
This is the **published version** of the bachelor thesis:

Carpena Barranco, Hugo; Sanz Fernández, Àlex, dir. Análisis de factores condicionantes en la probabilidad de participación y nivel de gasto en lotería. 2021. 39 pag. (952 Grau en Economia)

This version is available at <https://ddd.uab.cat/record/247557>

under the terms of the  license



**Universitat Autònoma
de Barcelona**

**TÍTULO: Análisis de factores condicionantes en la
probabilidad de participación y nivel de gasto en lotería**

AUTOR DEL TFG: HUGO CARPENA BARRANCO

GRADO: ECONOMÍA

TUTOR DEL TFG: ÀLEX SANZ FERNÁNDEZ

FECHA: 25 DE MAYO DE 2021

***Agradecimientos:** Dar las gracias a mi tutor por todo la ayuda proporcionada durante todo el año y por animarme a seguir con mi camino académico. De la misma manera, agradecer a todas las personas que han contribuido de una manera u otra en el desarrollo del trabajo.

RESUMEN EJECUTIVO

En la actualidad, los mercados de juegos de azar y loterías han crecido de manera significativa en la mayoría de economías más desarrolladas. El desarrollo de este mercado ha conllevado efectos negativos derivados del mismo, como por ejemplo la adicción al juego o el juego patológico. Esta problemática ha conllevado la realización de diferentes estudios empíricos para examinar qué factores condicionan la compra de lotería. Este trabajo estudia y examina detalladamente a través de un modelo de doble valla, qué factores condicionan la probabilidad de participación y el nivel de gasto en lotería. El estudio tiene el objetivo de proporcionar información para el diseño de políticas encaminadas a lidiar con el problema del juego patológico. Se observa que la edad, el género del individuo y el nivel educativo son determinantes significativos que explican la probabilidad de participar en lotería y la decisión de cuánto gastar en la misma.

Palabras clave: Lotería, Juegos de azar, Modelo de doble valla, Nivel educativo.

EXECUTIVE SUMMARY

Nowadays, the gambling and lottery markets have grown significantly in most of the developed economies. The development of this market has led to negative effects derived from it, such as gambling addiction. This problem has entailed different empirical studies to examine what factors are related to gambling expenditure. This article examines extensively what are the determinants of gambling expenditure by using a double hurdle model. The aim of the study is to provide information about the probability of purchase and expenditure on lottery to design policies to deal with gambling addiction. The empirical findings show that age, gender and level of education are among significant determinants of gambling expenditure.

Keywords: Lottery, Gambling, Double hurdle model, Educational level.

Índice

1. INTRODUCCIÓN	4
2. REVISIÓN DE LA LITERATURA	7
3. BASE DE DATOS	9
3.1. Creación de la Base de Datos	9
3.2. Variables propuestas para el análisis	10
3.3. Variable dependiente	11
3.4. Variables independientes	11
3.5. Análisis descriptivo	13
4. METODOLOGÍA	17
4.1. Modelo econométrico	18
4.2. Tests	19
4.2.1. Normalidad de residuos	19
4.2.2. Multicolinealidad	19
4.3. Modificación del modelo	19
5. RESULTADOS	21
6. CONCLUSIONES	25
Referencias	26

Índice de tablas

1. Estadísticos principales sobre las variables numéricas	13
2. Estadísticos principales sobre las variables categóricas	14
3. Resultados de la estimación	21
4. Efectos marginales sobre la variable dependiente	23
5. Variables propuestas para el análisis	28
6. Matriz de correlación	35
7. Output de Resultados en STATA	37
8. Output de Efectos Marginales en STATA	38

Índice de figuras

1.	Participación en lotería por nivel educativo	15
2.	Gasto de lotería por género	16
3.	Modelo de Encuesta	31
4.	Participación en lotería por tramos de renta	32
5.	Participación por consumo de tabaco	32
6.	Participación por consumo de alcohol	33
7.	Participación por Estado Civil	33
8.	Participación por facilidad al acceso a locales de juegos de azar por lugar de residencia	34
9.	Distribución provincial de la muestra	34
10.	Densidad sobre la variable G_LOTT	35
11.	Test de Normalidad	35
12.	Normalidad Corregida	36

1. INTRODUCCIÓN

El surgimiento de la lotería se remonta a la Antigua Roma y a regiones de China, donde ya significaba una fuente de ingresos para la financiación de ciertos proyectos públicos como por ejemplo, la construcción de la muralla China. La introducción en Europa de la misma no surgió hasta el siglo XV cuando se extendió de forma muy popular. Sin embargo, en España no apareció hasta finales de 1763 con creación de una especie de lotería primitiva.¹

Una lotería, en sí, se define como un juego de azar en el que los participantes tienen una cierta probabilidad de ganar un premio cuando se efectúa el sorteo condicionado a la estocasticidad del mismo. Además, el desarrollo de este mercado y el crecimiento de su popularidad llevó a la creación de nuevas loterías llamadas loterías modernas que sucedieron a otros tipos de lotería como la lotería primitiva para el caso de España.

En general, la mayoría de las loterías ofrecidas son de tipo pasivo², como por ejemplo un décimo de lotería. Otras, son de tipo activo o semiactivo³, tales como las loterías de tipo *lotto*, las instantáneas o las de “rasca y gana”.

En la actualidad, los mercados de juegos de azar y loterías han crecido sustancialmente en la mayoría de economías más desarrolladas. A nivel mundial, el mercado del juego ha crecido tanto en términos absolutos desde el punto de vista de las ventas totales efectuadas como en términos relativos a través del consumo *per cápita* o las ventas totales en relación con el PIB. Países como Reino Unido, Italia, España y Francia se encuentran en la parte más alta de la clasificación en cuanto a ventas efectuadas. Por otra parte, la relación entre las ventas totales y el PIB más elevada se encuentra en Australia y Nueva Zelanda, Europa y América con un 0,57 %, 0,55 % y 0,5 % respectivamente (Garret, 2001).

En el caso de España, el principal operador en el mercado de juegos de azar, como en la mayoría de países europeos, es de ámbito estatal. Así mismo, este se ha ido incrementando en los últimos años.

¹La lotería en España llegó gracias a Carlos III e importándola desde Italia. Esta lotería era igual a la lotería primitiva que conocemos hoy en día. Esta funcionaba de manera que el jugador escoge seis números de una matriz de cuarenta y nueve.

²Una lotería pasiva es un juego condicionado a la adquisición de un billete o una fracción del mismo que lleva un número predefinido. Los premios se establecen con antelación independientemente del volumen de ventas efectuado.

³Los juegos activos se basan en un formato de apuestas mutuas, en el cual el jugador decide su propia suerte escogiendo una combinación de números. El valor monetario esperado depende de las ventas efectuadas. La mayoría son matrices 6/36 o 6/49 en la que esta última el jugador seis números de cuarenta y nueve disponibles.

Las ventas en 2019 crecieron un 2,85% respecto el 2018, reportando un beneficio neto de más de dos mil millones de euros y una aportación de 0,74% al PIB español (SELAE, 2019).

Esencialmente, existen una variedad de factores que caracteriza este mercado y que de alguna forma lo hace especial. Uno de estos, es el hecho de que las loterías pueden ser consideradas desde dos ámbitos económicos: como un bien de consumo o como una fuente de ingreso público siempre y cuando la titularidad del mercado sea pública o estatal (Cook y Cloetfelter, 1990). Además, el hecho de que estos mercados hayan alcanzado enormes índices de popularidad parece no estar en concordancia con las premisas de la teoría microeconómica que supone conducta racional, aversión al riesgo y maximizadora de la utilidad de los individuos (Pérez, 2009).

El desarrollo de este mercado y el consumo de lotería ha provocado ciertos efectos negativos derivados de este, como la adicción al juego. El juego patológico se considera un importante problema de salud mental en España y en la mayoría de los países desarrollados. Los estudios epidemiológicos sobre el juego patológico muestran una prevalencia del 1,4% al 1,7% en la población española. Así mismo, existe una cierta proporción de la población que tiene una implicación en conductas relacionadas con el consumo excesivo de juegos de azar (Becoña, 2004).

Pese a estos datos, en la actualidad las instituciones políticas y sociosanitarias parecen subestimar el impacto del consumo excesivo en el juego. Esta falta de interés puede parecer que prevalezca el interés económico derivado del consumo del juego sobre la enfermedad ocasionada por el mismo. De otra manera, la lotería supone una fuente significativa para las arcas públicas por lo que elaborar ciertas leyes para restringir el consumo o la participación en la lotería podría suponer una reducción consistente sobre esta tipología de ingresos. Respecto a esto, desde el Gobierno ya se está empezando a elaborar un Anteproyecto de Ley del Juego, aunque no se clarifica si simplemente afectará a empresas privadas.

Esta problemática se ha considerado objeto de estudio desde diferentes ámbitos de conocimiento. Muchos de los estudios previos han ido encaminados a determinar funciones de demanda para este tipo de bienes a partir de los factores condicionantes que explican la participación y el consumo en juegos de azar a través de instrumentos microeconómicos, en los cuales se establecen vectores de variables de diferente índole para la especificación de un modelo. El análisis de estos factores son clave para entender las probables consecuencias que se pueden derivar de dicho consumo.

De hecho, varias investigaciones como las elaboradas por Layton y Worthington (1999) y Crowley *et al.* (2012) ya muestran que la mayoría de las variables que se introducen en el vector, como edad, nivel educativo, nacionalidad o etnia o la renta mensual percibida por el individuo tienen capacidad para explicar este fenómeno y ayuda a establecer ciertas conclusiones sobre la situación.

En líneas generales, existe una necesidad de estimular nuevas líneas de investigación en este campo, especialmente, para el caso español por los escasos estudios previos aplicados a la población española. De igual manera, el objetivo del trabajo consiste en determinar qué factores están asociados a la probabilidad de participar y cuáles están relacionados con el gasto efectuado en lotería con la finalidad de incentivar implicaciones políticas para paliar efectos negativos derivados de la participación y consumo. Para ello, a partir de la realización de una encuesta con una muestra de 220 observaciones se ha creado un modelo de doble valla con la introducción de ciertas variables dependientes de diferentes índoles como ahora, edad, género, nivel de educación, nacionalidad, situación ocupacional o estado civil.

Por otra parte, es importante contemplar otras variables en el modelo que ningún otro estudio había establecido hasta ahora. El juego patológico puede conllevar la adopción de otras conductas adictivas que causan dependencia, como ahora el consumo de alcohol o de tabaco y de otras sustancias psicotrópicas. Es por eso que es de especial relevancia introducir estas variables consideradas como factores relacionados cuyo objetivo es observar qué efectos tienen sobre la probabilidad de participación y la decisión de cuanto gastar en lotería.

Los principales resultados muestran la edad, el género del individuo y el nivel educativo como variables que tiene capacidad para explicar la probabilidad de participación y el gasto efectuado de manera conjunta. En cuanto a la primera, se observa que un aumento de la edad, lleva a un aumento de probabilidad la participación y el gasto efectuado. Por otro lado, en relación con el género de individuo, se observa que el mero hecho de ser hombre conlleva un aumento en el gasto efectuado en comparación al género femenino. Además, la probabilidad de participación y el gasto efectuado para niveles educativos superiores a la educación obligatoria es inferior en todos los casos.

A continuación, se procede a la revisión de la literatura previa. Posteriormente, se realiza la explicación de la creación de la base de datos así como las variables que se introducen en el modelo y el análisis descriptivo obtenido. Más tarde, se procede a una discusión de la metodología seguida que se acompaña con ciertos tests y modificaciones para solucionar problemas. Por último, se lleva a cabo un análisis detallado los resultados obtenidos así

como las conclusiones del análisis elaborado.

2. REVISIÓN DE LA LITERATURA

Es importante destacar que el objetivo de los estudios empíricos anteriores pasaba por observar qué efectos podían explicar la compra de lotería. Algunos de ellos, examinan los determinantes del gasto, únicamente, en ciertos tipos de lotería como las de tipo *lotto*, instantáneas o juegos ofrecidos en casas de apuestas y en los casinos. Otros, realizan una estimación de los factores influyentes en la compra del dicho bien llevando a cabo un análisis en diferentes ámbitos.

En cualquier caso, estos estudios utilizan un vector de factores socioeconómicos y demográficos para examinar los determinantes del gasto. Además, cabe destacar que en la mayoría de los análisis previos existe una convergencia en la utilización las variables. Entre estas, encontramos, edad, lugar de residencia, renta mensual percibida, nivel de educación, estado civil, situación ocupacional o nacionalidad u origen étnico del individuo.

Respecto a la edad, Layton y Worthington (1999) en su estudio encuentran que a medida que esta aumenta, la probabilidad de participación en lotería también aumenta, por lo que existe una relación positiva entre edad y participación. Otros, como Abdel-Ghany y Sharpe (2001) en el cual llevan a cabo un análisis regional, encuentran que la edad es significativa en ciertas regiones y tiene capacidad explicativa sobre la decisión de cuanto gastar en lotería, sin embargo, esta no explica la probabilidad de jugar. Por otra parte, Crowley *et al.* (2012) observan que en rangos de edades de 45 a 64 años la que la probabilidad de participar y el gasto efectuado es superior respecto a otras edades. Siguiendo con este trabajo, en relación con el género del individuo, se puede observar que el hecho de que para el género femenino existe más probabilidad de jugar, pero en cambio, el gasto efectuado es menor en comparación al género masculino.

En lo que concierne al lugar de residencia, en la literatura previa encontramos que existe significatividad en la localización geográfica (Layton y Worthington (1999); Abdel-Ghany y Sharpe (2001); Crowley *et al.* (2012)). Sin embargo, los segundos, encuentran que el lugar de residencia tiene más capacidad para explicar el gasto efectuado y que no es tan significativa en determinar la probabilidad de jugar. En cuanto al último, observamos que en áreas urbanas se efectúa más gasto en lotería y existe más probabilidad de participar.

Por otro lado, en lo que se refiere a la renta mensual percibida, la finalidad es observar

qué sucede respecto al gasto que cada individuo efectúa. Se observa que esta variable es significativa y tiene un efecto positivo en el gasto por lo que a medida que aumenta la renta de los individuos también aumenta la probabilidad de participar (Layton y Worthington, 1999). Otros estudios, como el de Crowley *et al.* (2012) observan que el hecho de que aumente el ingreso de los individuos lleva a estos mismos a gastar más proporción de su renta en lotería.

Otra variable relevante que se introduce en los diferentes estudios previos es el nivel educativo o los años de educación. En el análisis de esta variable es más bien homogénea pues prácticamente todos los estudios muestran la variable tiene capacidad explicativa y establece una relación negativa pues medida que aumenta el nivel educativo del individuo disminuye la probabilidad de jugar y menor gasto efectúa (Abdel-Ghany y Sharpe (2001); Crowley *et al.* (2012)).

Otro factor que se introduce en diferentes estudios anteriores es la ocupación del encuestado. Algunos artículos como el de Layton y Worthington (1999) muestran que los coeficientes que se derivan de la estimación tienen capacidad explicativa y son estadísticamente muy significativos y determinan que aquellos individuos que tienen un cargo inferior mayor es la probabilidad de participar en lotería, sin embargo no explican la decisión de cuánto gastar. Por otra parte el estudio elaborado por Abdel-Ghany y Sharpe (2001) muestran que empíricamente no existe un comportamiento o patrón claro para diferentes niveles de ocupación. Además, este último conjuntamente con el artículo de Crowley *et al.* (2012) introducen la variable que informa sobre el estado civil de la persona. Los resultados muestran que el simple hecho de estar soltero reduce considerablemente la probabilidad de jugar a lotería y aquellas personas solitarias reducen el gasto que llevan a cabo.

Asimismo, otra de las variables que contemplan los diferentes estudios es la etnia o nacionalidad del individuo encuestado. Layton y Worthington (1999) observan que se trata de un factor significativo y establece la capacidad de explicar el comportamiento diferencial sobre la participación y el gasto efectuado.

Por último, en relación con el estudio que realiza Balabanis (2002) se observa que los factores demográficos y los factores compulsivos como por ejemplo, el consumo de alcohol y de tabaco están relacionados con la participación frecuente en lotería. El análisis efectuado se basa en el establecimiento de ciertas hipótesis con base en la relación que puede (o no) existir entre estos factores y la participación del dicho bien. Los resultados muestran que existe una relación positiva en cuanto a la existencia de factores compulsivos en el individuo en relación con la participación y la compra de boletos.

3. BASE DE DATOS

3.1. Creación de la Base de Datos

Para el análisis, como en muchos estudios previos se ha requerido de la realización de una encuesta para obtener observaciones que capten los posibles efectos considerados en el estudio que se lleva a cabo. Primeramente, es destacable que el diseño de la encuesta ha pasado por diferentes fases y procesos de evaluación. Este diseño se ha basado en la creación de dos modelos de encuesta. Cada una de estas contenía diferentes preguntas que tenían el objetivo de conseguir la misma información para la estimación del modelo. Entre las preguntas se podía inducir información sobre la edad del individuo, el género, la nacionalidad, la situación ocupacional, el cargo que desempeña cada individuo en caso de que este perciba la remuneración del trabajo, las horas de trabajo que desempeña, el consumo de alcohol y de tabaco, la renta mensual percibida a partir de tramos o bien preguntando sobre su valor exacto o bien el nivel educativo del encuestado.

En cuanto al proceso de evaluación de estos dos modelos de encuesta se escogieron las preguntas más pertinentes basadas en el patrón de respuesta de los individuos de una muestra de aproximadamente 20 personas en cada modelo y por lo tanto, se efectuó una mezcla entre estas dos deviniendo en una encuesta definitiva más perfeccionada que captase de manera más adecuada la información que se quería recoger de los encuestados y evitar así, ciertos errores de muestreo debido a la cantidad de preguntas que son de ámbito personal y privado. Una vez elaborada esta encuesta final se procedió al envío masivo a través de la plataforma de «Google Forms» concluyendo la encuesta con 227 observaciones.⁴

Posteriormente, se realizó la observación y la limpieza de los datos eliminando ciertas incoherencias que existían en las observaciones, concluyendo así con 220 observaciones definitivas. Además, se decidió omitir la variable que informaba sobre la nacionalidad debido a que la gran mayoría de los encuestados tenían nacionalidad española por lo que, al no existir variabilidad en la misma, si se introdujese la variable en el modelo no tendría carácter explicativo ni significativo.

Otra de las variables omitidas es la referente a la ocupación. La justificación de la omisión se debe a que no se ha podido captar de manera adecuada la información referente a la ocupación del individuo para saber si empleaba un trabajo cualificado. Además, no todos los individuos han contestado a la pregunta y por lo tanto, se presentaba un problema de falta de observaciones. De todos modos, este hecho no es excesivamente preocupante, ya

⁴Los dos modelos de encuestas provisionales también se enviaron a través de la plataforma «Google Forms». El modelo de encuesta definitivo se puede consultar en la Figura 1 ubicada en el anexo.

que se introducen variables como el nivel educativo que está estrechamente relacionado con la ocupación del individuo.

No obstante, existe la consciencia de que esta metodología con la que se ha obtenido la información, aunque se haya hecho todo lo posible para prevenir ciertos problemas, está sujeta a ciertos sesgos y errores. Primeramente, la encuesta puede acarrear sesgos de observación en ciertas preguntas debido a la diferencia sistemática que puede existir entre el valor real y el que realmente registra el encuestado. Por otra parte, pueden existir sesgos de selección debido a que la muestra puede no ser suficientemente significativa como para poder determinar conclusiones en su inferencia estadística.

3.2. Variables propuestas para el análisis

Los anteriores estudios empíricos han utilizado diferentes variables para la estimación de la probabilidad de participación en lotería así como el consumo efectuado en esta. Entre estas variables podemos encontrar, edad, género, lugar de residencia como variables demográficas y otras como renta mensual percibida, nivel de educación, estructura familiar, como indicadores de índole socioeconómica. El análisis, en todos los casos, está sujeto a diferentes tipos de variables: continuas o discretas y también variables ficticias.

En el Cuadro 1 ubicado en el anexo se exponen las variables propuestas para análisis que se divide en variables de diferentes ámbitos. En cuanto a las variables demográficas, se encuentran, la edad del individuo encuestado, su género y la facilidad de acceso a juegos de azar a partir del lugar de residencia. La utilización de estas, se justifica por el hecho de que estudios anteriores determinan que existe una relación explicativa en lo que se refiere a la participación y cuánto se consume en lotería.

Además, la base de datos contempla variables socioeconómicas como la renta mensual percibida, el nivel de educación máximo alcanzado por el encuestado o estado civil. En lo que se refiere a estas, también existe evidencia de una cierta relación y causalidad, sin embargo, no todos los estudios indican una significatividad suficientemente alta sobre la capacidad explicativa de alguna de las variables de carácter socioeconómico.

Por otra parte, siguiendo el trabajo de Balabanis (2002), también se introducen variables que se consideran factores relacionados, como el consumo de alcohol y tabaco que explican la participación frecuente y el consumo excesivo en lotería.

3.3. Variable dependiente

En este trabajo, estamos determinando qué variables afectan a la probabilidad de participación y el gasto efectuado en lotería. Esencialmente, presuponemos que las variables que afectan a la decisión de participación, no tienen el mismo efecto que aquellas que explican el gasto efectuado. En cualquiera de los dos casos, la variable dependiente utilizada en el modelo econométrico es la de gasto semanal en lotería (G_LOTT). Esta variable explicada, por lo tanto, pretende informar sobre el gasto que se efectúa a la semana por cada individuo de la muestra.

3.4. Variables independientes

En cuanto a las variables explicativas, estas están formadas por un vector de diferentes indicadores. La mayoría, son variables que provienen del ámbito socioeconómico y demográfico. Por otra parte, también se incluyen factores relacionados con la excesiva participación y el consumo frecuente en lotería.

- Edad (ED): Se trata de una variable discreta que nos informa de la edad del individuo. Es una variable explicativa importante porque es esperable que sea significativa, ya que a medida que aumente la edad, aumentará la probabilidad de participación y el consumo efectuado en lotería. Por lo tanto, se presupone que pueda existir una relación causal positiva, aunque no está tan claro si la probabilidad de participación y el gasto efectuado sea significativo para todas las edades. Puede ser que la relación sea significativa en edades medianas o más bien, en edades avanzadas.
- Género (GNR): Esta variable discreta contiene información sobre el género de la persona encuestada. La introducción de esta variable se justifica por el hecho de que el comportamiento del consumidor puede diferir por género. En este caso, vamos a establecer el supuesto de que el hecho de ser hombre, conlleve que la probabilidad de jugar sea más elevada y que el gasto efectuado sea superior en comparación al género opuesto, ya que la problemática con los juegos de azar se ha descrito y relacionado con mayor prevalencia en los hombres.
- Ingreso mensual (RNT_MES): Esta variable nos informa sobre el ingreso mensual percibido por el encuestado. Principalmente, la utilización de esta se basa en el hecho de que muy probablemente sea una de las que muestre más relación y causalidad con la variable dependiente (G_LOTT). En general, esperamos que a mayor nivel de renta, más probabilidad de participar en lotería y más nivel de gasto.
- Facilidad de acceso a juegos de azar (URB_PROXIMIDAD): Se trata de una variable ficticia que informa si el individuo vive una localidad en la que existe oferta

de juegos de azar, como por ejemplo, casas de apuestas o casinos. Se introduce la siguiente variable debido a que puede existir un comportamiento espacial diferente sujeto a que el individuo resida en una área u otra. En cualquier caso, podemos pensar que existirá una significatividad y un aumento de la probabilidad de participar y la decisión de cuánto gastar sobre aquellos individuos que residen en localidades que tienen este tipo de oferta.⁵

- Educación (EDUC): La variable que nos informa de la educación del individuo también es de las más relevantes en el vector de variables independientes, esto se debe a que el nivel de educación tiene un efecto significativo sobre el consumo de lotería. Es esperable, entonces, que se establezca una relación negativa y significativa que nos indique que con un aumento en el nivel de educación del individuo encuestado, disminuya la probabilidad de jugar y el gasto efectuado. La justificación de este hecho pasa por pensar que a mayor nivel de educación existe más conocimiento sobre la teoría de la probabilidad y consecuentemente más conciencia sobre el consumo en la compra de juegos de azar.
- Estado Civil (E_CIVIL): El estado civil nos transmite si el individuo de la muestra está soltero, casado, divorciado o viudo. El comportamiento de esta variable es difícil de prever, aunque es cierto, que algunos de los estudios anteriores consideran que es una variable significativa y tiene cierta capacidad explicativa. Por una parte, podríamos pensar que aquellos individuos casados con cargas familiares, optarían por jugar de manera más recurrente a la lotería, por lo que se podría esperar que los casados tuvieran una probabilidad y nivel de gasto mayor en relación a los solteros. Sin embargo, aquellas personas que son viudas, tendrían menor probabilidad de participación debido a la no existencia de las cargas familiares. Por otra parte, el comportamiento de una persona soltera se puede asemejar en parte al comportamiento de una divorciada. En conclusión, no sabemos de qué manera se comportará la variable.
- Consumo de tabaco (TBC): Indica el consumo de tabaco efectuado por el individuo. Lo que se quiere observar con esta variable, es si existe un efecto diferencial de la participación y el gasto efectuado entre una persona fumadora y otra que no consume. Partiendo de las hipótesis de estudios anteriores, es esperable que el hecho de consumir tabaco conlleve una probabilidad más elevada en participación y un consumo más elevado en lotería.
- Consumo de alcohol (ALC): Indica el consumo de alcohol que realiza el encuestado. La introducción de esta variable se pretende analizar bajo el mismo supuesto que la

⁵La información de esta variable se obtiene a partir del lugar de residencia. Por cada lugar de residencia, se ha buscado a través de Google Maps si en el área comercial de la localidad, que en muchos casos coincide con el centro urbano, existía oferta de juegos de azar y cómo de aglomerada estaba la oferta.

variable de consumo de tabaco (TBC). Por lo tanto, esperamos que en su estimación el efecto sea el mismo.

3.5. Análisis descriptivo

Sobre el análisis que se lleva a cabo *ex ante* a la estimación del modelo, cabe destacar que la principal idea sobre metodología que se ha utilizado es la diferenciación entre variables numéricas y variables categóricas.⁶

Variable	Máx.	Mín.	Media	Desviación estándar
Muestra entera: $n = 220$				
Jugadores de Lotería: $n = 113$				
G_LOTT	20 euros	0 euros	2,32 euros	3,76 euros
ED	69 años	18 años	41,82 años	13,82 años
RNT_MES	3701 euros	0 euros	1532,30 euros	767,80 euros

Tabla 1: Estadísticos principales sobre las variables numéricas

En la Tabla 1 se muestran los estadísticos principales de las variables numéricas. Se puede ver que el análisis se ha realizado con una muestra de 220 individuos en los que 113 de estos han reportado su participación en lotería, lo que supone un 51,36 % de la muestra entera.

En lo que concierne a las variables categóricas, se puede encontrar toda la información sobre los estadísticos principales en la siguiente tabla. En este estudio, el conjunto de las categóricas está formado por variables como educación, estado civil, lugar de residencia, consumo de alcohol y tabaco, nacionalidad y género.

Variable	Categorías	Frecuencia Absoluta	Frecuencia Relativa
GNR	Masculino	115	50,88 %
	Femenino	111	49,12 %
E_CIVIL	Soltero/a	88	38,94 %
	Casado/a	117	51,77 %
	Divorciado/a	18	7,96 %
	Viudo/a	3	1,33 %

⁶Una variable numérica es aquella que toma valores discretos o continuos. Las variables categóricas son aquellas formadas por modalidades y no representan números.

EDUC	Bachillerato	26	11,5 %
	Educación Obligatoria	35	15,49 %
	CFGM	27	11,95 %
	CFGS	53	23,45 %
	Universidad: Grado	54	23,89 %
	Universidad: Máster	29	12,83 %
	Universidad: Doctorado	2	0,88 %
URB_PROXIMIDAD	Vive en una localidad donde existe oferta de juegos de azar	73	32,27 %
	No vive en una localidad donde existe oferta de juegos de azar	153	67,73 %
TBC	Reporta ser fumador	115	50,88 %
	No reporta ser fumador	111	49,12 %
ALC	Afirma ser consumidor de alcohol	80	35,40 %
	Rechaza ser consumidor de alcohol	146	64,60 %

Tabla 2: Estadísticos principales sobre las variables categóricas

Respecto al primero, se observa que el nivel de educación más predominante sobre la muestra es grado universitario, seguido del Ciclo de Formación Superior (CFGS). Estos representan un 24 % y 23,5 % respectivamente. Por otra parte, en relación con el estado civil, es destacable que más de la mitad de la muestra está casado o casada. Además, se observa que la distribución entre género masculino y femenino es bastante homogénea.

En cuanto a la variable que informa sobre la facilidad de acceso a juegos de azar, a la que llamamos URB_PROXIMIDAD, se observa que alrededor de un 33 % de los encuestados viven en localidades que poseen locales de juegos de azar.⁷

En relación con el análisis *a priori* de los factores relacionados con la participación y el consumo de lotería se observa que aproximadamente, un 51 % de los individuos de la muestra consume tabaco, aunque no todos lo hacen de forma diaria. Por otra parte, un

⁷Para la variable URB_PROXIMIDAD, se ha utilizado el criterio de diferenciación sobre la oferta de juegos de azar que posee la localidad en la cual vive el encuestado. Al elaborar esta variable se ha podido observar que la mayoría de municipios que poseen este tipo de locales son del ámbito de las apuestas deportivas de empresas como SPORTIUM o Codere.

35,4% afirma ser consumidor de alcohol.

Otro elemento importante observar la relación inicial que puede tener la participación y el nivel de gasto en lotería respecto a las diferentes variables. Para esto, se ha procedido a la elaboración de ciertos gráficos para ilustrar el impacto.⁸

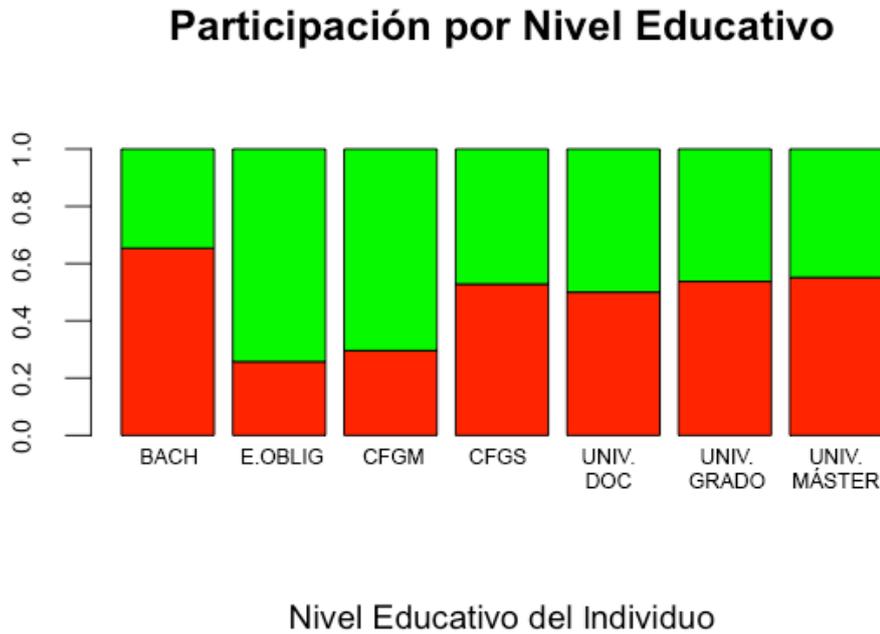
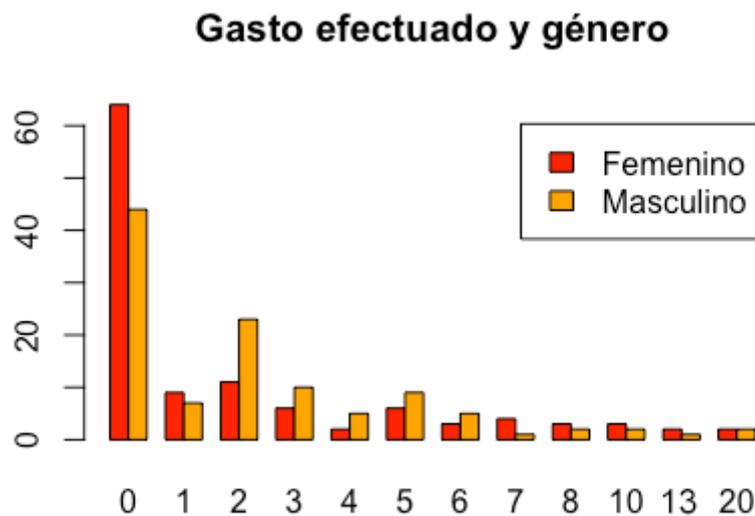


Figura 1: Participación en lotería por nivel educativo

En la Figura 1 se expone la participación en lotería por cada nivel educativo. Se puede observar que la participación es más elevada en niveles educativos inferiores, como ahora, la educación obligatoria o en aquellos individuos con un Ciclo Formativo de Grado Medio. En ambos casos, la participación en lotería por cada nivel educativo supera 65%. En otras palabras, aquellos porcentajes de participación en lotería más elevados se encuentran en niveles de educación inferiores. En niveles educativos más elevados, el comportamiento de la participación sobre estos es bastante homogénea y no hay tanta diferenciación de la misma por nivel educativo, a excepción del Bachillerato donde se observa que es más baja.

⁸Todos los gráficos se han elaborado con el programa «RStudio».



Gasto en lotería en relación al género del individuo

Figura 2: Gasto de lotería por género

En la Figura anterior se muestra el gasto en lotería desglosado por género. Podemos observar que existe un mayor número de personas con género femenino que no juegan a la lotería. Por consiguiente, en aquellos individuos jugadores se observa que el género masculino efectúa un gasto de entre 1 y 6 euros semanales en comparación a las mujeres que tienen un nivel de consumo más elevado de entre 7 y 13 euros. En cuanto al gasto semanal máximo (20 euros) se destaca que hay la misma cantidad de hombres y mujeres.

Una vez realizado el análisis descriptivo⁹, se procede a realizar un análisis empírico más detallado para ver si existen las relaciones establecidas anteriormente. Para ello, es necesario explicar el comportamiento del gasto en lotería que ayuda a justificar la metodología seguida en la estimación del modelo. En la Figura 10 ubicada en el anexo, se muestra la distribución de la densidad sobre G_LOTT.

Se puede observar que una gran cantidad de observaciones toman valor cero y posteriormente, para valores superiores a este, la densidad de las observaciones va disminuyendo consistentemente y por lo tanto, la estimación del modelo requerirá la utilización de modelos que contemplen estimaciones a partir de variables dependientes limitadas.

⁹Para ver un análisis más detallado de la participación por cada variable se puede consultar en los anexos otros gráficos explicativos.

4. METODOLOGÍA

Previamente a la explicación del procedimiento seguido en la estimación, es importante tener en cuenta varias consideraciones econométricas desde el punto de vista metodológico. Por eso, se debe destacar el hecho de que no existe un criterio homogéneo por el cual se estiman los parámetros de las variables propuestas para el análisis. En casos en los que se lleva a cabo este tipo de análisis, la estimación de un modelo lineal por Mínimos Cuadrados Ordinarios (MCO) nos lleva a $\hat{\beta}^{MCO}$ sesgados e inconsistentes debido a la especial caracterización de la variable explicada. Por lo tanto, existen diferentes modelos y métodos para la estimación de los parámetros con variables censuradas y variables dependientes limitadas, como por ejemplo, modelos de doble valla (Cragg, 1971) o Tobit (Tobin, 1958).

En modelos Tobit, la principal premisa se basa en que la variable dependiente toma valores no-negativos, con el cero reservado para observaciones censuradas, por lo que se tiene en cuenta que la variable explicada vale cero para una parte considerable de la muestra, pero se distribuye, aproximadamente, de forma continua para los valores positivos.

En cuanto a los modelos de doble valla, se sigue la principal premisa de los modelos Tobit y además se tiene en cuenta la diferencia entre decisión y participación. En otras palabras, este método de estimación tiene en cuenta que los factores que se asocian a decisión o probabilidad de participación pueden no ser los mismos que se asocian participación en sí.

Por otra parte, existe otro método de estimación que se utiliza en diversos campos como ahora, la economía de la salud o la economía del transporte. La metodología que se propone se basa en el modelo Poisson de ceros inflados que fueron propuestos por Lambert (1992). Este modelo asigna una probabilidad p en aquellas observaciones que declaran cero y una probabilidad $1 - p$ para las otras observaciones que siguen una distribución Poisson. Sin embargo, esta metodología puede no resultar del todo adecuada debido a la cantidad de ceros que se pueden presentar los datos y que presenta inconsistencia con el modelo de Poisson.

En general, para el análisis que vamos a realizar, es válido y correcto utilizar cualquiera de los tres métodos de estimación. Sin embargo, el método más apropiado es utilizar un modelo de doble valla, ya que la intención del análisis es determinar qué factores influyen en la probabilidad de participación en lotería y el gasto efectuado, sujeto a que estos factores pueden o no tener un efecto diferente en la decisión de participar y el consumo que realizan los individuos y por lo tanto, se ajusta perfectamente al objetivo del trabajo.

4.1. Modelo econométrico

El modelo econométrico propuesto se basa en un modelo de doble valla¹⁰ que nos permite separar dos procesos estocásticos diferentes: la decisión de participar y la decisión de consumir. Además, este tipo de modelos se caracterizan por la relación $y_i = s_i h_i^*$ donde s_i es la variable de selección que contempla la decisión de participación, entonces:

$$s_i = \begin{cases} 1 & \text{si } z_i' \eta + \epsilon_i > 0 \\ 0 & \text{en otro caso} \end{cases}$$

donde z_i es el vector de variables explicativas, η es el vector de coeficientes y ϵ_i el error estándar se distribuye como $N \sim (0, 1)$. Por otra parte, la variable discreta latente¹¹ y_i^* se observa cuando $s_i = 1$ y informa sobre el consumo en lotería que realiza el individuo y por lo tanto:

$$y_i^* = \exp(x_i' \beta + v_i)$$

donde x_i es el vector de variables explicativas, β es el vector de coeficientes y v_i son los errores estándar que se distribuyen como $N \sim (0, \sigma^2)$. Además, los términos de error ϵ_i y v_i son independientes.¹²

Por otro lado, los coeficientes se estiman por máxima verosimilitud a través de la maximización del logaritmo de la siguiente función:

$$\text{Log} - L = \sum_0 \ln \left[1 - \Phi(z_i' \eta) \Phi \left(\frac{x_i' \beta}{\sigma_i} \right) \right] + \sum_+ \ln \left[\Phi(z_i' \eta) \frac{1}{\sigma_i} \phi \left(\frac{y_i - x_i' \beta}{\sigma_i} \right) \right]$$

donde $\Phi(\cdot)$ y $\phi(\cdot)$ son las funciones de distribución acumulativa y densidad estándar normal, respectivamente.¹³

En resumen, la introducción de este método nos permite tener estimadores η^{MV} y β^{MV} consistentes e insesgados. Además, el modelo econométrico y el método de estimación van acompañados de diferentes tests para supervisar que todos los supuestos se cumplen

¹⁰Como se ha comentado anteriormente, se ha realizado un análisis de la utilización del modelo más apropiado. Para ello, se ha realizado la estimación por Tobit, Probit y regresión truncada. Observando el logaritmo de la función de verosimilitud, se ha realizado el siguiente test:

$$\lambda = [\text{Log} - L_{\text{Probit}} + \text{Log} - L_{\text{Truncreg}} - (L - \text{Log}T_{\text{obit}})] \sim \chi_{n,p}^2$$

donde n son los grados de libertad, tantos como variables independientes tenga el modelo y p el nivel de significatividad. El resultado del test muestra que $\lambda = 16,32563$ excede el valor crítico apropiado $P[\chi^2 > 16,32563 = 0,1768]$. Por lo tanto, la estimación del modelo será más fiable si se realiza a partir de un modelo de doble valla.

¹¹La variable latente es aquella que no es observable y se infiere a partir de otras que sí son observables.

¹²El modelo de doble valla se puede estimar de dos maneras: linealmente o exponencialmente. En este caso, la estimación más adecuada es la exponencial debido a que la variable dependiente es discreta y no continua. Además, en algunos casos se ha observado que ciertas variables no se comportan de manera lineal.

¹³Teoría basada en el artículo de Cragg, J. (1971). Some Statistical Models for Limited Dependent Variables with Application to the Demand for Durable Goods. *Econometrica*, 39(5), 829-844. doi:10.2307/1909582

correctamente.¹⁴

4.2. Tests

4.2.1. Normalidad de residuos

Tal y como hemos comentado anteriormente, los términos de error ϵ_i y v_i se distribuyen de la siguiente manera:

$$\begin{pmatrix} \epsilon_i \\ v_i \end{pmatrix} = N \sim \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & \sigma^2 \end{pmatrix}$$

Para evaluar la normalidad que sucede de la estimación se ha utilizado el comando `sktest` en STATA que relaciona la asimetría y la curtosis en relación a la distribución normal. El test está formado por las siguientes hipótesis:

H_0 : *Los errores se distribuyen de forma normal*

H_1 : *Los errores no se distribuyen de forma normal*

Al realizar el test, se recibe un output en STATA que se muestra en la Tabla 4 ubicado en el anexo. El resultado obtenido del p-valor nos lleva a rechazar, para cualquier valor de significatividad, H_0 y por lo tanto, los errores no se distribuyen de forma normal. En todo caso, esto nos lleva a realizar cambios en el modelo ya que existe la violación de al menos un supuesto: la normalidad de los residuos.

4.2.2. Multicolinealidad

La presencia de multicolinealidad en la regresión efectuada puede hacer que los coeficientes β^{MV} sean inestables provocando una cierta no significatividad en algunos coeficientes incluso cuando hay una relación entre la variable dependiente y la explicativa. La multicolinealidad sucede cuando una o más de una variable independiente está relacionada con otra variable predictora. Para la detección de colinealidad, se ha realizado la matriz de correlaciones de todas las variables independientes introducidas en el modelo utilizando el comando `corr` en STATA (véase Tabla 6). Generalmente, al observar los coeficientes de correlación se concluye no existe presencia significativa de colinealidad entre variables para que tenga efectos no deseados en los coeficientes.

4.3. Modificación del modelo

Para no violar los supuestos que se contemplan en el modelo de doble valla debemos realizar modificaciones para evitar obtener estimadores inconsistentes y sesgados estima-

¹⁴Para la estimación, que se ha realizado con STATA, se ha utilizado el comando `churd1e`.

dos por máxima verosimilitud. Para ello, se realiza una transformación T en la variable dependiente G_LOTT de manera que:

$$T(\theta y_i) = \log \left[\theta y_i + \left(\theta^2 y_i^2 + 1 \right)^{1/2} \right] / \theta = \operatorname{senh}^{-1}(\theta y_i) / \theta$$

donde θ es un parámetro desconocido. Una vez realizada la transformación de la inversa del seno hiperbólico (IHS)¹⁵, la función de verosimilitud cambia de tal manera que:

$$\operatorname{Log} - L = \sum_0 \ln \left[1 - \Phi(z'_i \eta) \Phi \left(\frac{x'_i \beta}{\sigma_i} \right) \right] + \sum_+ \ln \left[(1 + \theta^2 y_i^2)^{-1/2} \Phi(z'_i \eta) \frac{1}{\sigma_i} \phi \left(\frac{T(\theta y_i) - x'_i \beta}{\sigma_i} \right) \right]$$

En definitiva, al efectuar este cambio, se puede observar en el Cuadro 6 los tests realizados donde ya se corrige la normalidad de residuos y por lo tanto, la estimación con esta modificación permite obtener estimadores consistentes e insesgados. Por otra parte, la modificación de la variable no afecta a la situación de no multicolinealidad explicado previamente. Además, otra de las ventajas que sostiene esta modificación es que soluciona problemas de heterogeneidad no observada en el modelo.

¹⁵Transformación realizada a través de la metodología realizada por Yen, S. T., & Jensen, H. H. (1996). Determinants of household expenditures on alcohol. *Journal of Consumer Affairs*, 30(1), 48–67. <https://doi.org/10.1111/j.1745-6606.1996.tb00725.x>

5. RESULTADOS

	Modelo de Doble Valla (IHS)	
	Participación	Consumo
ihs_G_LOTT		
ED	0.0277*** (0.010)	-0.000400 (0.004)
URB_PROXIMIDAD	0.124 (0.200)	0.0947 (0.077)
GNR_MASCULINO	0.429** (0.194)	-0.00746 (0.081)
E_CIVIL_CASADO	-0.136 (0.260)	0.155 (0.097)
E_CIVIL_DIVORCIADO	-0.0431 (0.428)	0.0686 (0.147)
E_CIVIL_VIUDO	3.628 (203.9)	0.192 (0.256)
EDUC_BACH	-0.524 (0.410)	-0.370** (0.167)
EDUC_FPMEDIO	0.0693 (0.376)	-0.140 (0.122)
EDUC_FPSUP	-0.346 (0.319)	-0.182 (0.120)
EDUC_UNIV_GRADO	-0.297 (0.326)	-0.288** (0.125)
EDUC_UNIV_SUPERIOR	-0.365 (0.356)	-0.164 (0.157)
RNT_MES_MEDIA		-0.0000423 (0.000)
TBC	0.270 (0.190)	-0.164** (0.081)
ALC	-0.309 (0.207)	0.0937 (0.085)
Constante	-1.074** (0.530)	0.780*** (0.212)
Log-verosimilitud		-140.32077

El asterisco (*) muestra la significatividad del coeficiente — 1% (***), 5% (**), 10% (*). La estimación de la ecuación de la participación, como restricción de STATA, al menos debe tener $k - 1$ variables independientes introducidas en la ecuación del consumo. En este caso, se ha decidido omitir la variable RNT_MES_MEDIA. En paréntesis, se muestra la desviación estándar de cada coeficiente.

Tabla 3: Resultados de la estimación

En la Tabla anterior se muestra la regresión a partir del modelo de doble valla. La estimación se ha realizado a partir de la maximización de la función de verosimilitud que previamente se ha establecido en la explicación de la metodología.

En cuanto a los resultados, es destacable que los coeficientes obtenidos para la ecuación de la participación difieren de los obtenidos en el consumo. En otras palabras, se confirma el hecho de que las variables que tienen capacidad explicativa para explicar los dos procesos estocásticos son distintas.

En relación con las variables afectadas en la participación en lotería, la variable ED muestra una relación positiva por lo que con un incremento en la edad del individuo, se produce en consecuencia un aumento de la probabilidad de jugar, siendo esta una variable con un coeficiente significativo para los tres niveles de significatividad. Otra de las variables con capacidad explicativa es la relativa al género del individuo. Los resultados muestran una significatividad del 5% y 10% y podemos considerar, como era esperable, que el género masculino tiene más probabilidad de jugar que el femenino.

En general, es necesario destacar que todas las otras variables introducidas en la ecuación de la participación no son significativas y por lo tanto no tienen ninguna capacidad explicativa en el modelo.

En cuanto a las variables que afectan al consumo, se observan efectos significativos para alguna de las variables que informan sobre el nivel de educación y consumo de tabaco. Respecto a la primera, los niveles de educación que tienen capacidad para explicar el gasto en lotería son el Bachillerato y el Grado Universitario, con una significatividad del 5% y el 10%. Estos resultados establecen una relación negativa y muestran que los individuos que declaran tener estos niveles de educación efectúan un menor gasto en comparación al nivel educativo más bajo (Educación Obligatoria).

Por otra parte, el consumo de tabaco, es significativo también para el 5% y el 10%. El coeficiente muestra una relación negativa entre el consumo de tabaco y el de lotería. En otras palabras, se observa que aquellos individuos que son consumidores de tabaco tienen un gasto comparativamente inferior a aquellos que no consumen. En este sentido, esto puede ser debido por un efecto sustitución que provoca el consumo de tabaco y por lo tanto aquellos individuos que fuman, tienen más preferencia por el tabaco que por la lotería.

En cualquier caso, los resultados no son del todo extrapolables y es por eso que se ha

realizado la derivada parcial de la variable dependiente respecto las independientes para captar las elasticidades y el efecto combinado de la probabilidad y la participación para poder realizar un análisis más adecuado.

Efectos marginales de la variable <i>ihs_G_LOTT</i>	
ED	0.019** (0.007)
URB_PROXIMIDAD	0.167 (0.152)
GNR_MASCULINO	0.284* (0.145)
E_CIVIL_CASADO	0.424 (0.196)
E_CIVIL_DIVORCIADO	0.031 (0.317)
E_CIVIL_VIUDO	2.630 (138.44)
EDUC_BACH	-0.679** (0.315)
EDUC_FPMEDIO	-0.075 (0.274)
EDUC_FPSUP	-0.393 (0.240)
EDUC_UNIV_GRADO	-0.452* (0.245)
EDUC_UNIV_SUPERIOR	-0.390 (0.275)
RNT_MES_MEDIA	-0.0004 (0.000)
TBC	0.040 (0.146)
ALC	-0.128 (0.157)

El asterisco (*) muestra la significatividad del coeficiente — 1% (***) , 5% (**), 10% (*).
En paréntesis, se muestra la desviación estándar de cada coeficiente.

Tabla 4: Efectos marginales sobre la variable dependiente

Los resultados obtenidos en esta tabla muestran que la edad, el género del individuo y ciertos niveles educativos tienen capacidad explicativa en el modelo. En cuanto a la edad, siendo significativa al 5% y 10%, se muestra que con un aumento de la misma, el gasto

en lotería aumenta en 0,019 euros. Asimismo, los hombres efectúan un gasto superior en lotería de 0,28 euros en contraposición con las mujeres. Respecto el nivel educativo, aquellos individuos que tienen un nivel máximo de Bachillerato muestran un gasto de 0,68 euros inferior en comparación al nivel educativo más bajo. De la misma manera sucede para los individuos con Grado Universitario, donde el gasto es inferior en 0,45 euros.

Por otra parte, las variables que no muestran significatividad y no tienen capacidad suficiente para explicar el nivel de gasto o participación son las relativas a la proximidad a locales de azar respecto el lugar de residencia, el estado civil, la renta mensual, el consumo de tabaco y alcohol aunque el signo que se obtiene de los coeficientes es el esperable. En cualquiera de los casos, los resultados obtenidos muestran que las variables que afectan a la participación y el nivel de gasto son la edad, el género del individuo y algunos niveles educativos.

6. CONCLUSIONES

En las últimas décadas el mercado de juegos de azar y loterías han experimentado un crecimiento muy elevado en la mayoría de economías desarrolladas. El desarrollo de este mercado ha conllevado efectos negativos derivados del mismo, como por ejemplo la adicción al juego o el juego patológico. Este hecho se considera un importante problema de salud mental en España y en muchos de los países más desarrollados.

Para analizar esta problemática se ha llevado a cabo un estudio que proporciona un análisis sobre los factores condicionantes a la probabilidad de participación y el consumo efectuado en lotería. Para ello, se ha creado un modelo econométrico con un vector de variables de diferente índole como edad, género, nivel educativo, estado civil, consumo de tabaco y alcohol, entre otras.

La estimación se realiza a partir de un modelo de doble valla para lidiar con variables dependientes limitadas. La metodología seguida nos permite obtener y analizar dos procesos estocásticos diferentes: la decisión de participar y la decisión de cuánto consumir en lotería. Además, se ha realizado una modificación de este modelo a partir de la modificación de la variable dependiente, que se ha observado en estudios previos, para solucionar problemas de no normalidad de residuos o heterogeneidad no observada y así obtener estimadores consistentes e insesgados.

Los resultados empíricos obtenidos muestran que existe una diferenciación entre las variables que explican la probabilidad de participar y aquellas que captan el efecto del gasto efectuado. Primeramente, se observa que a niveles más elevados de educación, en comparación al nivel educativo más bajo se percibe una probabilidad más baja de participación y también un menor gasto en lotería.

De la misma manera, se percibe que con un aumento de la edad, mayor probabilidad de participación y mayor nivel de gasto en este tipo de juegos de azar. Además, el género también condiciona la probabilidad de participación y el consumo en lotería, ya que los resultados muestran que el género masculino tiene mayor propensión a la compra de lotería y efectúa un nivel de gasto superior en comparación a las mujeres.

En cualquier caso, las características socioeconómicas, como el nivel educativo, juegan un papel relativamente importante en la explicación del nivel de gasto en lotería. Las diferencias entre los efectos sobre la probabilidad de participación y el nivel de gasto sugieren la importancia del uso de la educación como instrumento fundamental en el diseño de políticas para la reducción de la participación y consumo frecuente en lotería.

Referencias

- [1] Abdel-Ghany M, Sharpe DL. Lottery expenditures in Canada: Regional analysis of probability of purchase, amount of purchase, and incidence. *Family and Consumer Sciences Research Journal*. 2001;30(1):64–78.
- [2] Becoña E. El juego patológico: prevalencia en España. *Salud y drogas*. 2004;4(2):9–34.
- [3] Burke WJ. Fitting and interpreting Cragg’s tobit alternative using Stata. *Stata Journal*. 2009;.
- [4] Chang W. *R Graphics Cookbook*; 2012.
- [5] Clotfelter CT, Cook PJ. On the Economics of State Lotteries. *Journal of Economic Perspectives*. 1990;4(4):105–119.
- [6] Cragg JG. Some Statistical Models for Limited Dependent Variables with Application to the Demand for Durable Goods. *Econometrica*. 1971;39(5):829–844. Available from: <http://www.jstor.org/stable/1909582>.
- [7] Crowley F, Eakins J, Jordan D. Participation, expenditure and regressivity in the Irish lottery: Evidence from Irish household Budget Survey 2004/2005. *Economic and Social Review*. 2012;43(2):199–225.
- [8] García J, Labeaga JM. Alternative approaches to modelling zero expenditure: An application to Spanish demand for tobacco. *Oxford Bulletin of Economics and Statistics*. 1996;58(3):489–506.
- [9] Garrett TA. An international comparison and analysis of lotteries and the distribution of lottery expenditures. *International Review of Applied Economics*. 2001;15(2):213–227.
- [10] George B. The relationship between lottery ticket and scratch-card buying behaviour, personality and other compulsive behaviours. *Journal of Consumer Behaviour*. 2002;2(1):7–22.
- [11] Guryan J, Kearney MS. Is lottery gambling addictive? *American Economic Journal: Economic Policy*. 2010;2(3):90–110.
- [12] Lambert D. Zero-inflated poisson regression, with an application to defects in manufacturing. *Technometrics*. 1992;34(1):1–14.
- [13] Layton A, Worthington A. The impact of socio-economic factors on gambling expenditure. *International Journal of Social Economics*. 1999;26(1-3):430–440.

- [14] Newman C, Henchion M, Matthews A. A double-hurdle model of Irish household expenditure on prepared meals. *Applied Economics*. 2003;35(9):1053–1061. Available from: <https://doi.org/10.1080/0003684032000079170>.
- [15] Oster E. Are All Lotteries Regressive? Evidence from the Powerball. *National Tax Journal*. 2004;57(2):179–187. Available from: <http://www.jstor.org/stable/41790148>.
- [16] Paradis E. R para Principiantes. Ahumada JA, editor; 2002. Available from: https://cran.r-project.org/doc/contrib/rdebuts_es.pdf.
- [17] Pérez Carcedo L. El consumo de lotería: ¿Qué nos dice la evidencia empírica? *RAE: Revista Asturiana de Economía*. 2009;(43):95–115.
- [18] Ronnie Davis J, Filer JE, Moak DL. The lottery as an alternative source of state revenue. *Atlantic Economic Journal*. 1992;20(2):1–10.
- [19] Sociedad Estatal Loterías y Apuestas de Estado, S M E SA. Memoria integrada. Loterías y Apuestas del Estado;. Available from: <https://www.selae.es/es/web-corporativa/responsabilidad-social/memoria-anual/memoria-anual>.
- [20] Vuong QH. Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses. *Econometrica*. 1989;57(2):307–333. Available from: <http://www.jstor.org/stable/1912557>.
- [21] Worthington AC, Brown KA, Crawford MC, Pickernell D. Socioeconomic and demographic determinants of household gambling in Australia. Discussion Paper No 156. 2003;61(0).
- [22] Worthington A, Brown K, Crawford M, Pickernell D. Gambling participation in Australia: Findings from the national Household Expenditure Survey. *Review of Economics of the Household*. 2007;5(2):209–221.
- [23] Yen ST, Jensen HH. Determinants of household expenditures on alcohol. *Journal of Consumer Affairs*. 1996;30(1):48–67.

ANEXOS

Variable	Definición
Variabes demográficas	
ED	Indica la edad del individuo
GNR	Variable ficticia que indica género del individuo. Toma el valor 1 si el individuo es hombre.
URB_PROXIMIDAD	Variable ficticia. Toma el valor 1 si el individuo vive en una localidad donde hay locales de casas de apuestas.
Variabes socioeconómicas	
G_LOTT	Variable discreta. Indica el gasto semanal en lotería efectuado por el individuo.
RNT_MES_MEDIA	Variable discreta. Indica la renta mensual percibida por el individuo utilizando la marca de clase
EDUC	Variable ficticia. Indica nivel de educación del encuestado. Se contempla los niveles de educación siguientes: Educación Obligatoria, Bachillerato, FP Medio, FP Superior, Grado Universitario y Máster y Doctorado en conjunto.
E_CIVIL	Variable ficticia. Indica el estado civil del individuo de la muestra. De esta variable se derivan los diferentes estados (Soltero, Casado, Divorciado y Viudo)
Factores relacionados	
TBC	Variable binaria. Indica si el individuo consume tabaco habitualmente. Toma el valor 1 si el encuestado es fumador.
ALC	Variable binaria. Indica si el individuo consume alcohol habitualmente. Toma el valor 1 si el encuestado es consumidor.

Tabla 5: Variables propuestas para el análisis

1. Indique su edad: _____

2. Indique su género:

■ Masculino

■ Femenino

3. Indique su nacionalidad: _____

4. Indique su lugar de residencia: _____

5. Indique cuantas veces juega a lotería mensualmente (en caso de que no juegue ponga un 0): _____

6. Indique, aproximadamente, su gasto (en euros) semanal en lotería (en caso de que no juegue indique 0):

7. Indique su estado civil:

■ Soltero

■ Casado

■ Divorciado

■ Viudo

8. Indique su máximo nivel de estudios finalizados:

■ Educación Obligatoria

■ Educación superior: Ciclo FP Medio

■ Bachillerato

■ Educación superior: Ciclo FP Superior

■ Educación superior: Universidad (Grado)

■ Educación superior: Universidad (Máster/Posgrado)

■ Educación superior: Doctorado

9. Indique, aproximadamente, su ingreso mensual percibido:

- 0 euros
- De 1 a 500 euros
- De 501 a 900 euros
- De 901 a 1300 euros
- De 1301 a 1700 euros
- De 1701 a 2100 euros
- De 2101 a 2500 eurps
- De 2501 a 2900 euros
- De 2901 a 3300 euros
- De 3301 a 3700 euros
- Más de 3701 euros

10. ¿Actualmente está trabajando?

- Sí
- No

10.1 Si no trabaja:

- Estoy buscando
- No estoy buscando

10.2 Si no está buscando:

- Estoy estudiando
- Estoy prejubilado o jubilado
- No trabajo por otras causas

10.3 Indique si trabaja:

- A tiempo parcial
- A tiempo completo

10.4 ¿Cuántos trabajos tiene?

- 1
- 2
- Más de 2

10.5 ¿Cuántas horas, aproximadamente, trabaja semanalmente?

- Entre 35 y 40 horas
- Entre 20 y 35 horas
- Menos de 20 horas

10.6 Indique su puesto de trabajo: _____

11. ¿Consume tabaco?

- Sí
- No

11.1 En caso afirmativo en la pregunta anterior, ¿Consume tabaco diariamente?

- Sí
- No

11.2 Si ha respondido que NO en la anterior pregunta, ¿Cuántos días, aproximadamente, fuma a la semana?

12. ¿Consume alcohol?

- Sí
- No

12.1 En caso afirmativo en la pregunta anterior, ¿Consume alcohol diariamente?

- Sí
- No

12.2 Si ha respondido que NO en la anterior pregunta, ¿Cuántos días, aproximadamente, bebe alcohol a la semana? _____

Figura 3: Modelo de Encuesta

Participación por Tramos de Renta

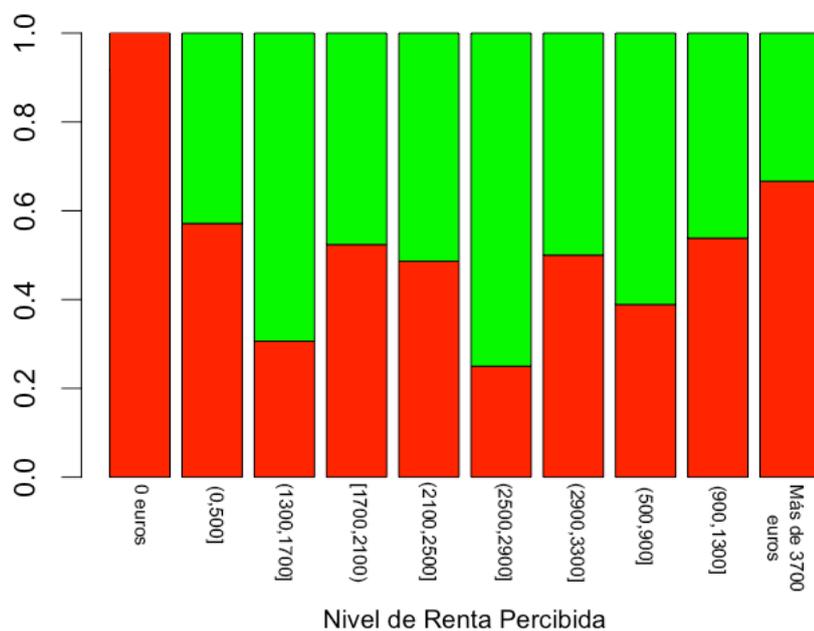


Figura 4: Participación en lotería por tramos de renta

Participación por consumo de Tabaco

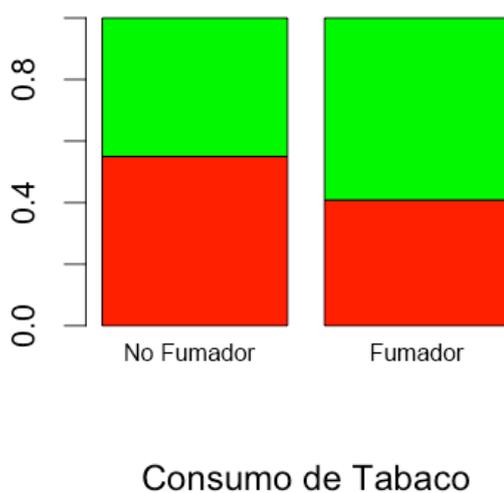


Figura 5: Participación por consumo de tabaco

Participación por consumo de Alcohol

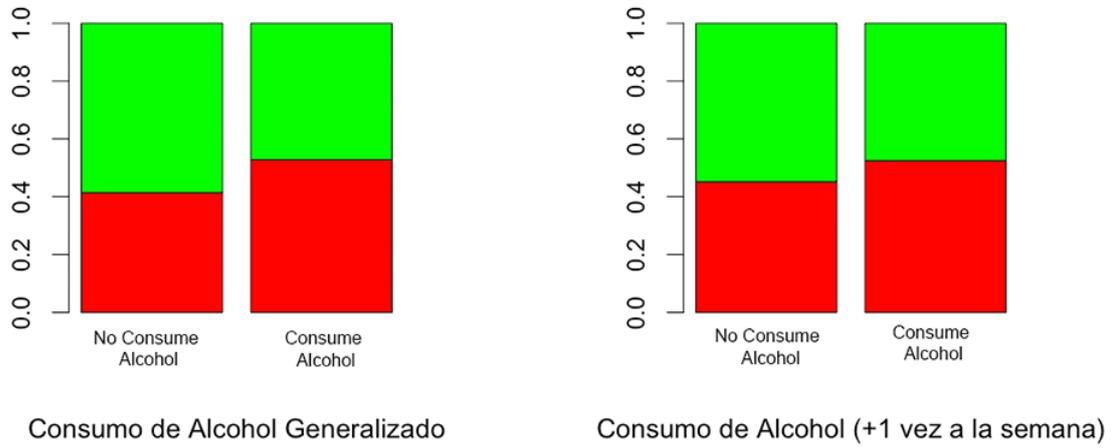


Figura 6: Participación por consumo de alcohol

Participación por Estado Civil

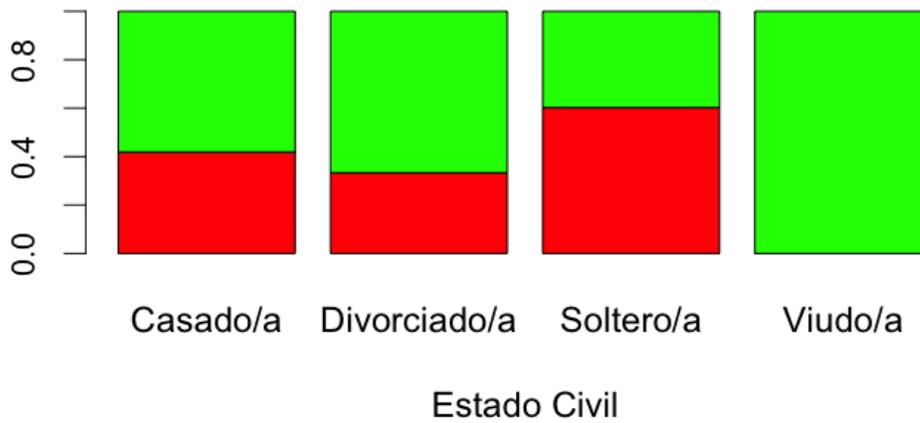


Figura 7: Participación por Estado Civil

Participación por facilidad al acceso a locales de juegos de azar

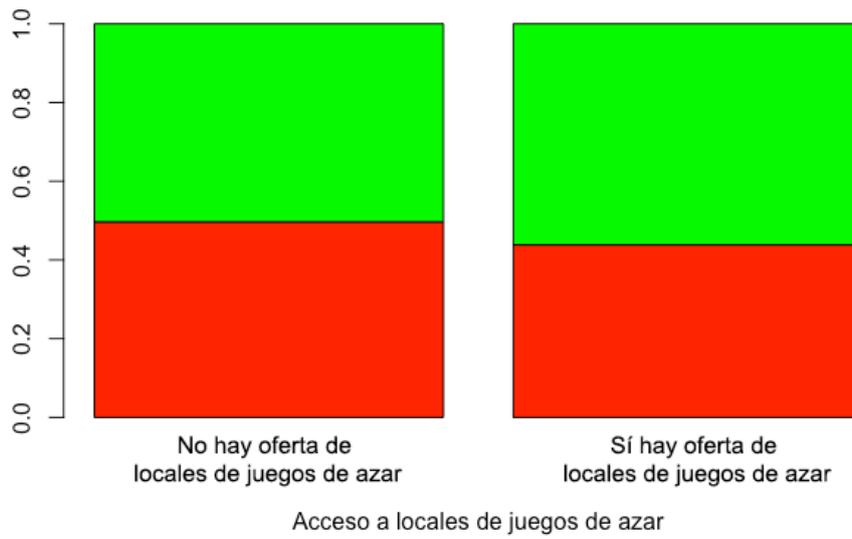


Figura 8: Participación por facilidad al acceso a locales de juegos de azar por lugar de residencia

Distribución provincial de los encuestados

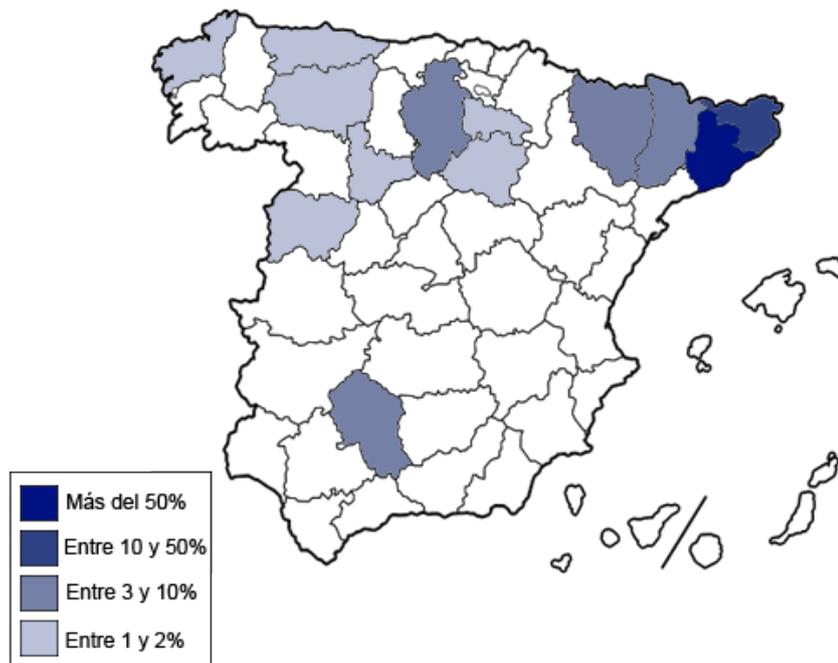


Figura 9: Distribución provincial de la muestra

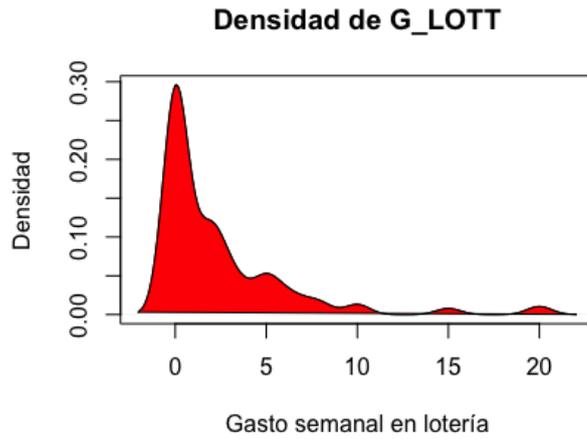


Figura 10: Densidad sobre la variable G_LOTT

Skewness/Kurtosis tests for Normality

Variable	Obs	Pr(Skewness)	Pr(Kurtosis)	adj chi2(2)	joint Prob>chi2
resd_1	220	0.0000	0.1648	18.97	0.0001

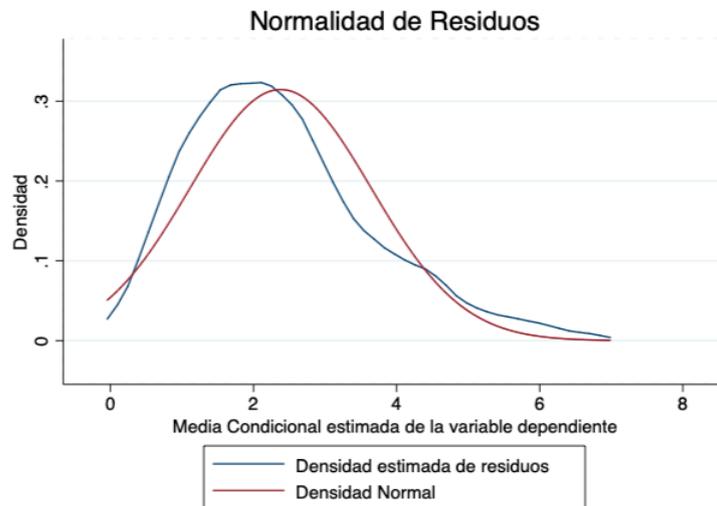


Figura 11: Test de Normalidad

	ED	URB_PR~D	GNR_MA~0	E_C~SADO	E_C~IADO	E_CI~UDO	EDUC_B~H	EDUC_F~0	EDUC_F~P	EDUC_U~0	EDUC~OR	RNT_ME~A	TBC_DEF	ALC_DEF
ED	1.0000													
URB_PROXIM~D	-0.0957	1.0000												
GNR_MASCUL~0	-0.0020	0.1458	1.0000											
E_CIVIL_CA~0	0.4924	-0.0801	-0.0455	1.0000										
E_CIVIL_DI~0	0.2790	-0.0177	-0.0170	-0.3029	1.0000									
E_CIVIL_VI~0	0.2062	0.0027	0.1176	-0.1231	-0.0340	1.0000								
EDUC_BACH	-0.2706	-0.0232	-0.1458	-0.1619	-0.0467	-0.0411	1.0000							
EDUC_FPMEDIO	0.0914	0.1087	0.0282	0.1243	0.0523	-0.0430	-0.1281	1.0000						
EDUC_FPSUP	-0.0567	0.0508	-0.0214	-0.0253	0.0393	-0.0654	-0.1947	-0.2037	1.0000					
EDUC_UNIV~0	-0.1714	-0.0322	0.0845	-0.0894	-0.0068	-0.0671	-0.1996	-0.2088	-0.3173	1.0000				
EDUC_UNIV~OR	0.0867	-0.0760	-0.1060	0.0615	-0.0158	-0.0467	-0.1390	-0.1455	-0.2211	-0.2266	1.0000			
RNT_MES_ME~A	0.4159	-0.0617	0.1436	0.3118	0.0802	0.0156	-0.3144	-0.1200	-0.1020	0.1714	0.4158	1.0000		
TBC_DEF	0.0646	0.1146	0.2275	0.0344	-0.0302	0.1124	-0.0451	0.0961	-0.0253	-0.0894	-0.1507	-0.0839	1.0000	
ALC_DEF	-0.0965	-0.0708	0.2748	-0.2143	0.0673	-0.0063	0.0116	-0.0686	0.0296	-0.0086	-0.1042	-0.1084	0.1651	1.0000

Tabla 6: Matriz de correlación

Skewness/Kurtosis tests for Normality

Variable	Obs	Pr(Skewness)	Pr(Kurtosis)	adj chi2(2)	joint Prob>chi2
r	220	0.0021	0.9821	8.64	0.0133

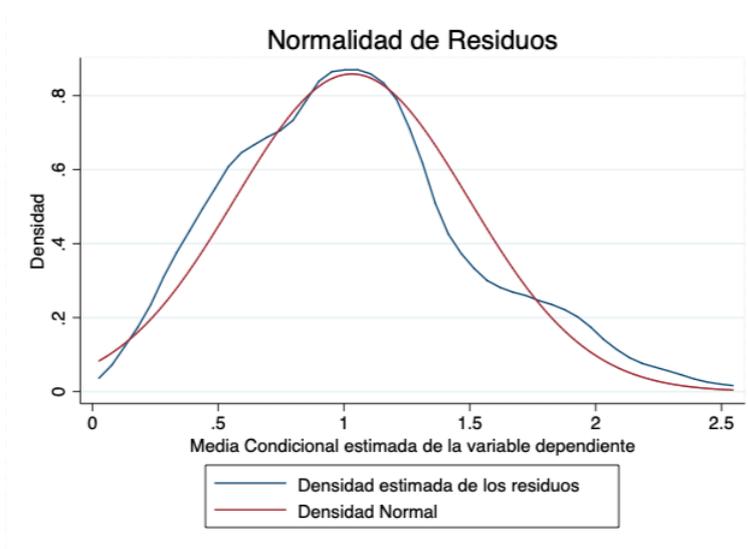


Figura 12: Normalidad Corregida

Cragg hurdle regression		Number of obs	=	220
Log likelihood = -140.32077		LR chi2(14)	=	56.68
		Prob > chi2	=	0.0000
		Pseudo R2	=	0.1680

ihsg_LOTT	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
ihsg_LOTT						
ED	-.0003997	.0038867	-0.10	0.918	-.0080175	.0072181
URB_PROXIMIDAD	.0947208	.0773737	1.22	0.221	-.0569289	.2463706
GNR_MASCULINO	-.0074623	.0805618	-0.09	0.926	-.1653606	.1504359
E_CIVIL_CASADO	.1547638	.0969199	1.60	0.110	-.0351957	.3447234
E_CIVIL_DIVO~0	.0686493	.1466954	0.47	0.640	-.2188684	.356167
E_CIVIL_VIUDO	.1924321	.2525632	0.76	0.446	-.3025826	.6874468
EDUC_BACH	-.3703623	.1672555	-2.21	0.027	-.698177	-.0425476
EDUC_FPMEDIO	-.1399733	.1223988	-1.14	0.253	-.3798706	.099924
EDUC_FPSUP	-.1815908	.1204217	-1.51	0.132	-.4176131	.0544314
EDUC_UNIV_GR~0	-.2878255	.12408	-2.32	0.020	-.5310177	-.0446332
EDUC_UNIV_SU~R	-.1639634	.1568823	-1.05	0.296	-.4714471	.1435202
RNT_MES_MEDIA	-.0000423	.0000672	-0.63	0.529	-.0001739	.0000894
TBC_DEF	-.1641534	.0806825	-2.03	0.042	-.3222883	-.0060185
ALC_DEF	.0936852	.0849913	1.10	0.270	-.0728947	.260265
_cons	.7796433	.2117623	3.68	0.000	.3645967	1.19469
selection_ll						
ED	.0276813	.0102471	2.70	0.007	.0075973	.0477653
URB_PROXIMIDAD	.1239113	.1996748	0.62	0.535	-.2674441	.5152667
GNR_MASCULINO	.4286148	.1937815	2.21	0.027	.04881	.8084196
TBC_DEF	.270087	.1894866	1.43	0.154	-.1012999	.6414739
ALC_DEF	-.3088042	.2065347	-1.50	0.135	-.7136048	.0959963
E_CIVIL_CASADO	-.1361445	.2593569	-0.52	0.600	-.6444748	.3721858
E_CIVIL_DIVO~0	-.0430998	.4277323	-0.10	0.920	-.8814396	.79524
E_CIVIL_VIUDO	3.628427	203.996	0.02	0.986	-396.1963	403.4532
EDUC_BACH	-.5243901	.4095156	-1.28	0.200	-1.327026	.2782456
EDUC_FPMEDIO	.0693344	.3715022	0.19	0.852	-.6587965	.7974653
EDUC_FPSUP	-.3464529	.31899	-1.09	0.277	-.9716618	.278756
EDUC_UNIV_GR~0	-.2969724	.3259272	-0.91	0.362	-.935778	.3418332
EDUC_UNIV_SU~R	-.3646024	.3529902	-1.03	0.302	-1.056451	.3272457
_cons	-1.073828	.5295077	-2.03	0.043	-2.111644	-.0360121
lnsigma						
_cons	-1.021284	.0665171	-15.35	0.000	-1.151655	-.8909129
/sigma	.3601322	.0239549			.3161131	.410281

Tabla 7: Output de Resultados en STATA

Expression : Conditional mean estimates of dependent variable, predict()
 dy/dx w.r.t. : ED URB_PROXIMIDAD GNR_MASCULINO E_CIVIL_CASADO E_CIVIL_DIVORCIADO E_CIVIL_VIUDO EDUC_BACH
 EDUC_FPMEDIO EDUC_FPSUP EDUC_UNIV_GRADO EDUC_UNIV_SUPERIOR RNT_MES_MEDIA TBC_DEF ALC_DEF

	Delta-method				
	dy/dx	Std. Err.	z	P> z	[95% Conf. Interval]
ED	.0184377	.0074154	2.49	0.013	.0039037 .0329717
URB_PROXIMIDAD	.1666329	.1516825	1.10	0.272	-.1306593 .4639251
GNR_MASCULINO	.2843769	.1453818	1.96	0.050	-.0005662 .5693201
E_CIVIL_CASADO	.0424678	.1958083	0.22	0.828	-.3413094 .426245
E_CIVIL_DIVORCIADO	.0305718	.3172679	0.10	0.923	-.5912619 .6524055
E_CIVIL_VIUDO	2.630121	138.4422	0.02	0.985	-268.7115 273.9718
EDUC_BACH	-.678614	.3150964	-2.15	0.031	-1.296192 -.0610365
EDUC_FPMEDIO	-.0749199	.2743074	-0.27	0.785	-.6125526 .4627128
EDUC_FPSUP	-.39336	.2398135	-1.64	0.101	-.8633858 .0766658
EDUC_UNIV_GRADO	-.4523536	.2479187	-1.82	0.068	-.9382653 .033558
EDUC_UNIV_SUPERIOR	-.3903165	.2749459	-1.42	0.156	-.9292006 .1485676
RNT_MES_MEDIA	-.0000368	.0000587	-0.63	0.530	-.0001519 .0000782
TBC_DEF	.0402502	.1458514	0.28	0.783	-.2456132 .3261136
ALC_DEF	-.1279322	.1568093	-0.82	0.415	-.4352728 .1794083

Tabla 8: Output de Efectos Marginales en STATA