



Universidad de Valladolid

**Facultad de Ciencias
Económicas y Empresariales**

Trabajo de Fin de Grado

**Grado en Administración y Dirección
de Empresas**

Determinantes de la pobreza

Presentado por:

Estefanía Vegas Losada

Tutelado por:

M^a Dolores de Prada Moraga

Valladolid, 06 de Julio de 2017

RESUMEN

La pobreza, entendida como la situación de los individuos que supone la dificultad para acceder a los bienes y servicios básicos que garantizan una calidad de vida digna, es uno de los aspectos que mayor preocupación genera a nivel mundial. Si bien los índices de pobreza se han reducido considerablemente en los últimos años, la cantidad de población que aún vive en la pobreza extrema es inaceptable. A partir de la Encuesta de Condiciones de Vida del año 2015, en este trabajo se realiza un análisis econométrico de la pobreza aplicando un modelo de regresión logística binaria que permita conocer el efecto de determinadas variables del individuo sobre la condición de pobre. Los resultados del modelo se han obtenido mediante el programa informático Eviews, concluyendo que variables como la edad, la nacionalidad o el desempleo presentan una influencia importante sobre la pobreza de la población en España.

Palabras clave: pobreza, modelo Logit, condiciones de vida, econometría.

Clasificación JEL: C25, I32, C51

ABSTRACT

Poverty, understood as individuals' situation that implies the difficulty to access the basic goods and services that guarantee a decent life quality, is one of the aspects that cause greatest concern worldwide. Although poverty rates have been significantly reduced in the last years, the number of people still living in extreme poverty is unacceptable. Based on the Living Conditions Survey of the year 2015, this paper presents an econometric analysis of poverty using a binary logistic regression model that allows us to know the effect of certain individuals' variables over the condition of poor. The model's results were obtained through the Eviews software, concluding that variables such as age, nationality or unemployment have a significant influence on the population's poverty in Spain.

Key words: poverty, Logit model, life conditions, econometrics.

JEL Classification: C25, I32, C51

ÍNDICE DE CONTENIDO

1. INTRODUCCIÓN.....	7
2. LA POBREZA EN ESPAÑA. MEDIDAS DE POBREZA.....	8
2.1. Definición y enfoques de pobreza.....	8
2.2. Líneas de pobreza.....	10
2.2.1. Pobreza absoluta.....	11
2.2.2. Pobreza relativa.....	12
2.2.2.1. Medición de la pobreza relativa.....	12
3. METODOLOGÍA.....	13
3.1. Encuesta de Condiciones de Vida.....	13
3.2. Modelos de elección discreta. El modelo Logit.....	15
4. ESTIMACIÓN DEL MODELO.....	19
4.1. Elección y preparación de los datos.....	19
4.2. Variables del modelo.....	20
4.2.1. Variable dependiente.....	20
4.2.2. Variables independientes o explicativas.....	21
4.3. El modelo Logit estimado.....	23
5. INTERPRETACIÓN DE LOS RESULTADOS.....	25
5.1. Especificación y estimación del modelo.....	25
5.2. Análisis de la bondad de ajuste.....	27
5.3. Efecto marginal de las variables explicativas.....	28
5.4. Escenarios.....	31
5.5. Curvas de respuesta.....	33
6. CONCLUSIONES.....	35
7. BIBLIOGRAFÍA.....	37
ANEXO I. ESCENARIOS.....	39
ANEXO II. CURVAS DE RESPUESTA.....	40

ÍNDICE DE TABLAS

Tabla 3.1. Metodología sintetizada de la ECV.....	15
Tabla 5.1. Estimación del modelo Logit.....	25
Tabla 5.2. Aciertos y errores en la predicción.	28
Tabla 5.3. Efectos marginales estimados de las variables cualitativas.	30
Tabla 5.4. Escenarios de probabilidad	39

ÍNDICE DE GRÁFICOS

Gráfico 5.1. Curva de respuesta de la edad y el grado de urbanización.	40
Gráfico 5.2. Curva de respuesta de la edad en función del estado de salud. ..	40
Gráfico 5.3. Curva de respuesta de la edad en función del estado civil.	40
Gráfico 5.4. Curva de respuesta de la edad en función de la nacionalidad.....	41
Gráfico 5.5. Curva de respuesta de la edad en función de los estudios.....	41
Gráfico 5.6. Curva de respuesta de la edad en función de la situación laboral.	41

1. INTRODUCCIÓN.

Una de las cuestiones que ha adquirido especial relevancia en el nuevo contexto social y económico es la relación entre la evolución de las condiciones macroeconómicas y la tendencia de la pobreza. Gracias a la ayuda de diferentes elaboraciones y publicaciones de datos del Instituto Nacional de Estadística y otros organismos estadísticos españoles y europeos, se conoce que las etapas de mayor crecimiento en España se han correspondido con considerables reducciones de los indicadores de pobreza y desigualdad, pero los últimos años de recesión económica no han provocado lo mismo.

En 2013, el Banco Mundial publicaba en su informe anual que 1.220 millones de personas vivían con menos de 1,25 dólares diarios. Más recientemente, en el informe de 2016, las expectativas del Banco Mundial mejoraban, indicando que el número de personas que vivían en situación de pobreza extrema se situaba por debajo del 10% de la población mundial. Sin embargo, se sigue manteniendo que millones de personas viven con menos de 1,90 dólares al día. Aunque la mayor parte de estas personas residen en los países que conocemos como el “tercer mundo”, en los “países ricos” también existen millones de personas que viven en inferioridad de condiciones con respecto al nivel de vida medio de estos países, lo que se conoce como “cuarto mundo”.

En España, el aumento de las familias en riesgo de pobreza o exclusión social es una de las graves secuelas que la crisis ha dejado. Generalmente, asociamos la pobreza a la existencia de una renta insuficiente para acceder a los bienes y servicios básicos o esenciales para disponer de una calidad y desarrollo de vida dignos. Sin embargo, la pobreza es un fenómeno más complejo que se ve influenciado por multitud de factores y que puede ser interpretada desde diversos enfoques.

En este sentido, el estudio que a continuación se presenta puede resultar relevante para conocer los grupos de población con mayor probabilidad de vivir en situación de pobreza y, de esta forma, poder elaborar políticas económicas y sociales que ayuden a solventar este problema.

Dada la complejidad y trascendencia de este fenómeno, el presente trabajo tiene como objetivo realizar una estimación del efecto que determinadas variables del individuo como la edad o la nacionalidad, entre otras, tienen sobre la probabilidad de que dicho individuo sea, o no, pobre, tomando como umbral de pobreza el que fija actualmente el EUROSTAT, que se corresponde con el 60% de la mediana de la distribución de los ingresos por unidad de consumo. Las técnicas utilizadas en este estudio permitirán, también, conocer la probabilidad estimada de que un individuo determinado se clasifique en el grupo de pobres, en función de los valores que tomen las variables anteriores para el mismo.

Teniendo en cuenta las características de la variable pobreza, la estimación planteada en los objetivos se realizará a partir de un modelo de elección discreta. En concreto, se estimará un modelo Logit, cuyas características principales se desarrollan en el apartado tercero de este trabajo. Los resultados de la estimación se obtendrán utilizando el programa estadístico Eviews a partir de los datos de la Encuesta de Condiciones de Vida 2015, publicados en el INE.

El trabajo se estructura en varias secciones. En primer lugar, se explica el concepto de pobreza atendiendo a distintas fuentes de información y se define la medición de la misma de acuerdo con los organismos europeos. A continuación se hace referencia a la metodología utilizada, tanto a la fuente de datos analizada como a la técnica de estimación empleada. En tercer lugar, se procede a explicar detalladamente el proceso de construcción y estimación del modelo, así como la interpretación de los resultados obtenidos y, por último, se establece un conjunto de conclusiones sobre la consecución del objetivo inicialmente planteado con este trabajo.

2. LA POBREZA EN ESPAÑA. MEDIDAS DE POBREZA.

2.1. Definición y enfoques de pobreza.

La pobreza, según del Diccionario de la Lengua Española de la Real Academia Española, se define como una situación de “necesidad, estrechez, carestía y falta de lo necesario para el sustento de la vida”. En otros términos, es la situación que supone la dificultad de los individuos para acceder a los bienes y servicios

básicos que garantizan una calidad de vida digna, así como un adecuado desarrollo personal y social.

Junto a las ideas clásicas de pobreza, que se basan en el nivel de vida material y en una renta insuficiente como factores determinantes de la misma, son muchos los autores que han aportado sus propios enfoques acerca de este fenómeno, como aquellos que entienden la pobreza en términos de capacidades o en términos de exclusión social. Entre los más actuales, cabe hacer especial referencia a la figura de Amartya Sen, nombrado Premio Nobel en Ciencias Económicas en 1998, y a su obra *“Pobreza y hambruna: Un ensayo sobre el derecho y la privación”*, que se ha convertido en una referencia obligatoria en el análisis económico de la pobreza. Amartya Sen define la pobreza atendiendo a las capacidades, de tal manera que una persona adquiere la condición de pobre cuando presenta cierto grado de privación de sus capacidades básicas para insertarse en la sociedad, a través del ejercicio de su voluntad.

Ante la necesidad de evaluar la situación en la que se encuentran los distintos países con respecto a la pobreza y de realizar comparaciones entre ellos, el Banco Mundial ha mostrado un gran interés en establecer una perspectiva universal de pobreza. Este organismo define la pobreza como “la incapacidad para alcanzar un nivel de vida mínimo”.

Dada la gran cantidad de variables que pueden influir en el nivel de pobreza de un país, es difícil realizar una medición exacta del mismo, pudiéndose adoptar diferentes puntos de vista en función de los aspectos o variables más relevantes para el investigador social, del horizonte temporal tomado para el análisis, etc.

De forma general, en el estudio de la pobreza existen dos grandes enfoques: el que considera la pobreza desde el punto de vista objetivo, basado en las variables que resultan observables por el investigador tales como los ingresos o gastos de los hogares, y el enfoque que considera la pobreza en términos subjetivos, teniendo en cuenta las valoraciones que los propios individuos hacen de su situación económica y social. También puede analizarse la pobreza de forma estática, surgiendo así la pobreza transversal, o de forma dinámica si se analiza la llamada pobreza persistente.

Finalmente, existe el concepto de privación multidimensional o pobreza carencial que está relacionado con la imposibilidad de acceder a ciertos bienes y servicios considerados necesarios por la sociedad y cuya privación puede dar lugar a la falta de integración social o exclusión social.

Todos los planteamientos anteriores aportan información útil al estudio de la pobreza, siendo lo más recomendado por los expertos la combinación de todos ellos para obtener una visión lo más completa posible de la situación de pobreza del territorio o país analizado.

Atendiendo al enfoque objetivo, se diferencian dos aproximaciones al concepto de pobreza o líneas de pobreza: la pobreza absoluta y la pobreza relativa.

2.2. Líneas de pobreza.

De acuerdo con la metodología que presenta el Instituto Nacional de Estadística y otros organismos, como el Banco Mundial, dentro del enfoque objetivo se utilizan, como instrumentos de medida, las llamadas líneas de pobreza.

Las líneas de pobreza se expresan a través de valores, generalmente monetarios, que actúan como umbral para clasificar a las personas del ámbito geográfico objeto de estudio como pobres y no pobres, en función del valor que tome la variable de referencia representativa de la pobreza para cada una de esas personas.

Pueden construirse dos líneas de pobreza: las líneas de pobreza absoluta y las líneas de pobreza relativa, siendo estas últimas la que sigue el Instituto Nacional de Estadística para elaborar los indicadores¹ de pobreza a partir de la Encuesta de Condiciones de Vida.

¹ Con periodicidad anual, el Instituto Nacional de Estadística, a partir de los datos de la Encuesta de Condiciones de Vida, produce cuatro indicadores principales de pobreza: la renta media del hogar, el riesgo de pobreza, el indicador de carencia material y la tasa de riesgo de pobreza o exclusión social (Indicador AROPE).

2.2.1. Pobreza absoluta.

Para que una medida sea considerada línea de pobreza absoluta debe ser consistente a lo largo del tiempo y entre países, es decir, estas líneas se refieren al valor de los recursos mínimos necesarios por una persona para garantizar un mínimo bienestar con independencia de la época o país en que se encuentre. En definitiva, medir la pobreza en términos absolutos supone considerar que un individuo requiere la misma cantidad de recursos en cualquier parte del mundo para su supervivencia.

La línea absoluta comúnmente utilizada es la que establece que el valor de esos recursos mínimos necesarios se sitúa en un dólar per cápita al día, de tal forma que, cualquier persona que viva con menos de un dólar al día será considerada pobre sea cual sea el país en el que habite.

Además de la anterior, existen varias líneas de pobreza absoluta construidas por distintos autores a lo largo del tiempo.

En 1901, Rowntree definió un umbral de pobreza absoluta fijado en el valor monetario de una cesta de bienes y servicios imprescindibles para satisfacer las necesidades básicas de los hogares más otra cantidad monetaria destinada a cubrir otros gastos como alquileres o gasolina. De acuerdo con esta línea de pobreza, cualquier hogar cuyos ingresos sean inferiores a la suma de dichas cantidades será clasificado como pobre. Sin embargo, la medida de Rowntree no ha sido bien aceptada, pues muchos críticos consideran que los gastos añadidos a la cesta varían de unas sociedades a otras, de tal manera que su consideración a la hora de definir el umbral tendería más hacia una línea de pobreza relativa que a una de pobreza absoluta.

Otra línea de pobreza absoluta a destacar es la de Mollie Orshanski, actualmente aplicada en Estados Unidos con alguna modificación. Siguiendo esta medida, el umbral de pobreza se obtiene multiplicando el valor de la cesta de alimentos básicos por el inverso de la proporción que supone el gasto en alimentación de los hogares sobre el gasto total. Al igual que la línea de Rowntree, esta medida ha sido objeto de críticas, en este caso, por el hecho de que demostraciones

empíricas han afirmado que, cuando mejora el desarrollo económico de un país, la proporción del gasto en alimentación sobre el gasto total disminuye, conduciendo a esta línea hacia la pobreza relativa.

La conclusión que se extrae de todo lo anterior es la dificultad para elaborar una línea de pobreza absoluta válida para todos los países y consistente a lo largo del tiempo. Por este motivo, la utilización de estas medidas de pobreza es escasa en los países más desarrollados frente a su amplia aceptación en los países en vías de desarrollo.

2.2.2. Pobreza relativa.

Las líneas de pobreza relativa definen la pobreza de un individuo en relación a la situación del resto de personas de la sociedad. A diferencia de la pobreza absoluta, en términos relativos no existe un umbral de pobreza fijo y consistente en el tiempo y entre países. De esta forma, una misma persona, con los mismos recursos, puede ser considerada como no pobre en un país y como pobre en otro país más rico, puesto que el valor de clasificación entre pobres y el resto de individuos será mayor en el segundo país que en el primero.

2.2.2.1. *Medición de la pobreza relativa.*

Para medir la pobreza relativa, normalmente, se establece un nivel mínimo de una variable monetaria, de modo que cualquier persona que presente un valor inferior a dicho nivel será clasificada como pobre y viceversa. Las variables monetarias que habitualmente se utilizan para fijar este umbral son el ingreso y el gasto, aunque ambas presentan algunos inconvenientes cuando se utilizan para medir la pobreza relativa.

Por un lado, cuando se toma como variable de referencia el gasto de los hogares se utiliza una medida más estable, pues los hogares no tienden a modificar sus gastos habituales cuando se reducen puntualmente sus ingresos. Sin embargo, en ocasiones, el gasto no está directamente relacionado con los ingresos que percibe el hogar sino que depende de sus costumbres o hábitos de consumo, lo que no reflejaría la capacidad económica real de los hogares.

Por otra parte, el ingreso anual, aunque es un buen reflejo de la capacidad económica del hogar, puede ser muy variable en el tiempo, por la existencia de ingresos excepcionales, además de dejar a un lado otros activos que influyen en la riqueza en el hogar. Actualmente, en Europa se utiliza la variable ingreso para la producción de los indicadores y estadísticas de pobreza y exclusión social.

Una vez seleccionada la variable monetaria, para construir la línea de pobreza relativa se hace necesario definir la unidad de análisis, es decir, el elemento que se toma como base para calcular el ingreso. En este sentido, la estadística oficial de la Unión Europea muestra su preferencia por tomar como unidad de análisis el ingreso por unidad de consumo, que resulta de dividir el ingreso total del hogar entre el número de unidades de consumo del hogar. Estas unidades de consumo se calculan a través de escalas de equivalencia que tienen en cuenta factores como la existencia de unidades de consumo equivalentes o las economías de escala en los hogares. La escala utilizada por el EUROSTAT es la escala de la OCDE².

Finalmente, la línea de pobreza se fija en el 60% de la mediana de la distribución de los ingresos por unidad de consumo. Este umbral clasifica como pobres a todas las personas cuyos ingresos por unidad de consumo sean inferiores a ese valor.

3. METODOLOGÍA.

3.1. Encuesta de Condiciones de Vida.

El análisis de la pobreza que se pretende en esta investigación se realizará a partir de los datos de la Encuesta de Condiciones de Vida publicados por el INE que se corresponden con el año 2015. Por este motivo, se hace necesario mencionar alguno de los aspectos más destacables de esta encuesta.

La Encuesta de Condiciones de Vida (ECV) se viene realizando, anualmente, por el Instituto Nacional de Estadística desde el año 2004 con el objetivo de recopilar información sobre las condiciones de vida de la sociedad española.

² $N^{\circ} \text{ u.c} = 1 + (a-1) \times 0.5 + b \times 0.3$; donde a es el número de adultos y b es el número de menores.

Anteriormente, desde 1994, a nivel europeo se realizaba una encuesta armonizada en todos los países miembros, el Panel de Hogares de la Unión Europea, pero, dada la necesidad de mejorar la medición estadística de la pobreza y obtener datos con un alto grado de comparabilidad entre países, se diseñó una nueva fuente estadística (EU SILC), que en España se denominó como Encuesta de Condiciones de Vida.

Las principales variables sobre las que el INE pretende proporcionar información con este estudio estadístico se pueden agrupar en los siguientes conceptos:

- Ingresos de los hogares privados y su situación económica.
- Pobreza, privación, protección mínima e igualdad de trato.
- Evolución del mercado de trabajo.
- Jubilaciones, pensiones y sobre la situación socioeconómica de las personas mayores.
- Aspectos relacionados con la vivienda.
- Desarrollo regional y movimientos migratorios.
- Formación, educación y salud.

En cuanto a la recogida de los datos, la ECV se compone de tres cuestionarios básicos que garantizan obtener la máxima información posible. Concretamente, consta de una ficha del hogar, que recoge los datos básicos de los miembros del hogar, un cuestionario individual para cada miembro, así como un cuestionario que recopila información sobre el hogar. Las características metodológicas de la encuesta se sintetizan en el siguiente cuadro.

Tabla 3.1. Metodología sintetizada de la ECV.

Tipo de encuesta	Encuesta de periodicidad anual.
Periodo base	2013
Ámbito poblacional	Población que reside en viviendas familiares principales.
Periodo de referencia de los ingresos	El año natural anterior a la entrevista.
Tamaño muestral	Alrededor de 13.000 hogares y 35.000 personas.
Tipo de muestreo	Bietápico con estratificación en las unidades de primera etapa. Las unidades de primera etapa son las secciones censales y las de segunda etapa son las viviendas familiares principales.
Método de recogida	Principalmente mediante entrevista personal. Los datos relativos a ingresos del hogar se construyen combinando la información proporcionada por el informante con registros administrativos.

Fuente: Instituto Nacional de Estadística (INE).

3.2. Modelos de elección discreta. El modelo Logit.

El objetivo principal de este trabajo es conocer la influencia de determinadas variables, cuantitativas y cualitativas, sobre la probabilidad de que un individuo adquiera la condición de pobre. En este análisis se distinguen dos tipos de variables: las variables exógenas o independientes y la variable pobreza, que es la variable dependiente, cuyo comportamiento se explica en función del resto.

La pobreza es una variable cualitativa, o variable dummy, que puede tomar dos valores: 1 si el individuo es pobre y 0 si no lo es. Para explicar este tipo de variables en función de otras, en el análisis de datos económicos existen los llamados modelos de elección binaria:

- El modelo lineal de probabilidad
- Los modelos Logit y Probit

El modelo lineal de probabilidad, que se asemeja a un modelo de regresión lineal pero teniendo en cuenta las características de la variable dependiente, asume la siguiente función:

$$Y_i = X'_i\beta + \varepsilon_i \quad i = 1 \dots N; \text{ siendo } E(\varepsilon_i) = 0$$

Donde Y_i es una variable de Bernoulli:

$$Y_i = \begin{cases} 1 & p_i = p(Y_i = 1) \\ 0 & 1 - p_i = p(Y_i = 0) \end{cases}$$

Con esperanza y varianza:

$$E(Y_i) = 1 p_i + 0(1 - p_i) = p_i \quad \text{Var}(Y_i) = p_i(1 - p_i)$$

Por tanto, en este modelo, la probabilidad de que ocurra el suceso de la variable endógena, viene definida por:

$$p_i = E(Y_i) = E(X'_i\beta) + E(\varepsilon_i) = X'_i\beta$$

Y la probabilidad estimada:

$$\hat{p}_i = \widehat{E(Y_i)} = X'_i\hat{\beta}$$

Dada la forma lineal de este modelo, el método de estimación asociado será el de Mínimos Cuadrados Ordinarios. Sin embargo, este método presenta varios problemas que hacen inadecuado el modelo para estudiar el comportamiento de la variable dicotómica dependiente:

- Puesto que los únicos valores que puede tomar Y_i son 0 y 1, las perturbaciones también tomarán únicamente dos posibles valores y su distribución no será una distribución normal, provocando que los estimadores del modelo no sean eficientes.

$$\varepsilon_i = Y_i - X'_i\beta = \begin{cases} 1 - X'_i\beta \\ -X'_i\beta \end{cases}$$

- La varianza de las perturbaciones no se mantiene constante para todas las observaciones. El modelo presenta heteroscedasticidad.

$$\text{Var}(\varepsilon_i) = (1 - p_i)^2 p_i + (1 - p_i) p_i^2 = p_i(1 - p_i)$$

- El problema más importante de la estimación por MCO es que los valores ajustados pueden salirse del intervalo $[0,1]$, de forma que la interpretación

de los resultados como la probabilidad de que ocurra el suceso (que el individuo sea pobre, en este caso) no es válida.

Por tanto, para el análisis de regresión que se pretende, se ha descartado el modelo lineal de probabilidad. La solución a los problemas se encuentra en la utilización de modelos no lineales, es decir, los modelos Logit y Probit. En estos modelos, cualquier valor estimado pertenece al intervalo $[0,1]$ ya que se formulan mediante una función F del siguiente modo:

$$Y_i = F(X_i' \beta) + \varepsilon_i$$

Donde $0 \leq F(X_i' \beta) \leq 1$ y siendo:

- Y_i la variable dependiente dicotómica que, en este caso, será:

$$Y_i = \begin{cases} 1 & \text{si el individuo es pobre} \\ 0 & \text{si no lo es} \end{cases}$$

- X_i el conjunto de variables independientes que influyen en Y_i
- ε_i las perturbaciones, que recogen todos aquellos factores que influyen en Y_i no incluidos en el modelo.

La probabilidad de que Y_i tome los valores 0 y 1 será:

$$p(Y_i = 1) = E(Y_i) = p_i = F(X_i' \beta)$$

$$p(Y_i = 0) = (1 - p_i) = 1 - F(X_i' \beta)$$

Aunque ambos modelos parten de la misma idea básica y se estiman por el criterio de máxima verosimilitud, la diferencia entre Logit y Probit radica en el tipo de función que asignan a F . El modelo Probit asume una función de distribución normal de parámetros ($\mu = 0; \sigma = 1$), mientras que el modelo Logit toma como F una función de distribución logística de parámetros ($\alpha = 0; \beta = 1$) y desviación típica $\frac{\pi}{\sqrt{3}}$. Las características de este último son las que se resumen a continuación, y será el utilizado en este trabajo.

En la regresión logística la función de distribución $F(X_i' \beta)$ es:

$$F(X_i' \beta) = \frac{e^{X_i' \beta}}{1 + e^{X_i' \beta}}$$

Así, el modelo resultante será el siguiente:

$$Y_i = F(X_i' \beta) + \varepsilon_i = \frac{e^{X_i' \beta}}{1 + e^{X_i' \beta}} + \varepsilon_i$$

En consecuencia, las probabilidades de ocurrencia de Y_i resultan:

$$p(Y_i = 1) = E(Y_i) = p_i = F(X_i' \beta) = \frac{e^{X_i' \beta}}{1 + e^{X_i' \beta}}$$

$$p(Y_i = 0) = (1 - p_i) = 1 - F(X_i' \beta) = 1 - \frac{e^{X_i' \beta}}{1 + e^{X_i' \beta}}$$

Existen dos posibles métodos de estimación de un modelo Logit: la estimación máximo verosímil y la estimación por MCO no lineales, siendo la más utilizada la estimación de máxima verosimilitud porque proporciona estimadores con buenas propiedades asintóticas. El estimador máximo verosímil del modelo Logit es consistente, asintóticamente eficiente y asintóticamente normal.

La función de verosimilitud asociada a los modelos no lineales es:

$$L(\beta) = \prod_{i=1}^N f(Y_i) = \prod_{i=1}^N [F(X_i' \beta)]^{Y_i} [1 - F(X_i' \beta)]^{1-Y_i}$$

Entonces, el estimador máximo verosímil de β se obtiene maximizando el logaritmo neperiano de la función de verosimilitud.

$$\max_{\beta} \ln(L(\beta))$$

$$\ln(L(\beta)) = \sum_{i=1}^N [Y_i \ln(F(X_i' \beta)) + (1 - Y_i) \ln(1 - F(X_i' \beta))]$$

Una de las características particulares del Logit es que permite obtener, mediante la aplicación de logaritmos, una función lineal de las variables explicativas o independientes que facilite la estimación e interpretación del modelo.

$$\ln\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 X_{1i} + \dots + \beta_k X_{ki} = X_i' \beta$$

Donde $\frac{p_i}{1-p_i}$ se conoce como ODD ratio y expresa la probabilidad de que el individuo "i" pertenezca a una población ($Y_i = 1$) frente a que pertenezca a la otra ($Y_i = 0$).

4. ESTIMACIÓN DEL MODELO.

En este apartado se desarrolla una explicación del procedimiento llevado a cabo para estimar el modelo de regresión logística que nos ocupa.

4.1. Elección y preparación de los datos.

Como ya se ha mencionado en anteriormente, la fuente de datos utilizada es la Encuesta de Condiciones de Vida 2015 del Instituto Nacional de Estadística, que proporciona ficheros de datos de corte transversal y datos longitudinales, o de panel, de un gran número de variables que se consideran influyentes sobre la situación de pobreza para el conjunto de individuos entrevistados. En este estudio se trabajará con los datos de corte transversal, es decir, información referida a un ciclo determinado, correspondientes a una muestra³ de 26.906 individuos para el año 2015.

El INE permite la descarga de los microdatos transversales de la Encuesta de Condiciones de Vida desglosados en cuatro archivos en formato csv:

- Fichero de datos básicos del hogar (esudb15d)
- Fichero de datos básicos de la persona (esudb15r)
- Fichero de datos detallados del hogar (esudb15h)
- Fichero de datos detallados de los adultos (esudb15p)

Dada la complejidad de este formato, ha sido necesario realizar una serie de transformaciones para simplificar y organizar el conjunto de datos válido para estimar el modelo Logit en Eviews.

En primer lugar, se ha utilizado el programa informático SPSS para fusionar la información relativa a los hogares y a las personas, asignado a cada individuo los datos de cada variable correspondientes al hogar al que pertenecen.

³ Aunque los microdatos de la ECV proporcionan información de 32.380 individuos, la muestra finalmente analizada en este trabajo se corresponde con 26.906 observaciones por haberse eliminado algunos casos por falta de datos en algunas de las variables manejadas.

Finalmente, del conjunto de datos resultante se han seleccionado aquellas variables que se consideran más relevantes para explicar la situación de pobreza de un individuo, habiendo sido necesaria la recodificación de algunas de ellas para su posterior tratamiento en Eviews. La descripción de estas variables se detalla en el apartado siguiente.

4.2. Variables del modelo

La Encuesta de Condiciones de Vida del INE recoge información acerca de aproximadamente 300 variables del individuo y del hogar que permiten producir estadísticas sobre las condiciones de vida de la población española. Sin embargo, para el análisis que nos ocupa este trabajo se ha seleccionado un conjunto de variables más reducido, teniendo en cuenta aquellas que se consideran que afectan con mayor importancia a la pobreza. Estas serán las que formen parte del modelo de regresión logística elaborado.

4.2.1. Variable dependiente.

La variable dependiente es la población en riesgo de pobreza o exclusión social que, de acuerdo con la metodología utilizada por la Encuesta de Condiciones de Vida siguiendo los criterios marcados por el EUROSTAT, son todos aquellos individuos que se encuentran en alguna de las siguientes situaciones:

- En riesgo de pobreza: El EUROSTAT fija el umbral de pobreza en el 60% de la mediana de los ingresos por unidad de consumo del año anterior al de la encuesta. De esta forma, un individuo se encontrará en riesgo de pobreza cuando sus ingresos sean inferiores a la línea de pobreza.
- En carencia material severa: serán aquellas personas pertenecientes a hogares que presenten al menos cuatro de las siguientes características:
 - No puede permitirse ir de vacaciones al menos una semana al año.
 - No puede permitirse una comida de carne, pollo o pescado al menos cada dos días.
 - No puede permitirse mantener la vivienda con una temperatura adecuada.

- No tiene capacidad para afrontar gastos imprevistos.
 - Ha tenido retrasos en el pago de gastos relacionados con la vivienda principal o en compras a plazos en los últimos 12 meses.
 - No puede permitirse disponer de un automóvil.
 - No puede permitirse disponer de teléfono.
 - No puede permitirse disponer de un televisor en color.
 - No puede permitirse disponer de una lavadora.
- En hogares sin empleo o con baja intensidad en el empleo: también se considerarán personas en riesgo de pobreza o exclusión social aquellas que pertenezcan a hogares cuyos miembros en edad de trabajar estuvieron ocupados menos del 20% del total de su potencial de trabajo durante el año anterior al de la encuesta, es decir, cuando el ratio entre el número de meses que los miembros del hogar trabajaron en el año de referencia y el total de meses que podrían haber trabajado sea inferior al 20%.

Teniendo en cuenta estos indicadores, en los microdatos de la ECV el INE proporciona una variable dicotómica (vrEU2020) que toma el valor 1 cuando la persona se encuentra en riesgo de pobreza o exclusión social y valor 0 cuando no se encuentra en dicha situación. En el análisis de este trabajo esta variable adquiere el nombre de POBREZA y se define de la siguiente forma:

$$POBREZA_i = \begin{cases} 1 & \text{si el individuo está en riesgo de pobreza o exclusión social} \\ 0 & \text{si no lo está} \end{cases}$$

4.2.2. Variables independientes o explicativas.

Atendiendo al objetivo perseguido con este estudio, analizar el efecto que tienen determinadas variables sobre la probabilidad de un individuo de encontrarse en riesgo de pobreza, a partir de los datos de la ECV se han construido las siguientes variables que serán las que formen parte del modelo Logit.

- Edad: Es una variable cuantitativa que ha sido elaborada a partir de la variable RB080 de los microdatos de la ECV, que proporciona información

sobre el año de nacimiento de las personas entrevistadas. En el modelo se corresponde con la variable $EDAD_i$.

- Situación laboral: La situación laboral viene medida por la variable RB210 en los microdatos de la ECV y proporciona información acerca de la situación de los individuos de la muestra en la actividad durante la semana anterior a la de la encuesta en cuatro modalidades (trabajando, parado, jubilado o jubilado anticipadamente y otra clase de inactividad económica) Para el estudio se ha simplificado esta información a tres categorías (ocupados, parados e inactivos), generando dos variables ficticias que se incluirán en el modelo de la siguiente forma:

$$OCUP1_i = \begin{cases} 1 & \text{si el individuo está ocupado} \\ 0 & \text{en otro caso} \end{cases}$$

$$OCUP2_i = \begin{cases} 1 & \text{si el individuo está parado} \\ 0 & \text{en otro caso} \end{cases}$$

- Nacionalidad: Se trata de una variable cualitativa que se ha incluido en el modelo como una variable dicotómica de valores 0 y 1, siendo el valor 1 para los individuos españoles y 0 para los extranjeros.

$$NAC_i = \begin{cases} 1 & \text{si es español} \\ 0 & \text{si es extranjero} \end{cases}$$

- Grado de urbanización: Mide la cantidad de población existente en áreas urbanas con respecto a la población total del lugar en el que se reside el individuo. Se trata de una variable ficticia que presenta dos categorías y que se incluye en el modelo de la siguiente forma:

$$URB_i = \begin{cases} 1 & \text{si el individuo reside en zona poco o medio poblada} \\ 0 & \text{si reside en zona muy poblada} \end{cases}$$

- Salud: Esta variable cualitativa se refiere a si el individuo tiene alguna enfermedad o problema de salud crónicos.

$$SALUD_i = \begin{cases} 1 & \text{si tiene enfermedad crónica} \\ 0 & \text{si no la tiene} \end{cases}$$

- Estado Civil: Es una variable cualitativa que presenta dos modalidades (casado o no casado) y, por tanto, se incluye en el modelo mediante una variable dummy que adquiere el nombre de $ECIVIL_i$

$$ECIVIL_i = \begin{cases} 1 & \text{si el individuo está casado} \\ 0 & \text{si no lo está} \end{cases}$$

- Nivel de estudios: En la Encuesta de Condiciones de Vida, el nivel de estudios terminado por los individuos de la muestra viene medido por la variable PE040 teniendo en cuenta la clasificación CNED2014 y presenta 10 categorías. En este caso, se ha simplificado la variable, manteniendo dicha clasificación, a tres modalidades que se refieren a educación secundaria, educación superior y estudios inferiores. La categoría de educación secundaria abarca desde la segunda etapa de ese nivel hasta educación postsecundaria no superior y en la categoría de estudios inferiores se incluyen los individuos sin estudios, con educación primaria o con la primera etapa de educación secundaria. De esta forma, se han construido dos variables dicotómicas que se incluirán en el modelo Logit del siguiente modo:

$$ESEC_i = \begin{cases} 1 & \text{si nivel educativo es educación secundaria} \\ 0 & \text{en otro caso} \end{cases}$$

$$ESUP_i = \begin{cases} 1 & \text{si nivel educativo es estudios superiores} \\ 0 & \text{en otro caso} \end{cases}$$

Con respecto a esta selección de variables cabe mencionar que, evidentemente, la renta es una de las características más influyentes en la situación de pobreza de los individuos pero, dado que la variable dependiente está elaborada a partir de los ingresos por unidad de consumo no se ha incluido en el modelo por ser una variable que predice perfectamente la probabilidad de ser pobre.

4.3. El modelo Logit estimado.

Una vez definidas las variables que se incluirán en el análisis, se estima un modelo de regresión logística binaria en el que la probabilidad de un individuo de ser, o no, pobre se explica por su edad, sexo, estado civil, el nivel educativo, su situación en relación con la actividad, la nacionalidad y la renta.

Para ello, se ha utilizado el programa informático Eviews que posibilita realizar múltiples acciones sobre el modelo tales como el análisis de la bondad del ajuste de la estimación, de los efectos de cada variable exógena sobre la pobreza, así como permite realizar predicciones sobre la probabilidad de que un individuo con unas determinadas características de las variables independientes se encuentre en riesgo de pobreza o exclusión social.

La estimación del modelo y el análisis de los resultados se han realizado en varias fases.

- En primer lugar, se realiza la estimación propiamente dicha, obteniendo el valor de los coeficientes que acompañan a las variables independientes y cuyo signo determinará la relación existente entre las mismas y la variable dependiente.
- Antes de realizar cualquier interpretación del modelo, es necesario analizar la bondad del ajuste de la estimación, es decir, comprobar que se trata de un modelo adecuado donde las variables exógenas son las correctas para explicar el comportamiento de la pobreza. Para ello, además de evaluar la significación individual y conjunta de las variables explicativas, utilizaremos el criterio de aciertos y errores en la estimación.
- En tercer lugar, se estudiará el efecto marginal de los regresores en la probabilidad, que dependerá del resto de variables explicativas a través de la función de distribución logística. Aquí, hay que tener en cuenta que el procedimiento para obtener dicho efecto será distinto en las variables cuantitativas y en las discretas, por las características de uno y otro tipo.
- Finalmente, se plantearán posibles escenarios en los que el individuo adquiere determinados valores de las variables exógenas y se analizarán como cambia la probabilidad de ser pobre cuando se modifica alguno de esos valores manteniendo el resto constantes. Además, se elaborarán varias curvas de respuesta que permitirán analizar gráficamente el efecto de la edad sobre la probabilidad de ser pobre en función de las distintas variables dummy.

5. INTERPRETACIÓN DE LOS RESULTADOS.

5.1. Especificación y estimación del modelo

El punto de partida en el análisis de resultados será definir el modelo y estimarlo mediante la técnica elegida. En este caso, el modelo Logit que se ha construido se especifica de la siguiente forma:

$$Y_i = \frac{e^{\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \beta_5 X_{5i} + \beta_6 X_{6i} + \beta_7 X_{7i} + \beta_8 X_{8i} + \beta_9 X_{9i}}}{1 + e^{\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \beta_5 X_{5i} + \beta_6 X_{6i} + \beta_7 X_{7i} + \beta_8 X_{8i} + \beta_9 X_{9i}}} + \varepsilon_i$$

Siendo: Y_i la variable dependiente = *POBREZA*_{*i*}

$$\begin{array}{lll} X_{1i} = EDAD_i & X_{4i} = ESUP_i & X_{7i} = OCUP2_i \\ X_{2i} = NAC_i & X_{5i} = ECIVIL_i & X_{8i} = SALUD_i \\ X_{3i} = ESEC_i & X_{6i} = OCUP1_i & X_{9i} = URB_i \end{array}$$

Y los resultados de la estimación son los que se presentan a continuación.

Tabla 5.1. Estimación del modelo Logit

Dependent Variable: POBREZA
Method: ML - Binary Logit (Quadratic hill climbing)
Date: 06/23/17 Time: 10:40
Sample: 1 26906
Included observations: 26906
Convergence achieved after 5 iterations
Covariance matrix computed using second derivatives

Variable	Coefficient	Std. Error	z-Statistic	Prob.
EDAD	-0.025401	0.001022	-24.85552	0.0000
NAC	-1.284630	0.065990	-19.46703	0.0000
ESEC	-0.652096	0.040751	-16.00197	0.0000
ESUP	-1.309381	0.045138	-29.00825	0.0000
ECIVIL	-0.134205	0.033732	-3.978579	0.0001
OCUP1	-0.451371	0.038120	-11.84078	0.0000
OCUP2	1.360980	0.043858	31.03145	0.0000
SALUD	0.218221	0.035608	6.128365	0.0000
URB	0.170959	0.031112	5.494975	0.0000
C	1.675777	0.085728	19.54766	0.0000
McFadden R-squared	0.152649	Mean dependent var	0.261206	
S.D. dependent var	0.439300	S.E. of regression	0.398079	
Akaike info criterion	0.974035	Sum squared resid	4262.126	
Schwarz criterion	0.977083	Log likelihood	-13093.69	
Hannan-Quinn criter.	0.975018	Deviance	26187.39	
Restr. deviance	30905.00	Restr. log likelihood	-15452.50	
LR statistic	4717.610	Avg. log likelihood	-0.486646	
Prob(LR statistic)	0.000000			

Fuente: Elaboración propia a partir de los datos de la ECV del INE

De los coeficientes estimados se deduce la ecuación del modelo.

$$\hat{Y}_i = \frac{e^{1,68-0,03X_{1i}-1,28X_{2i}-0,65X_{3i}-1,3X_{4i}-0,13X_{5i}-0,45X_{6i}+1,36X_{7i}+0,218X_{8i}+0,17X_{9i}}}{1 + e^{1,68-0,03X_{1i}-1,28X_{2i}-0,65X_{3i}-1,3X_{4i}-0,13X_{5i}-0,45X_{6i}+1,36X_{7i}+0,218X_{8i}+0,17X_{9i}}}$$

Del modelo obtenido se pueden realizar varios análisis importantes. Por un lado, los signos que acompañan a los coeficientes del modelo indican la dirección en la que afecta la variable a la que están asociados a la probabilidad de encontrarse en riesgo de pobreza o exclusión social. Así, se puede observar que las variables edad, nacionalidad, educación secundaria y superior, estado civil y ocupación presentan un signo negativo, por lo que un aumento de la edad y poseer la característica frente a no tenerla, para el caso de las variables ficticias, disminuyen la probabilidad de ser pobre. Por ejemplo, un individuo con estudios superiores tendrá menor probabilidad de estar en riesgo de pobreza que otro individuo sin estudios. En cambio, habitar en una zona poco urbanizada, tener alguna enfermedad crónica y estar en paro frente a ser inactivo son variables que aumentan dicha probabilidad. En los modelos de elección discreta solo es posible realizar esta interpretación de los parámetros, siendo incorrecto medir el efecto en la probabilidad a través de la magnitud de los mismos.

Por otro lado, es importante analizar la significación de las variables, tanto individual como conjunta, pues esto ayudará a determinar si se trata de variables adecuadas para explicar la pobreza. Para medir la significación individual se plantean las siguientes hipótesis:

$$H_0: \beta_j = 0 \rightarrow \text{si no se rechaza, } X_j \text{ no será significativa}$$

$$H_1: \beta_j \neq 0 \rightarrow \text{si se acepta, } X_j \text{ será significativa}$$

En este caso, dado el valor de los estadísticos Z asociados a cada variable y el p-valor correspondiente que toma valor 0 para todas las variables, excepto para el estado civil que toma valor 0.0001, adoptando un nivel de significación del 5% se rechazan las hipótesis nulas para cada uno de los β_j y, por tanto, todas las variables del modelo son significativas para explicar el comportamiento de la pobreza.

Finalmente, para analizar la significación conjunta se plantean las siguientes hipótesis:

$$H_0: \beta_1 = \beta_2 = \beta_3 = \dots = \beta_9 = 0$$

$$H_1: \text{Algún } \beta_j \neq 0$$

Teniendo en cuenta el estadístico de la razón de verosimilitud $LR = 4717,610$ y el p-valor asociado de 0, para el nivel de significación del 5% se rechaza la hipótesis nula y, por tanto, la regresión es significativa conjuntamente para explicar la pobreza.

5.2. Análisis de la bondad de ajuste.

Para evaluar la bondad del ajuste de un modelo Logit puede emplearse el criterio de porcentaje de aciertos y errores, así como la interpretación del pseudo- R^2 de McFadden. Este último compara la verosimilitud del modelo que solo incluye término constante β_0 (L_R) y la verosimilitud del modelo con todos los parámetros estimados (L) de la siguiente forma:

$$R^2 McFadden = 1 - \frac{\ln L}{\ln L_R}$$

Como se puede observar en la tabla anterior (Tabla 5.1) el valor del R^2 de McFadden en el modelo construido es de 0,152649 lo que nos indicaría una buena capacidad explicativa del modelo aunque no puede asegurarse una interpretación directa de este valor.

En cuanto al criterio de aciertos y errores, consiste en una tabla de clasificación en la que se presenta el porcentaje de aciertos de la estimación en relación con los valores observados. Generalmente, se toma como umbral una probabilidad de 0,5, de tal manera que aquel individuo para el cual se estime una probabilidad \hat{p}_i mayor o igual a 0,5 se clasificará en el grupo de pobres ($Y_i = 1$). Los resultados se muestran en la siguiente tabla (Tabla 5.2).

Tabla 5.2. Aciertos y errores en la predicción.

Expectation-Prediction Evaluation for Binary Specification

Equation: EQ01

Date: 06/23/17 Time: 10:40

Success cutoff: C = 0.5

	Estimated Equation			Constant Probability		
	Dep=0	Dep=1	Total	Dep=0	Dep=1	Total
P(Dep=1)≤C	18538	4701	23239	19878	7028	26906
P(Dep=1)>C	1340	2327	3667	0	0	0
Total	19878	7028	26906	19878	7028	26906
Correct	18538	2327	20865	19878	0	19878
% Correct	93.26	33.11	77.55	100.00	0.00	73.88
% Incorrect	6.74	66.89	22.45	0.00	100.00	26.12
Total Gain*	-6.74	33.11	3.67			
Percent Gain**	NA	33.11	14.04			

Fuente: Elaboración propia a partir de los datos de la ECV.

Los resultados de este criterio indican que de las 26.906 observaciones de la muestra, el modelo construido predice bien 20.865, que se corresponde con el 77,55% del total. Concretamente, de los 7028 individuos que realmente tienen la condición de pobre, el modelo clasifica adecuadamente al 33,11%, mientras que de los que no se encuentran en riesgo de pobreza se clasifican correctamente el 93,26% de las observaciones. Se puede afirmar, por tanto, que el modelo predice una menor cantidad de individuos pobres de los que realmente existen en la muestra y, aunque no se trata de un modelo perfecto para explicar la pobreza, presenta una adecuada bondad de ajuste.

5.3. Efecto marginal de las variables explicativas.

Uno de los análisis que nos permite realizar el modelo construido, y que constituye uno de los objetivos principales de este trabajo, es el de evaluar el efecto que provoca cada una de las variables independientes sobre la probabilidad de estar en riesgo de pobreza o exclusión social. Esto se realiza a través del cálculo del efecto marginal de X_{ji} , que mide el cambio en la variable dependiente, en la probabilidad de ser pobre en este caso, cuando se produce un cambio en una unidad en la variable exógena que se esté evaluando.

En los modelos lineales, el valor de los parámetros que acompañan a los regresores mide el efecto marginal estimado, que permanece constante, pero en los no lineales, como el modelo Logit, el efecto parcial no se corresponde con el

valor de los coeficientes ni tampoco es constante, sino que viene definido de la siguiente forma:

$$\frac{d\widehat{p}_i}{dX_{i,j}} = \frac{dF(\widehat{X}_i'\beta)}{dX_{i,j}} = f(\widehat{X}_i'\beta)\widehat{\beta}_j$$

Siendo $f(\cdot)$, en este caso, la función de densidad de la distribución logística.

Dado que la función de densidad es siempre positiva, el signo del efecto de la variable explicativa dependerá del coeficiente β que le acompaña.

Cabe destacar que el cálculo del efecto anteriormente descrito únicamente se realizará para aquellas variables explicativas de carácter cuantitativo, la edad en este caso, debiendo utilizar otro método en el caso de las cualitativas que toman valores 0 y 1 para las que un cambio unitario de la variable no tiene mucho sentido. Este método consistirá en determinar la diferencia en la probabilidad estimada cuando la variable dummy toma valor 1 y la probabilidad estimada cuando toma el valor 0, manteniendo el resto de variables en la media.

Los resultados obtenidos a través de Eviews indican los siguientes efectos marginales de la variable edad sobre la variable dependiente:

$$\frac{d\widehat{p}_i}{dEDAD_i} = \frac{dF(\widehat{X}_i'\beta)}{dEDAD_i} = f(\widehat{X}_i'\beta)\widehat{\beta}_1 = -0,004011$$

Esto implica que un aumento unitario de la edad disminuye la probabilidad de estar en situación de pobreza o exclusión social en un 0,401%. El signo negativo del efecto viene dado por el coeficiente β asociado a la edad (-0,025401).

Para el caso de las variables cualitativas, los efectos marginales de cada una de ellas se recogen en la tabla que se presenta a continuación (Tabla 5.3).

Tabla 5.3. Efectos marginales estimados de las variables cualitativas.

Variables	Probabilidad estimada ($X_j = 1$)	Probabilidad estimada ($X_j = 0$)	Efecto marginal
Nacionalidad	0,216008	0,498887	-0,282879
E. Secundaria	0,196911	0,320034	-0,123123
E. Superior	0,112746	0,320034	-0,207288
Estado Civil	0,215990	0,239580	-0,02359
Ocupación 1	0,157288	0,226676	-0,069388
Ocupación 2	0,533401	0,226676	0,306725
Salud	0,251927	0,213059	0,038868
Urbanización	0,241208	0,211314	0,02985

Fuente: Elaboración propia a partir de los datos de la ECV.

Dado que en este tipo de variables ya se ha comentado que un aumento unitario no tiene demasiado sentido, la interpretación de estos efectos se refiere a la diferencia en la probabilidad de ser pobre cuando el individuo presenta la característica de la variable ficticia frente a otro individuo que no la presenta. Así, por ejemplo, para el caso de la variable ocupación, la probabilidad estimada de estar en riesgo de pobreza o exclusión social de un individuo que se encuentre trabajando disminuirá en un 6,93% con respecto a dicha probabilidad de un individuo en situación de inactividad, mientras que para una persona parada la probabilidad de ser pobre aumentará en un 30,67% frente a un inactivo.

De igual modo ocurre para el resto de variables. Según la variable nacionalidad, la probabilidad de adquirir la condición de pobre de un individuo español frente a otro extranjero será menor, cuantificando esa diferencia en un 28,2879%. Atendiendo al estado de salud, tener alguna enfermedad crónica aumentará la probabilidad de riesgo de pobreza en un 3,8868%, igual que aumentará en un 2,985% si el individuo habita en una zona poco o medio poblada. El estado civil, por su parte, afectará positivamente a la situación de pobreza, disminuyendo la probabilidad de ser pobre cuando el individuo está casado que cuando no lo está y cuantificándose esa diferencia en 0,02359.

Finalmente, disponer de estudios, tanto secundarios como superiores, disminuye también la probabilidad de encontrarse en riesgo de pobreza o exclusión social con respecto a no tener estudios, si bien un individuo con estudios superiores disminuirá en mayor cuantía su probabilidad (una disminución del 12,31% para una persona con estudios secundarios frente a una disminución del 20,73% con educación superior).

5.4. Escenarios.

En el apartado que se desarrolla a continuación, se han planteado varias situaciones en las que el individuo adquiere diferentes características de las variables exógenas y para las que se pretende determinar la probabilidad de que dicho individuo se clasifique en el grupo de pobres. Por tanto, el objetivo que se persigue será el de evaluar con mayor profundidad la influencia de las variables explicativas sobre la pobreza. Además, este análisis permitirá realizar una interpretación más precisa de los efectos de las variables de carácter cualitativo, fijándose estas en un valor concreto y no en su valor medio como en los casos anteriores. Concretamente, se han planteado varios escenarios⁴ para distintos rangos de edad, variando en cada uno de ellos algunas características del individuo correspondientes a la nacionalidad, el estado civil, los estudios, la ocupación, el estado de salud o el grado de urbanización y manteniendo el resto constantes, analizando así la variación en la probabilidad estimada cuando se modifica dicha característica.

A la vista de los resultados, existen escenarios que destacan considerablemente en cuanto que presentan probabilidades muy elevadas de riesgo de pobreza. El más significativo se corresponde con un individuo de 25 años que reside en zona poco o medio urbanizada, extranjero, no casado, sin estudios, desempleado y con algún problema de salud crónico, tomando su probabilidad un valor del 94,2183%. Si se compara esta situación con la de una persona de nacionalidad española, se puede observar como la probabilidad de clasificarse como pobre, aunque sigue siendo elevada, disminuye en un 12,37% a una cifra del 81,8511%.

⁴ Los escenarios planteados y sus respectivas probabilidades estimadas se presentan en la Tabla 5.4. incluida en el Anexo I.

Tomando ahora el rango de edad de 51 años, la edad media de la muestra, y analizando las mismas características, la probabilidad de riesgo de pobreza se reduce considerablemente respecto a la misma situación correspondiente para 25 años, siendo del 69,9701% para un individuo español y del 89,3833% para el caso de un extranjero. Por otra parte, para el grupo de 80 años, estos escenarios en los que el individuo se encuentra desempleado no se han tenido en cuenta por considerarse que en esta edad la mayoría de población se encuentra en fase de jubilación. Sin embargo, se puede comparar la probabilidad de clasificarse en el grupo de pobres de una persona española de 80 años, sin estudios, jubilada, no casada, habitante en zona poco poblada y con enfermedad crónica con la probabilidad para el caso de un ciudadano de 25 o 51 años inactivo y con las mismas características, observando así que las probabilidades estimadas se reducen a media que aumenta la edad (52,63%, 37,4% y 22,24% para 25, 51 y 80 años, respectivamente). De todo lo anterior, puede deducirse que el riesgo de pobreza es menor para los rangos de edad más altos pero la nacionalidad es un factor clave en la determinación de dicho riesgo, perjudicando negativamente a los individuos extranjeros. Atendiendo a la ocupación, un individuo parado frente a otro que se encuentre inactivo, tanto para el rango de 25 como de 51 años y manteniendo el resto de variables en un valor constante, presenta probabilidades significativamente más elevadas de pobreza (escenarios 1, 2, 8 y 9), por lo que el desempleo puede considerarse también como una variable importante para determinar las condiciones de vida de la población. Cabe destacar, a su vez, que el estado civil, la salud y el grado de urbanización de la zona residencial no son variables que afecten especialmente a la pobreza dado que las probabilidades cuando se modifican estas características no varían en gran medida con respecto a como lo hacen cuando se modifican otras como la ocupación o los estudios. Por ejemplo, comparando el primer escenario con el tercero donde el individuo de 25 años pasa a estar casado, habitar en una zona urbanizada y no tener enfermedades crónicas solamente disminuye la probabilidad de ser pobre en un 9,08%.

Por último, analizando las menores probabilidades de riesgo de pobreza, el escenario más destacado es el que pertenece a una persona de 80 años, con estudios superiores, casada, inactiva, sin problema de salud crónico y residente

en zona muy urbanizada, cuya probabilidad alcanza el 4,37%. Para los rangos de edad de 25 y 51 años estas probabilidades se corresponden también con individuos casados, con estudios superiores, sin enfermedades, de zonas muy pobladas pero que se encuentran trabajando. Por tanto, se concluye que la educación superior y el empleo, sobre todo, favorecen las condiciones de vida de la población española.

5.5. Curvas de respuesta.

Para terminar con este estudio, se han elaborado una serie de gráficos o curvas de respuesta que permitirán evaluar el efecto de las variables cualitativas sobre la probabilidad de ser pobre cuando se producen cambios en una variable cuantitativa explicativa del modelo, en este caso, la edad. Para una mejor comprensión del apartado se hace necesario visualizar, en el Anexo II que se incluye al final de este trabajo, los gráficos a los que se hará referencia a lo largo del texto.

El procedimiento para elaborar estas curvas de respuesta ha consistido en representar gráficamente las probabilidades estimadas cuando el individuo presenta la característica de la variable dicotómica seleccionada y cuando no la presenta, junto con la edad, de tal manera que se obtenga una visión directa del efecto sobre la pobreza de cada una de las variables exógenas. Además, cabe destacar que para elaborar estas representaciones gráficas ha sido necesario construir una nueva variable edad, a fin de evitar valores repetidos en la muestra y obtener datos de forma continua.

En primer lugar, se ha observado que el estado civil, el grado de urbanización y el estado de salud no generan grandes diferencias en la probabilidad, al menos, no tan significativamente como otras variables como el desempleo o los estudios. Por este motivo, primero se analizarán brevemente sus curvas de respuesta para después interpretar más profundamente los gráficos correspondientes al resto de variables explicativas.

Observando las curvas de respuesta⁵ de las tres variables mencionadas con anterioridad, lo primero que debe destacarse es que la probabilidad de encontrarse en riesgo de pobreza o exclusión social es, para cualquier edad, superior para los individuos no casados, que presenten algún problema de salud crónico y que residan en zonas poco o medio pobladas. Además, la diferencia en probabilidad cuando se presenta la característica de la variable cualitativa frente a cuando no se presenta no es muy elevada, manteniéndose a lo largo de todo el periodo más o menos constante. Esto quiere decir que tanto el estado civil como el grado de urbanización de la zona residencial y la salud de la persona son variables que afectan a su situación de pobreza pero no en una medida importante.

A continuación, se procede a realizar la interpretación de las curvas de respuesta de las variables más determinantes de la pobreza de la población. La primera de ellas se corresponde con la nacionalidad⁶. En este caso, se observa la existencia de grandes diferencias en la probabilidad de encontrarse en riesgo de pobreza o exclusión social atendiendo a la nacionalidad del individuo, presentando una mayor probabilidad los ciudadanos extranjeros. Además, para los grupos de edad más avanzados, estas diferencias se van reduciendo, de tal manera que se puede afirmar que los jóvenes extranjeros residentes en España serán los más afectados por la pobreza frente al resto de individuos nacionales. Finalmente, a medida que avanza la edad, aunque siguen situándose en valores elevados, las diferencias se van reduciendo, pudiendo cuantificarse la variación, aproximadamente, desde un 30% para los más jóvenes hasta un 19% para los rangos de edad entre 80 y 90 años.

En tercer lugar, para los estudios se ha representado⁷, en una misma curva de respuesta, el efecto de la edad en función de las tres categorías asociadas a esta característica: sin estudios o con estudios inferiores, con educación secundaria y educación superior.

⁵ Ver Gráfico 5.1, Gráfico 5.2 y Gráfico 5.3 en Anexo II.

⁶ Ver Gráfico 5.4 en Anexo II.

⁷ Ver Gráfico 5.5 en Anexo II.

En este caso, la mayor probabilidad de riesgo de pobreza se asigna a aquellos que no poseen estudios, mientras que los individuos cuyo nivel máximo de estudios es la Educación Superior presentan menores probabilidades para cualquier rango de edad, situándose, por tanto, en una posición intermedia los que únicamente poseen el título de Educación Secundaria. Atendiendo a la influencia de la edad, se destaca que los grupos más jóvenes se verían más afectados por el riesgo de pobreza que los grupos más adultos, pues en este último caso la probabilidad de encontrarse en riesgo de pobreza se ve reducida en las tres categorías. Además, las diferencias serán mucho más acentuadas entre los jóvenes sin estudios con respecto a los que tienen estudios superiores que en relación a los que únicamente poseen Educación Secundaria.

Por último, se analiza la curva de respuesta de la edad atendiendo a la situación laboral⁸ de los individuos. Claramente, se puede afirmar que el desempleo es la situación laboral que afecta en mayor medida dado que las diferencias en la probabilidad de ser pobre entre un individuo parado y otro inactivo son significativamente más altas que las diferencias entre un individuo ocupado y un inactivo. Además la relación es inversa: la probabilidad de encontrarse en riesgo de pobreza será superior para un desempleado que para un inactivo, mientras que será inferior para una persona ocupada. En este caso, la diferencia en probabilidad también se ve reducida para los rangos de edad más alto, sobre todo coincidiendo con las etapas de jubilación (entre 65 y 86 años).

6. CONCLUSIONES

Tras haber realizado un estudio desde el punto de vista econométrico del fenómeno de la pobreza, que tanta preocupación genera a nivel mundial se han podido conocer cuáles son algunos de los determinantes más destacados de esta situación para el caso de España, todo ello a partir de los datos de la Encuesta de Condiciones de Vida del año 2015 elaborada por el INE y teniendo en cuenta que existen múltiples formas de medición de la pobreza pero se ha utilizado la metodología fijada por el Eurostat (umbral de pobreza en el 60% de la mediana de los ingresos).

⁸ Ver Gráfico 5.6 en Anexo II.

Se concluye, por un lado, que el modelo Logit que se ha estimado, aunque no es un modelo perfecto puesto que solo predice correctamente el 77,55% de las observaciones, presenta una buena bondad de ajuste, todas las variables seleccionadas son significativas para explicar la pobreza y, por tanto, la interpretación de los resultados obtenidos es válida.

Los resultados muestran que a medida que aumenta la edad disminuye la probabilidad de encontrarse en riesgo de pobreza. También disminuye esta probabilidad el hecho de ser español frente a ser extranjero y estar casado frente a no estarlo. Además cuando el nivel de estudios máximo del individuo es Educación Secundaria el riesgo de pobreza se reduce en relación a un individuo sin estudios, aunque esta reducción es mucho mayor cuando el nivel de estudios es Educación Superior. Disponer de empleo también es una variable que presenta relación inversa con la variable dependiente. Por otro lado, encontrarse en situación de desempleo es uno de los factores más determinantes de la pobreza, provocando un gran aumento de la probabilidad de ser pobre con respecto a la probabilidad cuando se está inactivo. Las enfermedades crónicas y residir en zonas poco urbanizadas también afectan negativamente a la pobreza aunque no son los factores más significativos. En base a esto, se concluye que los individuos con mayores probabilidades de encontrarse en riesgo de pobreza serán los jóvenes extranjeros, no casados, sin estudios, en paro, con algún problema de salud crónico y residentes en zonas poco o medio urbanizadas. Mediante el análisis gráfico, también se ha extraído la conclusión de que para cualquiera de las variables dicotómicas, las diferencias en probabilidad entre los individuos que presentan una u otra característica se reducen a medida que aumenta la edad, si bien estas diferencias se mantienen muy elevadas para determinadas variables como el desempleo o los estudios superiores.

En definitiva, y haciendo referencia al título y objetivo principal de este Trabajo Fin de Grado, se ha podido concluir que la edad, la nacionalidad, el nivel de estudios, la situación laboral, el estado civil, el estado de salud y el grado de urbanización de la zona de residencia son factores determinantes de la pobreza, destacando con mayor influencia sobre este fenómeno el desempleo, la nacionalidad y el nivel de estudios.

7. BIBLIOGRAFÍA

Goldberger, A.S. (2001). *Introducción a la Econometría*, Ariel, Barcelona.

Peña, D. (2002). *Análisis de datos multivariantes*, McGraw-Hill, Madrid.

INE (Instituto Nacional de Estadística). *Encuesta de Condiciones de Vida 2015*.

Disponible en:

http://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica_C&cid=1254736176807&menu=ultiDatos&idp=1254735976608 (última consulta: 28/06/2017)

INE (Instituto Nacional de Estadística). Ficheros transversales de usuario de la *Encuesta de Condiciones de Vida 2015*.

Mathus Robles, M.A: Principales aportaciones teóricas sobre la pobreza, en *Contribuciones a las Ciencias Sociales*, julio 2008. Disponible en:

www.eumed.net/rev/cccs (última fecha de consulta: 28/06/2017).

Sen, Amartya (1976). *Poverty: An Ordinal Approach to Measurement*. *Econometría*. Vol.44, no 2, 219-231.

Ureña, Carmen (1999). *Contraste entre medidas objetivas y subjetivas de pobreza*. Documento presentado en la Reunión del Grupo Río. Lisboa.

Herrero, C., A. Soler & A. Villar (2013). *La pobreza en España y sus comunidades autónomas: 2006-2011*. Valencia: Ivie, 38 pp. Disponible en:

http://dx.doi.org/10.12842/poverty_2012 (última fecha de consulta: 28/06/2017).

INE (Instituto Nacional de Estadística). *La pobreza y su medición: presentación de diversos métodos de obtención de medidas de pobreza*.

Ayala Cañón, L., Martínez López, R. & Sastre García, M. *La pobreza en España: un retrato a partir del Panel de Hogares de la Unión Europea*.

Banco Mundial. Disponible en: www.bancomundial.org

(Última fecha de consulta: 28/06/2017)

ANEXOS

ANEXO I. ESCENARIOS.

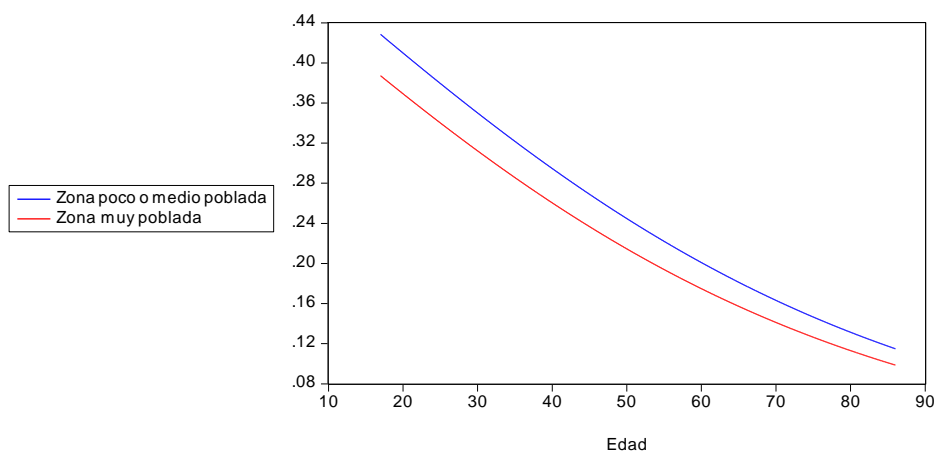
Tabla 5.4. Escenarios de probabilidad.

Edad	Nac.	E.Sec.	E.Sup.	E.Civil	Ocup.1	Ocup.2	Salud	Urb.	\hat{p}_i
25	1	No	No	0	0	1	1	1	0,818511
25	1	No	No	0	0	0	1	1	0,536264
25	1	No	No	1	0	1	0	0	0,727685
25	0	No	No	0	0	1	1	1	0,942183
25	1	Sí	No	0	0	1	1	1	0,701447
25	1	0	Sí	0	0	1	1	1	0,549069
25	1	0	Sí	1	1	0	0	0	0,105380
51	1	No	No	0	0	1	1	1	0,699701
51	1	No	No	0	0	0	1	1	0,373998
51	1	No	No	1	0	1	0	0	0,579931
51	0	No	No	0	0	1	1	1	0,893833
51	1	Sí	No	0	0	1	1	1	0,548294
51	1	0	Sí	0	0	1	1	1	0,386155
51	1	0	Sí	1	1	0	0	0	0,057365
80	1	No	No	0	0	0	1	1	0,222402
80	1	No	No	1	0	0	0	0	0,144909
80	0	No	No	0	0	0	1	1	0,508226
80	1	Sí	No	0	0	0	1	1	0,129677
80	1	0	Sí	0	0	0	1	1	0,071684
80	1	0	Sí	1	0	0	0	0	0,043752

Fuente: Elaboración propia a partir de los datos de la ECV.

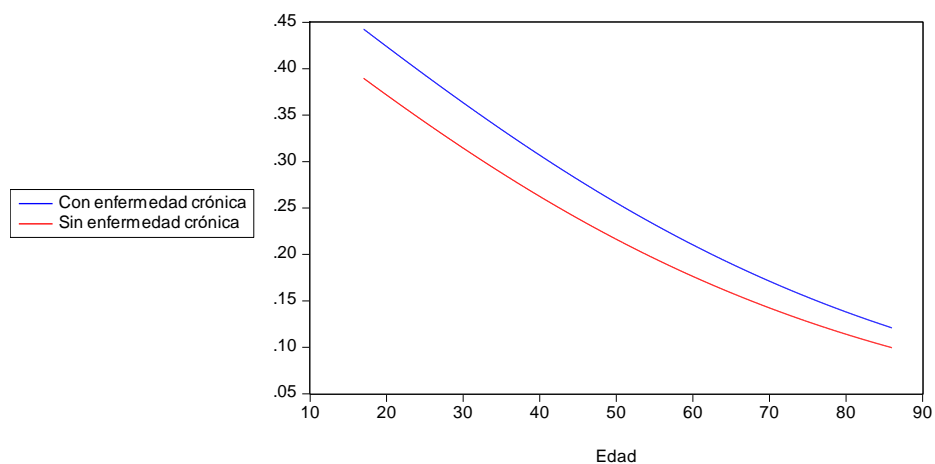
ANEXO II. CURVAS DE RESPUESTA.

Gráfico 5.1. Curva de respuesta de la edad y el grado de urbanización.



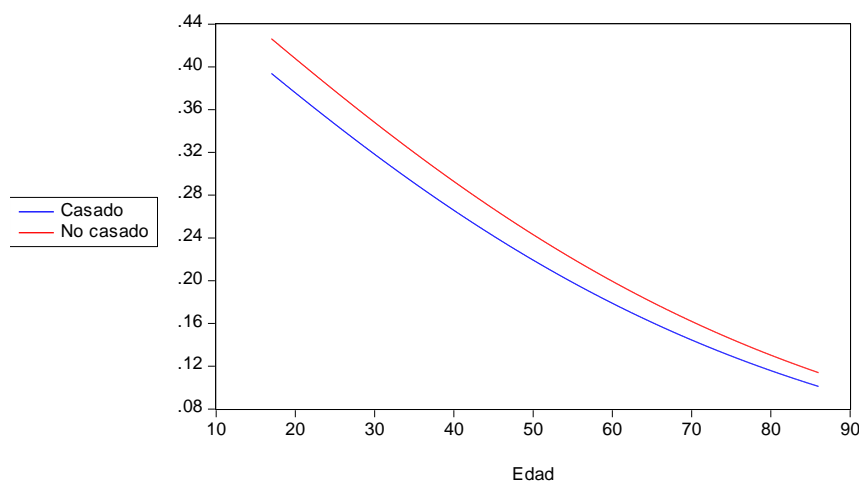
Fuente: Elaboración propia a partir de los datos de la ECV.

Gráfico 5.2. Curva de respuesta de la edad en función del estado de salud.



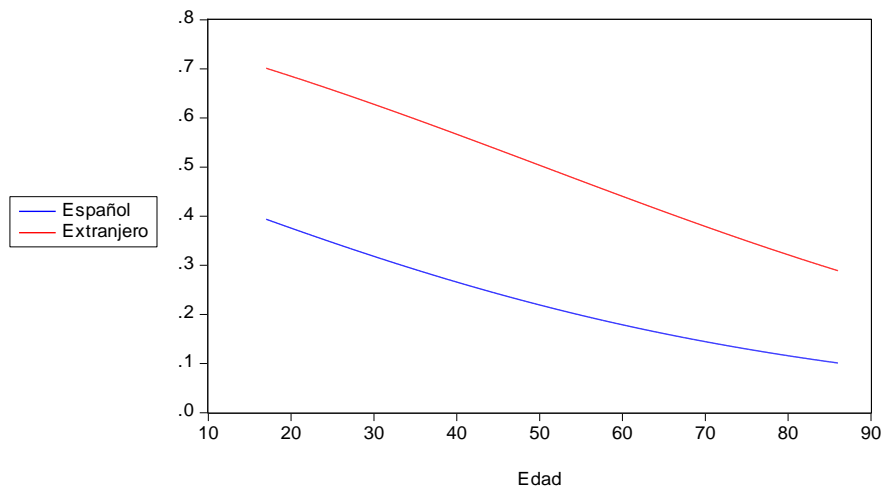
Fuente: Elaboración propia a partir de los datos de la ECV.

Gráfico 5.3. Curva de respuesta de la edad en función del estado civil.



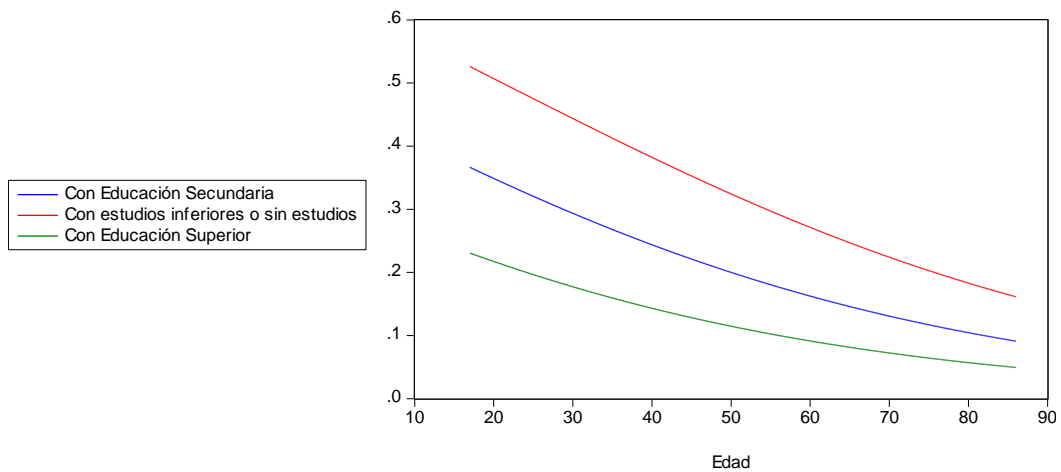
Fuente: Elaboración propia a partir de los datos de la ECV.

Gráfico 5.4. Curva de respuesta de la edad en función de la nacionalidad.



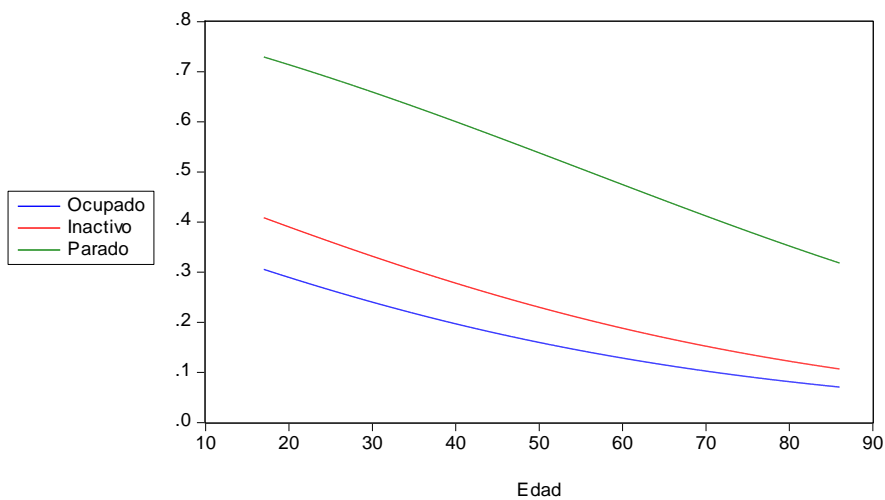
Fuente: Elaboración propia a partir de los datos de la ECV.

Gráfico 5.5. Curva de respuesta de la edad en función de los estudios.



Fuente: Elaboración propia a partir de los datos de la ECV.

Gráfico 5.6. Curva de respuesta de la edad en función de la situación laboral.



Fuente: Elaboración propia a partir de los datos de la ECV.