



**Universitat de les
Illes Balears**

Facultad de Economía y Empresa

Memoria del Trabajo de Fin de Grado

Análisis y modelización del gasto de los hogares a partir de microdatos de la ECPF.

Francisco Javier Cueto Vicens

Grado de Economía
Año académico 2013-14

DNI del alumno: 43176291R

Trabajo tutelado por Magdalena Cladera
Departamento de Economía Aplicada

El autor no autoriza el acceso público a este Trabajo de Fin de Grado.

Índice

·INTRODUCCIÓN.....	3
·PARTE PRIMERA:.....	4
1.1 Marco teórico del modelo.....	4
1.2 Base de datos.....	4
·PARTE SEGUNDA:.....	6
2.1 Descripción de las variables.....	6
2.2 Modelización.....	10
·PARTE TERCERA:.....	16
3.1 Efectos reflejados en el modelo.....	16
·CONCLUSIONES.....	19
·BIBLIOGRAFÍA.....	20
·ANEXO.....	21

Introducción

En el presente proyecto pretendo llevar a cabo un análisis y modelización del gasto de los hogares españoles del año 2012. El fin de este trabajo es analizar los efectos de distintas variables sobre el *gasto anual total* de las familias. El método que usaré para la estimación será por MCO (mínimos cuadrados ordinarios) y la base de datos que utilizada para dicha estimación provendrá de la EPF (encuesta de presupuestos familiares), obtenida de la web del Instituto Nacional de Estadística (INE).

El proyecto se dividirá en tres partes:

- Parte primera, marco teórico del modelo y base de datos, donde realizaré una explicación teórica del modelo de MCO y una explicación de la base de datos a usar.
- Parte segunda, modelización, en este apartado llevaré a cabo una explicación de cómo he estimado el modelo y una explicación de las variables seleccionadas.
- Parte tercera, análisis, donde analizaré los efectos reflejados en el modelo.

PARTE PRIMERA.

1.1 Marco teórico del modelo.

Como es bien conocido el modelo MCO o regresión lineal, es un modelo matemático que pretende obtener el valor de la variable estimada a través otras variables independientes o explicativas multiplicadas por su efecto marginal y un término de error¹. Este modelo proyecta el siguiente esquema:

$$Y_t = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_n X_n + \varepsilon$$

Donde cada término representa;

Y_t : es la variable estimada.

X_1, X_2, X_3, X_n : son las variables independientes o explicativas.

$\beta_1, \beta_2, \beta_3, \beta_n$: indican el efecto de cada variable sobre la variable estimada.

ε : es el término de error.

β_0 : representa el valor de una constante.

β_i ($i > 0$): donde cada subíndice indica que β pertenece a cada variable.

n : indica el número de variables que forman el modelo.

1.2 Base de datos.

La base de datos que usaré es la EPF del año 2012 (INE, 2013). La EPF tiene como objetivo principal recopilar información anual sobre la naturaleza y destino de los gastos de consumo, así como las diversas características relativas a las condiciones de vida de los hogares. El tamaño de la muestra es de aproximadamente 24.000 hogares, que estos permanecen en la muestra dos años consecutivos, renovándose cada año la mitad de la muestra. Esta base de datos empezó a utilizarse desde el año 2006, cuya antecesora es la Encuesta Continua de Presupuestos Familiares (ECPF).

En el año 1997 el INE empezó a realizar la ECPF (INE, 2013)² donde trataba de recopilar información sobre la naturaleza y el destino de las rentas de los hogares. Esta base de datos estuvo en vigor hasta el cuarto trimestre del 2005, a partir de ahí, cambió la metodología y pasó a ser la EPF, la cual tenía una periodicidad trimestral. El tamaño

¹ Para más información consultar Wooldridge, J. M. (2006). *Introducción a la econometría: un enfoque moderno*. Editorial Paraninfo.

² Información obtenida de la base de datos de la EPF del año 2012, en la que se comparan las características de la EPF y la ECPF.

de la muestra era de unos 8.000 hogares, a los cuales se les pedía una colaboración de ocho trimestres seguidos. Con lo que la muestra anual, teniendo en cuenta las rotaciones de los hogares, era de aproximadamente 11.000 hogares.

Con la EPF se pretende seguir las recomendaciones del EUROSTAT y peticiones de algunos usuarios. También, con la nueva metodología se pretende reducir el esfuerzo de las familias a la hora de colaborar.

PARTE SEGUNDA.

2.1 Descripción de las variables.

Como ya he comentado anteriormente, el objetivo de este trabajo es modelizar el gasto anual de los hogares. A la hora de decidir las variables he tomado como referencia los artículos publicados por los siguientes autores: Lázaro, Moltó & Sanchez (2000); Alegre, Mateo & Pou (2013); Burton, Dorsett & Young (2000); Ketkar & Ketkar (1987); Costa-Font, Kanavos & Rovira (2007); y Aristei & Pieroni (2008), ya que todos ellos tienen como objetivo modelizar distintos tipos de gastos. Para ello, he tratado que las variables del modelo estimado, siguieran ciertas características comunes con los trabajos anteriores.

Las características que siguen las variables seleccionadas para el proyecto contienen: información general, características relativas al hogar, datos del sustentador principal, características de la vivienda principal, gasto de consumo del hogar e ingresos regulares mensuales del hogar.

El modelo está formado por un total de 39 variables, de las cuales 38 son variables independientes y una variable dependiente. De las 38 variables independientes que forman el modelo, cuatro, son variables cuantitativas y treinta y cuatro, son variables cualitativas, estas se han transformado en variables ficticias para realizar la modelización.

Al haber variables cualitativas, las variables siguen cierta codificación³. El nombre de las variables son las siglas de la característica que representan, seguidas por un número, ese número indica en qué caso la variable toma valor 1 (por ejemplo, la variable que indica el tamaño del municipio de residencia: TAMAMU1, TAMAMU2, TAMAMU3 y TAMAMU4). Luego, las variables cuantitativas incluyen el valor que toma cada individuo en la variable.

La variable dependiente es LOG_GASTOT, esta variable representa en la base de datos el importe total del gasto anual del hogar monetario y no monetario. Esta variable ha sido previamente modificada. Primero se dividió entre la variable FACTOR, para conseguir un efecto uniforme del factor poblacional, y seguidamente se pasó al logaritmo, para buscar suavizar los valores extremos de la muestra.

³ Consultar el Anexo para ver una descripción más detallada de cada variable.

La variable que aporta información general es la variable TAMAMU, indica el tamaño del municipio de residencia. Como se ha comentado anteriormente, las variables que son cualitativas siguen una codificación que está desglosada en el Anexo.

Las características relativas al hogar vienen representadas por las variables: NMIEMB, que indica el número de miembros del hogar; NUMOCUP, indica el número de miembros ocupados del hogar; NUMESTU, indica el número de estudiantes en el hogar; y la variable SITUOCUHOG, que clasifica la situación del hogar respecto a la ocupación. De estas variables, NMIEMB, NUMOCUP y NUMESTU, son variables cuantitativas, por tanto no siguen una codificación. En cambio la variable SITUOCUHOG es una variable cualitativa y esta sí sigue una codificación.

En cuanto a datos del sustentado principal, las variables que aportan información de este son: ESTUCOMSP, la cual indica los estudios completados por el sustentador principal; TIPOCONT, esta clasifica el tipo de contrato de trabajo del sustentador principal; y por último, SITSOCISP, indica la situación socioeconómica del sustentador principal. Estas tres variables son cualitativas.

Aportando información sobre las características de la vivienda principal: REGTEN, clasifica el régimen de tenencia de la casa de residencia principal; y TIPOCASA, indica el tipo de casa de la vivienda principal. Ambas son variables cualitativas.

Las variables FUENPRIN (clasifica la fuente principal de ingresos) y LOG_IMPEXAC (indica el importe exacto de los ingresos mensuales netos totales del hogar), aportan información de los ingresos regulares mensuales del hogar. Siendo la primera cualitativa y la segunda cuantitativa. La variable LOG_IMPEXAC es el logaritmo de la variable IMPEXAC, se ha tomado como logaritmo para suavizar los efectos extremos de la muestra.

Tabla 1. Media, Mediana y Moda de las variables.

VARIABLES	Media	Mediana	Moda
FUENPRIN		2	2
SITUOCUHOG		5	8
TAMAMU		3	1
SITSOCISP		3	5
TIPOCONT		1	1
TIPOCASA		2	2
REGTEN		1	1
ESTUCOMSP		2	2
NUMESTU	0.19	0	0
NUMOCUP	0.99	1	1
NMIEMB	2.77	3	2
LOG_IMPEXAC	7.39	7.44	(7.33-7.54)
LOG_GASTOT	10.13	10.15	(10.07-10.25)

Elaboración propia, datos obtenidos de la base de datos EPF 2012.

En la tabla anterior se recoge información sobre la media, la mediana y la moda de las variables del modelo. Como se puede observar, en las variables cualitativas no se ha aportado la media, ya que, al ser variables ficticias una media aritmética no aporta información relevante, en cambio la mediana y la moda sí.

La variable FUENPRIN tiene como mediana valor 2, eso significa, la observación que deja el 50% de las observaciones por debajo y el otro 50% por encima toma valor 2. La moda, toma también el valor 2, siendo por tanto este valor el que presenta una frecuencia mayor. Teniendo en cuenta los valores que tomaba esta variable antes de transformarla en una variable ficticia, el valor 2 significaba que el sustentador principal trabaja por cuenta ajena.

La variable SITUOCUHOG tiene como mediana el valor 5, por tanto, la característica que deja el 50% de observaciones por debajo y el otro 50% por encima es que el sustentador principal o el cónyuge se encuentran ocupados y ninguno de los otros

miembros ocupados (si los hay). Mientras que tiene como moda el valor 8, la característica que tiene mayor frecuencia, es que no hay ningún ocupado en el hogar.

La variable TAMAMU tiene como mediana que el municipio tiene entre 20.000 y 50.000 habitantes. Por otro lado la moda, representa que el municipio tiene más de 100.000 habitantes.

La variable SITSOCISP tiene como mediana que el sustentador principal es un trabajador independiente y agricultor o trabajador a la agricultura. Mientras que como moda tiene que el sustentador principal está retirado o jubilado.

En lo que se refiere a la variable TIPOCONT, se observa que la media es la misma que la moda. Por tanto, la característica que deja a ambos lados el 50% de las observaciones y que tiene la mayor frecuencia, es que el sustentador principal tiene contrato indefinido.

Al igual que la variable TIPOCONT, la variable TIPOCASA tiene como mediana y moda el mismo valor. Por tanto, la mediana y la moda representan ambas que la principal vivienda de la familia es una casa media.

La variable REGTEN muestra una misma mediana y moda. La característica que representa el valor 1, es que la vivienda principal del hogar no tiene hipoteca en curso.

La variable ESTUCOMSP, la última variable cualitativa, muestra la misma mediana y moda, como algunas de las anteriores variables. Esta mediana y moda es 2, esto representa que el sustentador principal tiene terminado el primer ciclo de la educación secundaria.

En las variables NMIEMB, NUMOCUP, NUMESTU, LOG_IMPEXAC y LOG_GASTOT, se ha tenido en cuenta la media aritmética, ya que son variables cuantitativas.

Por lo que se refiere a la variable NMIEMB, la media indica que hay 2.77 miembros por hogar. La mediana de esta variable es 3. Mientras que la cantidad de miembros de un hogar que tiene una frecuencia mayor es 2.

Según la tabla, la variable NUMOCUP tiene una media de 0.99 miembros del hogar ocupados, tanto el valor de la mediana como el de la frecuencia mayor es 1.

La variable NUMESTU indica que la media de miembros que estudian en un hogar es 0.19, mientras que la mediana y la moda son 0.

Las dos últimas variables a comentar son LOG_IMPEXAC y LOG_GASTOT. Esta última es la variable dependiente del modelo. Al estar las dos variables en logaritmos, la relación que explica una variable con la otra es la elasticidad, esto se comentará más adelante en la parte tercera.

La media de la variable LOG_IMPEXAC es 7.39, mientras que la mediana es 7.44. Por lo que se refiere a la moda, esta está representada en intervalos, ya que es una variable continua. La moda está en el intervalo 7.33-7.54.

La variable dependiente LOG_GASTOT tiene una media de 10.13, una mediana de 10.15 y el intervalo que representa la moda es 10.07-10.25.

2.2 Modelización.

En este apartado se pretende presentar el modelo y llevar a cabo una explicación de los pasos seguidos para concluir en el modelo final.

Como ya se ha comentado anteriormente, el método que he usado para estimar el modelo es por MCO, y el programa con el que se ha trabajado ha sido el *Gretl 1.9.5*.⁴. La primera estimación que llevé a cabo del modelo por MCO, presentó diversos problemas, tanto de heterocedasticidad y no normalidad de los residuos.

El programa usado para llevar a cabo la estimación ofrecía diversos métodos para solucionar distintos tipos de heterocedasticidad, al no tener conocimiento del tipo que presentaba el modelo, no se han usado.

El procedimiento utilizado, ha sido la estimación robusta de los errores estándar, que como se sabe, son la raíz de la diagonal de la matriz de varianzas y covarianzas. Este tipo de estimaciones se usan cuando no se tiene conocimiento del tipo de heterocedasticidad que presenta el modelo. El programa aporta diversas instrucciones en la “Guía del Usuario” para elaborar la estimación robusta de la matriz de covarianzas. En el apartado 17 de la misma, se encuentra la explicación de cómo llevar a cabo la estimación robusta, más concretamente, en el apartado 17.2 se explica cómo hacer la estimación robusta para datos de sección cruzada, para la cual, hay que hacer diversas modificaciones en las herramientas. Los pasos a seguir son: herramientas – preferencias – general – HCCME. Una vez llegado al apartado HCCME, el programa ofrece tres

⁴ Gretl es un software econométrico obtenido gratuitamente en <http://gretl.sourceforge.net/>.

métodos para realizar la estimación robusta (según la “Guía del Usuario” es indiferente usar cualquiera de las tres vías que se ofrecen), el método seleccionado fue HC1. En la “Guía del Usuario” de *Gretl* se puede leer:

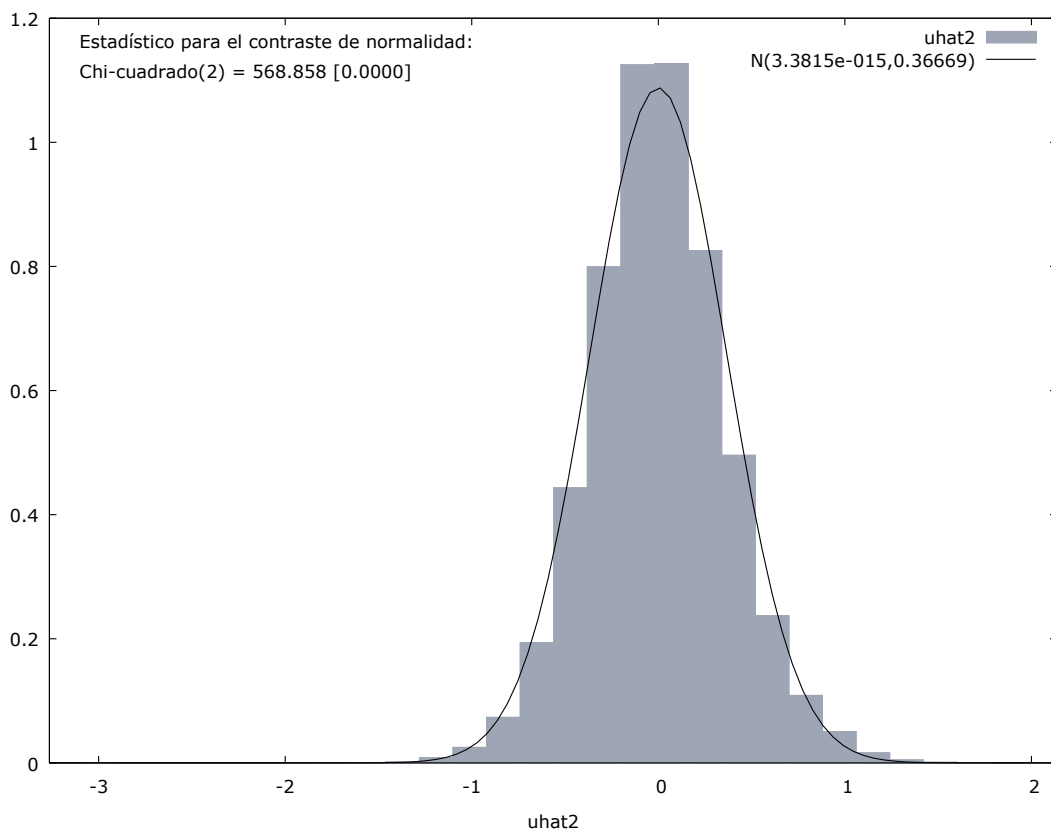
“*HC1: Applies a degrees-of-freedom correction, multiplying the HC0 matrix by $T/(T-k)$.*”

Donde HC0, es la estimación sin usar la regresión robusta de la matriz de covarianzas.

Una vez realizados los cambios, la estimación resultante, es la estimación robusta, la cual se mostrará más adelante.

A la hora de analizar la normalidad de los residuos de dicha estimación se obtiene:

Gráfico 1. Distribución de los residuos.



Elaboración propia, datos elaborados con *Gretl*.

Contraste de la hipótesis nula de distribución normal:
Chi-cuadrado(2) = 568.858 con valor p 0.00000

Como se ve en el contraste, el valor p es menor de 0.05, por tanto se rechaza la hipótesis nula, lo cual indica que los errores no siguen una distribución normal.

Al observar el gráfico, da la impresión que sigue una distribución muy similar a la normal, y a demás, al tener una muestra tan grande y tener tantas observaciones puede

que sumadas den un valor muy alto al estadístico de contraste de normalidad, que puede que sea el motivo de rechazar la hipótesis nula. Por tanto, no lo consideraremos un problema importante.

Por otro lado, en el apartado de la colinealidad, el contraste arroja la siguiente información:

Tabla 2. Contraste de colinealidad.

Factores de inflación de varianza (VIF)

Mínimo valor posible = 1.0

Valores mayores que 10.0 pueden indicar un problema de colinealidad.

FUENPRIN1	1.074
FUENPRIN2	5.150
FUENPRIN3	8.473
SITUOCUHOG1	3.711
SITUOCUHOG2	5.259
SITUOCUHOG3	1.748
SITUOCUHOG4	1.629
SITUOCUHOG5	2.487
SITUOCUHOG6	1.883
SITUOCUHOG7	1.319
SITUOCUHOG8	4.457
TAMAMU1	1.753
TAMAMU2	1.353
TAMAMU3	1.413
TAMAMU4	1.310
SITSOCISP1	2.638
SITSOCISP2	2.125
SITSOCISP3	9.160
SITSOCISP4	3.637
SITSOCISP5	7.454
TIPOCONT1	9.777
TIPOCONT2	2.462
TIPOCONT3	7.442
TIPOCASA1	1.002
TIPOCASA2	2.907
TIPOCASA3	2.740
REGTEN1	1.519
REGTEN2	1.337
REGTEN3	5.139
REGTEN4	1.520
REGTEN5	2.160
ESTUCOMSP1	2.144
ESTUCOMSP2	2.088
ESTUCOMSP3	1.507
NUMESTU	1.221
NUMOCUP	9.408
NMIEMB	1.849
LOG_IMPEXAC	2.312

Elaboración propia, datos elaborados con *Gretl*.

Según el contraste de colinealidad, los factores de inflación de la varianza no están lo suficientemente inflados como para considerar que haya dicho problema.

Por lo que se refiere a la especificación correcta del modelo, se indica a través del test RESET (Ramsey, 1969), el cual arroja la siguiente información:

Tabla 3. Contraste de especificación de RESET.

Contraste de especificación RESET (cuadrados y cubos)
 Estadístico de contraste: $F = 0.942010$,
 con valor $p = P(F(2,20844) > 0.94201) = 0.39$

Contraste de especificación RESET (cuadrados sólo)
 Estadístico de contraste: $F = 0.913995$,
 con valor $p = P(F(1,20845) > 0.913995) = 0.339$

Contraste de especificación RESET (cubos sólo)
 Estadístico de contraste: $F = 0.866945$,
 con valor $p = P(F(1,20845) > 0.866945) = 0.352$

Elaboración propia, datos elaborados con *Gretl*.

Como se puede observar, en los tres casos, se obtiene un valor p mayor a 0.05, el cual indica que la especificación es correcta.

Tras el análisis realizado y los resultados obtenidos, el modelo se considera válido. El cual queda reflejado a continuación:

Tabla 4. Modelo estimado por MCO.

Modelo por MCO, usando las observaciones 1-21808 ($n = 20885$)
 Se han quitado las observaciones ausentes o incompletas: 923
 Variable dependiente: LOG_GASTOT
 Desviaciones típicas robustas ante heterocedasticidad, variante HC1

	<i>Coefficiente</i>	<i>Desv. Típica</i>	<i>Estadístico t</i>	<i>Valor p</i>	
Const	6.49718	0.0574458	113.1010	<0.00001	***
FUENPRIN1	-0.0802008	0.034262	-2.3408	0.01925	**
FUENPRIN2	0.0756931	0.0187571	4.0354	0.00005	***
FUENPRIN3	0.0257916	0.0154849	1.6656	0.09581	*
SITUOCUHOG1	0.0225759	0.0300921	0.7502	0.45313	
SITUOCUHOG2	0.0453788	0.0140157	3.2377	0.00121	***
SITUOCUHOG3	0.0518507	0.0192432	2.6945	0.00706	***
SITUOCUHOG4	0.0151828	0.0436366	0.3479	0.72789	
SITUOCUHOG5	0.0847606	0.00997066	8.5010	<0.00001	***
SITUOCUHOG6	0.0313521	0.0203543	1.5403	0.12350	
SITUOCUHOG7	-0.015149	0.0395484	-0.3830	0.70169	
SITUOCUHOG8	0.0348162	0.0210939	1.6505	0.09885	*
TAMAMU1	0.0664908	0.00696675	9.5440	<0.00001	***
TAMAMU2	0.0528103	0.00941921	5.6067	<0.00001	***

TAMAMU3	0.0415797	0.00842215	4.9369	<0.00001	***
TAMAMU4	0.0346084	0.0093442	3.7037	0.00021	***
SITSOCISP1	-0.0666693	0.0217367	-3.0671	0.00216	***
SITSOCISP2	-0.0433229	0.0223584	-1.9377	0.05268	*
SITSOCISP3	-0.0535156	0.0242829	-2.2038	0.02755	**
SITSOCISP4	0.0248927	0.0179919	1.3835	0.16651	
SITSOCISP5	0.00675301	0.0159975	0.4221	0.67293	
TIPOCONT1	0.0851771	0.0207175	4.1114	0.00004	***
TIPOCONT2	0.0809318	0.0187867	4.3079	0.00002	***
TIPOCONT3	0.00677951	0.0192188	0.3528	0.72428	
TIPOCASA1	0.176162	0.0165094	10.6704	<0.00001	***
TIPOCASA2	0.255793	0.0138713	18.4405	<0.00001	***
TIPOCASA3	0.100137	0.0116398	8.6030	<0.00001	***
REGTEN1	0.057392	0.0180089	3.1869	0.00144	***
REGTEN2	-0.0187466	0.018045	-1.0389	0.29887	
REGTEN3	-0.179434	0.0193686	-9.2642	<0.00001	***
REGTEN4	-0.305051	0.0350471	-8.7040	<0.00001	***
REGTEN5	0.00858933	0.0227978	0.3768	0.70635	
ESTUCOMSP1	-0.191476	0.010154	-18.8571	<0.00001	***
ESTUCOMSP2	-0.0919122	0.00772701	-11.8949	<0.00001	***
ESTUCOMSP3	-0.056398	0.00790932	-7.1306	<0.00001	***
NUMESTU	0.0731811	0.00556288	13.1553	<0.00001	***
NUMOCUP	0.0457968	0.0135426	3.3817	0.00072	***
NMIEMB	0.0705263	0.00299295	23.5642	<0.00001	***
LOG_IMPEXAC	0.434263	0.00707484	61.3813	<0.00001	***

Media de la vble. dep.	10.15983	D.T. de la vble. dep.	0.555241
Suma de cuad. Residuos	2802.994	D.T. de la regresión	0.366691
R-cuadrado	0.564644	R-cuadrado corregido	0.563850
F(38, 20846)	1359.314	Valor p (de F)	0.000000
Log-verosimilitud	-8662.410	Criterio de Akaike	17402.82
Criterio de Schwarz	17712.74	Crit. de Hannan-Quinn	17504.01

Elaboración propia, datos elaborados con *Gretl*.

Al obtener en algunas variables, el valor p superior a 0,05 y por tanto, siendo así, consideradas no significativas, las he mantenido en el modelo, al ser correcta la especificación del mismo y para mantener las características de las variables cualitativas.

A la derecha del valor p de cada variable se pueden ver asteriscos. Un asterisco significa que el valor p está entre 0.05 y 0.1, dos asteriscos significan que el valor p está entre 0.01 y 0.05 y tres asteriscos significan que el valor p está por debajo de 0.01.

Por lo que se refiere a la bondad de ajuste, el R-cuadrado indica un valor de 0.56, mientras que el R-cuadrado corregido también indica un valor de 0.56. Eso significa que más del 55% de las observaciones se ven explicadas por este modelo.

PARTE TERCERA.

3.1 Efectos reflejados en el modelo.

En esta tercera parte del proyecto, comentaré los efectos de cada variable explicativa sobre la variable dependiente. Como ya he comentado en el marco teórico del modelo, los β_i son los efectos marginales de cada variable. Este β_i queda reflejado en el modelo como el coeficiente, mientras que el β_0 representa la constante.

El valor de la constante es 6.49, este es el valor que toma la función cuando las otras variables toman valor cero.

Los efectos marginales que tienen las variables con respecto a la variable dependiente son: cuando “no consta” la principal fuente de ingresos, tiene un efecto de -8.02%, es decir, tiene un efecto negativo sobre la variable dependiente. Cuando la principal fuente de ingresos es por trabajo por cuenta propia, y rentas de la propiedad y del capital, el gasto total aumenta un 7.56%; mientras que cuando es por trabajo por cuenta ajena la variable dependiente aumenta sólo un 2.57%.

En cuanto a la situación del hogar respecto a la ocupación se observa que: cuando el sustentador principal y cónyuge están ocupados, y al menos otro miembro también está ocupado, la variable dependiente aumenta un 2.25%. Cuando el sustentador principal y cónyuge están ocupados, y ninguno de los otros miembros está ocupado (si los hay) aumenta un 4.53%. Cuando el sustentador principal o el cónyuge ocupado, y otro de los miembros del hogar ocupados, el efecto es de 5.18% sobre la dependiente. Cuando el sustentador principal o cónyuge ocupado, y al menos otros dos miembros ocupados, el efecto sobre la variable dependiente es de 1.51%. Cuando sólo se encuentra el sustentador principal o el cónyuge, ocupados, el efecto es de 8.47%. Si ambos se encuentran desocupados, pero otro miembro del hogar si está ocupado, el efecto es de 3.13%. En la misma situación que la anterior, pero con al menos otros dos miembros del hogar ocupados, es de -1.51%. Y cuando no hay ningún ocupado en el hogar, el efecto es de 3.48%.

Por lo que se refiere al efecto del tamaño del municipio en el que se reside, se observa que: si se reside en un municipio con más de 100.000 habitantes el gasto total aumenta en un 6.64%. Si el municipio tiene entre 50.000 y 100.000 habitantes, el gasto aumenta en 5.28 %. Si el municipio tiene entre 20.000 y 50.000 habitantes, el gasto aumenta en 4.15%. Y si el municipio tiene entre 10.000 y 20.000 habitantes el efecto es de 3.46%.

Dependiendo de la situación socioeconómica del sustentador principal encontramos que: si el sustentador principal es trabajador manual, excepto agricultura, el efecto es -6.66%. Si no es un trabajador manual, excepto agricultura, el efecto se reduce a -4.33%. Si es un trabajador independiente y agricultor o trabajador a la agricultura este efecto es de -5.53%. Si el sustentador principal está parado, el gasto total aumenta un 2.48%. Mientras que si está jubilado o retirado el gasto total sólo aumenta un 0.67%.

Por lo que se refiere al tipo de contrato de trabajo del sustentador principal. Si no consta el tipo de contrato, el efecto es de 8.51%. Si el sustentador principal tiene un contrato de carácter indefinido, el gasto total aumenta un 8.09%. Mientras que si tiene un contrato de carácter eventual o temporal, el efecto es mínimo, sólo aumenta un 0.67%.

Dependiendo del tipo de casa de residencia habitual de la familia encontramos que: si no consta el tipo de casa, el efecto es un aumento del 17.61%. Si el tipo de casa es un chalé o casa grande el gasto total aumenta en un 25.57%. Mientras que si es una casa media, este efecto se reduce a tan solo un 10.01%.

Según el régimen de tenencia de la vivienda habitual observamos que, si la propiedad no tiene hipoteca en curso, el gasto aumenta un 5.73%. Si la propiedad tiene hipoteca en curso, este efecto se reduce hasta un -1.87%. Si el régimen de tenencia es alquiler, el gasto se reduce a -17.94%. Si es alquiler, pero alquiler reducido o renta antigua, el efecto se reduce hasta un -30.5%. Mientras que si el régimen de tenencia es cesión semigratuita, el gasto aumenta un -0.85%.

Por lo que hace a los estudios completados por el sustentador principal: si no tiene estudios o sólo con estudios de primer grado, el efecto es de -19.1%. Si tiene terminada la educación secundaria pero sólo hasta el primer ciclo, el efecto es de -9.19% y si tiene la educación secundaria hasta segundo ciclo, el efecto cambia hasta -5.63%.

La interpretación de las variables cuantitativas cambia un poco con respecto a las variables cualitativas. Las variables cuantitativas pueden tomar valores distintos a 0 o 1. Por tanto, por cada miembro del hogar que estudia, el gasto aumenta un 7.31%. Por cada miembro ocupado, el aumento es del 4.57% y por cada miembro del hogar, el gasto total aumenta un 7.05%.

A la hora de interpretar el efecto entre la variable LOG_IMPEXAC y la variable dependiente (LOG_GASTOT), al estar ambas representadas en logaritmos, el efecto se

interpreta como la elasticidad. La elasticidad representa el cambio porcentual de la variable dependiente cuando la variable explicativa varía un 1%⁵.

$$\begin{aligned} \text{Elasticidad } d_{y/x} = \beta_1 = (\Delta y/y) / (\Delta x/x) &\longrightarrow (\Delta y/y) = \beta_1 * (\Delta x/x) \\ &\approx \log(y) = \beta_1 * \log(x) \end{aligned}$$

Por tanto, en este modelo, el parámetro 0.43 representa la elasticidad entre el gasto total y los ingresos totales (GASTOT/INGRESO). Significa que por cada variación de un 1% de los ingresos mensuales totales de las familias, el gasto total anual varía un 0.43%.

⁵ Consultar Wooldridge, J. M. (2006). *Introducción a la econometría: un enfoque moderno*. Editorial Paraninfo.

Conclusiones.

Para concluir comentaré algunos aspectos que considero importantes resaltar del efecto de las variables.

Por lo que se refiere al tamaño del municipio de residencia, se observa que según se va reduciendo el tamaño del municipio se reduce el gasto total de los hogares. Alrededor de un 1% entre cada grupo de clasificación.

En cuanto al tipo de contrato, es notoria la diferencia entre un contrato indefinido o un contrato eventual, al aumentar el gasto total anual de los hogares en 8.09% y 0.67% respectivamente. Bajo mi punto de vista, esto podría deberse a poseer una mayor seguridad en la capacidad de afrontar gastos futuros por parte del trabajador indefinido frente al eventual, siendo así el gasto total anual más elevado.

Dependiendo del tipo de vivienda de residencia habitual de la familia, la diferencia existente entre casa grande o casa mediana es de más del 150% entre ambos. Esta diferencia sobre la variable dependiente, puede ser debida a que los hogares que poseen mayor capacidad económica llevan un nivel de vida mayor.

También considero importante destacar, el elevado aumento del gasto total anual de las familias, que conlleva la existencia de estudiantes en el hogar, siendo este de un 7.31% por cada estudiante. Y en contraposición, el reducido aumento, en un 4.57% por cada miembro ocupado del hogar, suponiendo la situación de ocupado un ingreso “extra” que se aporta al presupuesto familiar.

Por último, me referiré al efecto que produce que el sustentador principal se encuentre desocupado, ya que dado a la crisis española de los últimos tiempos, es una situación que está a la orden del día. Sería de esperar que si el sustentador principal está desocupado, el gasto total disminuyera, sin embargo, el gasto total en esta situación, aumenta en un 2.48%.

Bibliografía.

- Alegre, J., Mateo, S., & Pou, L. (2013). *Tourism participation and expenditure by Spanish households: The effects of the economic crisis and unemployment*. *Tourism Management*, 39, 37-49.
- Aristei, D., & Pieroni, L. (2008). *A double-hurdle approach to modelling tobacco consumption in Italy*. *Applied Economics*, 40(19), 2463-2476.
- Burton, M., Dorsett, R., & Young, T. (2000). *An investigation of the increasing prevalence of non purchase of meat by British households*. *Applied Economics*, 32(15), 1985-1991.
- Costa-Font, J., Kanavos, P., & Rovira, J. (2007). *Determinants of out-of-pocket pharmaceutical expenditure and access to drugs in Catalonia*. *Applied Economics*, 39(5), 541-551.
- <http://gretl.sourceforge.net/>. Página web desde donde se ha descargado *Gretl*. Última entrada 12-05-2014.
- INE (2013), *Encuesta de Presupuestos Familiares (2012)*, Fichero Base de datos, Instituto Nacional de Estadística, Madrid.
- Ketkar, K. W., & Ketkar, S. L. (1987). *Population dynamics and consumer demand*. *Applied economics*, 19(11), 1483-1495.
- Lazaro, N., Molto, M., & Sanchez, R. (2000). *Unemployment and consumption patterns*. *Applied Economics*, 32(3), 367-379.
- Ramsey, J. B. (1969). *Tests for specification errors in classical linear least-squares regression analysis*. *Journal of the Royal Statistical Society. Series B (Methodological)*, 350-371.
- Vidal Díaz de Rada Igúzquiza (1994), *Alternativas metodológicas al análisis de regresión lineal*, en Cuadernos de Economía. Vol. 22 321-346 1994. Universidad Pública de Navarra.
- Wooldridge, J. M. (2006). *Introducción a la econometría: un enfoque moderno*. Editorial Paraninfo.

Anexo.

FUENPRIN: fuente principal de ingresos.

FUENPRIN1 – No consta.

FUENPRIN2 – Trabajo por cuenta propia, y rentas de la propiedad y del capital.

FUENPRIN3 – Trabajo por cuenta ajena.

SITUOCUHOG: situación del hogar respecto a la ocupación.

SITUOCUHOG1 – SP y cónyuge ocupados, al menos otro miembro también ocupado.

SITUOCUHOG2 – SP y cónyuge ocupados, ningún otro miembro ocupado (si los hay).

SITUOCUHOG3 – SP o cónyuge ocupado, otro de los miembros ocupados.

SITUOCUHOG4 – SP o cónyuge ocupado, al menos otros dos miembros ocupados.

SITUOCUHOG5 – SP o cónyuge ocupado, ningún otro miembro ocupado (si los hay).

SITUOCUHOG6 – Ni el SP ni cónyuge ocupado, otro miembro ocupado.

SITUOCUHOG7 – Ni el SP ni cónyuge ocupado, al menos otros dos miembros ocupados.

SITUOCUHOG8 – Ningún ocupado en el hogar.

TAMAMU: tamaño del municipio de residencia.

TAMAMU1 – Municipio con más de 100.000 habitantes.

TAMAMU2 – Municipio entre 50.000 y 100.000 habitantes.

TAMAMU3 – Municipio entre 20.000 y 50.000 habitantes.

TAMAMU4 – Municipio entre 10.000 y 20.000 habitantes.

SITSOCISP: situación socioeconómica del sustentador principal

SITSOCISP1 – Trabajador manual, excepto agricultura.

SITSOCISP2 – Trabajador no manual, excepto agricultura.

SITSOCISP3 – Trabajador independiente y agricultor o trabajador a la agricultura.

SITSOCISP4 – Parado.

SITSOCISP5 – Retirado o jubilado.

TIPOCONT: tipo de contrato del sustentador principal.

TIPOCONT1 – No consta.

TIPOCONT2 – Indefinido.

TIPOCONT3 – Eventual/Temporal.

TIPOCASA: tipo casa de la vivienda principal.

TIPOCASA1 – No consta.

TIPOCASA2 – Chalé o casa grande.

TIPOCASA3 – Casa media.

REGTEN: régimen de tenencia de la vivienda principal.

REGTEN1 – Propiedad sin hipoteca en curso.

REGTEN2 – Propiedad con hipoteca en curso.

REGTEN3 – Alquiler.

REGTEN4 – Alquiler reducido (renta antigua).

REGTEN5 – Cesión semigratuita.

ESTUCOMSP: estudios completados del sustentador principal.

ESTUCOMSP1 – Sin estudios o con estudios de 1° grado.

ESTUCOMSP2 – Educación secundaria, primer ciclo.

ESTUCOMSP3 – Educación secundaria, segundo ciclo.

NUMESTU: número de estudiantes en el hogar.

NUMOCUP: número de miembros ocupados en el hogar.

NMIEMB: número de miembros del hogar.

LOG_IMPEXAC: logaritmo del importe exacto de los ingresos mensuales netos totales del hogar.