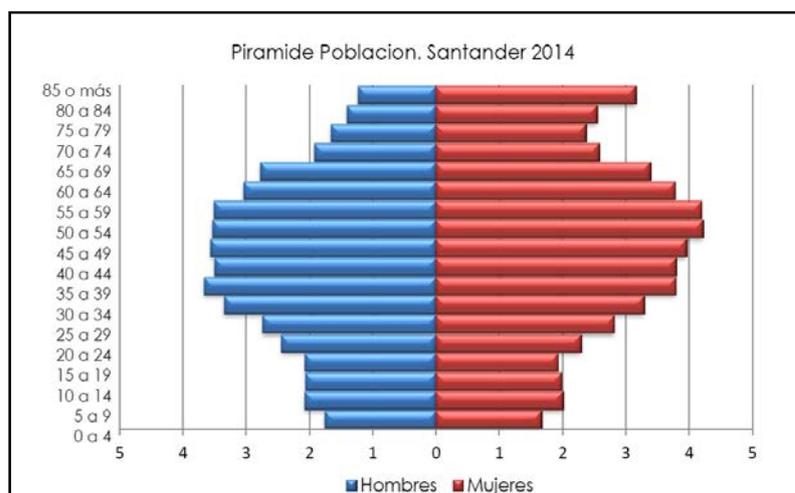


Gráfico 5. Pirámide de Población de Santander (2014) Fuente: Padrón Municipal de Habitantes a 1 de enero de 2014. Ayuntamiento de Santander

Si nos fijamos más detenidamente en la pirámide de población de Santander (véase Gráfico 5), se puede observar cómo existe también un gran número de habitantes que se concentran en los grupos comprendidos entre los 50 y los 65 años, lo cual indica que en los próximos 10-15 años el porcentaje de población mayor de 65 años se incrementará de forma notable. Por otro lado las franjas que representan los grupos de edad joven (menores de 20 a 24) se van estrechando cada vez más lo que hace que estemos ante una típica pirámide relacionada con sociedades desarrolladas.

Con una extensión regional de 5.341,39 Km² el término municipal en el que se emplaza la capital autonómica representa menos del 1% de su superficie. Se trata por ello de un término relativamente pequeño, de aproximadamente 36 Km² ubicado en la zona central de la franja costera de la región. (Véase Ilustración 15).



Ilustración 15. Situación del término municipal de Santander. Fuente: PGOU de Santander.

Sobre Santander recae la capitalidad de la comunidad autónoma de Cantabria, lo cual la convierte en el principal núcleo de población de la región llegando a aglutinar algo más del 30% de la misma y con unos niveles de densidad de población cercana a los 5.000 habitantes por kilómetro cuadrado, lo que representa casi un tercio de la población regional ocupando apenas un 1% de la superficie de la comunidad autónoma (véase Ilustración 16). La importancia del municipio de Santander no viene dada por su extensión, sino por el peso demográfico, cultural y económico que posee en el contexto regional, proceso en el que no es ajeno el hecho de ser la capital de la comunidad autónoma de Cantabria.

En

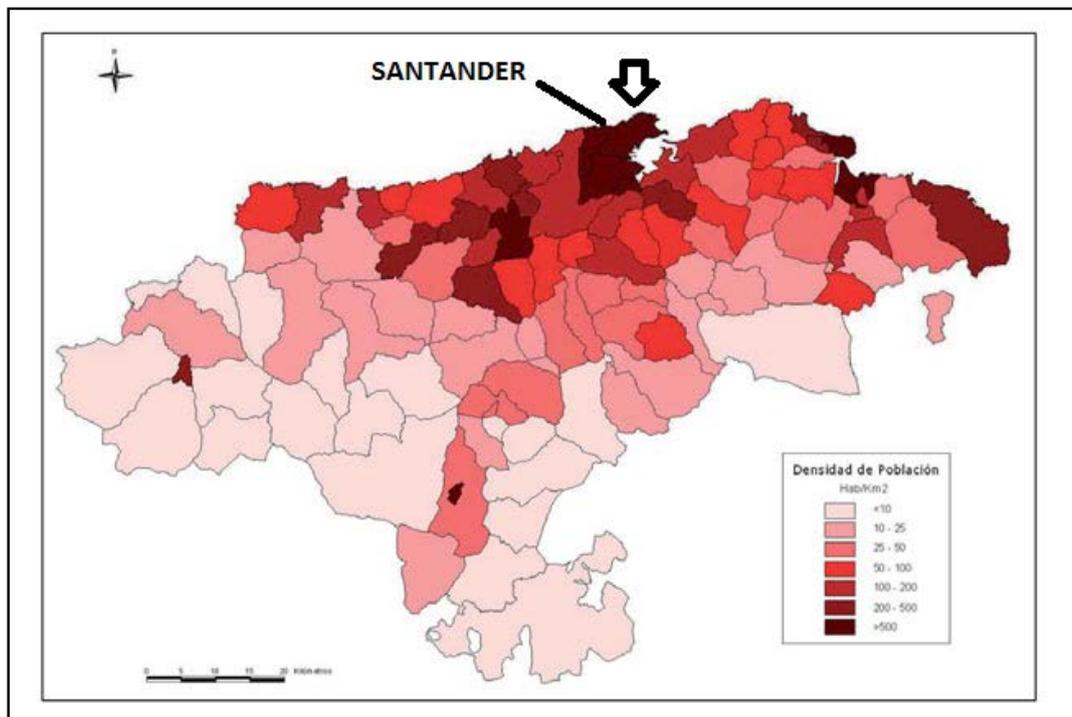


Ilustración 16. Densidad de población de los municipios de Cantabria. Fuente: ICANE.

cuanto a su estructura espacial, Santander se compone de un núcleo tradicional fuertemente consolidado, establecido en su origen en el cerro de Somorrostro (ubicación actual de la Catedral), zonas de reciente expansión (Valdenoja, Alisal y algunas áreas de Peñacastillo), y por las entidades periurbanas de Cueto, Monte, San Román y Peñacastillo (véase Ilustración 17).



Ilustración 17. Estructura del término municipal de Santander. Entidades de población.

Fuente: PGOU

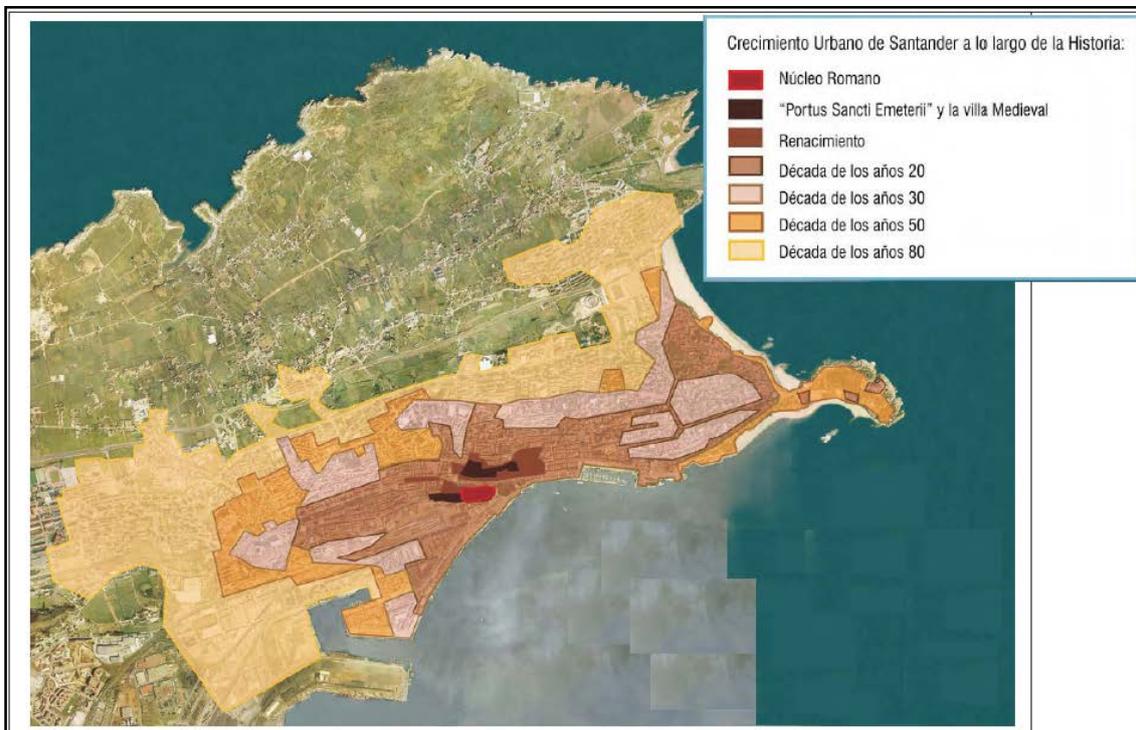


Ilustración 18. Evolución histórica del poblamiento de Santander. Fuente P.G.O.U de Santander.



El crecimiento del espacio urbano de la ciudad tuvo su mayor evolución a lo largo de la década de los años 20 del Siglo XX donde se fueron ocupando principalmente los espacios ubicados en torno a las playas del Sardinero así como la mayor parte de la vertiente sur del término municipal y que da a la bahía en una franja comprendida entre la zona de Puerto Chico y hasta la zona del centro incluyendo la calle Burgos y la zona de la Calle Alta. El periodo de mayor crecimiento urbano se dará en la década de los 50 a partir del desarrollo del planeamiento urbano a partir de 1955 (Plan Comarcal).

Este instrumento supondrá el avance de la expansión urbana a través de la clasificación de los suelos rústicos en urbanizables y urbanos. Utilizando estas categorías de suelo, la actuación del sector de la construcción influyó de manera importante en la sucesiva desaparición de los espacios rurales del municipio de Santander, el incremento del número de viviendas en las localidades de su entorno y, como consecuencia, en la configuración de un modelo urbano metropolitano (Véase Ilustración 18. (Cos Guerra, 2007).

El crecimiento fue canalizado por el Plan Comarcal de 1955, marco de la planificación urbanística de la ciudad hasta 1987, que propuso rodear la ciudad consolidada de áreas residenciales y situar nuevas piezas urbanas en los espacios rurales de Cueto, Monte, San Román y Peñacastillo. Los municipios que conforman su área de influencia son principalmente Santa Cruz de Bezana al oeste y Camargo junto a Astillero al sur. (véase Ilustración 19).

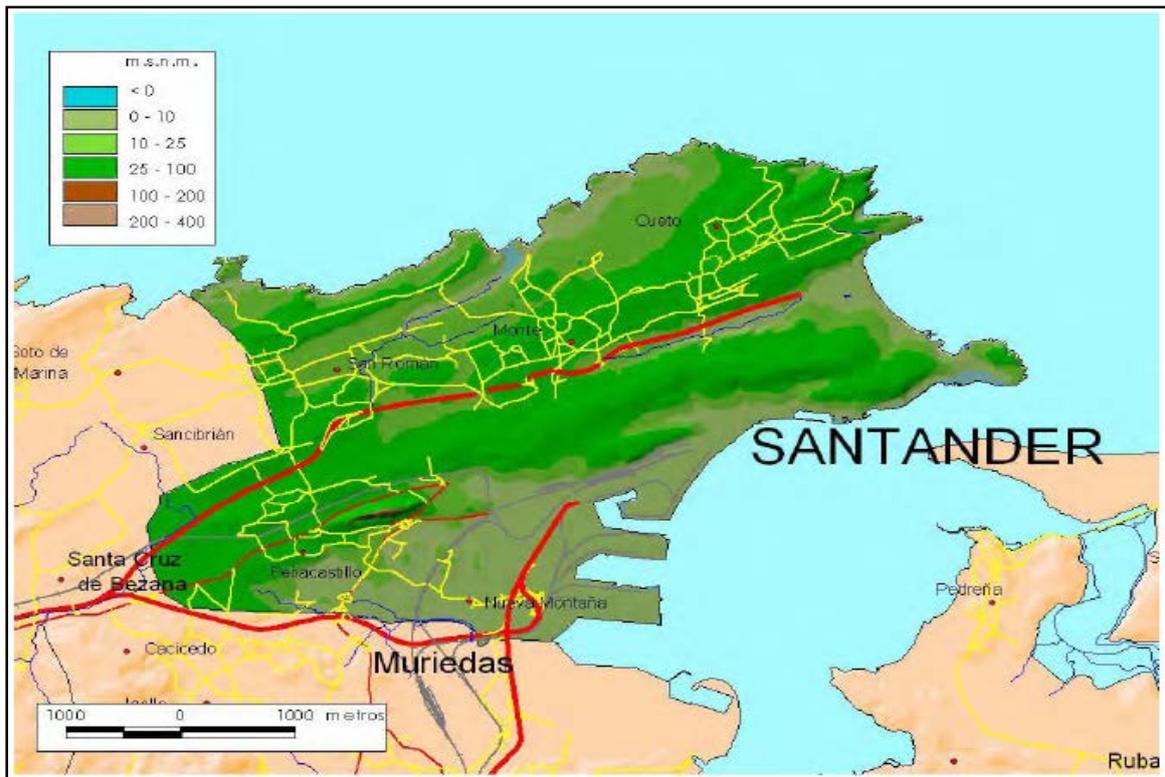


Ilustración 19. Término municipal de Santander y su área de influencia. Fuente: ICANE (Instituto Cántabro de Estadística) a partir de Cartografía Regional del Gobierno de Cantabria y datos del INE.

Santander desarrolla básicamente funciones terciarias. Se ha convertido en un centro administrativo y comercial, en centro financiero y cultural. Como municipio de referencia administrativa y económica ha experimentado importantes transformaciones espaciales derivadas del proceso urbanizador desarrollado, como se puede comprobar en el mapa anterior mente citado sobre la evolución del poblamiento (Véase Ilustración 18). Es por tanto una ciudad de eminente orientación terciaria como lo acredita el hecho de que aproximadamente un 75% de la población ocupada desarrolla su actividad laboral precisamente en este sector.

4.2. Datos utilizados

En esta sección se presenta la formulación teórica de los modelos de demanda de transporte público en bus incorporando variables macroeconómicas para cuantificar el



impacto del ciclo económico como la tasa de paro y el ingreso per cápita. Se presentan así mismo el área de estudio y los datos disponibles para la estimación de los modelos y los resultados obtenidos.

4.2.1. Origen de los datos

Las principales fuentes de información a las que se ha recurrido para la elaboración de una base de datos han sido principalmente dos, la primera de ellas de naturaleza estadística como el INE o el ICANE y otros organismos institucionales como el Ministerio de Industria, Turismo y Comercio o el propio ayuntamiento de la ciudad objeto de estudio y la segunda el Servicio Municipal de Transportes de Santander (TUS). De la primera de ellas se obtuvieron datos de Santander relacionados con el ámbito demográfico y socio-económico como por ejemplo, datos de población en Santander y en el resto de entidades de población integrados dentro del municipio, las cifras de paro en la ciudad, el número de visitantes que tuvo la ciudad además del número de pernoctaciones registradas, el número de actividades según su rama de actividad, los precios medios mensuales del litro de combustible en Cantabria, los valores del IPC en la región, el valor del PIB regional, la Renta Per Cápita y el censo de vehículos.

FUENTES CONSULTADAS Y TIPO DE DATOS	
FUENTE	TIPO DE DATOS
INE. Instituto Nacional de Estadística	Datos Demográficos, datos económicos (IPC y PIB).
ICANE. Instituto Cántabro de Estadística	Datos poblamiento, Número de visitantes y Pernoctaciones en alojamientos hosteleros.
Ayto de Santander	Censo de vehículos
T.U.S (Transportes Urbanos de Santander).	Datos del nivel de servicio (viajeros, líneas, kilómetros, nº de buses, recaudación, tarifas...)
Ministerio de Industria Turismo y Comercio	Precios promedios mensuales de los combustibles (Gasoil y Gasolina)
Servicio Cántabro de Empleo	Nº de Desempleados (totales, por edades, sexo y ramas de actividad)

Tabla 4. Fuentes principales consultadas para la obtención de la información.

De la segunda, el TUS, se recopilaban toda una serie de datos relacionados con el propio servicio de autobuses como son, la demanda de viajeros (número de viajeros), las tarifas, la recaudación, el número de vehículos/autobuses, los kilómetros recorridos, las plazas ofertadas y los litros de combustible consumidos.

4.2.2. EL TUS (Transportes Urbanos de Santander).

El sistema de transporte público de la ciudad está gestionado por los Transportes Urbanos de Santander (TUS), una empresa pública que cuenta con un total de 17 líneas diurnas y 3 líneas nocturnas (dell'Olio, Ibeas et al. 2011). Cinco de las líneas diurnas y la totalidad de las líneas nocturnas están concesionadas a una empresa privada y no se tendrán en cuenta en el presente estudio en el que se trabajará únicamente con las líneas municipales que representan en torno al 88% del tráfico total de viajeros (véase Ilustración 19).



Ilustración 20. Líneas de Autobús en el área de estudio.

El conjunto de líneas actuales de gestión municipal y que se han tenido en cuenta en la elaboración de esta tesis son:



- Línea 1: Parque Tecnológico - Adarzo - Valdenoja - A. Odriozola
- Línea 2: Corbán - Consuelo Bergés
- Línea 3: Ojáiz
- Línea 4: Bº Pesquero - Brisas
- Línea 5C-1: P. Italia - Miranda - P. Los Remedios - G. Davila
- Línea 5C-2: P. Italia - Miranda - Davila - Valdecilla
- Línea 6C-1: Complejo deportivo
- Línea 6C-2: Complejo - Tunel Puertochico - Av. Cantabria - Complejo
- Línea 7C-1: L. Quintanilla Isasi - Avda. Los Castros
- Línea 7C-2: J. Bustamante - Valdecilla
- Línea 11: Avda. Valdecilla - C/ Alta
- Línea 12: Carrefour - Canalejas
- Línea 13: Lluja - Cueto
- Línea 14: P. Estaciones - Residencia

4.2.3. Evolución de las líneas del TUS (2001-2012)

A lo largo de los años que abarcan el estudio (2001-2012), las líneas de transporte han ido experimentando diferentes cambios en cuanto a su estructura fundamentalmente basados en dos procesos, o bien la ampliación de recorridos en algunos casos o en la fusión de líneas por otra. Unos cambios que han supuesto cambios en las variables relacionadas con el nivel de servicio ya que progresivamente se han ido incrementando las cifras de kilómetros recorridos, el número de buses y por tanto el número de plazas ofertadas.

A partir del año 2002 el Servicio de transportes urbanos de Santander (TUS) inició un estudio de cara a la mejora del servicio en las zonas de Valdenoja y San Román resultando un desdoblamiento de la línea 1 en dos, así como la modificación de los recorridos de las líneas 6 (C1 y C2) y la antigua 9 que se fusionó con la antigua 8 para conformar la actual línea 13, una fusión que no se hizo efectiva hasta el mes de Marzo del año 2003.



El desdoblamiento de la línea 1 en las citadas 1 y 2 las convirtió en las líneas que transportaron un mayor número de viajeros. Tras el desdoblamiento la línea 1 cubría el recorrido Glorieta de Adarzo, centro de Santander, Avda Reina Victoria, El Sardinero y Valdenoja en el sentido de ida y Valdenoja, Agustinos, El Sardinero, Avda Reina Victoria, Centro de Santander y Glorieta de Adarzo en el sentido de vuelta transportando una cantidad que superó levemente los 2,2 millones de viajeros. La línea 2 con recorrido Corbán, Alisal, centro de Santander, Avda Reina Victoria, El Sardinero, Valdenoja en el sentido de ida y Valdenoja, Agustinos, El Sardinero, Avda Reina Victoria, Centro de Santander, Cazoña, Cruce con Rucandial y Corbán en el sentido de vuelta acumuló un valor cercano a los 2,5 millones de viajeros. En la actualidad la estructura de estas líneas es esencialmente la misma y tan sólo ha experimentado alargamientos de su recorrido en la zona donde se ubica el PCTCAN (Parque Tecnológico), alargamiento que se produjo a partir del mes de Marzo del año 2010.

En el año 2004 se establecieron nuevos precios con respecto al año anterior. La tarifa ordinaria subió de 0,90 céntimos de euro a 1 euro; el viaje con bonobús pasó de 0,45 a 0,50 céntimos de euro y los viajes para pensionistas pasaron a ser gratuitos. A su vez en 2004 el ahorro para los usuarios fue mayor debido a la medida adoptada en referencia a los transbordos entre líneas que pasaron a ser gratuitos a la hora de utilizar la tarjeta bonobús. En lo que a tarifas se refiere en el año 2005 permanecieron invariables y tan sólo el precio de los bonobuses se encareció unos 20 céntimos de euro y el carné joven dos euros. Con respecto a las distintas variaciones de las tarifas, éstas fueron recogidas en la base de datos como variable independiente y relacionadas con el nivel de servicio.

Cabe señalar que en el año 2007 entran en funcionamiento dos nuevas líneas en funcionamiento y que no se tuvieron en cuenta a la hora de realizar el actual estudio debido a la inexistencia de datos para el periodo 2001-2007. Esas nuevas líneas que entraron en vigor fueron la línea 16 cuyo recorrido discurre por la ladera sur del eje "General Dávila" y la ampliación y reordenación de la cobertura a toda la ciudad del servicio nocturno. La línea 16 con servicio entre Plaza de los Remedios y General Dávila



por la cuesta de la Atalaya, Via Cornelia y Calle monte que se realiza con microbuses. La línea 16 comenzó a funcionar unos días antes del verano, el 28 de Junio de 2007.

El año 2007 es un año importante ya que el ayuntamiento de Santander llega a un acuerdo con la empresa ALSA de manera que las tres líneas que esta empresa de transporte gestionaba (17, 18 y 19 que se extienden a San Román, Monte, La Albericia y la zona del Primero de Mayo) pasarían a ser gestionadas por el TUS lo cual suponía que los vecinos de las áreas periféricas conseguían mayores ventajas entre las cuales destacar que los trasbordos de esas líneas a las propias del TUS pasarían a ser gratuitos entre distintas líneas y el descenso del precio del billete ordinario que pasaba de 1,10 a 1 Euro. La concesión de esas líneas 17, 18 y 19 seguían siendo explotadas por ALSA como empresa privada. La unificación de la gestión de estas líneas con el resto de las líneas de concesión conllevó la unificación de tarifas de los autobuses en todo el término municipal, lo que supuso que los usuarios de las líneas 17, 18 y 19 pudieran recibir las mismas bonificaciones que los viajeros del resto de líneas municipales.

La línea 17 comunicaba la Plaza de las Estaciones con Corbán y Ciriego realizando dos recorridos diferentes, uno por Barrio la Torre y el otro por Avda del Deporte. La línea que pasaba a unir Puertochico con Monte pasó a denominarse línea 18 y la que conectaba la Plaza de las Estaciones con el Primero de Mayo, pasando por la zona de Mercasantander y el Zoco pasó a denominarse línea 19.

Otra línea, esta sí incluida en el desarrollo de esta tesis, fue la línea 11. En el mes de Marzo de 2007 se modificó el recorrido de esta línea, cuya cabecera se situaba en Valdecilla. En su recorrido se añadieron nuevas paradas y se mejoró el servicio en ambas direcciones de la calle Alta con nuevas paradas y la circulación de los autobuses de la línea por la calle Ruamayor. El objetivo de esta modificación fue el de facilitar un mayor acceso de los usuarios al hospital universitario Marqués de Valdecilla, reducir los tiempos de viaje con origen y destino en el centro de la ciudad desde la Calle Alta incorporando el doble sentido de circulación, mejorar el servicio prestado a los vecinos



del Grupo Velarde y mantener la cobertura de la línea en el centro de la ciudad (Jesús de Monasterio).

Para lograr una mejora del TUS (Transportes Urbanos de Santander), una mejora en la optimización del servicio y conseguir mejorar la calidad para los viajeros, el ayuntamiento contrató en 2006 con la universidad de Cantabria toda una serie de estudios para la reorganización de toda la red de líneas, adecuándola a las necesidades futuras de Santander. Este estudio, realizado por el Grupo de investigación de sistemas de Transporte (GIST) dirigido por el Profesor Ángel Ibeas, concluyó que era necesario que hasta 8 líneas de autobús de las existentes modificaran sus recorridos para poder dar servicio a los nuevos barrios de la periferia de la ciudad. Dichas modificaciones se irían realizando de forma progresiva en los tres años siguientes a partir de la realización del estudio, años 2007, 2008 y 2009.

Otro hito importante a considerar tuvo lugar a partir del 1 de Enero de 2008, fecha en la que se adoptó que todos los miembros de familias numerosas de la ciudad viajarían gratis en los autobuses municipales. Esta medida afectaría a casi 8.000 habitantes de la ciudad. Hasta 2007, las familias numerosas disfrutaban de una bonificación del 50% en el transporte municipal y realizaban una media de aproximadamente 350.000 viajes al año. Se trató de un importante avance de carácter social del servicio de transportes del Ayto de Santander. Además de las familias numerosas, también los pensionistas con ingresos limitados y las personas con discapacidad superior al 66% podían hacer uso de la red municipal de transportes de forma gratuita. También destacaron las bonificaciones existentes para los jóvenes mediante el carné trimestral y los transbordos gratuitos.

El año 2008 fue también un año en el que se llevaron a cabo importantes reordenaciones de líneas. Hasta el mes de Junio existieron las líneas 8 con recorrido de Correos a LLuja, la 9 con itinerario de Cueto a Valdecilla, la línea 10 con recorrido comprendido entre Jesús de monasterio y Canalejas y la antigua línea 12 que se extendía desde Alfonso XIII a Nueva Montaña. A partir del mes de Julio se fusionaron las líneas 8 y 9 dando lugar a



la actual línea 13 (LLuja – Cueto) y la fusión de las líneas 10 y 12 que dieron paso a la actual línea 12 (Carrefour – Canalejas).

A partir del 1 de julio de 2010 entrará en funcionamiento una nueva línea cuya concesión pertenece a la empresa ALSA pero que será gestionada también por el TUS (Transportes Urbanos de Santander), esa línea se trata de la línea 20 cuyo recorrido estará comprendido entre la zona de Monte y la zona de las Estaciones.

En el año 2011 algunas líneas sufrieron también pequeñas modificaciones en sus recorridos entre las que cabe señalar la línea 2 cuyo recorrido se amplía hasta dar servicio a la zona del Barrio La Sierra. Igualmente se modificarán recorridos de las líneas 6C1, 6C2, 7C1, 7C2 y 13 que pasan a hacer una pequeña prolongación en su recorrido habitual para entrar en la zona en la que se ubica la parada de Ricardo León y en el caso de la 7C1 su recorrido se ampliará hasta su nueva cabecera que pasará a ser Luis Quintanilla Isasi cuando hasta entonces había sido en Manuel Llano. En la línea 13 la modificación afectará a su recorrido con destino a la zona del Cementerio de LLuja que se mantendrá pero modificando su trayecto dando servicio a la zona del Parque Tecnológico.

La línea 14 cambia también su recorrido dejando de pasar por la Calle Alta para pasar a recorrer la calle Jerónimo Saiz de La Maza, bajo el Mercado de Mexico. La línea 23 comenzará a funcionar también a mediados de ese mismo año cubriendo el itinerario Estaciones – Carrefour aunque de forma temporal hasta que se completaran las obras del vial de Camarreal. Una vez finalizadas las obras de dicho vial su recorrido cambió hasta el actual Estaciones – Camarreal.

4.3. Construcción y tratamiento de los datos

Una vez realizada la labor de recopilación de datos el siguiente paso se centró en la construcción de la base de datos. El objetivo es que los datos puedan representarse de forma que constituyan una serie temporal de escala mensual con origen en el mes de Enero de 2001 como mes inicial y Diciembre de 2012 como mes final de la serie. Dado



que el objetivo de este estudio es el desarrollo de un modelo de previsión de la demanda en transporte público y más concretamente en el caso del servicio municipal de transportes urbanos de Santander se procedió a introducir el número total de viajeros registrados para cada uno de los meses de la serie planteada. El total de viajeros es el resultante de la suma del número de viajeros mensuales habidos en cada una de las líneas consideradas dentro del estudio. El número de viajeros se convertirá en la variable dependiente o variable a estimar y a partir de ella en esa misma base de datos se introducirán los valores del resto de variables explicativas o independientes con el objetivo de estudiar su comportamiento y si este influye en mayor o menor medida en la demanda de viajeros. De la misma manera se procedió a construir una base de datos en la que se recogiera de forma individual, a modo de series temporales, el número de viajeros mensuales por línea con el posterior objetivo de poder prever también la demanda por líneas.

4.3.1. Homogeneidad de los datos

A la hora de construir la base de datos el principal problema que se planteaba era el de poder obtener una base de datos en la cual los datos de las variables a estudiar fueran homogéneos ya que para el caso de la demanda de viajeros se tenían datos mensuales y para algunas otras variables de ámbito demográfico o económico no eran mensuales sino anuales o trimestrales, casos del censo para la obtención de datos demográficos o de datos económicos como los valores del PIB regional o el IPC cuyas cifras son trimestrales. De cara a solventar este escollo instrumental se crearon variables fruto de la combinación de las ya recogidas. Se optó por crear una variable que fuera Viajes/per cápita resultado de dividir el número de viajes entre la población de Santander y de esa forma obtener una nueva variable que integrara el dato de población. Tanto la demanda como las variables relacionadas con el nivel de servicio del transporte en bus se han considerado en forma per cápita, lo que evita introducir la población como variable independiente en el modelo.



En cuanto a los datos de la demanda de viajeros mensual para cada línea, hubo que homogeneizar los datos también puesto que dentro de la franja temporal considerada, de Enero de 2001 a Diciembre de 2012, algunas de las líneas tenían una estructura diferente. El número de viajeros de la línea 12 desde Enero de 2001 a Junio de 2008 se obtuvo a través de la suma de las antiguas líneas 10 y 12 puesto que a partir de Julio de 2008 la fusión del recorrido de ambas dio lugar a la actual línea 12. En el caso de la línea 13 se llevó a cabo el mismo procedimiento. El número de viajeros mensuales en la línea 13 desde Enero de 2001 a Junio de 2008 se calculó a partir de la suma de las antiguas líneas 8 y 9, actualmente inexistentes. La fusión del recorrido de ambas, líneas 8 y 9, dio como resultado a la 13 actual.

4.3.2. Deflatación de variables monetarias.

El siguiente paso a seguir dentro del análisis de los datos fue el de llevar a cabo la deflatación de las variables monetarias o aquellas expresadas en euros. Al trabajar con una serie estadística o serie temporal de datos que contienen alguna magnitud económica (renta per cápita, recaudación o ingresos, precios de tarifas, entre otros), lo conveniente es que la valoración monetaria de los datos se realice a precios corrientes de cada período, o también denominados precios nominales, estos son los precios establecidos; los que marcan la etiqueta. El problema radica en que si pretendemos comparar precios en el tiempo, en la medida en que los precios experimentan fluctuaciones de unos periodos a otros, la serie así representada no permite llevar a cabo comparaciones.

Para solventar esta disyuntiva se optó por expresar las variables monetarias en términos de precios constantes de un determinado periodo (año base). Los precios reales son los precios al que se le han descontado ya el efecto de la inflación. De esta forma podemos comparar los precios reales de un determinado producto en un año inicial y de uno similar en otro año final, descontando el “encarecimiento de la vida en esos años” (véase Gráfico 6).

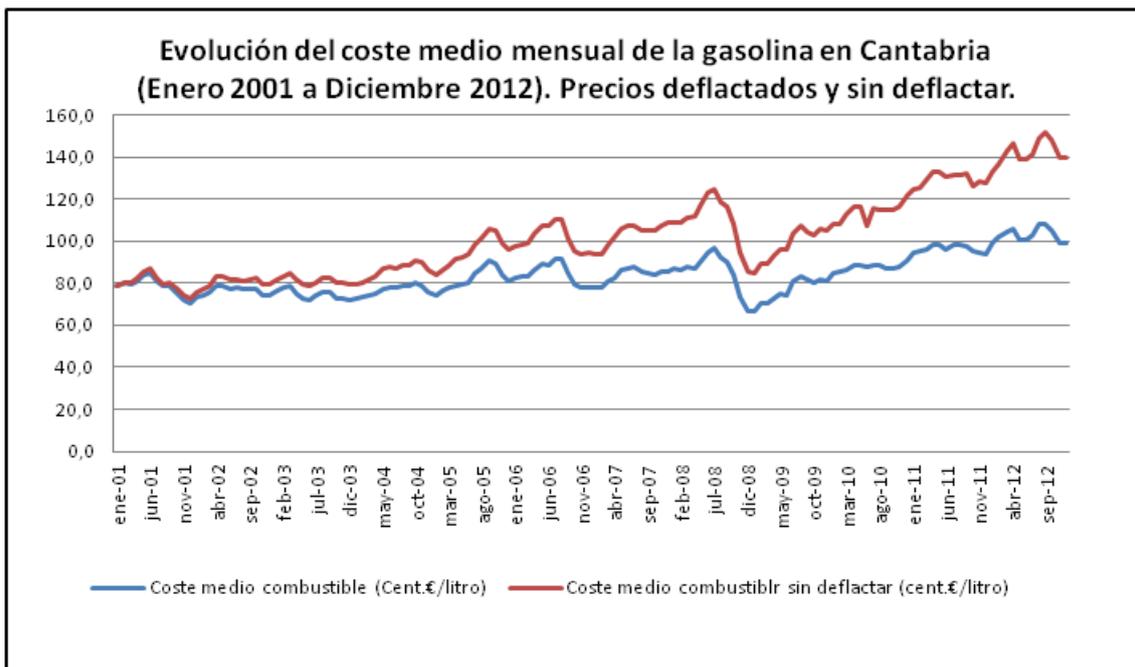


Gráfico 6. Elaboración propia a partir de los datos obtenidos del Ministerio de Industria, Turismo y Comercio. Evolución de los costes de los carburantes en Cantabria (Gasolina) medidos en promedios mensuales.

A la transformación de la serie original a la serie valorada en precios constantes se llama deflactación, y el índice a través del cual se puede pasar de una serie a la otra se llama deflactor. Las variables deflactadas fueron, la recaudación mensual, las tarifas, el precio medio mensual del litro de combustible en Cantabria, y la renta per cápita obtenida mediante la división del valor regional del PIB entre el número de habitantes de la región. Para llevar a cabo la deflactación se tuvo en cuenta como año base el año 2001. Otro problema añadido fue que durante los primeros doce meses de la serie, de Enero de 2001 a Diciembre de 2001, la moneda que aún circulaba era la peseta por lo que las variables monetarias para ese periodo se transformaron a euros.



4.3.3. Tendencia y Estacionalidad

Una vez homogeneizados los datos y deflactadas las variables monetarias se procedió a la representación gráfica de la variable a predecir, la del número de viajeros con el objeto de analizarla y tratar de buscar patrones en su comportamiento. La representación gráfica supone el primer paso de cara a iniciar el análisis de la serie temporal que se ha construido.

El objetivo del análisis de la serie temporal, en la que se representa el número de viajeros mensuales en el servicio municipal de transportes de Santander a lo largo de una serie comprendida entre Enero de 2001 y Diciembre de 2012, es doble. Por un lado se busca explicar las variaciones observadas en la serie en el pasado, tratando de determinar si responden a un determinado patrón de comportamiento. Y por otro, si se consigue definir ese patrón o modelo, se intentará pronosticar el comportamiento futuro de la misma.

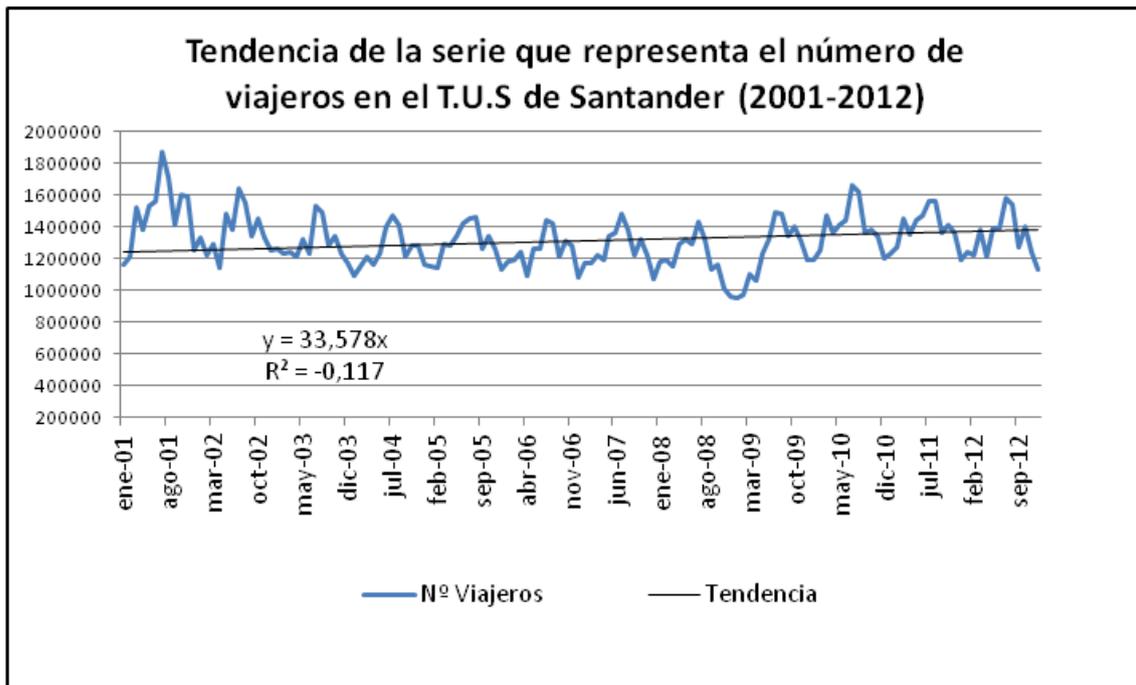


Gráfico 7. Tendencia de la serie. Fuente: Elaboración propia a partir de los datos aportados por el TUS de Santander.



El análisis de esta serie supone la valoración de cuatro aspectos importantes como son, la tendencia, la estacionariedad, la estacionalidad y las variaciones cíclicas si existieran. Analizada la tendencia se observa que existe un comportamiento suave de la serie a largo plazo algo que nos lo permite el hecho de contar con una serie suficientemente larga conformada por 144 meses u observaciones (12 años).

Atendiendo a la tendencia en el gráfico (véase Gráfico 7) puede observarse como en los primeros cuatro años de la serie (2001-2004) se da un decrecimiento progresivo de los valores máximos pasando de un promedio de prácticamente 1.8 millones de viajeros en los meses de Julio y Agosto de 2001 a 1.44 millones de viajeros en el mismo periodo estival del año 2005 lo que representa un descenso de casi el 20%. A partir de 2004 los valores máximos correspondientes a los periodos de verano tienden a estabilizarse y las oscilaciones variarán entre los 1.44 millones de viajeros de promedio en los meses de 2004 a los 1.43 en 2007. En el periodo estival de 2008 se registra el promedio de viajes más bajo de la serie llegando hasta los 1.37 millones de viajeros. A partir de 2008 vuelve a producirse un tercer periodo comprendido entre 2009 y 2012 en el que los valores máximos vuelven a recuperarse de forma progresiva hasta alcanzar valores superiores al periodo (2001-2004). En el periodo 2009-2012 el promedio de viajes en los meses de Julio-Agosto pasarán de 1.48 millones de 2009 hasta los 1.56 de 2012 siendo el verano de 2010 el que mayor número de viajes acumuló con 1.64 millones. Las diferencias entre el valor máximo (1.64 millones en 2010) y el mínimo (1.37 millones en 2008) no es demasiado importante y ambas se encuentran cercanas al valor promedio de los doce años de la serie situado en 1.52 millones de viajeros durante los meses de Julio y Agosto.

En virtud de la descripción de los datos anteriores y de su representación gráfica podemos afirmar que se trata de una serie estacionaria ya que se comporta de una forma estable y tanto la media como la variabilidad son constantes a lo largo del tiempo. Esto se refleja gráficamente en que los valores de la serie, de la demanda de viajeros en Santander, tienden a oscilar alrededor de una media constante y la variabilidad con respecto a esa media también permanece constante en el tiempo. Se trata por tanto de



una serie básicamente estable a lo largo del tiempo, sin que se aprecien aumentos o disminuciones sistemáticos de sus valores. El hecho de ser una serie estacionaria significa que posee un grado importante de validez de cara a poder hacer posteriores previsiones. Para el propósito de este estudio resulta conveniente el hecho de poder contar con una serie estacionaria ya que las predicciones resultan más sencillas de obtener, la media se puede estimar teniendo en cuenta la totalidad de los datos y utilizar con el objetivo de obtener nuevas observaciones e igualmente se pueden obtener intervalos de confianza para las predicciones. Sin embargo como se verá posteriormente la serie está caracterizada por presentar rasgos evidentes de estacionalidad.

En el presente estudio para poder captar la tendencia de la serie se introdujo una variable explicativa con referencia numerada acorde al orden de las observaciones en el tiempo

4.3.4. La Estacionalidad

Además de la estacionariedad, se analizaron también las variaciones estacionales o la estacionalidad, que hace alusión a los movimientos de la serie que se repiten de forma periódica. La razón de estas variaciones se basan en causas de tipo climatológico o de ordenación del tiempo (los días de la semana condicionan el comportamiento de ciertas series). En el caso de los datos de Santander la serie es claramente estacional (véase Gráfico 8) pues se muestra cómo se dan picos de demanda en los meses correspondientes al periodo veraniego y las demandas más bajas se producen a lo largo de los meses de otoño-invierno apareciendo algún pequeño repunte en algunos meses de Abril o Mayo, dependiendo de cuando tuvo lugar la Semana Santa, el segundo periodo vacacional más largo del año tras el verano.

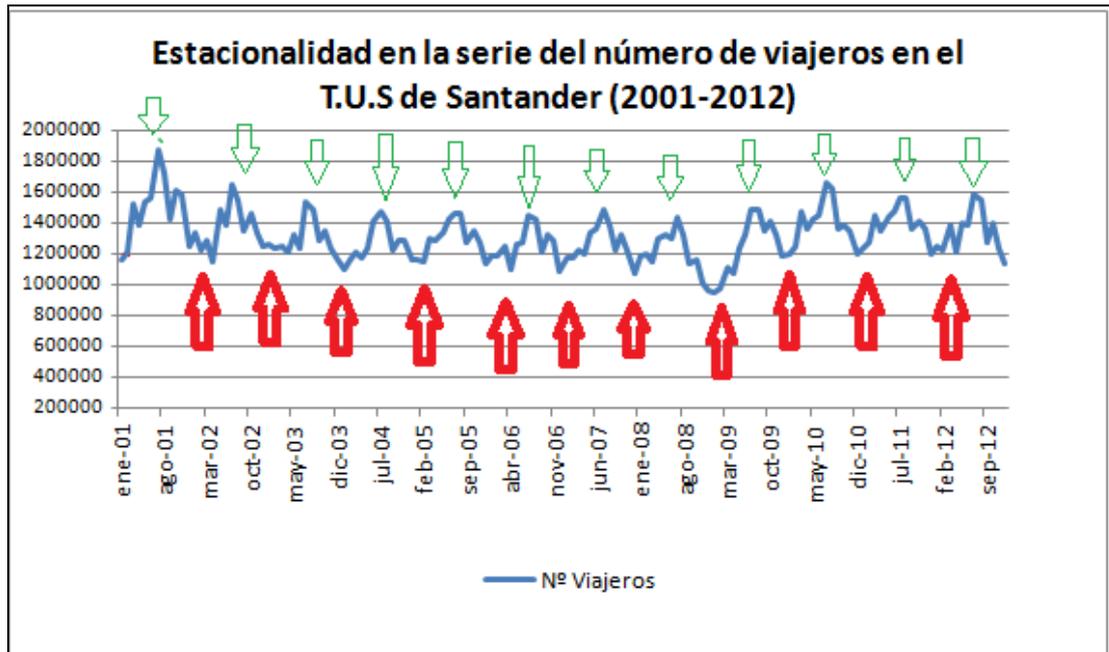


Gráfico 8. Elaboración propia a partir de los datos obtenidos por el T.U.S (Transportes Urbanos de Santander). Análisis de la estacionalidad.

El objetivo de determinar las componentes estacionales puede ser: en primer lugar para conocerlas y posteriormente eliminando posibles influencias estacionales. A ese proceso de eliminación de efectos estacionales en una serie temporal, como es la de los datos utilizados en este estudio, se le denomina “Desestacionalización”.

4.3.5. Desestacionalización de los datos y análisis de la correlación.

La Desestacionalización de la serie temporal permite analizar una serie cronológica de datos sin tener en cuenta la componente estacional. La Desestacionalización o corrección estacional evita las distorsiones en los valores medios de los datos. En este caso aplicado a Santander el método de desestacionalización utilizado fue el de la inclusión de las llamadas variables “dummy” que ayudan a realizar un ajuste estacional. Estas variables “dummy” son unas variables artificiales a las cuales se les otorga unos valores discretos generalmente de 0 y 1. En el caso de los datos de Santander se trabaja con datos mensuales por lo que se introduce junto al resto de variables independientes un total de 12 variables dummy, una por cada mes del año q1 (Enero), q2 (Febrero), q3 (Marzo)..... q12 (Diciembre). El problema de añadir estas doce variables es que tendrían



como resultado una combinación lineal exacta con la constante produciéndose así un caso de multicolinealidad y que el determinante de la matriz fuera igual a cero, lo cual presentaría problemas de cara a estimar los coeficientes del modelo de regresión. Para evitarlo en vez de incluir las doce variables lo que se hace es introducir un total de once además de la constante. Desestimando una de las variables dummy en la matriz se evitaría el efecto de colinealidad perfecta con la columna de la constante asociada con el intercepto.

Continuando con el proceso de análisis y tratamiento de los datos otro apartado importante es el relacionado con el análisis de la correlación que pudiera haber entre distintas variables ya que puede existir una correlación alta tanto positiva como negativa entre las variables independientes consideradas, algo que habitualmente puede darse en variables relacionadas con la población o con la renta per cápita. Cuando distintas variables están altamente correlacionadas entre sí es cuando se produce un efecto denominado multicolinealidad. Para comprobar si existe colinealidad en los datos se procede a obtener los coeficientes de correlación para las variables explicativas contenidas en la base de datos. Una vez obtenidos se eliminaron aquellas variables independientes que presentaran un alto grado de correlación entre sí (valores por encima de 0,5 y por debajo de -0,5) de forma que la bondad del ajuste se viera lo menos afectada posible. Se desestimaron las variables altamente correlacionadas entre sí para evitar el efecto de multicolinealidad.

4.4. Análisis de datos.

4.4.1. Evolución de la demanda de viajeros y el desempleo

Una de las principales variables tenidas en cuenta en este estudio y que está íntimamente ligada a periodos tanto recesivos como expansivos de la economía, es el desempleo o el número de parados. En el Gráfico 9 se muestra el comportamiento tanto



del paro como de la demanda de transporte en Santander que a continuación se procede a describir.

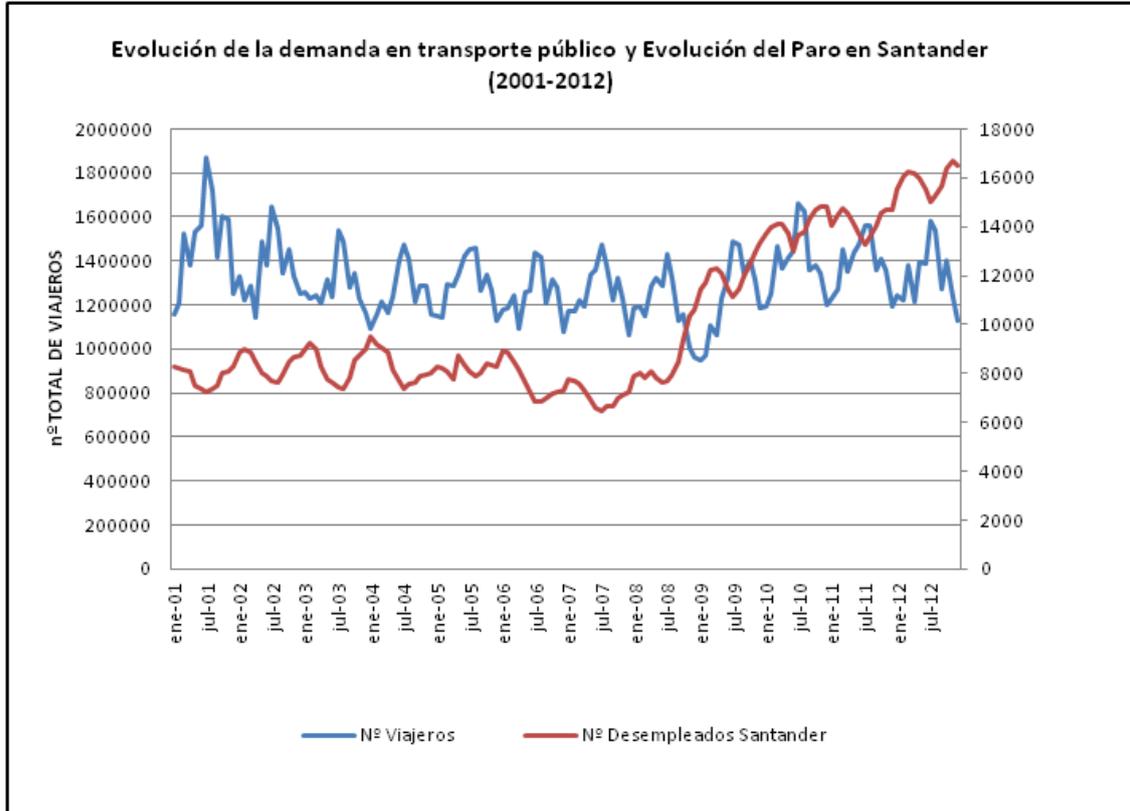


Gráfico 9. Elaboración propia a partir de los datos obtenidos del TUS de Santander y del INE.

En el caso de la demanda de transporte se puede comprobar cómo en el periodo comprendido entre 2001 y 2003 se produce un progresivo descenso en el número de viajes en los periodos de verano, periodos en los que se registran siempre el mayor volumen de viajeros debido al número de visitantes que recibe la ciudad de Santander debido a su orientación turística. Si en los meses de verano de 2001 la demanda superó levemente los 1.8 millones de viajeros, en 2003 la demanda estuvo por debajo de los 1.6 millones. A partir del año 2004 el número de viajes mensuales se estabiliza y se mantiene constante por encima del 1.4 millones de viajeros, una tendencia que se mantendrá hasta el año 2008. En el año 2009 habrá un pequeño repunte, con una cifra cercana a los 1.5 millones, que seguirá al alza en el periodo estival de 2010, en el que llegan a superarse los 1.6 millones de viajeros, un hito al que no se llegaba al menos desde el



verano de 2002. A partir de 2010 vuelve a registrarse un leve descenso con cifras muy cercanas a los 1.6 millones de viajeros en los años 2011 y 2012.

En cuanto a los periodos de invierno el número de viajeros tiende a descender de una manera sostenida en el periodo comprendido entre 2001 y 2009, siendo este último en el que se registró un menor número de viajeros. La media mensual de viajeros para el periodo que va de Diciembre de 2008 a Febrero de 2009 fue en donde más se sufrió la caída de la demanda con una media cercana a los 0.96 millones de viajeros, único periodo de toda la serie en el que la demanda está por debajo del millón de viajes. Ese periodo en el que se registró una menor demanda de viajeros coincide aproximadamente con el inicio de la crisis económica como así lo atestiguan los datos del paro que arrojan un gran repunte precisamente a partir de finales del año 2008 en adelante. Desde el año 2001 hasta finales de 2008 el paro en Santander se había comportado de una manera estacional debido al periodo de verano pero tanto los datos de desempleo para verano como para invierno en esos ocho años se comportaban de una forma estable y regular.

El promedio de parados los meses de verano en Santander para el periodo 2001-2008 fue de 7.381 y a partir de 2009 ese promedio se incrementó de forma notable pasándose a un promedio de 11.296 parados en el verano de 2009, 13.564 en 2010-2011 y 15.127 en el periodo estival de 2012. Si tenemos en cuenta, como base, el promedio de desempleados habidos entre 2001 y 2008 (7.381), para 2009 el paro se incrementó en casi cuatro mil personas, un incremento de un 53% en tan sólo un año. Para el periodo 2010-2011 el incremento continuó de una forma menos incisiva pero igualmente al alza registrándose un promedio de casi 6.200 parados más (un 84 % más con respecto al promedio de los años 2001-2008 y un 20% con respecto a 2009). Las cifras de paro para el año final de la serie, 2012, siguieron aumentando pasando a un promedio en los meses de Julio y Agosto de algo más de Quince mil personas (un 104% más si lo comparamos con los datos acumulados entre 2001 y 2008). Estos datos reflejan que los datos de paro en el periodo vacacional de 2012 llegaron a ser el doble de lo que habían sido hasta el año 2008. Esto supone que a partir del 2008 se produjo el impacto brusco



de la crisis económica en Santander materializándose en forma de aumento generalizado de la tasa de desempleo, una crisis que a lo largo de los años posteriores siguió contribuyendo a la destrucción de empleo de forma progresiva.

Analizando los datos de invierno los promedios se comportan de una forma más estable que los de verano oscilando entre 1.1 millones y 1.2 millones de promedio entre 2001 y 2008. Es en 2009 cuando ese promedio baja por primera vez del millón de viajeros y a partir de ahí hasta 2012 los promedios de invierno volverán a establecerse en los 1.2 millones de viajeros mensuales de promedio.

Si se comparan los datos de paro con los datos de demanda, se aprecia cómo en el periodo 2001-2008 a medida que los datos de paro son menores, la demanda de viajeros va menguando de forma progresiva pero sin caídas importantes. A partir de 2008 se produce un incremento espectacular del paro y eso se ve reflejado en la demanda que llega a experimentar incluso subidas, unas subidas que pueden estar relacionadas por un lado con el número de visitantes en la ciudad los meses de verano (un promedio que oscila entre los 40.000-45.000 visitantes de promedio en los meses de verano de 2008 a 2012) y por otro con el empleo del uso de transporte público frente al privado por su menor coste. Por otro lado en los periodos de invierno la demanda promedio mensual permaneció estable frente al incremento del paro como es lógico las menores demandas se registran en invierno por efecto de la estacionalidad y las cifras más altas de paro también se registran en invierno por el mismo motivo por lo que en la gráfica los picos de paro coinciden con los valles de demanda y viceversa.

De este gráfico se desprende a partir de qué momento repercutió en Santander la influencia de la crisis económica de carácter global. El origen de la crisis económica mundial actual tuvo lugar aproximadamente a comienzos del año 2007 en Estados Unidos con las hipotecas "subprime", de alto riesgo, como principal factor. Su origen comienza con el inicio de una crisis financiera, la cual comenzó a propagarse por el resto de la economía mundial. La crisis supuso una desconfianza crediticia que comenzó por extenderse a los mercados financieros norteamericanos que pusieron su punto de mira



en las “hipotecas basura” europeas y que acabó degenerando en una crisis bursátil que se considera la principal detonante de la crisis económica y financiera de 2008. A raíz de esa crisis hipotecaria y financiera se produjeron numerosas quiebras, nacionalizaciones bancarias, constantes intervenciones de los Bancos Centrales, enormes descensos en las cotizaciones bursátiles y deterioro de la economía real, entrando varias economías fuertemente industrializadas en recesión (García Mora, Hervás Ortega et al. 2007).

En España a partir de 2008 comenzaron a experimentarse deterioros en los principales indicadores macroeconómicos y sus consecuencias llegan hasta la actualidad, además en España el impacto de la crisis internacional se vio amplificada por el estallido de la conocida como “burbuja inmobiliaria” que provocó un fuerte aumento del desempleo y una contracción generalizada del consumo que se vio agravada con el aumento del IVA en el año 2009 (Ocón Galilea 2013). Consecuencia del impacto de la crisis económica mundial y del estallido de la burbuja inmobiliaria los datos de paro en Santander comenzaron a incrementarse enormemente a partir de 2008 precisamente como puede verse en el Gráfico 9. Un ejemplo de lo que supuso el estallido de la burbuja inmobiliaria en Santander queda reflejada en que el promedio del paro en el sector de la construcción pasó de 1.059 desempleados en el periodo (2006-2007) a 1.590 desempleados de promedio mensual en 2008 (primer año de la crisis) y a partir de ahí los promedios siguen subiendo de forma muy significativa hasta los 2.529 de promedio para los años 2009 y 2010. O lo que es lo mismo, el paro en la construcción aumentó en un 50% en tan sólo un año, de 2008 a 2009 y ese aumento supera ampliamente el 100% si comparamos los datos de 2008 con los de 2012.

Otra de las consecuencias de la crisis económica en España y por tanto en Santander fue la contracción del consumo, lo cual derivó en el aumento del paro en el sector comercial y de servicios. Los años 2006 y 2007 tuvieron un promedio mensual de 5.087 y 4.702 parados respectivamente en el sector comercial de Santander y ya en el año 2008 esas cifras se elevaron hasta los 5.406. Los años 2009 y 2010 reflejaron de la misma forma el fuerte incremento del paro que llegó a ser de 7.697 parados de media mensual para 2009 y 8.964 en 2010, lo que representaron aumentos porcentuales comprendidos entre un 30 y un 40% en tan sólo dos años. Como puede verse en el Gráfico 10, el



comportamiento del paro en el sector servicios es más estacional y experimenta descensos en periodos de verano.

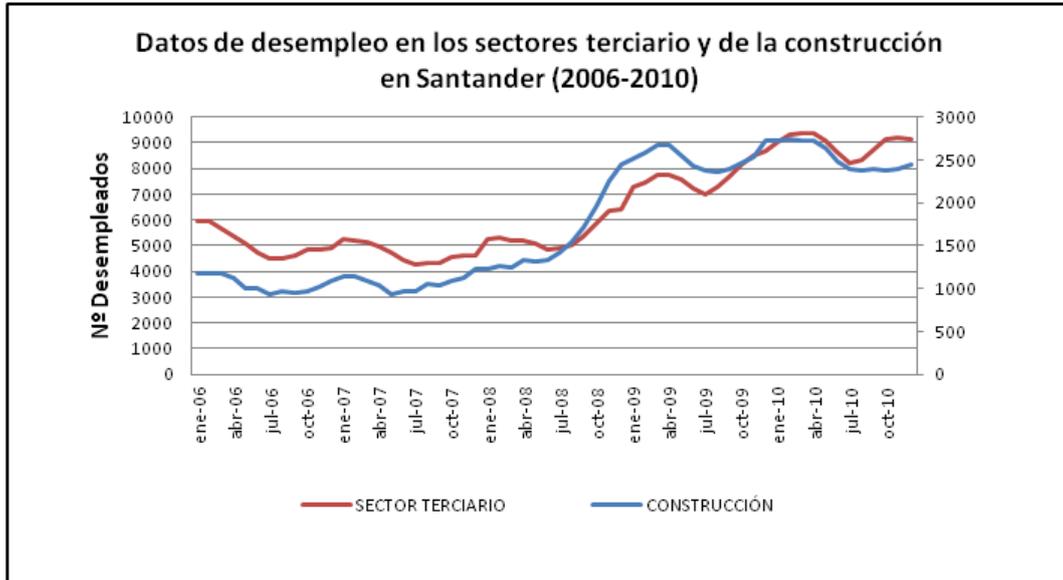


Gráfico 10. Elaboración propia a partir de los datos obtenidos del TUS de Santander y del INE.

4.4.2. Evolución de la demanda de viajeros y nivel de ingresos

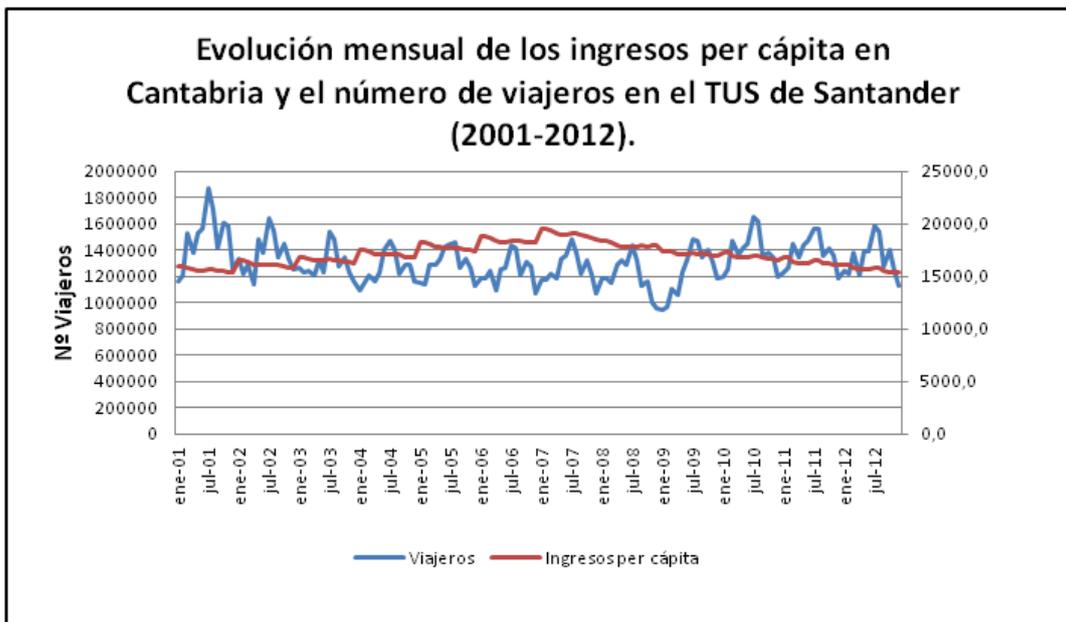


Gráfico 11. Elaboración propia a partir de los datos obtenidos del TUS de Santander y del INE.

Otra de las variables con indudable orientación económica es la del PIB (Producto Interior Bruto). El PIB es un indicador económico que refleja la producción total de



bienes y servicios generalmente asociada a un país durante un determinado periodo de tiempo (Mundial 2001). Este indicador, convertido en variable de estudio para la elaboración de esta tesis, es empleado a nivel internacional para valorar la actividad económica o riqueza de un país. Para calibrar o medir el nivel de vida de la población de un determinado país se suele utilizar este indicador que dividido entre la población, lo que se denomina como PIB per cápita.

En el caso de la ciudad de Santander se utilizó el valor del PIB regional obtenido del INE (Instituto Nacional de Estadística) para obtener el nivel de ingresos per cápita en Cantabria. En el Gráfico 11 se puede comprobar la evolución de ese nivel de ingresos en comparación a la evolución de la demanda para el mismo periodo (2001-2012). En el periodo comprendido entre 2001 y finales de 2007 se aprecia como el nivel de ingresos tiene una tendencia positiva a lo largo de esos seis años. El aumento del nivel de ingresos se va dando de una forma regular y de una forma sostenida que hace que se pase de un promedio anual de 15.600 euros/per cápita en 2001 a los 18.000 en 2008. En el tramo comprendido entre 2001 y finales de 2003 se puede observar cómo a medida que el nivel de ingresos va aumentando el número de viajeros va descendiendo, un hecho que se refleja de forma más clara en los periodos de verano mientras que en los periodos de invierno la demanda de viajeros permanece prácticamente estable. A partir del 2004 el nivel de ingresos adquiere una dinámica aún más alcista que llegará a su máximo nivel hacia el periodo estival del año 2007 con unos 19.000 euros/per cápita, un nivel que iniciará un claro y marcado descenso que llegará hasta el final de la serie en el año 2012 en donde el nivel de ingreso se establecerá aproximadamente en los 15.700 euros/per cápita, un valor que devolvía el nivel de ingresos a lo que prácticamente fue en el año 2001. En apenas cinco años, de 2007 a 2012 el nivel de ingresos se redujo en algo más de un 17%. Una caída de rentas que queda explicada si tenemos en cuenta el escenario de la crisis económica iniciada en el año 2008 y cuyas consecuencias siguen perdurando en la actualidad.

En la gráfica se pueden reconocer cuatro fases, en la primera de ellas a medida que el nivel de ingresos se incrementa la demanda de viajeros comienza a descender (periodo



2001-2003), en una segunda fase coincidiendo con la tendencia al alza en el nivel de ingresos y por tanto de una mayor capacidad adquisitiva la demanda comienza a caer de una forma más importante (periodo 2003-2007). La tercera fase tiene su inicio a partir de finales de 2007 y coincide con el mayor nivel de ingresos per/cápita registrados en toda la serie (19.000 euros/per cápita). Esta tercera fase tiene como característica que a medida que el nivel de ingreso disminuye la demanda de transporte de viajeros tiende a estabilizarse e incluso a aumentar aunque no de manera notable. Una cuarta fase podría fijarse a partir del año 2009 en donde la caída acelerada del nivel de ingresos no repercute tan negativamente en la demanda de viajeros como había sucedido en 2008 con el impacto de la llegada de la crisis y su directa implicación en el paro, responsable del descenso en el nivel de ingresos, (véase Gráfico 12).

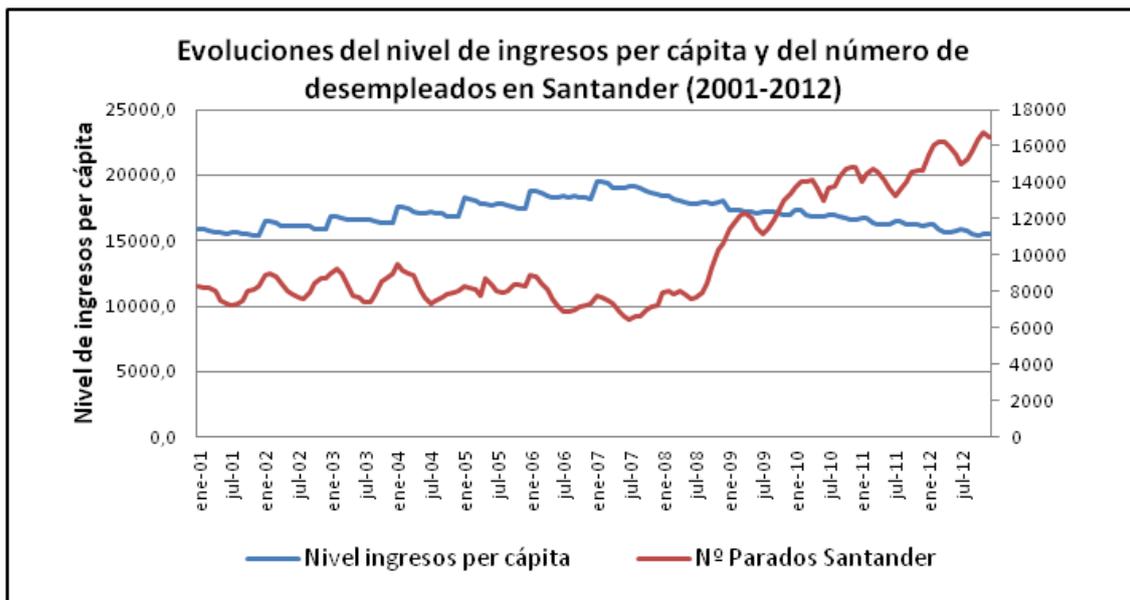


Gráfico 12. Elaboración propia a partir de los datos obtenidos del INE y el ICANE.

4.4.3. Demanda de viajeros y evolución del coste del combustible.

Otro de los indicadores que reflejan cual es el estado de la economía es el de los precios del petróleo cuyas variaciones tienen como consecuencia el aumento o disminución de



los precios de los combustibles como la gasolina o el gasóleo entre otros. En el Gráfico 13, se muestra el precio medio deflactado del litro de combustible en Cantabria (gasolina) para el periodo 2001-2012. Los precios, cuantificados en céntimos de euro están deflactados utilizando como base el año 2001 por lo que cuando se haga mención a los precios del resto de la serie se habla de precios deflactados y no precios reales.

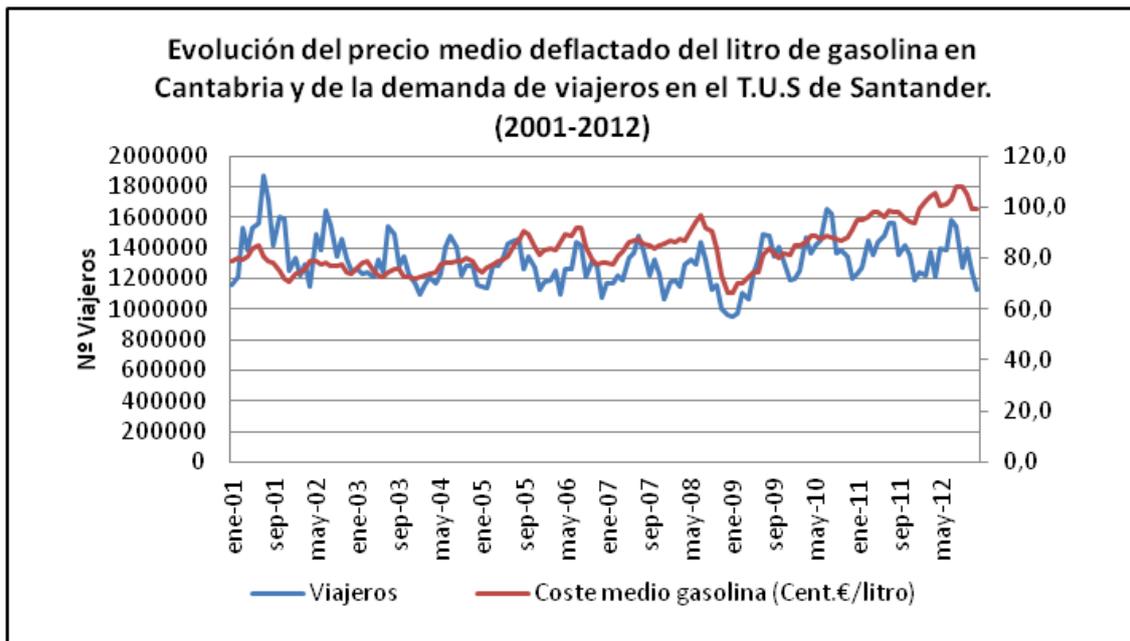


Gráfico 13. Elaboración propia a partir de los datos obtenidos del TUS de Santander, del INE y del Ministerio de Industria, Turismo y Comercio.

En el gráfico se aprecia cómo el precio medio del combustible en la región de Cantabria se mantiene bastante estable los primeros cuatro años de la serie, de Enero de 2001 a finales del verano de 2005 con un promedio anual de 77,3 cent/€ por litro de combustible. Los meses de Julio y Agosto de los años comprendidos entre 2001 y 2005 el promedio aumenta levemente por encima de dicha media situándose alrededor de los 79 cent/€, algo que sucede normalmente en periodos estivales en donde los precios de los combustibles experimentan una subida en comparación a meses precedentes para después volver a precios similares a los anteriores a verano. La diferencia del coste medio mensual entre los meses de verano y los del resto del año no suponía un



incremento demasiado acentuado, apenas dos céntimos de euro en ese periodo 2001-2005.

En lo que al precio medio mensual del litro de combustible a partir de 2005 comenzará a producirse un aumento de la diferencia entre los precios registrados en verano de los registrados el resto del año. De 2005 a 2008 el precio medio mensual del litro de combustible aumentó hasta los 84,2 cent/€, lo que supone una subida de un 9% con respecto al periodo anterior (2001-2005). Ese aumento de los precios del combustible en 2005 y 2008 coinciden también con un periodo en el cual el nivel de ingresos también se mantiene con una tendencia al alza como se podía observar en el Gráfico 11. Si se tiene en cuenta tan sólo en cuenta los meses de verano de los años 2005 a 2008 el precio medio del litro de combustible aumenta hasta los 89 cent/€, una diferencia de casi 6 céntimos de euro con respecto a los meses no estivales (83,6 cent.€). Con el impacto de la crisis iniciada a partir de 2008, los precios medios del combustible experimentan una fuerte subida. La irrupción de la “crisis” a lo largo del año 2008 trae como consecuencia que la diferencia en cuanto a precios con el año anterior en 2007 se incrementa, sobre todo en los meses de verano que son los que tradicionalmente experimentan subidas de los precios de los carburantes todos los años. Sin embargo se registra un fuerte incremento si observamos los precios medios habidos en el verano de 2007 con respecto a los de 2008. Si el promedio del coste del litro de combustible en los veranos de 2005 y 2006 era de 87,6 cent.€/litro y 89,3 respectivamente, en 2007 la media caía levemente a 86 cent.€/litro para posteriormente en el verano de 2008 llegar a los prácticamente 93.0 cent.€/litro lo que suponía un aumento del 8% con respecto al año anterior en el mismo periodo, (véase Gráfico 14).

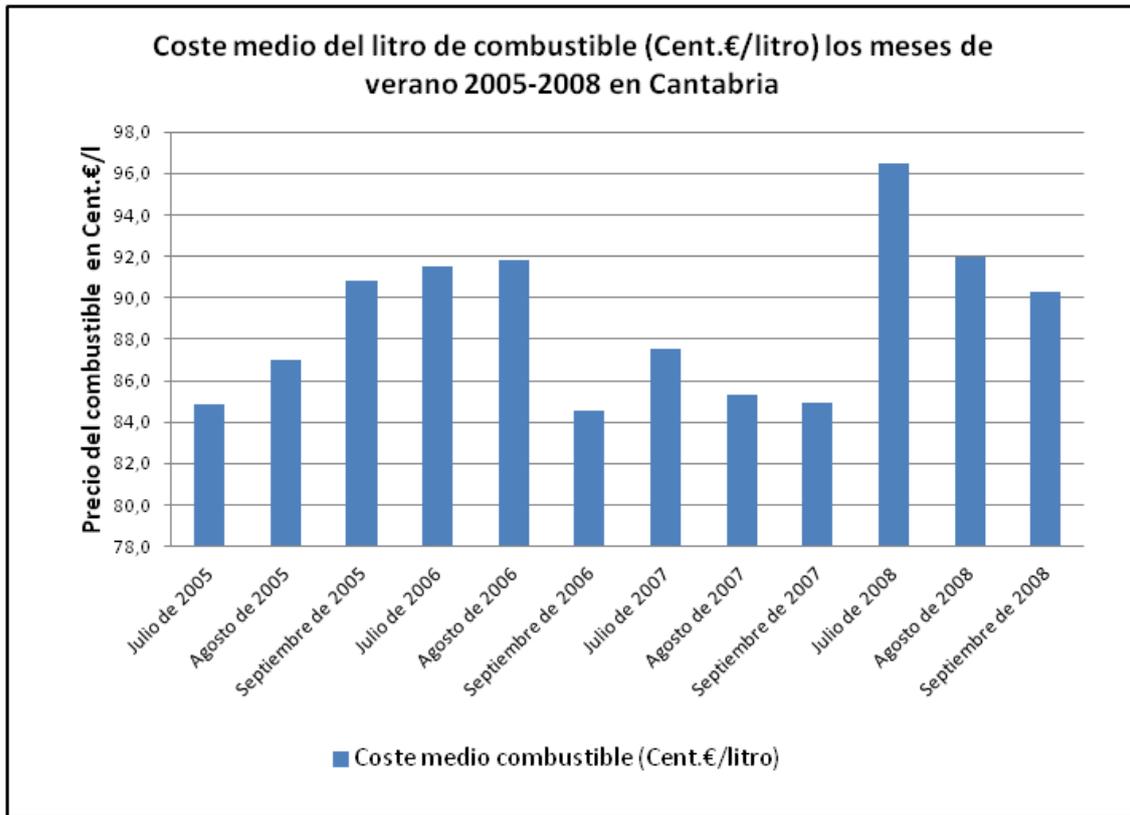


Gráfico 14. Elaboración propia a partir de los datos obtenidos del Ministerio de Industria, Turismo y Comercio.

A partir de los meses centrales del año 2008 el coste medio mensual de los combustibles comienza a descender. En el último trimestre del año 2008 los precios caen de forma significativa hasta situarse en una media de 74 cent.€/litro lo que representa casi un 29% menos. Esta vertiginosa caída se mantendrá durante el primer semestre del año 2009 cuando la media aún se reducirá hasta llegar a los 71,5 cent.€/litro. Será el segundo semestre del año 2009 el que suponga el inicio de una tendencia alcista en lo que a los precios del combustible se refiere puesto que el promedio se elevará a 81 cent.€/litro, un 14% con respecto al semestre anterior. La tendencia alcanzará a los siguientes semestres de manera que el impacto de la crisis en Cantabria coincide con una subida generalizada del precio de los combustibles. En el primer semestre del año 2010 el precio medio se elevaba a 87 cent.€/litro y en el segundo hasta 88,5 cent.€/litro. Para los dos semestres de 2011 vuelve a producirse un repunte significativo con medias del precio establecidas en 96,4 cent.€/litro en el primer semestre y 96,2 cent.€/litro en el segundo. En tan sólo dos años, desde el año 2009 el aumento del precio medio del



combustible había supuesto un 35%. En el año final de la serie se alcanzan los máximos valores con 102,1 cent.€/litro de media en el primer semestre y 104 cent.€/litro en el segundo.

Tras este breve repaso por la evolución del precio deflactado del combustible (gasolina), se podría establecer de igual manera la evolución de los precios reales sin deflactar, (véase

Gráfico 15).

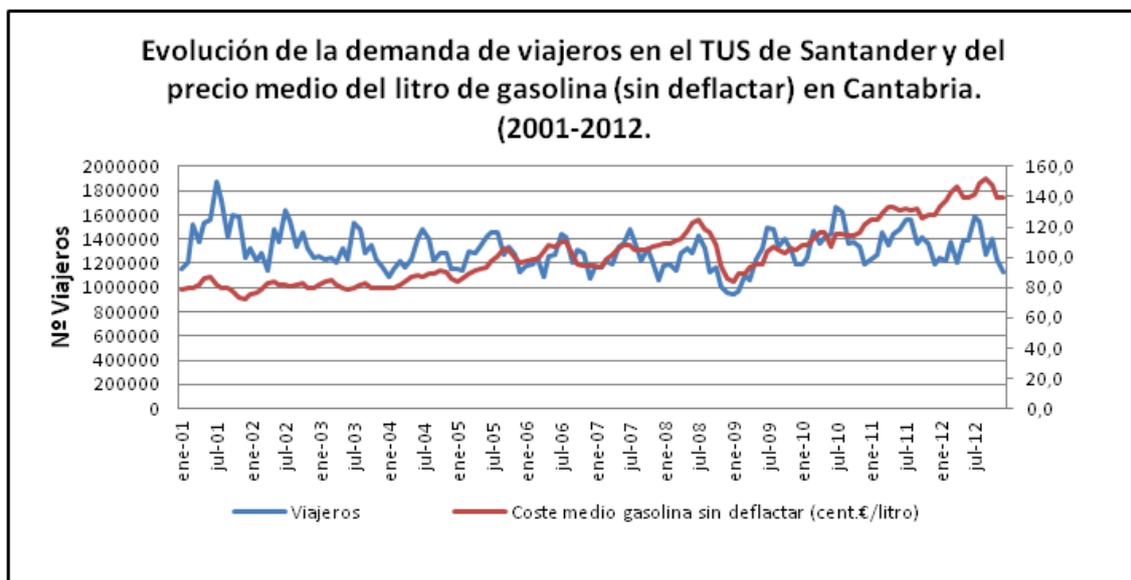


Gráfico 15. . Elaboración propia a partir de los datos obtenidos del TUS de Santander, del INE y del Ministerio de Industria, Turismo y Comercio.

Teniendo en cuenta la evolución de la demanda de viajeros con respecto a la evolución del coste medio del litro de combustible (gasolina) en Cantabria, sin ser este deflactado y midiéndose en valor real (véase

Gráfico 15), se puede ver cómo entre los años 2001 y 2003 el coste medio del litro de combustible (gasolina) se mantenía de forma regular en torno a los 80 céntimos de euro el litro, exceptuando un pequeño repunte del precio en los meses de Mayo a Julio de 2001 donde el precio medio alcanzó los 85 céntimos de euro de media. En ese mismo periodo la demanda superaba el pico de los 18 millones de viajeros siendo el más alto registrado en la serie para el periodo final de primavera y comienzo de verano de ese



año 2001. A partir de Enero de 2004 y hasta el verano de 2008 el estancamiento de los valores máximos de la demanda entorno a los 14-15 millones de viajeros en los periodos estivales y los 11,5 – 12 millones en los periodos invernales se da un progresivo aumento de los precios del combustible cuyos máximos valores coincidirán con los periodos de verano a diferencia del periodo anteriormente citado donde los precios de los carburantes no poseían tanta variabilidad en los precios en función de si era invierno o verano y permanecían más estables.

Es en el periodo final del año 2008 cuando se da una caída pronunciada de los precios de los combustibles que a su vez coincide con una bajada significativa de la demanda. Los últimos meses de 2008 junto con los primeros de 2009 coinciden con un periodo de invierno, lo que explica la bajada tanto del precio medio del litro de combustible como el descenso en el número de viajes que siempre encuentra en estos meses un descenso en el número de viajeros, eso sí en esta ocasión más acentuado ya que coincide con el impacto de la crisis económica y sus consecuencias.

Tras el impacto inicial de la crisis que deja sentir sus consecuencias tanto en el incremento del precio de los carburantes como en el estancamiento de la demanda, sobre todo en los meses de invierno y primavera, se observa que el precio de la gasolina comienza una tendencia progresivamente ascendente que irá desde los 88,5 cent/€ en Diciembre de 2008 hasta 1,05 €/litro en el mismo mes pero del año 2009, lo que supone un incremento del 18% en apenas un año. Esta tendencia en los precios medios del litro de la gasolina continuará creciendo llegando a cifrarse en 1,21 €/litro de media en Diciembre de 2010, lo que supone casi un 37% con respecto al periodo en el que podría definirse como el de impacto de la crisis en la economía. En el año 2011 y el precio medio del litro de gasolina continuó aumentando aunque de una forma más sostenida con un coste medio anual de 1,29 €/litro (1,32€/litro durante los meses de verano) y en el año 2012 se registran los mayores repuntes sufridos por el precio medio de la gasolina ya que de los 1,29€/litro de media anual se pasa a 1,39€/litro en 2012. El incremento tan elevado del precio medio de la gasolina en el año 2012 se debe a las fuertes subidas experimentadas en periodos anómalos del año ya que si normalmente los precios subían



en los meses de verano en el año 2012 los mayores precios coincidieron en los meses de primavera (de Marzo a Junio de 2012 con 1,43 €/litro de media) y con los meses de otoño (De Septiembre a Noviembre con valores cercanos a 1,48€/litro).

COSTE MEDIO SEMESTRAL DEL LITRO DE COMBUSTIBLE (Gasolina) EN CANTABRIA. PRECIOS REALES Y PRECIOS DEFLACTADOS		
PERIODO	PRECIO REAL (cents.€)	PRECIO DEFLACTADO (cents.€)
Primer Semestre 2001	82,2	81,3
Segundo Semestre 2001	77,7	75,9
Primer Semestre 2002	80,1	76,6
Segundo Semestre 2002	81,0	76,2
Primer Semestre 2003	81,6	75,5
Segundo Semestre 2003	80,7	73,8
Primer Semestre 2004	83,3	75,0
Segundo Semestre 2004	88,6	78,4
Primer Semestre 2005	89,3	77,9
Segundo Semestre 2005	100,9	86,3
Primer Semestre 2006	102,2	85,6
Segundo Semestre 2006	101,1	83,8
Primer Semestre 2007	100,3	82,0
Segundo Semestre 2007	106,4	85,4
Primer Semestre 2008	113,6	88,9
Segundo Semestre 2008	108,1	83,7
Primer Semestre 2009	91,2	71,4
Segundo Semestre 2009	104,9	81,5
Primer Semestre 2010	111,5	86,8
Segundo Semestre 2010	116,3	88,4
Primer Semestre 2011	129,4	96,4
Segundo Semestre 2011	129,6	96,2
Primer Semestre 2012	139,7	102,1
Segundo Semestre 2012	144,7	103,8

Tabla 5. Elaboración propia a partir de los datos obtenidos del Ministerio de Industria, Turismo y Comercio. Comparativa entre el coste real del litro de gasolina y el coste deflactado de forma semestral.

Los meses correspondientes al año 2012 en los que se registraron un coste medio del litro del combustible por encima de 1,40€/litro fueron Marzo (1,43€/litro), Abril (1,46€/litro) y de Agosto a Octubre con 1,49 , 1,51 y 1,48 respectivamente siendo el mes de Septiembre en el que más caro estuvo el litro de gasolina de toda la serie objeto de



estudio (1,51€/litro). La media anual para 2012 quedó establecida en 1,42€/litro. En cuatro años, (2009-2012) el precio medio del litro de gasolina en Cantabria había aumentado hasta un 60%. En la Tabla 5 se puede comprobar cuál fue la evolución de los precios medios de la gasolina por semestres, teniendo en cuenta por un lado los precios deflactados y los precios sin deflactar.

4.4.4. Demanda de viajeros y nivel de Motorización

Otra de las variables tenidas en cuenta en este estudio es el de la motorización, obtenida a partir del número de vehículos censados en la ciudad de Santander. En el Gráfico 16 puede verse la evolución anual tanto del número de vehículos censados en el ayuntamiento de Santander como el número de viajeros desde 2001 hasta 2012.

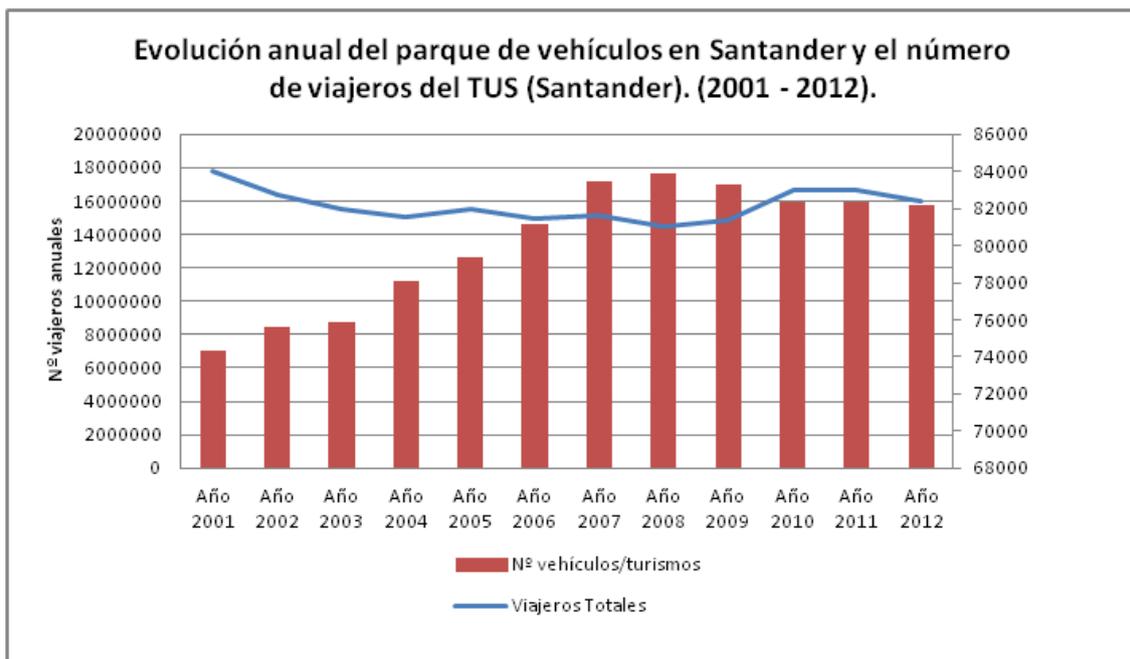


Gráfico 16. Elaboración propia a partir de los datos obtenidos del ayuntamiento de Santander y del TUS.

En el gráfico pueden reconocerse tres fases diferenciadas. En la primera de ellas queda patente como a medida que aumenta el número de vehículos (turismos) censados en la ciudad, la demanda de viajeros va descendiendo paulatinamente. Esa primera fase aglutinaría a los años



que van desde 2001 hasta 2008. Si en el año 2001 el parque de turismos censados en Santander era de 74.327 vehículos, en el año 2008 se llegaba al número máximo de vehículos censados con casi 83.900 vehículos, lo cual supone un aumento en porcentaje del 12%. Para ese mismo periodo el número de viajeros totales anuales se redujo de los aproximadamente 17,8 millones de 2001 a los 14,4 millones de 2008, una caída de un 19%.

En una segunda fase, de nuevo el impacto de la crisis vuelve a notarse en las cifras de turismos censados ya que a partir de 2009 el parque de vehículos comienza a descender de forma lenta. En 2009 el número de turismos censados era de 83.300, unos 600 vehículos menos que el año anterior que había sido el que mayor número de vehículos registrados tuvo de toda la serie de estudio. Ese leve descenso de vehículos censados coincide con una pequeña subida del número de viajeros en el año 2009 que pasa a 14,8 millones, lo que representa un aumento total de 400.000 viajeros. La tercera y última fase comprende los años 2010 al 2012, donde las cifras de vehículos censados se estabilizan en torno a los 83.200 y el número de viajeros experimenta una importante subida pasándose de los 14,8 millones de 2009 a los 16,7 millones de media para los años 2010 y 2011. En el año 2012 el número de vehículos apenas sobrepasa los 82.200 vehículos una cifra similar a la del año 2006 mientras que el número de viajeros caía hasta los 16 millones, un dato cercano al obtenido en el año 2002 en el que el número de viajeros fue de 16,4 millones.

4.4.5. Evolución de la Demanda en el periodo de Crisis Económica (Fase Contractiva) (2009-2012)

A tenor de los datos analizados de algunas de las principales variables tenidas en cuenta para la elaboración de este estudio es importante remarcar el punto de inflexión que supuso el final del año 2008 y el comienzo de 2009. A partir de ese punto de inflexión se concatenaron de forma sucesiva, por un lado, los incrementos de la tasa de paro y de los precios de los combustibles y por otro los descensos más o menos significativos tanto del nivel de ingresos como del número de vehículos censados. En vista de los resultados podría situarse el inicio de la crisis en Santander a partir de los meses finales del año 2008 y primeros meses del año 2009.



En el caso de los datos de viajeros en el TUS los meses de tránsito entre 2008 y 2009 también supusieron un periodo en el que la demanda cayó pero que posteriormente en los años siguientes de la crisis hasta 2012 pudo frenar dicha caída e incluso pudo incrementarse levemente el número de viajeros debido a una serie de causas que se explicarán posteriormente (véase Gráfico 17). Sin embargo si se tiene en cuenta la recaudación se produce un descenso de la misma a medida que se incrementa levemente la demanda.

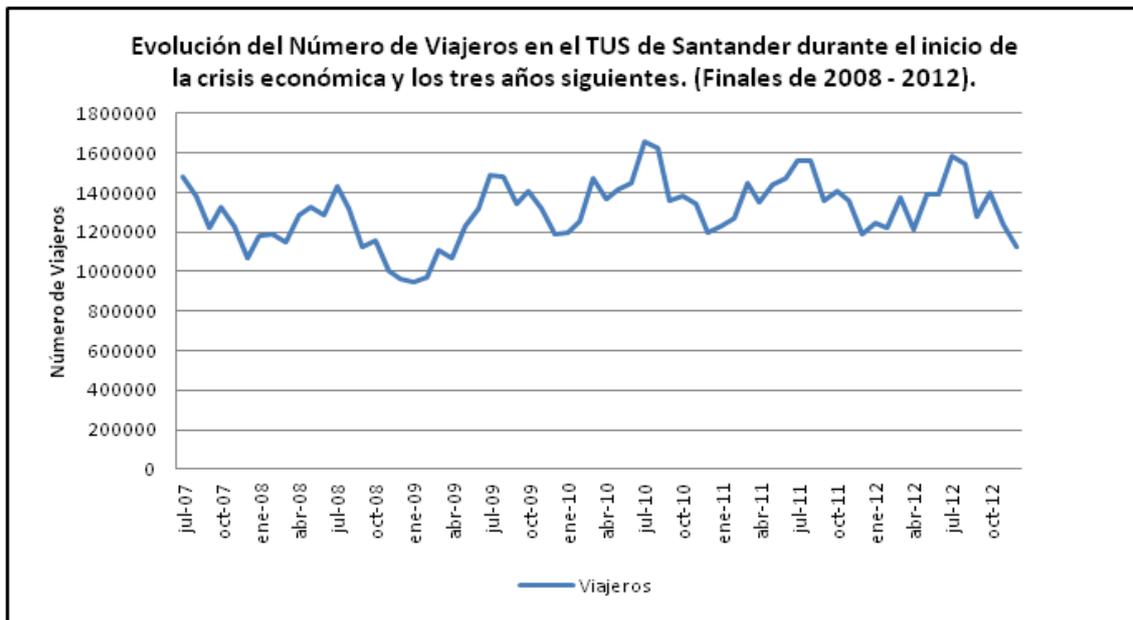


Gráfico 17. Elaboración propia a partir de los datos obtenidos del TUS.

En el Gráfico 17 se aprecia un gran “valle” que va desde aproximadamente el mes de Noviembre de 2008 hasta el mes de Febrero de 2009. En esos tres meses se registraron las cifras de viajeros más bajas de toda la serie objeto de estudio (2001-2012). En el mes de Noviembre de 2008 apenas se superó el millón de viajeros (1.006.891) y de Diciembre de 2008 a Febrero de 2009 ni siquiera se llegó a registrar el millón con 960.014, 948.444 y 969.060 respectivamente.

El número anormalmente bajo de viajeros de esos meses se debió fundamentalmente a que a partir del mes de Noviembre de 2008 desaparece el bono bus ordinario (cartón) como sistema de pago. Hasta el mes de Noviembre de 2008 la recogida de los datos de



viajeros que usaban ese sistema de pago habían sido contabilizados pero a partir de ahí y hasta el mes de Marzo los usuarios que aún contaban con dicho bono bus no fueron contabilizados ya que en ese mismo periodo se da la implantación de un nuevo sistema de recogida de datos siendo en ese transcurso cuando se pierden los datos de usuarios de bono bus. Teniendo en cuenta el número de viajeros que utilizaron bono bus de cartón durante los tres primeros meses del año 2008 con 189.996 en Enero, 172.121 en Febrero y 161.593 en Marzo se estima que pudieron perderse una media de 170.000 viajeros cada uno de los meses que conforman el “valle” citado anteriormente por lo que las cifras reales en cada uno de esos meses comprendidos entre Noviembre de 2008 y Febrero de 2009 habrían rondado entre 1,10 Y 1,15 millones de viajeros. A pesar de no contabilizarse ese número de usuarios, en caso de que hubieran sido contabilizados el periodo de tránsito entre 2008 y 2009 igualmente hubiera sido el que habría registrado el menor número de viajes de toda la serie.

En relación a la evolución de la demanda, cabe destacar que tras el impacto de la crisis económica en Santander, que se puede fijar en función de los distintos datos analizados a finales de 2008 y comienzos de 2009, la demanda entra en una fase de estancamiento con una leve tendencia a incrementarse en 2009 y a mantenerse en guarismos estables para el trienio (2010-2012), algo que entra en contradicción con los datos de demanda en transporte público de otras ciudades de España en donde la demanda tiende a caer. El promedio mensual para el año 2009 es de casi 1,23 millones de viajeros mientras que para 2008 ese promedio fue de 1,20 millones, lo que supone que en el primer año de la crisis el volumen promedio de viajeros aumentó en aproximadamente 35.000 viajeros de promedio mensual. A medida que continuaba la crisis, en el año 2010 el número promedio mensual de viajeros continuó aumentando hasta los 1,39 millones, unos guarismos que tenderán a estancarse en torno a una media de 1,35 millones para los años 2011 y 2012.

El aumento de la demanda durante el periodo de crisis económica coincide con distintas decisiones que se tomaron por tanto del Ayuntamiento de Santander en relación a las políticas tarifarias y que supusieron un importante avance de carácter social. La

aplicación de esas políticas consiguieron frenar la caída de la demanda aunque a costa de la reducción de los ingresos recaudados. A partir de 2008 se aplicó como novedad que todos los miembros de familias numerosas de la ciudad podrían viajar de forma gratuita. Hasta 2008 las familias numerosas disfrutaban de una bonificación del 50 % en el transporte municipal. Además de las familias numerosas también se verían beneficiados otros colectivos sociales como los pensionistas con ingresos limitados, las personas con una discapacidad superior al 66% y bonificaciones para jóvenes a través del uso de un carné trimestral, (véase Gráfico 18).

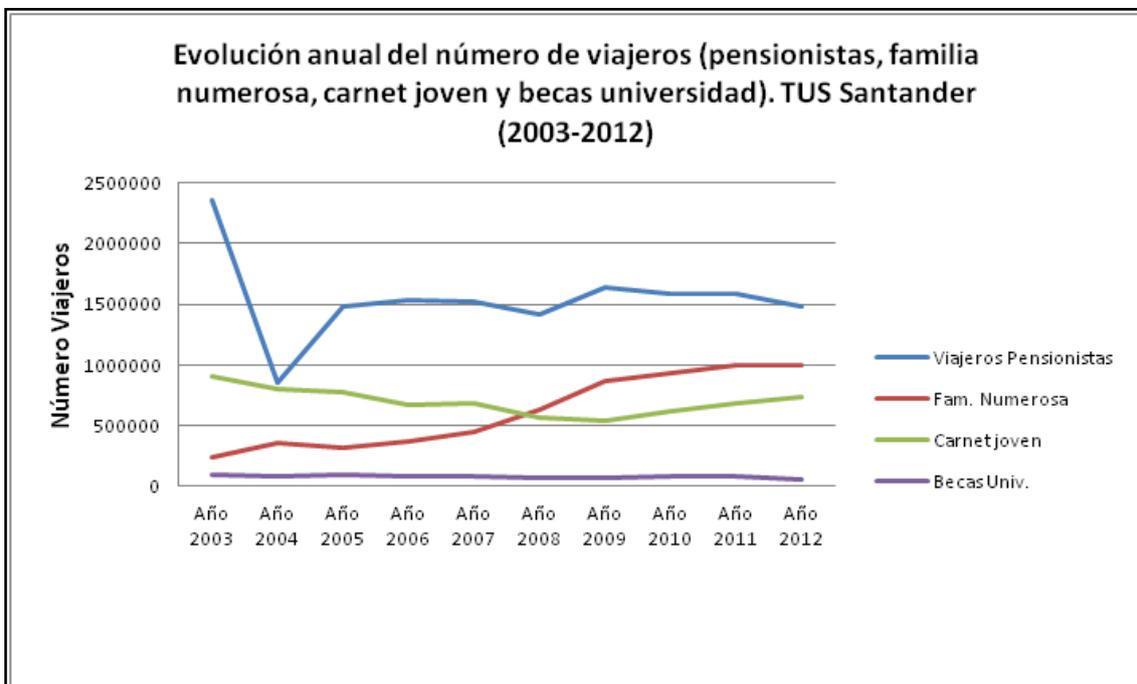


Gráfico 18. Elaboración propia a partir de los datos del TUS de Santander. Número de viajeros según perfiles.

Tal y como se muestra en el gráfico anterior a partir del año 2004 se da un aumento importante en el número de viajeros cuyo perfil se corresponde con el de pensionistas pasándose de algo más de 0,9 millones en el año 2004 a aproximadamente 1,5 millones al año siguiente. La cifra de viajeros pensionistas se mantendrá estable hasta el año 2007, en torno a esa cifra 1.5 millones y en el 2008 sufrirá una leve caída que la situará por debajo de dicha cifra. A partir de 2008 comienza a darse un pequeño ascenso que se mantendrá hasta 2010, especialmente en al año 2009 con 1.636.175 viajeros. En los



dos últimos años 2011-2012 la demanda de viajeros vuelve a estar por debajo de los 1.5 millones.

PORCENTAJE DE VIAJES ANUALES SEGÚN PERFILES EN EL TUS (SANTANDER) 2007-1012						
Año	Número Total de Viajeros	%. Pensionistas	%. Familia numerosa	%. Carne t Joven	%. Becas universidad	%. Total (Viajes subvencionados)
2007	15.327.094	9,9	2,9	4,4	0,52	17,8
2008	14.444.079	9,8	4,4	3,9	0,49	18,6
2009	15.009.917	10,9	5,8	3,6	0,46	20,7
2010	16.648.788	9,5	5,6	3,7	0,47	19,3
2011	17.037.259	9,3	5,8	4	0,47	19,6
2012	16.000.485	9,3	6,3	4,6	0,36	20,5

Tabla 6. Elaboración propia a partir de los datos del TUS de Santander. Porcentajes de tipos de viaje según perfil.

En la Tabla 6 aparecen recogidos los porcentajes de los distintos perfiles de viajeros que se han visto favorecidos por las políticas tarifarias aplicadas por parte del ayuntamiento de Santander. Teniendo como base la demanda total anual para el periodo 2007-2012 se han calculado los porcentajes de cada tipo de perfil con respecto a cada una de las demandas totales para cada año. Para el caso de los pensionistas el porcentaje oscila entre el 9,3 y el 10% de la demanda total de viajeros para cada año del periodo 2007-



2012 siendo en 2009 cuando represente el porcentaje mayor con casi un 11% del total de la demanda para ese mismo año. En cuanto a las tarifas el coste del viaje a los pensionistas es gratuito desde el año 2004 hasta la actualidad.

En el caso de viajes referentes a familias numerosas el porcentaje de este tipo de viajeros se va incrementando de manera progresiva y cada año aumenta en cuanto a porcentaje con respecto a la demanda total. En el año 2007 la cifra de viajeros de familias numerosas se situó en 449.039, un 2,9% sobre el total de viajeros de ese año y al año siguiente la cifra se incrementó de manera notable hasta los 633.802 una subida del 41% con respecto al número de viajeros del año anterior y que supusieron el 4,4% de la demanda total de viajeros en 2008. Apuntar que en cuanto al coste de los viajes desde el año 2004 al 2007 el coste variaba entre los 0,22 cent/€ de 2004 y los 0,27 cent/€ en 2007 y desde 2008 hasta 2012 la tarifa pasó a ser gratuita. En los años sucesivos, 2009, 2010, 2011 y 2012 la tendencia del número de viajeros pertenecientes a familia numerosa seguirá creciendo de una manera notable con algo más de 935.000 viajeros en 2010, 997.000 en 2011 y superando el millón en 2012, exactamente 1.004.779. Tomando como base la cifra dada de viajeros en 2007 para 2012 el porcentaje de este perfil de viajeros se incrementó de una forma muy notoria. En cuatro años el número de viajeros de familias numerosas se incrementó en más del doble, más de un 120%. En cuanto a porcentaje total con respecto a la demanda anual los viajeros de familia numerosa pasan de suponer el 5,8% en el año 2009 al 6,3% en 2012.

El número de viajeros cuyo perfil equivale al de carnet joven tendrá un comportamiento distinto al de pensionistas y familias numerosas ya que del año 2007 al año 2010 el número de viajeros de este colectivo tendrá una tendencia decreciente. Si en el año 2007 el número de viajeros de este perfil fueron 682.900, en 2008 esas cifras se vieron mermadas hasta situarse en los 567.644 y en 2009 en 540.771, unas caídas porcentuales del 16,9% y un 20,8% respectivamente. A partir del año 2010 la tendencia se invertirá y para ese mismo año se superan los 620.000 viajes, una tendencia que se mantendrá al alza con casi 684.000 viajeros en 2011 y rozando los 739.000 en 2012. De 2010 a 2012 la demanda de viajeros con carnet joven se incrementó inicialmente en un 10,3% en



2011 y un 19,1% en 2012. Al igual que en el caso de los perfiles de usuarios anteriormente mencionados la tarifa pasó a ser gratuita a partir del año 2008.

El porcentaje de usuarios de carnet joven con respecto a la demanda total de 2007 era de un 4,4%, un porcentaje que se estabilizaría entre el 3,6% y el 4% para los tres años posteriores. El porcentaje de viajeros con carnet joven con respecto a la demanda total aumentará en los dos últimos años del periodo cuando experimente un salto de casi un 1% pasándose del 3,7% de viajeros en 2010 al 4% en 2011 y 4,6% en 2012. En cuanto a las cifras de usuarios de tarjeta universitaria apenas sufren cambios rondando el 0,5% de la demanda total para todos los años excepto para el 2012 que no llega al 0,4%.

Teniendo en cuenta el porcentaje total de viajeros favorecidos por las políticas tarifarias se puede verificar que el porcentaje de dichos viajeros pasa de una media de poco más del 18% en 2007 (algo más de 3,6 millones de usuarios frente a los 15,8 totales) a casi el 21% en el año 2009 (3,1 millones de 15,1 millones de viajeros totales ese año). El aumento de ese porcentaje de sectores sociales favorecidos se refleja sobre todo a partir de la gratuidad de los viajes que se aplica a partir del año 2008, año que coincide con el inicio de un periodo contractivo de la economía tanto a nivel regional como nacional. En los años 2010 y 2011 el porcentaje de viajeros con gratuidad en los viajes se situarán en torno a un 19,5% de la demanda total y para el el año 2012 un 20,5%.

ya que el número total de viajeros en 2008 fue de 14,4 millones y en 2009 esa cifra se incrementó hasta los 15,1 lo que representa un aumento aproximado de 565.000 viajeros más, lo que supone un aumento en 2009 con respecto a 2008 de un 4%. En los años 2010 y 2011 el porcentaje se sitúa en torno al 19,5% rebasando el 20% de nuevo para el año 2012

4.5. Análisis de datos incluyendo variable demográfica. Per cápita.

Datos generales.

En este epígrafe tanto la demanda como el nivel de servicio del transporte en bus urbano se han considerado en forma per cápita, lo que evita introducir la población como variable independiente en el modelo (Holmgren 2007).



Gráfico 19 Evolución del número anual de viajeros per cápita (color rojo) y vehículos/km per cápita (color verde). Periodo 2001-2012.

En el Gráfico 19 puede observarse la evolución de la demanda de transporte público per cápita en el periodo 2001 – 2012 teniendo en cuenta que esa demanda incluye la totalidad de las líneas objeto de estudio. En un primer intervalo 2001 – 2008, la demanda cae progresivamente desde los 96 pasajeros por habitante hasta alrededor de 79 viajes por habitante en 2008. Más adelante, entre 2009 y 2012 el número de pasajeros por habitante vuelve a incrementarse hasta alcanzar la cifra de 90. En cuanto al nivel de servicio medido como vehículos – km/cápita, estos se incrementaron progresivamente desde los 19 en 2001 hasta los 20,5 en el año 2007. Posteriormente se da un descenso en el nivel de servicio hasta situarse en 2009 en los 19 vehículos – km/cápita, con un incremento final de la serie hasta los 20,2 en 2012.



Así mismo en los siguientes gráficos, (véase Gráfico 20 y Gráfico 21), se muestran también la evolución de cuatro variables estructurales durante el período 2001-2012: Tasa de paro o desempleo, renta per cápita o ingresos per cápita, vehículos privados por habitante y precios deflactados del combustible (gasolina). Cabe destacar el aumento de la tasa de desempleo y el estancamiento de los vehículos privados por habitante en el área de estudio inicialmente desde el año 2007 y un incremento importante del paro sobre todo a partir del año 2008.

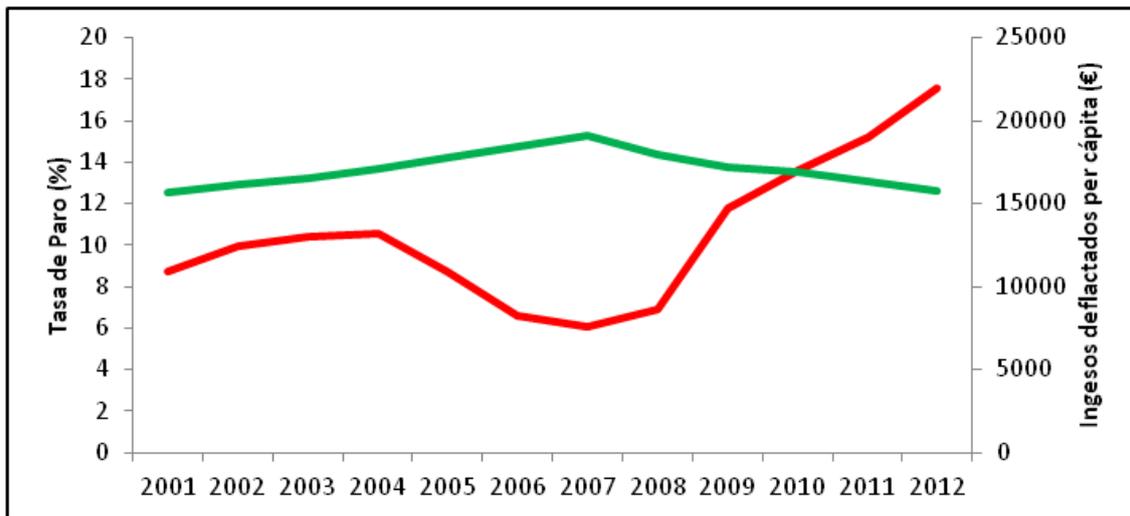


Gráfico 20. Tasa de Paro (color rojo) e Ingresos per cápita deflactados (color verde), Periodo 2001-2012.

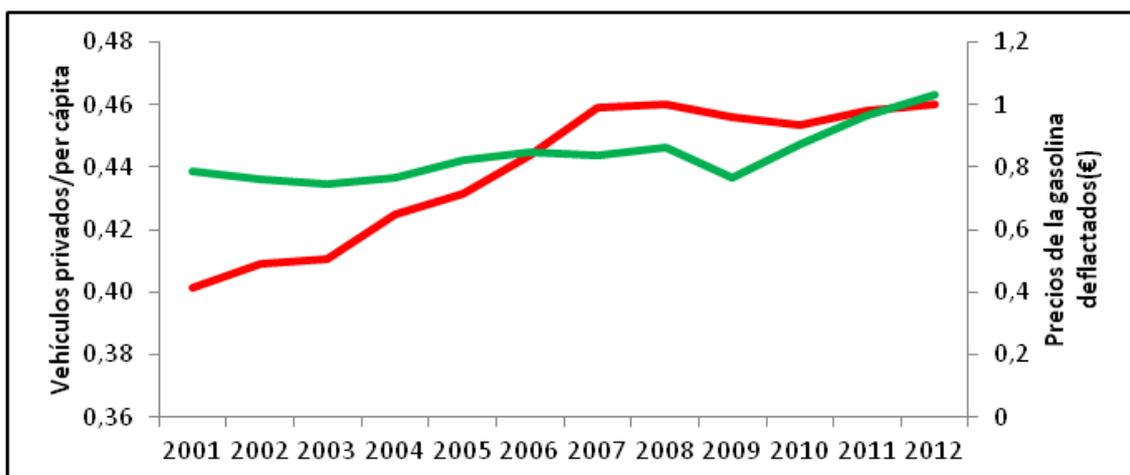


Gráfico 21. Tasa de vehículos privados per cápita (color rojo) y el precio medio deflactado de los combustibles (color verde). Periodo 2001-2012.



Para llevar a cabo este análisis se ha contado con un total de 144 observaciones, cada una de ellas refiriéndose a un dato mensual. La serie compuesta por esas 144 observaciones engloba los 12 años considerados. Los estadísticos descriptivos de las variables contenidas en la base de datos pueden consultarse en la Tabla 7.

Variable	Valor Min.	Valor Max.	Media	Std. Desviación	Descripción
TRIPPC	5.19	10.11	7.19	0.85	Viajeros-km/per capita
FARE	0.29	0.50	0.40	0.04	Recaudación (deflactada €/viaje)
VEHKM	1.38	1.79	1.64	0.07	Buses – km/per capita
FUEL	0.67	1.08	0.84	0.09	Precio deflactado del combustible (€)
URATE	4.70	19.20	10.50	3.47	Tasa de Paro
INCOME	15401	19508	17072	1053	Ingresos deflactados per cápita (€)
POP	178659	185231	182694	1907	Población
VEH	0.40	0.46	0.44	0.02	Vehículos privados/per capita

Tabla 7. Estadísticas descriptivas de las variables contenidas en la base de datos mensual de las líneas objeto de estudio agregadas (N=144)

Todas las variables monetarias fueron deflactadas de cara a tener en cuenta el efecto de la inflación tomando como año base 2001. La variable dependiente se corresponde con TRIPPC, es decir, los pasajeros per cápita realizados en el sistema de transporte público. La variable VEHKM representa la oferta de transporte público disponible y está expresada como buses – km per cápita. Las variables URATE e INCOME, centrales para los objetivos de esta tesis, representan respectivamente la tasa de paro obtenida como el cociente entre la población parada y la población activa por 100 y el GDP per cápita respectivamente. Ambos indicadores han sido obtenidos a partir de datos proporcionados por el Instituto Nacional de Estadística de España (INE 2013). Todas las variables fueron transformadas obteniendo su logaritmo natural para que los parámetros puedan interpretarse directamente como elasticidades.



La variable FARE representa la recaudación total deflactada entre el número total de viajes realizados. Esta recaudación media por pasajero es un proxy ampliamente utilizado de las tarifas reales si bien tiene limitaciones en áreas con un sistema de tarifas complejo (De Rus 1990). En el área estudiada esta estructura no es excesivamente compleja diferenciándose entre los billetes pagados en metálico con tarifa ordinaria, los billetes multiticket con una tarifa en torno al 60% de la tarifa ordinaria y los billetes gratuitos para viajeros de colectivos especiales (personas jubiladas principalmente). Este último tipo de usuario, no se ve afectado por la evolución de la variable de FARE ya que representa alrededor del 30% de la demanda total. Su inclusión en el cálculo de los ingresos medios por viaje no es problemático, ya que su cuota de mercado se ha mantenido estable durante el período 2001-2012.

4.6. Análisis de datos incluyendo variable demográfica. Per cápita. Datos por línea. Estadísticos Descriptivos.

Al igual que para los datos generales también se ha tenido en cuenta tanto la demanda como el nivel de servicio del transporte en bus en forma per cápita, lo que evita introducir la población como variable independiente en el modelo.

4.6.1. LINEA 1 - Parque Tecnológico (PCTCAN) – Adarzo – Arsenio Odriozola.

En el Gráfico 22 puede comprobarse la evolución de la demanda de transporte público per cápita en el periodo 2001 – 2012 para la línea 1. En los dos primeros años de la serie, de 2001 y 2002, la demanda se mantiene en torno a los 30 y 27 pasajeros por habitante respectivamente para en el año siguiente caer de forma abrupta hasta prácticamente la mitad, 14 pasajeros por habitante. En 2004 llega a su mínimo con algo más de 12 y a partir de ahí la cifra de pasajeros por habitante va aumentando progresivamente y de manera sostenida hasta el año 2010 cuando logra un máximo de 17 pasajeros por

habitante. Del año 2010 al 2012 vuelve a producirse un leve descenso poniéndose fin a la tendencia iniciada en 2004 y de los 17 del año 2010 baja a algo menos de 15 en 2012.



Ilustración 21. Recorrido de la línea 1. Fuente: Aplicación cartográfica del (S.A.E). Sistema de ayuda a la explotación del TUS de Santander.

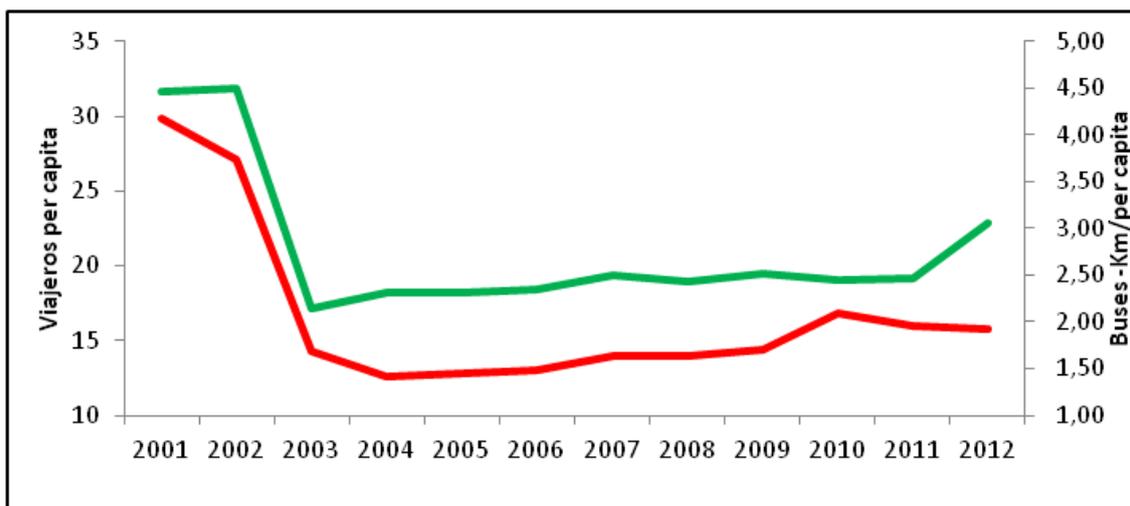


Gráfico 22. Evolución del número medio anual de pasajeros – km per cápita (color rojo, eje izquierdo) y de los buses – kilómetro per cápita (color verde, eje derecho) en la línea 1 de Transportes Urbanos de Santander. Periodo 2001 - 2012.



En cuanto al nivel de servicio medido como vehículos – km/cápita, estos se mantuvieron estables alrededor de los 4,5 de media en los dos primeros años de la serie para después en 2003 caer hasta los 2,15 el valor anual más bajo en lo que a vehículos-km/cápita se refiere. En el periodo 2003 – 2006 ese nivel de servicio se estabilizará en torno a los 2,35 y a partir de 2006 hasta 2011 experimentará un pequeño aumento que lo situará en un valor cercano a los 2,5. Un pequeño salto porcentual hará que en el último año de la serie el nivel de servicio llegue a 3, siendo este valor el más alto desde la fuerte caída experimentada en el año 2003.

Como hecho destacable cabe señalar que en el segundo año de la serie, 2002, a lo largo del primer semestre convivieron dos tipos de moneda como forma de pago en metálico, la peseta y el euro. La coincidencia de ambas monedas en el mismo periodo pudo haber repercutido en cierta manera en el descenso de la demanda ya que si en 2001 con la peseta como moneda la cifra era prácticamente la treintena para el año 2002 con la llegada de la nueva moneda común europea esos guarismos bajaron a 27.

La fuerte caída de la demanda y del nivel de servicio para el año 2003 se explica tras la puesta en marcha de un estudio cuyo objeto era la mejora del transporte urbano en las zonas de Valdenoja y San Román y que se materializaba en 2003 de forma que la línea 1 sufría un desdoblamiento en dos siendo una parte la línea 1 modificada y su desdoblamiento que pasaría a denominarse como línea 2.

Eso explica que tanto demanda como nivel de servicio cayeran prácticamente la mitad de 2002 a 2003 y la razón se explica por ese cambio a nivel de reestructuración de las líneas de transporte. Unos cambios que también acabarían afectando a los recorridos de otras líneas como las 6C1, 6C2 y otras que conllevaron fusiones como las de la 8 y la 9 para acabar formando la 13.



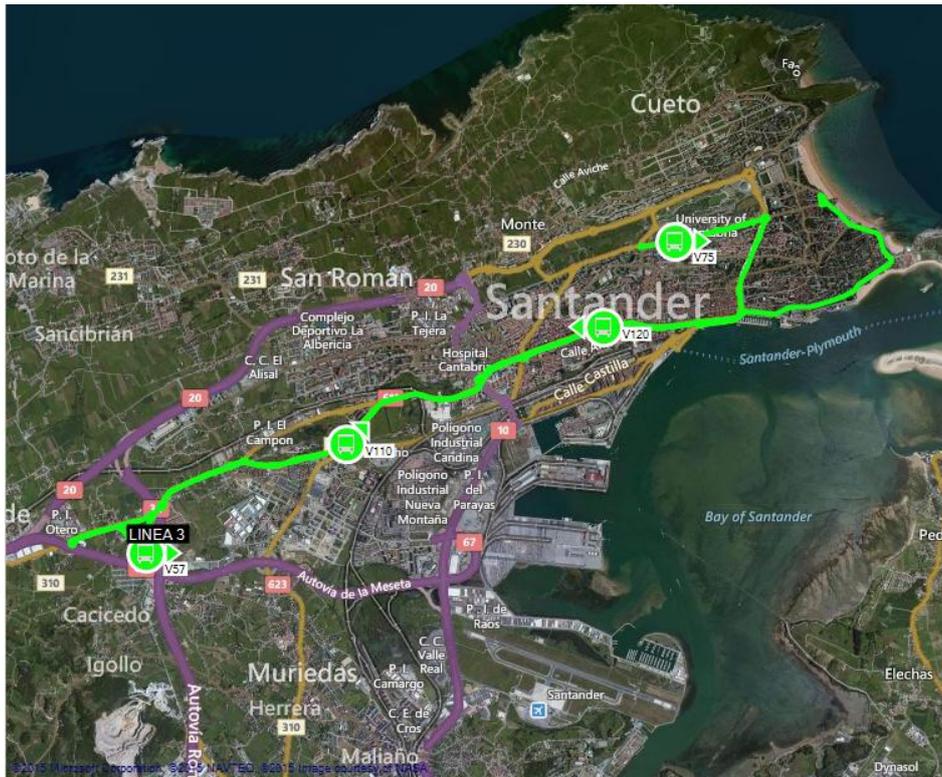
Variable	Valor Min.	Valor Max.	Media	Std. Desviación	Descripción
TRIPPC	0.84	2.99	1.39	0.50	Viajeros-km/per capita
FARE	0.25	0.50	0.37	0.04	Recaudación (deflactada €/viaje)
VEHKM	0.14	0.40	0.23	0.07	Buses – km/per capita
FUEL	0.67	1.08	0.84	0.09	Precio deflactado del combustible (€)
URATE	4.70	19.20	10.5	3.47	Tasa de Paro
INCOME	15401	19508	17072	1053	Ingresos deflactados per cápita (€)
POP	178659	185231	182694	1907	Población
VEH	0.40	0.46	0.44	0.02	Vehículos privados/per capita

Tabla 8. Estadísticas descriptivas de las variables contenidas en la base de datos correspondiente a la línea 1. (N=144).

Al igual que en el caso de los datos de líneas agregados se ha partido con un total de 144 observaciones mensuales de la línea 1 para los 12 años considerados. Los estadísticos descriptivos de las variables contenidas en la base de datos pueden consultarse en la Tabla 8. Estadísticas descriptivas de las variables contenidas en la base de datos correspondiente a la línea 1. (N=144).

4.6.2. LINEA 3 - Ojáiz – Paseo Pereda – Rectorado (Universidad)

En la gráfica correspondiente a la línea 3, (ver Gráfico 23), puede observarse la evolución de la demanda de transporte público per cápita en el periodo 2001 – 2012 en la línea 3. En 2001, primer año de la serie, coincide el valor máximo de demanda per cápita de la serie con casi 8 pasajeros por habitante. A partir de ese primer año se da un paulatino descenso hasta el año 2004 con un valor sensiblemente por encima de los 6 pasajeros per cápita. A partir de 2004 se da una alternancia de subidas y bajadas en el número de pasajeros per cápita y no se denota una tendencia clara tanto de ascenso como de descenso. En 2005 se da un leve aumento hasta casi los 7 pasajeros pero ese aumento no consigue mantenerse y en 2006 las cifras vuelven a caer hasta 6,1, el valor más bajo de la serie.



**Ilustración 22. Recorrido de la línea 3. Fuente: Aplicación cartográfica del (S.A.E).
Sistema de ayuda a la explotación del TUS de Santander.**

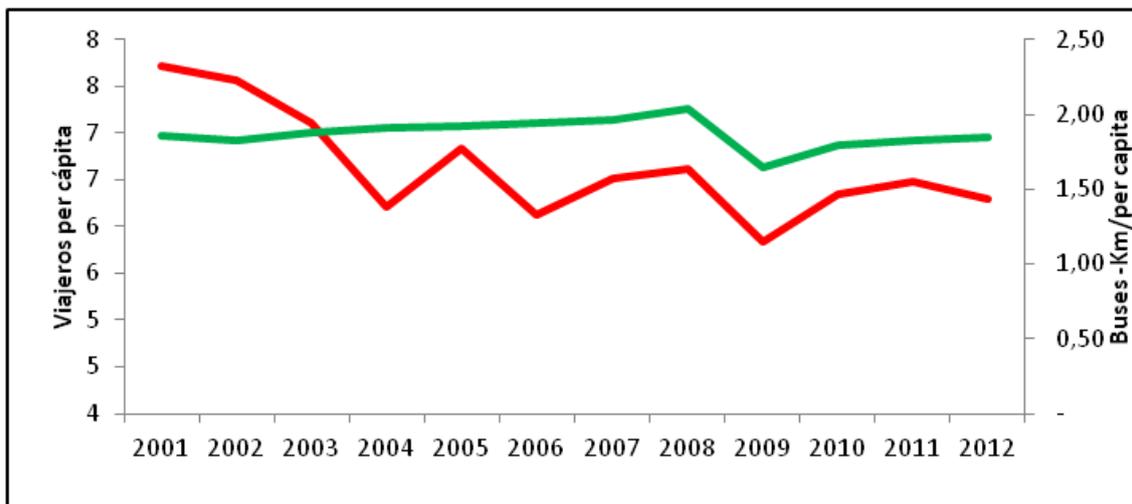


Gráfico 23. Evolución del número medio anual de pasajeros – km per cápita (color rojo, eje izquierdo) y de los buses – kilómetro per cápita (color verde, eje derecho) en la Línea 3 de Transportes Urbanos de Santander. Periodo 2001 - 2012.



Desde 2006 los datos experimentan una lenta mejoría y en los dos años sucesivos, 2007 y 2008, el número de viajeros per cápita vuelve a establecerse por encima de los 6,5. A partir de 2008 de nuevo un leve descenso en 2009 y un pequeño repunte en 2010 y 2011 con valores cercanos al 6,5. En 2012 las cifras se estancan de nuevo a la baja con 6,3 , una de las cifras más bajas de toda la serie. De 2010 a 2012 se da un leve incremento en el primer año y un estancamiento en los dos años siguientes.

En cuanto al nivel de servicio medido como vehículos – km/cápita, este se mantuvo muy estable a lo largo de toda la serie y con mínimas oscilaciones. El promedio de vehículos – km/cápita para los doce años de la serie es de 1,87. De 2001 a 2007 las cifras oscilan entre 1,8 y 1,9 y es en 2008 cuando se da el valor más alto superando de forma escasa los 2 vehículos – km/cápita. Es precisamente a partir de 2008 cuando se produce una bajada de prácticamente medio punto llegándose a 1,6 en 2009, el valor más bajo de la serie. A partir de 2009 las cifras se recuperan paulatinamente y con tendencia positiva hasta volver de nuevo a establecerse unas cifras similares a las de los primeros años de la serie con 1,8.

Señalar que la Línea 3 cuenta con una de sus cabeceras de línea en Ojáiz por lo que se trata de la línea principal que da servicio al núcleo urbano de Peñacastillo. Resulta llamativo que a pesar del fuerte incremento de la población en esta zona, de algo más de 10.000 habitantes en 2001 a cerca de 16.000 en 2010, no parece que eso se haya traducido en una mayor número de viajeros en este línea aunque a su vez el nivel de servicio ha permanecido prácticamente inalterable a pesar de que Peñacastillo aumentó en más de un 50% su población en apenas diez años.

Para la línea 3 se cuenta igualmente con un total de 144 observaciones mensuales para los 12 años considerados. Los estadísticos descriptivos de las variables contenidas en la base de datos pueden consultarse en la Tabla 9.

Variable	Valor Min.	Valor Max.	Media	Std. Desviación	Descripción
TRIPPC	0.31	0.84	0.55	0.08	Viajeros-km/per capita
FARE	0.35	0.50	0.41	0.03	Recaudación (deflactada €/viaje)
VEHKM	0.05	0.21	0.15	0.01	Buses – km/per capita
FUEL	0.67	1.08	0.84	0.09	Precio deflactado del combustible (€)
URATE	4.70	19.20	10.5	3.47	Tasa de Paro
INCOME	15401	19508	17072	1053	Ingresos deflactados per cápita (€)
POP	178659	185231	182694	1907	Población
VEH	0.40	0.46	0.44	0.02	Vehículos privados/per capita

Tabla 9. Estadísticas descriptivas de las variables contenidas en la base de datos correspondiente a la línea 3.(N=144).

4.6.3. LINEA 5C1 – Miranda/Plaza Italia – Plaza de los Remedios – General Dávila

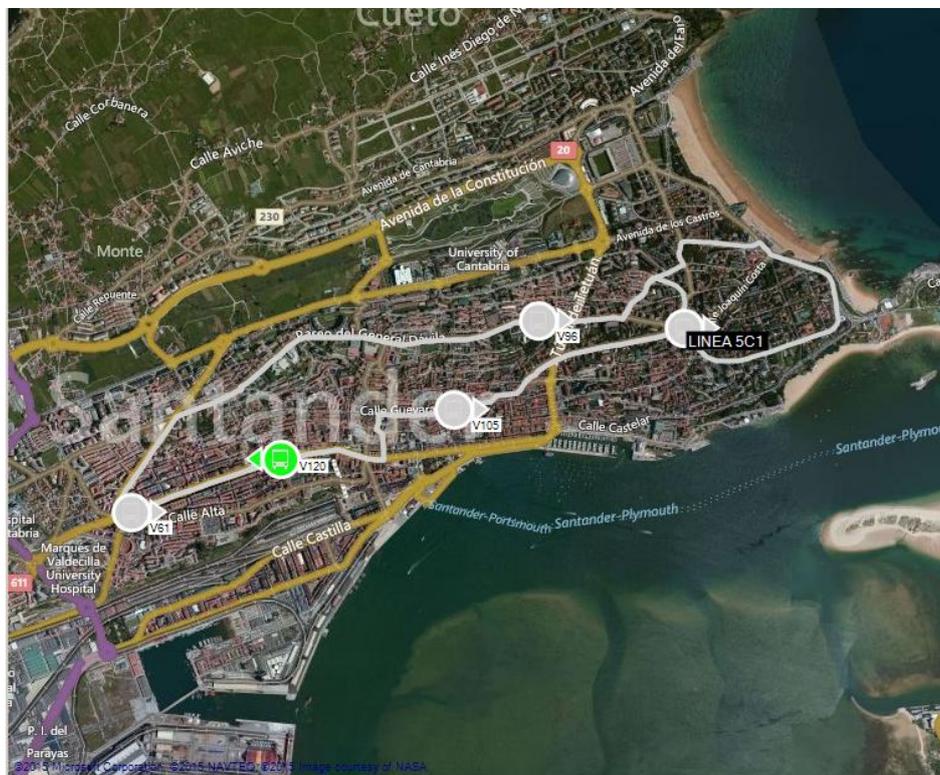


Ilustración 23. Recorrido de la línea 5C1. Fuente: Aplicación cartográfica del (S.A.E).
Sistema de ayuda a la explotación del TUS de Santander.



El Gráfico 24 muestra los datos pertenecientes a la línea 5C1 donde puede observarse la evolución de la demanda de transporte público per cápita en el periodo 2001 – 2012. El promedio de viajeros per cápita para los doce años de la serie es de 7,8. En 2001, primer año de la serie, coincide el valor máximo de demanda per cápita de la serie con guarismos levemente por encima de los 8,5 viajeros por habitante. Entre los años 2002-2003 la media de viajeros por habitante apenas varía en comparación al año 2001 cifrándose 8,1 y 8,3 respectivamente. A partir de 2003 ese primer año se da una marcada tendencia a la baja que llega hasta el año 2004 con un valor ligeramente por debajo de los 7,5 pasajeros per cápita. A partir de 2004 y hasta 2007 el número de viajeros per cápita permanecerá estable en torno a los 7,2-7,5 viajeros per cápita, siendo 2008 el que registró un menor número per cápita de viajeros cayendo por debajo de 7, lo que confiere al año 2008 como el año en el que se registra el mínimo de la serie histórica analizada.

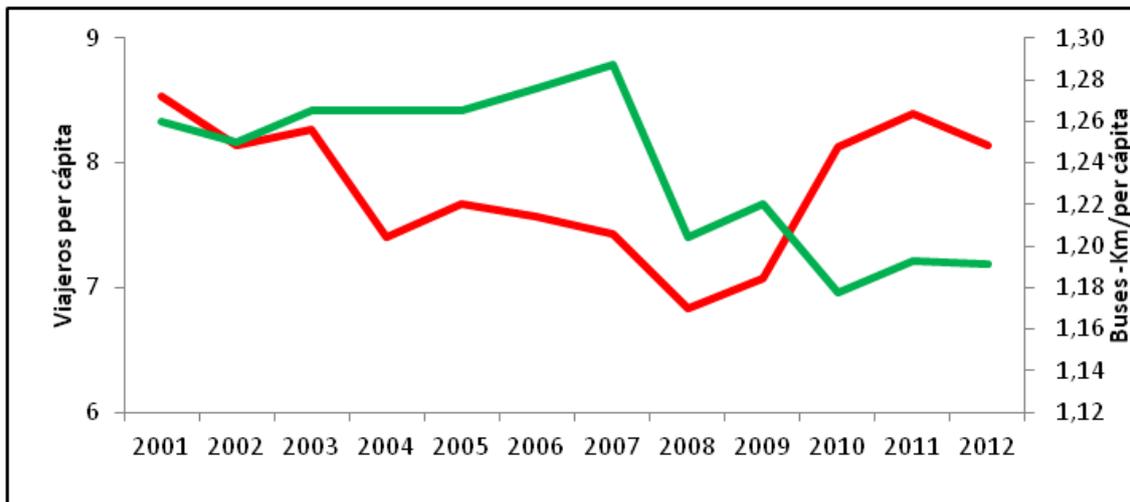


Gráfico 24. Evolución del número medio anual de pasajeros – km per cápita (color rojo, eje izquierdo) y de los buses – kilómetro per cápita (color verde, eje derecho) en la Línea 5C1 de Transportes Urbanos de Santander. Periodo 2001 - 2012.

A partir de 2008 comienza una tendencia progresiva que hará que en los años sucesivos hasta llegar al año 2012 los datos de viajeros per cápita se incrementen pasando a ser de 6,8 pasajeros per cápita en 2008 a los 8,1 de 2010 y 8,4 en 2011, un dato que sitúa a este año como el segundo más alto de la serie con una media de viajes per cápita muy



similar al del año 2001. En el año 2012 la media vuelve a descender pero manteniéndose por encima de la barrera de los 8 pasajeros per cápita.

En cuanto al nivel de servicio medido como vehículos o buses – km/per cápita, se pueden reconocer dos fases a lo largo de la serie. En la primera de esas fases hay una estabilidad de los datos que arrojan de 2001 a 2007 un promedio de 1,27 buses/kilómetro per cápita y en donde el valor máximo tiene lugar en el año 2006 con 1,28. En la segunda fase comprendida entre 2008 y 2012 la gráfica muestra un descenso de la media que para los cuatro años finales de la serie será de 1,20 buses/kilómetro per cápita siendo el año 2010 con 1,18 el que menor nivel de servicio se registra para esta línea.

Para la línea 5C1 se parte, igual que en los casos de las líneas anteriores, de 144 observaciones mensuales para los 12 años considerados. Los estadísticos descriptivos de las variables contenidas en la base de datos pueden consultarse en la Tabla 10.

Variable	Valor Min.	Valor Max.	Media	Std. Desviación	Descripción
TRIPPC	0.34	0.93	0.65	0.08	Viajeros-km/per capita
FARE	0.31	0.50	0.38	0.03	Recaudación (deflactada €/viajes)
VEHKM	0.09	0.11	0.10	0.01	Buses – km/per capita
FUEL	0.67	1.08	0.84	0.09	Precio deflactado del combustible (€)
URATE	4.70	19.20	10.5	3.47	Tasa de Paro
INCOME	15401	19508	17072	1053	Ingresos deflactados per cápita (€)
POP	178659	185231	18.694	1907	Población
VEH	0.40	0.46	0.44	0.02	Vehículos privados/per capita

Tabla 10. Estadísticas descriptivas de las variables contenidas en la base de datos correspondiente a la línea 5C1. (N=144)

4.6.4. LINEA 12 – Carrefour - Canalejas

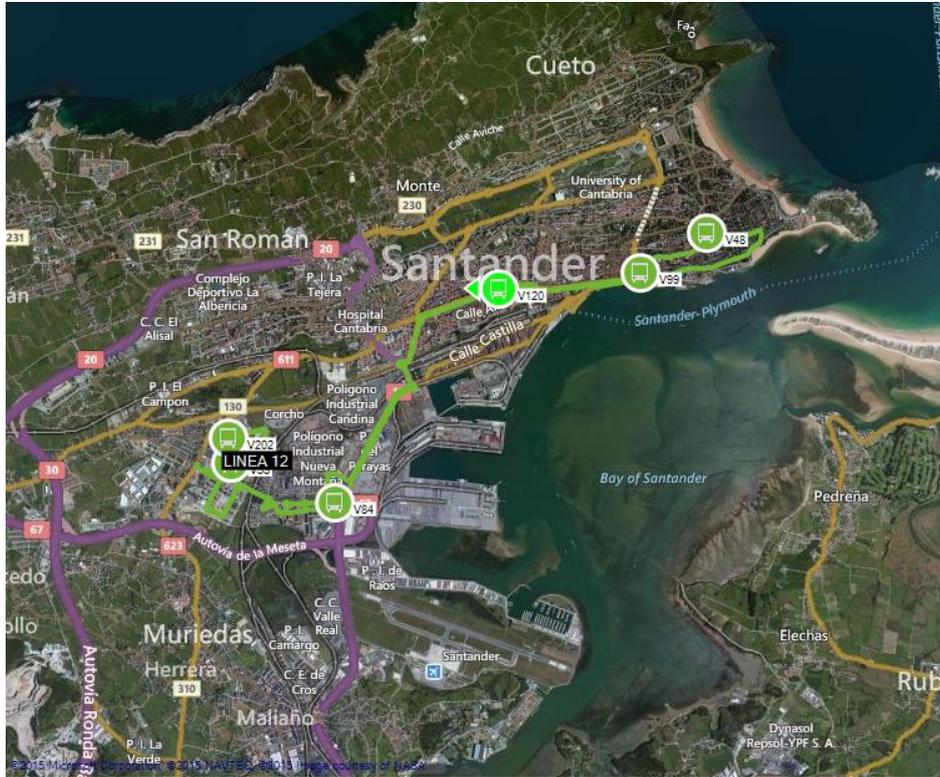


Ilustración 24. Recorrido de la línea 12. Fuente: Aplicación cartográfica del (S.A.E). Sistema de ayuda a la explotación del TUS de Santander.

Para la línea 12 puede observarse la evolución de la demanda de transporte público per cápita en el periodo 2001 – 2012. En primer lugar señalar que la línea 12 es una de las líneas cuya estructura de modificó a partir de Julio de 2008. Inicialmente la línea 12 estaba dividida en dos líneas independientes, una de ellas fue la línea 10 y la otra la 12. A partir de Julio de 2008 la nueva línea 12 surgía de la fusión entra las citadas líneas 10 y 12 (antigua). Por ello a nivel de datos, para el periodo de tiempo comprendido entre los años 2001 y Julio de 2008, se procedió a la suma del número de viajeros de las líneas antiguas 10 y 12 ya que la suma de ambas conformaba la 12 posterior. De esa forma los datos de la línea 12 se toman como si la línea 12 hubiera sido la misma desde el año 2001 homogeneizando así los datos. La demanda de transporte y el nivel de servicio de esta línea quedan expresados en el Gráfico 25.



El promedio de viajeros per cápita para los doce años de la serie es de 5,2. En el bienio inicial 2001-2002 la media queda fijada en una media de 4,3 viajeros per cápita siendo en el año 2002 cuando se registra el dato más bajo de toda la serie con 4,26. A partir del año 2002 el promedio de viajeros per cápita comienza a subir hasta alcanzar los 4,7 de media el bienio siguiente (2003-2004). En el trienio 2005-2008 la tendencia volverá a ser de progresivo crecimiento hasta superar los 5 viajeros de promedio per cápita en el año 2007. En el año 2008 se da un primer descenso acusado del número de viajeros promedio, un descenso marcado por la entrada en vigor de la nueva estructura de la línea 12 que se describía anteriormente y que repercutirá marcando una cifra media de 4,4 viajeros per cápita, unos datos similares a los del bienio 2001-2002 cuando aún coexistían las líneas 10 y 12. A partir del año 2009 las medias de viajeros per cápita se vuelven a recuperar de forma generalizada situándose cercanas a los 6,5 viajeros de promedio per cápita. Los años 2010, 2011 y 2012 son los años con una media mayor de viajeros per cápita con 6,4, 6,5 y 6,4 respectivamente.

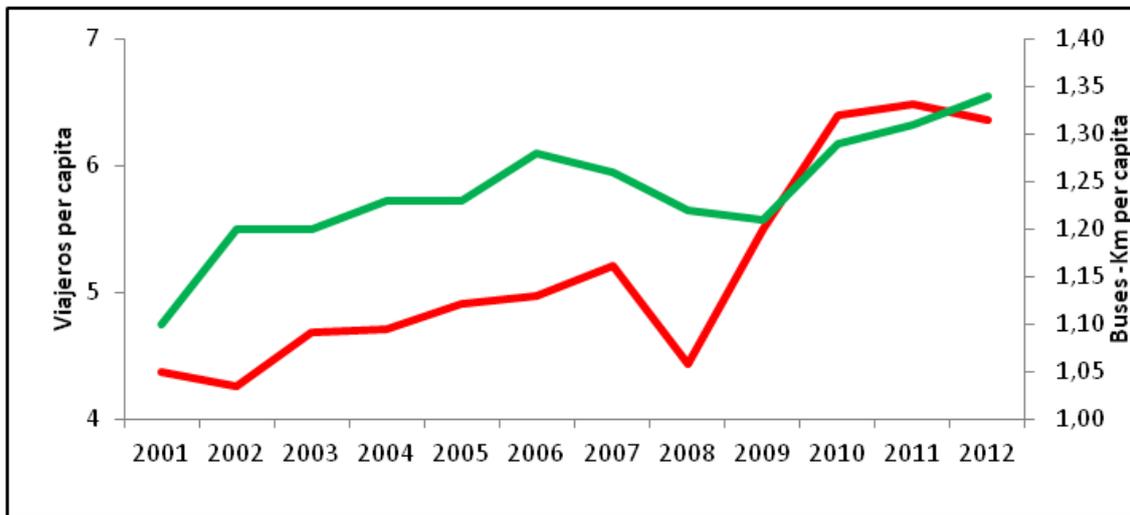


Gráfico 25. Evolución del número medio anual de pasajeros – km per cápita (color rojo, eje izquierdo) y de los buses – kilómetro per cápita (color verde, eje derecho) en la Línea 12 de Transportes Urbanos de Santander. Periodo 2001 - 2012.

En cuanto al nivel de servicio medido como vehículos o buses – km/per cápita, los datos a lo largo de la serie se comportan también de una forma regular aunque se puede señalar tres partes. En el año 2001 se muestra la cifra más baja con 1,1 buses-km/per



cápita y a partir de ahí comienza un periodo comprendido entre 2002 y 2009 en donde las cifras se estabilizan en torno a 1,23 y en donde el valor máximo dentro de ese periodo se da en el año 2006 con 1,28. Otra de las partes a señalar es la conformada por el bienio 2008-2009 donde las cifras vuelven a descender levemente hasta los 1,21. Los últimos tres años de la serie son los que registran un mayor nivel de servicio, concentrándose ahí los mayores niveles de servicio de esta línea 12. Para esa última fase se llega a superar la cifra de 1,3 buses-km/per cápita siendo 2012 el año con el guarismo más alto con 1,34.

Para la línea 12 se parte, igual que en los casos de las líneas anteriores, de 144 observaciones mensuales para los 12 años considerados. Los estadísticos descriptivos de las variables contenidas en la base de datos pueden consultarse en la Tabla 11.

Variable	Valor Min.	Valor Max.	Media	Std. Desviación	Descripción
TRIPPC	0.31	0.59	0.43	0.07	Viajeros-km/per capita
FARE	0.26	0.48	0.35	0.04	Recaudación (deflactada recaudación €/viajes)
VEHKM	0.06	0.11	0.10	0.01	Buses – km/per capita
FUEL	0.67	1.08	0.84	0.09	Precio deflactado del combustible (€)
URATE	4.70	19.20	10.5	3.47	Tasa de Paro
INCOME	15401	19508	17072	1053	Ingresos deflactados per cápita (€)
POP	178659	185231	182694	1907	Población
VEH	0.40	0.46	0.44	0.02	Vehículos privados/per capita

Tabla 11. Estadísticas descriptivas de las variables contenidas en la base de datos correspondiente a la línea 12.(N=144)



CAPÍTULO 5

Análisis y discusión de resultados



5. ANÁLISIS Y DISCUSIÓN DE RESULTADOS

5.1. Metodología y aplicación (Modelo teórico)

El modelo básico empleado presenta una forma funcional de tipo log – log tal como:

$$\ln Y_t = \alpha_0 + \sum_i \alpha_{i1} \ln X_{it} + \sum_{j=1}^{11} \alpha_{j2} D_{jt} + \alpha_3 T_t + \varepsilon_t \quad (79)$$

Donde Y_t es la variable dependiente correspondiente a la demanda en el tiempo t , X_{it} corresponden a las variables independientes o explicativas i en el tiempo t (Tasa de paro, Renta Per Cápita...etc...), D_{jt} son las $j-1$ variables dummy mensuales, i , T_t corresponde a la variable de tendencia, α_0 es un término constante, α_{i1} son los parámetros a estimar para cada una de las variables independientes i , α_{j2} son los parámetros a estimar para las $j-1$ variables dummy mensuales, α_3 es el parámetro de la tendencia, ε_t es el término de error independiente e idénticamente distribuido (IID).

En la ecuación 79 los parámetros α_{i1} se pueden interpretar directamente como elasticidades de equilibrio. Las point elasticities se definen como:

$$E_{x_i}^{Point} = \frac{Limit}{\Delta x_i \rightarrow 0} \left(\frac{\Delta y}{y} / \frac{\Delta x_i}{x_i} \right) = \frac{x_i}{y} \left(\frac{\partial y}{\partial x_i} \right) \quad (80)$$

Si bien dada la forma funcional log – log la point elasticity es equivalente a la de arco y constante en toda la función de demanda.

Para estimar los impactos a largo plazo se debe incluir el siguiente término adicional en la ecuación (81) : $\alpha_4 \ln(y_{t-1})$. De esta forma las elasticidades a largo plazo se pueden



calcular mediante $\alpha_{i1} / (1 - \alpha_4)$. Esta especificación dinámica es recomendada por autores como Goodwin (1992).

Dado que el modelo se estima con datos de una serie temporal, la especificación incluye once variables dummy D_{jt} cuyos parámetros recogen los posibles efectos estacionales de la demanda de transporte. Adicionalmente el parámetro de la variable T_t permite capturar la tendencia general de la serie.

El modelo dado puede presentar autocorrelación serial en el término de error ε_t . Si se da este caso, el cual puede determinarse mediante el “test d” de Durbin – Watson (Durbin and Watson 1950), los parámetros estimados estarán insesgados pero serán ineficientes. Para evitar este problema puede descomponerse el término de error en dos partes:

$$\varepsilon_t = \rho\varepsilon_{t-1} + u_t \tag{82}$$

Donde ρ es un parámetro que mide la autocorrelación del residuo t con el residuo de la observación $t - 1$ y u_t son residuos IID. Para realizar la estimación del modelo con autocorrelación puede recurrirse a distintos procedimientos como el algoritmo de Prais – Winsten que permite estimar ρ de forma iterativa. Este método es útil en muestras pequeñas ya que no requiere eliminar la primera observación a diferencia de otros métodos disponibles como el de Cochrane–Orcutt (Prais and Winsten 1954).

La influencia recíproca de la demanda y el nivel de servicio ofertado crea un problema de endogeneidad que viola la hipótesis de independencia entre las variables independientes y los residuos del modelo (Gujarati and Porter 2009). Esta endogeneidad puede testarse mediante el test de Wu (Wu 1973). La estimación de los modelos con endogeneidad puede realizarse mediante mínimos cuadrados en dos etapas (2SLS). Este procedimiento aporta parámetros consistentes a diferencia de la estimación directa



mediante mínimos cuadrados ordinarios (OLS). La dificultad de aplicación de 2SLS se deriva de la dificultad de encontrar una o varias variables instrumentales que correlacionando con la variable independiente de interés (nivel de servicio) a la vez que no correlacionen con el término de error del modelo.

Dado que se está trabajando con datos de series temporales, para utilizar el modelo en modo predictivo, hay que comprobar su estacionariedad débil, es decir, la estabilidad en la media y la varianza así como la independencia de la covarianza de t en el proceso estocástico. Esto puede realizarse mediante un test de raíz unitaria como el Augmented de Dickey-Fuller (ADF) que permite además la existencia de correlación en los residuos (Greene 2003).

Este método se aplicó a los datos de series temporales que considerando de forma agrupada la demanda de viajeros, sin distinguir entre las diferentes líneas del servicio existentes, y posteriormente se aplicó el modelo resultante al resto de líneas de forma individual para verificar los valores predictivos de las mismas y especificar cuáles habían obtenido un mejor resultado tras llevar a cabo la respectiva validación.

5.2. Resultados y discusión

5.2.1. Resultados del modelo agregado de líneas

La Tabla 12 presenta los parámetros estimados utilizando el método de Mínimos Cuadrados Ordinarios u OLS y de Mínimos Cuadrados Ordinarios en dos etapas o 2SLS. La hipótesis nula de exogeneidad de la variable VEHKM (Vehículos/Km per cápita) se testeó mediante el test de Wu rechazándose ésta con un valor F de 39.53 (p – valor 0.000). Los modelos con la denominación final AR(1) tienen en cuenta además la autocorrelación de orden uno en los residuos y han sido estimados mediante el método de Prais – Winsten. Debido a la importante correlación existente entre URATE e INCOME (R^2 cercana a 0.7) se ha diferenciado entre los modelos que consideran únicamente



URATE como variable independiente (MLR1) y los modelos que consideran como variable independiente INCOME (MLR2). En la Tabla 12 bajo los parámetros estimados pueden consultarse entre paréntesis los p – valor de los test t.

Variables	MLR1- OLS	MLR2- OLS	MLR1- AR(1)	MLR2- AR(1)	MLR1- 2SLS	MLR2- 2SLS	MLR1- 2SLS AR(1)	MLR2- 2SLS AR(1)
(Constant)	1.880 (.000)	7.088 (.000)	1.262 (.000)	3.654 (.062)	2.059 (.000)	6.469 (.000)	1.355 (.002)	6.363 (.019)
FARE	-0.059 (.618)	-0.081 (.485)	-0.197 (.006)	-0.198 (.007)	-0.040 (.730)	-0.062 (.596)	-0.133 (.088)	-0.117 (.152)
VEHKM	-0.205 (.277)	-0.074 (.710)	0.478 (.003)	0.497 (.003)	-0.465 (.626)	-0.410 (.638)	0.469 (.538)	0.810 (.257)
FUEL	0.523 (.000)	0.397 (.000)	0.190 (.147)	0.158 (.249)	0.558 (.000)	0.462 (.002)	0.272 (.048)	0.166 (.315)
VEH	-	-	-	-	-	-	-	-
URATE	0.100 (.000)	-	0.125 (.008)	-	0.090 (.051)	-	0.133 (.013)	-
INCOME	-	-0.527 (.000)	-	-0.222 (.272)	-	-0.442 (.080)	-	-0.505 (.098)
T	-0.001 (.000)	-0.001 (.003)	-0.001 (.011)	-0.001 (.179)	-0.001 (.000)	-0.001 (.000)	-0.001 (.004)	-0.001 (.236)
DJAN	-	-	-	-	-	-	-	-
DFEB	0.017 (.551)	0.010 (.747)	0.038 (.059)	0.044 (.038)	-0.045 (.568)	-0.030 (.070)	-0.020 (.751)	0.059 (.372)
DMAR	0.089 (.000)	0.102 (.000)	0.098 (.000)	0.103 (.000)	0.083 (.000)	0.092 (.000)	0.058 (.003)	0.061 (.000)
DAPR	0.027 (.259)	0.036 (.128)	0.061 (.005)	0.062 (.006)	-	-	-	-
DMAY	0.130 (.000)	0.132 (.000)	0.139 (.000)	0.138 (.000)	0.126 (.000)	0.126 (.000)	0.091 (.001)	0.079 (.000)
DJUN	0.144 (.000)	0.148 (.000)	0.181 (.000)	0.178 (.000)	0.131 (.000)	0.131 (.000)	0.138 (.000)	0.134 (.000)
DJUL	0.270 (.000)	0.271 (.000)	0.287 (.000)	0.282 (.000)	0.269 (.000)	0.270 (.000)	0.244 (.000)	0.223 (.000)
DAUG	0.236 (.000)	0.234 (.000)	0.256 (.000)	0.249 (.000)	0.235 (.000)	0.233 (.000)	0.218 (.000)	0.194 (.000)
DSEP	0.084 (.000)	0.077 (.000)	0.109 (.000)	0.099 (.000)	0.078 (.000)	0.070 (.013)	0.080 (.002)	0.061 (.026)
DOCT	0.152 (.000)	0.141 (.000)	0.161 (.000)	0.156 (.000)	0.147 (.000)	0.136 (.000)	0.144 (.000)	0.128 (.000)
DNOV	0.095 (.000)	0.085 (.000)	0.114 (.000)	0.111 (.000)	0.083 (.002)	0.071 (.002)	0.106 (.000)	0.103 (.000)
DDIC	-	-	-	-	-	-	-	-



ρ	-	-	0.748 (.000)	0.761 (.000)	-	-	0.696 (.000)	0.670 (.000)
R2	.745	.748	.745	.748	.740	.739	.740	.739
R2adj	.716	.718	.716	.718	.712	.711	.712	.711
F	25.5 (.000)	25.3 (.000)	25.5 (.000)	25.3 (.000)	26.2 (.000)	26.1 (.000)	26.2 (.000)	26.1 (.000)
Durbin - Watson	0.787	0.808	2.353	2.393	0.881	0.910	2.291	2.277

Tabla 12. Parámetros estimados de los modelos de demanda. Modelos teniendo en cuenta los datos agregados de las líneas objeto de estudio (ρ valor del test t aparece entre paréntesis, N=144).

Comparando los ajustes de los modelos MLR1 y MLR2 considerando respectivamente URATE o INCOME, estos fueron similares tanto en los modelos estimados mediante OLS como en los modelos estimados mediante 2SLS. El ajuste de los modelos estimados mediante OLS fue superior al de los 2SLS, algo esperado dado que la estimación en dos etapas depende de la capacidad predictiva de las variables instrumentales sobre la variable endógena.

Esta característica de la estimación en dos etapas puede afectar también a los errores estándar de los parámetros ya que ésta proporciona parámetros consistentes pero no eficientes (Gujarati and Porter 2009). El test de Durbin – Watson señala como en los modelos con estimación mediante OLS o 2SLS hay un grado considerable de autocorrelación en los residuos, ampliamente reducida en los modelos con autocorrelación de orden 1 con valores mucho más cercanos a 2 (ausencia de autocorrelación).

El parámetro de la variable referente a la tarifa (FARE) presentó el signo negativo esperado y elasticidades de equilibrio cercanas al -0.2 en los modelos estimados mediante OLS considerando autocorrelación en los residuos (MLR1 – AR(1) y MLR2 – AR(2)) y cercanas a -0.1 en los modelos estimados mediante 2SLS (MLR1 – 2SLS AR(1) y MLR2 – 2SLS AR(2)). El parámetro de la variable VEHKM también presentó el signo



esperado, en este caso positivo, y valores cercanos a 0,5 en la estimación mediante OLS considerando autocorrelación en los residuos e inferiores y cercanas a 0.5 y 0.8 en la estimación mediante 2SLS. La diferencia en la magnitud del parámetro en la estimación en dos etapas es un efecto esperado, ya que al existir endogeneidad y correlación positiva entre VEHKM y los residuos de la regresión, el parámetro en la estimación mediante OLS puede estar sesgado (Hayashi 2000).

Los parámetros estimados para la variable FUEL no en todos los casos fueron significativos aunque en general las estimaciones se situaron entre 1.5 y 3 en los modelos considerando autocorrelación. El signo positivo obtenido es consistente con las expectativas ya que un incremento en el precio de los carburantes implica un aumento de la demanda de transporte público. La variable VEH en cambio no fue significativa en ninguna de las estimaciones y se desechó de los modelos para evitar problemas de inclusión de variable irrelevante que pudieran sesgar el resto de parámetros. Sin embargo, la variable VEH tiene una correlación moderada con la variable de ingreso ($r = 0.43$) por lo que parte de su influencia puede ser capturada por este último. Por lo tanto, el efecto negativo de los ingresos (INCOME) (o positivo de URATE – Tasa de Paro) en la demanda de transporte también podrían estar asociados al estancamiento en la propiedad de automóviles, una variable que es bien conocida por afectar al número de viajes en autobús urbano significativamente (Holmgren 2007).

Los parámetros de las dos variables principales para los objetivos de este estudio, URATE e INCOME, mostraron un signo positivo en el caso de URATE y negativo en el caso de INCOME en todas las especificaciones. Esto proporciona evidencia adicional respecto al carácter de bien inferior del transporte público ya apuntado por otros autores (Holmgren 2007). En el caso de los bienes inferiores, los incrementos de renta de los usuarios no dan lugar a incrementos en la demanda como el de los bienes normales, sino a reducciones en el consumo (Varian and Repcheck 2010). La significatividad de las variables URATE e INCOME fue cambiante en las distintas especificaciones. En los modelos MLR1-AR(1) y MLR1-2SLS AR(1), los parámetros estimados de URATE fueron significativos a un nivel de confianza del 95% presentando valores entre 0.125 y 0.133



de incremento en los pasajeros – km per cápita por cada incremento del 1 por ciento en la tasa de paro.

El resto de parámetros estimados corresponden a la tendencia, la cual fue o bien ligeramente negativa (-0.001) o bien no significativamente distinta de cero. Las dummy mensuales en su mayor parte fueron positivas y significativas y especialmente elevadas durante los meses estivales de Julio y Agosto cuando hay una mayor actividad comercial y turística en la ciudad debido a que Santander, área de estudio, es uno de los principales lugares de destino del norte de España en relación a su modelo turístico de sol y playa además de otros importantes reclamos culturales.

La estimación mediante 2SLS (Mínimos Cuadrados en dos etapas) se realizó recurriendo al número de pernoctaciones realizadas en la ciudad como variable instrumental. Esta variable presenta una correlación positiva con VEHKM y en una primera etapa fue significativa al 95% del nivel de confianza presentando la regresión un valor del test F de 83.26 y un R^2 de 0.37. Estos valores apuntan a que la estimación en dos etapas posee una calidad suficiente si se siguen las recomendaciones de Bound, Jaeger et al. (1995).

Las elasticidades de equilibrio obtenidas mediante el modelo estático pueden compararse con las elasticidades a corto y largo plazo obtenidas en la especificación dinámica (véase Tabla 13).

En todos los modelos habían aumentado las elasticidades estimadas si comparamos las estimaciones a corto plazo con las estimaciones a largo plazo. Este aumento de las elasticidades a largo plazo, especialmente importante en el caso de la variable VEHKM, es más consistente con los procesos de cambio que conllevan un periodo de adaptación por parte de los usuarios. Las elasticidades pueden aumentar progresivamente por varios efectos como el procesado de la nueva información por los usuarios, la ruptura de los patrones de conducta (hábitos) o el cambio en las restricciones de las elecciones



de transporte provocadas por el cambio en elecciones de largo plazo como aquellas que tienen que ver con el lugar de residencia o de trabajo (Balcombe, Mackett et al. 2004).

Variables	MLR1- AR(1)		MLR2- AR(1)		MLR1- 2SLS AR(1)		MLR2- 2SLS AR(1)	
	SR	LR	SR	LR	SR	LR	SR	LR
FARE	-0.197	-0.259	-0.198	-0.332	-0.145	-0.196	-0.126	-0.159
VEHKM	0.348	0.456	0.238	0.400	1.519	2.059	1.708	2.159
FUEL	0.140	0.183	0.072	0.121	0.080	0.109	0.030	0.040
URATE	0.109	0.143	-	-	0.155	0.210	-	-
INCOME	-	-	-0.193	-0.326	-	-	-0.681	-0.861
R ²	.750		.752		.777		.778	
R ² _{adj}	.718		.721		.751		.752	
F	23.6		23.9		29.5		29.7	
	(.000)		(.000)		(.000)		(.000)	

Tabla 13. Cálculo de las elasticidades de equilibrio para las variables estructurales del estudio. Estimaciones a corto plazo (SR) y a largo plazo (LR).

Comparando los resultados obtenidos con las elasticidades calculadas para la variable FARE en otros estudios como el de De Rus (1990) para España, Paulley, Balcombe et al. (2006) para el Reino Unido o Bresson, Dargay et al. (2004) para Francia que, generalmente, han estimado elasticidades de equilibrio en el rango -0.3 a -0.4, éstas estuvieron ligeramente por debajo especialmente en la estimación mediante 2SLS. Esto pudo deberse a la poca variabilidad de las tarifas en el periodo considerado con decrementos en el precio deflactado desde los 0.43 euros por viajero al inicio de la serie hasta los 0.35 de Diciembre de 2012.

Para la variable VEHKM, referente al nivel de servicio, los estudios disponibles han obtenido resultados más variados, desde la elasticidad a largo plazo de 0.68 obtenida por Bresson, Dargay et al. (2004) para Francia pasado por el 0.71 obtenido por De Rus (1990) para España hasta el 1.38 obtenido por Holmgren (2007) en su meta – análisis internacional. Los resultados calculados en los modelos de equilibrio se encontraron por lo tanto más cerca de los de Bresson, Dargay et al. (2004) y De Rus (1990).



Las estimaciones obtenidas en la variable INCOME fueron en general no significativas especialmente en el caso de la estimación del modelo MLR2 – AR(1). Bresson, Dargay et al. (2004) estimaron una elasticidad de -0.23, similar aunque algo inferior a la obtenida en la estimación de equilibrio mediante 2SLS – AR(1). Por el contrario Holmgren (2007) en cambio obtuvo de la recopilación de estudios un parámetro ligeramente superior de -0.62 de su evaluación de otras investigaciones que era similar, al igual que Paulley, Balcombe et al. (2006) que estimaron elasticidades a largo plazo también con signo negativo en un rango entre -0.5 y -1.0 a largo plazo. Todos los estudios coinciden por lo tanto en la característica de bien inferior del transporte público, algo apoyado en el presente estudio por las estimaciones realizadas con el signo positivo de la variable URATE.

Las estimaciones obtenidas en la variable FUEL (combustible-gasolina) pueden compararse con las obtenidas por Holmgren (2007) de 0.73, bastante superiores a las elasticidades a largo plazo obtenidas por Bresson, Dargay et al. (2004) (0.14) que son más similares a las obtenidas en el presente estudio y a las estimadas por Goodwin (1992) a partir de 120 estudios previos (0.34).

5.2.2. Resultados del modelo desagregado por líneas

De la misma forma en la que se han expuesto los resultados del modelo con datos agregados de las líneas, en este epígrafe se citarán aquellas líneas cuyos datos se ajustaron mejor al modelo generado y que por tanto contaban con unos valores predictivos idóneos. Las líneas más significativas han sido la 1, la 3, la 5C1 y la 12. A continuación se muestran los resultados para esas líneas que han obtenido una mayor significatividad. Al igual que en el caso del modelo agregado se tiene en cuenta la hipótesis nula de exogeneidad de la variable VEHKM testeándose mediante el test de Wu y rechazándose ésta con un valor F de 39.53 (p – valor 0.000).

5.2.2.1. LINEA 1. Parque Tecnológico (PCTCAN) – Adarzo –Arsenio Odriozola

A continuación aparecen expuestos los parámetros estimados utilizando OLS (Mínimos Cuadrados Ordinarios) y 2SLS (Mínimos Cuadrados en dos etapas) en esta ocasión considerando la línea 1.

Los modelos con la denominación final AR(1) tienen en cuenta la autocorrelación de orden uno en los residuos y se estimaron mediante el método de Prais – Winsten. Debido a la importante correlación existente entre URATE e INCOME (R2 cercana a 0.8) se ha diferenciado entre los modelos que consideran únicamente URATE como variable independiente (MLR1) y los modelos que consideran como variable independiente INCOME (MLR2) véase Tabla 14 .

Variables	MLR1- OLS	MLR2- OLS	MLR1- AR(1)	MLR2- AR(1)	MLR1- 2SLS	MLR2- 2SLS	MLR1- 2SLS AR(1)	MLR2- 2SLS AR(1)
(Constant)	1.494 (.000)	8.857 (.000)	0.387 (.355)	-2.004 (.669)	1.724 (.004)	5.878 (.293)	0.326 (.539)	8.117 (.411)
FARE	-0.078 (.308)	-0.078 (.303)	0.014 (.674)	0.013 (.670)	-0.083 (.260)	-0.082 (.267)	0.014 (.674)	0.016 (.627)
VEHKM	0.959 (.000)	0.887 (.000)	0.142 (.145)	0.146 (.144)	1.115 (.007)	1.096 (.012)	0.099 (.666)	0.066 (.790)
FUEL	0.344 (.100)	0.255 (.212)	0.006 (.983)	-0.022 (.937)	0.205 (.639)	0.134 (.705)	-0.181 (.587)	-0.306 (.369)
VEH	-	-	-	-	-	-	-	-
URATE	0.104 (.028)	-	0.145 (.372)	-	0.069 (.501)	-	0.123 (.480)	-
INCOME	-	-0.747 (.006)	-	0.277 (.567)	-	-0.417 (.500)	-	-0.783 (.448)
T	-0.001 (.016)	-0.001 (.057)	-0.003 (.114)	-0.003 (.229)	-0.0008 (.691)	-0.000 (.782)	-0.003 (.220)	-0.002 (.272)
DJAN	-	-	-	-	-	-	-	-
DFEB	0.110	0.122	0.010	0.014	0.124	0.131	0.015	0.016



	(.049)	(.027)	(.700)	(.614)	(.072)	(.032)	(.687)	(.721)
DMAR	0.081	0.092	0.045	0.053	0.084	0.091	0.055	0.057
	(.141)	(.090)	(.203)	(.124)	(.118)	(.072)	(.020)	(.010)
DAPR	0.016	0.016	-0.011	-0.004	-	-	-	-
	(.771)	(.768)	(.776)	(.908)				
DMAY	0.048	0.050	0.094	0.098	0.042	0.044	0.108	0.107
	(.409)	(.383)	(.038)	(.030)	(.421)	(.413)	(.000)	(.000)
DJUN	0.146	0.142	0.148	0.150	0.145	0.147	0.159	0.158
	(.009)	(.010)	(.001)	(.001)	(.005)	(.011)	(.000)	(.000)
DJUL	0.171	0.169	0.182	0.177	0.169	0.170	0.195	0.187
	(.003)	(.002)	(.0002)	(.000)	(.001)	(.002)	(.000)	(.000)
DAUG	0.129	0.122	0.147	0.139	0.125	0.123	0.158	0.148
	(.022)	(.028)	(.002)	(.003)	(.015)	(.022)	(.000)	(.000)
DSEP	0.030	0.016	0.035	0.025	0.031	0.026	0.425	0.034
	(.581)	(.765)	(.434)	(.569)	(.557)	(.683)	(.310)	(.430)
DOCT	0.109	0.094	0.138	0.134	0.101	0.094	0.141	0.140
	(.048)	(.082)	(.0001)	(.000)	(.003)	(.072)	(.000)	(.000)
DNOV	0.056	0.040	0.077	0.076	0.045	0.038	0.074	0.077
	(.301)	(.459)	(.003)	(.003)	(.377)	(.462)	(.005)	(.002)
DDIC	-	-	-	-	-	-	-	-
ρ	-	-	0.936	0.947	-	-	0.943	0.950
			(.000)	(.000)			(.000)	(.000)
R2	.771	.777	.771	.777	.760	.761	.760	.761
R2adj	.745	.750	.745	.750	.734	.735	.734	.735
F	28.8	29.7	28.8	29.7	29.2	29.4	29.2	29.4
	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)
Durbin - Watson	0.585	0.599	2.175	2.196	0.646	0.658	2.159	2.202

Tabla 14. Parámetros estimados de los modelos de demanda. Modelos teniendo en cuenta los datos de la línea 1 (ρ valor del test t aparece entre paréntesis, N=144).

Comparando los ajustes de los modelos MLR1 y MLR2 considerando respectivamente URATE o INCOME, estos fueron también fueron similares tanto en los modelos estimados mediante OLS como en los modelos estimados mediante 2SLS. El ajuste de los modelos estimados mediante OLS fue superior al de los 2SLS, algo que ya se ha citado para el caso del resultado de los modelos agregados, debido a que la estimación en dos etapas depende de la capacidad predictiva de las variables instrumentales sobre la variable endógena.



El test de Durbin-Watson muestra que en los modelos cuya estimación se ha realizado mediante mínimos cuadrados ordinarios (OLS) o mínimos cuadrados en dos etapas (2SLS) existe un cierto grado de autocorrelación con valores comprendidos entre 0.58 y 0.60 para los modelos estimados mediante OLS y 0.65-0.66 para los estimados utilizando 2SLS. Teniendo en cuenta los modelos con autocorrelación de orden 1 los valores se acercan mucho más a 2, lo que equivale a no presencia de autocorrelación.

El parámetro (FARE) presentó signo negativo pero sólo en el caso de los modelos estimados mediante OLS y 2SLS mientras que para los modelos estimados con autocorrelación de orden 1 fueron positivos. Las elasticidades de equilibrio fueron cercanas a 0.01 en los modelos estimados mediante OLS considerando autocorrelación en los residuos (MLR1-AR(1) y (MLR2-AR(1)). Para los modelos estimados a partir de 2SLS las elasticidades de equilibrio se aproximaron también a 0.001.

El parámetro de la variable VEHKM igualmente mostró el signo positivo esperado con valores cercanos a 0.15 en la estimación mediante OLS considerando autocorrelación en los residuos y cercanas a 0.001 en la estimación mediante 2SLS.

La variable VEH en cambio no fue significativa tampoco en ninguna de las estimaciones halladas para la línea 1 y por tanto se desechó también de los modelos para evitar la influencia por parte de una variable irrelevante que pudieran alterar el resto de parámetros.

Los parámetros de las variables, URATE (Tasa de Paro) e INCOME (Ingresos deflactados per cápita), mostraron signos positivos para el caso de de la primera y negativos en el caso de la segunda excepto para el modelo estimado en MLR2 AR(1) que mostró signo positivo cercano a 0.3.

5.2.2.2. LINEA 3. Ojáiz – Paseo Pereda –Rectorado (Universidad)

A continuación aparecen expuestos los parámetros estimados utilizando OLS (Mínimos Cuadrados Ordinarios) y 2SLS (Mínimos Cuadrados en dos etapas) en esta ocasión considerando la línea 3 (véase Tabla 15).

Variables	MLR1- OLS	MLR2- OLS	MLR1- AR(1)	MLR2- AR(1)	MLR1- 2SLS	MLR2- 2SLS	MLR1- 2SLS AR(1)	MLR2- 2SLS AR(1)
(Constant)	-0.831 (.000)	0.593 (.741)	-0.809 (.001)	-0.331 (.895)	-1.085 (.161)	-2.394 (.628)	-4.414 (.560)	2.427 (.637)
FARE	-0.659 (.001)	-0.563 (.007)	-0.492 (.039)	-0.506 (.040)	-0.579 (.005)	-0.612 (.005)	-0.500 (.043)	-0.430 (.118)
VEHKM	0.104 (.210)	0.125 (.139)	0.092 (.150)	0.090 (.162)	-0.132 (.820)	-0.135 (.819)	0.331 (.517)	0.331 (.522)
FUEL	0.536 (.000)	0.500 (.0001)	0.411 (.020)	0.407 (.023)	0.701 (.048)	0.727 (.090)	0.353 (.286)	0.301 (.444)
VEH		-	-	-	-	-	-	-
URATE	-0.001 (.979)	-	0.021 (.685)	-	-0.020 (.757)	-	0.044 (.534)	-
INCOME	-	-0.135 (.426)	-	-0.046 (.851)	-	0.128 (.758)	-	-0.278 (.540)
T	-0.003 (.000)	-0.002 (.000)	-0.002 (.000)	-0.002 (.000)	-0.003 (.000)	-0.003 (.003)	-0.002 (.0006)	-0.002 (.053)
DJAN	-			-	-	-	-	-
DFEB	-0.023 (.477)	-0.018 (.582)	-0.022 (.368)	-0.022 (.391)	-0.058 (.265)	-0.061 (.309)	-0.022 (.612)	-0.014 (.781)
DMAR	0.118 (.000)	0.123 (.000)	0.125 (.000)	0.126 (.000)	0.087 (.061)	0.082 (.116)	0.102 (.003)	0.108 (.008)
DAPR	0.051 (.117)	0.051 (.118)	0.059 (.078)	0.059 (.075)	-	-	-	-
DMAY	0.148 (.000)	0.148 (.000)	0.160 (.000)	0.160 (.000)	0.132 (.000)	0.133 (.000)	0.109 (.001)	0.109 (.000)
DJUN	0.210 (.000)	0.206 (.000)	0.217 (.000)	0.216 (.000)	0.182 (.000)	0.183 (.000)	0.184 (.000)	0.182 (.000)
DJUL	0.345 (.000)	0.333 (.000)	0.340 (.000)	0.341 (.000)	0.318 (.000)	0.322 (.000)	0.304 (.000)	0.297 (.000)
DAUG	0.303 (.000)	0.289 (.000)	0.295 (.000)	0.296 (.000)	0.275 (.000)	0.280 (.000)	0.263 (.000)	0.253 (.000)



DSEP	0.096 (.009)	0.089 (.018)	0.100 (.011)	0.099 (.014)	0.069 (.052)	0.073 (.034)	0.076 (.040)	0.067 (.074)
DOCT	0.164 (.000)	0.163 (.000)	0.178 (.000)	0.177 (.000)	0.150 (.000)	0.153 (.000)	0.154 (.000)	0.150 (.000)
DNOV	0.070 (.036)	0.073 (.028)	0.086 (.000)	0.085 (.001)	0.053 (.092)	0.054 (.078)	0.078 (.003)	0.075 (.003)
DDIC	-	-	-	-	-	-	-	-
ρ	-	-	0.575 (.000)	0.570 (.000)	-	-	0.546 (.000)	0.531 (.000)
R2	.685	.687	.685	.687	.661	.658	.661	.658
R2adj	.648	.650	.648	.650	.624	.621	.624	.621
F	18.6 (.000)	18.7 (.000)	18.6 (.000)	18.7 (.000)	18.0 (.000)	17.8 (.000)	18.0 (.000)	17.8 (.000)
Durbin - Watson	0.885	0.892	2.119	2.119	0.927	0.918	2.126	2.131

Tabla 15. Parámetros estimados de los modelos de demanda. Modelos teniendo en cuenta los datos de la línea 3 (ρ valor del test t aparece entre paréntesis, N=144).

En el caso de la línea 3 las variables de interés (URATE e INCOME) no fueron significativas, con valores aún inferiores que los obtenidos para el caso de datos agregados de las líneas. La variable URATE (Tasa de paro) mostró signo positivo para los modelos estimados a partir de MLR1 AR(1) y MLR1 2SLS AR(1) con valores de 0.02 y 0.04 respectivamente. Para los modelos MLR1 OLS y MLR1-2SLS en cambio el signo fue negativo con valores, siendo el segundo el de mayor significatividad con -0.02.

Para la variable INCOME los valores tampoco fueron significativos y los signos fueron para todos los casos negativos, igual que en los modelos estimados para las líneas agregadas, a excepción del modelo estimado por MLR2-2SLS que para la línea 3 fue de signo positivo.

En el caso de los ajustes de los modelos los valores fueron todos ellos muy similares oscilando entre 0.65 y 0.68 lo que supone que los modelos estimados para la línea 3 ajusten sensiblemente peor que los obtenidos en el modelo agregado de líneas.

5.2.2.3. LINEA 5C1. Miranda/Plaza de Italia – Plaza de los Remedios- General Dávila.

A continuación se pueden ver los parámetros estimados utilizando OLS (Mínimos Cuadrados Ordinarios) y 2SLS (Mínimos Cuadrados en dos etapas) en esta ocasión considerando la línea 3 (véase Tabla 16).

VARIABLES	MLR1- OLS	MLR2- OLS	MLR1- AR(1)	MLR2- AR(1)	MLR1- 2SLS	MLR2- 2SLS	MLR1- 2SLS AR(1)	MLR2- 2SLS AR(1)
(Constant)	1.290 (.034)	10.959 (.000)	2.029 (.001)	10.967 (.000)	-1.586 (.549)	4.578 (.405)	1.007 (.688)	9.442 (.068)
FARE	-0.049 (.224)	-0.073 (.074)	-0.033 (.433)	-0.039 (.357)	-0.049 (.245)	-0.067 (.118)	-0.019 (.674)	-0.045 (.311)
VEHKM	0.881 (.002)	1.099 (.000)	1.250 (.000)	1.357 (.000)	-0.458 (.707)	-0.653 (.587)	0.743 (.518)	0.458 (.672)
FUEL	0.676 (.000)	0.495 (.000)	0.515 (.001)	0.373 (.022)	0.683 (.000)	0.554 (.000)	0.613 (.000)	0.436 (.008)
VEH	-	-	-	-	-	-	-	-
URATE	0.176 (.000)	-	0.191 (.000)	-	0.123 (.024)	-	0.170 (.006)	-
INCOME	-	-0.914 (.000)	-	-0.859 (.000)	-	-0.658 (.032)	-	-0.904 (.006)
T	-0.002 (.000)	-0.000 (.083)	-0.001 (.011)	-0.000 (.792)	-0.002 (.001)	-0.002 (.042)	-0.001 (.019)	-0.001 (.352)
DJAN	-	-	-	-	-	-	-	-
DFEB	0.044 (.253)	0.090 (.030)	0.089 (.014)	0.113 (.003)	-0.076 (.444)	-0.076 (.472)	0.023 (.811)	0.022 (.815)
DMAR	0.083 (.008)	0.109 (.001)	0.090 (.003)	0.107 (.000)	0.086 (.015)	0.104 (.002)	0.057 (.086)	0.083 (.005)
DAPR	0.050 (.119)	0.068 (.038)	0.072 (.032)	0.080 (.018)	-	-	-	-
DMAY	0.095 (.002)	0.100 (.001)	0.107 (.001)	0.109 (.001)	0.098 (.006)	0.102 (.004)	0.071 (.049)	0.078 (.026)
DJUN	0.088 (.005)	0.094 (.003)	0.108 (.001)	0.109 (.001)	0.057 (.071)	0.094 (.057)	0.065 (.029)	0.062 (.036)
DJUL	0.147 (.000)	0.136 (.000)	0.138 (.000)	0.131 (.001)	0.234 (.019)	0.251 (.011)	0.143 (.148)	0.167 (.074)
DAUG	0.124 (.001)	0.107 (.004)	0.115 (.003)	0.102 (.009)	0.211 (.034)	0.223 (.026)	0.121 (.215)	0.139 (.138)



DSEP	0.012 (.717)	-0.010 (.759)	0.011 (.740)	-0.007 (.844)	0.060 (.366)	0.058 (.407)	0.007 (.917)	0.005 (.934)
DOCT	0.115 (.000)	0.096 (.002)	0.120 (.000)	0.106 (.000)	0.124 (.001)	0.113 (.007)	0.107 (.004)	0.092 (.020)
DNOV	0.147 (.000)	0.130 (.000)	0.155 (.000)	0.148 (.000)	0.109 (.002)	0.087 (.009)	0.134 (.000)	0.104 (.000)
DDIC	-	-	-	-	-	-	-	-
ρ	-	-	0.503 (.000)	0.495 (.000)	-	-	0.457 (.000)	0.451 (.000)
R2	.626	.620	.626	.620	.557	.510	.557	.510
R2adj	.582	.576	.582	.576	.509	.457	.509	.457
F	14.3 (.000)	13.9 (.000)	14.3 (.000)	13.9 (.000)	11.6 (.000)	9.6 (.000)	11.6 (.000)	9.6 (.000)
Durbin - Watson	1.074	1.077	2.098	2.127	1.180	1.170	2.088	2.155

Tabla 16. Parámetros estimados de los modelos de demanda. Modelos teniendo en cuenta los datos de la línea 5C1 (ρ valor del test t aparece entre paréntesis, N=144).

En el caso de la línea 5C1 las variables de interés (URATE e INCOME) no fueron demasiado significativas, pero sin embargo fueron superiores a los valores de las mismas para el caso de datos agregados de las líneas. El modelo cuya estimación obtuvo un mayor peso fue el de MLR1-AR(1) con un valor cercano a 0.2. La variable URATE mostró un signo positivo para todos los modelos estimados, signo esperado al igual que para los valores obtenidos en los modelos para el conjunto de líneas agregadas.

Los parámetros estimados para la variable INCOME resultaron bastante más significativos que los obtenidos en el agregado de líneas y todos ellos con signo negativo como era de esperar. Los parámetros de mayor estimación los arrojaron los modelos MLR2-OLS y MLR2-2SLS AR(1) superando con 0.91 y 0.90 respectivamente.

En el caso de los ajustes de los modelos los valores fueron todos ellos muy similares oscilando entre 0.51 y 0.62 lo que supone que los modelos estimados para la línea 5C1 cuentan con peor capacidad de ajuste que los obtenidos en el modelo agregado de líneas.



5.2.2.4. LINEA 12. Carrefour – Canalejas.

A continuación se pueden ver los parámetros estimados utilizando OLS (Mínimos Cuadrados Ordinarios) y 2SLS (Mínimos Cuadrados en dos etapas) en esta ocasión considerando la línea 3 (véase Tabla 17).

Variables	MLR1- OLS	MLR2- OLS	MLR1- AR(1)	MLR2- AR(1)	MLR1- 2SLS	MLR2- 2SLS	MLR1- 2SLS AR(1)	MLR2- 2SLS AR(1)
(Constant)	-1.250 (.000)	3.413 (.011)	-1.480 (.000)	1.944 (.420)	-0.571 (.971)	6.937 (.012)	0.842 (.625)	7.487 (.023)
FARE	-0.037 (.434)	-0.011 (.830)	-0.013 (.671)	-0.011 (.714)	-0.036 (.436)	-0.032 (.518)	-0.023 (.425)	-0.019 (.531)
VEHKM	0.039 (.727)	0.103 (.404)	-0.023 (.817)	0.003 (.997)	0.340 (.718)	0.188 (.852)	0.958 (.199)	0.883 (.228)
FUEL	0.225 (.039)	0.103 (.384)	0.034 (.832)	-0.037 (.832)	0.181 (.167)	0.027 (.851)	-0.022 (.887)	-0.223 (.216)
VEH URATE	0.139 (.000)		0.123 (.014)		0.139 (.000)		0.128 (.002)	
INCOME		-0.436 (.001)		-0.315 (.195)		-0.783 (.000)		-0.680 (.026)
T	0.002 (.000)	0.003 (.000)	0.002 (.000)	0.003 (.000)	0.002 (.031)	0.003 (.001)	0.001 (.035)	0.002 (.001)
DJAN								
DFEB	-0.058 (.012)	-0.035 (.285)	-0.054 (.012)	-0.046 (.033)	-0.031 (.597)	-0.018 (.768)	0.010 (.837)	0.024 (.601)
DMAR	0.010 (.676)	0.015 (.631)	0.010 (.675)	0.017 (.505)	0.005 (.869)	0.024 (.483)	-0.001 (.953)	0.013 (.588)
DAPR	-0.018 (.516)	-0.019 (.552)	-0.018 (.516)	-0.016 (.579)				
DMAY	0.017 (.580)	0.027 (.428)	0.034 (.262)	0.035 (.264)	0.027 (.351)	0.030 (.323)	0.033 (.148)	0.036 (.115)
DJUN	0.005 (.873)	0.007 (.823)	0.015 (.630)	0.014 (.657)	0.022 (.536)	0.017 (.661)	0.043 (.111)	0.040 (.128)
DJUL	0.118 (.0001)	0.117 (.0004)	0.132 (.000)	0.130 (.0001)	0.125 (.000)	0.129 (.000)	0.127 (.000)	0.131 (.000)
DAUG	0.044	0.039	0.058	0.053	0.052	0.050	0.052	0.052



	(.141)	(.231)	(.066)	(.100)	(.069)	(.102)	(.065)	(.082)
DSEP	0.018	0.008	0.028	0.020	0.032	0.018	0.046	0.036
	(.553)	(.792)	(.345)	(.498)	(.294)	(.579)	(.094)	(.194)
DOCT	0.086	0.076	0.095	0.090	0.089	0.077	0.077	0.065
	(.004)	(.017)	(.0002)	(.0005)	(.009)	(.035)	(.010)	(.029)
DNOV	0.067	0.060	0.066	0.064	0.090	0.063	0.115	0.094
	(.027)	(.064)	(.001)	(.001)	(.081)	(.251)	(.004)	(.018)
DDIC	-	-	-	-	-	-	-	-
ρ	-	-	0.661	0.707	-	-	0.588	0.694
			(.000)	(.000)			(.000)	(.000)
R2	.775	.741	.775	.741	.762	.726	.762	.726
R2adj	.749	.711	.749	.711	.737	.696	.737	.696
F	29.5	24.4	29.5	24.4	29.6	24.4	29.6	24.4
	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)
Durbin - Watson	0.723	0.623	2.390	2.454	0.759	0.611	1.968	2.008

Tabla 17. Parámetros estimados de los modelos de demanda. Modelos teniendo en cuenta los datos de la línea 12 (ρ valor del test t aparece entre paréntesis, N=144).

En el caso de la línea 12 las variables de interés (URATE e INCOME) no fueron demasiado significativas pero sus valores estuvieron algo por encima de los registrados en los modelos para el caso de datos agregados de las líneas. La variable URATE mostró un signo positivo para todos los modelos estimados, manteniéndose la lógica de las líneas anteriores y de los modelos agregados.

Los parámetros estimados para la variable INCOME resultaron similares a los obtenidos en el agregado de líneas y todos ellos con signo negativo como era de esperar. Los parámetros de mayor estimación los arrojaron los modelos MLR2-OLS y MLR2-2SLS AR(1) superando con 0.78 y 0.68 respectivamente.

En el caso de los ajustes de los modelos los valores fueron todos ellos muy similares oscilando entre 0.72 y 0.77 lo que supone que los modelos estimados para la línea 12 cuentan con un ajuste prácticamente igual que el del modelo agregado de líneas.



5.3. Validación del modelo

5.3.1. Validación del modelo con datos agregados

Teniendo en cuenta los datos de demanda agregados, los modelos fueron validados usando datos referentes a 2013 de cara a comprobar su capacidad predictiva. Previamente a la validación se aplicó el test de ADF para comprobar la estacionariedad débil de la serie temporal con término constante y variable de tendencia. Los resultados del test fueron en todos los casos cercanos a -3.5, inferior al valor crítico de -3.41 para un nivel de confianza del 95%. De ahí se deriva que se puede rechazar la hipótesis nula de raíz unitaria y por lo tanto de no estacionariedad débil de la serie.

En la Tabla 18 se recogen los errores cuadráticos medios (MSE) de los modelos de equilibrio.

Modelo	MSE
MLR1 - AR(1)	0.008
MLR2 - AR(1)	0.012
MLR1 - 2SLS AR(1)	0.006
MLR2 - 2SLS AR(1)	0.009

Tabla 18. MSE de los resultados de los modelos frente a las observaciones de 2013.

En general los errores cuadráticos medios (MSE) se situaron por debajo de 0.010 excepto en el caso del modelo MLR2 – AR(1) que tuvo una capacidad predictiva claramente inferior al resto. Los modelos estimados mediante 2SLS presentaron un error cuadrático medio ligeramente menor siendo el modelo más cercano a los datos de 2013 el MLR1 – 2SLS AR(1) estimado mediante mínimos cuadrados en dos etapas y con la tasa de paro como variable independiente. En términos de error absoluto medio en el número de viajeros, este modelo arrojó un error de 88.757 pasajeros, lo que supone un 6,72% del promedio mensual.



5.3.2. Validación del modelo con datos desagregados. Por líneas. Las más significativas.

A continuación se muestran los datos de demanda desagregados por líneas, siendo estas las más representativas en cuanto a los resultados obtenidos.

5.3.2.1. Validación del modelo. Línea 1.

En la Tabla 18 se recogen los errores cuadráticos medios (MSE) de los modelos de equilibrio para la línea 1.

Modelo	MSE
MLR1 - AR(1)	0.003
MLR2 - AR(1)	0.032
MLR1 - 2SLS AR(1)	0.017
MLR2 - 2SLS AR(1)	0.003

Tabla 19. MSE de los resultados de los modelos frente a las observaciones de 2013 de la línea 1.

En general, para la línea 1 los errores cuadráticos medios (MSE) se situaron por debajo de 0.030 excepto en el caso del modelo MLR2 – AR(1) que tuvo una capacidad predictiva algo inferior al modelo MLR1 - 2SLS AR(1) y aún menor en comparación al resto. Los modelos estimados mediante OLS como el de MLR1 - AR(1) y 2SLS como el de MLR2 - 2SLS AR(1) presentaron el mismo error cuadrático medio siendo sensiblemente menor que los anteriormente citados. Se optó por el modelo MLR2 - 2SLS AR(1) como el más cercano a los datos de 2013 estimado mediante mínimos cuadrados en dos etapas y con la recaudación o ingresos deflactados per cápita como variable independiente. En términos de error absoluto medio en el número de viajeros, este modelo arrojó un error de 10.040 pasajeros, lo que representa un 4,34% del promedio mensual.

5.3.2.2. Validación del modelo. Línea 3.

En la Tabla 18 se recogen los errores cuadráticos medios (MSE) de los modelos de equilibrio para la línea 3.

Modelo	MSE
MLR1 - AR(1)	0.006
MLR2 - AR(1)	0.004
MLR1 - 2SLS AR(1)	15.573
MLR2 - 2SLS AR(1)	0.003

Tabla 20. MSE de los resultados de los modelos frente a las observaciones de 2013 de la línea 3.

En el caso de la línea 3 los errores cuadráticos medios (MSE) fueron razonablemente correctos en tres de los cuatro modelos estimados siendo el modelo MLR1 - 2SLS AR(1) el que contaba claramente con una menor capacidad predictiva por su alto valor del error. En los tres tipos de modelos restantes los resultados de los modelos están por debajo de 0.005 siendo el MLR2 - 2SLS AR(1) el idóneo una vez se estableció la comparación con los datos de 2013. Al igual que en la línea 1 el modelo de mejor calidad predictiva fue el estimado mediante mínimos cuadrados en dos etapas y la recaudación o ingresos deflactados per cápita como variable independiente. En términos de error absoluto medio en el número de viajeros, este modelo arrojó un error de 3.713 pasajeros, lo que representa un 4,02% del promedio mensual.

5.3.2.3. Validación del modelo. Línea 5C1.

En la Tabla 18 se recogen los errores cuadráticos medios (MSE) de los modelos de equilibrio para la línea 5C1.

En el caso de la línea 5C1 los errores cuadráticos medios (MSE) se mostraron regulares siendo el modelo MLR1 - 2SLS AR(1) con peor capacidad predictiva. En general los errores cuadráticos medios (MSE) se situaron por debajo de 0.012. Los mejores modelos son los MLR2 - AR(1) y MLR2 - 2SLS AR(1) siendo este último, estimado mediante mínimos cuadrados



en dos etapas y la recaudación o ingresos deflactados per cápita como variable independiente, el más sólido de cara a hacer proyecciones de demanda. En términos de error absoluto medio en el número de viajeros, este modelo arrojó un error de 3.889 pasajeros, lo que representa un 3,17% del promedio mensual.

Modelo	MSE
MLR1 - AR(1)	0.004
MLR2 - AR(1)	0.002
MLR1 - 2SLS AR(1)	0.012
MLR2 - 2SLS AR(1)	0.001

Tabla 21. MSE de los resultados de los modelos frente a las observaciones de 2013 de la línea 5C1.

5.3.2.4. Validación del modelo. Línea 12.

En la Tabla 18 se recogen los errores cuadráticos medios (MSE) de los modelos de equilibrio para la línea 12.

Modelo	MSE
MLR1 - AR(1)	0.009
MLR2 - AR(1)	0.002
MLR1 - 2SLS AR(1)	0.005
MLR2 - 2SLS AR(1)	0.003

Tabla 22. MSE de los resultados de los modelos frente a las observaciones de 2013 de la línea 12.

Para la línea 12 los errores cuadráticos medios (MSE) se situaron por debajo de 0.01 siendo el modelo MLR1 – AR(1) el que posee una peor capacidad predictiva. El mejor modelo fue son los MLR2 - AR(1) , estimado mediante mínimos cuadrados ordinarios y la recaudación o ingresos deflactados per cápita como variable independiente. En términos de error absoluto medio en el número de viajeros, este modelo tuvo como



resultado un error de 3.870 pasajeros, lo que representa un 4,11% del promedio mensual.



CAPÍTULO 6

Conclusiones y líneas futuras de investigación



6. CONCLUSIONES Y LINEAS FUTURAS DE INVESTIGACIÓN

En la presente tesis se han estimado modelos para comprobar la influencia de variables relacionadas con el ciclo económico como la renta per cápita y la tasa de paro en la demanda de transporte público en bus en la ciudad media de Santander. Los modelos han sido estimados mediante OLS y 2SLS de cara a tener en cuenta en estos últimos la endogeneidad del nivel de servicio. Los modelos han considerado además la existencia de autocorrelación en los residuos aplicando la estimación propuesta por Prais – Winsten.

Los resultados obtenidos permiten afirmar que las fases descendentes del ciclo económico pueden tener un impacto significativo en el número de viajeros. En los modelos de equilibrio, el aumento de la tasa de paro implicó de forma significativa un aumento en la demanda de transporte público en bus. Por el contrario si se considera el GDP per cápita, el parámetro no fue claramente significativo en todos los modelos, aunque si presentó un signo negativo que implica la caída de la demanda al aumentar los ingresos y que indica el carácter de bien inferior del transporte público. Estos resultados fueron similares también en los modelos a largo plazo.

Considerando la magnitud de las elasticidades, el modelo MLR1-2SLS AR(1) presentó una estimación en la tasa de paro de 0.133 en el modelo estático de equilibrio y de 0.210 en el modelo dinámico a largo plazo. Esto significa que un incremento en un 1% en la tasa de paro puede implicar un aumento en el número de viajeros en el entorno del 0.13-0.2%. En España, donde la tasa de paro se ha elevado en la fase recesiva del ciclo en más de 15 puntos, esto puede suponer un incremento de un 2-3% en la demanda de transporte en bus, una cifra quizás no excesivamente elevada pero que puede resultar de interés para los planificadores de transporte de cara al ajuste en la gestión de sus servicios. La elasticidad obtenida para los ingresos no fue significativa en todos los modelos, con parámetro estimado en el modelo estático de -0.505 y de -0.861 en el modelo dinámico a largo plazo. Estos cálculos implican un aumento en la demanda similar a la dada por la tasa de paro si se considera una reducción del GDP de más de 6 puntos porcentuales como la que se ha dado en España.



Estos efectos podrían deberse en parte al estancamiento experimentado de la propiedad de automóviles que se produjo durante la fase de de recesión en la zona de estudio. La correlación moderada entre VEH e INCOME indica que este último puede captar algunos de los efectos. Este efecto estaría de acuerdo con los resultados conocidos en la literatura relacionada con el fuerte impacto de la propiedad de automóviles sobre el uso del autobús como medio de transporte público.

La comparación entre la capacidad predictiva de los distintos modelos permitió comprobar como los estimados mediante 2SLS presentaron un menor error medio cuadrático (MSE). Además el mejor modelo en términos predictivos fue el especificado con la tasa de paro como variable independiente en lugar del GDP per cápita. Estos resultados son también relevantes de cara realizar estimaciones sobre la demanda futura de transporte que considere el efecto de la dinámica macroeconómica.

En cuanto a la aplicación del modelo estimado al resto de líneas que forman parte del estudio, fueron cuatro las líneas que obtuvieron unos mejores ajustes como fueron las líneas 1, 3, 5C1 y 12. Dichas líneas mostraron parámetros con estimaciones en general no demasiado significativas pero cuyo signo coincidía con los de los parámetros estimados para el modelo agregado de líneas.

Esta tesis está abierta a otras líneas de investigación futuras que podrían complementar aún más los resultados obtenidos. De la misma manera que se ha optado por estimar modelos utilizando variables de tipo socio económico en fases contractivas de la economía, podrían también estimarse modelos a partir de la combinación de variables demográficas y variables espaciales y así poder estimar modelos de previsión de la demanda en función de las variaciones demográficas en distintas áreas de la ciudad pudiendo utilizar como base las secciones y los distritos de la ciudad de Santander.

Otra línea de investigación interesante podría orientarse a la estimación de modelos a partir de variables de tipo climatológico. Esos modelos podrían servir para hacer previsiones de carácter diario si se tuvieran como referencia los datos de viajeros distribuidos por franjas horarias y paradas en función de diferentes tipos de tiempo. En



lo que a modelos estimados a partir de datos climatológicos se refiere se podrían desarrollar estimaciones a partir de datos obtenidos en días especialmente adversos del año. Igualmente podría resultar interesante la estimación de la demanda en días anormalmente adversos en periodos de verano al igual que de días anormalmente benignos en periodos de invierno.

Con los modelos estimados a partir de de los diferentes tipos de datos (socio-económicos, demográficos, espaciales y climatológicos) se podría tener un conocimiento más global de las diferentes variables que mayor influencia tienen sobre la demanda de transporte público y así poder ayudar a la toma de decisiones para mejorar los datos de demanda.

Una última línea de investigación se basaría en la aplicación de la metodología de esta tesis a los datos de demanda de distintas ciudades de tamaño similar a la ciudad de Santander y así poder indagar cómo influyen las mismas variables en distintos entornos urbanos. Podrían clasificarse distintas áreas geográficas en España a partir de la cuales poder aplicar esta metodología en distintas ciudades de esas áreas, como por ejemplo ciudades en la franja norte del país, ciudades en la franja mediterránea, ciudades en la franja sur y ciudades del centro del país así como de territorios insulares y a partir de los resultados realizar una clasificación de las diferentes áreas en función de la influencia que ejercen sobre la demanda unas y otras variables.



Bibliografía



BIBLIOGRAFÍA

- Abuín, J. R. (2007). "Regresión lineal múltiple." *IdEyGdM-Ld Estadística*, Editor: 32.
- Aguirre Jaime, A. (1994). "Introducción al tratamiento de series temporales: Aplicación a las ciencias de la salud." Introducción al tratamiento de series temporales: Aplicación a las ciencias de la salud.
- Arellano, M. and O. Bover (1990). "IA ECONOMETRIA DE DATOS DE PANEL." Investigaciones Económicas (segunda época) **14**(1): 3-45.
- Arellano, M. and B. de España (1992). Introducción al análisis econométrico con datos de panel, Banco de España.
- Balcombe, R., et al. (2004). "The demand for public transport: a practical guide."
- Bound, J., et al. (1995). "Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak." Journal of the American Statistical Association **90**(430): 443-450.
- Box, G. E. and G. M. Jenkins (1976). Time series analysis: forecasting and control, revised ed, Holden-Day.
- Bresson, G., et al. (2004). "Economic and structural determinants of the demand for public transport: an analysis on a panel of French urban areas using shrinkage estimators." Transportation Research Part A: Policy and Practice **38**(4): 269-285.
- Cameron, A. C. and P. K. Trivedi (2005). Microeconometrics: methods and applications, Cambridge university press.
- Castellacci, F. (2008). "Technology clubs, technology gaps and growth trajectories." Structural Change and Economic Dynamics **19**(4): 301-314.
- Causelo, P. Á. (2012). *Apuntes de Microeconomía I*. U. d. Cantabria.
- Coppola, P., et al. (2013). "A LUTI Model for the Metropolitan Area of Santander." Journal of Urban Planning and Development **0**(ja): null.
- Chatfield, C. (2013). The analysis of time series: an introduction, CRC press.



Cos Guerra, O., Meer Lecha-Marzo, Á. D., & Martín Latorre, E. (2007). Planeamiento urbanístico y crecimiento urbano: importancia de la escala metropolitana en la definición del modelo territorial.

Dagum, E. (1988). "The X-11-ARIMA/88 Seasonal Adjustment Method, Methodology Branch." Statistics Canada.

Dargay, J. M. and M. Hanly (2002). "The demand for local bus services in England." Journal of transport economics and policy: 73-91.

DE, E. D. D. "Serie Apuntes de Clase."

de Rivera, D. P. S. (2005). Análisis de series temporales, Alianza Editorial.

De Rus, G. (1990). "Public transport demand elasticities in Spain." Journal of Transport Economics and Policy: 189-201.

dell'Olio, L., et al. (2011). "The quality of service desired by public transport users." Transport Policy **18**(1): 217-227.

Durbin, J. and G. S. Watson (1950). "Testing for Serial Correlation in Least Squares Regression: I." Biometrika **37**(3/4): 409-428.

Durbin, J. and G. S. Watson (1951). "Testing for serial correlation in least squares regression. II." Biometrika: 159-177.

Figueroa, O. (2005). "Transporte urbano y globalización: Políticas y efectos en América Latina." EURE (Santiago) **31**(94): 41-53.

Filardo, A. J. (1994). "Business-cycle phases and their transitional dynamics." Journal of Business & Economic Statistics **12**(3): 299-308.

Frank, R. H. and R. Piras (2003). Microeconomía, McGraw-Hill.

García Mora, A., et al. (2007). "El sistema financiero español ante la crisis crediticia internacional." Ekonomiaz **3**(66): 92-125.

Glass, G. V. (1966). "Testing homogeneity of variances." American Educational Research Journal **3**(3): 187-190.



Goodwin, P. B. (1992). "A review of new demand elasticities with special reference to short and long run effects of price changes." Journal of transport economics and policy: 155-169.

Greene, W. H. (2003). Econometric analysis, Pearson Education India.

Gujarati, D. N. and D. C. Porter (2009). Basic econometrics. Boston, McGraw-Hill Irwin.

Hardin, J. W. (1996). "Prais-Winsten regression." Stata Technical Bulletin 5(25).

Hausman, J. A. (1978). "Specification tests in econometrics." Econometrica: Journal of the Econometric Society: 1251-1271.

Hayashi, F. (2000). "Econometrics." Princeton University Press. Section 1: 60-69.

Holmgren, J. (2007). "Meta-analysis of public transport demand." Transportation Research Part A: Policy and Practice 41(10): 1021-1035.

Hylleberg, S., et al. (1990). "Seasonal integration and cointegration." Journal of econometrics 44(1): 215-238.

INE (2013). Encuesta de Población Activa (EPA), Instituto Nacional de Estadística.

Irgoin, C. H. A. (2011). "Análisis De Series De Tiempo." Contribuciones a la Economía(2011-02).

Jiménez, E. U. and A. P. Giménez (2000). Introducción al análisis de series temporales.

Johnston, J. and J. Dinardo (2001). Métodos de econometría, Vicens Vives.

Juan, Á. A., et al. "Modelos de ecuaciones simultáneas." Universitat Obertura de Catalunya.

Kikut, A. C. and C. Torres (1998). Variables estacionales en los modelos de regresión: una aplicación a la demanda por dinero en Costa Rica Departamento Investigaciones Económicas, DIE-NT-02-98, Banco Central de Costa Rica. En revista "Economía y Sociedad".

Kydland, F. E. and E. C. Prescott (1991). "El análisis econométrico del enfoque de equilibrio general de los ciclos económicos." Cuadernos Económicos del ICE(48): 2.



Labra, R. and C. Torrecillas (2014). "Guía CERO para datos de panel. Un enfoque práctico." UAM-Accenture Working Papers(2014/16): 1-57.

Litman, T. (2014). Transit Price Elasticities and Cross-Elasticities. Victoria, Victoria Transport Policy Institute.

Mahía, R. (2001). "Conceptos básicos de matemática económica para el manejo de series." Recuperado de: http://www.uam.es/personal_pdi/economicas/anadelsur/pdf/manejo.pdf.

Mayorga, M. and E. Muñoz (2000). "La técnica de datos de panel. Una guía para su uso e interpretación." Documento de trabajo. San José: Banco Central de Costa Rica, División Económica.

Medina Moral, E. (2002). "Introducción a la Econometría e Informática." Modelos Econométricos e Información Estadística.

Molina Ibáñez, J. G., et al. (2009). "Estadística (Estadística descriptiva en Psicología)(2009-2010)."

Mundial, B. (2001). "Correspondencia sobre las tasas de crecimiento per cápita del PIB." Marzo, Washington, DC.

Nijkamp, P. and G. Pepping (1998). "Meta-analysis for explaining the variance in public transport demand elasticities in Europe." Journal of Transportation and Statistics 1(1): 1-14.

Ocón Galilea, F. J. (2013). ""La Crisis Económica Española a partir de 2007"." Universidad de La Rioja: 44.

Oum, T. H., et al. (1992). "Concepts of price elasticities of transport demand and recent empirical estimates: an interpretative survey." Journal of transport economics and policy: 139-154.

Paulley, N., et al. (2006). "The demand for public transport: The effects of fares, quality of service, income and car ownership." Transport Policy 13(4): 295-306.

Peña Sánchez de Rivera, J. I. (1990). "Problemas de homogeneidad en modelos de series temporales."

Prais, S. and C. Winsten (1954). "Trend estimation and serial correlation." Cowles Commission.



Prais, S. J. and C. B. Winsten (1954). Trend estimators and serial correlation, Cowles Commission discussion paper.

Redondo, J. F. B. (2011). "La función de producción Cobb Douglas y la economía española." Revista de economía crítica(12): 9-38.

Reyes, G. M., et al. (1997). Introducción a la econometría, Prentice Hall.

Rodríguez Jaume, M. J. and R. Mora Catalá (2001). "Análisis de regresión múltiple."

Rodríguez, M. d. C. F., et al. (2010). "Combinación De Datos De Corte Transversal Y De Series Temporales: La Utilización De Datos De Panel." Contribuciones a la Economía(2010-10).

Roodman, D. (2009). "A note on the theme of too many instruments*." Oxford Bulletin of Economics and statistics **71**(1): 135-158.

Uriel, E. (2013). "Regresión lineal múltiple: estimación y propiedades." Universidad de Valencia Versión: 09-2013.

Uriel, E. and I. G. Rosat (1997). Econometría aplicada, Paraninfo.

Varian, H. R. and J. Repcheck (2010). Intermediate microeconomics: a modern approach, WW Norton & Company New York, NY.

Webster, F. V. and P. H. Bly (1981). "The demand for public transport." Transport Reviews **1**(4): 323-351.

Wu, D.-M. (1973). "Alternative Tests of Independence between Stochastic Regressors and Disturbances." Econometrica **41**(4): 733-750.

Zellner, A. (1978). Seasonal analysis of economic time series, US Department of Commerce, Bureau of the Census.