

«ESTUDIO DEL CONSUMO ELÉCTRICO DE LOS HOGARES A TRAVÉS DE LA ENCUESTA DE PRESUPUESTOS FAMILIARES»

by

«MERCEDES BLÁZQUEZ MONTERO»

A thesis submitted in conformity with the requirements
for the MSc in Economics, Finance and Computer Science

University of Huelva & International University of Andalusia



Septiembre 2023

«ESTUDIO DEL CONSUMO ELÉCTRICO DE LOS HOGARES A TRAVÉS DE LA ENCUESTA DE PRESUPUESTOS FAMILIARES»

MERCEDES BLÁZQUEZ MONTERO

Máster en Economía, Finanzas y Computación

Mónica Carmona Arango
Universidad de Huelva y Universidad Internacional de Andalucía

2023

ABSTRACT

This study presents an analysis of electricity consumption expenditures in Spanish households, based on information from family budget surveys (EPF). The objective is to verify whether there have been changes in consumption patterns over time, comparing the results of the same econometric model for the data from the Family Budget Survey in 2009 (Medina and Vicens, 2011), with an application of the model for the data in 2022. Obtaining as the main result the effect of household size and surface area, although remaining the variables that most affect electricity expenditure in the home, they reduce their importance.

JEL classification: I25, I28, J24, J28, O52.

Keywords: electric consumption, expenditure, spanish households, consumption patterns.

RESUMEN

En este estudio se presenta un análisis de gastos en consumo eléctrico en los hogares españoles, a partir de la información proveniente de las encuestas de presupuestos familiares (EPF). El objetivo es comprobar si se han producido cambios en los patrones de consumo a lo largo del tiempo, comparando los resultados del mismo modelo econométrico para los datos de la Encuesta de Presupuestos Familiares en 2009 (Medina y Vicens, 2011), con una aplicación del modelo para los datos en 2022. Obteniendo como principal resultado el efecto del tamaño del hogar y la superficie, aunque manteniéndose como las variables que más afectan al gasto eléctrico en el hogar, reducen su importancia.

Palabras claves: consumo eléctrico, gasto, hogares españoles, patrones de consumo.

AGRADECIMIENTOS

Agradecer a las personas que me han ayudado y han hecho posible este Trabajo de Fin de Master:

- A toda mi familia, por animarme siempre a conseguir las metas marcadas y creer siempre en mí, sus valores han hecho posible que llegue hasta aquí.
- Mis compañeros y amigos con los que he compartido tantas horas de biblioteca, risas, nervios y alegrías durante el Máster. Consiguiendo que este período haya sido una experiencia única en Huelva.
- A mi tutora Mónica Carmona Arango, por su ayuda, tiempo y haberme guiado para la realización del mismo.

Muchas gracias a todos.

TABLA DE CONTENIDOS

RESUMEN	iii
AGRADECIMIENTOS	iv
LISTA DE TABLAS.....	vi
LISTA DE FIGURAS.....	vii
1. INTRODUCCIÓN	1
2. OBJETIVO	2
3. REVISIÓN DE LA LITERATURA	2
4. METODOLOGÍA	5
4.1. Base de datos.....	5
4.2. Contenido de los ficheros	7
4.3. Variables.....	7
5. MODELO TEÓRICO	10
6. RESULTADOS.....	12
7. CONCLUSIONES	20
BIBLIOGRAFÍA	22
ANEXO.....	25

LISTA DE TABLAS

Tabla 4.1. Clasificación ECOICOP.	6
Tabla 4.2. Descripción de variables	9
Tabla 6.1. Resultados de la estimación MCO y regresión cuantílica	14
Tabla 6.2. Resultados obtenidos homocedasticidad MCO.....	15
Tabla 6.3. Resultados obtenidos contraste de normalidad	16
Tabla 6.4. Resultados obtenidos de mutlicolinealidad	17
Tabla 6.5. Resultados obtenidos de variables omitidas	18
Tabla 6.6. Resultados obtenidos error de especificación.....	18
Tabla 6.7. Resultados obtenidos significatividad conjunta	20

LISTA DE FIGURAS

Figura 6.2. Resultados obtenidos gráficamente normalidad de los residuos.... 16

Figura 6.3. Resultados obtenidos gráficamente presencia de anómalos 19

1. INTRODUCCIÓN

El presente estudio está enfocado a demostrar los conocimientos adquiridos en el Máster de Economía, Finanzas y Computación. De todos los conceptos estudiados durante dicha formación este estudio está centrado en aplicar técnicas de Minería de Datos (Data Mining), basado en analizar y reconocer patrones del consumo eléctrico en hogares españoles a través de la encuesta de presupuestos familiares (EPF) del INE.

Actualmente, debido a factores como la continua aparición de nuevas tecnologías, los cambios económicos y el crecimiento poblacional, el consumo de energía es cada vez mayor, lo que provoca cambios en sus tendencias de consumo, afectando a la demanda, por lo que es necesario tomar medidas. Teniendo en cuenta que la demanda de electricidad ha seguido creciendo en el sector residencial y otros sectores de consumo durante las últimas décadas, la electricidad representará casi una cuarta parte del consumo de energía final para 2040 agregando 8300TWh (Teravatios-hora) de generación de electricidad global. (Jestin-Fleury , 1994).

En el presente trabajo se analiza y explica el consumo de electricidad residencial en los hogares españoles mediante un modelo econométrico, junto a otras variables. De esta forma el modelo matemático juega un papel importante en la elaboración de estimaciones ya que permiten incluir los factores determinantes para el comportamiento de dicho consumo, como son el tipo de hogar, el número de personas del hogar, antigüedad de la vivienda, ingresos, etc.

Este documento se estructura de la siguiente manera:

Una vez introducido el tema de análisis, y se ha especificado los principales objetivos, en el tercer apartado se realiza una revisión literaria consultando fuentes de información secundarias acerca del gasto eléctrico. Posteriormente, se explica la metodología describiendo los datos que se han utilizado para el análisis, así como las distintas variables trabajadas, además de explicar el modelo teórico. Con todo esto pasaremos a presentar el modelo estimado junto con los resultados obtenidos.

Finalmente, se muestran conclusiones que se pueden extraer del estudio. Adjuntando por último bibliografías y anexos utilizados.

2. OBJETIVO

En esta sección se presentan los objetivos del presente trabajo fin de máter. Cuyo objetivo general se basa en analizar el consumo de energía eléctrica en los hogares españoles a través de un modelo econométrico, utilizando los datos de la EPF de 2022 y compararlo con el modelo econométrico obtenido por Mediana y Vicens para el consumo eléctrico en el año 2009.

3. REVISIÓN DE LA LITERATURA

El consumo de energía eléctrica es la demanda de una serie de equipos eléctricos del hogar, alumbrado, limpieza, cocina, calefacción, etc., por lo que la demanda se hace sobre los aparatos electrónicos y no sobre la electricidad de forma directa. La consecuencia es inmediata debido a que cambios en las características económicas, renta o precio, provocarán dos causas, uno a corto plazo sobre la utilización del equipamiento y otro sobre las existencias de equipos a largo plazo.

Para la revisión bibliográfica se han analizado estudios sobre la influencia de los precios y la influencia de los ingresos, una de las opciones como variable explicativa en los modelos desarrollados es el nivel de ingresos. Por ejemplo el estudio de Herriges y King (1995), se inclina por plantear un modelo de demanda con respuesta no lineal en los precios.

Por otro lado, artículos sobre elasticidad-precio y elasticidad-renta no son concluyentes y llevan a equivocaciones, como indica Espey y Espey (2004) sus valoraciones de elasticidades tienen un rango demasiado grande, para el precio entre -0.07 y -2.5 a largo plazo y entre -0.004 y -2.01 a corto plazo y para la renta entre 0.04 y 3.48 a corto plazo y entre 0.2 y 5.74 a largo plazo. Medina y Vicéns (2011) para estos datos a pesar de la alta variabilidad, llegan a dos conclusiones, primeramente que las elasticidades renta y precio son muy parecidas, en el corto plazo es mayor la elasticidad precio y a largo plazo es mayor la elasticidad

demanda. Por otra parte, que las elasticidades a corto plazo son más bajas y que no se pueden cambiar significativamente el gasto de los hogares en energía eléctrica variando el precio o los ingresos familiares.

Castro (2000) y Peña (1988), emplean análisis de series temporales sobre elasticidades, concluyendo que las estimaciones de las elasticidades superiores a la unidad no parecen aceptables.

Asimismo, Agostini, Plottier y Saavedra estudian la demanda residencial en Chile, seleccionando los datos de la Encuesta de Caracterización Socioeconómica Nacional, con un enfoque MCO. Concluyen que la elasticidad ingreso es inelástica, siendo menos inelástica la elasticidad precio de forma que una modificación en el precio tiene mayor impacto en el consumo de electricidad que un cambio en ingreso. (Agostini, Cecilia Plottier, & Saavedra, 2012).

Con el propósito de reconocer características socioeconómicas, geográficas o del equipamiento del hogar, encontramos análisis de Hancevic y Navajas, eligen Buenos Aires para evaluar la demanda residencial, mediante regresión cuantílica. Consiguen a partir la Encuesta Nacional de Gastos de los Hogares (ENGH) y del cálculo de la factura de electricidad residencial el consumo de kwh. Entre los resultados de este estudio deducen que los hogares que no tienen instalado el servicio de gas natural en promedio gastan más energía eléctrica que los que sí tienen, que los usuarios que tiene instalado en sus viviendas equipos de aire acondicionado tienen en promedio mayores niveles de consumo que los que no disponen y que los que habiten en casas en promedio consumen más que los que residen en departamentos (Hancevic & Navajas, 2015).

Las previsiones de la demanda residencial para 2030 según el Informe de la comisión de Expertos sobre Escenarios de transición Energética, con un supuesto crecimiento anual del PIB del 2% anual y un incremento del número de hogares del 4%, la demanda de energía se incrementaría en torno al 50% en los cuatro apartados del consumo doméstico: calefacción, agua caliente, iluminación y aire

acondicionado, pero sobre todo el equipo de aire acondicionado multiplica casi por cinco su gasto y en iluminación, aumentaría un 34% (De la Torre Prados, 2019).

Encontramos dificultades a la hora de recopilar información sobre el consumo eléctrico, pero existen fuentes fiables que permiten extraer contenido sobre la demanda eléctrica, aunque resultan insuficientes debido a que solo ofrecen información a nivel agregado, siendo difícil aproximar el comportamiento de sus componentes. A nivel agregado el más significativo es el sector residencial, estudios previos como el de Vicéns (2008), indica que el sector residencial es un 22% de energía utilizada.

Otra fuente de información para hacer el estudio sobre la energía eléctrica es el INE a través de la Encuesta de Presupuestos Familiares, facilita el consumo de los hogares a través de una encuesta. Basándose en esta información según el estudio de Medina y Vicéns (2011) indican que el consumo eléctrico medio de un hogar en la vivienda principal es de 288,7 kwh al mes, 49 euros, simbolizando un 2,4% de la renta del hogar. Estiman que el consumo eléctrico no se mantiene constante según el nivel de renta, los hogares con rentas más altas consumen más que los hogares de rentas más bajas, 62€ al mes duplicando el gasto de los usuarios que obtienen menor rentas. Además, señalan que al ser un bien básico para el hogar, el peso del consumo eléctrico sobre el ingreso sigue el comportamiento de la Ley de Engel, esta ley observa una relación estrecha entre las variaciones en el gasto y el nivel de ingreso, a medida que las personas perciben un nivel de renta o ingreso mayor el consumo de bienes de primera necesidad es menor, de ser así la electricidad pasa a ser un 6% en los hogares de menor ingresos a un 1 % en los de mayor renta.

Otra característica que destacan Medina y Vicéns (2011) es que 1% de los hogares destinan el 10% de sus ingresos al gasto eléctrico, concretamente 202 euros al mes, mientras que los hogares que reciben menor ingreso un 1% de los mismos dedica el 25% de su renta al consumo de energía eléctrica, cifra que aunque se va disminuyendo según sube el nivel de ingreso continúa siendo elevada al situarse por encima del tramo de renta media.

4. METODOLOGÍA

4.1. Base de datos

Como ya hemos mencionado, la base de datos procederá de la Encuesta de Presupuestos Familiares (EPF), ejecutada por el (INE) en España. Debido a la especial relevancia que tiene en este estudio, resulta conveniente realizar algunas consideraciones en torno a ella.

La EPF ha evolucionado a lo largo de los años desde que empezó a realizarse hasta llegar a existir como la conocemos actualmente. Se ha desarrollado en aspectos como el tamaño muestra, el tipo de población considera, el nivel de desagregación del gasto, el sistema de recogida o el sistema de cuestionarios. Esta nueva EPF, entró en vigor en enero de 2006.

Como indica el Instituto Nacional de Estadísticas (INE), la Encuesta de Presupuestos Familiares (EPF en adelante) ofrece información sobre el gasto de los hogares españoles, así como la distribución del mismo entre las distintas parcelas de consumo. La EPF, suministra evaluación anual, en la base de datos a estudiar 20.229 viviendas, con ello da a conocer información imprescindible para las estimaciones del gasto en consumo de los hogares de la Contabilidad Nacional y la actualización del Índice de Precios de Consumo (IPC). Los gastos de consumo se registran como flujo monetario que destina el hogar al pago de determinados bienes y servicios de consumo final, como al valor de determinados consumos no monetarios efectuados por los hogares. (INE, s.f.)

La encuesta está formada por una serie de preguntas relacionadas con los gastos familiares en el hogar, concretamente por 12 grupos de gastos que a su vez se desagregan en 225 partidas, dicha clasificación para recoger los datos se denomina European Classification of Individual Consumption by Purpose, o también ECOICOP.

Tabla 4.1. Clasificación ECOICOP.

1. Alimentos y bebidas no alcohólicas
2. Bebidas alcohólicas, tabaco y narcóticos
3. Artículos de vestir y calzado
4. Vivienda, agua, electricidad, gas y otros combustibles
5. Mobiliario, equipamiento del hogar y gastos corrientes de conservación de la vivienda
6. Salud
7. Transportes
8. Comunicaciones
9. Ocio, espectáculos y cultura
10. Enseñanza
11. Hoteles, cafés y restaurantes
12. Otro bienes y servicios

Fuente: (INE, 2012, pág. 9)

A diferencia de las encuestas realizadas con base 1997, las familias españolas que participan ofrecen la información imprescindible durante dos semanas consecutivas del año de estudio para poder hacer la estimación y para evitar así la posibilidad de escoger una semana en la que el gasto en las familias no sea el habitual. Deben contestar la encuesta en base a un listado de códigos que les facilita el INE. La identidad de los usuarios interrogados no puede ser publicada, pero se notifican de forma agregada y desagregada. Cualquier usuario puede conocer los datos ya que es de acceso libre.

Según el INE, el gasto en consumo final de los hogares se registra a precios de adquisición, es decir el precio que el comprador debería de pagar en el momento de la compra y según su precio al contado.

De la tabla ECOICOP se ha usados los datos del grupo 4 correspondiente a "Electricidad".

4.2. Contenido de los ficheros

Se va a trabajar con los microdatos que cede la Encuesta de Presupuestos Familiares de forma gratuita. Vienen en un fichero plano, por lo que necesitamos elaborar una serie de diccionarios en el software para extraer los datos, el programa que se va a usar es STATA16. Todo el proceso se presenta en el anexo de este trabajo.

La EPF está formada por tres ficheros, fichero de hogar, fichero de los miembros del hogar y fichero de gastos. El fichero de gastos ofrece datos sobre gastos de consumo y características relativas a las condiciones de vida de las viviendas, pero no reúne los precios de los bienes consumidos declarados por el hogar. Para dicho estudio se va a hacer uso del fichero de gastos y el fichero hogar.

En el fichero hogar nos ayuda a conocer las características de la vivienda, contiene tantos registros como hogares de la muestra que hayan participado en la encuesta. Cada registro o cada hogar se le determina mediante la variable número secuencial “**NÚMERO**”. (INE, 2021).

Por otra parte, el fichero de gastos se hace una clasificación cada año para todos los hogares de la muestra con gasto. Hay tantos registros como tipo de gastos diferentes que hayan tenido. Cada registro se reconoce por medio de la concatenación del número secuencial del hogar “**NÚMERO**”, pero en este caso seguida de la variable “**CÓDIGO**”, con esto hace referencia al código correspondiente al gasto (5 dígitos). Para el gasto de consumo eléctrico está definida por el código 04511. (INE, 2021).

4.3. Variables

Teniendo en cuenta que el objetivo de la investigación es analizar los patrones del consumo eléctrico, se introducen variables de consumo eléctrico presentadas en la revisión de la literatura.

Como ya hemos mencionado antes la Encuesta de Presupuestos Familiares permite conocer el gasto en consumo, así como las características de los hogares residentes en España, pero no suministra información sobre el precio de los bienes consumidos declarados por el hogar.

Se considera como variable dependiente el “GASTO”, que identifica la cantidad destinada al gasto en la partida clave.

Para explicar los distintos patrones de comportamiento que registra el gasto eléctrico de los hogares se utilizan variables independientes, factores socio-económicos de los miembros del hogar, características de la vivienda, y geografía o ubicación de la misma. Entre los factores socio-económicos que explican el gasto eléctrico de un hogar se encuentran el número de miembros y los ingresos mensuales del hogar.

Las características de la vivienda utilizadas para explicar el consumo eléctrico del hogar son:

- El tamaño de la vivienda en m²: año de construcción de la vivienda, distinguiendo si se ha construido en menos de 25 años o más de 25 años.
- Zona de residencia: urbana de lujo, urbana alta, urbana media, urbana inferior, rural industrial, rural pesquera, rural agraria.
- Tipo de hogar: chalé, casa media, casa económica.
- Geografía de la vivienda dentro del territorio nacional (CCAA). Variable que servirá para explicar aquellos factores no observados relacionados con la ubicación geográfica, como el clima que afecta tanto al uso de calefacción como de equipos de aire acondicionados.
- Calefacción eléctrica del hogar: para medir si usan o no la electricidad como fuente de energía.

Como resumen de las variables a investigar, este estudio utiliza variables continuas y categóricas, dos de ellas recodificadas con el programa STATA. La información para cada parámetro de datos usados se detalla a continuación en la tabla 4.2.

Tabla 4.2. Descripción de variables

Fichero de hogar	Naturaleza de la variable	Descripción de la variable	Valores
GASTO_ELECTRICO	Dependiente continua	Importe total del gasto del hogar en consumo eléctrico.	Gasto total del hogar. Creada a partir de la unión del fichero hogar y el fichero gasto.
CCAA	Independiente categórica	Comunidad autónoma de residencia	1 Andalucía 2 Aragón 3 Asturias 4 Baleares 5 Canarias 6 Cantabria 7 Castilla y León 8 Castilla – La Mancha 9 Cataluña 10 Comunidad Valenciana 11 Extremadura 12 Galicia 13 Madrid 14 Murcia 15 Navarra 16 País Vasco 17 La Rioja 18 Ceuta 19 Melilla
NM_HOGAR	Independiente categórica	Número de miembros del hogar	1-20
ZONA_RESIDENCIAL	Independiente categórica	Tipo de zona de residencia	1 Urbana de lujo 2 Urbana alta 3 Urbana media 4 Urbana inferior 5 Rural industrial 6 Rural pesquera 7 Rural agraria . No consta
TIPO_HOGAR	Independiente categórica	Tipo de casa	1 Chalé o casa grande 2 Casa media 3 Casa económica o alojamiento . No consta
ANTIGUEDAD	Independiente categórica	Fecha construcción edificio	1 Hace menos de 25 años 0 Hace 25 o más años . No consta Recodificada a partir de la variable original "ANNOCON"
CALEFACCIÓN	Independiente categórica.	Disposición de calefacción	1 No tener calefacción 0 Si tener calefacción . No consta Recodificada a partir de la variable original "CALEF"
INGRESO_MENSUAL_HOGAR	Independiente categórica	Importe exacto de los ingresos mensuales netos totales del hogar	0-99999 Importe
SUPERFICIE	Independiente continua	Superficie útil de la vivienda	35 metros o menos 36-299 Metros 300 metros o más . No consta

Fuente: Elaboración propia a partir de EPF del INE.

Finalmente mencionar que en dicho trabajo vamos a estudiar las mismas variables que el estudio de Medina y Vicéns (2011).

5. MODELO TEÓRICO

Para el análisis de la demanda de energía eléctrica vamos a estimar el modelo de regresión lineal múltiple a través de Mínimos Cuadrados Ordinarios (MCO), su post-estimación y regresión cuantílica.

Para una mejor interpretación del modelo se han tomado logaritmos en las variables continuas, tanto dependientes como independientes. Esto permitirá interpretar estos coeficientes en términos de elasticidades.

Donde

Y_t es la variable dependiente, explicada o endógena.

X_t es la variable independiente, explicativa o exógena.

Por lo que la función del modelo de regresión lineal múltiple es:

$$Y: \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_p X_p + \varepsilon_i$$

- β_0 es el término independiente. Es el valor esperado de Y cuando X_1, \dots, X_p son cero.
- $\beta_1, \beta_2, \dots, \beta_p$ son los coeficientes parciales de la regresión:
 - β_1 mide el cambio en Y por cada cambio unitario en X_1 , manteniendo X_2, X_3, \dots, X_p constantes.
 - β_2 mide el cambio en Y por cada cambio unitario en X_2 , manteniendo X_1, X_3, \dots, X_p constantes.
 - ...
 - β_p mide el cambio en Y por cada cambio unitario en X_p , manteniendo X_1, \dots, X_{p-1} constantes.
- ε es el error de observación debido a variables no controladas

Suponiendo que se cumple el modelo de regresión lineal múltiple, estamos interesados en determinar qué variables X_j son significativas para explicar la variable respuesta Y .

$$H_0 : \beta_j = 0 \text{ (} X_j \text{ no influye sobre } Y \text{)}$$

$$H_1 : \beta_j \neq 0 \text{ (} X_j \text{ influye sobre } Y \text{)}$$

El modelo de regresión lineal múltiple propuesto en este trabajo busca determinar los patrones de consumo de energía eléctrica en hogares españoles. El modelo a estimar toma una forma logarítmica del tipo:

$$\begin{aligned} \text{Log (gasto eléctrico)}_i : & \beta_0 + \beta_1 \log (n^\circ \text{ miembros})_i + \beta_2 \log (\text{ingreso})_i + \beta_3 \log \\ & (\text{superficie})_i + \beta_4 \text{ antigüedad}_i + \beta_5 \text{ zona de residencia}_i + \beta_6 \text{ tipo de hogar}_i + \beta_7 \\ & \text{calefacción}_i + \beta_8 \text{ comunidad autónoma}_i + \varepsilon_i \end{aligned}$$

El método de regresión cuantílica es un tipo de análisis menos robusto que el MCO debido a la existencia de heterocedasticidad, cambios estructuras y datos atípicos. Este método se basa en la minimización de desviaciones absolutas ponderadas con pesos asimétricos sin estar dañado por datos extremos.

Una vez creado el modelo, nos disponemos a estudiar su post-estimación:

- **Homocedasticidad:** Los errores de estimación son constantes a lo largo de las observaciones, usamos el comando *rvfplotK* para verlo gráficamente y el Test de Breusch-Pagan y White. Es lo contrario a heterocedasticidad
- **Normalidad de los residuos:** Test de Shapiro-Wilks y gráfico Kdensity para ver la similitud de los residuos con una normal. Comando *swilk* y *kdensity*.
- **Multicolinealidad:** Para observar si existe multicolinealidad entre los regresores, es decir, si es un regresor es linealmente dependiente de otro. Si existe, se puede solucionar omitiendo la variable o las variables que causen problemas. El comando de Stata para contrastar la multicolinealidad es *vif* (variance inflation factor). Un $vif > 10$ o $1/vif < 0.10$ indica problemas. Para ello tras la regresión tendremos que escribir *vif*.

- **Variables omitidas:** Contraste *ovtest*, nos permite saber si hemos introducido o no todos los regresores necesarios. Contrastar el posible sesgo de variables omitidas es importante ya que uno de los supuestos es que el término de error y las variables explicativas no están correlacionadas ($E(e|X) = 0$). Se hace con el comando ***ovtest***.
- **Error de especificación:** a través del test *linktest*, para contrastar si necesitamos más variables en el modelo. El comando es ***linktest***.
- **Presencia de anómalos**
- **Test de significatividad conjunta:** Contrasta si dos o más coeficientes son conjuntamente distintos de cero. Usamos el comando ***test***.

6. RESULTADOS

En este capítulo se expone los resultados del modelo detallado en el apartado anterior. La tabla 6.1. representa los resultados de la estimación MCO y regresión cuantílica.

Como podemos observar en la tabla 6.1, la primera columna recoge los datos de la estimación de mínimos cuadrados ordinarios, sin embargo, las otras tres columnas indican los datos de los cuantiles 25, 50 y 75 de la distribución normal.

El p-value nos indica el nivel de significatividad de las variables de forma que: (***) p-value < 0,01; (**) p-value < 0.05; (*) p-value < 0.1, siendo el nivel de significatividad mayor el que contiene (***)).

Para medir la bondad conjunta del modelo en la regresión cuantílica (Pseudo R^2) se calcula con siguiente, siendo y_θ el cuantil analizado:

$$Pseudo R^2 = 1 - \frac{\sum_{i=1}^n |y_i - \hat{y}_i|}{\sum_{i=1}^n |y_i - \hat{y}_\theta|}$$

En el estudio de Mínimos Cuadrados Ordinarios, el Pseudo R^2 hace referencia al porcentaje de la varianza de la variable explicada, es decir, gasto eléctrico. El modelo explica el 17,72% de la varianza de la dependiente.

En primer lugar, las variables socio-económicas, superficie, tipo de vivienda y casi todas las variables geográficas (exceptuando Cataluña, Murcia, y Melilla) alcanzan parámetros estadísticamente significativos (***). Los parámetros de las variables antigüedad, zona de residencia y calefacción muestran ser no significativas, pero en el estudio de Medina y Vicens (2011) con datos del año 2009, dichas variables si son totalmente significativas, sin incluir zona de residencia que finalmente no incluyeron en el modelo.

El tamaño de la vivienda (m^2) y del hogar (número de miembros) son las variables que más influyen en el gasto de energía eléctrica, con una elasticidad de 0,24% y 0,19%. Las variables que menos influyen en la factura eléctrica es el año de construcción de la vivienda, zona de residencia y la calefacción.

Con respecto a las variables de las Comunidades Autónomas se observan diferencias en el consumo eléctrico. Andalucía ha sido la variable de referencia. En general los resultados nos permiten concluir que solo Murcia tienen más gasto eléctrico que Andalucía, concretamente un 3%. Sin embargo, el resto de comunidades es inferior entre un 2% y un 40%.

El ingreso del hogar tiene poca importancia en los razonamientos de las facturas eléctricas, toma un valor del 11%, similar al estudio de Medina y Vicens (16%). Además, Medina y Vicens indican que son valores inferiores a otras regiones como Dinamarca (28%), Noruega (22%) o Reino Unido (23%).

En el caso de la variable zona de residencia, la variable como referencia es urbana de lujo. El consumo eléctrico en residencias rurales es inferior al realizado en residencias urbanas, un 15% pesquera y 13% agraria.

En las estimaciones de la regresión cuantílica en el caso de antigüedad e ingreso podemos deducir que no tienen diferencias estadísticas significativas en los tres parámetros investigados.

Tabla 6.1. Resultados de la estimación MCO y regresión cuantílica

GASTO ELÉCTRICO												
	MCO			Q25			Q50			Q75		
	Parámetro	IC (95%)		Parámetro	IC (95%)		Parámetro	IC (95%)		Parámetro	IC (95%)	
_cons	5,13 ***	4,87	5,39	4,73 ***	4,41	5,05	5,26 ***	4,97	5,55	5,72 ***	5,39	6,05
Características socio-económicas												
log (Nº de miembros)	0,24 ***	0,22	0,25	0,24 ***	0,22	0,26	0,25 ***	0,23	0,26	0,24 ***	0,22	0,26
log (Ingresos)	0,11 ***	0,09	0,12	0,10***	0,08	0,12	0,07 ***	0,06	0,09	0,08 ***	0,06	0,10
Características de la vivienda												
log (Superficie (m2))	0.19 ***	0,16	0,21	0.203***	0,17	0,23	0,02 ***	0,01	0,23	0,19 ***	0,16	0,22
Antigüedad												
Menos de 25 años	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref
Más de 25 años	-0,01	-0,02	0,00	-0,00	-0,02	0,01	-0,02 *	-0,04	-0,00	-0,02	-0,04	-0,00
Zona de residencia												
Urbana de lujo	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref
Urbana alta	-0,16	-0,37	0,04	-0,09	-0,34	0,15	-0,17	-0,40	0,05	-0,20	-0,46	0,05
Urbana media	-0,21 *	-0,41	-0,01	-0,15	-0,40	0,09	-0,22	-0,45	0,00	-0,24	-0,50	0,00
Urbana inferior	-0,18	-0,39	-0,01	-0,15	-0,41	0,10	-0,19	-0,42	0,04	-0,16	-0,43	0,10
Rural industrial	-0,13	-0,35	0,03	-0,04	-0,30	0,22	-0,14	-0,38	0,09	-0,20	-0,47	0,06
Rural pesquera	-0,15	-0,37	0,07	-0,06	-0,34	0,20	-0,14	-0,39	0,11	-0,22	-0,51	0,05
Rural agraria	-0,13	-0,34	0,06	-0,08	-0,33	0,16	-0,13	-0,36	0,09	-0,15	-0,41	0,10
Tipo de vivienda												
Chalé o casa grande	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref
Casa media	-0,07 ***	-0,1	-0,04	-0,03 *	-0,07	-0,00	-0,06 ***	-0,09	-0,03	-0,10 ***	-0,13	-0,06
Casa económica	-0,10 ***	-0,15	-0,05	-0,06 *	-0,12	-0,00	-0,07 *	-0,12	-0,01	-0,13 ***	-0,19	-0,06
Calefacción eléctrica												
No tener calefacción	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref
Si tener calefacción	0,02 *	0	0,048	0,01	-0,13	0,04	0,02 *	0,00	0,05	0,01	-0,01	0,04
Comunidad Autónoma												
Andalucía	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref
Aragón	-0,21 ***	-0,26	-0,17	-0,20 ***	-0,26	-0,14	-0,22 ***	-0,28	-0,17	-0,25 ***	-0,31	-0,19
Asturias	-0,12 ***	-0,16	-0,07	-0,11 ***	-0,17	-0,06	-0,11 ***	-0,16	-0,05	-0,12 ***	-0,17	-0,06
Baleares	0,08**	0,02	0,13	0,08 *	0,02	0,15	0,08 **	0,02	0,15	0,08	0,01	0,15
Canarias	-0,17 ***	-0,21	-0,13	-0,16 ***	-0,21	-0,11	-0,16 ***	-0,21	-0,12	-0,21 ***	-0,26	-0,16
Cantabria	-0,22 ***	-0,27	-0,17	-0,17 ***	-0,23	-0,11	-0,20 ***	-0,26	-0,15	-0,28 ***	-0,34	-0,22
Castilla León	-0,32 ***	-0,36	-0,28	-0,31 ***	-0,36	-0,26	-0,32 ***	-0,37	-0,27	-0,33 ***	-0,38	-0,28
Castilla La Mancha	-0,08 ***	-0,12	-0,03	-0,12 ***	-0,17	-0,06	-0,08 **	-0,13	-0,03	-0,05 *	-0,11	-0,00
Cataluña	-0,03	-0,07	0,00	-0,01	-0,06	0,02	-0,04	-0,08	0,00	-0,05 *	-0,10	-0,00
Comunidad Valenciana	-0,09 ***	-0,12	-0,05	-0,08 ***	-0,12	-0,03	-0,07 ***	-0,11	-0,03	-0,08 ***	-0,12	-0,03
Extremadura	-0,09 ***	-0,14	-0,05	-0,12 ***	-0,17	-0,06	-0,09 ***	-0,14	-0,04	-0,09 ***	-0,15	-0,04
Galicia	-0,07 ***	-0,11	-0,03	-0,12 ***	-0,17	-0,07	-0,08 ***	-0,13	-0,03	-0,07 **	-0,12	-0,02
Madrid	-0,19 ***	-0,23	-0,15	-0,17 ***	-0,22	-0,13	-0,18 ***	-0,22	-0,13	-0,21 ***	-0,26	-0,16
Murcia	0,03	-0,01	0,07	0,04	-0,00	0,10	0,04	-0,00	0,09	0,06 *	0,01	0,12
Navarra	-0,28 ***	-0,33	-0,23	-0,27 ***	-0,33	-0,20	-0,30 ***	-0,36	-0,25	-0,33	-0,39	-0,26
País Vasco	-0,33 ***	-0,37	-0,29	-0,27 ***	-0,31	-0,22	-0,31 ***	-0,36	-0,27	-0,36 ***	-0,41	-0,31
La Rioja	-0,36***	-0,41	-0,31	-0,30 ***	-0,37	-0,24	-0,36 ***	-0,41	-0,30	-0,41 ***	-0,48	-0,34
Ceuta	-0,17 **	-0,27	-0,06	-0,11	-0,23	0,01	-0,21 ***	-0,33	-0,09	-0,22 ***	-0,35	-0,08
Melilla	-0,02	-0,13	0,086	-0,01	-0,15	0,11	-0,05	-0,17	0,06	-0,05	-0,19	0,07
Bondad conjunta del modelo:												
	R = 0,17			Pseudo R2= 0,08			Pseudo R2= 0,10			Pseudo R2= 0,10		

Fuente: Elaboración propia a partir de los datos EPF del INE (2022).

Respecto a los resultados de la post-estimación de MCO.

- **Homocedasticidad**

Test de Breusch-Pagan y White:

H_0 : Los residuos son homocedásticos

H_1 : Los residuos no son homocedásticos

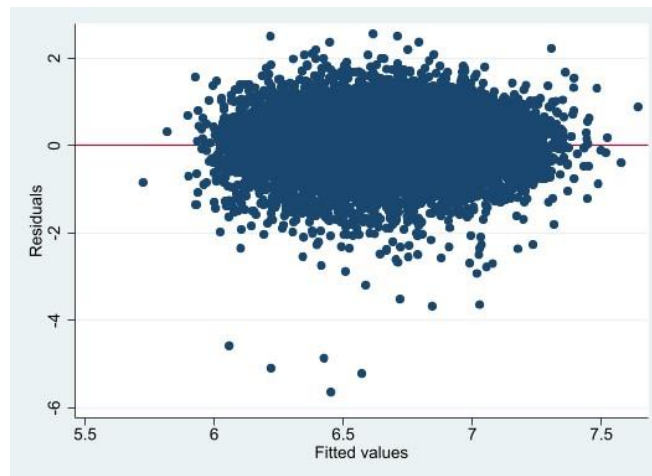
Tabla 6.2. Resultados obtenidos homocedasticidad MCO

H_0 : Constant variance
chi2(1) = 4.88
Prob > chi2 = 0.0272

Fuente: Elaboración propia a través del programa STATA.

Como P-value = 0.02, se rechaza la hipótesis nula, por tanto los residuos son heterocedásticos.

Figura 6.1. Resultados obtenidos gráficamente homocedasticidad MCO



Fuente: Elaboración propia a través del programa STATA

- **Normalidad de los residuos**

Contraste de Shapiro-Wilks:

H_0 : Los residuos siguen una distribución normal

H_1 : Los residuos NO siguen una distribución normal

Tabla 6.3. Resultados obtenidos contraste de normalidad

Shapiro-Wilk W test for normal data					
Variable	Obs	W	V	z	Prob>z
residuo	19,632	0.97559	215.007	14.620	0.00000

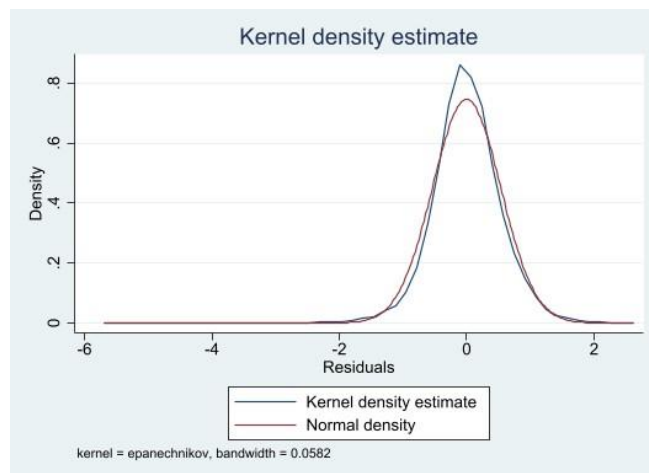
Note: The normal approximation to the sampling distribution of W' is valid for $4 < n <= 2000$.

Fuente: Elaboración propia a través del programa STATA

En este caso, se rechaza la hipótesis nula, por lo que no seguirían una normal.

Gráfico Kdensity: Gráfico utilizado para ver la similitud de los residuos con una normal. Si le añadimos la opción "normal" tendrá superpuesta la normal.

Figura 6.2. Resultados obtenidos gráficamente normalidad de los residuos



Fuente: Elaboración propia a través del programa STATA

- **Multicolinealidad**

Tabla 6.4. Resultados obtenidos de mutlicolinealidad

. vif		
Variable	VIF	1/VIF
_IZONA_RES~3	135.17	0.007398
_IZONA_RES~7	94.47	0.010586
_IZONA_RES~2	34.64	0.028868
_IZONA_RES~4	17.11	0.058437
_IZONA_RES~5	10.36	0.096509
_IZONA_RES~6	5.72	0.174706
_ICCAA_16	2.40	0.416675
_ICCAA_13	2.16	0.462895
_ICALEFACC~1	2.11	0.474887
_ITIPO_HOG~2	1.94	0.515595
_ICCAA_7	1.94	0.516008
_ICCAA_9	1.91	0.522364
_ICCAA_12	1.77	0.565346
_ITIPO_HOG~3	1.76	0.568920
_ICCAA_8	1.67	0.600467
_ICCAA_10	1.66	0.601722
log_SUPERF~E	1.66	0.602955
_ICCAA_3	1.63	0.611678
_ICCAA_2	1.60	0.625934
log_INGRES~R	1.57	0.635711
_ICCAA_6	1.48	0.676500
_ICCAA_15	1.46	0.684763
_ICCAA_11	1.45	0.691698
_ICCAA_17	1.44	0.694500
_ICCAA_5	1.37	0.728772
log_NM_HOGAR	1.37	0.730313
_ICCAA_14	1.36	0.737441
_ICCAA_4	1.23	0.809912
_ICCAA_18	1.07	0.937642
_IANTIGUED~1	1.06	0.940935
_ICCAA_19	1.04	0.957974
Mean VIF	10.89	

Fuente: Elaboración propia a través del programa STATA

Si $vif > 10$ o $1/vif < 0.1$ quiere decir que hay problemas. En nuestro caso, $vif = 10,89$.

- **Variables omitidas**

El contraste ovtest tiene como hipótesis:

H_0 : el modelo NO tiene variables omitidas

H_1 : el modelo tiene variables omitidas

Tabla 6.5. Resultados obtenidos de variables omitidas

H0: Model has no omitted variables
F(3, 19597) = 1.07
Prob > F = 0.3621

Fuente: Elaboración propia a través del programa STATA

En este caso, se rechaza la hipótesis nula, esto quiere decir que al modelo le faltan variables que le ayudarían a mejorar.

- **Error de especificación**

Test linktest:

H₀: No hay error de especificación

H₁: Hay error de especificación.

Tabla 6.6. Resultados obtenidos error de especificación

. linktest						
Source	SS	df	MS	Number of obs	=	19,632
Model	1221.65268	2	610.826339	F(2, 19629)	=	2132.74
Residual	5621.82278	19,629	.286403932	Prob > F	=	0.0000
				R-squared	=	0.1785
				Adj R-squared	=	0.1784
Total	6843.47545	19,631	.348605545	Root MSE	=	.53517
log_GASTO_~0	Coefficient	Std. err.	t	P> t	[95% conf. interval]	
_hat	.6623042	.6107873	1.08	0.278	-.5348907	1.859499
_hatsq	.0253969	.0459206	0.55	0.580	-.0646115	.1154052
_cons	1.120984	2.029437	0.55	0.581	-2.856885	5.098853

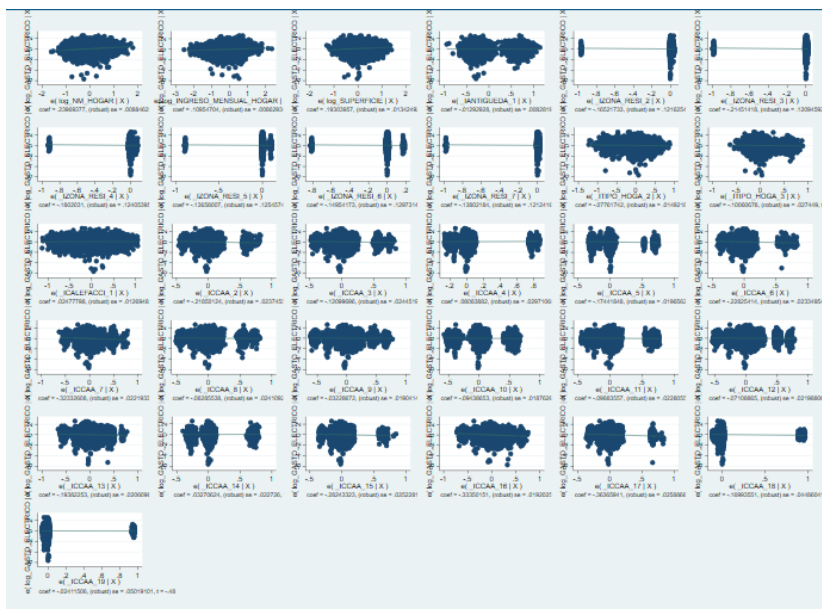
Fuente: Elaboración propia a través del programa STATA

Si se rechaza la hipótesis nula quiere decir que hay error de especificación y por tanto el modelo no estaría bien especificado. En nuestro caso pasa esto.

- **Presencia de anómalos**

Para chequear la presencia de datos anómalos usamos el comando avplot. Los outliers son puntos extremos que pueden tener efectos negativos en las estimaciones.

Figura 6.3. Resultados obtenidos gráficamente presencia de anómalos



Fuente: Elaboración propia a través del programa STATA

- **Test de significatividad conjunta**

Supongamos que queremos ver la significatividad conjunta de un par de variables, por ejemplo, INGRESO_MENSUAL_HOGAR NM_HOGAR. Para contrastar si dos o más coeficientes son distintos de cero usamos el comando test.

H_0 : Los coeficientes son conjuntamente cero

($\text{Beta_INGRESO_MENSUAL_HOGAR} = \text{Beta_NM_HOGAR} = 0$)

H_1 : Al menos uno de los coeficientes no es cero

Tabla 6.7. Resultados obtenidos significatividad conjunta

```

. test log_INGRESO_MENSUAL_HOGAR log_NM_HOGAR

( 1) log_INGRESO_MENSUAL_HOGAR = 0
( 2) log_NM_HOGAR = 0

      F( 2, 19600) = 735.88
      Prob > F = 0.0000

```

Fuente: Elaboración propia a través del programa STATA

Se rechaza la hipótesis nula, por lo que ambas variables tienen un efecto estadísticamente significativo sobre la variable dependiente GASTO_ELECTRICO.

7. CONCLUSIONES

En este trabajo se ha efectuado una primera aproximación al comportamiento del consumo eléctrico en los hogares españoles, creando y evaluando modelos econométricos como MCO, regresión cuantílica y su post-estimación a través de los datos proporcionados por la EPF del INE 2022. El objetivo general se basa en analizar el consumo de energía eléctrica en los hogares españoles a través de un modelo econométrico, y contrastar con el modelo econométrico obtenido por Medina y Vicens para el consumo eléctrico en el año 2009.

Una vez finalizado el estudio, observamos que coincide con muchos aspectos al modelo estudiado por Medina y Vicens. En primer lugar, los obstáculos encontrados a la hora de encontrar datos basados en mediciones reales en un sector tan importante como es la demanda eléctrica.

También, nos da la posibilidad de confirmar que el tamaño del número de miembros del hogar y la superficie de la vivienda en m^2 son las variables que más predominan en el gasto eléctrico en el hogar. Obteniendo elasticidades de 0,24% en el caso del tamaño del hogar medido a través del número de miembros, y 0,19% para la superficie de la vivienda. Frente al modelo de Medina y Vicens que obtiene un 0.31% y un 0.38% respectivamente.

Esta diferencia, menores elasticidades observadas en el año 2022, suponen un menor incremento del gasto en energía eléctrica con respecto al número de miembros en el hogar y la superficie del mismo, que pueden deberse a varios factores entre los que podríamos incluir una mayor eficiencia energética de los aparatos eléctricos de alto consumo en el hogar, responsables de la mayor parte del gasto energético del mismo (Trotta, Gram-Hanssen , & Lykke Jørgensen, 2020).

También podríamos apuntar que una menor elasticidad en relación con el número de miembros podría estar asociada a programas de subvenciones o bonificaciones en base al número de miembros de la familia, por ejemplo, bonificaciones en el precio de la electricidad en el caso de familias numerosas.

Por otro lado, una reducción de la elasticidad observada entre el primer estudio, con datos de 2009 (0.16%) y el presentado en este documento, que analiza datos de 2022 (0.11%) implica una demanda aún más rígida actualmente. Estas elasticidades tan reducidas se corresponden con bienes de primera necesidad.

Entre las limitaciones del presente estudio destaca la imposibilidad de trabajar con el consumo real energético, al disponer en la EPF del gasto monetario (PXQ) para cada una de las familias y no de la cantidad de energía consumida en el hogar (Q).

BIBLIOGRAFÍA

Agostini, C., Cecilia Plottier, M., & Saavedra, E. (2012). *ResearchGate*. Obtenido de https://www.researchgate.net/publication/46443774_La_Demanda_Residencial_por_Energia_Electrica_en_Chile

Castro Rodríguez, F. (2000). La demanda de electricidad de largo plazo. Principales determinantes en un marco regulado y repercusiones del proceso de liberalización. *Información Comercial Española Revista de Economía*, 783,2000, 119-131.

Considine, T. (2000). The impacts of weather variations on energy demand and carbon emissions. *Resource and Energy Economics*, 22, 295-314.

De la Torre Prados, I. (Septiembre-Diciembre de 2019). *Dialnet*. Obtenido de <https://dialnet.unirioja.es/servlet/articulo?codigo=7185491>

Dhar, A., Reddy, T., & Claridge, D. (1999). *Journal of Solar Energy Engineering*. Obtenido de http://auroenergy.com/wp-content/uploads/2016/05/1999_Dhar_ASME-JSEE_A-Fourier-Series-Model-to-Predict.pdf

Espey, J., & Espey, M. (2004). Turning on the lights: a meta-analysis of residential electricity demand elasticities. *Journal of Agricultural and Applied Economics*, 36 (1), 65-81.

Fernandez Villadangos, L. (2006). *Tesis doctoral, Universiad de Barcelona*. Obtenido de <http://hdl.handle.net/10803/2580>

Filippini, M., & Pachauri, S. (2004). *ScinceDirect*.

Hancevic, P., & Navajas, F. (2015). *Scielo*. Obtenido de https://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S2448-718X2015000400897

- Harris, J., & Mu Liu, L. (1993). Dynamic structural analysis and forecasting of residential electricity consumption. *International Journal of Forecasting*, 9, 437-455.
- Herriges, J., & King, K. (1995). Residential Demand for Electricity Under Inverted Block Rates: Evidence From a Controlled Experiment. *RAND Journal of Economics and Business Statistics*, 12, 419-430.
- Hyde, O., & Hodnett, P. (1997). An adaptable automated procedure for short-term electricity load forecasting. *IEEE Transactions on Power Systems*, 12(1), 84-93.
- INE. (s.f.). Obtenido de https://www.ine.es/prensa/epf_prensa.htm
- INE. (2010). Encuesta de Empleo del Tiempo 2009-2010. *Instituto Nacional de Estadística*, 4.
- INE. (2012). Encuesta de Presupuestos Familiares, Metodología. Obtenido de <https://www.ine.es/metodologia/t25/t2530p458.pdf>
- INE. (2021). Encuesta de presupuestos Familiares, Base 2006, Ficha de usuarios, Año 2021. Obtenido de https://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica_C&cid=1254736176806&menu=resultados&secc=1254736195147&idp=1254735976608#!tabs-1254736195147
- Instituto para la Diversificación y Ahorro de la Energía*. (s.f.). Obtenido de https://www.idae.es/uploads/documentos/documentos_Documentacion_Basica_Residencial_Unido_c93da537.pdf
- Jestin-Fleury, N. (1994). International Energy Agency. *World Energy Outlook*.
- Marrero, L., García-Santander, L., Carrizo, D., & Ulloa-Vásquez, F. (2021). Uso de algoritmo K-means para clasificar perfiles de clientes con datos. *Revista chilena de ingeniería*, vol. 29 N° 4, 778-787.

- Medina, E., & Vicéns, J. (2011). Factores determinantes de la demanda eléctrica de los hogares en España: Una aproximación mediante regresión cuantílica. *Estudios de Economía Aplicada*, 29-2, 515-538.
- Peña Sánchez de Rivera, J. (1988). Demanda de electricidad y precios en series temporales españolas. *Información Comercial Española Revista de Economía*, 663, 59-73.
- Red Eléctrica de España. (s.f.). Obtenido de <https://www.ree.es/es/actividades/demanda-y-produccion-en-tiempo-real>
- Suárez, Y. R. (2022). *Aplicación de técnicas de machine learning para la detección de patrones de consumo energético*.
- Tarifaluzhora by Selectra. (7 de febrero de 2023). Obtenido de Tarifaluzhora by Selectra: <https://tarifaluzhora.es/info/calcular-consumo-electrico-casa>
- Trotta, G., Gram-Hanssen, K., & Lykke Jørgensen, P. (2020). Heterogeneity of Electricity Consumption Patterns in. *Energies*.
- Valor, E., Caselles Costa, V., Pardo Tornero, A., & Meneu Ferrer, V. (2001). Consumo eléctrico y meteorología. *Revista Española de Física*, 15(4).
- Vicéns Otero, J. (2008). La demanda del sector residencial en España: Importancia del sector residencial y estrategias de gestión activa. *Club Español de la Energía*, 299-328.
- Westergren, K., & Norlen Urbano, H. (1999). Monitoring energy consumption in single-family houses. *Energy and Buildings*, 29, 247-257.

ANEXO

A continuación, se muestra, los comandos utilizado para la depuración de los datos y la transformación del modelo econométrico estimado con la ayuda de STATA.

*** Nos aseguramos de cerrar todo antes de empezar**

```
clear all
```

*** Abrimos primero la base de datos gastos, indicamos la ruta:**

```
use "C:\Users\F541\Desktop\xxx\MÁSTER EN ECONOMÍA FINANZAS Y  
COMPUTACIÓN\TFM\EPF Gastos en el hogar consumo electrico\Base de  
datos\datos_2022\EPFgastos\STATA\EPFgastos_2022.dta"
```

*** Luego abrimos la base de datos del hogar para saber las características y las observaciones, incicamos la ruta:**

```
use "C:\Users\F541\Desktop\xxx\MÁSTER EN ECONOMÍA FINANZAS Y  
COMPUTACIÓN\TFM\EPF Gastos en el hogar consumo electrico\Base de  
datos\BasesDeDatos_MercedesBlázquez_TFM\EPFhogar_2022.dta"
```

*** Unimos ambas bases de datos**

```
merge 1:m NUMERO using "C:\Users\F541\Desktop\xxx\MÁSTER EN ECONOMÍA  
FINANZAS Y COMPUTACIÓN\TFM\EPF Gastos en el hogar consumo  
electrico\Base de datos\datos_2022\EPFgastos\STATA\EPFgastos_2022.dta"
```

*** Eliminamos los códigos que no queremos y nos quedamos solo con 04511 que corresponde con el gasto en electricidad.**

```
tab CODIGO
```

```
keep if (CODIGO=="04511")
```


*** Guardamos la nueva base de datos**

save "GASTO ELECTRICO EPF 2022"

gen NUEVA=1 if (NUMERO==NUMERO[_n+1])

replace NUEVA=2 if (NUMERO==NUMERO[_n-1])

replace NUEVA=3 if ((NUMERO!=NUMERO[_n+1]) & NUMERO!=NUMERO[_n-1])

gen GASTO_ELECTRICO=.

replace GASTO_ELECTRICO=(GASTO+GASTO[_n+1]) if NUEVA ==1

replace GASTO_ELECTRICO=GASTO if NUEVA ==3

*** Para asegurarnos que ya solo tenemos el código 04511**

tab CODIGO

*** Dividimos entre el factor poblacional**

replace GASTO_ELECTRICO= GASTO_ELECTRICO/FACTOR

*** ESTUDIO DESCRIPTIVO DE LOS DATOS**

sum

describe

*** MODIFICAMOS EL FORMATO DE LAS VARIABLES**

destring NMIEMB, replace // N° de miembros

destring IMPEXAC, replace // Ingresos mensuales

destring ANNOCON, replace // Antigüedad

destring ZONARES, replace // Zona residencial

destring TIPOCASA, replace // Tipo de hogar

destring CALEF, replace // Disponen de calefacción

destring CCAA, replace // Disponen de CCA

destring SUPERF, replace // Disponen de SUPERF

*** AJUSTE EN LOS DATOS**

*** INPEXAC: valor -9 se sustituye por "missing values".**

replace IMPEXAC=. if IMPEXAC== -9

*** ZONARES: valor -9 se sustituye por "missing values".**

replace ZONARES=. if ZONARES== -9

*** TIPOCASA valor -9 se sustituye por "missing values".**

replace TIPOCASA=. if TIPOCASA== -9

*** ANNOCON: valor 6 se sustituye por 0 y el valor -9 "missing values".**

replace ANNOCON =0 if ANNOCON ==6

replace ANNOCON=. if ANNOCON== -9

*** CALEF: valor 6 se sustituye por 0 y el valor -9 "missing values".**

replace CALEF =0 if CALEF ==6

replace CALEF =. if CALEF == -9

*** SUPERF: el valor -9 "missing values".**

replace SUPERF =. if SUPERF == -9

*** RENOMBRAMOS LAS VARIABLES**

rename NMIEMB NM_HOGAR

rename IMPEXAC INGRESO_MENSUAL_HOGAR

rename SUPERF SUPERFICIE

rename ANNOCON ANTIGUEDAD

rename ZONARES ZONA_RESIDENCIAL

rename TIPOCASA TIPO_HOGAR

rename CALEF CALEFACCION

*** RENOMBRAMOS LAS CATEGORÍAS**

*** Primero creamos una "variable de etiqueta" y luego se la asignamos a la variable que le corresponde**

label define CCAA_etiqueta 1 "Andalucía" 2 "Aragón" 3 "Asturias" 4 "Baleares" 5 "Canarias" 6 "Cantabria" 7 "Castilla y León" 8 "Castilla – La Mancha" 9 "Cataluña" 10 "Comunitat Valenciana" 11 "Extremadura" 12 "Galicia" 13 "Madrid" 14 "Murcia" 15 "Navarra" 16 "País Vasco" 17 "La Rioja" 18 "Ceuta" 19 "Melilla" // **Etiquetas creadas**

label values CCAA CCAA_etiqueta // **Se las asignamos a la variable CCAA**

tab CCAA // **Comprobamos que están creadas y asignadas**

label define ANTIGUEDAD_etiqueta 1 "Hace menos de 25 años" 0 "Hace 25 ó más años"

label values ANTIGUEDAD ANTIGUEDAD_etiqueta

tab ANTIGUEDAD

label define ZONA_RESIDENCIAL_etiqueta 1 "Urbana de lujo" 2 "Urbana alta" 3 "Urbana media" 4 "Urbana inferior" 5 "Rural industrial" 6 "Rural pesquera" 7 "Rural agraria"

label values ZONA_RESIDENCIAL ZONA_RESIDENCIAL_etiqueta

tab ZONA_RESIDENCIAL

label define TIPO_HOGAR_etiqueta 1 "Chalé o casa grande" 2 "Casa media" 3 "Casa económica o alojamiento"

label values TIPO_HOGAR TIPO_HOGAR_etiqueta

tab TIPO_HOGAR

* REGRESIONES

* **Se transforman las variables a logaritmos para poder estimar la regresión:**

generate log_GASTO_ELECTRICO=ln(GASTO_ELECTRICO)

generate log_NM_HOGAR=ln(NM_HOGAR)

generate log_INGRESO_MENSUAL_HOGAR=ln(INGRESO_MENSUAL_HOGAR)

generate log_SUPERFICIE=ln(SUPERFICIE)

* **Regresión de forma logarítmica en MCO:**

```
regress          log_GASTO_ELECTRICO          log_NM_HOGAR
log_INGRESO_MENSUAL_HOGAR          log_SUPERFICIE  i.ANTIGUEDAD
i.ZONA_RESIDENCIAL i.TIPO_HOGAR i.CALEFACCION i.CCAA
```

eststo modelo1

esttab, r2 ar2 se scalar(rmse)

*** Estimaciones de Regresión por cuantiles en el cuantil 0.25**

```
qreg          log_GASTO_ELECTRICO          log_NM_HOGAR
log_INGRESO_MENSUAL_HOGAR          log_SUPERFICIE          i.ANTIGUEDAD
i.ZONA_RESIDENCIAL i.TIPO_HOGAR i.CALEFACCION i.CCAA, q(.25)
```

eststo modelo2

esttab, r2 ar2 se scalar(rmse)

*** Estimaciones de Regresión por cuantiles en el cuantil 0.5**

```
qreg          log_GASTO_ELECTRICO          log_NM_HOGAR
log_INGRESO_MENSUAL_HOGAR          log_SUPERFICIE          i.ANTIGUEDAD
i.ZONA_RESIDENCIAL i.TIPO_HOGAR i.CALEFACCION i.CCAA, q(.5)
```

eststo modelo3

esttab, r2 ar2 se scalar(rmse)

*** Estimaciones de Regresión por cuantiles en el cuantil 0.75**

```
qreg          log_GASTO_ELECTRICO          log_NM_HOGAR
log_INGRESO_MENSUAL_HOGAR          log_SUPERFICIE          i.ANTIGUEDAD
i.ZONA_RESIDENCIAL i.TIPO_HOGAR i.CALEFACCION i.CCAA, q(.75)
```

eststo modelo4

esttab, r2 ar2 se scalar(rmse)

* POST-ESTIMACIÓN MCO

- **HOMOCEASTICIDAD**

1) Gráficamente:

```
xi:   quietly   regress   log_GASTO_ELECTRICO   log_NM_HOGAR
log_INGRESO_MENSUAL_HOGAR   log_SUPERFICIE   i.ANTIGUEDAD
i.ZONA_RESIDENCIAL i.TIPO_HOGAR i.CALEFACCION i.CCAA
```

```
predict rselfhat
```

```
rvfplot, yline(0)
```

2) Test de Breusch-Pagan y White

```
xi:   quietly   regress   log_GASTO_ELECTRICO   log_NM_HOGAR
log_INGRESO_MENSUAL_HOGAR   log_SUPERFICIE   i.ANTIGUEDAD
i.ZONA_RESIDENCIAL i.TIPO_HOGAR i.CALEFACCION i.CCAA
```

```
estat hettest
```

*** Para arreglar problema de heterocedasticidad, simplemente añadimos al modelo la opción de robust que hace que los residuos sean homocedásticos.**

```
xi:   quietly   regress   log_GASTO_ELECTRICO   log_NM_HOGAR
log_INGRESO_MENSUAL_HOGAR   log_SUPERFICIE   i.ANTIGUEDAD
i.ZONA_RESIDENCIAL i.TIPO_HOGAR i.CALEFACCION i.CCAA, robust
```

```
predict rselfhatt
```

```
rvfplot, yline(0)
```

- **NORMALIDAD**

```
xi: quietly regress log_GASTO_ELECTRICO log_NM_HOGAR
log_INGRESO_MENSUAL_HOGAR log_SUPERFICIE i.ANTIGUEDAD
i.ZONA_RESIDENCIAL i.TIPO_HOGAR i.CALEFACCION i.CCAA, robust
```

```
predict residuo, resid
```

*** Gráfico Kdensity: Gráfico utilizado para ver la similitud de los residuos con una normal. Si le añadimos la opción "normal" tendrá superpuesta la normal. También podemos hacerlo con el comando histogram**

```
kdensity residuo
```

```
kdensity residuo, normal
```

```
histogram residuo, kdensity normal
```

*** Contraste de Shapiro-Wilks:**

```
swilk residuo
```

- **MULTICOLINEALIDAD**

```
xi: quietly regress log_GASTO_ELECTRICO log_NM_HOGAR
log_INGRESO_MENSUAL_HOGAR log_SUPERFICIE i.ANTIGUEDAD
i.ZONA_RESIDENCIAL i.TIPO_HOGAR i.CALEFACCION i.CCAA, robust
```

```
vif
```

- **VARIABLES OMITIDAS**

* El contraste ovtest

```
xi: quietly regress log_GASTO_ELECTRICO log_NM_HOGAR
log_INGRESO_MENSUAL_HOGAR log_SUPERFICIE i.ANTIGUEDAD
i.ZONA_RESIDENCIAL i.TIPO_HOGAR i.CALEFACCION i.CCAA, robust
```

```
ovtest
```

- **ERROR DE ESPECIFICACIÓN**

- * **El test linktest**

```
xi: quietly regress log_GASTO_ELECTRICO log_NM_HOGAR
log_INGRESO_MENSUAL_HOGAR log_SUPERFICIE i.ANTIGUEDAD
i.ZONA_RESIDENCIAL i.TIPO_HOGAR i.CALEFACCION i.CCAA, robust
```

```
linktest
```

- **PRESENCIA DE ANÓMALOS**

```
xi: quietly regress log_GASTO_ELECTRICO log_NM_HOGAR
log_INGRESO_MENSUAL_HOGAR log_SUPERFICIE i.ANTIGUEDAD
i.ZONA_RESIDENCIAL i.TIPO_HOGAR i.CALEFACCION i.CCAA, robust
```

```
avplots
```

- **TEST DE SIGNIFICATIVIDAD CONJUNTA**

* Supongamos que queremos ver la significatividad conjunta de un par de variables, por ejemplo, `INGRESO_MENSUAL_HOGAR` `NM_HOGAR`. Para contrastar si dos o más coeficientes son distintos de cero usamos el comando **test**.

```
test log_INGRESO_MENSUAL_HOGAR log_NM_HOGAR
```

// FIN //