UNIVERSIDAD DE OVIEDO

Departamento de Economía Aplicada

COST OF LIVING IN SPAIN: A SPATIAL PERSPECTIVE

Programa de Doctorado "Economía y Sociología de la Globalización" (Mención de Calidad)

Elena Lasarte Navamuel

(2014)

Universidad de Oviedo
Departamento de Economía Aplicada

COST OF LIVING IN SPAIN: A SPATIAL PERSPECTIVE

Memoria que, para la obtención del grado de doctor, presenta

Elena Lasarte Navamuel

Programa de Doctorado “Economía y Sociología de la Globalización” (Mención de Calidad)

Bajo la dirección de los profesores
Dr. Esteban Fernández Vázquez
Dr. Fernando Rubiera Morollón

RESUMEN DEL CONTENIDO DE TESIS DOCTORAL

1.- Título de la Tesis	
Español/Otro Idioma:	Inglés: COST OF LIVING IN SPAIN: A SPATIAL PERSPECTIVE

2.- Autor	
Nombre: ELENA LASARTE NAVAMUEL	DNI/Pasaporte/NIE:
Programa de Doctorado: Economía Aplicada y Sociología de la Globalización	
Órgano responsable: ECONOMIA APLICADA	

RESUMEN (en español)

Desde mediados del siglo pasado los científicos especializados en la Ciencia Regional han estado estudiando el efecto de las aglomeraciones, generando una rica literatura, tanto teórica como empírica, que ha evidenciado los efectos que tienen sobre toda actividad económica las diferencias regionales y las aglomeraciones. El estudio de las diferencias regionales y las economías de aglomeración en España van a ser el centro de esta tesis. En este contexto, las economías de aglomeración van a ser uno de los conceptos clave, entendiéndose por economías de aglomeración aquellas que se generan cuando la población se concentra en un espacio urbano, generando así una serie de efectos externos que son esenciales para entender las dinámicas espaciales en una economía. Estos efectos externos son, por ejemplo, la escasez de suelo urbano o la congestión. Como hemos dicho anteriormente existe una gran extensión de literatura económica dedicada al estudio de cómo las aglomeraciones afectan a las dinámicas de crecimiento e innovación, generación de especialización productiva, pero muy poco se ha estudiado acerca de cómo las economías de aglomeración afectan al coste de vida y al bienestar.

Si las economías de aglomeración son capaces de explicar por qué las actividades económicas se localizan en un lugar determinado o por qué las ciudades grandes siguen diferentes patrones de comportamiento que las ciudades medianas o los entornos rurales, parece lógico pensar que las economías de aglomeración serán un concepto clave a la hora de estudiar las dinámicas espaciales de los patrones de consumo, los precios y el coste de vida. Hay estudios empíricos previos que sugieren que el lugar de residencia e incluso el tamaño de la ciudad de residencia afecta a los patrones de consumo. Es lógico, ya que las ciudades más grandes ofrecen una gran variedad de productos y servicios, además, de una gran calidad

que no se encuentran en ciudades más pequeñas haciendo que se genere un estilo de vida diferente. Como resultado los comportamientos de los consumidores serán diferentes en estas grandes ciudades a los que se generan en las zonas rurales. Además, se espera que todas estas características de las grandes metrópolis hagan que el coste de vida sea mayor en ellas.

Aunque las disparidades regionales en el coste de vida juegan un importante papel en el estudio de la economía regional y en la aplicación de políticas regionales, en general, no existen datos oficiales sobre costes de vida a nivel espacial. En España, esta falta de datos es relevante ya que las características del país impulsan a tener en cuenta el espacio y la geografía a la hora de estudiar las dinámicas económicas. Aunque España no es un país especialmente grande, sí está caracterizado por diferencias bastante grandes en los aspectos culturales, sociológicos y climatológicos. Además, España es un país particularmente atractivo para este tipo de estudios ya que comprende todo tipo de ciudades, desde grandes metrópolis, como Madrid y Barcelona, hasta numerosas ciudades medianas todas ellas rodeadas por áreas muy rurales. Todo este entramado urbano tan completo se caracteriza además por comprender tanto ciudades centrales como periféricas, ciudades costeras y turísticas como ciudades interiores, además de la especial situación de las Islas Baleares y las Islas Canarias.

España es un sistema federal *de facto* en donde la mayoría de las políticas que afectan al bienestar de los ciudadanos son responsabilidad de los gobiernos regionales y locales. Debido a la falta de información sobre índices de coste de vida a nivel espacial, estas políticas se toman apoyándose en indicadores de renta a nivel nominal. Los ajustes en estos indicadores de bienestar por diferencias en el coste de vida son muy relevantes para el estudio de las desigualdades regionales, comparaciones de salarios o análisis de pobreza. Por todo esto, esta tesis va a proporcionar nueva información acerca de los costes de vida a nivel espacial, basándose en la teoría económica del consumidor y en técnicas microeconométricas que proporcionarán información fiable y precisa sobre los costes de vida en España.

En el Capítulo 2 de esta tesis se explicará de forma sucinta la metodología aplicada para la construcción de Índices de Coste de Vida Espaciales. La idea central que subyace en la metodología aplicada es que un "verdadero" índice de coste de vida debe comparar el coste monetario de adquirir dos cestas de bienes diferentes conectadas únicamente por la condición de que proporcionan el mismo nivel de bienestar (Konus, 1939). Este enfoque entraña muchas dificultades computacionales debido a que la función de utilidad de un consumidor no es directamente observable, por eso se va a aplicar una metodología específica llamada Almost Ideal Demand System (Sistema de Demanda Cuasi-Ideal) desarrollada por Deaton y Muellbauer (1980). Para aplicar esta metodología se usará la Encuesta de Presupuestos Familiares publicada por el Instituto Nacional de Estadística. Esta encuesta recoge hábitos de consumo y características socioeconómicas de los hogares españoles.

Como se ha señalado anteriormente, las disparidades en el coste de vida a nivel espacial juegan un importante papel en la Economía Regional y Urbana. A pesar de esto no existe información de niveles de precios en España, y, en general, en ningún país Europeo. Los esfuerzos y los datos disponibles son destinados a la construcción del Índice de Precios al Consumo (IPC), pero este índice es simplemente un indicador de inflación que no permite comparaciones de precios entre regiones, solo permite conocer la evolución de los precios en una región a lo largo de un periodo de tiempo. En el Capítulo 3 de la tesis se construirá el Índice de Coste de Vida Espacial considerando como regiones las Comunidades Autónomas Españolas y para el periodo 2008-2012. Este índice permitirá hacer comparaciones de costes de vida tanto a lo largo del tiempo como a través del espacio y su importancia radica, además de que va a poner de relieve las disparidades entre Comunidades Autónomas, en que se utilizará para la evaluación de la pobreza de las regiones en España. Es importante centrar la atención en el análisis de la pobreza ya que, en España y debido a la grave crisis económica, ésta ha aumentado considerablemente tanto a nivel nacional como en cada una de las regiones. Lo que se pretende en este capítulo es proporcionar información sobre cómo la incidencia de la pobreza es diferente entre regiones y cómo se ve afectada cuando se tienen en cuenta las diferencias de coste de vida regionales.

Las aglomeraciones se producen a diversas escalas geográficas. Las disparidades regionales dentro de un mismo país hacen que se produzcan aglomeraciones a una escala geográfica más pequeña que las regiones. Este será el centro del Capítulo 4 de esta tesis. A pesar de la rigurosidad económica de los resultados obtenidos en el

Capítulo 3, hay varias debilidades en este análisis que se derivan de la existencia de una diversidad muy rica de ciudades dentro de cada Comunidad Autónoma. Los datos usados no proporcionan información a una escala más desagregada que las Comunidades Autónomas, por lo que no es posible calcular el Índice de Coste de Vida Espacial a nivel provincial o a nivel municipal, lo que nos daría una panorámica bastante más fiel del coste de vida en España. A pesar de esto se intentará explotar la información de la encuesta y buscar sus fortalezas para poder estudiar uno de los más importantes asuntos en la Economía Regional y Urbana esto es los efectos de las aglomeraciones sobre el consumo y el coste de vida. Es más, la información proporcionada en la encuesta nos permitirá calcular un Índice de Coste de Vida para las zonas rurales y las zonas urbanas así como para toda la estructura urbana del país.

El primer objetivo de este capítulo será el análisis de los patrones de consumo a lo largo de los distintos tamaños de ciudad a través de la estimación de las elasticidades de demanda. La literatura teórica nos ofrece varias explicaciones de cómo el tamaño de la ciudad puede afectar a los precios y a la conducta del consumidor. Los modelos estándar de la Nueva Geografía Económica (NEG) predicen que los costes de vida son menores en las ciudades que en la periferia, este el caso del modelo de Krugman (1991) así como de otros modelos de la NEG. Pero, en cambio, otros modelos teóricos predicen que los costes de vida tienden a ser mayores en las aglomeraciones debido a la escasez de suelo urbano y los costes de vivienda más altos (Tabuchi, 2001). Ante esta controversia hay que acercarse a la literatura empírica para saber qué se concluye acerca del efecto de las aglomeraciones sobre el coste de vida. La mayoría de los trabajos empíricos han sido realizados para los Estados Unidos y también la mayoría de ellos concluyen que hay una relación positiva entre el tamaño de la ciudad y el coste de vida. El segundo objetivo de este capítulo es calcular un índice de coste de vida para cada tamaño de ciudad según la clasificación de las mismas en la Encuesta de Presupuestos Familiares ¿Pero qué es lo que hace que existan diferencias persistentes en el coste de vida? El tercer objetivo de este capítulo es contestar a esta pregunta a través de la postulación de un modelo que sea capaz de explicar los determinantes del coste de vida en España. Debido a que las grandes ciudades atraen a una población con unas características particulares; de mayor renta, mayor nivel educativo, etc., en éstas se producirán unos patrones de consumo que no se darán en las zonas menos urbanas haciendo que la demanda de ciertos productos crezca y como consecuencia de la presión sobre la demanda los precios sean más altos. Este proceso hace que las aglomeraciones sean más caras, pero al mismo tiempo puede ocurrir que existan características intrínsecas a las aglomeraciones que hacen que los precios sean mayores en éstas, es decir, un mismo individuo con unas particulares características puede tener un comportamiento diferente en las aglomeraciones que en las zonas rurales. Para contrastar estas hipótesis se calculará el coste de vida de cada hogar de la muestra que será la variable dependiente de un modelo de regresión en el que se incluirán tanto variables geográficas y demográficas como variables socioeconómicas del hogar. En cuanto a las variables geográficas se incluirán variables de regiones y una variable que representará el efecto de las aglomeraciones sobre el coste de vida. Respecto a las variables socioeconómicas se incluirán aquellas relativas a las características del hogar y de los individuos, como nivel educativo, nivel de ingresos, tamaño del hogar, entre otras. El propósito de estas variables es aislar el puro efecto de las aglomeraciones sobre el coste de vida a través del control de las características de los hogares e individuos.

Por último, en el Capítulo 5, se expondrán las principales conclusiones derivadas de los resultados obtenidos en esta tesis, se plasmarán las implicaciones de política económica y se propondrán futuras líneas de investigación derivadas de esta tesis.

Bibliografía

Deaton, A., & Muellbauer, J. (1980). An Almost Ideal Demand System. *The American Economic Review, 70*(3).

Konus, A. A. (1939). The Problem of the True Index of the Cost of Living. *Econometrica, 7*(1), 10-29.

Krugman, P. (1991). Increasing returns and econmic geography. *Journal of Political Economy, 99*, 483-499.

Tabuchi, T. (2001). On interregional price differentials. *The Japanese Economic Review, 52*(1).

RESUMEN (en Inglés)

Since the 40s of the 20th century regional scientists have been working on agglomeration economies. This research has generated a rich world of theoretical literature and empirical evidence clearly showing the regional differences and agglomerations affect all kinds of economic activities. Regional differences and

agglomeration economies are a crucial issue that will be the center of this thesis. Agglomeration economies are one of the key concepts in this context; as the population becomes more concentrated into populated urban spaces a series of external effects essential for understanding spatial dynamics, like urban land constrains, commuting costs or interregional transportation costs, are triggered off. There exists a large body of literature devoted to study how agglomerations affect the dynamics of growth and innovation, trends in productive specialization or processes of spatial concentration of economic activity but very little has been done on how agglomerations determines the standard of living.

If agglomeration economies are capable of explaining why economic activities locates in some places or explain why large cities follow paths different from those of medium-sized or small rural areas, it would seem logical that the agglomeration economies is a fundamental concept to understand the spatial dynamics of other aspects such as consumption patterns, price dynamics and costs of living. There are previous empirical studies that suggest that the place of residence, in an urban or rural area even depending on the size of the city, affects consumption behavior. Large cities offer a greater variety and higher quality of goods attracting people with particular characteristics and generating a different style of life. As a result the consumption patterns generated in the metropolis are different than those generated in rural areas. Moreover, is also hopped that local amenities founded in metropolitan areas makes that the cost of living is higher.

Although disparities in the cost of living across the space play a crucial role in regional economics and regional policy, in general, comprehensive regional cost of living data are not available from a governmental source. In Spain, this is particularly relevant due to the characteristics of the country. Although is not a big country, is characterized by striking cultural, sociological and climatological differences both between Autonomous Communities and within Autonomous Communities. Spain regions comprise all kinds of cities, since the biggest ones to very rural areas and also characterized by the existence of much more medium-size cities in central and peripheral positions as well as coastal and inland positions, all of them surrounding by very rural areas.

Spain is a *de facto* federal system in which most of the welfare and developing programs have become the responsibility of regional and local governments. On account of the lack of information, regional studies have to rely on nominal indicators. Adjustments in income measures to take into account the local cost of living are immediately relevant for economic analysis such as inequality studies, wage comparisons and assessments of poverty. The rationale of this thesis is to provide new information about regional costs of living in Spain based on the demand theory and on the micro-econometric analysis.

The aim of this thesis is to estimate spatial cost of living indices with a microeconomic approach. In the Chapter 2 the methodology to construct these indices will be explained in detail. The methodology followed along this thesis is based on Konus (1939) who raised the problem of the "true index of cost of living". This idea emphasized in when computing a "true index of cost of living" it should be compared the monetary cost of two different basket of goods which are connected by the solely condition that they provided the same standard of living. This approach entails computational problems due to the impossibility of knowing directly the utility function. To address this problem it will be used the methodology developed by Deaton and Muellbauer (1980) called Almost Ideal Demand System (AIDS). This methodology permits the estimation of a flexible demand system with several properties without the necessity of know the form of the utility function. With this approach we can know the not observable utility function, and, consequently, to derive the cost function and to calculate a cost of living index (COLI) by fixing the utility level instead the basket of consumer goods as in the case of the CPIs calculated by statistical agencies. For estimating the AIDS we use expenditure data from the Spanish Household Budget Survey (HBS) provided by the National Statistical Institute (INE).

Although disparities in cost of living across the space play a crucial role in Regional and Urban Economics, knowledge on regional price levels is scarce in all European countries, especially in Spain. The price data are usually conducted to construct the CPI at the national or regional level but the CPI is only an inflation measure and does not allow for interregional price comparisons. All this leads to the necessity of constructing a Spatial Index of the Cost of Living (SCOLI). This is the purpose of Chapter 3 of this thesis: to construct a SCOLI at the regional level considering as regions the administrative Autonomous Communities of Spain. This index will allow knowing the existence and magnitude of the differences in cost of living across Spanish regions in year 2012, the last year of available data in the HBS. Moreover, a temporal SCOLI will also be calculated in this chapter for the period 2008-2012. This spatio-temporal index will allow making comparisons both along the time and across the space, simultaneously. Along al the chapter comparisons of these indices with the CPI will be made in order to highlights the strengths of the indices and the limitations of the CPI.

An important issue derived to the lack of price information is that regional studies have to rely on nominal indicators. Adjustments in income measures to take into account the local price levels are immediately relevant for economic analysis such as inequality studies, wage comparisons and assessments of poverty. Other objective of this Chapter 3 is to use this applied micro-economic research to evaluate the impact of price differentials across Spanish regions over welfare measures, emphasizing in poverty analysis. What is intended to do in this chapter is to provide information on how the incidence of poverty has been different across regions and how this poverty incidence is affected when the regional costs of living differentials are taking into account.

Agglomeration occurs at many geographical scales in diverse ways. But strong regional disparities within the same country imply the existence of agglomerations at another spatial scale. This issue has been the center of the Chapter 3. Despite the logic and the economic rigor of the results obtained in this chapter there are some critical remarks which weaken, in some, way, the conclusions established: regional agglomerations are also reflected in the diversity of the cities. Spain, as it said before, is characterized by a complete urban system formed by big metropolitan areas, several medium-size cities, which at the same time are classified as central and peripherical cities, all surrounding by an important extension of rural areas.

The data of the HBS does not permit identify all these kinds of cities and calculate their costs of living, something that would give us a different and more accurate view of spatial differences in the costs of living. Despite this, it is tried to exploit the HBS data and searches its strengths. The strengths that offer the survey are exploited in the Chapter 4 to studying an important topic in the Regional and Urban Economics that is the effects of the agglomerations over the consumption patterns and the costs of living. Indeed, the data allow estimating the differences in the consumption behavior and prices both between urban and rural areas and along all the urban structure of the country.

The first aim of the Chapter 4 is to propose an analysis of the consumption patterns of households along the different size of the cities and in the context of a developed country where the urbanization process it is supposed to be completed. This analysis will be made through the estimation of demand elasticities in the context of the AIDS methodology.

Theoretical literature gives us some important clues to explain why the size of the cities could affect prices and consumption patterns. Standard models of the New Economic Geography (NEG) predict that the costs of living are lower in the core than in the periphery, this is the case of the Krugman's (1991) model as well as other NEG models. But in reality the costs of living tend to be high in the agglomerations areas due to the land scarcity and the higher housing prices (Tabuchi, 2001). But what tell about the empirical evidence? Most of the applications concerned in evaluating city size effect conclude that there is a positive relationship between the city size and the cost of living. But all these researches use as costs of living the official CPIs or other kinds of price data generally provided by official statistical agencies, neither of them apply strictly the theory of the "true" cost of living index. The second aim of the Chapter 4 is to calculate a "true" COLI for the different city sizes in Spain following the methodology applied in the Chapter 3 for calculating the regional COLI. These estimations will provide a comparison of the costs of living between urban and rural areas as well as along all the urban hierarchy including the medium-sized cities. But the question remains as to whether there may be some common factors that contribute to the explanation of the spatial differences in the costs of living.

The third aim of this Chapter 4 is to develop a model which is capable to explain the determinants of the cost of living in a place. Because the biggest cities attract a particular population with particular characteristics this make that a particular consumption patterns take place in agglomerations, in turn, this consumer behavior makes that the demand of certain goods rises exerting a pressure over prices of all goods and services. This process makes that agglomerations are more expensive to live in. But at the same time it could be observed that individuals with the same characteristics have a different consumption behavior by the fact those agglomerations promote a particular consumption which is not found in small areas. To contrast this hypothesis a micro-cost-of-living will be calculated for each individual to regress it through a quantile regression over several regional, demographic and socio-economic variables. The objective of this exercise is to isolate the pure effect of agglomerations over the cost of living by controlling for individuals' and households' characteristics.

In the last chapter of this thesis, Chapter 5 it will be summarized the main lessons to be drawn and, also, lay out several implications for the application and effectiveness of the regional policies. The results obtained in this study encourage us to continue with this research and thus future lines of research will be proposed at the end of this Chapter 5.

REFERENCES

Deaton, A., & Muellbauer, J. (1980). An Almost Ideal Demand System. *The American Economic Review, 70*(3).

Konus, A. A. (1939). The Problem of the True Index of the Cost of Living. *Econometrica, 7*(1), 10-29.

Krugman, P. (1991). Increasing returns and econmic geography. *Journal of Political Economy, 99*, 483-499.

Tabuchi, T. (2001). On interregional price differentials. *The Japanese Economic Review, 52*(1).

A mis padres,

Bernardo y Ana María

RESUMEN EN ESPAÑOL

Desde mediados del siglo pasado los investigadores especializados en la Ciencia Regional han estado estudiando el efecto de las aglomeraciones, generando una rica literatura, tanto teórica como empírica, que ha evidenciado los efectos que tienen sobre toda actividad económica las diferencias regionales y las aglomeraciones. El estudio de las diferencias regionales y las economías de aglomeración en España van a ser el centro de esta tesis. En este contexto, las economías de aglomeración van a ser uno de los conceptos clave, entendiéndose por economías de aglomeración aquellas que se generan cuando la población se concentra en un espacio, generando así una serie de efectos externos que son esenciales para entender las dinámicas espaciales en una economía. Estos efectos externos son, por ejemplo, la escasez de suelo urbano o la congestión. Existe una gran extensión de literatura económica dedicada al estudio de cómo las aglomeraciones afectan a las dinámicas de crecimiento e innovación y a la generación de especialización productiva, pero muy poco se ha estudiado acerca de cómo las economías de aglomeración afectan al coste de vida y al bienestar.

Si las economías de aglomeración son capaces de explicar por qué las actividades económicas se localizan en un lugar determinado o por qué las ciudades grandes siguen diferentes patrones de comportamiento que las ciudades medianas o los entornos rurales, parece lógico pensar que las economías de aglomeración serán un concepto clave a la hora de estudiar las dinámicas espaciales de los patrones de consumo, los precios y el coste de vida. Hay estudios empíricos previos que sugieren que las características generales del lugar de residencia, y en particular el tamaño de la ciudad de residencia afectan a los patrones de consumo. Esta afirmación puede parecer lógica, ya que las ciudades más grandes ofrecen una gran variedad de productos y servicios, que a su vez son normalmente de una mayor calidad de los que se encuentran en ciudades más pequeñas haciendo que se genere un estilo de vida diferente. Como resultado, los comportamientos de los consumidores serán diferentes en estas grandes ciudades a los que se observan en las zonas rurales. Además, se espera que todas estas características de las grandes metrópolis hagan que el coste de vida sea mayor en ellas.

Aunque las disparidades regionales en el coste de vida juegan un importante papel en el estudio de la economía regional y en la aplicación de políticas regionales, en general, no existen datos oficiales sobre costes de vida a nivel espacial. En España, esta falta de datos es especialmente importante, ya que las características del país impulsan a tener en cuenta el espacio y la geografía a la hora de estudiar las dinámicas económicas. Aunque España no es un país especialmente grande, sí está caracterizado por diferencias bastante grandes en los aspectos culturales, sociológicos y climatológicos. Además, España es un país particularmente atractivo para este tipo de estudios ya que comprende todo tipo de ciudades, desde grandes metrópolis, como Madrid y Barcelona, hasta numerosas ciudades medianas todas ellas rodeadas por áreas muy rurales. Todo este entramado urbano tan completo se caracteriza además por comprender tanto ciudades centrales como periféricas, ciudades costeras y turísticas como ciudades interiores, además de la especial situación de las Islas Baleares y las Islas Canarias.

España es un sistema federal *de facto* en donde la mayoría de las políticas que afectan al bienestar de los ciudadanos son responsabilidad de los gobiernos regionales y locales. Debido a la falta de información sobre índices de coste de vida a nivel espacial, estas políticas se toman apoyándose en indicadores de renta a nivel nominal. Ajustar estos indicadores de bienestar teniendo en cuenta posibles diferencias en el coste de vida resulta una cuestión interesante para el estudio de las desigualdades regionales, comparaciones de salarios o análisis de pobreza. Por todo esto, esta tesis va a tratar de proporcionar nueva información empírica acerca de los costes de vida a nivel espacial, basándose en la teoría económica del consumidor y en técnicas micro-econométricas que estimarán y cuantificarán las variaciones espaciales en el coste de vida en España.

En el Capítulo 2 de esta tesis se explicará de forma sucinta la metodología aplicada para la construcción de Índices de Coste de Vida Espaciales. La idea central que subyace en la metodología aplicada es que un “verdadero” índice de coste de vida debe comparar el coste monetario de adquirir dos cestas de bienes diferentes conectadas únicamente por la condición de que proporcionan el mismo nivel de bienestar (Konus, 1939). Este enfoque entraña muchas dificultades computacionales debido a que la función de utilidad de un consumidor no es directamente observable, por eso se va a aplicar una metodología específica llamada Almost Ideal Demand System (Sistema de Demanda Cuasi-Ideal) desarrollada por Deaton y Muellbauer (1980). Para aplicar esta metodología se usará la Encuesta de Presupuestos Familiares publicada por el Instituto Nacional de Estadística. Esta encuesta recoge hábitos de consumo y características socioeconómicas de los hogares españoles.

Como se ha señalado anteriormente, las disparidades en el coste de vida a nivel espacial juegan un importante papel en la Economía Regional y Urbana. A pesar de esto no existe información de niveles de precios en España, y, en general, en ningún país Europeo. Los esfuerzos y los datos disponibles son destinados a la construcción del Índice de Precios al Consumo (IPC), pero este índice es simplemente un indicador de inflación que no permite comparaciones de precios entre regiones, solo permite conocer la evolución de los precios de una cesta de la compra tipo en una región a lo largo de un periodo de tiempo. En el Capítulo 3 de la tesis se construirá el Índice de Coste de Vida Espacial considerando como regiones las Comunidades Autónomas Españolas y para el periodo 2008-2012. Este índice permitirá realizar comparaciones de costes de vida tanto a lo largo del tiempo como a través del espacio y su importancia radica, además de que va a poner de relieve las disparidades entre Comunidades Autónomas, en que se utilizará para la el estudio de las diferencias en la distribución de renta entre las regiones en España. En particular, es importante centrar la atención en el análisis de la pobreza ya que, en España y debido a la grave crisis económica experimentada en años recientes, ésta ha aumentado considerablemente tanto a nivel nacional como en cada una de las regiones. Lo que se pretende en esta sección de este capítulo es proporcionar información sobre cómo la incidencia de la pobreza es diferente entre regiones y cómo se ve afectada cuando se tienen en cuenta las diferencias de coste de vida regionales.

Las aglomeraciones se producen a diversas escalas geográficas. Las disparidades regionales dentro de un mismo país hacen que se produzcan aglomeraciones a una escala geográfica más pequeña que las regiones. Este será el centro del Capítulo 4 de esta tesis. Si bien los resultados obtenidos en el Capítulo 3 permiten llevar a cabo comparaciones bajo una perspectiva espacial, en este análisis deberían tenerse en cuenta igualmente las consecuencias que se derivan de la existencia de una diversidad muy rica de ciudades dentro de cada Comunidad Autónoma. Los datos usados no proporcionan información a una escala más desagregada que las Comunidades Autónomas, por lo que no es posible calcular el Índice de Coste de Vida Espacial a nivel provincial o a nivel municipal, lo que nos daría una panorámica más detallada del coste de vida en España. A pesar de ello, se explotará la información de la encuesta y se aprovechará toda la información disponible en ella para poder estudiar uno de los más importantes asuntos en la Economía Regional y Urbana: los efectos de las aglomeraciones sobre el consumo y el coste de vida. En particular, la información proporcionada en la encuesta nos permitirá cuantificar las diferencias en coste de vida para las zonas rurales y las zonas urbanas así como para toda la estructura urbana del país.

Más específicamente, el primer objetivo de este capítulo será el análisis de los patrones de consumo a lo largo de los distintos tamaños de ciudad a través de la estimación de las elasticidades precio y renta de la demanda. La literatura teórica nos ofrece varias explicaciones de cómo el tamaño de la ciudad puede afectar a los precios y a la conducta del consumidor. Los modelos estándar de la Nueva Geografía Económica (NEG) predicen que los costes de vida son menores en las ciudades que en la periferia, este el caso del modelo de Krugman (1991) así como de otros modelos de la NEG. Sin embargo, otros modelos teóricos predicen que los costes de vida tienden a ser mayores en las aglomeraciones debido a la escasez de suelo urbano y los costes de vivienda más altos (Tabuchi, 2001). Ante esta controversia hay que acercarse a la literatura empírica para saber qué se concluye acerca del efecto de las aglomeraciones sobre el coste de vida. La mayoría de los trabajos empíricos han sido realizados para los Estados Unidos y la mayoría de ellos concluyen que hay una relación positiva entre el tamaño de la ciudad y el coste de vida. El segundo objetivo de este capítulo es calcular un índice de coste de vida para cada tamaño de ciudad según la clasificación de las mismas que se recoge en la Encuesta de Presupuestos Familiares. Una vez calculado este índice y observadas las diferencias notables entre el coste de vida estimado para las ciudades grandes frente a las áreas más pequeñas, una pregunta natural que surge es: ¿qué es lo que hace que existan diferencias persistentes en el coste de vida? El tercer objetivo de este capítulo es contestar a esta pregunta a través de la estimación de un modelo que explique los factores determinantes del coste de vida en España. Debido a que las grandes ciudades atraen a una población con unas características particulares (mayor renta, mayor nivel educativo, etc.), en éstas se producirán unos patrones de consumo que no se darán en las zonas menos urbanas haciendo que la demanda de ciertos productos crezca y como consecuencia de la presión sobre la demanda los precios sean más altos. Este proceso hace que las aglomeraciones sean más caras, pero al mismo tiempo puede ocurrir que existan características intrínsecas a las aglomeraciones que hacen que los precios sean mayores en éstas. En otras palabras, un mismo individuo con unas características particulares puede tener un comportamiento diferente en las aglomeraciones que en las zonas rurales. Para contrastar estas hipótesis se calculará el coste de vida de cada hogar de la muestra para un mismo nivel de utilidad, que será la variable dependiente de un modelo de regresión en el que se incluirán tanto variables geográficas y demográficas como variables socioeconómicas del hogar. En concreto, las variables geográficas incluirán variables dummy de regiones y una variable que representará el efecto de las aglomeraciones sobre el coste de vida. Respecto a las variables socioeconómicas particulares de cada hogar se incluirán aquellas relativas a las características del hogar y de los individuoss. El propósito de incluir estas variables es el de aislar el efecto puro de las aglomeraciones sobre el coste de vida a través del control de las características de los hogares e individuos.

Por último, en el Capítulo 5, se expondrán las principales conclusiones derivadas de los resultados obtenidos en esta tesis, se plasmarán las implicaciones de política económica y se propondrán futuras líneas de investigación derivadas de esta investigación.

TABLE OF CONTENTS

1 INTRODUCTION

Since the 40s of the 20th century regional scientists have been working on agglomeration economies. This research has generated a rich world of theoretical literature and empirical evidence clearly showing the regional differences and agglomerations affect all kinds of economic activities. Regional differences and agglomeration economies are a crucial issue that will be the center of this thesis. Agglomeration economies are one of the key concepts in this context; as the population becomes more concentrated into populated urban spaces a series of external effects essential for understanding spatial dynamics, like urban land constrains, commuting costs or interregional transportation costs, are triggered off. There exists a large body of literature devoted to study how agglomerations affect the dynamics of growth and innovation, trends in productive specialization or processes of spatial concentration of economic activity but very little has been done on how agglomerations determines the standard of living.

If agglomeration economies are capable of explaining why economic activities locates in some places or explain why large cities follow paths different from those of medium-sized or small rural areas, it would seem logical that the agglomeration economies is a fundamental concept to understand the spatial dynamics of other aspects such as consumption patterns, price dynamics and costs of living. There are previous empirical studies that suggest that the place of residence, in an urban or rural area even depending on the size of the city, affects consumption behavior. Large cities offer a greater variety and higher quality of goods attracting people with particular characteristics and generating a different style of life. As a result the consumption patterns generated in the metropolis are different than those generated in rural areas. Moreover, is also hopped that local amenities founded in metropolitan areas makes that the cost of living is higher.

Although disparities in the cost of living across the space play a crucial role in regional economics and regional policy, in general, comprehensive regional cost of living data are not available from a governmental source. In Spain, this is particularly relevant due to the characteristics of the country. Since an administrative point of view, Spain is divided into seventeen Autonomous Communities (NUTS-II) which include several provinces (NUTS-III), each province in turn divided into municipalities. Although is not a big country, is characterized by striking cultural, sociological and climatological differences both between Autonomous Communities and within Autonomous Communities. Spain regions comprise all kinds of cities, since the biggest ones to very rural areas and also characterized by the existence of much more medium-size cities in central and peripheral positions as well as coastal and inland positions, all of them surrounding by very rural areas.

Spain is a *de facto* federal system in which most of the welfare and developing programs have become the responsibility of regional and local governments. On account of the lack of information, regional studies have to rely on nominal indicators. Adjustments in income measures to take into account the local cost of living are immediately relevant for economic analysis such as inequality studies, wage comparisons and assessments of poverty. The rationale of this thesis is to provide new information about regional costs of living in Spain based on the demand theory and on the micro-econometric analysis. In Spain fewer studies, specifically, Alberola and Marqués (2001) and Garrido-Yserte *et al*. (2012), have pointed out the relevance of regional price disparities and their persistence over time. But neither calculates costs of living in a microeconomic framework; the formers use the available Consumer Price Index (CPI) official data to evidence regional and persistent price differences, and the latters proposed a cost of living index simply incorporating the cost of acquiring or living in owned housing to the CPI. Also, disparities in well-being are evidenced by the Spanish Autonomous Communities by some authors, such as Ayala *et al*. (2011), García-Luque *et al*. (2009), Jurado and Pérez-Mayo (2012), Pérez-Mayo (2008), or Poggi (2007).

The aim of this thesis is to estimate spatial cost of living indices with a microeconomic approach. In the Chapter 2 the methodology to construct these indices will be explained in detail. The methodology followed along this thesis is based on Konus (1939) who raised the problem of the “true index of cost of living”. This idea emphasized in when computing a “true index of cost of living” it should be compared the monetary cost of two different basket of goods which are connected by the solely condition that they provided the same standard of living. This approach entails computational problems due to the impossibility of knowing directly the utility function. To address this problem it will be used the methodology developed by Deaton and Muellbauer (1980) called Almost Ideal Demand System (AIDS). This methodology permits the estimation of a flexible demand system with several properties without the necessity of know the form of the utility function. Their approach has important advantages over other procedures; the most important of these being the fact that it considers non-homothetic preferences for each household income group. This makes it especially valuable for comparisons across space, given that it allows the basket of preferences between consumers of different regions or areas to be varied. With this approach we can know the not observable utility function, and, consequently, to derive the cost function and to calculate a cost of living index (COLI) by fixing the utility level instead the basket of consumer goods as in the case of the CPIs calculated by statistical agencies.

For estimating the AIDS we use expenditure data from the Spanish Household Budget Survey (HBS) provided by the National Statistical Institute (INE). This survey record the expenditure of many items but it has information neither about quantities purchased nor about prices except for the food and energy groups of goods. Because the COLI estimates require information about prices faced by consumers we are only able to calculate the COLI for food products and excluding the energy groups because the aforementioned group only represents a 4% of the total expenditure of a Spanish household. Despite this limitation we can obtain the advantage of knowing regional price differences in dairy products which are supposed to have less price dispersion than others. Differences in price levels are obvious in goods such as housing, but the critical question is whether the dispersion in other representative consumer goods is pervasive and of sufficient magnitude to influence households' costs of living significantly (Slesnick, 2002).

Although disparities in cost of living across the space play a crucial role in Regional and Urban Economics, knowledge on regional price levels is scarce in all European countries, especially in Spain. The price data are usually conducted to construct the CPI at the national or regional level but the CPI is only an inflation measure and does not allow for interregional price comparisons. All this leads to the necessity of constructing a Spatial Index of the Cost of Living (SCOLI). This is the purpose of Chapter 3 of this thesis: to construct a SCOLI at the regional level considering as regions the administrative Autonomous Communities of Spain. This index will allow knowing the existence and magnitude of the differences in cost of living across Spanish regions in year 2012, the last year of available data in the HBS. Moreover, a temporal SCOLI will also be calculated in this chapter for the period 2008-2012. This spatio-temporal index will allow making comparisons both along the time and across the space, simultaneously. Along al the chapter comparisons of these indices with the CPI will be made in order to highlights the strengths of the indices and the limitations of the CPI.

An important issue derived to the lack of price information is that regional studies have to rely on nominal indicators. Adjustments in income measures to take into account the local price levels are immediately relevant for economic analysis such as inequality studies, wage comparisons and assessments of poverty. Other objective of this Chapter 3 is to use this applied micro-economic research to evaluate the impact of price differentials across Spanish regions over welfare measures, emphasizing in poverty analysis.

The construction of spatial deflators is related to the literature on poverty lines. The importance of having a proper index for comparisons of poverty across the space has been underlined by Ravallion and Van de Walle (1991), Booth (1993), Nord and Cook (1995), Asra (1999) and Jolliffe (2006). It is important to focus the attention in poverty because in the last years the poverty has been aggravated in Spain due to the strong economic crisis. This crisis has made that poverty increases both in the whole country and in each of the regions. What is intended to do in this chapter is to provide information on how the incidence of poverty has been different across regions and how this poverty incidence is affected when the regional costs of living differentials are taking into account. This objective will be accomplished taking as reference the work of Herrero *et al*. (2013) and deflating the poverty lines not only in a temporal way, but also in a spatial fashion.

Agglomeration occurs at many geographical scales in diverse ways, strong regional disparities within the same country imply the existence of agglomerations at another spatial scale. This issue has been the center of the Chapter 3. Despite the logic and the economic rigor of the results obtained in this chapter there are some critical remarks which weaken, in some, way, the conclusions established: regional agglomerations are also reflected in the diversity of the cities. Although within most developed countries the urban hierarchy is rather stable, in Spain, instead, this does not occurs. Spain, as it said before, is characterized by a complete urban system formed by big metropolitan areas, several medium-size cities, which at the same time are classified as central and peripheral cities, all surrounding by an important extension of rural areas.

The data of the HBS does not permit identify all these kinds of cities and calculate their costs of living, something that would give us a different and more accurate view of spatial differences in the costs of living. Despite this, it is tried to exploit the HBS data and searches its strengths. The strengths that offer the survey are exploited in the Chapter 4 to studying an important topic in the Regional and Urban Economics that is the effects of the agglomerations over the consumption patterns and the costs of living. Indeed, the data allow estimating the differences in the consumption behavior and prices both between urban and rural areas and along all the urban structure of the country.

There are previous empirical studies that suggest that the place of residence and even depending on the size of the city affects consumer behavior (Burney and Akmal, 1991; Lewis and Andrews, 1998; Abdulai *et al*., 1999; Alfonzo and Peterson, 2006; and Haq *et al*., 2008) but all of them are focused in the context of a developing country. The first aim of the Chapter 4 is to propose an analysis of the consumption patterns of households along the different size of the cities and in the context of a developed country where the urbanization process it is supposed to be completed. This analysis will be made through the estimation of demand elasticities in the context of the AIDS methodology.

Theoretical literature gives us some important clues to explain why the size of the cities could affect prices and consumption patterns. Standard models of the New Economic Geography (NEG) predict that the costs of living are lower in the core than in the periphery, this is the case of the Krugman's (1991) model as well as other NEG models. But in reality the costs of living tend to be high in the agglomerations areas due to the land scarcity and the higher housing prices (Tabuchi, 2001)[1]. But what tell about the empirical evidence? Most of the applications concerned in evaluating city size effect conclude that there is a positive relationship between the city size and the cost of living. The majority of these applications are made for the US (Alonso and Fajans, 1970; Haworth and Rasmussen, 1973; Simon and Love, 1990; Walden, 1998; Kurre, 2003; and Cebula and Todd, 2004); fewer for the UE like Blien *et al*. (2009) and Roos (2006); and others for developing countries like Asra (1999), Katwani and Hill (2002) and Majumber *et al*. (2012).

But all these quoted researches use as costs of living the official CPIs or other kinds of price data generally provided by official statistical agencies, neither of them apply strictly the theory of the "true" cost of living index. The second aim of the Chapter 4 is to calculate a "true" COLI for the different city sizes in Spain following the methodology applied in the Chapter 3 for calculating the regional COLI. These estimations will provide a comparison of the costs of living between urban and rural areas as well as along all the urban hierarchy including the medium-sized cities. But the question remains as to whether there may be some common factors that contribute to the explanation of the spatial differences in the costs of living.

The third aim of this Chapter 4 is to develop a model which is capable to explain the determinants of the cost of living in a place. To do this it is necessary to move away from a scenario where a representative consumer of each place is chosen to a scenario where the characteristic of all households and individuals are considered. Because the biggest cities attract a particular population with particular characteristics this make that a particular consumption patterns take place in agglomerations, in turn, this consumer behavior makes that the demand of certain goods rises exerting a pressure over prices of all goods and services. This process makes that agglomerations are more expensive to live in. But at the same time it could be observed that individuals with the same characteristics have a different consumption behavior by the fact those agglomerations promote a particular consumption which is not found in small areas. To contrast this hypothesis a micro-cost-of-living will be calculated for each individual to regress it through a

[1] A more detailed debate about the theoretical discussion of the NEG models is developed in the Chapter 4.

quantile regression over several regional, demographic and socio-economic variables. The objective of this exercise is to isolate the pure effect of agglomerations over the cost of living by controlling for individuals' and households' characteristics.

In the last chapter of this thesis, Chapter 5, named *Summary and Principal Findings*, it will be summarized the main lessons to be drawn and, also, lay out several implications for the application and effectiveness of the regional policies. The results obtained in this study encourage us to continue with this research and thus future lines of research will be proposed at the end of this Chapter 5.

2 METHODOLOGY

2.1 Introduction

The theory of the "true" Cost of Living Index (COLI) was first developed by Konüs (1939). He defines a "true index of the cost of living" as the ratio of the Cost of Living (COL) at one period of time and the COL at the other period. The expression would be:

$$\boldsymbol{COLI_{ij}} = \frac{\boldsymbol{COL_i}}{\boldsymbol{COL_j}} \tag{1}$$

Where $\boldsymbol{COL_i}$ and $\boldsymbol{COL_j}$ are the "cost of living" in period *i* and *j*, respectively. Its mean the monetary value of the goods consumed in the period *i* and *j* of time by a household which are necessary for the maintenance of a certain standard of living. Thus in computing a true COLI it is compared the monetary cost of two different combinations of goods which are connected solely by the condition that, during the consumption of the two combinations, the standard of living provided by both is exactly the same.

However, the usual method of calculating the COLI is the so-called method of aggregates. This consists on calculate the cost of a given basket of goods corresponding to the average or normal consumption and at prices prevailing at a given time, and dividing it by the cost of the same basket of goods at prices of another period. But this index does not show exactly the changes in the COL because there is the assumption that while prices change consumption does not change. But, in reality, consumers change its consumption due to rises and falls in prices in order to maintain its standard of living.

In order to construct a "true" index of cost of living it is necessary to know which combination of goods yield a given standard of living despite price changes. For this purpose it is used the concept of *indirect utility function,* the consumer is going to maximize its utility function at a given prices and subject to a budget restriction. The reformulation of the COLI would be:

$$\boldsymbol{COLI_{ij}} = \frac{\boldsymbol{c(\bar{p}_i, u)}}{\boldsymbol{c(\bar{p}_j, u)}} \tag{2}$$

Where $\bar{\boldsymbol{p}}_{\boldsymbol{i}}$ and $\bar{\boldsymbol{p}}_{\boldsymbol{j}}$ are price vectors in the periods *i* and *j,* respectively. Where *u* is he utility function to be reached by the consumer, and *c* is the cost of attaining the utility level *u* at prices $\bar{\boldsymbol{p}}_{\boldsymbol{i}}$ and $\bar{\boldsymbol{p}}_{\boldsymbol{j}}$.

The major problem arises from the unknown and not observable utility function, and without knowing the utility function is impossible to derive the cost function and to calculate the COLI. The typical solution to address this problem is to follow a flexible function demand system with several convenient properties. These flexible functional forms permit the estimation of demand equations without knowing explicitly the functional form of the utility function. The flexible functional form to be used in this research will be the Almost Ideal Demand System (AIDS) proposed by Deaton and Muellbauer (1980). The model possesses a set of advantages which take place simultaneously. As Deaton (1980) describe in his work: "... it gives an arbitrary first-order approximation to any demand system; it satisfies the axioms of choice exactly; it aggregates perfectly over consumers without invoking parallel linear Engel curves; it has a functional form which is consistent with known household-budget data; it is simple to estimate, largely avoiding the need for non-linear estimation; and, it can be used to test the restrictions of homogeneity and symmetry through linear restrictions on fixed parameters".

2.2 The Almost Ideal Demand System (AIDS)

2.2.1 Specification of the AIDS

The conventional consumer choice is to maximize a utility function $u(q)$ subject to the budget constraint $p'q \leq x$ where *q* is a vector of quantities of goods and services, *p* is a vector of prices of the goods and services, and *x* is the income available to the consumer. $u(q)$ is the utility function which is increasing with q and is quasi-concave.

The solution to this maximization problem gives the marshallian demand equations q=*q(x, p)* which are homogeneous of degree zero in income and prices.

Substituing the Marshallian functions into the utility function gives the indirect utility function:

$$u = \varphi(x, p) \tag{3}$$

Which on solving for *x* in terms of *u* and *p* yields the cost or expenditure function:

$$x = c(u, p) \tag{4}$$

Which is interpreted as the minimum cost of buying the utility level *u* at the price vector *p.*

This function has the following properties:

- $c(u, p)$ is an increasing function of *u* for all *p*
- $c(u, p)$ is an increasing and concave function of prices for all *u*
- $c(u, p)$ is positively linear homogeneous in *p* for every *u*

The starting point is from a specific class of preferences which permit exact aggregation over consumers. These preferences are known as the PIGLOG class and represented through the cost or expenditure function proposed by Muellbauer (1975) and named price independent generalized logarithmic cost function (PIGLOG). This can be written:

$$log\, c(p, u) = (1 - u)\, log\big(a(p)\big) + u\, log\big(b(p)\big) \tag{5}$$

Where *c* is the expenditure function, *p* is the price vector and *u* is the utility level, *u* can take the value 0 which represents the subsistence level and the value 1 which represents the bliss level, so *log(a(p))* and *log(b(p))* can be considered as the log of the costs of subsistence and bliss, respectively. Their respective functional forms are:

$$log\big(a(p)\big) = \alpha_0 + \sum_{i=1}^{n} \alpha_i\, log\, p_i + \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \gamma_{ij}\, log\, p_i\, log\, p_j \tag{6}$$

$$log\big(b(p)\big) = log\big(a(p)\big) + \beta_0 \prod_i {p_i}^{\beta_i} \tag{7}$$

So the AIDS cost function is written:

$$log\, c(p,u) = \alpha_0 + \sum_{i=1}^{n} \alpha_i \, log\, p_i + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \gamma_{ij} \, log\, p_i \, log\, p_j + u\beta_0 \prod_i {p_i}^{\beta_i} \quad (8)$$

Where α_i, β_i and γ_{ij} are parameters.

By applying the Shepard's lemma to (8) (i.e., price derivate are equal to the quantities demanded) and multiplying both sides of the equation by $p_i/c(u,p)$, we obtain:

$$\frac{\partial\, log\big(c(p,u)\big)}{\partial\, log\, p_i} = \frac{p_i q_i}{c(p,u)} = w_i \quad (9)$$

Where p_i is the price vector, $q_i$is a vector that represents quantities purchased and c(p,u) is the cost function From (9)we obtain w_i which is the budget share of good i:

$$w_i = \alpha_i + \sum_{j=1}^{n} \gamma_{ij} \, log\, p_j + \beta_i u \beta_0 \prod_i p_i^{\beta_i} \quad (10)$$

To obtain an estimable system we need to solve for u as a function of observed and known parameters from equation (8):

$$u = \frac{log\, c(u,p) - \alpha_0 - \sum_{i=1}^{n} \alpha_i \, log\, p_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \gamma_{ij} \, log\, p_i \, log\, p_j}{\beta_0 \prod_i {p_i}^{\beta_i}} \quad (11)$$

Substituting u in equation (10) we obtain:

$$w_i = \alpha_i + \sum_{j=1}^{n} \gamma_{ij} \, log\, p_j + \beta_i \Big(log\, c(p,u) - \alpha_0 - \sum_{i=1}^{n} \alpha_i \, log\, p_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \gamma_{ij} \, log\, p_i \, log\, p_j\Big) \quad (12)$$

The shares in (12) are determined from prices and the expenditure function, plus a set of parameters to be estimated. These shares are the AIDS demand functions and they can be expressed as:

$$w_i = \alpha_i + \sum_{j=1}^{n} \gamma_{ij} \log p_j + \beta_i \log\{x/P\} \tag{13}$$

Where α, $\beta$and γ are the parameters to be estimated, x is the total expenditure on the food group and P is a price index defined as:

$$\log P = \alpha_0 + \sum_{j=1}^{n} \alpha_j \log p_j + \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \gamma_{ij} \log p_i \log p_j \tag{14}$$

The restrictions on the parameters of (10) imply restrictions on the parameters of the AIDS equation (13). Firstly, they must hold the adding-up restriction ($\sum_{i=1}^{n} w_i = 1$), which requires equality of the sum of individual commodity expenditures and the total expenditures:

$$\sum_{i=1}^{n} \alpha_i = 1, \quad \sum_{i=1}^{n} \gamma_{ij} = 0, \quad \sum_{i=1}^{n} \beta_i = 0 \tag{15}$$

Furthermore, the equations of the AIDS model are homogeneous of degree zero in prices and total expenditure taken together. This means that if prices and total expenditure increase by the same amount the demand remains unchanged:

$$\sum_{j=1}^{n} \gamma_{ji} = 0, \tag{16}$$

Moreover, the total expenditure must verify the Slutsky symmetry, which requires that the compensated cross-price derivative of commodity *i* with respect to commodity *j* equals the compensated cross-price derivative of commodity *j* with respect to commodity *i*:

$$\gamma_{ij} = \gamma_{ji} \tag{17}$$

The β and γ parameters can be interpreted in economic terms. The γ_{ij} elements quantify the effect of changes in relative prices, representing the percentage of change on the *i*th budget share produced by a 1% increase in the price of the *j*th product, being (x/P) held constant. The effects of changes in the real expenditure operate through the β_i coefficients, which are positive for luxuries and negative for necessities, (Deaton and Muellbauer, 1980).

2.2.2 Generality of the model

The flexible functional form property of the AIDS cost function implies that demand functions derived from it are first order approximations to any set of demand functions derived from utility maximizing behavior. The AIDS is thus as general as other flexible functional forms such as the translog or the Rotterdam models. However, if maximizing behavior is not assumed and it is simply held that demand are continuous functions of budget and prices, then the AIDS demand functions without the restrictions (16) and (17) can still provide a first order approximation. Without maximizing assumptions the budget shares are considered unknown functions of *logp* and *logx*. From (13) and (14) the AIDS has derivatives $\partial w_i/\partial logx = \beta_i$ and $\partial w_i/\partial logp_j = \gamma_{ij} - \beta_i\alpha_j - \beta_i \sum \gamma_{jk} logp_k$ so that β and γ can be chosen so that the derivatives of the AIDS will be identical to those of any true model. Given that the α parameters act as intercepts, the AIDS can provide a local first-order approximation to any true demand system derived from the theory of choice or not. This property is important because it means that tests of homogeneity of symmetry are set within a maintained hypothesis which makes sense and would be widely accepted it its own right.

But generality is not without shortcomings, there are a large number of parameters and on most data sets these are unlikely to be all well determined. In the AIDS this can be solved by placing whatever restrictions on γ_{ij} parameters are thought to be empirically or theoretically plausible.

2.2.3 Restrictions

Starting from equations (13) and (14) it can be examined the effects of the restrictions (15)-(17) which are required to make the model consistent with the theory of demand. The condition (15) is the *adding up* restriction which ensures that $\sum w_i = 1$. *Homogeneity* of the demand functions requires restriction (16) which can be tested equation by equation. *Slutsky symmetry* is satisfied if and only if the symmetry restriction holds. As in other flexible functional forms, *negativity* cannot be ensured by any restrictions of the parameters alone. It can be checked by calculating the eigenvalues of the Slustky matrix s_{ij}. In practice, it is easier to use instead s_{ij}, $k_{ij} = p_i p_j s_{ij}/x$ with eigenvalues with the same sign of s_{ij} and given by:

$$\boldsymbol{k_{ij} = \gamma_{ij} + \beta_i \beta_j log\frac{x}{P} - w_i \delta_{ij} + w_i w_j} \tag{18}$$

Where δ_{ij} is the Kronecker delta. Apart from this negativity condition, all the restrictions are expressible as linear constraints involving only the parameters and so can be imposed globally by standard techniques.

2.2.4 Estimation

Estimation can be done by substituting (14) in (13):

$$\boldsymbol{w_i = \alpha_i + \sum_{j=1}^{n} \gamma_{ij} \log p_j + \beta_i \left\{ logx - \sum_k \alpha_k \, logp_k - \frac{1}{2} \sum_k \sum_j \gamma_{kj} logp_k logp_j \right\}} \tag{19}$$

Although all the parameters in (19) are identified given sufficient variation in the independent variables, in many examples the practical identification of $\boldsymbol{\alpha_0}$ is likely to be problematical. This parameter only identified from the $\boldsymbol{\alpha_i}$s in (19) by the presence of these latter inside the term in braces, originally the formula for $\boldsymbol{log\,P}$ in (14). However, in situations where individual prices are closely collinear, $\boldsymbol{log\,P}$ is unlikely to be very sensitive to its weights so that changes in the intercept term in (19) due to variations in $\boldsymbol{\alpha_0}$ can be offset in the $\boldsymbol{\alpha_i}$s with minimal effect on $\boldsymbol{log\,P}$. This can be overcome in practice by assigning a value to $\boldsymbol{\alpha_0}$ a priori. Since the parameter can be interpreted as the outlay required for a minimal standard of living when prices are unity choosing a plausible value is not very difficult.

In many situations it is possible to exploit the collinearity of the prices to get a much simpler estimation process. From (13) if prices P is known, the model would be linear in the parameters, β and γ. In situations where prices are closely collinear, it may be adequate to approximate P to some known index. One usual approximation is the Stone Price Index:

$$logP^* = \sum w_k \, logp_k \tag{20}$$

If (20) is substituted in (13) it is obtained the Linear approximation of the AIDS model, namely LAIDS. In this framework the α_i parameters are identified only up to a scalar multiple of β_i. However, it must be emphasized that the LAIDS model is only an approximation to the AIDS and will only be accurate in specific circumstances, albeit widely occur in time series estimation where usually prices are almost collinear.

2.3 Application to the Spanish Data

The data used in this analysis are obtained from the Household Budget Survey (HBS)[2] from the National Statistical Institute (INE), a survey that provides information about the patterns of consumption of Spanish households, income and other socioeconomic and demographic characteristics at the household level. The dataset is formed by more than 20,000 observations which are disaggregated across the 17 regions at the NUTS-II level.

The HBS is composed by three data files: the Household file, which provided information about the Autonomous Community where the household belongs to, the city size, the density of population, the income level, number of members, number of employees, some characteristics about the household head like age, sex, education level, marital status, among others, and some information about the characteristics of the house; the Expenditure file which represents all households with any expenditure. The expenditure codes are shown at the maximum level of COICOP disaggregation (5 digits) and for each code are provided the expenditure made in this code and the quantity purchased in physic units; and, the Household members file where it can be found all the information about each member that compounds the household.

The variable that represents household income in the analysis is the total expenditure in the goods analyzed and not the total expenditure in all goods consumed, so the demand functions estimated are conditional demand functions, and the elasticities estimated are conditional elasticities (Pollack and Wales, 1992).

The individual household responses often provided a useful source of price data. Households not only report their expenditures in each good, but also the physical amount they bought. The estimation of the AIDS requires data on prices, physical

[2] http://www.ine.es/jaxi/menu.do?type=pcaxis&path=%2Ft25%2Fp458&file=inebase&L=0

quantities purchased and monetary household expenditure. Because all the prices must be observable to estimate the AIDS system, individual prices at which households purchase the commodities can be recovered by dividing monetary expenditures by physical quantities consumed, being these "prices" known as *unit values* (Deaton, 1988). Although the unit values may not be directly assumed as true market prices, taking them as proxy of prices is a widely accepted method for obtaining unitary prices because they depend on actual market prices. The attraction of such measure is the amount of data provided. But, by the other side, unit values are not the same thing as prices, and are affected by the choice of quality as well as by the actual prices that consumers face in the market (Deaton, 1997), but the empirical experience demonstrated that unit values are very useful as an indicator of time and spatial price variations.

Despite the source of price information derived of using unit values, the HBS still having limitations because not for all goods are reported quantities. Only for the Food and non- alcoholic beverages, Alcoholic beverages and cigarettes and Energy quantities are reported. The price of Alcoholic beverages and cigarettes is not being used in this analysis because their price are strongly regulated, respect to the Energy group, also it is going to be excluded because the energy expenditure only represents a 3% of the total household expenditure. They are used 64 food products at five digits level in the COICOP classification which are aggregate at three digits level sub-categories. Consequently, the AIDS is estimated for food products which are assigned to ten sub-groups belonging to the category of "Food and non-alcoholic beverages" in the HBS classification, namely: (1) Bread and cereals, (2) Meat, (3) Fish, (4) Milk, cheese and eggs, (5) Oil, (6) Fruits, (7) Vegetables, (8) Sugar, (9) Coffee, tea and cacao; and (10) Mineral water and other soft drinks. For each group i=1,…,10 the observed budget share w_i of equation (15) in each household is calculated by dividing the expenditure of the household in this specific group by the total household expenditure in food.

Due to the analysis is restringed to the food group, it supposed the hypothesis of *weakly separability* in the consumer's preferences as Fajardo *et al*. (1995) and Ramajo (1996) supposed for the same analysis for the Autonomous Community of Extremadura. This hypothesis implies that the food consumption is independent of the rest of consumption of other goods which do not belong to the food group. This means that the households make their budget process in two steps: first, the consumer allocates its income among the different expenditure groups, and then, the consumer reallocates the quantities assigned to each group among the goods belonging to each group. This implies that the marginal substitution rates among the goods belonging to the same group are independent of changes in relative prices out of the group.

An additional problem in the estimation process is the existence of households that report zero consumption of some type of product i. Consequently, prices are not available for all items in all households. This situation can happen when the consumed quantities are not reported by a household, or because the household do not really consume that specific group, being the consequence that the price of the item cannot be obtained by means of unit values. In both cases the price of the item is replaced by a geometric mean of the prices of this item in the same region and the same city size. So the price is replaced by the average price of the same item in in the region where the household is located and distinguishing by the size of the municipality which the household belongs to. Controlling by the effect of the size of the city is expected to produce more accurate estimates because prices in biggest cities are assumed different than prices in the rest of the cities. This is assumed because the biggest cities are normally the main cities of the region, which results in an agglomeration process of public and private services in this type of locations.

The modeling of demand systems with household-level microdata has the advantage of providing a large and statistically rich sample avoiding the problem of aggregation over consumers. In the other hand the modelization of the consumption patterns with detailed microdata is more complex due to the existence of zero in commodity purchases, especially when a very detailed classification for the commodities is used.

According to the description in Ramajo (1996), zero consumption in household surveys is due to the following reasons. The first one is because the existence of *corner solutions:* when a change in the relative price of two goods occurs and provokes that this price is superior to the substitution rate, the good with the highest price is not consumed. Also, there exist socio-demographic and geographic characteristics that can influence the preferences of the consumer to a particular good. Another factor could be that the consumer does not perceived the good like strictly economic, that is, its consumption does not depend on price or income, but depends on the characteristics of the consumers like, for example, the non-smokers, non-drivers or vegetarian consumers. And the third factor is when the consumer only buys the good infrequently so that over the limited period of the survey no purchase is recorded for some households. For example, due to the survey observe the households only for a short period, 7-14 days for foods; it is possible that household eats from food stocks.

Since a statistical point of view the zero consumption biases the estimation of the parameters of the model and it may produce a selection bias if we do not incorporate these observations into the estimation process. Dealing with censored data is more complicated in the case of demand systems than in a case of the econometric estimation of one single equation. The complication arises from the necessity of ensuring nonnegative estimates of the quantities consumed; the requirement of including the constraints imposed by economic theory; and the numerical problem of having to evaluate high-dimension cumulative density functions during the estimation (Dong *et al.*, 2004).

2.4 The AIDS with Censored Dependent Variable

2.4.1 The two step procedure of Shonkwiler and Yen

In the single equations applications with limited dependent variable, maximum-likelihood estimation of the Tobit model is common and straightforward. However, when censoring occurs in multiple equations the application of direct maximum-likelihood remains difficult.

Two general approaches have been used to estimate micro-level demand systems: the Kuhn-Tucker approach proposed by Wales and Woodland (1983) and associated dual approach suggested by Lee and Pitt (1986); and, the Amemiya-Tobin approach proposed by Amemiya (1974). Given the complexity of estimating these approaches, a number of alternative two-step models have been adopted for the estimation of censored demand systems (Heien and Wessells, 1990; Shonkwiler and Yen, 1999; Yen *et al*., 2002; and, Perali and Chavas, 2000).

To address these problems we will follow the two-step method proposed by Shonkwiler and Yen (1999), which improves the previous "favorite" two-step estimation procedure of Heien and Wessells (1990).

Consider the following system of equations with limited dependent variables:

$$\boldsymbol{y^*_{it} = f(x_{it}, \beta_i) + \epsilon_{it}}, \qquad \boldsymbol{d^*_{it} = z'_{it}\alpha_i + v_{it}} \tag{21}$$

$$\boldsymbol{d_{it} = \begin{Bmatrix} 1 \; if \; d^*_{it} > 0 \\ 0 \; if \; d^*_{it} = 0 \end{Bmatrix}}, \qquad \boldsymbol{y_{it} = d_{it} y^*_{it}}$$

Where for the ith equation and tth observation, y_{it} and d_{it} are the observed dependent variables, y^*_{it} and d^*_{it} are corresponding latent variables, x_{it} and z_{it} are vectors of exogenous variables, β_i and α_i are vectors of parameters, and ϵ_{it} and v_{it} are random errors.

The system of equations (21) implies that for the ith equation the single dependent variable y_{it} is observed with nonnegative values. If a nonnegligible proportion of its values are identically zero then it likely cannot be properly represented with a continuous distribution. Direct maximum-likelihood estimation of equation (21) is difficult when the error terms are allowed to be contemporaneous correlated. Alternatively, can be estimated using individual maximum-likelihood probit estimators when the v_{it} are normally distributed. This has led to consider two-step estimators which apply the probit estimator and then apply method-of-moments estimators to the observed y_{it} while augmenting its regressors with a correction factor obtained from the first step. Heien and Wessells (1990) proceed as follows: first, they obtain the probit estimates for each equation i based on the binary outcomes of d_{it}; and, second, they estimate with SUR the system with each equation augmented by an inverse Mills ratio defined as:

$$\phi(k_{it}\mathrm{z}'_{\boldsymbol{it}}\alpha_{\boldsymbol{i}})/\Phi(k_{it}\mathrm{z}'_{\boldsymbol{it}}\alpha_{\boldsymbol{i}}) \tag{22}$$

Where k_{it}=$2d_{it}$ – 1 and $\phi(.)$ and $\Phi(.)$ are univariate standard normal probability density function and cumulative function, respectively.

This procedure implies $\mathrm{E}(y_{it}|\mathrm{x}_{it}, \mathrm{z}_{it}; v_{it} > -\mathrm{z'}_{it}\alpha_i) = f(\mathrm{x}_{it}, \beta_i) + \delta_i[\phi(\mathrm{z'}_{it}\alpha_i)/\Phi(\mathrm{z'}_{it}\alpha_i)$ and $\mathrm{E}(y_{it}|\mathrm{x}_{it}, \mathrm{z}_{it}; v_{it} \leq -\mathrm{z'}_{it}\alpha_i) = f(\mathrm{x}_{it}, \beta_i) + \delta_i\{\frac{\phi(\mathrm{z'}_{it}\alpha_i)}{[1-\Phi(\mathrm{z'}_{it}\alpha_i)]}\}$ where the scalar δ_i is the coefficient of the correction factor of the *i*th equation in the second step. These in turn imply the unconditional expectation:

$$\mathrm{E}(y_{it}|\mathrm{x}_{it}, \mathrm{z}_{it}) = = f(\mathrm{x}_{it}, \beta_i) + 2\delta_i\, \phi(\mathrm{z'}_{it}\alpha_i) \tag{23}$$

Shonkwiler and Yen (1999) evidenced an inconsistency in the Heien and Wessells (1990) model demonstrated that considering expression (23) this implies $\mathrm{z'}_{it}\alpha_i \to -\infty$, namely, that the unconditional expectation of y_{it} is $f(\mathrm{x}_{it}, \beta_i)$, but the system in equation (21) suggests that as $\mathrm{z'}_{it}\alpha_i \to -\infty$, $y_{it} \to 0$.

To motivate an alternative two-step procedure Shonkwiler and Yen (1999) based it in the results of Lee (1976) and Wales and Woodland (1980). Then the conditional mean of y_{it} is:

$$\mathrm{E}(y_{it}|\mathrm{x}_{it}, \mathrm{z}_{it}; v_{it} > -\mathrm{z'}_{it}\alpha_i) = f(\mathrm{x}_{it}, \beta_i) + \delta_i \frac{\phi(\mathrm{z'}_{it}\alpha_i)}{\Phi(\mathrm{z'}_{it}\alpha_i)} \tag{24}$$

Because $\mathrm{E}(y_{it}|\mathrm{x}_{it}, \mathrm{z}_{it}; v_{it} \leq -\mathrm{z'}_{it}\alpha_i) = 0,$ the unconditional mean of y_{it} is:

$$\mathrm{E}(y_{it}|\mathrm{x}_{it}, \mathrm{z}_{it}) = \Phi(\mathrm{z'}_{it}\alpha_i) f(\mathrm{x}_{it}, \beta_i) + \delta_i\, \phi(\mathrm{z'}_{it}\alpha_i) \tag{25}$$

Which differs from expression (23). Based on this equation (25) for each *i*, the system of equations (21) is rewritten:

$$\begin{gathered} y_{it} = \Phi(\mathrm{z'}_{it}\alpha_i) f(\mathrm{x}_{it}, \beta_i) + \delta_i\, \phi(\mathrm{z'}_{it}\alpha_i) + \xi_{it} \\ (i = 1,2, \dots, m; t = 1,2, \dots, T) \end{gathered} \tag{26}$$

Where $\xi_{it} = y_{it} - \mathrm{E}(y_{it}|\mathrm{x}_{it}, \mathrm{z}_{it})$. The system (26) can be estimated by a two-step procedure using all observations: first, obtained the probit estimations of α_i using the binary outcome d_{it}=1 and d_{it}=0 for each *i;* and, second calculate $\Phi(\mathrm{z'}_{it}\alpha_i)$ and $\phi(\mathrm{z'}_{it}\alpha_i)$ and estimate β_i and δ_i in the system.

A large number of studies used the Heien and Wessells (1990) procedure. To demonstrate the magnitude of the inconsistency Shonkwiler and Yen (1999) made Monte Carlo simulations to compare they alternative estimator. Using a three equation linear system Monte Carlo results suggest that they proposed procedure performs well while the Heien and Wessells (1990) perform poorly.

2.4.2 The AIDS model with the Shonkwiler and Yen (1999) procedure

For the reasons evidenced above, in this thesis the Shonkwiler and Yen (1999) procedure is used. Adapting all the development described before to the AIDS model imply in the first step estimate a probit regression with a dependent binary variable that represents the household decision of consuming or not, which takes the value of 1 if the household purchases the commodity and the value of 0 if not, which depends on a set of socioeconomic variables that are used as regressors. The probit model determines the probability that a given household consumes a given good and it is used to estimate the normal cumulative distribution function (Φ) and the density function (ϕ). The second step includes the cumulative function $\Phi(x)$ as a scalar in the equations for shares, while the density function $\phi(x)$is included as an extra explanatory variable. The reformulation of the AIDS is:

$$w_i = \Phi(x)\left[\alpha_i + \sum_{j=1}^{n} \gamma_{ij} \, log \, p_j + \beta_i \, log\{x/P\}\right] + \delta_i \phi(x) \tag{27}$$

Where δ_i is an extra parameter associated with the density function and defined as the coefficient of the correction factor for the *i*th equation Note how the definition of $log\,P$ (14) implies that Equation (27) must be estimated using a non-linear procedure.

The set of n–1 equations like (27) conform the demand system, where n is the number of shares, being the last share recovered as a residual of the remaining n–1 ones. Once this demand system is estimated, the parameters are used to recover the expenditure function of a representative household COL index defined in (2) is calculated.

2.5 Summary

Following the Konus' theory of the "true index of the cost of living" it has been defined a COLI in the sense that it compares two different combinations of goods at given prices which the condition that both combinations provide the same standard of living or utility level. The way in which the difficulties arise of this calculation has been addressed with the methodology developed by Deaton and Muellbauer (1980) called Almost Ideal Demand System (AIDS) which establish the proper way to estimate a flexible functional demand system with several convenient properties and the most relevant characteristic is that eliminates the necessity of knowing explicitly the functional form of the utility function.

The data used to apply this methodology provides from the Household Budget Survey (HBS) from the National Statistical Institute (INE) which recover information about consumption patterns, income and other socioeconomic characteristics of the Spanish households. Due to the characteristics of the survey a problem in the estimation process arise. This problem is due to households that report zero consumption in some of the goods analyzed and known as censored data. The censored dependent variable makes that the AIDS in the standard way as have been developed by Deaton and Muellbauer (1980) cannot be applied. This issue biases the estimation of the parameters of the model what makes us to follow the two step procedure developed by Shonkwiler and Yen (1999) which addressed the problem by estimating in the first step a Probit model in which the dependent variable is the decision of consumption a particular item for incorporating the normal cumulative and density functions in the AIDS in the second step.

3 REGIONAL PRICE DIFFERENCES AND WELFARE IMPLICATIONS

3.1 Introduction

Computing a spatial version of the Cost of Living Index (COLI) is an issue of great interest since it allows measuring expenditure differentials across regions. Despite this, the main reason which explains the lack of studies in this area for the Spanish case is the limitations of data availability. In Spain, like in the majority of the countries, it is available the official Consumer Price Index (CPI) and it is often used for these purposes. The Spanish National Statistical Institute (INE) estimates the CPI as a chain-Laspeyres index between the current period and the base period, but the CPI would allow a comparison of the changes in the price levels over time across different regions but do not permit a comparison of absolute price levels between regions at a given point in time. Moreover, the way in which the CPI is calculated offers a biased overview of the costs of living. This CPI evaluates changes in the average prices for the acquisition of a basket of goods and services which is considered to be representative of the expenditure of all consumers ignoring consumer`s substitution because of the changes in its preferences. The seminal work of the Boskin Commission (Boskin *et al*., 1996) estimates that the annual bias in the US CPI is 1.1 points per annum concluding that over-indexation of federal tax and expenditure programs could add $1.07 trillion to US national debt by 2008.

This characteristic makes that the CPI is a limited indicator for measuring how the cost of living evolves over time because even in the case that consumers face the same nominal prices, variations in the cost of living can arise because of differences in expenditure patterns. It can be highlight the importance of the existing bias between the two indices taking into account the words of Allan Greenspan in 1995 in the U.S. Congress. He declared that he suspected that the Price Consumer Index overstated the cost of living by 0.5 to 1.5% points annually. The bias of 1.1% a year up to 2006 would generate an increase of US$ 691 billion in the public deficit (Fava, 2010).

Recent literature shows that the substitution bias is more important in spatial comparisons than in a time series context, Paredes and Iturra (2013) find that spatial substitution bias is higher than the substitution bias over time and the bias is larger when smaller are the spatial units considered. There are two main reasons that aggravate the bias problem in this spatial context: (i) even though price variations could be marginal over a short period of time, transportation costs affect current prices differently across areas. Even if the market price could be balanced by supply and demand adjustments, transportation costs imply price differentials among regions; and, (ii) the composition of the consumption basket is affected by geographical factors and weather conditions. The geography of Spain is characterized by different climatic conditions between the north and the south and more extremely differences take place in regions like the Canary Islands. For these reasons, the assumption that the consumption basket is fixed could be a highly unrealistic assumption when spatial price variations are studied.

These two limitations of the CPI point out the importance of constructing an alternative index which measures the cost of living in an accurate way, and also permits comparisons not only along the time, but also across the space. Recent research in the field of regional economics estimates their own COLI according to different approaches: (i) estimation of a regression model on the factors that explain cost of living in an area (Kurre, 2003), this technique involves identifying the factors that cause the cost of living differences between places. This approach relies on the preexistence of a cost of living database for a set of areas to fits a structural equation to those data then data for areas that did not participate in the initial study can be applied to this equation to estimate their cost of living values. Many authors use a variation of this approach (Walden, 1997, 1998; Cebula, 1993; Cebula *et al*., 1992; Kurre, 1992; and McMahon, 1991); (ii) estimation of cost of living indicators from expenditure data (Voicu and Lahr, 1999). This approach is based on the premise that variations in expenditures can be used to approximate prices alone, but the problem is that quantities vary as prices do and as a result, changes in expenditure tend to reflect both price and quantity changes; and (iii) estimation of a complete set of demand equations for all commodities in all places (Paredes and Iturra, 2013), this approach is based solidly in microeconomic theory and starts with a set of demand equations, one for each commodity consumed.

While this last approach is considered as technically complicated and it has not been broadly applied due to its data requirements, it is strongly grounded on the consumer theory and it will be applied in this thesis. It will be followed Konüs' idea described in the *Methodology* when studying price differentials across space. Despite the vast literature on cost of living indices over time, the spatial dimension has received less attention (Desai, 1969; Nelson, 1991; Timmins, 2006 and Atuesta and Paredes, 2012). For the specific case of Spain, no previous attempt of computing a Spatial Cost of Living Index (SCOLI) has been made in the context of the microeconomic theory; it could be found some works like Alberola and Marqués (2001) and Garrido-Yserte *et al*. (2012) which are some approximations to spatial cost of living indices in the way that they make price comparisons across the space.

The estimates of the SCOLI give us the ability to deflate regional nominal measures of welfare which if they are not spatially deflated would produce biased information. Adjustments in income measures to take into account the local price level are relevant for inequality studies, wage comparisons or assessments of poverty. This is another purpose of this chapter, income and wages will be deflated by SCOLI differences and it will be made a deeper analysis of inequality through the evaluation of poverty measures.

The economic crisis have arisen the poverty levels in Spain. But this poverty incidence is not uniform across the space. The usual analyses do not take into account the importance of the space; it is assumed that costs of living and consumption patterns are uniform across the regions. However, in the real world living standards are very heterogeneous across the space. The estimation of spatio-temporal price indices allows identify how the impact of poverty incidence is across regions. Once the spatio-temporal price index is constructed it will be used for reviewing the poverty lines and its incidence in Spain. The aim of this exercise is to estimate regional poverty incidence using the SCOLI to compare this results with those based in the official CPI. It is tried to answer several questions relating with poverty incidence. First, in which magnitude the poverty incidence is affected by differences between the CPI and the COLI in a given moment of time; second, it is asked if poverty could be modified due to the different costs of living across the space, the poverty incidence could be reduced or increased if we take into account the spatial cost of living differences. The results show that poverty incidence has risen in the whole country more than the usual analyses reflect. Also, it can be observed important differences across regions in the way that the richest and more touristic regions have more poverty incidence due to the higher cost of living that they support. Or in the same way, we can find that certain regions with lower costs of living reduces they poverty incidence when adjusted by cost of living differences.

The rest of the Chapter 3 is structured as follows. The section 3.2 explains how to construct a Spatial Cost of Living Index. The section 3.3 provides the results of the application to the whole country, to the Autonomous Communities and for the period 2008-2012, always comparing with the results provide by the CPI. The welfare implications over the wage level and a deeper analysis of poverty can be found in section 3.4 and, lastly, a set of conclusions are summarized in section 3.5.

3.2 The construction of a Spatial Index of Cost of Living

The SCOLI is calculated by estimating an Almost Ideal Demand System (AIDS) developed by Deaton and Muellbauer (1980) following a microeconomic approach consistent with the consumer theory. As it has been described in the *Methodology*, the AIDS estimates an expenditure function as a function of prices and a given utility level. After the parametric construction, the expenditure ratio between the average prices of two regions is directly estimated. In this section is provided an estimation of the SCOLI for the regions of Spain following this particular approach.

The data required for the estimation are obtained from the Household Budget Survey (HBS) at the household level taking as reference years 2008 to 2012. It is important to note that our analysis will be partially affected by data availability. Firstly, because the SCOLI estimated will be calculated only for the food group, given that information about monetary expenditure and physical quantities consumed for each type of good at the household level is required for its computation. This information allows for recovering prices, something that cannot be properly achieved in the rest of the expenditure groups. Even when focusing only on food consumption could be seen as a limitation of our study, it has to be considered that the food group is the most important group in terms of household consumption in Spain, with an expenditure of more than one quarter of total household budget in 2010 (Spanish National Statistical Institute, INE).[3] Moreover, the prices in the food products are especially important for the recipients of social welfare (Breuer, 2007) and, consequently, for the evaluation of welfare policies. One reason is because food products are commodities for which prices typically present huge variability across space and along the time. Consequently, large variations on prices of these products are likely to boost inequality (Loughrey *et al.*, 2012). Also, differences in the cost of food reflect real variations in supply costs and these variations are real components in disparities in the quality of life (Curran *et al.*, 2006). On the contrary, regional differences in housing prices are likely to reflect only amenity and disamenity differentials (Kaplow, 1995). Secondly, the analysis is limited because it estimates the SCOLI for the 17 Spanish NUTS-II regions (Autonomous Communities), since the Household Budget Survey provides data on household consumption at this level of geographical disaggregation.

In order to calculate the SCOLI it will be followed the general economic approach of Deaton and Muellbauer (1980) but applying to the spatial case. The estimation of the AIDS is used to recover the expenditure function and to calculate the cost of living for a representative household in each region. The definition of a SCOLI between regions *i* and *j* is:

$$SCOLI_{ij} = \frac{c(\bar{p}_i, u)}{c(\bar{p}_j, u)} \tag{28}$$

Where c is the cost function that represent the cost of living in Euros, $\bar{p}_i$ and $\bar{p}_j$are vectors containing the prices paid by the reference consumer in the regions *i* and *j*, respectively; and u is a utility level set as common for both regions. In this approach median prices are chosen instead of mean prices in order to avoid the variability caused by outliers.

[3]Housing is normally considered as a very important group in relative terms of the household budgets, but the characteristics of our database, which only include information on rents paid (and not payments of mortgages, for example) prevent from including housing in our COLI.

It is formulated the spatial version of the AIDS model with censored data define in (27):

$$w_i = \Phi(x)\left[\alpha_i + \sum_{j=1}^{n} \gamma_{ij} \log p_j + \beta_i \log\{x/P\}\right] + \sum_k c_k\, CS_k + r_h R_h + \delta_i \phi(x) \qquad (29)$$

where δ_i is a parameter associated with the density function, CS_k are dummy variables for different urban sizes and R_h is a regional dummy for each one of the NUTS-II regions of Spain, and c_k and r_h are the parameters associated with each type of dummy, respectively, with the aim of recover the idiosyncratic components inherent to each region and type of city. Note that this is a particular characteristic in our formulation, since we incorporate a factor to consider the spatial heterogeneity, in terms of unobservable city size and regional characteristics. The information contained in the HBS allows for distinguishing between five types of municipalities according to their population sizes:

1. Municipalities of more than 100,000 inhabitants
2. Municipalities between 50,000 and 100,000 inhabitants
3. Municipalities between 20,000 and 50,000 inhabitants
4. Municipalities between 10,000 and 20,000 inhabitants
5. Municipalities of less than 10,000 inhabitants

The set of n– 1 equations like (29) conform the demand system, where n is the number of shares, being the last share recovered as a residual of the remaining n– 1 ones. Once this demand system is estimated, the parameters are used to recover the expenditure function of a representative household and to calculate the index defined in Equation (28) by dividing the expenditure function of the consumer by the expenditure function of the consumer taken as reference.

The estimates of the Probit model for the first step of the Shonkwiller and Yen (1999) methodology are shown in Appendix 1-Appendix 5. A binary variable which represents the decision of consumption of each one of the ten food groups at the sample is regressed as a function of the socioeconomic variables like the log of the expenditure level of the household, the household income level, the household head age, sex and marital status, the education level and the size of the household and the number of employed; and geographic variables represented as region dummies and one dummy that takes the value 1 if the household is living in a capital city and 0 otherwise.

The estimation results show that all the socio-economic variables are significant at the 1% level. There is evidence that there are remarkably different purchase patterns across NUTS-II regions, given that all the regional dummies (with the exception of only a few) are significant at the 1% level for all the commodities. The estimates also show that the effect of being located in a capital city is a negative factor in the decision of consumption of all commodities, since in all the commodity regressions the coefficient of the capital city dummy is negative and significant at the 1% level. The results of the Probit model (Appendix 1 to Appendix 5) will be used to calculate the cumulative (Φ) and the density (ϕ) functions, which are included as regressors in the second step in the estimation of the AIDS.

The parameters of the AIDS model are recovered by applying Nonlinear Seemingly Unrelated Regression (NLSUR), which estimates a system of nonlinear equations by Feasible Generalized Nonlinear Least Squares (FGNLS). The parameters estimates are shown in Appendix 6, being most of the estimates significant at the 1% level. These estimates are required in order to recover the utility level and the expenditure equation described in (5) as a function of prices and income. Specifically, the household in the median of the expenditure distribution in each region is taken as the representative household, being the prices in (5) set to the median prices on each region. Once the value for this expenditure function is calculated, it is used to calculate the SCOLI defined in equation (28).

3.3 Results and discussion

3.3.1 An overview of the cost of living in Spain

Before start with the spatial analysis, the evolution of the COLI for the period 2008 to 2012 is calculated. To see the evolution of the cost of living in Spain since the beginning of the crisis until its lasts years the expenditure function is recovered for each year and constructed with the median prices, the median expenditure level and the common utility level that it is set as the utility level of the median consumer in 2008. After the calculation of the expenditure functions, which represents the median cost of living of each year, it is obtained the COLI defined in (2) by dividing the cost of living of each year by the cost of living of the year 2011, this is the reason why the year 2011 presents a value of 1. In Figure 1 the evolution of this COLI is represented.

Figure 1 Evolution of the COLI in Spain with the utility level of 2008 (2008-2012)

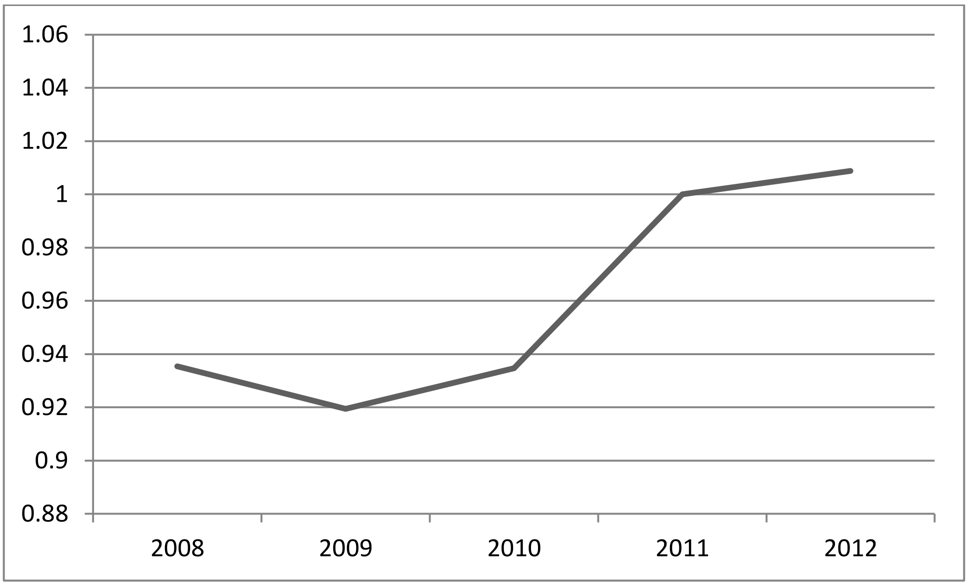
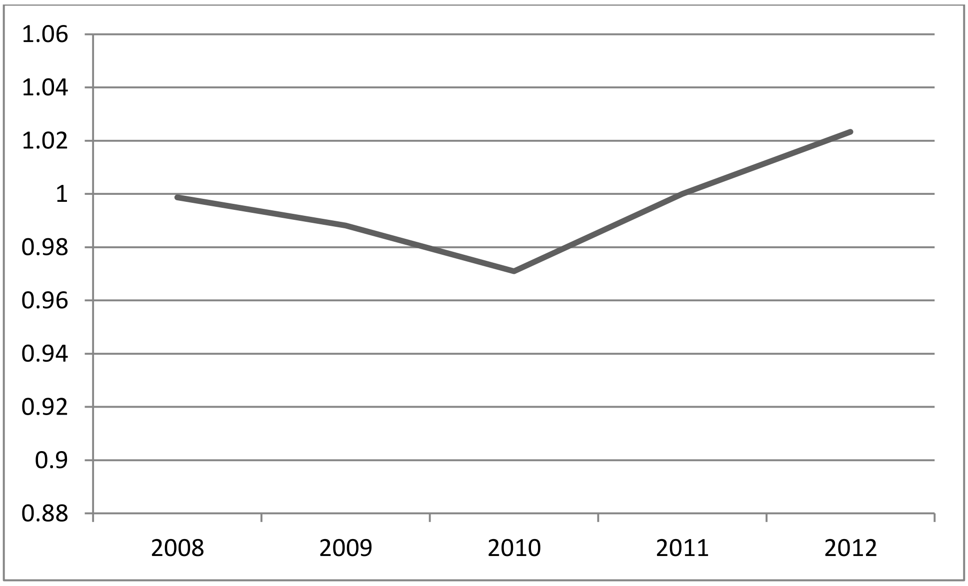


Figure 2 Evolution of the Food CPI in Spain (2008-2012). Base year 2011



Source: INE

The Figure 1 represents the evolution of the median cost of living in Spain. The utility level of reference is the median utility level of the year 2008 but with the base year of reference in 2011; this year is chosen for comparison purposes with the CPI in Base 2011 which is represented in Figure 2.

The evolution of both indices is very different despite both represent the food prices. In Figure 1 it can be observed a clear upward trend of the COLI since 2009 being the increment in all the period of almost 7.3%, this increment in cost of living is remarkably high in the period 2010-2011. Instead the CPI showed in Figure 2 shows a decrement in prices in the period 2008-2010 and then a continuous and smooth increment until 2012. In the overall period 2008-2012 the increment in prices was of 2.4%, a very low value compared to the increment of the 10% than throw out the COLI. In the Table 1 it is showed the values of the cost of living in Euros for attaining the median utility level of the year 2008, indeed in the second row it is showed the value of the COLI for all the period.

Table 2 Evolution of the Cost of living for Spain with the utility level of 2008 (2008-2012)

	2008	2009	2010	2011	2012
Cost of living in Euros	4058.29	3988.94	4055.19	4338.77	4376.95
Cost of living index	0.9354	0.9194	0.9346	1.0000	1.0088

These results could give us a general overview of the evolution of prices along the period 2008-2012 in Spain, but assuming that all regions follow the same evolution. As we discuss previously, this is a very strong assumption. To demonstrate the necessity of taking into account spatial differentials a spatial version of the COLI is calculated for each Autonomous Communities.

3.3.2 A Spatial Cost of Living Index for the Autonomous Communities

A Spatial Cost of Living Index (SCOLI) for the 17 Autonomous Communities is calculated here applying the methodology explained in section 3.2. In this case the spatial factor is now included, this implied that for obtaining the SCOLI defined in (28) it is necessary calculate the cost function with the median prices and the median expenditure level of each Autonomous Community, and the utility level set in this case will be the utility level of Madrid. Dividing each expenditure function of each Autonomous Community by the expenditure function of Madrid we obtain the SCOLI with respect to Madrid, for this reason Madrid takes the value 1. The results of the SCOLI are shown in Table 3.

Table 3 The Spatial Cost of Living Index by regions, 2012

	SCOL in Euros	SCOL Index
ANDALUSIA	3334.63	0.9207
ARAGON	3627.06	1.0014
ASTURIAS	3176.51	0.8770
BALEARIC ISLANDS	3610.42	0.9968
CANARY ISLANDS	3789.83	1.0464
CANTABRIA	3345.68	0.9237
CASTILE LEON	3088.99	0.8529
CASTILE LA MANCHA	2950.67	0.8147
CATALONIA	3877.83	1.0707
VALENCIA	3514.60	0.9704
EXTREMADURA	3115.53	0.8602
GALICIA	3070.08	0.8477
MADRID	3621.85	1.0000
MURCIA	3841.27	1.0606
NAVARRA	3871.45	1.0689
BASQUE COUNTRY	3721.12	1.0274
LA RIOJA	3354.68	0.9262

The results of Table 3, show the smallest value (0.8147) in Castile-La Mancha and the highest one (1.0707) in Catalonia. The results could be interpreted as follows: for example, the cost in food products required to attain the same utility level as the median household in Madrid is around 19% lower for the median household in Castile-La Mancha. Similarly, achieving this utility level for the equivalent households living in Catalonia is 7% more expensive respect to Madrid. It is important to note that differences in consumption patterns between households are allowed in the estimation of SCOLI, provided that their utility level is the same.

One important question is whether our SCOLI is invariant to the choice of the utility level of the base region. For evaluating its sensitivity to the choice of the region taken as reference we have computed the index with the utility level attained by the household in the median of the expenditure distribution in the region of Castile La Mancha, which happened to be the region with the lowest cost of living. The results are reported in Table 4 being, in general, very similar to those with the utility of the median household in Madrid.

Table 4 Comparison of the SCOL index with different utility levels

	SCOL index (1)	SCOL index (2)	Income per capita 2012, € (3)
ANDALUSIA	0.9207	0.9198	**16,744**
ARAGON	1.0014	1.0013	**24,812**
ASTURIAS	0.8770	0.8776	**20,867**
BALEARIC ISLANDS	0.9968	0.9964	**23,596**
CANARY ISLANDS	1.0464	1.0451	**18,940**
CANTABRIA	0.9237	0.9243	**21,698**
CASTILE LEON	0.8529	0.8533	**17,693**
CASTILE LA MANCHA	0.8147	0.8145	**22,000**
CATALONIA	1.0707	1.0700	**26,419**
VALENCIA	0.9704	0.9700	**19,485**
EXTREMADURA	0.8602	0.8592	**15,133**
GALICIA	0.8477	0.8475	**20,336**
MADRID	1.0000	1.0000	**28,914**
MURCIA	1.0606	1.0609	**18,032**
NAVARRA	1.0689	1.0691	**28,499**
BASQUE COUNTRY	1.0274	1.0271	**30,051**
LA RIOJA	0.9262	0.9277	**25,191**
Correlation coefficient SCOL/Income per capita	0.4971	0.4999	

(1) SCOLI calculated with the utility level of the median household in Madrid as in Table 1

(2) SCOLI calculated with the utility level of Castile La Mancha and expressed in terms of Madrid

(3) National Statistical Institute (INE)

Independently on the region taken as reference for the comparison, regional differences in the cost of living seem to be quite relevant, especially if it is taken into account the relative small size of the country and the geographical proximity between some of the regions.

One interesting result is that the outcomes in Table 3 or Table 4 suggest a positive correlation between the cost of living figures estimated and the income level. The average income per capita in 2012 in the most developed regions like Madrid, Basque Country and Navarra was about 29,000 euros while the average income level in regions like Castile La Mancha, Andalusia or Extremadura was less than 18,000 euros.[4] Note that the former three rich regions are those with the highest SCOLI, whereas the last low-income regions of Spain present the smallest SCOLI estimates.

The SCOLI estimates can be also useful to study the relation between the type of region and its cost of living. Our results suggest that in regions like Catalonia or the Basque Country the cost of consuming food products is higher than in the region of Madrid, which is the region with the largest city of Spain. Even when large city sizes are normally linked to higher incomes and, consequently, higher costs, one should bear in mind that the results are only observable at the NUTS-II level. In other words, our results are just an average of all the households living in the region of Madrid, which includes Madrid City but much smaller towns and villages as well. Catalonia, the Basque Country and Navarra do not contain such a large city, even when Barcelona in Catalonia or the metropolitan area of Bilbao in the Basque Country are considerably large. However, they are regions containing many more urban areas on average than the region of Madrid. Furthermore, these regions are located in the so-called Ebro-Axis, which is the area with the most traditionally developed areas of Spain and also with relatively higher incomes per capita. This result is in line with previous literature that concludes that income is expected to affect the cost of living in the way that the richer region would experience a larger demand for most goods with an upward pressure on prices (Kurre, 2003). One interesting exception is the case of the Canary Islands, where the cost of consuming food products is estimated to be larger than in many comparatively richer NUTS-II regions. The results of the SCOLI for the Canary Islands are 3.4% higher than Madrid and very similar to the cost of living estimated for the Basque Country. Even when this could be somehow surprising, since Canary Islands are considered as relatively poor within Spain, this could be a consequence of the particular transportation and climate factors in this specific region.[5]

[4]These data are available in the Regional Accounts published by the INE:http://www.ine.es/jaxi/menu.do?type=pcaxis&path=%2Ft35%2Fp010&file=inebase&L=0

[5]They are located just off the northwest coast of mainland Africa.

Another interesting insight comes from comparing our SCOLI results with an axiomatic Laspeyres Price Index. In this comparison, a spatial Laspeyres index is calculated by setting a fixed basket, taking the average shares calculated by the National Statistical Institute for the whole country and the median prices (unit values) in each region:

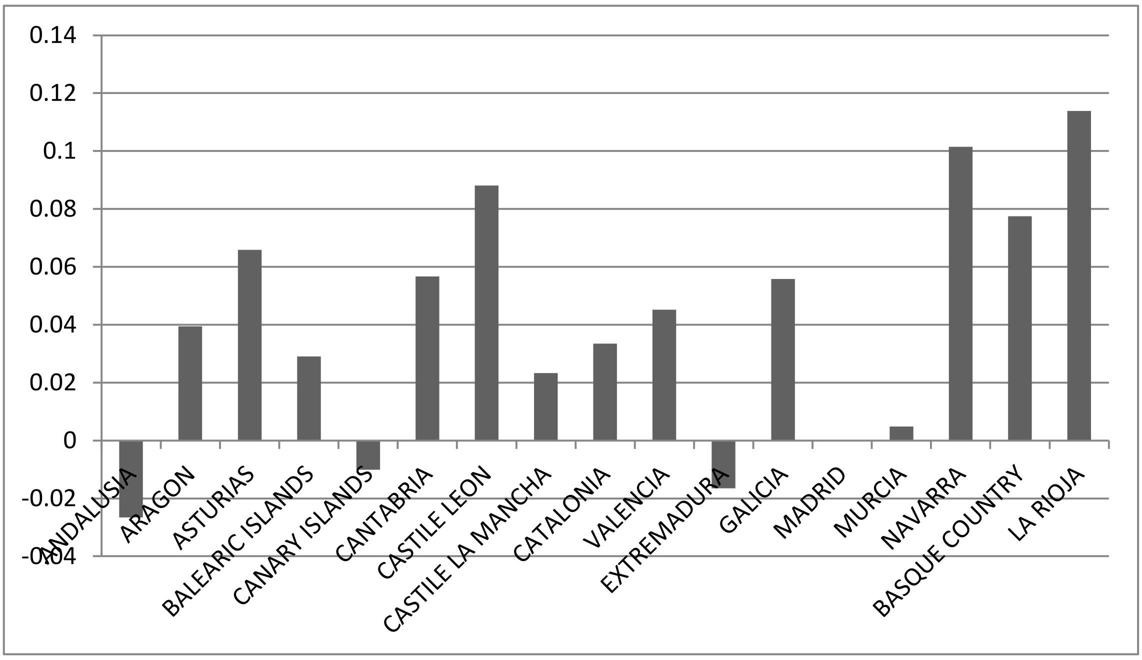
$$\mathrm{L}_{hr} = \frac{\sum_j \bar{p}_{jh}\, w_{jr}}{\sum_j \bar{p}_{jr}\, w_{jr}}; j = 1, \dots, n \qquad (30)$$

Being $\bar{p}_h$ and $\bar{p}_r$ the vectors with the median prices paid by the households in areas h and r for the j=1,…n food products, respectively; and w_r the vector that contains the average national budget shares of the INE (weights).

The substitution bias is calculated as the difference between the Spatial Laspeyres Price Index and our SCOLI [(30)-(28)] A positive difference between the two indices would mean that a fixed basket approach would over-estimate the cost of living. Oppositely, a negative substitution bias indicates that the fixed basket approach under-estimates the cost of living.

Regional substitution bias is shown in Figure 3, where remarkably discrepancies between the two indices can be observed. All the regions but the Canary Islands, Andalusia and Extremadura show a positive substitution bias, indicating that for most regions the setting of a common consumption pattern would result in an overestimation of the "true" cost of living. The results also suggest that the differentials in the cost of living among Spanish regions depend not only on differences in price levels, but also on different consumption patterns and substitution among regions.

Figure 3 Regional substitution bias 2012



Given these results it is evident the importance and the magnitude of the spatial substitution bias, and a discussion emerges that deals with the validity of using a fixed basket approach as is currently the case of the majority of the Statistical Agencies. The CPI becomes playing more and more important roles in policy decisions which implies that the accurate measurement of the CPI has been more important for an economy. Therefore, the substitution bias should be taken into consideration.

All previous results show evidence that there exist relevant differences in price levels across the regions in Spain. Once a SCOLI for the Spanish regions has been constructed, now is interesting to know how the evolution of the regional costs of living is. Is the evolution the same in the Autonomous Communities than in the whole country? Does all Autonomous Communities have the same evolution?

3.3.3 Evolution of the Spatial Cost of Living Index 2008-2012

For asking these questions it will be constructed the SCOLI for the 17 Autonomous Communities for the period 2008 to 2012. Table 5 shows this spatio-temporal cost of living.

Table 5 Cost of Living in Euros of the utility level of Madrid-2008 for Spanish Regions (2008-2012)

	2008	2009	2010	2011	2012
ANDALUSIA	3999.69	3634.63	3782.51	4090.30	4143.24
ARAGON	3994.97	4062.12	4059.20	4340.74	4514.88
ASTURIAS	3990.92	3491.14	3598.02	3972.70	3960.31
BALEARIC ISLANDS	4010.88	4038.06	4026.64	4386.01	4490.91
CANARY ISLANDS	4042.65	4157.02	4128.27	4697.92	4706.47
CANTABRIA	3973.08	3700.74	3875.91	4260.02	4171.32
CASTILE LEON	4014.34	3476.55	3494.56	3780.85	3850.33
CASTILE LAMANCHA	4030.68	3249.49	3271.49	3534.06	3672.40
CATALONIA	4024.63	4366.69	4418.83	4823.84	4821.78
VALENCIA	4046.64	4214.54	4074.56	4409.64	4372.74
EXTREMADURA	4004.71	3701.13	3675.05	3826.55	3869.80
GALICIA	4016.28	3461.18	3501.31	3795.18	3820.64
MADRID	3961.15	4047.24	4053.15	4414.44	4509.55
MURCIA	3980.55	4299.32	4297.53	4724.99	4785.61
NAVARRA	3959.58	4213.77	4358.76	4739.14	4822.14
BASQUE COUNTRY	3993.36	4179.08	4252.44	4619.52	4630.68
LA RIOJA	4067.34	3946.12	3949.29	4331.20	4190.49

These data reflect the spatial and the temporal dimension, so they permit comparisons across the space in different time moments simultaneously. More exactly, shows the cost of living in Euros of attaining the utility level of Madrid in year 2008. These quantities show the total annual expenditure that the median household of each region needs to acquire the utility level of reference, Madrid in 2008. The same results are presented in Table 6 but in the form of an index. We have divided all the expenditure functions of all the Spanish regions by the expenditure function of the region of Madrid (that contains the capital and the largest city of the country) in year 2008, for this reason Madrid 2008 takes the value 1.

Table 6 Cost of Living Index with the utility level of Madrid-2008 for Spanish Regions (2008-2012)

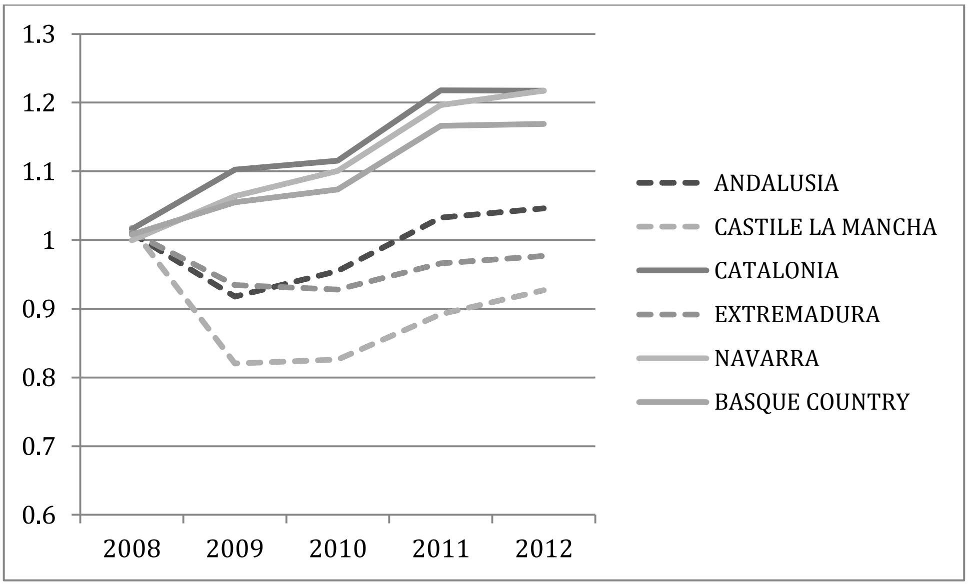
	2008	2009	2010	2011	2012
ANDALUSIA	1.0097	0.9176	0.9549	1.0326	1.0460
ARAGON	1.0085	1.0255	1.0248	1.0958	1.1398
ASTURIAS	1.0075	0.8813	0.9083	1.0029	0.9998
BALEARIC ISLANDS	1.0126	1.0194	1.0165	1.1073	1.1337
CANARY ISLANDS	1.0206	1.0494	1.0422	1.1860	1.1882
CANTABRIA	1.0030	0.9343	0.9785	1.0754	1.0531
CASTILE LEON	1.0134	0.8777	0.8822	0.9545	0.9720
CASTILE LAMANCHA	1.0176	0.8203	0.8259	0.8922	0.9271
CATALONIA	1.0160	1.1024	1.1155	1.2178	1.2173
VALENCIA	1.0216	1.0640	1.0286	1.1132	1.1039
EXTREMADURA	1.0110	0.9344	0.9278	0.9660	0.9769
GALICIA	1.0139	0.8738	0.8839	0.9581	0.9645
MADRID	1.0000	1.0217	1.0232	1.1144	1.1384
MURCIA	1.0049	1.0854	1.0849	1.1928	1.2081
NAVARRA	0.9996	1.0638	1.1004	1.1964	1.2174
BASQUE COUNTRY	1.0081	1.0550	1.0735	1.1662	1.1690
LA RIOJA	1.0268	0.9962	0.9970	1.0934	1.0579

It can be seen the evolution of the SCOLI during the entire period that shows the different evolution of the regions depending on the group that these regions belong to. For the poorest regions it can be observed a downward trend in the cost of living between 2008-2010, contrary, in the richest regions it can be seen an upward trend in the cost of living in the same period. Then, since 2010 until 2012 both types of regions present an increase in the cost of living. The region most expensive is Catalonia in 2011 an almost a 40% higher than the cheapest region, Castile La Mancha in 2009. This means that a consumer in the Catalonia pay a 40% more than a consumer in Castile La Mancha for attaining the same utility level.

In the works of Alberola and Marqués (2001) and Garrido-Yserte *et al*. (2012) it is found similar conclusions for the Spanish case. The former arrived to these conclusions with their constructed series of regional CPIs finding that regional price divergences at a regional level are evident and persistent. Their study uses quarterly data of the INE from 1961 to 1998 and like us they find Zamora in Castile Leon the lowest inflation region and Vizcaya in the Basque Country the highest one. Garrido-Yserte *et al*. (2012) proposed a Cost of Living Index that incorporates the cost of acquiring or living in owned housing. Their data proceed from the Ministry of Housing from 1995 to 2007 and they note that the regional differences have grown over time.

In Figure 4 the evolution of the cost of living during the entire period is represented. This figure shows the SCOLI for the three richest regions and the three poorest regions of Spain.

Figure 4 Cost of Living Index by regions for 2008-2012

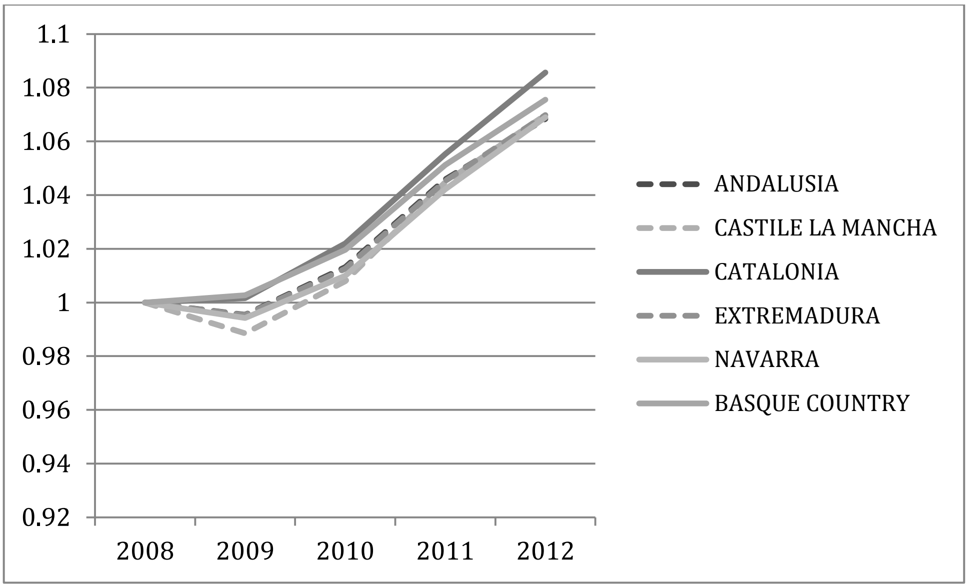


It is appreciated considerable differences in the SCOLI among Spanish regions, more precisely, between the poorest (with the discontinuous line) and the richest regions (with the solid line). It can be appreciated this heterogeneity in the way of an existing high-price area formed by Navarra, Catalonia and Basque Country and a low-price area formed by Andalusia, Castile La Mancha and Extremadura.

The Figure 4 shows up an important proposal of the regional economics, which is that the high income regions support higher prices than lower income regions (Kosfeld *et al*., 2007). Suedekum (2006) in his paper also indicates a strong correlation between cost of living indices and nominal earnings indicators such as income and wage, these proposals cannot be inferred from the CPI official data.

If the evolution of the CPI between the richest and the poorest regions is represented it cannot be observed any differences between both. This can be seen in Figure 5.

Figure 5 Consumer Price Index by regions for 2008-2012. Base 2008

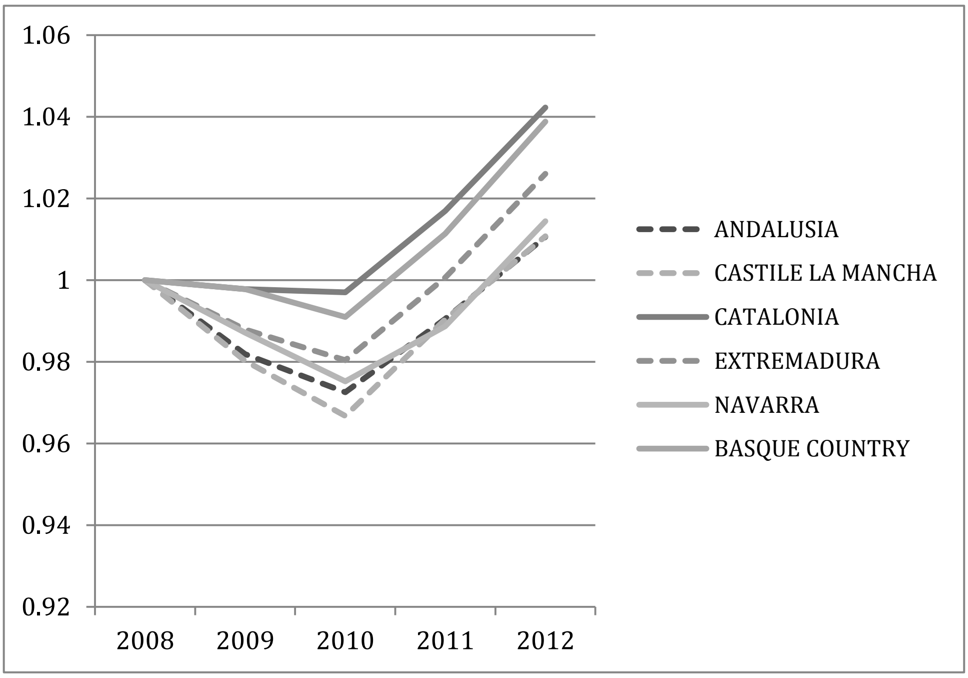


Source: INE

It can be observed that the overall trends are the same across regions. There is no gap between the poorest (Andalusia, Castile La Mancha and Extremadura) and the richest (Catalonia, Navarra and Basque Country) throughout the sample period. This result is not in the line neither with the literature about regional prices in Spain, nor with the regional economics. Alberola and Marqués (2001) and Garrido-Yserte *et al*. (2012) using other alternative price indices evidence substantial and permanent differences in prices among Spanish regions.

One could think that an important reason behind the strike differences between two indices could be that the CPI data represents the entire consumption basket with all kind of goods and services, in contrast with the SCOLI estimate in this thesis which represents only the food group. In Figure 6 we can contrast this issue.

Figure 6 Food CPI for regions for 2008-2012, Base 2008



Source: INE

The food CPI data show some more variation than the whole CPI represented in Figure 5, but these variations are no longer than 4%, indeed there is no difference in the behavior among regions, all of them evolves in the same way. After that, the SCOLI represented in Figure 4 could be considered a better and more realistic indicator of the economic situation both for the whole country and at a regional level.

Finally, a simple comparison of the two indices is made here. It is compared the official CPI data for the period 2008-2012 in base year 2011 as the INE publish. Then we will construct the SCOLI for the same period but with the utility level of each Autonomous Community in 2011. This is the best way to make both indices comparable.

Table 7 SCOL index with the utility level of each Autonomous Community in 2011 (2008-2012)

	2008	2009	2010	2011	2012
ANDALUSIA	0.9760	0.8893	0.9260	1.0000	1.0127
ARAGON	0.9203	0.9358	0.9361	1.0000	1.0397
ASTURIAS	1.0050	0.8789	0.9071	1.0000	0.9965
BALEARIC ISLANDS	0.9134	0.9216	0.9218	1.0000	1.0230
CANARY ISLANDS	0.8562	0.8891	0.8831	1.0000	0.9989
CANTABRIA	0.9343	0.8692	0.9114	1.0000	0.9784
CASTILE LEON	1.0614	0.9199	0.9250	1.0000	1.0180
CASTILE LAMANCHA	1.1381	0.9196	0.9259	1.0000	1.0391
CATALONIA	0.8331	0.9063	0.9182	1.0000	0.9988
VALENCIA	0.9167	0.9575	0.9262	1.0000	0.9902
EXTREMADURA	1.0444	0.9687	0.9627	1.0000	1.0105
GALICIA	1.0566	0.9118	0.9221	1.0000	1.0071
MADRID	0.8960	0.9173	0.9195	1.0000	1.0203
MURCIA	0.8414	0.9111	0.9118	1.0000	1.0112
NAVARRA	0.8336	0.8895	0.9211	1.0000	1.0169
BASQUE COUNTRY	0.8638	0.9049	0.9215	1.0000	1.0019
LA RIOJA	0.9403	0.9124	0.9136	1.0000	0.9657

Table 8 CPI index for the period 2008-2012. Base year 2011

	2008	2009	2010	2011	2012
ANDALUSIA	0.9562	0.9518	0.9688	1.0000	1.0216
ARAGON	0.9560	0.9522	0.9693	1.0000	1.0236
ASTURIAS	0.9514	0.9471	0.9657	1.0000	1.0227
BALEARIC ISLANDS	0.9576	0.9560	0.9722	1.0000	1.0250
CANARY ISLANDS	0.9752	0.9653	0.9746	1.0000	1.0203
CANTABRIA	0.9524	0.9492	0.9671	1.0000	1.0260
CASTILE LEON	0.9554	0.9486	0.9666	1.0000	1.0279
CASTILE LAMANCHA	0.9571	0.9461	0.9647	1.0000	1.0240
CATALONIA	0.9476	0.9492	0.9685	1.0000	1.0288
VALENCIA	0.9568	0.9545	0.9700	1.0000	1.0245
EXTREMADURA	0.9567	0.9522	0.9687	1.0000	1.0235
GALICIA	0.9551	0.9495	0.9673	1.0000	1.0237
MADRID	0.9530	0.9524	0.9700	1.0000	1.0238
MURCIA	0.9592	0.9534	0.9698	1.0000	1.0232
NAVARRA	0.9596	0.9540	0.9692	1.0000	1.0257
BASQUE COUNTRY	0.9513	0.9538	0.9700	1.0000	1.0231
LA RIOJA	0.9552	0.9507	0.9666	1.0000	1.0255

Source: INE

If both indices are compared it can be observed that the CPI shows a smooth and homogeneous increments during the period 2008-2012 in all the Autonomous Communities around 6%-8%. The evolution of the cost of living with the SCOLI is, in turn, very heterogeneous among Autonomous Communities, some of them, those with the highest COL, experimented increases in the COL in 2012 since 2008 up to 18% as the case of Navarra. In contrast, others, like Castile Leon, Castile La Mancha and Extremadura, experienced a decrease in the COL in 2012 since 2008 around 3%-9%.

Summarizing, if one takes the CPI as reference of the evolution of the cost of living, could see a distorted view of this evolution that is with the CPI all Autonomous Communities increase a little its prices in the period 2008-2012. The SCOL index shows both increases and decreases in prices in the Autonomous Communities and the magnitude of these movements is very heterogeneous among regions.

3.4 Some welfare implications: an analysis of the poverty incidence in Spanish Autonomous Communities.

Studies of welfare in Spain do not take into account appropriated price indices. To compare the standard of living of people in different localities, appropriated price indices among localities are needed. Not adjusting the nominal indicators of welfare yields misleading results for economic analysis and policy decisions such as inequality studies, wage comparisons or assessment of poverty.

For example, policies designed to alleviate regional income disparities that do not consider potential geographical differences in cost of living could result in benefits to regions that in real terms would not need these benefits. Most of the literature in regional wage differentials point out that one of the limitations in estimation of real wage differentials is the lack of an index that reflects the "true" cost of living by accounting for substitution or differences in preferences (Dumond *et al*., 1999 and Blien *et al*., 2009). This problem can be alleviated by means of a SCOLI. As an illustrative example, Table 9 shows the average wages across regions in nominal terms for 2011[6] and the same values when they are adjusted by the estimated SCOLI[7].

[6]We choose the year 2011 because is the last year available by the INE

[7] Information on regional nominal wages can be founded at: http://www.ine.es/jaxi/menu.do?type=pcaxis&path=/t22/e308_mnu&file=inebase&N=&L=0

Table 9 Nominal and adjusted wages in 2011 by regions

	Ranking (1)	Nominal Wage	Ranking (2)	Wage adjusted by SCOL
ANDALUSIA	10	21,351	11	23,213
ARAGON	13	22,333	5	22,304
ASTURIAS	12	22,286	15	25,394
BALEARIC ISLANDS	11	21,351	3	21,429
CANARY ISLANDS	1	19,517	1	18,675
CANTABRIA	5	20,932	6	22,645
CASTILE LEON	7	21,029	13	24,644
CASTILE LAMANCHA	4	20,665	14	25,370
CATALONIA	15	24,499	9	22,897
VALENCIA	9	21,316	4	21,974
EXTREMADURA	2	19,879	10	23,136
GALICIA	3	19,970	12	23,565
MADRID	16	25,845	17	25,845
MURCIA	8	21,077	2	19,867
NAVARRA	14	24,385	8	22,809
BASQUE COUNTRY	17	26,370	16	25,673
LA RIOJA	6	20,997	7	22,634
Standard deviation		2,059		1,975

In this Table 9 it can be viewed in the column of named *Nominal Wage* the average wage per worker in 2011, according to the Regional Accounts published by the INE and before adjusting for regional cost of living differences. It can be observed a low-wage area formed, among others, by Castile Leon, Castile La Mancha and Extremadura, and a high-wage area formed by Catalonia, Basque Country, Navarra or Madrid. Note that most of these areas correspond to the regions with the lowest and highest costs of living, respectively. The last column shows these values adjusted by the SCOLI estimated in the previous section, i.e., expressed in equivalent Euros of the median household in Madrid. By applying this adjustment is relatively easy to detect that the regional disparities in wages are lower: last row in Table 9 shows the standard deviations, being the dispersion in the adjusted wages by COL index lower. For example, the mean wage in Madrid in 2011 was approximately 20% higher than in Castile La Mancha; however if differences in cost of living are corrected by using the estimated SCOLI, this difference vanishes. The ranking of the regions according to their nominal or SCOLI-adjusted wages, however, does not vary very much in some regions like Basque Country or Madrid which are in the same position in both cases. But there are Autonomous Communities that change their positions dramatically that is the case of Castile La Mancha which moves its position from the fourth lowest wage to position number 14, also Extremadura changes from the second lowest wage to the position 10. Increased interest in regional labor market adjustment processes has led to a number of studies which employ a regional cost of living to deflate nominal wages (e.g. Hayes, 2005). The reason for doing this adjustment is based on the hypothesis that a regional price index instead a national aggregate helps to capture the endogeneity between regional goods and labor markets, provoking a deeper understanding of regional labor markets. The adjustment of wages by differentials in regional cost of living could be helpful when studying the comparative flexibility of regional labor markets: for example, the existence of an elastic wage curve at the regional level in Spain has been rejected in some recent literature (Garcia-Mainar and Montuenga-Gomez., 2003).

Following with the assessment of welfare, a deeper analysis on poverty will be made. It is important to focus the attention in poverty because in the last years the poverty incidence in Spain has been aggravated due to the strong economic crisis. It can be observed an increase in the percentage of families considered as poor. This increase in poverty incidence is observed both in the whole country and in each of the regions, although the incidence is different among regions.

This situation has caused the proliferation of many researches in Spain dedicated to the study of poverty: Ayala (2013); García-Serrano and Arranz (2013); and, Pérez (2013). Particularly, the study of Pérez (2013) highlights the different incidence and evolution of poverty across the space. Also, the National Statistical Institute (INE) analyses the heterogeneity of poverty evolution across the Autonomous Communities.

The aim of this section is to measure the impact of regional price differentials on poverty. This can be measured comparing the estimates of poverty which use the official CPI to deflate (Herrero *et al.*, 2013) with the estimates of poverty deflating by the SCOLI. To make this, the regional poverty lines are deflated not only in a temporal way, but also in a spatial fashion.

The usual procedure to measure poverty is based on a delimitation of a poverty line that is considered to be the 60% of the median or the 50% of the mean of the household's income or expenditure. These measures are known as "poverty lines". Usually these analyses assume that cost of living is uniform across the space and that do not exist different consumption patterns among regions. This makes difficult to compare household welfare across the space, so some corrections for differences in the cost of living are needed to make an accurate analysis of the poverty incidence. Some studies show that estimates of poverty are heavily dependent upon the inflation rates used (Sigit and Surbakti, 1999; Frankenberg *et al*., 1999; and Asra, 1999).

Herrero *et al*. (2013) make a poverty analysis in the Spanish Autonomous Communities over the period 2006-2011 using the same data as is used in this thesis that is the HBS. The authors use a poverty line very usual in these kinds of works that is the percentage of households below the 60% of the median of the household expenditure in 2006 and adjusted by equivalence scales of consumption. They adjust the poverty line of 2006 with the CPI and they obtain the poverty line in constant Euros of 2011. With this approach the results of the percentage of poor households are shown in Table 10.

Table 10 Percentage of poor households below the 60% of the median expenditure in 2006 by Autonomous Communities

	2006	2007	2008	2009	2010	2011
ANDALUSIA	25.1	18.5	15.8	18.2	20.0	21.3
ARAGON	21.7	15.9	11.3	15.3	17.3	18.3
ASTURIAS	20.3	18.3	13.5	11.6	10.5	15.0
BALEARIC ISLANDS	20.7	12.9	16.2	25.9	25.4	27.6
CANARY ISLANDS	21.6	18.5	18.1	24.2	27.5	29.8
CANTABRIA	27.6	18.7	15.6	16.1	16.3	15.2
CASTILE LEON	24.9	20.2	14.7	16.7	18.9	16.0
CASTILE LAMANCHA	22.2	15.8	14.5	16.9	17.2	15.3
CATALONIA	23.1	19.4	18.0	20.2	22.0	21.6
VALENCIA	23.2	18.0	19.1	21.6	23.4	26.6
EXTREMADURA	23.5	17.2	14.6	16.5	17.6	13.9
GALICIA	21.1	17.7	15.3	17.3	16.7	17.0
MADRID	22.2	19.0	15.7	18.9	19.2	19.6
MURCIA	19.0	14.7	18.4	23.4	23.6	22.8
NAVARRA	18.8	16.1	13.5	16.1	16.4	16.1
BASQUE COUNTRY	17.2	13.4	11.7	13.0	12.4	10.6
LA RIOJA	20.7	18.4	14.8	14.8	16.1	15.4

Source: Herrero *et al.* (2013)

In Table 10 we can be observed a very smooth increase in the poverty in Spain (around 2%), also we can observe some intermediate periods with a decrease in poverty. But this pattern of growth shows relevant differences in regional behavior pointing out the importance of the spatial analysis in this context. Some Autonomous Communities like Canary and Balearic Islands and Valencia, suffer from remarkably increments in poverty higher than the national mean, for example, in Canary Island the increment is about 30%. In contrast, regions like Castile Leon and Extremadura reduce they poverty incidence considerably. These results are discussed widely in Herrero *et al*. (2013) but are shown here to highlights the importance of the spatial analysis in poverty in Spain.

As we can see the poverty analysis are based in the delimitation of a poverty line measured through a percentage of the mean or the median of the household income or expenditure. But these analyses suppose that the costs of living are uniform across the space being a postulation very far from the reality.

The existence of regional prices and costs of living variations imply the necessity of re-estimate the poverty lines using a Spatial Cost of Living Index. For this purpose, first, we calculate the SCOLI with the utility level of each Autonomous Community in 2011. The reason because we choose the utility level of each Autonomous Community in 2011 is to make the results comparable with those of Herrero *et al* (2013). The results of the SCOLI can be seen in Table 11.

Table 11 SCOL for Autonomous Communities, in euros, with the utility level of each Autonomous Community in 2011

	2006	2011	SCOLI
ANDALUSIA	2,841.60	3,515.34	1.24
ARAGON	2,780.75	3,467.11	1.24
ASTURIAS	2,569.80	3,282.45	1.28
BALEARIC ISLANDS	2,742.42	3,142.48	1.15
CANARY ISLANDS	2,442.58	3,087.62	1.26
CANTABRIA	2,587.12	3,157.44	1.22
CASTILE LEON	2,707.68	3,342.28	1.23
CASTILE LAMANCHA	2,762.99	3,285.85	1.19
CATALONIA	2,650.65	3,594.00	1.35
VALENCIA	2,841.90	3,208.50	1.12
EXTREMADURA	2,509.29	3,027.66	1.21
GALICIA	2,797.95	3,651.23	1.30
MADRID	2,675.89	3,330.04	1.24
MURCIA	2,927.18	3,500.07	1.19
NAVARRA	2,982.75	3,862.60	1.29
BASQUE COUNTRY	3,043.03	3,964.00	1.30
LA RIOJA	2,871.25	3,373.65	1.17

In the last column of Table 11 we have the SCOLI between 2006 and 2011. It can be concluded that there exists an important increment in the expenditure level needed to attain the constant utility level. The mean increase is about 20% but it can be observed that these increments are not homogeneous across the space: some regions have increases above the mean; these regions are those with the highest per capita income (Catalonia, Navarra and Basque Country). The reality that these data throw out is very different to that reflected with the CPI. These results confirm that in the economic crisis a relevant increase in the cost of living have been taken place.

It is easy to note that these different results will affect in a relevant sense to the poverty analysis.

The analysis starts applying our SCOLI calculated in Table 11 to the poverty lines of 2006. This mean that we have the poverty line in terms of 2011 and spatially deflated, the derived poverty lines are given in Table 12 Poverty line for each Autonomous Community temporal and spatially deflated Table 12.

Table 12 Poverty line for each Autonomous Community temporal and spatially deflated

	2006
ANDALUSIA	9,873.37
ARAGON	10,599.05
ASTURIAS	10,112.34
BALEARIC ISLANDS	11,244.36
CANARY ISLANDS	9,975.37
CANTABRIA	10,518.02
CASTILE LEON	10,027.63
CASTILE LAMANCHA	8,632.13
CATALONIA	13,549.05
VALENCIA	9,881.81
EXTREMADURA	7,874.44
GALICIA	10,292.66
MADRID	12,927.61
MURCIA	9,407.00
NAVARRA	13,187.05
BASQUE COUNTRY	13,197.98
LA RIOJA	9,151.77

As can be observed, the poverty line in the Autonomous Communities with highest cost of living, like Catalonia, Navarra, Basque Country and Madrid, is much higher than the poverty line in the Autonomous Communities with lowest cost of living, like Extremadura, Castile La Mancha and Andalusia. If this poverty line is taken to calculate the incidence of poverty, the results changes considerably. These results can be seen in Table 13.

Table 13 Percentage of poor households below the 60% of the expenditure median deflated by the SCOL index

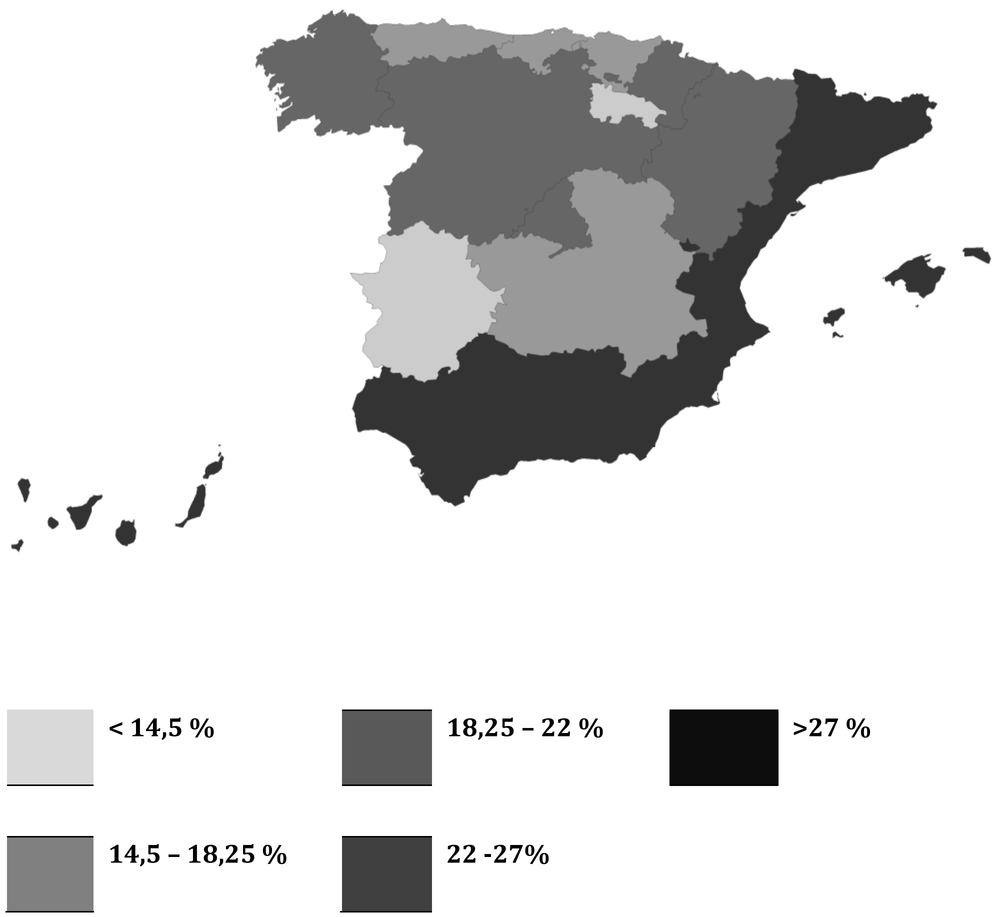
	2006	2007	2008	2009	2010	2011
ANDALUSIA	31.25	23.70	21.68	24.23	25.60	27.51
ARAGON	26.86	20.98	16.32	21.10	22.99	24.36
ASTURIAS	29.38	25.31	18.46	18.08	16.56	20.35
BALEARIC ISLANDS	23.84	15.35	17.49	27.05	26.93	29.49
CANARY ISLANDS	30.94	27.55	25.50	34.16	36.90	39.32
CANTABRIA	32.49	22.67	21.05	19.76	21.15	20.79
CASTILE LEON	29.79	25.62	20.49	21.51	24.47	22.25
CASTILE LAMANCHA	25.74	19.04	18.60	20.89	20.09	18.88
CATALONIA	35.22	31.69	27.96	31.46	33.67	33.03
VALENCIA	24.37	18.56	19.47	22.13	24.18	27.63
EXTREMADURA	28.58	21.76	18.24	21.61	21.86	17.67
GALICIA	31.39	25.74	24.14	26.90	25.72	25.16
MADRID	28.75	24.80	23.23	25.61	25.67	25.65
MURCIA	23.36	18.01	22.10	26.31	28.43	27.60
NAVARRA	29.88	25.20	21.65	25.58	25.87	25.77
BASQUE COUNTRY	27.32	23.78	21.98	22.03	20.32	20.37
LA RIOJA	19.78	18.11	13.80	13.81	15.81	14.86

The results before applying the SCOLI to the poverty line show that the poverty levels in Spain are higher than the levels calculated with the standard CPI. All Autonomous Communities increase their poverty rate and in the majority of the cases above the 25%. This result was hoped due to the increases in the COLI are higher than the increases in the CPI. In the other hand, it can be observed the different behavior across the space. The regions with highest costs of living suffer from the highest increases in poverty incidence respect to the results obtained with the CPI, in some cases next to the 10% increase respect to the results of Herrero *et al*. (2013), this occurs in Catalonia, Basque Country and Navarra. If these results are mapped it can be seen clearly the different results obtained with both methodologies.

Figure 7 Map of poverty incidence according with the standard procedure and according with the application of the SCOL index, 2011

Incidence of poverty according with the standard procedure (Table 4)

Incidence of poverty with the SCOL index procedure



It can be checked that accounting for spatial differences in the cost of living throws out a very different map of poverty than the one obtained with the standard methods. In general, poverty has increased in the whole country. It must be stand the high levels supported by the Mediterranean regions plus the Canary and the Balearic Islands. Also, the poverty incidence is high, but in a lesser extent, in Navarra, Aragon, Madrid, Castile Leon and Galicia. In contrast, the lowest levels of poverty are found in Extremadura and La Rioja.

These results underline the importance of measure accurately welfare indicators for the designing of social policies to alleviate poverty. The results presented in this section are in line with others for other countries where it is observed that in places where the wealth is concentrated are also the places where the incidence of poverty is higher. These results also are useful to reflect that if a proper price index is not be used an underestimation of poverty take place in some regions with high costs of living, while the reverse is true for regions with low costs of living. Spain faces a poverty incidence higher than that showed by the official studies, due to the increase in the cost of living together with a reduction of the household's income, this means that poverty rates over 25% have been taken place in many regions of Spain.

3.5 Conclusions

Computing a Spatial version of a Cost of Living Index (SCOLI) is of a great importance since prices are not uniform across space due to the existence of transportation costs or barriers to trade, indeed it permits measuring expenditure differentials across regions. Although the CPI published by the INE is used as an approximation of a "true" cost of living index, it does not permit neither make comparisons of prices across the space in a given moment of time, nor measure different expenditure patterns across regions. The main reason because the CPI is biased in this way is because it is calculated using an axiomatic approach that is fixing a representative basket of goods and pricing it in different places. This characteristic makes the CPI a bad approximation of a "true" cost of living, due to it does not take into account the changes in the preferences of the consumers. Another limitation of the CPI is that it does not permit comparisons of prices across the space in a given moment of time; it only permits comparisons of prices in different regions along a period of time respect to a base year of reference.

For solving this lack of accurate knowledge about prices, first, it was computed a true COLI for the median consumer of Spain for the period 2008-2012 through the AIDS approach of Deaton and Muellbauer (1980) and applying the two-step methodology of Shonkwiller and Yen (1999) to address the problems derived of having censored data in the survey. This COLI allows us to study the differences arising between the two approaches, the one of the INE (fixed basket approach) and the one followed in this thesis (fixed utility approach), due this differences to the fact that it is taken into account the substitution made by consumers because of the changes in its preferences. Secondly, a SCOLI for the 17 Autonomous Communities of Spain is calculated adapting the general AIDS methodology to the spatial case. This SCOLI allows us to know an important feature in regional economics that is know the differences in price levels across the regions in the same moment of time to make comparisons among them. Lastly, it was merged the two indices and it has been calculated a SCOLI for the period 2008-2012. This spatio-temporal index allows making comparisons across the space and along the time simultaneously. Indeed the equivalent CPI index was represented to compare the evolution of prices in the different Autonomous Communities with both indices.

All of this shows us that the differences between the evolution of the COLI and the CPI in the period studied are considerable. With the calculated SCOLI it can be found differences in 2012 between the lowest cost of living region, Castile-La Mancha (0.8104), and the highest cost of living region, Catalonia (1.0707), of around 26%. The SCOLI also shows different behaviors among regions according to the regional economics, we can see a gap between the richest regions and the poorest ones, clearly distinguishing a high price area formed by the regions with the highest costs of living (Catalonia, Basque Country and Navarra) and a low price area formed by the regions with the lowest costs of living (Andalusia, Extremadura and Castile La Mancha). This result is in line with an important proposal of the regional economics: the high income regions support higher prices than the lower income regions and the strong correlation between nominal indicators such as wages and income and the cost of living.

Indeed, it has been evidenced striking differences in the behavior of regions in all the period of study contrary to the evidence obtained with the CPI series, which do not reflect any regional price differences. With the SCOLI the evolution of the poorest Autonomous Communities and the richest ones is markedly different. While the CPI shows a homogeneous increment of prices of around a 6%-8%, the SCOLI shows that the Autonomous Communities with the highest costs of living and highest income per capita (Catalonia, Basque Country and Navarra) increase their cost of living between 2008 and 2012 up to 25%. In contrast, the poorest Autonomous Communities (Andalusia, Extremadura and Castile La Mancha) and with the lowest costs of living experienced a decrease in their cost of living since 2008.

Spatial cost of living figures were used to assess different features of the standard of living. We examine how the percentage of poor households evolves over the period 2006-2011 in Spain. The way to make this study was the recalculation of the poverty incidence in the Autonomous Communities through a new poverty line deflated by the SCOLI proposed here. It was found that price variation has a substantial influence on the estimates of poverty levels. The estimated poverty incidence changes dramatically once regional cost of living differences are incorporated. With the standard CPI the poverty incidence in high income regions like Catalonia, Basque Country or Navarra had been underestimated, while the contrary happens in low income regions like Extremadura or Castile La Mancha. Moreover, an overall substantial increase in the regional poverty levels can be observed after adjusting for SCOLI and compared to the adjustments for CPI.

In this section it has been observed empirically that substantial variations in prices across regions exist. Through the estimation of a demand system, proposed by Deaton and Muellbauer (1980), for Spanish households and for ten food groups of consumption it has been calculated first, a COLI for the median consumer in Spain and, second, a SCOLI for the 17 Autonomous Communities and for the period 2008-2012. Since the Spanish CPI provides a poor approximation of the costs of living, our proposed index is a better estimate because is consistent with the microeconomic theory and it maintains the utility level constant instead of the basket of goods as the CPI does. Moreover, the CPI only permits comparisons along time taken a base year as reference, but with our SCOLI comparisons across the space and along the time are allowed, simultaneously.

As it have been demonstrated regional variations in costs of living clearly affect the conclusions concerning regional differences in the standard of living: income levels, wage levels and poverty rates change once account is taken of variations in the cost of living across space. From a public policy point of view, more accurate information on regional prices at different spatial scales is crucial to assessments related to the regional effects of policies. The ideal framework would be to calculate a SCOLI for the different provinces and municipalities in Spain since the Autonomous Communities compel both biggest cities and rural areas as well. In fact, it can be observed that the highest cost of living regions are those that comprise the biggest and more urbanized cities of the country: Madrid, Barcelona in Catalonia, Bilbao in the Basque Country; while the lowest cost of living Autonomous Communities are those that have the highest proportion of rural population: Extremadura, Andalusia and Castile La Mancha.

The existence of regional disparities between Spanish regions implies the existence of agglomerations at a small spatial scale. These regional agglomerations are reflected in the variety of cities, the Autonomous Communities of Spain comprise big metropolitan areas and small rural areas as well, and also these cities are characterized for being both central and peripheral. For these reasons the estimation of a SCOLI at a small spatial scale would be the ideal framework to study the spatial dynamics of the cost of living. Since the official statistics only provide information at a large scale this goal is thwarted. But at least the HBS provide information about the type of the city in which the household resides that is if it is a big city, a medium-sized city or a small one. This information allows calculating the effect of agglomerations over consumption patterns and costs of living, something that will be studied in the next chapter.

4 COST OF LIVING AND CITY SIZE

4.1 Introduction

In previous chapter the work has been focused in the existence of regional price differences at a large scale. But in the spatial context, there is a crucial issue that has been the center of many researches that is not only whether are important the differences between regions, but also the variations within these regions. It is reasonable that if cost of living varies between Autonomous Communities, it should also varies among provinces or municipalities within an Autonomous Community. More specifically, what is of a great interest is to go beyond the regional dimension and focus the study in the differences between urban and rural areas.

It is evident that urbanization, income, agricultural productivity and industrialization are deeply interconnected processes (Bairoch, 1988). The directions of causality between these phenomena are manifold and exert a feedback effect in the process of economic development of a nation (Polèse, 2005). Economic modernization and nascent industrial development lead to increases in agricultural productivity. These increases have facilitated the migration of the rural population to urban areas where industrial development is concentrated. Increases in income transform consumption structures (Engel's Law, 1821-1896); there is demand for more industrial and tertiary products and a decrease in the participation of agricultural products in overall income. In the process of sectoral change, when all these dynamics come into play, significant gaps in income between rural and urban settings may result in very different consumption patterns between the two contexts.

The connection between the level of household income and consumption structure was examined by Houthakker and Taylor (1970). These authors used US data for the first three decades of the twentieth century and showed that estimates of own-price and expenditure elasticities were significantly different between income quartiles. Ahmed and Shams (1994), worked only with foodstuff demand in the case of Bangladesh and found striking differences between the estimated elasticities of poor and rich consumers, showing that low-income households are more price responsive than those with high incomes. Another example can be seen in a study of the consumption of animal products in Turkey over five income groups, which also shows considerable differences in consumption between these groups (Armagan and Akbay, 2008). Based on all this evidence, it is therefore likely that changes in prices or income affect high-income and lower-income households very differently. Inasmuch as there are strong spatial differences in both income levels and the distribution of income groups, it is also understandable that such adjustments will have different effects within a nation. However, the focus of this chapter explores whether there are other spatial patterns in addition to those that may result from a simple concentration of income in space.

There is an extensive body of empirical literature on how own-price elasticities and expenditure elasticities vary depending on household income or life cycle, among other factors. However, the number of studies exploring the relationship between space and elasticities or, more generally, consumption patterns, is very limited. The first objective of this chapter proposes and analyses the consumption patterns of households along different city sizes in the context of a developed country. If the influence of the spatial setting is relevant, this would imply that measures that affect prices, or situations that reduce income, such as the current crisis, will generate sharply differentiated spatial effects depending on the size of the city. It will be explored specifically the degree to which the demand elasticities vary across the city sizes in Spain. The study, again, focuses only on food consumption due to the statistical limitations; however, the characteristics of the food products and markets make them suitable for an analysis like this.

Theoretical literature gives us some important clues to explain why the size of the cities could affect the prices and consumption patterns. Classical approaches in Regional Economics put their attention in how the distance and, consequently, the transportation costs affect the land prices and specialization across space. The classical works of Christaller, Lösch or Von Thünen predict higher costs in the central (larger) places. The monocentric city model proposed by Alonso (1964), Mills (1967), Muth (1969) and Wheaton (1974), which could be considered as an integration of all these classical theories, explains how the center of cities concentrates the activities that generate the highest added value per square meter with strong pressures over the land prices and, consequently, pushing up the prices in general. According with it, higher cost of living is expected in large cities, especially in the center and/or its neighbor areas close to it (see Haworth and Rasmussen, 1973).

The idea of the pressure over land prices explained by the monocentric model takes relevance if we combine it with the concept of agglomeration economies. The agglomeration economies, first proposed by Weber (1909) and later developed by Ohlin (1933), Hoover (1937) and Isard (1956), connect the attractiveness of a place with their size in terms of population and business density. When in a reduced space there is a large concentration of people and business economic activity, the possibilities of interaction between economic agents are much more intense. Proximity encourages formal and informal exchanges of ideas which nourish innovation and contributes to the diffusion of knowledge; large numbers of potential customers make specialization to be possible, which enhances the variety of specialized skills offered; a sense of community is the basis on which to pool public services and achieve economies of scale in their production; and so on. According with this concept, the pressure over the land in these large metropolises with stronger agglomeration economies will be higher and, consequently, largest cities should be remarkably more expensive. The level of differences in cost of living gives indirect information about how relevant the agglomeration economies are. General equilibrium models of systems of cities building on Henderson (1974, 1987) also show that cities specializing in sectors with stronger agglomeration economies have more expensive land, which offsets the higher wages resulting from agglomeration economies.

The New Economic Geography (NEG) framework (Krugman, 1991; Krugman and Venables, 1995 and Fujita and Krugman, 1995) gives a different view of the same processes of concentration and its effects over the economic dynamics but less focused in the urban perspective. The original models consider two regions, core and periphery, with two types of products: transportable (manufacturing) and non-transportable (agricultural) goods. Under the general conditions and assumptions of the basic model, if transportation costs are high-enough, all the producers of transportable goods will be agglomerate in the core (the main urban area). So the core will become larger, and the transportation cost supported there will be lower. The periphery, (rural places), will produce just non transportable goods (agriculture) so the prices in transportable ones (manufactures) will be higher. Under these conditions, cost of living in rural areas should be higher and it should be lower in large central areas. This conclusion of NEG initial models is at odds with real-world experiences. Some products are certainly cheaper in agglomeration areas like New York, London or Madrid in the case of Spain. But on aggregate average, however, these places tend to have higher costs of living than locations in the remote rural periphery. Taking this into account, recent contributions also within the NEG framework open different perspectives more adjusted to the real world behavior. Essentially, the first NEG models make this prediction of lower prices in agglomerated central places because they ignore land scarcity and higher housing prices, the main reason why agglomerations are more expensive (Tabuchi, 2001). Models that incorporate these factors do not reach the same conclusion. Helpman (1998) replaces the standard agricultural sector in Krugman (1991) with an immobile housing stock concluding that there are higher costs of living in central areas. Recently, an extensive literature has been tried to incorporate into the core-peripheral model the housing and the non-tradable services effects (see Tabuchi, 1998; Tabuchi and Thisse, 2003; Tabuchi *et al*., 2003; Cavaihès *et al*., 2004 and or Suedekum, 2006; among others) and being the general conclusion the same: the variety of goods and services is larger in central places but the cost of living is higher.

Summarizing, theoretical models predict, in general, a city size effect on the cost of living that makes large cities more expensive places to live in. But these models are focused in the effect of land pressure and they do not consider how the higher variety of products and firms could affect the prices through a greater market competition or by means of changes in the consumption. At this point we need to review what empirical analysis tells about this question.

Although the empirical evidence seems conclusively in line with theoretical approaches all the researches use as a measure of cost of living a Consumer Price

Index or other kind of price index generally published by governmental sources. In order to make a proper comparison of the cost of living in different places a "true" cost of living index should be calculated. As in the previous chapter it will be followed the Konüs (1939) approach for solving these limitations by calculating an indicator of cost of living at the same level of utility. Cost of living at the same utility level has the advantage, when compared with the calculation of cost of living with the same basket of commodities, of avoiding the substitution bias derived from neglecting the substitutions made by consumers in response to price variations. Considering that the large cities offer more variety of goods and consumption possibilities that could be used by the inhabitants to avoid the higher prices, this improvement in the way of measurement the "true" cost of living could be much more precise and give different conclusions. The number of applications that analyze the city size effect with a "true" cost of living strategy is certainly reduced. Ravallion and Van de Walle (1991), in their study for Indonesia use an AIDS model and found that the urban cost of living for food staples is considerably higher than in the rural areas. No previous attempts have been made to calculate a "true" COLI for different city sizes. In previous chapter it was evidenced how the more urbanized regions of the Spanish economy present higher cost of living, but from that study we cannot evaluate the relation between city size and cost of living since their results were presented only at a NUTS-II level. This chapter tries to extend this analysis by explicitly calculating the "true" cost of living for the different municipality sizes.

Different consumption patterns exist depending on whether the household resides in a large city, a small or medium-sized town or a rural area in the case of a country like Spain. The changes affect own-price elasticities more than expenditure elasticities. Although the latter vary depending on different city sizes, there are exceptions with respect to types of products. However, own-price elasticities are always higher in larger cities. Price sensitivity is higher in large cities compared to smaller towns. Additionally, not only consumption patterns change across the urban structure but, also striking differences in the cost of living between the biggest cities and the smallest ones. The highest cost of living is found in the biggest municipalities with more than 100,000 inhabitants, contrary, the lowest costs of living are found in the smallest municipalities with less than 10,000 inhabitants.

It has been evidenced that there exist relevant variations in the consumption patterns and the cost of living across the size of the cities. But which are the causes of these differences? Why some regions have higher costs of living than others? Why is more expensive to live in large cities? It could be that the biggest cities attract a particular population, young people with the highest incomes, the highest wages and the most qualified people making that the demand of goods rise and then the pressure over prices is higher in agglomerations. But also, it could be that the same household or individual with the same characteristics faces higher costs of living and presents different consumption behaviors in the biggest cities than in the smallest ones by the fact that the agglomerations *per se* have higher prices and promote a particular consumption which is not found in small areas.

To ask this question the estimation of a quantile regression model of the factors that explain the cost of living is going to be estimated. We estimate a quantile regression using our own estimates of cost of living at a household level. This way, a cost of living for each household is calculated to regress it over some geographical variables (like income per capita, agglomeration variable and regional variables); and other variables related to the household characteristics (like number of members, income level of the household, education level of the household head, age and number of employees and dependents). These socioeconomic variables help to isolate the pure effects of the geographic variables over the cost of living.

The main purpose of this chapter is to show empirical evidence about the importance of geography in the determination of the costs of living, confirming the spatial patterns of the costs of living in Spain. The rest of the chapter is organized as follows. In Section 4.2, the demand elasticities by city size are calculated, and the consumption patterns by city size are evaluated. In Section 4.3 a COLI by city size is calculated. In Section 4.4 it is tried to determine the factors that influence in the cost of living of the households, focusing the attention in the effect of the agglomeration over this cost of living. Finally, it is concluded in Section 4.5 with a synthesis of the whole chapter.

4.2 City size and consumption patterns: food elasticities in Spain

There have been previous empirical studies that suggest that the place of residence, in an urban or rural area and even depending on the size of the city, also affects consumer behavior in a similar fashion to variations in income levels or stages in the life cycle. Nevertheless, the empirical evidence is limited; most of the studies are focused in developing country contexts, and researchers have found different results depending on the country and period of the analysis. For instance, Abdulai *et al*. (1999) find very different patterns of food consumption that they explain as a consequence of the different demographic and sociological characteristics of the urban areas in comparison with traditional rural India. Burney and Akmal (1991) estimate own-price and expenditure elasticities in Pakistan, separating samples between rural and urban areas as well as into six different income levels. They found that the demand responses for different food items vary between urban and rural areas, as well as by household income. In general, urban areas are more sensitive to the income changes; regarding own-price elasticities, their magnitude depends on the product characteristics. Haq *et al*. (2008) likewise conducted an estimation exercise for Pakistan nearly two decades later. Their study, which focused on the impact of rising food prices on poverty in rural and urban areas, once again identified significant differences between rural and urban contexts with similar conclusions as the previous study of Burney and Akmal (1991). Similar conclusions were drawn by Alfonzo and Peterson (2006) for Paraguay. Finally, Lewis and Andrews (1998) conducted an analysis for China concluding that there are important differences between China's largest metropolis and rural areas. This study shows that the city size, not only the distinction between rural or urban areas, also has an effect on elasticities and consumption patterns. In all cases, the differences between rural and urban areas are significant at a comparable level to the influence that has been found for the variations in the income level of the family. Normally, comparing for the same level of income, cities appear more sensitive to the changes in prices or income.

Based on prior work, what can be said about variations across space apart from urban-rural differences that have been the focus of much of the literature to date? For example, why would one expect larger cities to have different consumption patterns than those of small towns or rural areas? Is this something that occurs only in developing countries and is it also present in already developed countries that have much higher levels of urbanization? To what extent are these differences relevant and do they have an impact on urban dynamics? Will these variations generate a differential spatial impact among different cities (according to their size) from policy measures?

4.2.1 An overview of the consumption patterns in Spain

In this section it will be explored specifically the degree to which the demand elasticities vary across the city sizes in Spain. The study focuses only on food consumption due to the statistical limitations; however, the characteristics of the food products and markets make them suitable for an analysis like this.

The derivations of elasticity formulas for the AIDS model are found in Green and Alston (1990). The expenditure elasticity (η_i) and the uncompensated cross- and own-price elasticities (ε_{ij}) take the following form:

$$\eta_i = 1 + \frac{\beta_i}{w_i} \tag{31}$$

$$\varepsilon_{ij} = -\delta_{ij} + \frac{\gamma_{ij}}{w_i} - \frac{\beta_i \alpha_j}{w_i} - \frac{\beta_i}{w_i} \sum_k \gamma_{ij} \log P_k \tag{32}$$

Where δ_{ij} is the Kronecker delta, which takes the value of 1 if *i=j*, and the value 0 for *i ≠j*.

Table 14 shows the income and own-price elasticities for the whole sample, comprising almost 22,000 households, for Spain in 2010. It should be remembered that the expenditure elasticity measures the change in demand in terms of quantity and quality given a change in expenditure. Meanwhile, the own-price elasticity measures the change in demand given a change in the price of the product.

The results for the Spanish economy are similar to those found in previous studies that used data from a decade ago (Gracia *et al*., 1998 and Angulo *et al*., 2002). Several food groups, such as *Meat, Fish, Oils & fats, Fruits* and *Vegetables* have expenditure elasticities greater than 1; however, most of the products present expenditure elasticities lower than 1 confirming the Engel law. There are also several groups of goods with own-price elasticities greater than or close to 1: *Oils & fats, Fruits, Vegetables, Coffee, tea & cocoa and Mineral waters & soft drinks*. This means that Spanish households are quite responsive to changes in those prices. Nevertheless, most of the products present own-price elasticities smaller than 1. Those that have the lower values, close to 0.5, are the staple food products in a standard family diet like *Bread & cereals, Meat*, or *Milk, eggs & cheese.*

In general, the estimates for Spain are more similar to those reported by Pakistan, India and China by the aforementioned authors such as Alfonzo and Peterson. (2006), Sheng *et al*. (2008) and Huq and Arshad (2010) than those found by the United States Department of Agriculture (2007) for Belgium, the USA, Canada, France, Denmark, Germany and the UK. However, many of these differences may be attributable to the methodological approach. Overall, the estimates based on the AIDS procedure tend to yield higher elasticities than those found using other estimation methods.

Table 14 Expenditure and own-price elasticities for the whole sample (2010)

	Whole Sample		
	Expenditure	Own-price	Share
Bread & cereals	0.609	-0.540	15.281
Meat	1.231	-0.644	25.010
Fish	1.442	-0.905	12.103
Milk, eggs & cheese	0.868	-0.717	14.583
Oils & fats	1.462	-1.301	2.297
Fruits	1.013	-1.006	10.649
Vegetables	1.025	-0.953	10.705
Sugar	0.768	-0.763	2.826
Coffee, tea & cocoa	0.843	-1.079	1.823
Mineral waters & soft drinks	0.086	-0.974	4.717

Table 15 presents the results for the same ten groups of products, though now divided into two sub-samples: the first three columns correspond to households that are in the first income quartile, i.e., lower-income earners, and then, the last three columns, indicate households that are in the third income quartile, i.e., higher-income earners. As Table 14, Table 15 shows: expenditure elasticities, first and fourth columns; own-price elasticities, second and fifth columns; and expenditure shares, third and sixth columns. The results match expectations. In most food groups, there are no changes in sign or changes from elastic to inelastic products in terms of aggregate behavior. However, for many groups of goods, there appear to be remarkably differences between the first and third income quartiles.

Considering expenditure elasticities, note that regarding highly elastic goods with expenditure elasticities greater than 1, such as *Meat*, *Fish* or *Vegetables*; the expenditure elasticities for lower-income households are remarkably higher than those for higher-income households. This also occurs in products that, even though they are not elastic, present expenditure elasticities approaching the value of 1, such as *Milk, eggs & cheese* and *Fruits*. For more inelastic products, with expenditure elasticities far below 1, like *Bread & cereals* or *Sugar,* the first quartile expenditure elasticities are lower than those of the third quartile. This indicates that reductions in earnings in lower-income households would lead to a rapid reduction in the consumption of more luxury foodstuffs, with these households consuming cheaper products than higher-income households. In the context of an economic crisis, when reductions in earnings are generalized, it can be noted that the purchasing structure and the quality of consumption is affected much more rapidly in lower-income households with a reduction of the consumption of higher quality and higher protein value products compared to high-income households, whose consumption structure is less sensitive to changes in earnings.

The behavior pattern is similar with respect to own-price elasticities, though even more pronounced. Except for *Vegetables*, all own-price elasticities are higher for lower-income households. Low-income households are much price responsive than high-income households. This greater sensitivity is more pronounced the greater the elastic of the product.

All these results are consistent with microeconomic theory and empirical evidence found in similar studies cited previously, such as Bouis (1990), Ahmed and Shams. (1994), Abdulai *et al.* (1999) and Alfonzo and Peterson*.* (2006) and Menezes *et al*. (2008). They are also consistent with the evidence found in the Spanish case by Gracia *et al*. (1998) and Angulo *et al*. (2002).

Table 15 Expenditure and own-price elasticities by income quartiles (2010)

	1st quartile global			**3rd quartile global**		
	Income	**Own-price**	**Share**	**Income**	**Own-price**	**Share**
Bread & cereals	0.422	-0.582	0.192	0.732	-0.526	0.126
Meat	1.290	-0.792	0.213	1.201	-0.522	0.275
Fish	1.526	-1.233	0.098	0.992	-0.670	0.144
Milk, eggs & cheese	1.007	-0.721	0.164	0.880	-0.679	0.130
Oils & fats	1.612	-1.673	0.022	1.767	-1.081	0.025
Fruits	1.102	-1.074	0.123	0.849	-1.003	0.101
Vegetables	1.125	-0.948	0.114	1.094	-0.992	0.102
Sugar	0.556	-0.904	0.029	0.751	-0.673	0.028
Coffee, tea & cocoa	0.828	-1.353	0.018	0.771	-0.840	0.018
Mineral waters & soft drinks	-0.103	-1.498	0.025	0.862	-0.949	0.050

4.2.2 Elasticities by city size

To what degree is there a city size affect in the variation of the elasticities? This is the objective whose results are presented in Table 16, where once again the same ten product groups divided between the 1st and 3rd quartiles are presented, with the additional dimension of municipalities by size according to the scale that the database facilitates. The results show that the size of the city where the household resides has an influence almost as relevant as that of its level of income.

Although there are differences among the groups of goods, in general, with no relevant differences between levels of income, expenditure elasticities are higher in the larger cities. Lower values of expenditure elasticities are observed for smaller cities with less than 50,000 inhabitants, and this is particularly the case in rural areas with less than 10,000 inhabitants. The elasticities rise remarkably in medium-sized cities with more than 50,000 inhabitants. It is not possible to observe relevant increments in moving from medium cities to those of more than 100,000 inhabitants. This is not the case in certain groups of goods like *Coffee, tea & cocoa*, *Meat* or *Fruits* that maintain almost the same expenditure elasticity for all type of cities.

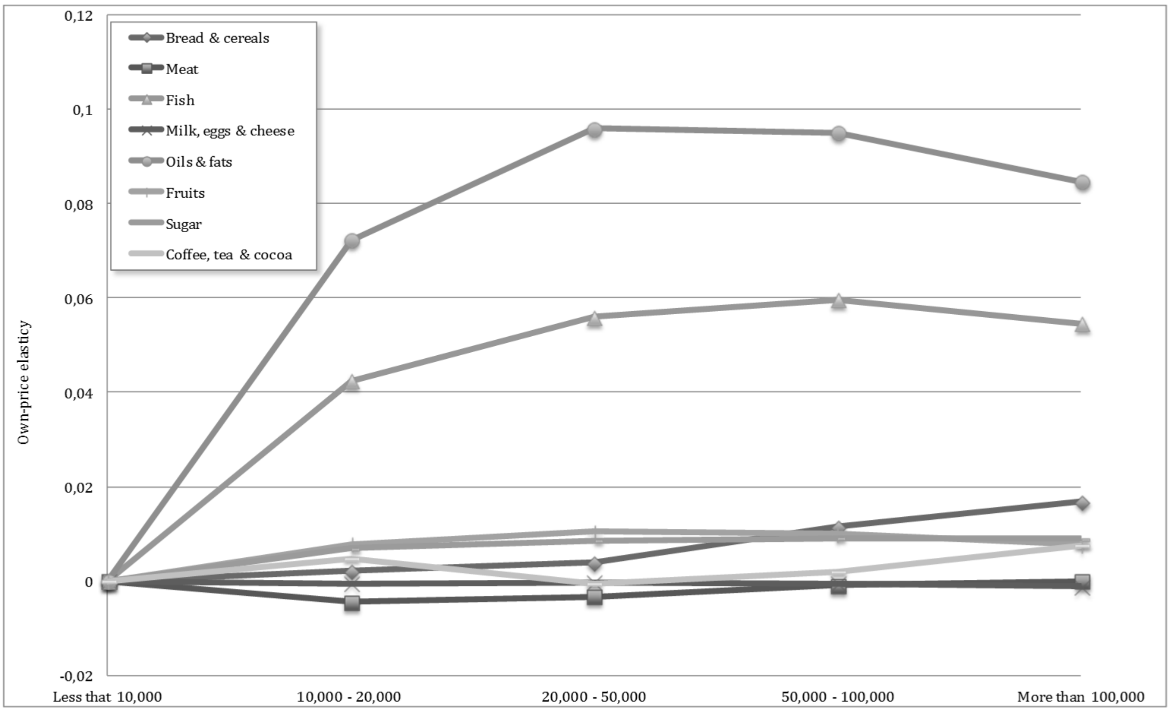
In virtually all cases, there are only two exceptions: own-price elasticities are more sensitive the higher the size of the city; the less elastic values are normally found in small towns (rural areas) with fewer than 10,000 inhabitants. This occurs for both income levels, but is more pronounced among lower-income households. The group *Fish* is one that presents larger differences depending on the income level of the household and the city size: in families with higher income levels, this product presents own-price elasticity close to 1 and the variations across city size are very small. In lower income level families, *Fish* is a very elastic good and the own-price elasticity is very sensitive to the size of the city, higher in medium and large cities than in the small ones or rural areas. Something similar happens with *Bread & Cereals* but less intensively: the evolution along the different city sizes is very similar for both quartiles, but the increments in own-price elasticities are higher for the lower-income level families. The rest of the groups of goods also show higher own-price elasticities in medium-size or large cities than in rural areas. The virtual exceptions are with respect to *Milk, cheese & eggs* and *Meat*.

Table 16 Expenditure and price elasticities by municipality size (2010)

	Expenditure elasticities					Own-price elasticities				
	1st quartile					1st quartile				
	more than 100,000	50,000-100,000	20,000-50,000	10,000-20,000	less than 10,000	more than 100,000	50,000-100,000	20,000-50,000	10,000-20,000	less than 10,000
Bread & cereals	0.429	0.425	0.419	0.415	0.394	-0.583	-0.580	-0.576	-0.575	-0.574
Meat	1.293	1.294	1.293	1.292	1.296	-0.795	-0.795	-0.793	-0.793	-0.796
Fish	1.557	1.566	1.563	1.543	1.481	-1.258	-1.264	-1.260	-1.244	-1.193
Milk, eggs & cheese	1.009	1.009	1.009	1.009	1.010	-0.721	-0.721	-0.722	-0.721	-0.722
Oils & fats	1.702	1.717	1.719	1.685	1.583	-1.757	-1.774	-1.775	-1.737	-1.612
Fruits	1.110	1.112	1.113	1.110	1.098	-1.085	-1.094	-1.097	-1.087	-1.049
Vegetables	1.133	1.134	1.135	1.133	1.127	-0.952	-0.955	-0.955	-0.953	-0.945
Sugar	0.566	0.570	0.569	0.553	0.493	-0.903	-0.904	-0.903	-0.902	-0.895
Coffee, tea & cocoa	0.839	0.842	0.843	0.840	0.843	-1.390	-1.382	-1.379	-1.386	-1.371
Mineral waters & soft drinks	0.273	0.392	0.426	0.324	-0.234	-1.296	-1.232	-1.212	-1.268	-1.547
	3rd quartile					3rd quartile				
Bread & cereals	0.742	0.735	0.742	0.728	0.714	-0.533	-0.527	-0.533	-0.521	-0.508
Meat	1.203	1.201	1.204	1.200	1.198	-0.524	-0.523	-0.522	-0.522	-0.520
Fish	0.992	0.992	0.991	0.992	0.993	-0.667	-0.669	-0.668	-0.671	-0.672
Milk, eggs & cheese	0.884	0.882	0.883	0.879	0.874	-0.680	-0.679	-0.680	-0.678	-0.675
Oils & fats	1.892	1.768	1.839	1.727	1.647	-1.102	-1.081	-1.093	-1.074	-1.059
Fruits	0.847	0.848	0.847	0.851	0.854	-1.006	-1.004	-1.006	-1.001	-0.997
Vegetables	1.098	1.095	1.098	1.093	1.089	-0.999	-0.994	-0.998	-0.989	-0.981
Sugar	0.763	0.754	0.754	0.747	0.737	-0.685	-0.676	-0.676	-0.668	-0.657
Coffee, tea & cocoa	0.784	0.774	0.777	0.771	0.751	-0.846	-0.841	-0.843	-0.840	-0.829
Mineral waters & soft drinks	0.873	0.858	0.875	0.843	0.815	-0.938	-0.944	-0.938	-0.949	-0.956

Figure 8 Own-price elasticity variations by municipality size (2010)

1st quartile, all groups of products



1st quartile, less sensitive groups of products

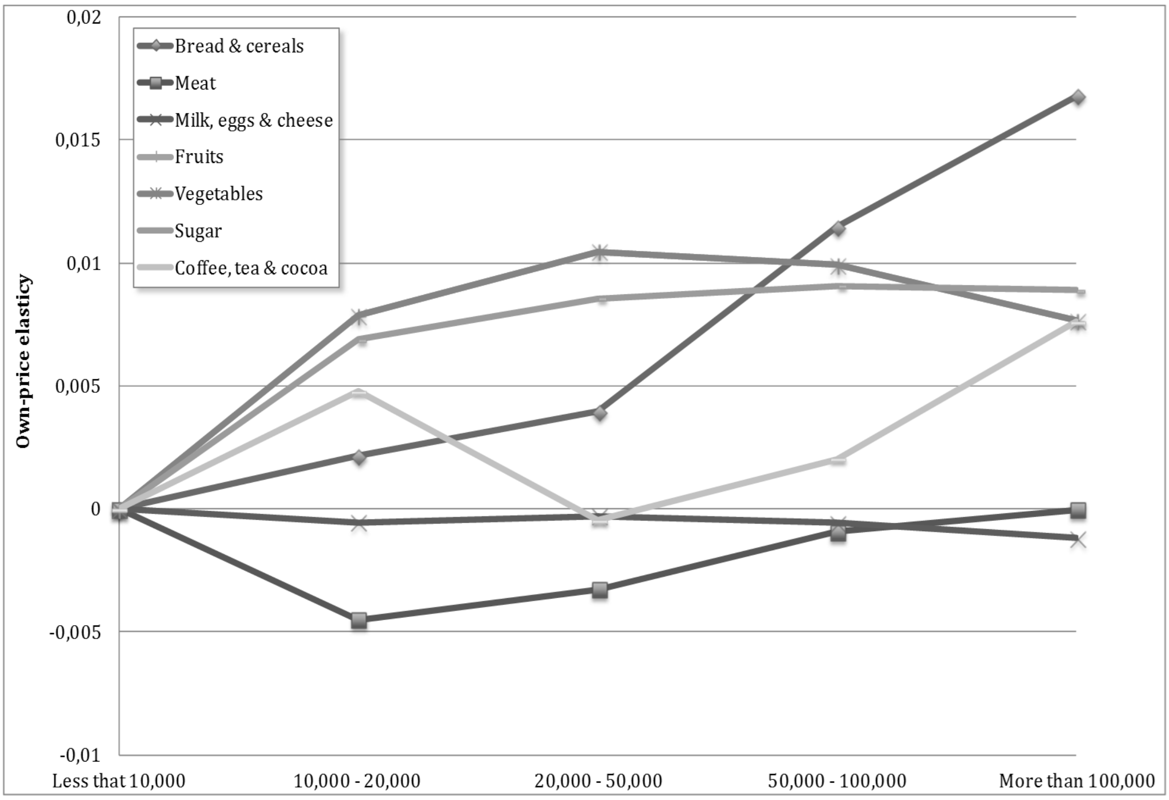


Figure 8 *graphically* summarizes these results, allowing them to be more clearly identified. Viewing this figure, one can conclude that when a price increase occurs in the largest cities, businesses and shops will be forced to respond to it with a more pronounced reduction in their profit margins than those similar businesses or shops in smaller cities or rural areas.

Summarizing the different results presented in the tables, the following conclusion can be offered for the Spanish case: (i) expenditure and own-price elasticities show that, in general, households with lower income levels are more sensitive to changes; (ii) also both elasticities, but much more clearly the own-price elasticities, in large cities are more reactive to the changes than in small cities or rural areas. The database does not distinguish between areas or neighborhoods within cities and there are likely to be remarkably intra-city variations especially in the largest cities where there is likely to be pronounced segregation by income. Inasmuch as the results revealed that lower-income households (1st quartile of earnings) present higher own-price elasticities than higher-income households (3rd quartile), one might posit that in neighborhoods or areas where lower-income households tend to concentrate in large cities, increases in costs or taxes will be mainly supported by businesses or shops, whereas they are passed on to a greater extent to the consumer in higher-income areas. This differential response may be more clearly pronounced in small cities or rural areas. This will result in a highly differentiated business landscape within the city and different prices within it depending on areas or neighborhoods as well as the size of the city. Although this is the general presumption, we also have to consider that the process may break down if high-income consumers shop in neighborhoods where prices are lower. During the recession, consumer surveys showed that higher income households adjusted to the real reduction in their income by shopping at discount stores: lower income households responded by consuming less.

The economic policy conclusions from these results suggest that the increase in VAT rates may cause different effects that have to be analysed between areas (urban versus rural) and between types of city (larger versus smaller) together with variations associated with income levels. In large cities, particularly in neighborhoods where there is a concentration of lower-income households, the reaction to a price change will be very intense and probably it will likely be absorbed to a greater extent by businesses. But the behavior will be different, less sensitivity to the price changes, in small cities, especially if they belong to a region of high-income levels or if they belong to neighborhoods that concentrate the higher-income families.

4.3 The cost of Living by city size (2008-2012)

Most of the applications concerned in evaluating the city size effect conclude that there is a positive relationship between the size of the city and the costs of living. The majority of the studies are made for the US economy in different moments of time and using different sources of information for prices. One of the first studies was Alonso (1970), based on data from the US Bureau of Labor Statistics (BLS) for all US urban system and for 1967. He found that the cost of living was positively correlated with the urban size, especially for higher incomes. With the same dataset but for 1970, Haworth and Rasmussen (1973) studied the determinants of the differences in cost of living among cities with a model that explained over 60% of its variations among metropolitan areas using three independent variables: the city size, the city form and regional dummies. Using again the dataset from BLS but with microdata for 193 commodities in the US metropolitan areas Simon and Love (1990) found that there are more items with prices positively correlated with the size of the city than cases where this relation was negative. Cebula and Todd (2004) made an analogous approach with similar conclusions, but just for the Florida state and using another source of price information: the Statistical Abstract of the United States. Walden (1998) analysed the ACCRA[8] dataset, obtaining additional evidence of locational price differences. This author studied data of 20 North Carolina communities for the period 1991-1994, finding again a positive relation between population and a price index. More recently, Kurre (2003), found the same relation between cost of living and population size with the same dataset for all the US counties.

[8] American Chamber of Commerce Researchers Association. Although this data is called cost of living index, it is a price index that it calculates the cost of buying the same products and services in different locations.

There are a remarkably lower number of applications for other countries. In Europe most of the studies are made within a regional context, like Blien *et al*. (2009) or Roos (2006) for Germany, which base on price indices from the German Federal Statistical Office. In both cases they find higher price levels in regions with large agglomerations. Other authors apply this approach for developing countries, finding stronger differences between the urban and rural locations than in US or Europe, which was expected because these countries are experiencing intense changes due to the process of industrial development and urbanization. For instance, Asra (1999) studies the impact of poverty measures in Indonesia on the urban-rural cost of living differentials by applying a chain Laspeyres price index and concludes that urban areas had higher costs of living. The same conclusion is found for the case of Thailand by Kakwani and Hill (2002) using several price indices. More recently, Majumder *et al.* (2012) calculate the rural-urban differentials in food price in India through the estimation of unit values as a proxy of prices.

4.3.1 Estimation of the Cost of Living Index by city size

All the empirical works quoted previously use as a cost of living official statistical sources published by statistical agencies, in reality the number of applications that analyze the city size effect with a "true" cost of living is very reduced. In the estimation of the COLI by city size proposed in this thesis it will be followed the Konus' approach of the "true index of the cost of living". As in many other countries, the estimation of the COLI by city size to the Spanish case entails the difficulties arising from the lack of available data. As we say before, the estimation of the AIDS requires information on prices, quantities and expenditures at the household level. As all the prices must be observable to estimate the model, the unitary values at which households purchase the commodities are recovered by dividing expenditures by quantities. All these information requirements limit the estimation to be feasible only for the food group, being the only type of product studied in the HBS with detailed information about the variables required. The data of these products are classified into ten food sub-groups: (i) *Bread and cereals*, (ii) *Meat*, (iii) *Fish*, (iv) *Milk, cheese and eggs*, (v) *Oil*, (vi) *Fruits*, (vii) *Vegetables*, (viii) *Sugar*, (ix) *Coffee, tea and cacao*; and (x) *Mineral water and soft drinks*. As in previous cases, the problem of the censored data is still present in this estimation. Again, it will be solved following the procedure of Shonkwiller and Yen (1999).

In this case it is estimated the same model as explained in section 3.2 of the previous chapter, we incorporate a factor to consider the spatial heterogeneity, in terms of unobservable city size and regional characteristics.

$$w_i = \Phi(x)\left[\alpha_i + \sum_{j=1}^{n} \gamma_{ij}\, log\, p_j + \beta_i\, log\{x/P\}\right] + \sum_k c_k\, CS_k + r_h R_h + \delta_i \phi(x) \qquad (33)$$

where δ_i is a parameter associated with the density function, CS_k are dummy variables for different urban sizes and R_h is a regional dummy for each one of the NUTS-II regions of Spain, and c_k and r_h are the parameters associated with each type of dummy, respectively, with the aim of recover the idiosyncratic components inherent to each region and type of city. Note that this is a particular characteristic in our formulation, since we incorporate a factor to consider the spatial heterogeneity, in terms of unobservable city size and regional characteristics. The information contained in the HBS allows for distinguishing between five types of municipalities according to their population sizes:

1. Municipalities of more than 100,000 inhabitants
2. Municipalities between 50,000 and 100,000 inhabitants
3. Municipalities between 20,000 and 50,000 inhabitants
4. Municipalities between 10,000 and 20,000 inhabitants
5. Municipalities of less than 10,000 inhabitants

The set of n– 1 equations like (33) conform the demand system, where n is the number of shares, being the last share recovered as a residual of the remaining n– 1 ones. Once this demand system is estimated, the parameters are used to recover the expenditure function of a representative household for each city size and to calculate the index defined in Equation (28) by dividing the expenditure function of the consumer of each city in each year by the expenditure function of the consumer taken as reference. In this case, the common utility level set in this calculation is that of the municipalities of more than 100,000 inhabitants in 2008.

The estimation process and the estimator used is the same as in the case of the SCOLI for Autonomous Communities. Also, all the parameter estimates of the Probit and AIDS models for each year can be found in the Appendix 1 to Appendix 7.

4.3.2 Results

Once the value of the expenditure functions is calculated, they are used to calculate the index defined in Equation (28). The utility level of the representative household in the municipalities with more than 100,000 inhabitants in 2008 is taken as reference for the computation of the index. This implies that the results reflect the cost of attaining the utility level of the household with the median expenditure in the municipalities with more than 100,000 inhabitants in that year. Results are presented in Table 17.

Table 17 Cost of Living (€) by type of municipality, 2008-20012

	2008	**2009**	**2010**	**2011**	**2012**
>100,000	4041.96	4096.62	4117.31	4500.84	4524.83
50,000-100,000	4073.26	3866.95	3959.44	4296.07	4253.70
20,000-50,000	4060.88	4035.65	4047.84	4357.48	4413.17
10,000-20,000	4045.12	3882.30	3959.00	4261.49	4311.40
<10,000	4033.50	3786.56	3844.46	4142.00	4176.87

Results in Table 17 should be read as the expenditure required in each type of municipality and each year to attain the same level of utility as the household of reference in 2008. In order to see more clearly the cost of living figures, the same results reported in Table 17 are presented in Table 18 in the form of an index. In this table, estimated costs of living have been divided by the estimate cost of living for the representative household.

Table 18 Cost of living index by type of municipality (2008-2012). Base: representative household in cities with population >100,000 in 2008

	2008	2009	2010	2011	2012
>100,000	1.0000	1.0135	1.0186	1.1135	1.1195
50,000-100,000	1.0077	0.9567	0.9796	1.0629	1.0524
20,000-50,000	1.0047	0.9984	1.0015	1.0781	1.0918
10,000-20,000	1.0008	0.9605	0.9795	1.0543	1.0667
<10,000	0.9979	0.9368	0.9511	1.0247	1.0334

Results in both tables suggest that the smallest areas benefit from reduced costs of living when compared with the largest cities of Spain. The estimates of cost of living by city size seem to be coherent with the expectations about the effects of agglomeration economies in recent literature of New Economic Geography (see, for example, Helpman, 1998; Tabuchi, 2001; Tabuchi and Thisse, 2003; Cavaihès *et al*., 2004; and Suedekum, 2006). The reported results are consistent not only with this theoretical literature, but also with many of the empirical works, most of them for the US (see, for example, Alonso, 1970; Haworth and Rasmussen, 1973; Simon and Love, 1990; Walden, 1998; Kurre, 2003; Cebula and Todd, 2004). The estimates indicate that the largest cities have suffered the highest cost of living all along the period under study, being the smallest municipalities the areas where these estimates get the lowest values. These differences range from around more than 10% in 2011, suggesting that the higher market competition and the wider variety of products present in large cities are not enough to offset the spatial competition and land pressure that characterize these big municipalities.

Differentials in costs of living between small and medium-size cities (those in the intervals between 20,000 and 100,000 inhabitants) are smaller and fluctuating along time. Within this category there are a large number of municipalities located close to the main metropolitan areas of the country: Madrid and Barcelona mainly and also Bilbao, Valencia, Seville and Zaragoza. According with Polèse *et al*. (2007) or, more recently, Viñuela *et al*. (2010), almost 2/3 of the Spanish population live in central areas. These areas are defined as those strongly interconnected with a large metropolis and located closer than one hour driving from them. This type of municipalities will be directly influenced by the higher prices on the main metropolis, which could explain why in some cases smaller places present higher costs of living.

Besides spatial comparisons among types of cities, the results allow for identifying some dynamics in the cost of living during the period 2008-2012. In general, the estimates reveal an upward trend in the cost of living between the initial and the final year with available data. Figure 9 plots the indexes reported in Table 2 along the period 2008-2012.

Figure 9 Cost of living index by type of municipality (2008-2012). Base: representative household in cities with population >100,000 in 2008

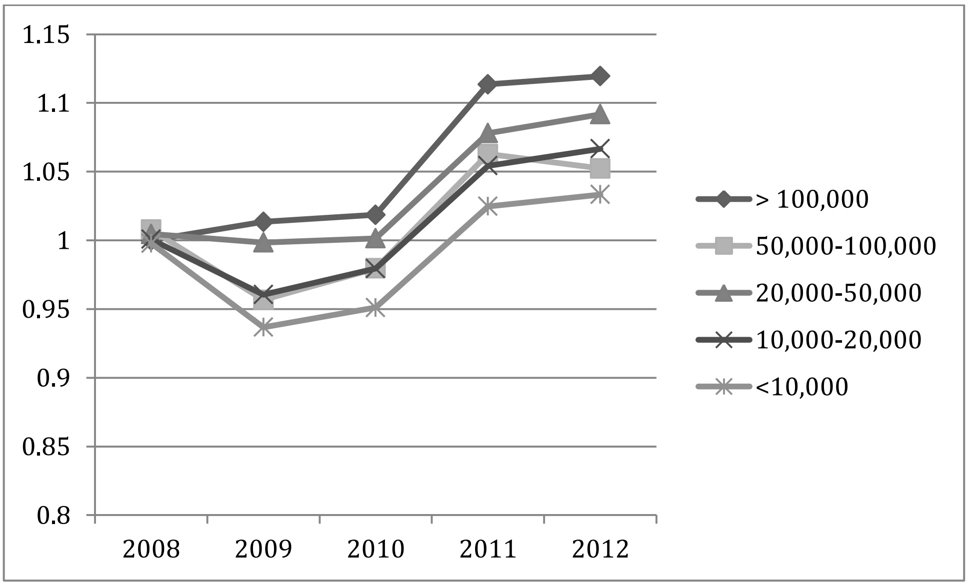


Figure 9 shows how the highest increases in costs of living between the initial and the final year have happened in the largest municipalities (more than 11%), being this growth slightly lower rises in less populated locations (no more than 3.5% in the municipalities of less than 10,000 inhabitants). The estimates also indicate how the last year of the period studied, which has been when the economic crisis had its most severe effects on the Spanish economy, is when the increases in the cost of living seem to be more relevant: the annual changes from 2010 to 2011 range between 7% and 10% approximately, being these annual changes remarkably smaller in previous years, especially at the beginning of the period under study where it can be found decreases around 6% in the smallest cities.

One important issue at this point is quantifying to what extent our estimates of cost of living, which have been derived by applying the proposed AIDS-based methodology, differ from those obtained by using a traditional approach with a fixed basket of commodities. In order to give an answer to this question, we will compare our estimates with a Laspeyres index that sets as the reference basket the average basket consumed in the municipalities with population larger than 100,000 in 2008. More specifically, the results of applying the index depicted in (28) will be confronted with the results of applying the following index:

$$\mathrm{L}_{hr} = \frac{\sum_j \bar{p}_{jh}\, w_{jr}}{\sum_j \bar{p}_{jr}\, w_{jr}}; j = 1, \dots, n \qquad (34)$$

Being $\overline{p_h}$ and $\overline{p_r}$ the vectors with the median prices paid by the households in areas h and r for the $j = 1, .. n$ food products, respectively; and w_r the vector that contains the average budget shares (weights) in the reference area (in our case, municipalities with population larger than 100,000 in 2008). Table 19 reports the results obtained by the Laspeyres price index in (34).

Table 19 Laspeyres index by type of municipality (2008-2012). Base: cities with population >100,000 in 2008

	2008	2009	2010	2011	2012
>100,000	1.0000	0.9764	0.9894	1.0103	0.9986
50,000-100,000	0.9601	0.9196	0.9518	0.9504	0.9455
20,000-50,000	0.9871	0.9500	0.9577	0.9768	0.9733
10,000-20,000	0.9787	0.9305	0.9501	0.9576	0.9654
<10,000	0.9672	0.9184	0.9297	0.9434	0.9330

A comparison of the index numbers reported in Table 18 and Table 19 reveals relevant differences between the two approaches. The Laspeyres index shows only modest differences in the cost of living between different types of municipalities (around 9% at the most). Moreover, the evolution of this index along time follows a very stable behavior, in contrast with the striking growths estimated by applying the proposed COLI at constant utilities. This comparison is the consequence of computing a fixed-basket index, which compares the cost of acquiring the same bunch of products in two different locations and two periods of time, only considering variations in the prices of these goods but neglecting the changes in living standards derived from keeping the same consumption patterns. In other words, the estimated results in the proposed COLI are growing along time because keeping the same utility level as in 2008 is estimated to be higher in the following years.

Cities are more expensive to live in but what we will try to do in the next section is to explore in a deeper way the agglomeration effect to discover whether is an effect provoked by the particular people who are agglomerated in cities what make these more expensive, or if the cities have intrinsic characteristics to be more expensive. By controlling for the characteristics of the households and the individuals we explore the effect of the agglomeration and other geographic effects over the cost of living of the households.

4.4 Determinants of the Cost of Living

4.4.1 A model of Cost of Living

In order to examine the determinants of the cost of living (COL) variation in Spain we postulated a model of COL data at a household level on a set of geographic and socioeconomic regressors. This is the first contribution of the study; our own data on COL will be estimated following a microeconomic approach and at a household level. The advantage of working at a micro level *versus* at an aggregate level, like all the previous works cited in this chapter, is that the more disaggregated COL allow us to isolate the model of the factors inherent to the households and to the individuals and then explore the pure effect of the geographic variables. The second contribution of the study is that this is the first time that quantile regressions are used for this purpose. This method not only allow us to know how the determinants included in the model influence the COL, but for whom these determinants influence more.

Basic economic theory could be used to find the determining factors of the cost of living variations. As Kurre (2003) explain, the fundamental idea is that factors that increase the demand of goods cause prices to be higher; those which tend to increase supply cause prices to be lower. Additionally, there exist idiosyncratic factors of a region which can influence the cost of living, for example the climate conditions or the situation in the country, as is the case of the Canary and Balearic Islands.

Based on this, the key variables examined are: income per capita in the Autonomous Community; a dummy variable which represents if the household belongs to a city of more than 100,000 inhabitants; one dummy for each region of the country at NUTS-I level; and a set of variables representing various characteristics of the household, like the size of the household, the number of employed, the number of dependents; and of the household head, like the age, the income level and the level of education. The variables can be simplified as:

$$COL_i = f(Agglomeration, X, Z) \tag{35}$$

Where COL_i is the Cost of Living in Euros of each household; X is a set of geographic and regional variables relating to each region at which the households belong to; and, Z is a set of households' and individuals' characteristics variables.

Previous analyses in the empirical literature have demonstrated the strong relation between income and costs of living. The low income areas have the lowest COL and the high income areas have the highest one, in general, the richer the area, the higher the demand for goods, so the higher the pressure on prices. This relationship is found strongly remarkably in works such as Hogan and Rex (1984), McMahon (1991), Kurre (2003) and Kosfeld *et al*. (2008).

Is not immediately clear the effect of the population over the cost of living, the magnitude of the city's population could affect the cost of living in at least three magnitudes (Haworth and Rasmussen, 1973): 1) economies or diseconomies of scale in the provision of public services; 2) externalities affecting the compensation of those employed in the city; and 3) the cost of land. On the one hand, if there is more population the demand of the goods rise and, consequently, the price of the goods rise too. But, on the other hand, large population can produce economies of scale in the production process and lead to lower prices. Cebula (1980 and 1989) finds that the second factor predominate over the first one, so the more the population, the lower the cost of living. In contrast, other authors like Blien *et al*.

(2009) find that larger cities are more expensive to live in. In the model proposed here is introduced a dummy variable which represents the agglomeration effect, this dummy variable is that of Municipalities of more than 100,000 inhabitants. The reason for choosing this is because in the previous section it could be seen that the most striking differences took place between this municipalities and the rest ones.

The influence of geographic variations over the cost of living is also well documented. In Hogan (1984) is revised some empirical works in this issue, for example, Shefer (1970) and Sherwood (1975) evidence highest cost of living in the North East and lower in the South; and Haworth and Rasmussen (1973) found lower living cost in the South. Gradually, more evidences have emerged; McMahon and Melton (1978) and McMahon (1991) concluded that the Southern US benefits from lower costs of living compared to the Eastern Seaboard and the Northeast. In Europe, Hayes (2005) found a great impact of regional price variations in the South East Region of the UK; Kosfeld *et al*. (2008) find strong evidence for the presence of spatial price effects using Consumer Price Index for the Bavarian districts. In this work we also hope to find remarkably differences between the regions included in the model, this regions are included in form of a dummy variable, one for each region (NUTS-I) that is Northwest, Northeast, Region of Madrid, Central Region, East Region, South Region and Canary Islands.

The rest of the variables which compose the vector Z in equation (35) are include as control variables to try to isolate the pure effect of the size of the city over the COL. These variables are expected to have the effects that predict the consumer theory.

4.4.2 Data and Estimation

This study uses detailed information on household expenditure in 2012 for ten expenditure food items. Indeed, we use the information of the HBS for obtaining the geographic and household characteristics variables included in equation (35). The income per capita data of the region is obtained from the Regional National Accounts of the INE. The dataset is formed by 21,484 observations which are disaggregated across the 17 regions at the NUTS-II level.

The dependent variable of the regression (35) is the Cost of Living (COL) at the individual level provided by our own estimations. The COL data are obtained through an AIDS model, as always, for products which are assigned to ten sub-groups belonging to the category of "Food and non-alcoholic beverages" in the HBS classification, namely: (1) Bread and cereals, (2) Meat, (3) Fish, (4) Milk, cheese and eggs, (5) Oil, (6) Fruits, (7) Vegetables, (8) Sugar, (9) Coffee, tea and cacao; and (10) Mineral water and other soft drinks. For each group $i = 1, \dots ,10$ the observed budget share w_i of equation (11) in each household is calculated by dividing the expenditure of the household in this specific group by the total household expenditure in food.

In this case the AIDS model to be estimated is of the form:

$$w_i = \Phi(x)\left[\alpha_i + \sum_{j=1}^{n} \gamma_{ij}\, log\, p_j + \beta_i\, log\{x/P\}\right] + \sum_{k} c_k\, CS_k + r_h R_h + \delta_i \phi(x) \qquad (36)$$

where δ_i is a parameter associated with the density function, CS_k are dummy variables for different urban sizes and R_h is a regional dummy for each one of the NUTS-II regions of Spain, and c_k and r_h are the parameters associated with each type of dummy, respectively, with the aim of recover the idiosyncratic components inherent to each region and type of city. The estimation of the parameters is made by applying Nonlinear Seemingly Unrelated Regression (NLSUR), which estimates a system of nonlinear equations by Feasible Generalized Nonlinear Least Squares (FGNLS). With the parameters estimated (Appendix 5 and Appendix 6) we recover the expenditure functions for each household defined as in Equation (8):

$$log\, c(p,u) = \alpha_0 + \textstyle\sum_{i=1}^{n} \alpha_i\, log\, p_i + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \gamma_{ij}\, log\, p_i\, log\, p_j + u\beta_0 \prod_i {p_i}^{\beta_i} \qquad 38$$

The $log\, c(p,u)$ represents the COL for each household in Euros needed to attain the median utility level of the country as a whole. More precisely, this COL is calculated with the prices faces by each household, with the expenditure level of each household applying the median utility level of the country. In the next table are summarized the main statistics of the estimated COL.

Table 20 Summary statistics of the estimated individual Cost of Living

Percentiles	
1%	2544.631
5%	2855.942
10%	3002.871
25%	3229.804
50%	3461.429
75%	3699.711
90%	3931.433
95%	4092.787
99%	4461.623
Number of Observations	21,484
Std. Dev.	382.1568
Variance	146043.9

Table 20 shows the value of the COL in Euros of the percentiles 1 to 99. Indeed, the Number of Observations indicates the number of households which compose the sample. Also, the Standard Deviation and the Variance are reported in the Table 20.

Once the COL for the 21,484 households is calculated it is proceeded to estimate the full regression (35) for quantiles 1-99:

$$Q_\theta[COL|Agglomeration, X, Z] = \beta_\theta^0 + Agglomeration\beta_\theta^1 + X\beta_\theta^2 + Z\beta_\theta^3 \qquad (38)$$

Where COL is the log of the Cost of Living in Euros of each household, $Q_\theta[COL|X, Z]$ is the θth conditional quantile of COL, β_θ^0 is the regression intercept, $Agglomeration$ represents the cities of more than 100,000 inhabitants, X and Z are covariates matrix which include all geographic and household regressors, respectively, and, the coefficients β_θ represent the returns to covariates at the θth quantile.

The process yields a sample of 21,484 observations. The intercept X recovers the Income per capita of the Autonomous Community of the household, a dummy variable that represents the agglomeration effect and is measured as the municipalities with more than 100,000 inhabitants, and the set of dummies of Spanish regions; the intercept Z recovers the Household Size measured as the number of members of the household, the Number of Employees in the household measured in number of people employed, the Age of the household head measured as a continuous variable that represents the number of years old, the Number of Dependents in the household, Education Level of the household head which is divided into four categories: no studies, first cycle studies, second cycle studies and high degree studies, and the income level of the household which is divided into seven categories which range from less than 500 net Euros per month to more than 3,000 net Euros per month.

With the described model it is estimated a quantile regression model (Koenker and Basset, 1978) which fits quantiles to a linear function of covariates. In its simplest form, the least absolute deviation estimator fits medians to a linear function of covariates. The method of quantile regression is more attractive because medians and quartiles are less sensitive to outliers than means, and therefore Ordinary Least Squares (OLS). Indeed, the likelihood estimator is more efficient than the OLS one. Quantile regressions allow that different solutions at different quantiles may be interpreted as differences in the response of the dependent variable to changes in the regressors, thus, quantile regressions detect asymmetries in the data which cannot be detected by OLS. But the most important feature is that quantile regression analyzes the similarity or dissimilarity of regression coefficients at different points of the dependent variable, which in this case is the household COL; it allows one to take into account the possible heterogeneity across COL levels. The model is estimated in using the least-absolute value minimization technique and bootstrap estimates of the asymptotic variances of the quantile coefficients are calculated with 20 repetitions.

4.4.3 Results

Before starting with the estimation model described above it will be reported the results of the quantile regression taking as dependent variable the data on household expenditure reported by the HBS. These estimations will give us a first view of how the geographical variables behave over the expenditure level. So it is regressed the model described in (38) but replacing the COL variable described in this equation by the expenditure level of the HBS of each household. The quantile regression estimates are reported in Table 21 for comparison purposes we also report OLS estimates.

The dependent variable is the logarithm of the Expenditure Level in the entire food group in each household. As the results show, no variable is significant (except a few) nor in the OLS estimation or in the quantile estimation. This means that the official available data does not reflect the “true” cost of living due to these expenditures represent different utility levels. From the official data it cannot be inferred any postulate of the Regional and Urban Economics explained above.

After these disappointing results it is proceeded to estimate the model defined in (38) taking as dependent variable the COL estimates for each household which warrants that represent the same standard of living for all the household of the survey. The results are shown in Table 22.

Table 21 Estimates of the OLS and Quantile Regression over the Expenditure Level provided by the HBS

	OLS		QUANTILE REGRESSION									
			10		25		50		75		90	
Expenditure HBS	Coef.	t	Coef.	t	Coef.	t	Coef.	t	Coef.	t	Coef.	t
Agglomeration	-0.0011	-0.10	0.0017	0.06	-0.0052	-0.32	-0.0064	-0.48	-0.0056	-0.63	-0.0024	-0.22
Income	0.0176	0.32	0.0020	0.02	-0.1008	-1.45	0.0199	0.28	0.0493	1.2	0.1196	1.67
Northwest	-0.0376	-1.21	-0.0280	-0.42	-0.0587	-1.27	-0.0502*	-1.53	-0.0185	-0.61	-0.0253	-1.02
Northeast	-0.0505*	-2.15	-0.0839	-1.26	-0.0400	-1.21	-0.0571	-2.07	-0.0235	-0.92	-0.0572***	-3.46
Central	-0.0139	-0.40	-0.0628	-0.81	-0.0566	-1.14	-0.0237	-0.64	0.0061	0.17	0.0293	0.77
East	-0.0439*	-1.65	0.0105	0.15	-0.0441	-1.21	-0.0772**	-2.76	-0.0308	-1.06	-0.0299	-1.14
South	-0.0179	-0.47	-0.0375	-0.46	-0.0621	-1.35	-0.0347	-0.86	0.0113	0.42	0.0272	0.85
Canary Islands	-0.0100	-0.25	-0.0383	-0.47	-0.0264	-0.59	-0.0150	-0.46	0.0186	0.58	0.0078	0.2
Household Size	0.0126	1.65	0.0420	1.71	0.0138	1.01	0.0123	1.27	0.0019	0.26	-0.0078	-0.76
Number of employed	-0.0126	-1.39	-0.0374*	-1.78	-0.0122	-0.99	-0.0112	-1.45	0.0050	0.93	0.0053	0.55
Age	-0.0002	-0.51	-0.0004	-0.41	0.0003	0.61	-0.0001	-0.3	-0.0002	-0.61	-0.0004	-0.79
Number of dependents	-0.0001	-0.01	-0.0395	-1.35	-0.0004	-0.02	-0.0039	-0.34	0.0046	0.48	0.0073	0.43
First cycle studies	0.0134	0.81	0.0383	0.88	-0.0103	-0.45	0.0133	0.84	0.0057	0.37	-0.0085*	-0.51
Second cycle studies	-0.0064	-0.32	0.0183	0.28	0.0059	0.22	-0.0133	-0.59	-0.0130	-0.73	-0.0459	-1.91
High degree studies	0.0156	0.78	0.0478	1.1	-0.0099	-0.37	0.0198	0.83	0.0079	0.47	-0.0038	-0.18
500-1000 Euros	-0.0038	-0.13	-0.0758	-1.13	0.0041	0.12	0.0133	0.44	-0.0296	-1.23	-0.0113	-0.29
1000-1500 Euros	-0.0100	-0.35	-0.0897*	-2.02	-0.0004	-0.01	0.0052	0.2	-0.0431*	-1.78	0.0014	0.04
1500-2000 Euros	-0.0164	-0.55	-0.0827	-1.38	0.0048	0.13	0.0132	0.47	-0.0630***	-3.05	-0.0187	-0.52
2000-2500 Euros	0.0140	0.44	-0.0386	-0.68	0.0208	0.58	0.0354	1.34	-0.0206	-0.81	0.0145	0.31
2500-3000 Euros	-0.0537	-1.60	-0.1348*	-2.15	-0.0515	-1.39	-0.0375	-1.13	-0.0747**	-2.62	-0.0293	-0.73
More than 3000 Euros	-0.0188	-0.55	-0.0812	-1.41	0.0030	0.07	-0.0045	-0.19	-0.0497*	-1.9	-0.0195	-0.57
_cons	8.0740***	167.79	7.2006***	62.31	7.6942***	150.3	8.1733***	166.94	8.5820***	182.85	8.9062***	145.08

Note: *, ** and *** represent the level of significance to 10%, 5% and 1%, respectively.

Table 22 Estimates of the OLS and Quantile Regression with the COL estimated at household level

	OLS		QUANTILE REGRESSION									
			10		25		50		75		90	
COL	Coef.	t	Coef.	t	Coef.	t	Coef.	t	Coef.	t	Coef.	t
Agglomeration	0.0096***	2.72	-0.0014	-0.25	0.0035	0.91	0.0103**	2.39	0.0159***	3.53	0.0211***	3.61
Income	0.2013***	12.03	0.1995***	7.35	0.1758***	7.66	0.2065***	10.69	0.2231***	11.55	0.2205***	13.03
Northwest	-0.0519**	-5.54	-0.0115	-0.65	-0.0362**	-2.45	-0.0530***	-4.38	-0.0614***	-5.49	-0.0685***	-5.05
Northeast	0.0154***	2.16	0.0465***	2.99	0.0385***	3.12	0.0188**	2.16	0.0003	0.03	-0.0116	-0.72
Central	-0.0580***	-5.53	-0.0528*	-2.85	-0.0556***	-3.47	-0.0479***	-3.56	-0.0490***	-3.3	-0.0532*	-3.3
East	0.0709***	8.8	0.0845***	5.63	0.0870***	6.45	0.0778***	7.19	0.0697***	5.77	0.0530**	4
South	0.0795***	6.96	0.0910***	4.71	0.0807***	5.4	0.0825***	6.26	0.0785***	5.29	0.0686***	4.18
Canary Islands	0.1409***	11.78	0.1846***	8.79	0.1672***	11.91	0.1444***	10.06	0.1247***	8.02	0.0893***	4.25
Household Size	-0.0162***	-7.33	-0.0184***	-7.59	-0.0245***	-13.79	-0.0242***	-10.44	-0.0187***	-6.97	-0.0109**	-2.17
Number of employed	0.0132***	4.82	0.0126**	2.66	0.0105*	2.46	0.0138***	3.35	0.0110***	4.05	0.0150**	3.7
Age	0.0004***	2.86	0.0003	1.2	0.0005*	2.49	0.0006***	3.06	0.0005*	2.24	0.0004*	2.03
Number of dependents	0.0100***	3.63	0.0173***	4.65	0.0167***	6.35	0.0155***	4.43	0.0099***	3.33	0.0019	0.31
First cycle studies	0.0138***	2.75	0.0194**	2.21	0.0232***	3.77	0.0197***	5.29	0.0122*	1.89	0.0045	0.55
Second cycle studies	0.0400***	6.57	0.0433**	4.41	0.0522***	7.93	0.0461***	6.24	0.0448***	6.04	0.0409***	4.38
High degree studies	0.0557***	9.21	0.0486***	5.23	0.0606***	11.61	0.0620***	8.46	0.0593***	7.14	0.0545***	6.87
500-1000 Euros	0.0534***	6.02	0.0765***	5.9	0.0686***	5.35	0.0441***	4.42	0.0304***	2.92	0.0400*	2.95
1000-1500 Euros	0.0788***	8.94	0.1172***	14.75	0.1118***	11.57	0.0724***	8.45	0.0495***	5.14	0.0519*	3.35
1500-2000 Euros	0.1108***	12.09	0.1526***	15.25	0.1461***	13.25	0.1042***	10.05	0.0731***	7.65	0.0724***	6.61
2000-2500 Euros	0.1339***	13.87	0.1874***	19.23	0.1793***	17.18	0.1305***	13.88	0.0907***	8.67	0.0839***	4.46
2500-3000 Euros	0.1614***	15.8	0.2190***	16.45	0.2131***	15	0.1596***	13.88	0.1158***	8.96	0.1018***	6.32
More than 3000 Euros	0.1824***	17.6	0.2504***	22.09	0.2410***	19.77	0.1848***	15.28	0.1377***	11.24	0.1228***	6.38
_cons	5.9771***	34.77	5.6481***	20.53	6.0452***	26.8	5.9324***	30.96	5.9494***	29.82	6.1096***	34.78

Note: *, ** and *** represent the level of significance to 10%, 5% and 1%, respectively.

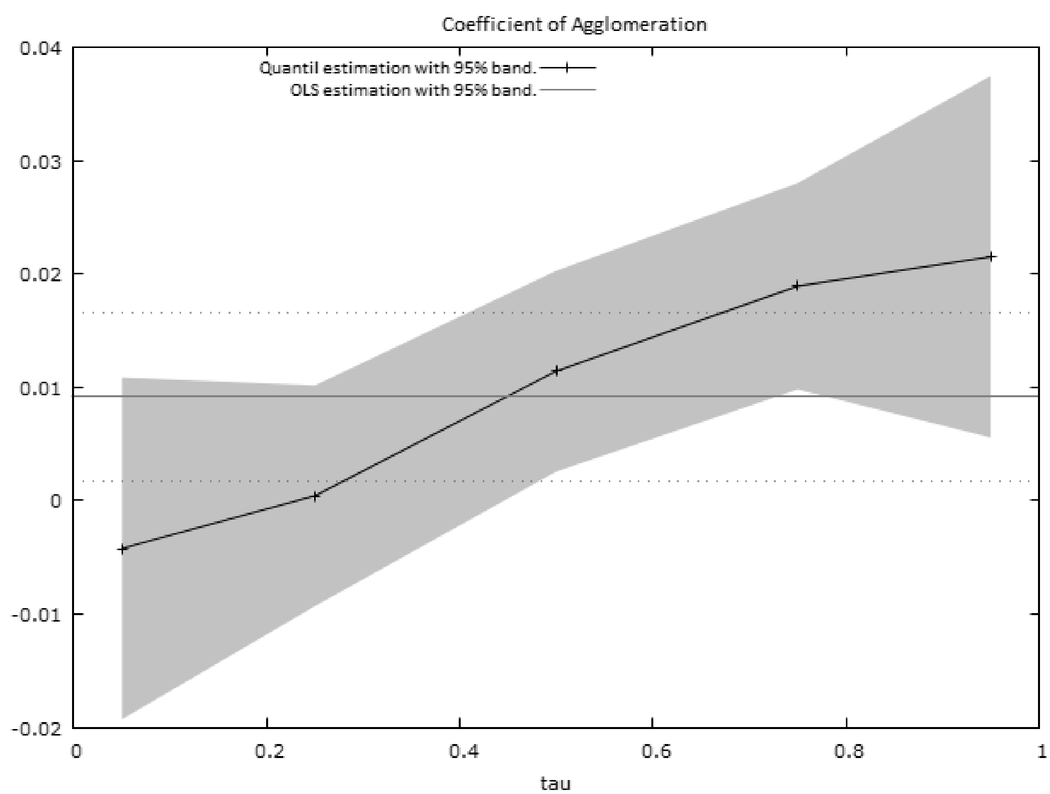
Table 22 gives us the results of the OLS and quantile regression estimations of the COL as a function of the regional and socioeconomic variables described above. The first column of Table 20 gives the results of the OLS regression, the successive columns gives the results of the 10, 25, 50, 75 and 90 quantiles, respectively.

The OLS estimates of Table 20 show that all variables are significant at 1% level except a few. The income per capita of the Autonomous Community and the agglomeration variable are both positive and statistically significant. Also, the regional NUTS-I variables are significant. All this tell that there is a significant effect of the geography and demography over the cost of living of the households, contrary with the results obtained in the Table 19 which no variables were significant. The socioeconomic variables are significant and have the expected effect over the COL. The household size, number of employed, the age and number of dependents are continuous variables. The level of education is represented with a set of dummy variables that indicate the effect of each degree of studies respect to individuals which have not studies or have basic studies. Respect to the income level the results are reported respect to the households which have less than 500 Euros of net monthly income.

In the rest of the columns of Table 20 are shown the estimates of the quantile regression. The coefficients of the set of socioeconomic variables are of the expected sign and most of them are statistically significant at the 1% level. Our focus is on the results of the geographical and demographic variables. The income variable represents the income per capita of the Autonomous Community at which the household belongs to. This variable is one of the most statistically significant showing a positive relationship between the income per capita of the Autonomous Community of residence and the COL of the household. Thus, the strong theoretical response of prices in income is supported by the data.

The agglomeration variable is represented by the municipalities of more than 100,000 inhabitants. The variable is statistically significant and positive in the upper budget level that is in 50, 75 and 90 percentiles, this means that the COL is higher in the biggest cities only for the rich. This result has sense because there are some kinds of goods which are only available in the biggest cities and are only consumed by high income households. Consequently, the biggest cities have a greater demand of the goods with income elastic demands which are only demanded by rich households and this cause an upward pressure on prices. In contrast, the price of inferior goods which composed the basket of the poor, are not affected as much as the price of superior goods. In other words, the poor will never consume superior goods and their basket of goods costs similarly in all city sizes. It can be seen graphically the evolution of the coefficient of the agglomeration variable in Figure 10.

Figure 10 Evolution of the Agglomeration coefficient along the quantile distribution



Regional dummy variables are represented at the level of NUTS-I. The omitted region is the Autonomous Community of Madrid, so the results are interpreted respect to this region. As we can see all regional dummies are statistically significant, the Northwest and Central dummies are negative and statistically significant; this means that living in those regions is cheaper than in the Autonomous Community of Madrid. The rest of the dummies are positive and statistically significant meaning that the COL in these regions is higher than in Autonomous Community of Madrid.

These results are in line with the expectations. The Northwest and Central regions include Autonomous Communities all of them with lower COL than Madrid, these Autonomous Communities are Galicia, Asturias and Cantabria in the Northwest and Extremadura, Castile Leon and Castile La Mancha in the Central region. In contrast, the rest of the regions have higher COL than Madrid, this can be explained by the fact that the Northeast region is formed by some of the richest Autonomous Communities that is Navarra and Basque Country. In the same way the East region is influenced by Catalonia which has a COL in 2012 7.07% higher than Madrid; the South region includes Autonomous Communities very touristic like Murcia and the Mediterranean side of Andalusia which make arise the COL respect to Madrid. Lastly, the particular position of the Canary Islands makes that the COL is remarkably higher than in Madrid mainly due to transportation costs.

Overall, geography matters a lot in determining the COL and, as expected, the COL is stronger linked with income. The novelty of the study is the empirical evidence for Spain of the effect of agglomeration over the COL using for the first time a quantile regression over a “true” Cost of Living calculated at a household level.

4.5 Conclusions

Prices and consumption patterns change across the space. There are geographical, weather, cultural, sociological and economic reasons to offer as explanations for the fact that the level of prices and the way of consume differ from one region to another. Particularly relevant are the potential effects of the size of the cities. Large cities are more competitive, offer a greater variety of goods and services and, among other factors, develop a different style of life… As a result, the response of consumers to changes in prices should be different in a small town in contrast to a large metropolis.

Although there is ample evidence of how consumption patterns are affected by factors such as the level of income or stage in their life cycle at which households find themselves, the empirical studies on spatial effects are limited and contradictory. Several studies have found significant differences in consumption patterns of households living in rural areas compared to those residing in urban areas. However, most of these studies refer to developing countries that have not completed the process of urbanization and where the realities of urban and rural life are clearly poles apart. There is little empirical evidence on similar differences in developed countries.

The first aim of this chapter was to estimate expenditure and own-price elasticities in Spain by income level and by city size. Spain is particularly suitable for a study of this type as it is characterized by an advanced level of urbanization and development. It possesses a very rich urban structure with several large cities, a large network of medium-sized towns and a rural setting that is still important. Furthermore, differences in earnings have worsened since the onset of the economic crisis and so the breach between high- and low-income households has become wider: the Gini index in Spain increase 3 points from 2008 to 2012[9].

The exploration in this section focused on variations in expenditure elasticities for 2010; data limitations restricted attention only to foodstuffs. The AIDS estimates were made to explore more precise comparisons across space. Although the study would be greatly enhanced if it could be extended to other goods and services, the fact that it only refers to foodstuffs goods is still valuable as these comprise one of the major items of household expenditure.

The results confirm that differences in income between households clearly affect their consumption patterns. As expected, higher-income households react to changes in income with moderate changes in the quantities consumed compared to lower-income households. Similarly, low-income households are much more sensitive to price changes than high-income households. An important finding is that the intensity of the effect of the size of the city of residence is of a similar order in magnitude to that

[9] http://epp.eurostat.ec.europa.eu/tgm/table.do?tab=table&language=en&pcode=tessi190

of the income level. This means that different consumption patterns exist depending on whether the household resides in a large city, a small or medium-sized town or a rural area in the case of a country like Spain. The changes affect own-price elasticities more than expenditure elasticities. Although the latter vary depending on different city sizes, there are exceptions with respect to types of products. However, own-price elasticities are always higher in larger cities. Price sensitivity is higher in large cities compared to smaller towns and in rural areas.

If consumption patterns change from one place to another, as was found in our analysis for the Spanish food demand, the conclusion derived is that policies that affect prices could produce different effects across space. Regional policies oriented to impulse the convergence among territories, urban planning, poverty policies, or programs designed to promote economic growth, productivity or competition should take into account how the consumption patterns and the cost of living change among cities and, in particular, how relevant the effect of the city size might be. Previous research in Urban and Regional Economics has pointed out the existence of substantial differences in costs of living among different sizes of cities, and, also a systematic relationship between the cost of living and the city size has been identified. Most of these studies have been applied for the US, but the number of contributions that analyze this city size effect in Europe is smaller due to data availability and the conclusions less clear. This lack of empirical studies is especially important for the case of Spain, where there is not any quantification of the effect of city size on the cost of living.

The key question asked in this chapter is whether the COL differs between city sizes. The answer is yes and it has been demonstrated through several ways. The first way was the estimation of the COL by municipality size along the period 2008-2012. The results showed that the smallest areas have lower COL consistent with the theoretical and empirical literature revised in previous sections. The difference between the smallest municipalities and the biggest ones is more than 8% in 2012. The second way corroborates the previous results through an alternative approach. In this approach a quantile regression model was used to determine the factors that influence the COL.

For this purpose a COL at a microlevel for each household of the HBS has been calculated to regress it over a set of socioeconomic variables and demographic and geographic variables. Among these variables it has been used the cities of more than 100,000 inhabitants to represent the effects of agglomeration over the COL. Through the estimation of a quantile regression it is found that the agglomerations raise the COL but only for the high income quartiles, this result is rational due to the kinds of goods that offers the biggest cities and are only consumed by the rich.

Developing and applying cost of living indicators that allow for spatial comparisons have important policy and welfare implications. Disparities on the average income between large cities and rural or small cities areas (urban premium) could be not as large as they seem if income is adjusted by cost of living differences. Another important implication of not having a proper index of cost of living is the possibility of obtaining misleading results in poverty analysis. A failure to account properly for cost of living differences between urban and rural or small cities areas may lead to regionally inconsistent poverty lines and may result in unwarranted policy interventions. Nominal poverty thresholds that are invariant across space result in an overestimation of the poverty in less urbanized areas compared with urban areas, affecting considerably the eligibility for benefits.

5 SUMMARY AND FINAL REMARKS

Previous urban and regional research has pointed out the existence of substantial interregional differences in costs of living across the space. These studies have also identified systematic relationship between certain socio-economic-demographic variables and living costs in an area. Even so, these kinds of studies still remain scarce due to the lack of available data. No Spanish study, as far as I know, have evaluated these tendencies and the consequences for overall measures of inequality. For this reason, this thesis examines the regional consumer price information in some detail in an attempt to identify the systematic patterns of the costs of living across Spain as well as its welfare implications.

The most widely used cost of living measure is the Consumer Price Index (CPI) compiled both at the national and regional level by the National Statistical Institute but this measure does not cease to be a simple measure of the level and the rate of inflation. Thus, the CPI is oriented towards the time dimension and it cannot be used for cross-regional comparisons. Therefore, for those interested in making cost of living comparisons at spatial scale alternative sources of data must be found. Another limitation of the use of the CPI as a cost of living index (COLI) is derived by the own way in which it is calculated. Although the CPI baskets of goods have been carefully designed to represent the consumption patterns of the average, these baskets do not provided the same level of standard of living when they are compared along the time. Following Konüs (1939) a "true" index of the cost of living is obtained when in the course of two periods of time the standard of living remains constant.

In this thesis are addressed both of these issues. On the one hand, a "true" COLI has been calculated following the Konüs' approach of fixed utility, and on the other hand, adapting this approach to the spatial case to make comparisons not only along time but also across the space. For both purposes the methodology of Deaton and Muellbauer (1980) called Almost Ideal Demand System (AIDS) is followed; and the data provided by the Household Budget Survey of the National Statistical Institute (INE) are used.

Firstly, striking differences between the national CPI and our true COLI, calculated for the whole country, have been found. The CPI provokes an underestimation of the rate of inflation along the period analyzed being the increment in the CPI between 2008 and 2012 of a 2.4% versus a 10% of increment in the COL. The following sub-sections have focused in the spatial analysis of both the CPI and the COL. A systematic pattern of living costs differentials were found across the 17 Autonomous Communities, with the overall costs of living highest in Catalonia, Navarra and Basque Country together with Murcia and the Canary Islands and lowest in Andalusia, Castile La Mancha and Extremadura. The magnitude of this difference is around 25% between the most expensive region (Catalonia) and the cheapest one (Castile La Mancha) in 2012. This pattern remains constant along all the period 2008-2012. Moreover, the evolution of both groups of regions is opposite. While the high cost of living regions present an upward trend in the COLI since 2008, the group of the low cost of living regions shows a downward trend in the COLI in 2008 with a smooth increase until 2012 and always below the formers. These results support those found by Alberola and Marqués (2001) and Garrido-Yserte *et al*. (2012) who using alternative price indices evidence substantial and permanent differences among Spanish regions. All these conclusions cannot be inferred from the official CPI. The representation of the CPI for all the Autonomous Communities shows that the overall trend of prices is the same among regions; there are no differences neither in the level of the CPI between Autonomous Communities or in the evolution of this CPI along the period 2008-2012.

A positive relationship have been identified between the income level and the costs of living that is the high income regions support higher prices contrary to the low income regions which benefit from the lowest costs of living. This proposal of the Regional Economics is supported by the estimates made in this thesis and also matches with the results of Kosfeld *et al*. (2007) for Germany and with the results of Suedekum (2006) for the US. It must be said that this analysis of the cost of living is somewhat biased because it is measured the average household of the Autonomous Communities which comprise big cities but many rural areas as well. For this reason price information at different spatial scales is crucial to an accurate assessment of the cost of living in which we were being able to calculate the cost of living at the province and municipality level. This is important because from the previous results it can be inferred that the highest costs of living are found in the richest regions which at the same time are the regions which comprise the biggest and more touristic cities where, in general, are agglomerated the highest incomes and highest wages. So not only is important how the cost of living varies between Autonomous Communities but also how this cost of living varies within the Autonomous Communities, specifically, how different are the costs of living in the urban areas respect to the rural ones.

As we cannot analyzed the differences within the Autonomous Communities due to the lack of data availability, we will comply with the possibility of analyzing a very important topic in the Regional and Urban Economics that is the effect of the city size over the prices and the costs of living. In the section 4.3 it has been made this analysis by explicitly calculated the "true" COL for the five different municipality sizes classified by the HBS. The results provided empirical evidence about the existence of differences in COL among different sizes on the cities. More precisely, the smallest cities benefit from the lowest COL while the biggest cities of more than 100,000 inhabitants support the highest COL. The reported results are consistent with both theoretical works like Helpman (1998), Tabuchi (2001), Tabuchi and Thisse (2003) Cavaihès *et al*. (2004) and Suedekum (2006), and empirical works like Alonso (1970), Haworth and Rasmussen (1973), Simon and Love (1990), Walden (1998), Kurre (2003) and Cebula and Todd (2004). The differences in the COLI between the biggest and the smallest municipalities reached the 9% in 2012, and if we attend to the period 2008-2012 it could be seen that the increase in the COLI in the whole period was higher in the biggest cities (11%) *versus* the smallest areas (2.7%).

It would not be unreasonable that if the cost of living varies across the space not only this variation was due to regional price differences but also due to different consumption patterns which are affected by the place of residence and even by the size of the city. For this purpose the demand elasticities by city size have been estimated for the 10 food categories resulting that both own-price and income elasticities changes remarkably between the city sizes being the expenditure elasticities lower in the small areas of less than 10,000 inhabitants and the own-price elasticities more sensitive the higher the size of the city. It is already proven that in larger cities exist different consumption patterns respect to the smallest areas, also than in these larger cities prices and costs of living are higher. But, what are the reasons for this to occur? On the one hand, larger cities agglomerate a set of consumers which are the youngest, the most qualified and, consequently, with the highest incomes and wages. This could be provoke that in agglomerations were generated a particular consumption patterns due to the possibilities offered by its agglomerations. Indeed, this kind of consumers raises de demand of certain goods provoking a pressure over prices making these prices to rise. But, on the other hand, it could be occur that a certain consumer with certain characteristics experience a different consumption patterns and support higher prices by the fact that agglomerations generate *per se* these different consumption patterns and these higher prices.

These postulates were investigated through a deeper analysis on what factors determine the cost of living in a household (COL). By calculating a micro-COL at a household level (not for a representative one) for all the households that comprise the HBS, it have been regressed trough a quantile regression this micro-COL over a set of individual and household characteristics plus a set of demographic and geographic variables. The aim of this exercise is to isolate the pure effect of the agglomeration and the regions to corroborate if, effectively, there is a spatial and city size effect in the determination of the COL. The robustness of the results confirms that the biggest cities support higher COL and that regions (NUTS-I) have statistically significant differences in the COL between them.

All these findings are considered of great importance because the differences in price levels across regions and between urban and rural areas matter for economic outcomes such as inequality and should be taken into account in all comparisons that involves measures of income. The rationale for this thesis is to explore the degree to which spatial price differences impacted upon the economic welfare of Spanish households.

As we noted earlier, income and price are positively correlated which may bias any calculation of income inequality. To illustrate this fact, it has been examined the role of the cost of living in the wage differentials across Autonomous Communities and it could be seen that adjusting for price differences has a large effect on regional wage differentials. The workers in the low cost of living regions like Extremadura, Andalusia and Castile La Mancha realize a substantial wage advantage when adjusting for cost of living differences and suggest a rough equality in wages between Autonomous Communities. The dispersion in approximate real wages (that is adjusted by SCOLI) is lower than the dispersion in nominal wages.

Another important implication in welfare, specifically in poverty, is how governments measure poverty, generally without accounting for geographic differences in the COL. In this thesis can be seen that adjusting the poverty line for the SCOLI under consideration systematically change the geographic distribution of poverty. The comparison is made with the results of Herrero *et al*. (2013) who adjust the poverty line with the official CPI. One important finding is that all regions increase its poverty rate in many cases above the 25%, this result was hopped since the increments in the COL were greater than the increments in the CPI. Another important insight is that the prevalence of poverty has been greater in the richest and most touristic regions, in general, these regions also have the biggest agglomerations, than in the poorest and more rural ones which present lower poverty rates than the official rates. These conclusions are in line with those finding for the US in North and Cook (1995), Slesnick (2002) and in Jolliffe (2006). All demonstrated that not adjusting for geographic differentials in the costs of living overestimates the rural poverty and underestimates the urban poverty, for example, one striking result of the study of Jolliffe (2006) which worth mentioning is that the poverty rate of the nonmetropolitan areas in 2001 with any spatial correction was 28% higher than in metropolitan areas; by contrast, when geographic correction was made the results show that the prevalence of poverty in nonmetropolitan areas was 12% lower than in metropolitan areas. Exactly the same conclusions were reached by Ravallion and Van de Walle (1991) and Asra (1999) in a context of a developing country, in both cases for Indonesia. Therefore, is should not be surprising the results achieved by the Spanish case and it must be highlight the substantial influence that spatial price variations have over the estimates of welfare.

To conclude, the estimation of COL figures manages to have explanatory power in Regional and Urban Economics. The use of these estimations has proved to offer a better understanding of the patterns of regional price differences as well as the implications of agglomeration economies, something that is not provided by the official CPI data. It has been proved several postulates of the Regional and Urban Economics which cannot be inferred from the official available data. Besides being useful for consumers who want to know how prices and costs of living vary across the space, there are many uses of the estimations made in this thesis. These include examining geographic disparity in real per capita incomes, comparing real wages among locations, and the assessment of the picture of poverty in Spain. All of these results point out the importance of having good COL information; perhaps it is time for the official statistics to consider the creation of an official spatial price index.

Future lines of research emerge in a framework where the National Statistical Institute would provide us information of the location of households at a more disaggregate geographical scale. This will allow us, for example, calculate the cost of living at a provincial level, this estimation will provide more realistic knowledge of the cost of living due to many Autonomous Communities comprise several poor provinces together with very rich provinces so assigning the same Autonomous Community cost of living to all provinces or locations is a very strong assumption. With the availability of these data another important insight could be resolved that is the effect of the proximity to a high cost of living areas. How these areas, like Madrid or Barcelona, could transmit higher prices to other medium-sized or small areas situated near these big metropolises. This economic process may give rise to spatial patterns in household demand by two ways. First, prices are expected to be spatially correlated because prices depend upon distance from some point. And second, whether households gain utility in consuming bundles similar to those consumed by their neighbors. Following the seminal work of Case (1991) it would be desirable an estimation scheme that allows for spatial interaction among households and testing the extent to which households look to a reference group when making decisions, the behavior of other households affects a given household behavior through social proximity.

6 REFERENCES

Abdulai, A., Jain, D., & Sharma, A. (1999). Household food demand analysis in India. *Journal of Agricultural Economics , 50*, 316-327.

Agriculture, U. D. (2007). *International Food Consumption Patterns.* Washington, DC: US Department of Agriculture.

Ahmed, A., & Shams, Y. (1994). Demand elasticities in rural Bangladesh: an application of the AIDS model. *The Bangladesh Development Studies , 22* (1), 1-25.

Alberola, E., & Marqués, J. (2001). On the evolution of relative prices and its nature at a regional level: the case of Spain. *Journal of Regional Science , 41* (3), 451-474.

Alfonzo, L., & Peterson, H. (2006). Estimating food demand in Paraguay for household survey data. *Agricultural Economics , 34*, 243-257.

Alonso, W. (1964). *Location and land use, toward a general theory of land rent.* Cambridge, MA: Harvard University Press.

Alonso, W., & Fajans, M. (1970). Cost of living and income by urban size. *Working Paper No. 128 Department of City and Regional Planning. University of California, Berkeley* .

Amemiya, T. (1974). Multivariate Regression and Simultaneous Regression Models when the dependent variables are truncated normal. *Econometrica , 42*, 999-1012.

Angulo, A. M., Gil, J., & Dhehibi, B. (2002). Town size and the consumer behavior of Spanish households: a panel data approach. *Applied Economics , 34*, 503-507.

Armagan, G., & Akbay, C. (2008). An econometric analysis of urban households' animal products consumption in Turkey. *Applied Economics , 40*, 2029-36.

Asra, A. (1999). Urban-Rural differences in cost of living and their impact on poverty measures. *Bulletin of Indonesian Economic Studies , 35* (3), 51-69.

Aten, B., & Menezes, T. (2002). Poverty price levels: an application to Barzilian metropolitan areas. *Paper presented at the World Bank ICP Conference.* Washington DC.

Atuesta, L., & Paredes, D. (2012). A spatial cost of living for Colombia using a microeconomic approach and cesored data. *Applied Economic Letters , 19* (18), 1799-1805.

Ayala, L. (2013). Crisis económica y distribución de la renta: una perspectiva comparada. *Papeles de Economía Española , 135*.

Ayala, L., Jurado, A., & Pedraja, F. (2011). Income poverty and multidimensional deprivation: lessons from ross-regional analysis. *Review of Income and Wealth , 57* (1), 40-60.

Bairoch, P. (1988). *Cities and development: from the drawn of hystori to the present.* Chicago, IL: University of Chicago Press.

Blien, U., Gartner , H., Stüber, H., & Wolf, K. (2009). Regional price levels and the agglomeration wage differential in western Germany. *Ann Reg Sci , 43*, 71-88.

Booth, A. (1993). Counting the poor in Indonesia. *Bulletin of Indonesian Economic Studies , 29* (1), 53-83.

Boskin, M., Dulberger, E., Gordon, R., Griliches, Z., & Jorgenson, D. (1996). Toward a more accurate measure of cost of living. *Final Report to the US Senate Finace Committee .* Washington, DC.

Bouis, H. (1990). Evaluating demand for calories for urban and rural populations in the Philippines: implication for nutrition policy under economic recovery. *World Development , 18*, 281-299.

Breuer, C. C. (2007). Cost of living indexes for Germany. Essen, Germany: University of Duisburg-Essen. Department of Economics.

Burney, N., & Akmal, M. (1991). Food demand in Pakistan: an application of the extended linear expenditure system. *Journal of Agricultural Economics , 42*, 185-195.

Case, A. (1991). Spatial patterns in household demand. *Econometrica , 59* (4), 953-965.

Cavailhès, J., Gaigne, C., & Thisse, J.-F. (2004). Trade cost versus urban cost. *CERP Disscussion Papers* (4400).

Cebula, R. (1980). Determinants of geographic living cost differentials in the United States: An empirical note. *Land Economics , 56* (4).

Cebula, R. (1989). The analysis of geographic living cost differentials: A brief empirical note. *Land Economics , 65* (1), 64-67.

Cebula, R. (1993). The impact of living costs on geographic mobility. *Quartely Review of Economics and Finance , 32* (4), 101-105.

Cebula, R., & Todd, S. (2004). An empirical note on determinants of geographic living - cost differentials for counties in the State of Florida, 2003. *The Review of Regional Studies , 34* (1), 112-119.

Cebula, R., Gigi, M., & James, V. (1992). Afurther note on determinants of geographic cost of living differentials. *Quartely Review of Economics and Finance , 32* (4), 144-147.

Coondoo, D., Majumder, A., & Ray, R. (2004). A method of calculating regional consumer price differentials with illustrative evidence from India. *Review of Income and Wealth , 50* (1), 51-68.

Crawford, I., Laisney, F., & Preston, I. (2003). Estimation of household demand systems with teoretically compatible Engel curves and unit value specifications. *Journal of Econometrics , 114*, 221-241.

Curran, L., Wolman, H., Hill, E., & Furdell , K. (2006). Economic wellbeing and where we live: accounting for geographical Cost-of-Living differences in the US. *Urban Studies , 43* (13).

Deaton, A. (1988). Quality, quantity and spatial variation of price. *American Economic Review , 70* (3), 418-430.

Deaton, A., & Muellbauer, J. (1980). An Almost Ideal Demand System. *The American Economic Review , 70* (3).

Deaton, A., & Tarozzi , A. (2000). Prices and poverty in India. En *Research Program in Development Studies.* Priceton University.

Desai, A. V. (1969). A spatial index of cost of living. *Economic and Political Weekly , 4* (27), 1079-1081.

Dong, D., Gould, B., & Kaiser, H. (2004). Food demand in Mexico: an application of the Anemiya-Tobin approach to the estimation of a censored food system. *American Journal of Agricultural Economics , 86* (4), 1094-1107.

Dumond, J., Hirsch, B., & Macpherson, D. (1999). Wage differentials across labor markets and workers: does cost of living matter? *Economic Inquirity , 37* (4), 577-598.

Fava, V. L. (2010). Price dispersion and price indexes. *Applied Economics , 42*, 23-36.

Frankenberg, E., Duncan Thomas , & Kathleen Beegle. (1999). *Te Real Cost of Indonesia`s economic crisis: preliminary findings from the Indonesian Family Life Survey.* Santa Monica: Rand Corporation.

Fujita, M., & Krugman, P. (1995). When is the economy monocentric? *Regional Science and Urban Economics , 25*, 505-528.

García-Luque, O., LaFuente, M., & Faura, U. (2009). Disparidad territorial de la pobreza dinámica en España. *Estudios de Economía Aplicada , 27* (2), 417-436.

García-Mainar, I., & Montuenga-Gomez, V. (2003). The Spanish wage curve: 1994–1996. *Regional Studies , 37* (9), 929-945.

García-Serrano, C., & Arranz, J. (2013). Crisi económica y desigualdad salarial. *Papeles de Economía Española , 135*.

Garrido-Yserte, R., Mañas-Alcón, E., & Gallo-Rivera, M. (2012). Housing and cost of living: application to the Spanish regions. *Journal of Housing Economics , 21* (3), 246-255.

Gracia, A., Gil, , J., & Angulo, A. (1998). El consumo de alimentos en España: el consumidor rural versus consumidor urbano. *Estudios Regionales , 50*, 119-129.

Green, R., & Alston, J. (1990). Elasticities in AIDS model. *American Journal of Agricultural Economics , 72*, 442-445.

Haq, Z., Nazli, H., & Meilke, K. (2008). Implications of high food prices for poverty in Pakistan. *Agricultural Economics , 39*, 477-484.

Haworth, C., & Rasmussen , D. (1973). Determinants of metropolitan cost of living variations. *Southern Economic Journal , 40* (2), 183-192.

Hayes, P. (2005). Estimating UK regional price indices, 1974-96. *Regional Studies , 39* (3), 333-344.

Heien, D., & Wessells, R. (1990). Demand systems estimation with microdata: a censored regression approach. *Journal of Business and Economic Statistics , 8* (3), 365-371.

Helpman, E. (1998). The size of regions. En D. Pines, E. Sadka, & I. Zilcha, *Topics in Public Economics.* Cambridge: University press.

Henderson, J. V. (1987). General equilibrium modeling systems of cities. En *Handbook of Regional and Urban Economics* (págs. 927-956). Amsterdam: Mills.

Henderson, J. V. (1974). The size and types of cities. *American Economic Review , 64* (4), 640-656.

Herrero, C., Soler, A., & Villar, A. (2013). La pobreza en España y sus Comunidades Autónomas: 2006-2011. *Valencia: Ivie, 38 pp. http://dx.doi.org/10.12842/poverty_2012* .

Hoang, L. (2009). Estimation of food demand from Husehold Survey Data in Vietnam. *DEPOCEN Working Paper Series* (12).

Hogan, T., & Rex, T. (1984). Intercity differences in cost of living. *Growth and Change , 15* (4), 16-23.

Hoover, E. M. (1937). *Location Theory and the Shoe and Leather industry.* Cambridge, MA: Harvard University Press.

Houthakker, H. S., & Taylor, L. (1970). *Consumer demand in the United States.* Cambridge, MA: Harvard University Press.

Huq, A., & Arshad, F. (2010). Demand elasticities for different food items in Bangladesh. *Journal of Applied Sciences , 10*, 2369-78.

Isard, W. (1956). *Location and Space-economy.* Cambridge, MA: The Technology Press of Massachusets.

Isserman, A. M. (2001). Competititve advantages of rural America in the next century. *International Regional Science Review , 24*, 38-58.

Jolliffe, D. (2006). Poverty, prices and place: how sensitive is the spatial distribution of poverty to cost of living adjustments? *Economic Inquiry , 44* (2), 296-310.

Jurado, A., & Pérez-Mayo, J. (2012). Construction and evolution of a multidimensional well-being index for the Spanish regions. *Social Indicators Research , 107* (2), 259-279.

Kakhki, M. D., Shahnoushi, N., & Rezapour, F. (2010). An experimental comparison between demand systems of major food groups in urban economics. *American Journal of Applied Science , 7* (8), 1164-1167.

Kakwani, N., & Hill, R. (2002). Economic theory of spatial cost of living indices with an application to Thailand. *Journal of Public Economics , 86*, 71-97.

Kalecki, M. (1948). Determinants of the increase in the cost of living in the United States. *The Review of Economic and Statistics , 30* (1), 22-24.

Kaplow, L. (1995). Regional cost-of-living adjustments in tax-transfer schemes (No. w5008). *NAtional Bureau of Economic Research* .

Koenker, R., & Basset Jr, G. (1978). Regression quantiles. *Econometrica , 46* (1), 33-50.

Konus, A. A. (1939). The Problem of the True Index of the Cost of Living. *Econometrica , 7* (1), 10-29.

Kosfeld, R., Eckey, H.-F., & Türk, M. (2008). New economic geography and regional price levels. *Jahrbuch für Regionalwissenschaft , 28*, 43-60.

Krugman, P. (1991). Increasing returns and econmic geography. *Journal of Political Economy , 99*, 483-499.

Krugman, P., & Venables, A. (1995). Globalization and the inequality of nations. *Quartely Journal of Political Economics , 60*, 857-880.

Kurre, J. A. (2003). Is the cost of living less in rural areas? *Internatioonal Regional Science Review , 26* (1), 86.

Kurre, J. (1992). *The cost of liivng in rural Pennsylvania.* Erie, Pennsylvania State University: Center for Rural Pennsylvania.

Lanaspla, L., Perdiguero, A., & Sanz, F. (2004). La distribución del tamaño de las ciudades en España. *Revista de Economía Aplicada , 46* (14), 5-16.

Lanaspla, L., Pueyo, F., & Sanz, F. (2003). Evolution of the Spanish Urban Structure during the twentieth century. *Urban Studies , 40* (3), 567-580.

Larsen, E. R. (2007). Does the CPI mirror the Cost of Living? Engel`s law suggests not in Norway. *Scandinavian Journal of Economics , 109* (1), 177-195.

Lasarte, E., Paredes, D., & Fernández, E. (2012). A true cost of living for Spain using a microeconomic approach and censored data. *Documentos de Trabajo en Economía y Ciencia Regional , 34*.

Lasarte, E., Rubiera, F., & Paredes, D. (2014). City size and household food consumption: demand elasticities in Spain. *Applied Economics , 46* (14), 1624-1641.

Lee, L. (1976). Estimation of limited dependent variable models with two-stage methods. *PhD dissertation* . University of Rochester.

Lee, L. F., & Pitt, M. (1986). Microeconometric demand systems with binding nonnegative constraints: the dual approach. *Econometrica , 5*, 1237-1242.

Lewis, P., & Amdrews, N. (1989). Household demand in China. *Applied Economics , 21*, 793-807.

Loughrey, J., & O'Donogue, C. (2012). The welfare impact of price changes on household welfare and inequality 1999-2011. *The Economic and Social Review , 43* (1), 31-66.

Majumder, A., Ray, R., & Sinha, K. (2012). Calculating rural-urban food price differentials from unit values in household expenditure surveys: a comparison with existing methods and a new procedure. *American Journal of Agricultural Economics , 94* (5), 1218-1235.

McHugh, R. (1990). *Cost of living differences in urban versus rural areas: An analysis with expenditure data.* Iowa State University: Center for Agricultural and Rural Development Staff Report 90-SR 49.

McMahon, W. (1991). Geographical cost of living differences: an update. *AUREUEA Journal , 19* (3).

McMahon, W., & Melton, C. (1978). Measuring Cost of Living Variation. *Industrial Relations , 17* (3).

Menezes, T., Azzoni, C., & Silveira, F. (2008). Demand elasticities for food products in Brazil: a two stage budgeting system. *Applied Economics , 40* (19), 2557-2572.

Mills, E. (1967). An aggregative model of resource allocation in metropolitan area. *American Economic Review , 57* (2), 197-210.

Muellbauer, J. (1975). Aggregation, income distribution and consumer demand. *Review of Economic Studies , 42* (4), 525-543.

Muth, R. F. (1969). *Cities and Housing.* Chicago: University of Chicago Press.

Nelson, F. (1991). An inter-state cost of living index. *Educational Evaluation and Policy Analysis , 13* (1), 103-111.

Nord, M. (2000). Does it cost less to live in rural areas? Evidence from new data on food security and hunger. *Rural Sociology , 65* (1), 104-125.

Nord, M., & Cook, P. (1995). Do the proposed revisions of the poverty measures matter for rural America? (D. U.S. Department of Agriculture. Whashington, Ed.) *Economic Research Service Staff* (9514).

O'Donnell, C. J., & Rao, D. (2007). Predicting expenditure shares for computing PPP exchange rates. *Unpublished, University of Queensland* . Brisbane, Australia.

Ohlin, B. (1933). *Interregional and Internal Trade.* Cambridge, MA: Harvard University Press.

Paredes, D., & Iturra, V. (2013). Substitution bias and the construction of a spatial cost of living index. *Papers in Regional Science , 92* (1), 103-117.

Perali, F., & Chavas , J. (2000). Estimation of censored demand equations from large cross-section data. *American Journal of Agricultural Economics , 82*, 1022-37.

Pérez, J. (2013). Crisis económica y territorio: el impacto de la crisis sobre la desigualdad de rentas en las distintas regiones españolas. *Papeles de Economía Española , 135*.

Pérez-Mayo, J. (2008). *La dimensión territorial de la pobreza y la privación en España.* Fundación Alternativas.

Poggi, A. (2007). Does persistence of social exclusion exist in Spain? *Journal of Economic Inequality , 5* (1), 53-72.

Polese, M. (2010). *The wealth and poverty of regions: why cities matters.* Chicago: The University of Chicago Press.

Polese, M., Rubiera, F., & Shearmur, R. (2007). Observing regularities in location patterns: an analysis of the spatial distribution of economic activity in Spain. *European Urban & Regional Studies , 14* (2), 157-180.

Pollack, R., & Wales, T. (1992). Price-augmenting returns to scale: an application to Nonseparable Two-Stage technologies. *The Review of Economics and Statistics , 74* (2), 213-220.

Ramajo, J. (1996). La demanda de productos alimenticios, bebidas y tabaco en Extremadura: un enfoque de regresiones censuradas. *Investigación Agraria , 11* (3).

Ravallion, M. (1992). *Poverty comparisons: guide to concepts and methods.* Washington, DC: World Bank.

Ravallion, M., & Van de Walle, D. (1991). Urban-Rural cost of living differentials in a developing economy. *Journal of Urban Econmics , 29*, 113-127.

Roos, M. (2006). Regional price levels in Germany. *Applied Economics , 38* (13), 1553-1566.

Rosen, S. (1974). Hedonic prices and implicitit markets: product differentiation in pure competititon. *Journal of Political Economy , 82*, 34-55.

Serwood, M. (1975). Family budgets and geographic differences in prices levels. *Monthly Labor Review , 98* (4), 8-15.

Shefer, D. (1970). Comparable living costs and urban size: a statistical analysis. *Journal of the American Institute of Planners , 36* (6), 417-21.

Sheng, T. Y., Shamsudim, M. N., & Mohamed, Z. (2008). A complet demand system for food in Malasya. *Journal of Agricultural Economics , 5*, 17-29.

Shonkwiler, J. S., & Yen, S. T. (1999). Two step estimation of a censored system equations. *American Journal of Agricultural Economics , 81* (4), 972-982.

Sigit, H., & Surbakti, S. (1999). Social impact of the economic crisis in Indonesia. *Finalization Conference for Assesing the Social Impact of the Financial Crisis in Selected Asian Developed Countries.* Manila: ADB.

Simon, J. L., & Love, D. (1990). City size, prices and efficency for individual goods and services. *Ann Reg Sci , 24*, 163-175.

Slesnick, D. (2002). Prices and regional variation in welfare. *Journal of Urban Economics , 51*, 446-468.

Suedekum, J. (2006). Agglomeration and regional cost of living. *Journal of Regional Science , 46* (3), 529-543.

Tabuchi, T. (2001). On interregional price differentials. *The Japanese Economic Review , 52* (1).

Tabuchi, T. (2001). On interregional price differentials. *The Japanese Economic Review , 52* (1).

Tabuchi, T. (1998). Urban agglomeration and dispersion: A synthesis of Alonso and Krugman. *Journal of Urban Economics , 44*, 333-351.

Tabuchi, T., & Thisse, J.-F. (2003). Regional especialization, urban hierarchy and commuting costs. *International Economic Review , 47* (4), 1295-1317.

Tabuchi, T., Thisse, J.-F., & Zeng, D.-Z. (2005). On the number and size of the cities. *Journal of Economic Geograhy , 5*, 423-448.

Timmins, C. (2006). Estimating Spatial Differences in the Brazilian Cost of Living with Household Location Choices. *Journal of Development Economics , 80*, 59-83.

Viñuela, A., Rubiera, F., & Cueto, B. (2010). An analysis of urban size and territorial location effects on employment probabilities: the spanish case. *Growth and Change , 41* (4), 495-519.

Voicu, A., & Lahr , M. (1999). Expenditure based inter-area cost of living index. *Paper presented at the 38th annual meeting of the Western Regional Science Association.* Ojai, CA.

Walden, M. F. (1998). Geographic variation in consumer prices: implications for local price indices. *The Journal of Consumer Affairs , 32* (2), 204-226.

Walden, M. (1996). How much income variation really exists within a state? *Review of Regional Studies , 27* (3), 237-250.

Wales, T. J., & Woodland, A. (1983). Estimation of consumer demand systems with binding non-negative constraints. *Journal of Econometrics , 21*, 263-285.

Weber, A. (1909). *Alfred Weber's Theory of the Locstion of Industries.* Chicago: University of Chicago Press.

Wheaton, W. C. (1974). A comparative static analysis of urban spatial structure. *Journal of Economic Theory , 9* (2), 223-237.

Yen, S., Kan, K., & Su, S. J. (2002). Household demand for fats and oils: two-step estimation of a censored demand system. *Applied Economics , 14*, 1799-1806.

7 APPENDIXES

Appendix 1 PROBIT ESTIMATES 2008

	Share 1	Share 2	Share3	Share 4	Share5	Share6	Share7	Share8	Share9	Share10
Log of Expenditure	0.5571***	0.9037***	0.8113***	0.7158***	0.5348***	0.7308***	0.8220***	0.5551***	0.5077***	0.4960***
Household Size	0.2840***	0.0062***	0.0285***	0.0816***	0.1588***	0.0587***	0.1121***	0.1568***	0.1627***	0.1248***
Age	-0.0040***	-0.0048***	0.0027***	-0.0062***	0.0024***	0.0106***	0.0007***	-0.0036***	0.0017***	-0.0152***
Sex	0.0129***	0.0228***	0.0129***	0.0176***	0.0047***	0.0260***	0.0485***	0.0044***	0.0114***	-0.0172***
0 employed	0.2544***	-0.1788***	-0.1385***	0.2135***	-0.1403***	-0.1401***	0.0164***	-0.1379***	-0.1616***	-0.2647***
1 employed	0.3273***	-0.2166***	-0.1809***	0.1128***	-0.1206***	-0.0993***	-0.0456***	-0.1088***	-0.1016***	-0.1024***
2 employed	0.4174***	-0.0816***	-0.2628***	0.2468***	-0.1677***	-0.0322***	-0.0148***	-0.0792***	-0.0848***	-0.0936***
More than 2 employed	omitted	omitted	omitted	omitted	omitted	omitted	omitted	omitted	omitted	omitted
No studies	0.0957***	0.1461***	0.0021***	0.0726***	0.0765***	-0.2133***	-0.0921***	-0.0716***	0.0572***	0.0041***
First cycle studies	-0.1158***	0.1808***	0.0776***	-0.0413***	0.0468***	-0.2104***	-0.1436***	-0.0267***	0.0520***	-0.0032*
Second cycle studies	0.0510***	0.1962***	-0.0454***	-0.0916***	0.0636***	-0.2040***	-0.1353***	0.0124***	0.0810***	0.0622***
High degree studies	omitted	omitted	omitted	omitted	omitted	omitted	omitted	omitted	omitted	omitted
Income less tan 500	0.5200***	0.0441***	-0.0291***	0.0257***	0.2601***	0.0109***	-0.1174***	0.1562***	0.1575***	0.1370***
Income 500-1,000	0.3482***	0.1738***	-0.0310***	0.0729***	0.1568***	-0.0273***	0.0168***	0.0839***	0.0582***	0.0803***
Income 1,000-1,500	0.2942***	0.1425***	-0.0537***	0.1496***	0.1399***	0.0608***	0.0993***	0.0790***	0.0924***	0.1758***
Income 1,500-2,000	0.1720***	0.2275***	0.0840***	0.1611***	0.1163***	-0.0206***	0.0982***	0.0386***	0.0251***	0.1138***
Income 2,000-2,500	0.1525***	0.1989***	0.0600***	0.0521***	0.0739***	-0.0155***	0.1115***	0.0305***	0.0495***	0.0668***
Income 2,500-3,000	0.0252***	0.1462***	0.0525***	0.3043***	0.0980***	-0.0710***	-0.1190***	0.0776***	0.0473***	0.0724***
Income more tan 3,000	omitted	omitted	omitted	omitted	omitted	omitted	omitted	omitted	omitted	omitted
Single	-0.0882***	0.0061	0.1635***	0.0448***	0.0322***	-0.0819***	-0.0506***	-0.1083***	-0.2181***	-0.0915***
Married	-0.0938***	0.0860***	0.3077***	0.1262***	-0.0262***	-0.0336***	0.0514***	-0.0563***	-0.2613***	0.0431***
Widower	-0.0720***	-0.0222***	0.1313***	0.0645***	-0.0431***	-0.1687***	-0.2218***	-0.0323***	-0.2716***	0.1691***
Separate	-0.1943***	0.1242***	0.2136***	0.0459***	0.0639***	-0.1528***	-0.0398***	0.0903***	-0.2275***	0.1475***
Divorced	omitted	omitted	omitted	omitted	omitted	omitted	omitted	omitted	omitted	omitted
Capital of province	0.0186***	-0.0028***	-0.0054***	-0.0037***	-0.0053***	-0.0172***	-0.0432***	-0.0023***	-0.0013***	-0.0115***

	Share 1	Share 2	Share3	Share 4	Share5	Share6	Share7	Share8	Share9	Share10
ANDALUSIA	0.3386***	0.0774***	0.2595***	0.2776***	0.2487***	0.1433***	0.2213***	0.2815***	0.3545***	0.6658***
ARAGON	0.4877***	0.0724***	0.0870***	0.2139***	0.0959***	0.3338***	0.2181***	0.2696***	0.2124***	0.2010***
ASTURIAS	-0.0097	-0.2279***	0.1736***	0.7021***	0.0429***	0.0799***	0.2248***	0.1084***	0.0935***	0.1851***
BALEARIC	0.2017***	-0.0546***	-0.3445***	0.1060***	-0.0242***	0.0743***	0.1759***	0.1500***	-0.0740***	0.5702***
CANARY	0.4437***	0.1825***	0.2449***	0.3746***	0.4721***	0.2363***	0.5406***	0.6414***	0.6097***	1.0623***
CANTABRIA	0.2708***	-0.2435***	-0.0297***	0.0851***	0.0101***	0.0893***	0.0834***	0.1575***	0.0086**	0.0963***
CASTILELEON	0.1855***	-0.2006***	0.1127***	0.2431***	-0.0884***	-0.1211***	-0.0533***	-0.0494***	-0.1178***	-0.0077
CASTILE LAMANCHA	0.4030***	-0.0004	-0.0254***	0.1390***	-0.2645***	0.0284***	0.1654***	-0.1127***	-0.0562***	0.2575***
CATALONIA	0.0593***	-0.2208***	-0.0537***	0.2460***	0.0071***	-0.0434***	0.2755***	0.0993***	0.0945***	0.3701***
VALENCIA	0.3052***	-0.0874***	-0.0521***	0.4580***	-0.1086***	-0.0903***	0.3884***	-0.0206***	0.0649***	0.5421***
EXTREMADURA	0.5713***	-0.0303***	0.5315***	0.2818***	0.4365***	0.2603***	0.2391***	0.4466***	0.6079***	0.7603***
GALICIA	-0.0939***	-0.2792***	-0.0400***	0.3907***	-0.0317***	-0.2146***	-0.1180***	0.1198***	-0.0553***	0.0492***
MADRID	0.3746***	0.1722***	0.4403***	0.4705***	0.4161***	0.4655***	0.5969***	0.5468***	0.5332***	0.5745***
MURCIA	0.2999***	0.2442***	0.4401***	0.2601***	0.5946***	0.4615***	0.4975***	0.5248***	0.6966***	0.8962***
BASQUE COUNTRY	0.3039***	-0.2965***	-0.2177***	0.1259***	-0.1109***	0.0348***	0.1028***	0.0951***	0.0150***	-0.2028***
NAVARRA	0.2213***	-0.1058***	0.0552***	0.2723***	-0.0590***	-0.0153***	0.3737***	0.0983***	-0.0104*	-0.0720***
LA RIOJA	omitted	omitted	omitted	Omitted	omitted	omitted	omitted	omitted	omitted	omitted
_cons	-3.0002***	-4.9726***	-5.4704***	-3.9373***	-4.3149***	-4.3440***	-4.8551***	-3.8306***	-4.0111***	-2.4876***

(1) *, ** and *** represent the level of significance to 10%, 5% and 1%, respectively.
(2) Share1= Bread & Cereals; Share2 =Meat; Share3=Fish; Share4=Milk, eggs & cheese; Share5=Oils & fats; Share6= Fruits; Share7=Vegetables; Share8=Sugar; Share9=Coffee, tea & cocoa; Share10=Mineral water & other soft drinks.
(3) Income variables are expressed in monthly net Euros per household

Appendix 2 PROBIT ESTIMATES 2009

	Share 1	Share 2	Share3	Share 4	Share5	Share6	Share7	Share8	Share9	Share10
Log of Expenditure	0.1091***	0.4865***	0.4844***	0.3822***	0.2776***	0.4137***	0.3824***	0.2832***	0.3420***	0.3131***
Household Size	0.4505***	0.2473***	0.1714***	0.2914***	0.2555***	0.2239***	0.2764***	0.2810***	0.2339***	0.2387***
Age	0.0014***	0.0029***	0.0110***	0.0075***	0.0075***	0.0180***	0.0087***	0.0078***	0.0079***	-0.0070***
Sex	0.0080***	0.0251***	0.0373***	0.0290***	0.0266***	0.0403***	0.0530***	0.0400***	0.0250***	-0.0005***
0 employed	0.3082***	0.1476***	-0.0748***	0.1066***	-0.1310***	-0.1028***	0.1125***	-0.1311***	-0.0987***	-0.2794***
1 employed	0.2793***	0.0452***	-0.1229***	0.0234***	-0.1235***	-0.1337***	0.1057***	0.0119***	-0.0432***	-0.1528***
2 employed	0.2586***	0.0441***	-0.1503***	0.0316***	-0.1897***	-0.1106***	0.0050	-0.0555***	-0.0533***	-0.1014***
More than 2 employed	Omitted	omitted	Omitted	Omitted	Omitted	Omitted	Omitted	Omitted	Omitted	Omitted
No studies	0.1828***	0.2339***	0.1246***	0.1430***	0.1430***	-0.0305***	0.0106***	0.0166***	0.1185***	0.0351***
First cycle studies	0.1236***	0.2047***	0.1913***	0.2097***	0.1695***	-0.0234***	0.1128***	0.0221***	0.0959***	0.0659***
Second cycle studies	-0.0558***	0.1431***	0.0299***	0.0236***	0.0662***	0.0553***	0.0827***	0.0344***	0.0978***	0.0717***
High degree studies	Omitted	omitted	Omitted	Omitted	Omitted	Omitted	Omitted	Omitted	Omitted	Omitted
Income less tan 500	0.0205***	0.0013	0.0028	0.3176***	0.1821***	-0.1019***	0.0749***	0.1849***	0.2811***	0.2479***
Income 500-1,000	0.1200***	0.0258***	0.0800***	0.1507***	0.1394***	0.0509***	0.0172***	0.2177***	0.1546***	0.2450***
Income 1,000-1,500	0.0727***	-0.0202***	0.0420***	0.0986***	0.1416***	-0.0091***	0.0095***	0.1610***	0.1862***	0.3100***
Income 1,500-2,000	0.1770***	0.0363***	0.1290***	0.2299***	0.1652***	0.1789***	0.0975***	0.1332***	0.2035***	0.2137***
Income 2,000-2,500	0.1382***	-0.0248***	0.0603***	0.0511***	0.1232***	0.0529***	0.0249***	0.1537***	0.1364***	0.1646***
Income 2,500-3,000	0.1402***	0.0482***	-0.0010	0.0828***	0.0301***	0.0278***	0.1505***	0.0674***	0.0798***	0.1454***
Income more tan 3,000	Omitted	omitted	Omitted	Omitted	Omitted	Omitted	Omitted	Omitted	Omitted	Omitted
Single	-0.2747***	0.0680***	0.0743***	0.1032***	0.0247***	0.2417***	0.2673***	0.0466***	0.0256***	0.0433***
Married	-0.1628***	0.1767***	0.3402***	0.2344***	0.0489***	0.2683***	0.3963***	0.1644***	0.0432***	0.1652***
Widower	-0.0823***	0.0566***	0.0332***	0.0464***	-0.0102***	0.0916***	0.1560***	-0.0536***	-0.0870***	0.1504***
Separate	-0.2407***	0.2121***	0.0493***	0.1707***	0.0882***	-0.0204***	0.1337***	-0.0075***	0.0592***	0.1886***
Divorced	Omitted	omitted	Omitted	Omitted	Omitted	Omitted	Omitted	Omitted	Omitted	Omitted
Capital of province	0.0095***	-0.0159***	-0.0023***	-0.0225***	-0.0006***	0.0063***	-0.0203***	0.0068***	-0.0013***	-0.0189***

	Share 1	Share 2	Share3	Share 4	Share5	Share6	Share7	Share8	Share9	Share10
ANDALUSIA	0.2590***	0.0730***	0.1704***	0.0684***	0.1747***	0.0531***	0.2260***	-0.0169***	0.1873***	0.5829***
ARAGON	0.2982***	-0.0281***	0.1595***	-0.0463***	0.0313***	0.0694***	0.1740***	0.0253***	0.0711***	0.2083***
ASTURIAS	-0.0356***	-0.1854***	0.0284***	0.1044***	-0.0527***	-0.1020***	-0.0903***	-0.1406***	-0.1472***	0.0583***
BALEARIC	-0.0312	-0.3163***	-0.2166***	-0.0263***	0.0468***	0.0465***	0.0106	-0.0620***	-0.1706***	0.5573***
CANARY	-0.0050***	-0.1108***	-0.0463***	-0.0239***	0.2082***	0.0151***	0.1686***	0.2584***	0.2471***	0.7149***
CANTABRIA	-0.0606***	-0.1928***	-0.0425***	-0.2626***	-0.0451***	-0.1812***	-0.1838***	-0.0539***	-0.1354***	0.0166***
CASTILELEON	0.1426***	-0.0215***	0.0527***	-0.2447***	-0.2022***	0.0563***	-0.1033***	-0.2661***	-0.2871***	-0.1263***
CASTILE LAMANCHA	0.1229***	0.0168***	-0.1683***	-0.0039	-0.3107***	-0.0420***	0.1177***	-0.3828***	-0.2652***	0.2609***
CATALONIA	0.0228***	-0.1812***	-0.1787***	-0.1741***	-0.0766***	-0.0700***	0.1852***	-0.1246***	-0.1495***	0.3903***
VALENCIA	-0.1205***	-0.0160***	-0.2282***	0.1086***	-0.2355***	-0.1707***	0.2786***	-0.3732***	-0.2614***	0.4423***
EXTREMADURA	0.2838***	0.0278***	0.0711***	0.0542***	0.1980***	-0.0237***	0.0944***	0.0163***	0.2239***	0.4362***
GALICIA	0.0631***	-0.1632***	-0.0451***	-0.0527***	-0.0971***	-0.0658***	-0.1391***	-0.1493***	-0.2787***	0.0016
MADRID	0.1164***	0.0474***	0.1816***	-0.0765***	0.2138***	0.0786***	0.1846***	0.1768***	0.2405***	0.4362***
MURCIA	-0.0341***	-0.0757***	0.0895***	-0.0351***	0.2470***	-0.0529***	0.0876***	0.1288***	0.2313***	0.4893***
BASQUE COUNTRY	0.2108***	-0.2257***	-0.1512***	-0.3691***	-0.0473***	-0.1406***	0.0284***	-0.1620***	-0.1239***	-0.1836***
NAVARRA	0.1140***	-0.0751***	-0.0221***	-0.0597***	-0.0587***	0.0117	0.1729***	-0.0579***	-0.2263***	-0.1353***
LA RIOJA	Omitted	Omitted	Omitted	Omitted	Omitted	Omitted	Omitted	Omitted	Omitted	Omitted
_cons	-0.2012***	-4.0312***	-4.9004***	-3.2098***	-3.4005***	-4.1011***	-3.6574***	-3.4687***	-4.1258***	-2.5475***

(1) *, ** and *** represent the level of significance to 10%, 5% and 1%, respectively.
(2) Share1= Bread & Cereals; Share2 =Meat; Share3=Fish; Share4=Milk, eggs & cheese; Share5=Oils & fats; Share6= Fruits; Share7=Vegetables; Share8=Sugar; Share9=Coffee, tea & cocoa; Share10=Mineral water & other soft drinks.
(3) Income variables are expressed in monthly net Euros per household

Appendix 3 PROBIT ESTIMATES 2010

	Share 1	Share 2	Share3	Share 4	Share5	Share6	Share7	Share8	Share9	Share10
Log of Expenditure	0.1679***	0.4292***	0.5535***	0.3852***	0.3852***	0.4147***	0.4640***	0.3967***	0.3710***	0.3133***
Household Size	0.3365***	0.2427***	0.1432***	0.2237***	0.2237***	0.1490***	0.2217***	0.2364***	0.2074***	0.2579***
Age	0.0017***	0.0103***	0.0152***	0.0041***	0.0041***	0.0187***	0.0114***	0.0047***	0.0075***	-0.0065***
Sex	0.0287***	0.0240***	0.0297***	0.0522***	0.0522***	0.0469***	0.0436***	0.0284***	0.0253***	0.0114***
0 employed	0.0287***	-0.4040***	-0.1851***	0.0659***	0.0659***	-0.1118***	-0.3638***	-0.1260***	-0.1685***	-0.0182***
1 employed	-0.0243***	-0.4115***	-0.1682***	-0.1336***	-0.1336***	-0.1756***	-0.4483***	-0.1139***	-0.1147***	-0.0318***
2 employed	-0.1491***	-0.2356***	-0.1395***	-0.1878***	-0.1878***	-0.1309***	-0.4583***	-0.0560***	-0.1153***	0.0290***
More than 2 employed	Omitted	Omitted	Omitted	Omitted	Omitted	Omitted	Omitted	Omitted	Omitted	Omitted
No studies	0.3275***	0.2224***	0.1454***	0.0096***	0.0096***	-0.0650***	-0.0981***	-0.0105***	0.0891***	0.1067***
First cycle	0.2552***	0.1708***	0.1610***	0.1259***	0.1259***	0.0453***	-0.0357***	0.0443***	0.1000***	0.1400***
First cycle studies	0.0150***	0.1424***	-0.0073***	-0.0191***	-0.0191***	-0.0087***	-0.0698***	0.0586***	0.0433***	0.1155***
Second cycle studies	Omitted	Omitted	Omitted	Omitted	Omitted	Omitted	Omitted	Omitted	Omitted	Omitted
High degree studies	-0.0854***	0.2526***	0.2710***	0.0502***	0.0502***	0.0978***	0.3156***	0.4310***	0.3190***	0.2898***
Income less tan 500	-0.2074***	0.2879***	0.1494***	0.0800***	0.0800***	0.1403***	0.2140***	0.3705***	0.1971***	0.3018***
Income 500-1,000	-0.1534***	0.2623***	0.0983***	0.0959***	0.0959***	0.1839***	0.1872***	0.3028***	0.1608***	0.2595***
Income 1,000-1,500	-0.0455***	0.2845***	0.0751***	0.0629***	0.0629***	0.0759***	0.1853***	0.1830***	0.1642***	0.2357***
Income 1,500-2,000	0.1613***	0.1765***	0.0017	0.0725***	0.0725***	0.0197***	-0.0586***	0.1315***	0.1287***	0.1411***
Income 2,000-2,500	-0.2715***	0.0272***	-0.0328***	-0.0430***	-0.0430***	-0.0574***	-0.1293***	0.0573***	0.1095***	0.1069***
Income 2,500-3,000	Omitted	Omitted	Omitted	Omitted	Omitted	Omitted	Omitted	Omitted	Omitted	Omitted
Income more tan 3,000	-0.1888***	-0.0520***	0.1467***	-0.0340***	-0.0340***	0.2193***	0.0598***	-0.0972***	-0.0369***	0.0213***
Single	0.0641***	0.1128***	0.3294***	0.2797***	0.2797***	0.3713***	0.2599***	0.0530***	0.0144***	0.1431***
Married	-0.1438***	0.0707***	0.0976***	0.0801***	0.0801***	0.0931***	0.0233***	-0.0581***	-0.0957***	0.0559***
Widower	-0.2087***	0.1631***	0.1082***	-0.0426***	-0.0426***	0.1807***	0.0809***	-0.0887***	-0.0446***	0.1729***
Divorced	Omitted	Omitted	Omitted	Omitted	Omitted	Omitted	Omitted	Omitted	Omitted	Omitted
Capital of province	-0.0116***	-0.0010***	0.0102***	0.0026***	0.0026***	-0.0042***	-0.0187***	0.0029***	0.0011***	-0.0168***

	Share 1	Share 2	Share3	Share 4	Share5	Share6	Share7	Share8	Share9	Share10
ANDALUSIA	0.0011***	-0.1330***	0.3040***	0.0694***	0.0694***	0.3636***	0.3990***	0.1237***	0.3503***	0.5210***
ARAGON	-0.0442***	-0.1055***	0.3039***	0.0883***	0.0883***	0.2793***	0.3207***	0.1305***	0.2505***	0.1657***
ASTURIAS	-0.0414***	-0.2542***	0.0714***	0.1863***	0.1863***	0.0772***	0.1063***	0.0083***	-0.0413***	0.0124***
BALEARIC	0.0003	-0.1653***	-0.0599***	0.1254***	0.1254***	0.2574***	0.3504***	0.0690***	-0.0129***	0.4963***
CANARY	0.0955***	-0.1729***	0.0223***	-0.0742***	-0.0742***	0.2261***	0.1421***	0.2204***	0.3207***	0.7453***
CANTABRIA	0.0094	-0.0997***	0.2404***	0.4546***	0.4546***	0.0541***	0.0582***	0.1423***	0.1789***	0.2188***
CASTILELEON	0.0619***	-0.1847***	0.1532***	-0.0860***	-0.0860***	0.1972***	-0.1114***	-0.2449***	-0.2213***	-0.1861***
CASTILE LAMANCHA	0.2031***	-0.2635***	0.0656***	-0.0318***	-0.0318***	0.1710***	0.1642***	-0.2186***	-0.1361***	0.0757***
CATALONIA	-0.0398***	-0.1741***	0.0007	0.0980***	0.0980***	0.3113***	0.3268***	0.0143***	0.0370***	0.3806***
VALENCIA	-0.0777***	-0.2024***	0.0762***	-0.1213***	-0.1213***	0.0507***	0.2595***	-0.1351***	-0.1140***	0.2941***
EXTREMADURA	-0.1575***	-0.1548***	0.2468***	0.0523***	0.0523***	0.0655***	0.0980***	0.2720***	0.4436***	0.3101***
GALICIA	0.2833***	0.0669***	0.2229***	0.4329***	0.4329***	0.1485***	0.1109***	0.0429***	-0.0442***	0.0500***
MADRID	-0.1141***	-0.1530***	0.2539***	-0.0140***	-0.0140***	0.4565***	0.3018***	0.1073***	0.2495***	0.3858***
MURCIA	0.0492***	0.0575***	0.2860***	0.1311***	0.1311***	0.2274***	0.3200***	0.1481***	0.3677***	0.5404***
BASQUE COUNTRY	0.0145***	-0.2046***	0.0206***	-0.3171***	-0.3171***	0.0462***	0.2137***	-0.0409***	0.0230***	-0.2025***
NAVARRA	0.1389***	-0.0533***	0.1504***	0.0717***	0.0717***	0.2120***	0.1773***	0.1313***	0.0111***	-0.1559***
LA RIOJA	Omitted	Omitted	Omitted	Omitted	Omitted	Omitted	Omitted	Omitted	Omitted	Omitted
_cons	-0.0767***	-3.3627***	-5.7210***	-2.7617***	-2.7617***	-4.0935***	-3.7997***	-4.0994***	-4.2803***	-2.7191***

(1) *, ** and *** represent the level of significance to 10%, 5% and 1%, respectively.
(2) Share1= Bread & Cereals; Share2 =Meat; Share3=Fish; Share4=Milk, eggs & cheese; Share5=Oils & fats; Share6= Fruits; Share7=Vegetables; Share8=Sugar; Share9=Coffee, tea & cocoa; Share10=Mineral water & other soft drinks.
(3) Income variables are expressed in monthly net Euros per household

Appendix 4 PROBIT ESTIMATES 2011

	Share 1	Share 2	Share3	Share 4	Share5	Share6	Share7	Share8	Share9	Share10
Log of Expenditure	-0.0325***	-0.0325***	0.5174***	0.2392***	0.3313***	0.4020***	0.2310***	0.3206***	0.3550***	0.2766***
Household Size	0.4766***	0.4766***	0.1871***	0.2832***	0.2494***	0.2300***	0.2589***	0.2433***	0.2013***	0.3148***
Age	0.0088***	0.0088***	0.0128***	0.0102***	0.0086***	0.0167***	0.0067***	0.0035***	0.0064***	-0.0107***
Sex	0.0231***	0.0231***	0.0246***	0.0285***	0.0102***	0.0289***	0.0206***	0.0289***	0.0206***	0.0072***
0 employed	0.3901***	0.3901***	0.1368***	-0.2376***	-0.0198***	0.2687***	0.0263***	-0.0334***	-0.1375***	0.0713***
1 employed	0.3047***	0.3047***	0.0981***	-0.1961***	-0.0405***	0.1628***	-0.0299***	-0.0065***	-0.1163***	0.0118***
2 employed	0.2620***	0.2620***	0.0615***	-0.1323***	-0.0266***	0.2700***	-0.0669***	0.0304***	-0.0717***	0.0641***
More than 2 employed	Omitted	Omitted	Omitted	Omitted	Omitted	Omitted	Omitted	Omitted	Omitted	Omitted
No studies	-0.0395***	-0.0395***	0.1695***	-0.0346***	0.1289***	-0.1151***	-0.0855***	0.0738***	0.1845***	0.0825***
First cycle studies	0.0457***	0.0457***	0.1980***	0.0626***	0.1276***	0.0666***	0.0032***	0.0703***	0.1345***	0.0417***
Second cycle studies	-0.1037***	-0.1037***	0.0950***	0.0703***	0.0997***	0.0771***	0.0656***	0.0456***	0.1115***	-0.0151***
High degree studies	Omitted	Omitted	Omitted	Omitted	Omitted	Omitted	Omitted	Omitted	Omitted	Omitted
Income less tan 500	-0.2313***	-0.2313***	0.0006	0.1862***	0.3195***	-0.1763***	0.1407***	0.1557***	0.1684***	0.1613***
Income 500-1,000	-0.0791***	-0.0791***	-0.0278***	0.1516***	0.1763***	0.1180***	0.0767***	0.1335***	0.0966***	0.1928***
Income 1,000-1,500	-0.0738***	-0.0738***	-0.0383***	0.1291***	0.1617***	0.0427***	0.0574***	0.1576***	0.1262***	0.2157***
Income 1,500-2,000	-0.0847***	-0.0847***	0.0033***	0.2385***	0.0781***	0.0784***	0.0828***	0.1237***	0.0988***	0.1898***
Income 2,000-2,500	-0.0343***	-0.0343***	-0.0462***	0.0488***	-0.0017	0.0061***	0.0920***	0.1310***	0.0271***	0.1524***
Income 2,500-3,000	-0.1030***	-0.1030***	0.0097***	0.0958***	0.0451***	-0.0654***	0.0416***	0.1347***	0.0493***	0.0189***
Income more tan 3,000	Omitted	Omitted	Omitted	Omitted	Omitted	Omitted	Omitted	Omitted	Omitted	Omitted
Single	0.0996***	0.0996***	0.0519***	-0.0628***	-0.0084***	-0.0104***	-0.0722***	-0.0073***	-0.0562***	-0.0774***
Married	0.2704***	0.2704***	0.1768***	0.0122***	0.0402***	0.1154***	0.1227***	0.1037***	0.0291***	0.0568***
Widower	-0.0835***	-0.0835***	-0.1240***	-0.2007***	-0.0400***	-0.0403***	-0.0371***	0.0238***	-0.1442***	0.1186***
Separate	-0.1289***	-0.1289***	-0.0376***	-0.0784***	0.0176***	-0.1736***	-0.1966***	0.0998***	-0.0216***	0.0692***
Divorced	Omitted	Omitted	Omitted	Omitted	Omitted	Omitted	Omitted	Omitted	Omitted	Omitted
Capital of province	-0.0328***	-0.0328***	-0.0196***	-0.0141***	-0.0082***	-0.0066***	-0.0130***	0.0053***	-0.0037***	-0.0171***

	Share 1	**Share 2**	**Share3**	**Share 4**	**Share5**	**Share6**	**Share7**	**Share8**	**Share9**	**Share10**
ANDALUSIA	0.1780***	0.1780***	0.1507***	0.4583***	0.2237***	0.1571***	0.3462***	0.2115***	0.2314***	0.6207***
ARAGON	0.3139***	0.3139***	0.0248***	0.0752***	0.0039	0.0485***	0.3611***	0.1087***	-0.0147***	0.2629***
ASTURIAS	0.4881***	0.4881***	-0.0519***	0.4017***	-0.0006	0.0508***	0.0950***	0.1665***	-0.0945***	0.1061***
BALEARIC	0.3366***	0.3366***	-0.3280***	0.4336***	-0.0454***	0.1614***	0.2699***	0.0314***	-0.0323***	0.5284***
CANARY	0.1120***	0.1120***	0.0053	0.1746***	0.2677***	0.1042***	0.1387***	0.4140***	0.3584***	0.6694***
CANTABRIA	-0.0044	-0.0044***	0.0413***	0.1772***	0.0260***	-0.0092	-0.0071	0.2337***	0.1081***	0.0994***
CASTILELEON	0.2695***	0.2695***	0.0209***	0.2387***	-0.1771***	0.0559***	-0.0994***	-0.1375***	-0.3258***	-0.1589***
CASTILE LAMANCHA	0.7633***	0.7633***	-0.0565***	0.1857***	-0.2929***	0.2501***	0.1921***	-0.2237***	-0.1757***	0.1205***
CATALONIA	0.4205***	0.4205***	-0.0241***	0.2034***	-0.0497***	0.1511***	0.4075***	0.0692***	-0.0111***	0.3942***
VALENCIA	0.2470***	0.2470***	-0.1760***	0.2879***	-0.3042***	0.0483***	0.2297***	-0.1366***	-0.1493***	0.3888***
EXTREMADURA	0.4289***	0.4289***	0.0824***	0.2634***	0.2064***	0.0880***	0.1295***	0.1667***	0.2719***	0.3480***
GALICIA	0.5530***	0.5530***	0.0881***	0.4583***	-0.0344***	0.0802***	0.1263***	0.1077***	-0.1388***	-0.0262***
MADRID	0.2981***	0.2981***	0.0660***	0.1885***	0.2703***	0.1759***	0.3303***	0.2514***	0.2041***	0.3481***
MURCIA	0.1943***	0.1943***	0.2270***	0.3697***	0.1059***	0.1484***	0.2805***	0.2666***	0.1928***	0.4966***
BASQUE COUNTRY	0.2856***	0.2856***	-0.1985***	0.0665***	-0.0867***	-0.0263***	-0.0104	0.0433***	-0.0212***	-0.0943***
NAVARRA	0.5446***	0.5446***	-0.0847***	0.2031***	-0.1204***	0.1488***	0.2746***	0.0852***	-0.0895***	-0.1727***
LA RIOJA	Omitted	Omitted	Omitted	Omitted	Omitted	Omitted	Omitted	Omitted	Omitted	Omitted
_cons	0.7249***	0.7249***	-5.3106***	-1.7642***	-3.9857***	-4.1679***	-1.6884***	-3.6809***	-4.0290***	-2.2056***

(1) *, ** and *** represent the level of significance to 10%, 5% and 1%, respectively.
(2) Share1= Bread & Cereals; Share2 =Meat; Share3=Fish; Share4=Milk, eggs & cheese; Share5=Oils & fats; Share6= Fruits; Share7=Vegetables; Share8=Sugar; Share9=Coffee, tea & cocoa; Share10=Mineral water & other soft drinks.
(3) Income variables are expressed in monthly net Euros per household

Appendix 5 PROBIT ESTIMATES 2012

	Share 1	Share 2	Share3	Share 4	Share5	Share6	Share7	Share8	Share9	Share10
Log of Expenditure	0.0881***	0.4487***	0.5790***	0.4922***	0.3664***	0.4699***	0.3213***	0.3814***	0.4609***	0.3649***
Household Size	0.3178***	0.2489***	0.1494***	0.2315***	0.2252***	0.1429***	0.2697***	0.2605***	0.1770***	0.2780***
Age	0.0006***	0.0011***	0.0114***	0.0079***	0.0070***	0.0131***	0.0053***	0.0023***	0.0042***	-0.0093***
Sex	0.0099***	0.0364***	0.0272***	0.0230***	0.0120***	0.0307***	0.0265***	0.0265***	0.0281***	-0.0045***
0 employed	0.4043***	-0.0261***	0.0004***	-0.2373***	-0.2343***	-0.0280***	0.3625***	0.0218***	-0.2247***	0.0472***
1 employed	0.3150***	-0.0938***	-0.0211	-0.2014***	-0.2213***	-0.0930***	0.3358***	0.0100***	-0.1862***	0.1190***
2 employed	0.4479***	0.0062	0.0194***	-0.2523***	-0.2400***	-0.1499***	0.2700***	-0.0518***	-0.1927***	0.0338***
More than 2 employed	Omitted	Omitted	Omitted	Omitted	Omitted	Omitted	Omitted	Omitted	Omitted	Omitted
No studies	0.2429***	0.1939***	0.1513***	0.1431***	0.1724***	-0.2111***	0.0609***	0.0480***	0.1005***	0.0229***
First cycle studies	0.1177***	0.1707***	0.0955***	0.0348***	0.0979***	-0.0763***	0.1238***	0.0293***	0.0609***	0.0028***
Second cycle studies	0.1821***	0.1698***	0.0761***	0.1056***	0.1616***	-0.0188***	0.1429***	0.0779***	0.1079***	0.0220***
High degree studies	Omitted	Omitted	Omitted	Omitted	Omitted	Omitted	Omitted	Omitted	Omitted	Omitted
Income less tan 500	-0.1312***	0.0542***	0.1042***	0.1784***	0.2498***	-0.1003***	-0.0624***	0.0055***	0.2923***	0.1865***
Income 500-1,000	-0.0367***	0.0727***	0.1650***	0.1249***	0.2105***	0.0715***	0.0502***	0.1764***	0.2344***	0.2364***
Income 1,000-1,500	0.0675***	0.1970***	0.2541***	0.0300***	0.2155***	0.1242***	0.1540***	0.1272***	0.2424***	0.1692***
Income 1,500-2,000	0.1339***	0.1497***	0.2703***	0.0680***	0.1970***	0.0188***	0.1763***	0.1194***	0.2296***	0.1601***
Income 2,000-2,500	0.1035***	0.2377***	0.1313***	-0.0356***	0.1621***	0.0184***	0.1793***	0.0676***	0.1788***	0.1208***
Income 2,500-3,000	0.0252***	0.0620***	0.1272***	-0.0697***	0.0753***	0.0647***	0.1389***	0.0658***	0.1158***	0.1021***
Income more tan 3,000	Omitted	Omitted	Omitted	Omitted	Omitted	Omitted	Omitted	Omitted	Omitted	Omitted
Single	-0.2660***	0.0432***	0.0502***	0.0351***	0.0210***	0.1311***	-0.1818***	0.0042***	0.1243***	-0.1441***
Married	0.1145***	0.2515***	0.2372***	0.1223***	0.0837***	0.4162***	0.1215***	0.1238***	0.2294***	-0.0172***
Widower	-0.0293***	0.0644***	-0.0415***	-0.0565***	0.0217***	0.1450***	-0.1309***	0.0001	0.1146***	0.0934***
Separate	0.1927***	0.2189***	-0.0345***	0.1165***	0.1210***	-0.0049	-0.0748***	0.0912***	0.2465***	0.1254***
Divorced	Omitted	Omitted	Omitted	Omitted	Omitted	Omitted	Omitted	Omitted	Omitted	Omitted
Capital of province	0.0024***	-0.0049***	-0.0057***	-0.0079***	0.0094***	-0.0087***	-0.0159***	0.0077***	0.0051***	0.0009***

	Share 1	Share 2	Share3	Share 4	Share5	Share6	Share7	Share8	Share9	Share10
ANDALUSIA	0.1777***	0.0755***	0.0906***	0.0792***	0.2155***	0.1913***	0.2588***	0.1678***	0.3035***	0.4679***
ARAGON	-0.0818***	0.1969***	-0.0018	0.0414***	-0.0115***	0.3764***	0.2716***	0.0559***	-0.1079***	0.1391***
ASTURIAS	0.0513***	-0.1641***	0.0146***	0.1037***	0.0695***	0.1155***	-0.1325***	0.1263***	0.0584***	0.0477***
BALEARIC	-0.1227***	-0.1817***	-0.3972***	-0.1067***	0.0299***	0.2378***	0.0888***	0.0626***	0.0144***	0.4140***
CANARY	-0.0228***	-0.0457***	-0.0980***	-0.0265***	0.2452***	0.3021***	0.1458***	0.2852***	0.4872***	0.8111***
CANTABRIA	-0.2729***	-0.1216***	-0.0824***	-0.0466***	0.1959***	0.0395***	-0.2473***	0.0129***	0.0846***	0.0255***
CASTILELEON	0.0295***	-0.1023***	-0.1256***	-0.1719***	-0.2494***	0.0684***	-0.2041***	-0.2535***	-0.2531***	-0.2833***
CASTILE LAMANCHA	-0.0450***	-0.1265***	-0.0431***	-0.1026***	-0.1576***	0.0932***	0.0437***	-0.1284***	-0.0659***	0.1626***
CATALONIA	0.0154	-0.2041***	-0.0760***	-0.2244***	-0.0776***	0.1588***	0.1678***	0.0165***	-0.0633***	0.2436***
VALENCIA	0.0383***	-0.0442***	-0.1714***	-0.1465***	-0.2749***	-0.0497***	0.0803***	-0.1710***	-0.1455***	0.2618***
EXTREMADURA	0.3932***	0.0484***	0.2638***	0.0979***	0.3014***	0.3082***	0.2685***	0.2058***	0.4030***	0.3635***
GALICIA	0.3835***	-0.1486***	0.0050	0.3088***	-0.0094***	0.0916***	-0.0453***	-0.0159***	-0.0267***	0.0474***
MADRID	-0.0523***	-0.1719***	-0.0768***	0.0024	0.1377***	0.2812***	0.0315***	0.1901***	0.3261***	0.3517***
MURCIA	-0.1394***	0.0460***	0.2624***	-0.1001***	0.2106***	0.2718***	0.3310***	0.1606***	0.3492***	0.4250***
BASQUE COUNTRY	0.1009***	-0.1233***	-0.1217***	-0.0850***	0.0172***	0.2158***	0.1169***	0.0110***	0.0242***	-0.1565***
NAVARRA	0.1917***	-0.2442***	-0.1301***	-0.1607***	-0.0745***	0.1829***	0.1581***	0.0044	-0.0833***	-0.2503***
LA RIOJA	Omitted	Omitted	Omitted	Omitted	Omitted	Omitted	Omitted	Omitted	Omitted	Omitted
_cons	0.1764***	-3.5991***	-5.9062***	-3.7662***	-4.2508***	-4.3525***	-2.8966***	-4.0540***	-5.1705***	-3.0038***

(1) *, ** and *** represent the level of significance to 10%, 5% and 1%, respectively.
(2) Share1= Bread & Cereals; Share2 =Meat; Share3=Fish; Share4=Milk, eggs & cheese; Share5=Oils & fats; Share6= Fruits; Share7=Vegetables; Share8=Sugar; Share9=Coffee, tea & cocoa; Share10=Mineral water & other soft drinks.
(3) Income variables are expressed in monthly net Euros per household

Appendix 6 AIDS ESTIMATES OF THE MODEL DEFINED IN EQUATION (29)

Coefficient	2008	2009	2010	2011	2012
α_1	0.0840***	0.0278***	0.0137***	0.0246***	0.0137***
α_2	0.4009***	0.3747***	0.3675***	0.3803***	0.3675***
α_3	0.1277***	0.1831***	0.1865***	0.2069***	0.1865***
α_4	0.1278***	0.0953***	0.1026***	0.0995***	0.1026***
α_5	0.0607***	0.0584***	0.0525***	0.0500***	0.0525***
α_6	0.0399***	0.0505***	0.0670***	0.0717***	0.0670***
α_7	0.0871***	0.0888***	0.1055***	0.0943***	0.1055***
α_8	0.0006***	0.0087***	0.0102***	0.0097***	0.0102***
α_9	0.0126***	0.0043***	-0.0003***	0.0037***	-0.0003***
β_1	-0.0255***	-0.0498***	-0.0543***	-0.0487***	-0.0543***
β_2	0.0666***	0.0634***	0.0601***	0.0584***	0.0601***
β_3	0.0099***	0.0279***	0.0251***	0.0299***	0.0251***
β_4	-0.0057***	-0.0163***	-0.0115***	-0.0129***	-0.0115***
β_5	0.0150***	0.0116***	0.0114***	0.0068***	0.0114***
β_6	-0.0225***	-0.0178***	-0.0156***	-0.0103***	-0.0156***
β_7	-0.0048***	-0.0029***	0.0002***	-0.0029***	0.0002***
β_8	-0.0080***	-0.0067***	-0.0052***	-0.0068***	-0.0052***
β_9	0.0008***	-0.0023***	-0.0029***	-0.0020***	-0.0029***
γ_{11}	0.0337***	0.0488***	0.0461***	0.0485***	0.0461***
γ_{12}	-0.0312***	-0.0365***	-0.0337***	-0.0320***	-0.0337***
γ_{13}	-0.0076***	-0.0113***	-0.0137***	-0.0131***	-0.0137***
γ_{14}	-0.0071***	-0.0035***	-0.0066***	-0.0066***	-0.0066***
γ_{15}	-0.0029***	-0.0015***	-0.0036***	-0.0004***	-0.0036***
γ_{16}	0.0031***	0.0034***	0.0037***	0.0013***	0.0037***
γ_{17}	0.0023***	-0.0019***	-0.0004***	-0.0018***	-0.0004***
γ_{18}	0.0050***	0.0024***	0.0040***	0.0012***	0.0040***
γ_{19}	0.0005***	0.0000***	0.0002***	-0.0008***	0.0002***
γ_{22}	0.0445***	0.0420***	0.0500***	0.0465***	0.0500***
γ_{23}	0.0115***	0.0131***	0.0096***	0.0110***	0.0096***
γ_{24}	-0.0010***	-0.0039***	-0.0023***	-0.0031***	-0.0023***
γ_{25}	0.0007***	0.0006***	0.0000***	-0.0021***	0.0000***
γ_{26}	0.0015***	0.0052***	0.0013***	0.0053***	0.0013***
γ_{27}	-0.0029***	-0.0012***	-0.0012***	-0.0015***	-0.0012***
γ_{28}	-0.0060***	-0.0048***	-0.0053***	-0.0066***	-0.0053***
γ_{29}	-0.0052***	-0.0058***	-0.0079***	-0.0081***	-0.0079***
γ_{33}	0.0156***	0.0175***	0.0195***	0.0140***	0.0195***
γ_{34}	-0.0041***	-0.0032***	-0.0026***	-0.0002***	-0.0026***
γ_{35}	0.0011***	-0.0026***	0.0007***	0.0002***	0.0007***
γ_{36}	-0.0074***	-0.0047***	-0.0049***	-0.0026***	-0.0049***
γ_{37}	0.0010***	0.0039***	0.0024***	0.0013***	0.0024***
γ_{38}	-0.0012***	-0.0024***	-0.0039***	-0.0044***	-0.0039***
γ_{39}	-0.0022***	-0.0035***	-0.0055***	-0.0033***	-0.0055***
γ_{44}	0.0159***	0.0160***	0.0151***	0.0160***	0.0151***
γ_{45}	-0.0002***	-0.0017***	-0.0003***	-0.0015***	-0.0003***
γ_{46}	-0.0028***	-0.0028***	-0.0025***	-0.0029***	-0.0025***
γ_{47}	-0.0011***	0.0003***	-0.0018***	-0.0014***	-0.0018***
γ_{48}	0.0002***	-0.0003***	0.0005***	0.0003***	0.0005***

Coefficient	2008	2009	2010	2011	2012
γ_{49}	-0.0009***	-0.0008***	-0.0005***	-0.0005***	-0.0005***
γ_{55}	0.0053***	0.0078***	0.0055***	0.0036***	0.0055***
γ_{56}	0.0031***	0.0002***	0.0022***	0.0016***	0.0022***
γ_{57}	0.0002***	-0.0022***	0.0005***	0.0006***	0.0005***
γ_{58}	-0.0048***	-0.0023***	-0.0028***	-0.0022***	-0.0028***
γ_{59}	0.0000***	0.0008***	-0.0008***	-0.0002***	-0.0008***
γ_{66}	-0.0017***	-0.0018***	-0.0007***	-0.0018***	-0.0007***
γ_{67}	-0.0024***	-0.0031***	-0.0028***	-0.0033***	-0.0028***
γ_{68}	0.0015***	0.0005***	0.0007***	0.0015***	0.0007***
γ_{69}	0.0013***	0.0009***	0.0011***	0.0001***	0.0011***
γ_{77}	0.0005***	-0.0017***	0.0004***	0.0020***	0.0004***
γ_{78}	-0.0014***	0.0013***	-0.0007***	0.0002***	-0.0007***
γ_{79}	0.0023***	0.0016***	0.0014***	0.0023***	0.0014***
γ_{88}	0.0092***	0.0096***	0.0117***	0.0129***	0.0117***
γ_{89}	-0.0019***	-0.0023***	-0.0022***	-0.0020***	-0.0022***
γ_{99}	0.0061***	0.0109***	0.0171***	0.0157***	0.0171***
r_1	0.0032***	-0.0013***	-0.0016***	-0.0013***	-0.0016***
r_2	0.0031***	0.0000***	-0.0002***	0.0005***	-0.0002***
r_3	0.0031***	0.0010***	0.0012***	0.0016***	0.0012***
r_4	0.0034***	-0.0020***	-0.0023***	-0.0025***	-0.0023***
r_5	0.0028***	-0.0034***	-0.0043***	-0.0041***	-0.0043***
r_6	0.0024***	0.0010***	0.0016***	0.0018***	0.0016***
r_7	0.0032***	0.0009***	0.0006***	0.0016***	0.0006***
r_8	0.0030***	0.0000***	-0.0014***	-0.0009***	-0.0014***
r_9	0.0030***	-0.0010***	-0.0008***	-0.0007***	-0.0008***
r_{10}	0.0032***	-0.0002***	-0.0014***	-0.0006***	-0.0014***
r_{11}	0.0034***	0.0005***	0.0009***	0.0004***	0.0009***
r_{12}	0.0030***	0.0006***	0.0013***	0.0011***	0.0013***
r_{13}	0.0032***	0.0003***	-0.0005***	-0.0001***	-0.0005***
r_{14}	0.0031***	-0.0006***	-0.0006***	-0.0012***	-0.0006***
r_{15}	0.0038***	0.0016***	0.0012***	0.0012***	0.0012***
r_{16}	0.0033***	0.0022***	0.0016***	0.0024***	0.0016***
δ_1	0.6720***	-0.2305***	-0.2460***	-0.0891***	-0.2460***
δ_2	0.2711***	0.4881***	0.3530***	0.3360***	0.3530***
δ_3	-0.0285***	0.0351***	0.0341***	0.0505***	0.0341***
δ_4	0.3031***	0.0504***	0.1596***	0.1106***	0.1596***
δ_5	0.0546***	0.0466***	0.0473***	0.0325***	0.0473***
δ_6	-0.0089***	-0.0590***	-0.0713***	-0.0052***	-0.0713***
δ_7	0.0250***	0.0593***	0.0945***	0.1036***	0.0945***
δ_8	0.0181***	0.0187***	0.0289***	0.0266***	0.0289***
δ_9	0.0322***	0.0277***	0.0305***	0.0291***	0.0305***
c_1	0.0003***	-0.0004***	-0.0003***	-0.0006***	-0.0003***
c_2	-0.0002***	-0.0009***	-0.0008***	-0.0010***	-0.0008***
c_3	0.0006***	-0.0004***	-0.0008***	-0.0011***	-0.0008***
c_4	0.0001***	-0.0006***	-0.0006***	-0.0007***	-0.0006***

Notes:
(1) *, ** and *** represent the level of significance to 10%, 5% and 1%, respectively.
(2) The subindexes correspond to the following goods: 1=Bread & Cereals; 2=Meat; 3=Fish; 4=Milk,eggs & cheese; 5=Oils & Fats; 6=Fruits; 7=Vegetables; 8=Sugar; 9=Coffee, tea & cocoa; 10=Mineral water & other soft drinks.
(3)$r_1 = Andalusia; r_2 = Aragón;\ r_3 = Asturias; r_4 = Balearic\ Islands; r_5 = Canary\ Islands;\ r_6 = Cantabria; r_7 = Castile\ y\ León; r_8 = Castile\ La\ Mancha; r_9 = Catalonia; r_{10} = Valencian\ Community; r_{11} = Extremadura; r_{12} = Galicia; r_{13} = Madrid; r_{14} = Murcia; r_{15} = Navarra; r_{16} = Basque\ Country; r_{17} = La\ Rioja$
(4)$c_1 = municipalities\ of\ more\ than\ 100{,}000\ inhabitants;\ c_2 = municipalities\ between\ 50{,}000 - 100{,}000;\ c_3 = municipalities\ between\ 20{,}000 - 50{,}000;\ c_4 = municipalities\ between\ 10{,}000 - 20{,}000;\ c_5 = municipalities\ of\ less\ than\ 10{,}000\ inhabitants$