Universitat Autònoma de Barcelona

Facultat d'Economia i Empresa UAB

Treball de Final de Grau

Facultat d'Economia i Empresa

TÍTOL: The income gradient of gambling in Spain: evidence from the Family Budget Survey (EPF) using microdata.

AUTOR/A: *(GUILLEM LLOBET MESALLES)*

TUTOR/A: *(GABRIEL A. FACCHINI PALMA)*

GRAU: ECONOMIA EN ANGLÈS

DATA: *(30/05/2022)*

Contents

1 Introduction **2**

2 Gambling sector in Spain **3**

3 Review of the literature **5**

4 Data and empirical methodology **8**

4.1 Data 8

4.2 Empirical methodology 18

4.2.1 The Ordinary Least Square Model 19

4.2.2 The Tobit Model 19

5 Empirical findings **21**

6 Concluding remarks **29**

Appendices **31**

1 Introduction

The gambling sector in Spain, is a subject worth studying due to its importance in the GDP and also the amount of jobs that it creates. But all that glitters is not gold, as this type of games can induce a serious addiction problem. In order to avoid that, as people should be free to spend their money in whatever they want. The only weapon remaining to fight that back is to be better informed, and this study is willing to learn more about it.

The study is going to be divided in two different phases, the first one determining what makes a household spend money on gambling, and the second one will aim to find what shifts the amount they spend on it. This is going to be achieved by conducting an Ordinary Least Square regression for the first case. And a Tobit regression, that will censure households that do not gamble, for the second one. Both models will be divided in three different stages, where socioeconomic and macroeconomic variables are going to be added in each one of them, in order to see how the regressand changes.

In particular, what is aimed to be found is the relationship that the response variables have with the income of the household. Obtaining this coefficient will allow to characterize its income elasticity, the type of good that gambling is and the kind of tax that it has, in Spain.

Apart from that, and due to the controls that are going to be added in each stage, the head of the family that maximize both independent variables will be revealed.

The main conclusions obtained are that income is significant and positive in the first model, so each increase of income makes the household more close to gambling. In the other model, the findings showed that gambling is a luxury good, that substantially increases with small changes in income. This increases do not follow a linear function, but one that flattens in higher levels. Do to this characteristics, the tax is progressive, making high income households contribute proportionally more than the ones with lower levels of income. Allowing for an accomplishment of vertical equity.

Lastly, it was found that the only differences between the head that maximize the probability of a household to gamble and the amount spent in this activity, are the gender and the level of education. For a household to gamble it does not matter the gender of the head, while regarding spending households headed by men spend more. The other one is the level of education, that while lower levels of education make households gamble more money, a medium one makes them more prone to spend something on it.

2 Gambling sector in Spain

The Spanish gambling sector is a subject worth studying due to its importance in the economy. To begin with, it accounted for 0,6% of the Spanish GDP in 2020, and gave employment to about 85.361 people. Apart from that, the sector reported a gross gaming revenue (GGR) of 6.771 Mill. Eur (the GGR is the amount played deducted by the amount of prizes). This numbers are much lower than the ones from the previous year, when represented the 0,8% of the GDP and reported a GGR of 10.226 Mill. Eur (a reduction of the 33,79%), becoming the 4th European country with the highest amount, only after Italy, Germany and France [1]; and the 6th if computed as a share of GDP [2]. Finally, and contrary to what would be thought due to the pandemic lockdown, it employed 0,66% more people.

In order to get deeper into the sector, we are going to consider and discuss the data collection made by Gómez and Lalanda 2022. To begin with, the sector should be divided between publicly and privately owned. The public one is the most important in size and is carried out

[1]OPAP. (June 16, 2020). Gross Gaming Revenue (GGR) in select European countries in 2019 (in billion euros) [Graph]. In Statista. Retrieved May 30, 2022, from https://www.statista.com/statistics/586185/gross-gambling-revenue-europe-by-country/

[2]OPAP. (June 16, 2020). Gross Gaming Revenue (GGR) as a share of GDP in select European countries in 2019 [Graph]. In Statista. Retrieved May 30, 2022, from https://www.statista.com/statistics/967875/gross-gambling-revenue-share-gdp-europe-by-country/

mainly through SELAE (78,3%) and ONCE (20%), but we can also find other minor sources like the games carried out at an autonomous community level (0,91%) and Cruz Roja (0,8%).

SELAE is in charge of games like "Lotera de Navidad", "El Niño", "Lotería semanal", "Primitivas" and "Apuestas Mutuas". In 2020 had a GGR of 2.673 mill. Eur (24,23% less than in 2019), levied 533 Mill. Eur in taxes (from which 41 Mill. Eur where special taxes applied on gambling), provided 1.456 Mill. Eur of its profits to the state and 27 to social entities (81% went to social projects, 1% to cultural ones and the remaining 18% to projects related with sports).

ONCE is in charge of games like "Cupones", "Loteras instantneas" and other kind of active games. In 2020 had a GGR of 684 mill. Eur (35,59% less than in 2019), levied 23 Mill. Eur in taxes (none of those came from special taxes applied on gambling, as they are the only ones that do not have to pay them) and provided 60 Mill. Eur of its profits to social causes.

The private should not be divided between companies, but between sources of gambling. Two types can be found, the in-person and the online gambling. The in-person has three main games: betting (40,6%), bingo (37,6%) and casino gambling (21,7%), and generated a GGR of 2.504 Mill. Eur (it decreased by 48,48%), paying about 1.065 Mill. Eur in taxes (from which 638 Mill. Eur came from special gambling taxes).

Finally, the online one generated 882 Mill. Eur (it grew by 13,66%, being the only one that generated higher profits with respect to the previous year), there is no information on the exact total amount of taxes that they paid, what we know is that 140 Mill. Eur were subtracted from their revenues in special gambling taxes.

All those games mentioned above gave a total of 14.363 Mill. Eur in prices, from which the government also levies taxes, accounting for 421 Mill. Eur (about the 68% of total prizes. This source of revenue for the government exists since 2011, as before no taxes were levied on prizes).

Lastly, the gambling sector has been really harmed by the Covid crises, suffering a reduction of the total GGR by 24,2%, even if the online gambling was not affected (it increased). Something remarkable is that the introduction of this new online games, is eating the share of the more traditional ones. Every year increases its share of the total GGR, representing a 4,8% in 2016 and reaching a 13,7% in 2020.

The last thing we should look for, is how does the regulation of this sector work. In order to do so, two different types of regulation should be differentiated. The one regarding the land-based and the one for the digital type; with this last one having three main products

regulated differently: Gaming (Casino games, Poker and Bingo), Betting and Lotteries.

The first one (land-based), is fully regulated by the relevant authority within the competent Autonomous Region where is performed. In the other hand, the digital one is regulated differently depending on the product. "Dirección General de Ordenación del Juego" (DGOJ) regulates Gaming and Betting, and the Spanish State does so in the case of Lotteries.

3 Review of the literature

Gambling is a field that has been studied many times before, specially there are many studies related to lotteries, and that is the starting point of this literature.

Firstly, I am going to discus a paper by Perez and Muñiz-Fernández 2021 that focus on a more general perspective of the game, they argue that lottery is a normal good worldwide that increases its income elasticity with the income quartile (making it a progressive tax), with the exception of Africa where is regressive. They also state that Latin America is the region with the highest range of income elasticity, that goes between 2,012 and 2,634 (the min. is the elasticity of the first quartile and the max. is the one of the third), followed by Asia that goes between 1,054 and 2,179, and Europe between 1,106 and 1,199, having a higher income elasticity in the first percentile than Asia, but a lower one in the highest income group. Finally, Africa's elasticity goes from 0,752 to 0,216, making it regressive as mentioned before.

Now, looking for the Spanish lottery sector, it can be seen that fulfils the general findings mentioned before, having an elasticity of 1,816 or 1,237 depending on the model used to compute it (the first model uses the income, and in the second one they use this income but divided by the number of members) (Humphreys and Perez 2011). Contrary to that, Martínez Martínez et al. 2013 found a regressive income tax, that does not fulfil the previous European findings.

The type of participants that spend more money in this kind of game are mainly middle aged men, whose spending increases as age increases, up to a certain point where it starts decreasing following an inverted U-shape (Humphreys and Perez 2011; Martínez Martínez et al. 2013). Married individuals spend more than singles (Humphreys and Perez 2011), but (Martínez Martínez et al. 2013) argues that this just happens in some games ("La primitiva" and "Bonoloto"). Finally, Humphreys and Perez 2011 mention that the level of education decreases spending, but as before Martínez Martínez et al. 2013 argues that this just happens on a particular game (in this case on "Euromillones").

To end up with the lottery analysis, should be taken into consideration the study of Garrett and Coughlin 2007 that say that total spending should not just be the only think taken into consideration to determine the total amount of lottery purchased, as it differs between different sources of income. Analysing the case of United States, they found that negative marginal spending gives a negative income elasticity of demand, so that lottery tickets are inferior goods contrary to the previous findings of Perez and Muñiz-Fernández 2021.

Apart from that, they found that income from transfer payment, retirement and income maintenance components, have a higher marginal spending than earnings income and wealth income.

Finally, the largest spending comes from transfer income. This kind of income, has a greater deadweight loss for the government (than other sources like earnings or wealth), this happens due to the amount of taxes and spend channels that needs to pass before they are able to allocate in wherever they desire.

Now, is the time to visit the literature of gambling, this one has not been studied as much, so it is not as extensive as the previous one, but much more significant for this study. As we are going to see as we get through it, does not differ much with respect to what we have seen until now (with the case of lotteries).

To begin with we are going to take a look at the Spanish gambling market, clearly and as seen before on the lottery market, men are more prone to gambling, an assumption that is confirmed by all the literature (Leal, López-Laborda, and Rodrigo 2014; Vallejo Gabasa and Rivera Torres 2015; Sarti and Triventi 2012; Kohler 2016), with the exception of Australia, where Worthington et al. 2007 found that gambling expenditure is distributed equally between men and women, and what really matters in order to identify a gender is the game being played, this also happens in Spain, where Vallejo Gabasa and Rivera Torres 2015 found that the "Bingo" game is mainly played by women. Even if women spend more on "Bingo", Vallejo Gabasa and Rivera Torres 2015 state that men double the amount spend by women, reaffirming what was said by the rest of the literature.

Apart from that, Leal, López-Laborda, and Rodrigo 2014 say that per capita GDP of the studied region affects positively to the amount spent on casinos, as it does the amount of tourists that the region receives.

Regarding to the income level, its effects have never been studied in the case of Spain (or at least at the visited literature), so the literature of other countries is the one that is going to be taken into consideration.

The first that is going to be considered is Italy, that in my opinion is the one that could be a closest estimation to our country (due to the living standards of its population), Sarti and Triventi 2012 found that as income increases, absolute spending increases. In contrast to that, in Australia gambling expenditure is unaffected by changes in income (Worthington et al. 2007). Finally, there are countries like Switzerland, that found it to be regressive (Kohler 2016).

The Spanish age characterization performed by Vallejo Gabasa and Rivera Torres 2015 shows us that population between 18 and 25 even if they play a lot they do not spend much, while population between 46 and 55 play less, but their amount spend is 7 times higher than the one of the youngest group. The traditional games player is aged between 26 and 35 years old, with the exception of "Bingo" that is a game more related to older people, with a player aged between 36 and 45 years old. In the other hand, people aged between 18 and 45 years old, are more into online gambling. Taking a look into other countries, in Australia gambling is performed by elderly population, while in Switzerland with a similar age as the one of the Spanish market, where spending also peaks on older individuals (Worthington et al. 2007; Kohler 2016).

The literature agrees that the level of education negatively affects the amount gambled (Leal, López-Laborda, and Rodrigo 2014; Sarti and Triventi 2012; Kohler 2016).

It also agrees that the introduction of online gambling has negatively affected the most traditional types of gambling (Leal, López-Laborda, and Rodrigo 2014; Vallejo Gabasa and Rivera Torres 2015).

Finally, the Spanish literature, or at least the one revised to do this study has not taken into consideration the role of household structure and composition, Leal, López-Laborda, and Rodrigo 2014 just found that the unemployment rate decreases the amount gambled (something also argued by Garrett and Coughlin 2007 for the United States lottery market). As this study wants to do a deeper research on it I think it's worth giving it some lines with the findings of studies from other countries. In the case of Australia, persons living alone, self-employed, households with pensions as the principal source of income, the ones with a higher amount of dependent people and couples or lone parents with children spend less on those kind of products (Worthington et al. 2007). In the case of Switzerland the variables previously mentioned do not have a significant effect (Kohler 2016).

The last thing that this literature is willing to take a look at is how does crisis periods affect this sector, this is interesting for the paper as the years that comprise this study have suffered the pandemic, that had an economic crises as an effect. In order to do so, we are going to

observe the findings that Chóliz and Manzón 2013 achieved by studying the Spanish sector after the 2008 crises.

To begin with, Chóliz and Manzón 2013 and corroborated by the findings seen before by Leal, López-Laborda, and Rodrigo 2014, per capita GDP increases the gambling expenditure. Then in crisis periods due to the decrease of the richness of the country, and as gambling is not an essential good but a consumption one, a decline on the total amount spend on this type of goods can be observed. Finally, a needed remark is that some games even though the crises are being more played. This is the case of active lottery games, that experience a very little increase and of gambling from type "B" machines that had a sharp increase. This last one is associated with pathological players, that due to their illness crisis periods do not reduce their necessity to gamble.

As far as I am aware, this paper represents the first attempt to test the determinants of gambling expenditures in Spain as well as its income elasticity, as it has just been done regarding the Lottery spending. Apart from that, this paper is interesting as it is going to have two different models, one that tries to find what makes a household spend money on gambling, and the other one that will show how much are they going to spend on it.

4 Data and empirical methodology

4.1 Data

To conduct the analysis a representative survey of the Spanish population was used, the Family Budget Survey (EPF). This is a yearly survey carried out by the INE, and representing a sample of 24.000 dwellings. For the purpose of this study, I have taken into consideration the ones done between 2016 and 2020, having a total of 104.739 entries. From which 70.476 are men and 34.263 women, comprising people aged between 16 and 85 years old (it needs to be taken into consideration that people 85 or older are all categorised at 85 years).

This survey is presented in 3 different documents, the first one has the data from all the spending categories (for the sake of this study, the only one taken into consideration is the gambling spending, that had the code "09430"), the second one collects socioeconomic data about the household and his main breadwinner; and the last one, socioeconomic data about the member of the household that is being surveyed (The present study is has not used this document).

The datasets presented before needed some modifications, and the addition of some other

variables, so that the sample could be more representative. The ones modified were the variables "GASTO" (yearly amount spend on gambling by the household) and "GASTOT" (yearly total spending of the household), that were raised to the population factor, so they required to be divided by the variable "FACTOR" (population factor) in order to achieve the result of a single household. At the same time, the variable "IMPEXAC" (net income of the household) was transformed from monthly to annual. This variable did not need any modification, as it changes nothing if it is monthly or annual form, but as all the other variables were computed annually, its interpretation will be more appealing. There is just one that it is not computed annually, the variable "COMIMH" (amount of meals eaten by the household members), that is biweekly, .

The added data is more macro/general and inform about the autonomous communities where the households are located. Firstly, the per capita gdp and the unemployment rate of each autonomous community were added (variables taken from INE). Apart from that, the other added refer to the amount of casinos, "salas de bingo", "salon de juegos" and "salas de apuestas" that the regions have (variables taken from "Dirección General de Ordenación de Juego" [3]).

Finally, it was observed that there were 697 entries that had no income, but they still had an entry on the gambling spending ("GASTO") and on total spending ("GASTOT"). So due to the fact that this is not sustainable over time, and it may represent a misinformation when the survey was conducted, they were taken away. This produced a reduction of the sample from 105.436 observations to the 104.739 mentioned at the beginning of this section.

All this variables added and the others used from the INE surveys can be seen on the following summary statistics table. This table shows its minimum, the 1st quartile, the median, the mean, the 3rd quartile (the dummy variables of both the years and the autonomous communities are not included).

Statistic	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Gambler	0.00	0.00	0.00	0.33	1.00	1.00
Gambling spending	0	0	0	161	35	40986
Income	0	14484	21432	25686	32760	264996
Total spending	0	17447	25716	29752	37620	299482
Age	16	45	55	56	67	85
Male	0.00	0.00	1.00	0.67	1.00	1.00

[3]Ordenacionjuego.es. 2021. Informe anual del juego en España 2020 — Dirección General de Ordenación del Juego. [online] Available at: ¡https://www.ordenacionjuego.es/es/noticia-datos-MJ-ESP-2020¿

Citizenship						
Spanish citizenship	0.00	1.00	1.00	0.93	1.00	1.00
Foreign citizenship	0.00	0.00	0.00	0.05	0.00	1.00
Dual citizenship	0.00	0.00	0.00	0.02	0.00	1.00
Level of education						
Less 1st stage of sec. educ.	0.00	0.00	0.00	0.19	0.00	1.00
1st stage of secondary educ.	0.00	0.00	0.00	0.29	1.00	1.00
2nd stage of secondary educ.	0.00	0.00	0.00	0.18	0.00	1.00
Higher education	0.00	0.00	0.00	0.33	1.00	1.00
Activity situation						
Working	0.00	0.00	1.00	0.55	1.00	1.00
Retired	0.0	0.0	0.0	0.3	1.0	1.0
Studying	0	0	0	0	0	1
Other situation	0.00	0.00	0.00	0.04	0.00	1.00
Full time	0.00	1.00	1.00	0.96	1.00	1.00
Source of income						
Freelance	0.00	0.00	0.00	0.12	0.00	1.00
Employee	0.00	0.00	0.00	0.47	1.00	1.00
Other sources	0.00	0.00	0.00	0.41	1.00	1.00
Working position						
Army	0.00	0.00	0.00	0.04	0.00	1.00
Manager/Director	0.00	0.00	0.00	0.04	0.00	1.00
Scientific	0.00	0.00	0.00	0.15	0.00	1.00
Technical/supporting	0.0	0.0	0.0	0.1	0.0	1.0
Accountant/Administrative	0.00	0.00	0.00	0.08	0.00	1.00
Restaurant/Security/Retailer	0.00	0.00	0.00	0.17	0.00	1.00
Agriculture/forestry/Fishing	0.00	0.00	0.00	0.04	0.00	1.00
Manufacturing/Construction	0.00	0.00	0.00	0.16	0.00	1.00
Plant/Machinery operators	0.0	0.0	0.0	0.1	0.0	1.0
Elementary occupations	0.00	0.00	0.00	0.12	0.00	1.00
Marital status						
Single	0.00	0.00	0.00	0.19	0.00	1.00
Married	0.0	0.0	1.0	0.6	1.0	1.0
Widower	0.00	0.00	0.00	0.12	0.00	1.00
Separated	0.00	0.00	0.00	0.03	0.00	1.00
Divorced	0.00	0.00	0.00	0.07	0.00	1.00
Household members	1.0	2.0	2.0	2.6	4.0	16.0
Household children	0	0	0	1	1	10
Tenure regime						
Without mortgages	0.00	0.00	1.00	0.53	1.00	1.00
Mortgage	0.00	0.00	0.00	0.28	1.00	1.00

Rent	0.00	0.00	0.00	0.13	0.00	1.00
Meals	0	36	56	62	84	352
Capital	0.00	0.00	0.00	0.34	1.00	1.00
Density of population						
Highly populated	0.00	0.00	0.00	0.47	1.00	1.00
Intermediately populated	0.00	0.00	0.00	0.24	0.00	1.00
Depopulated area	0.00	0.00	0.00	0.28	1.00	1.00
Per capita GDP	17117	20436	23130	24758	29727	36049
Unemployment rate	8.2	12.2	15.1	16.7	20.6	31.4
Casino	0	1	2	3	4	10
"Bingo"	1	7	16	24	36	62
"Salón de juegos"	5	97	161	252	366	897
"Sala de apuestas"	0	6	18	38	47	194

Table 1: Summary statistics

With the summary statistics table being presented, some explanations should be done so that the dataset can be better understood. To begin with, we can observe that there are more households that do not gamble, than households that do it. Apart from that, looking to the amount spent on it, we can observe that the 3rd quartile of the sample spends 35 Eur./year, while the mean is 161, this shows us that the majority of the sample spends really few money or anything (also as the 1st quartile and the median is located at 0) but the few households that spend money spend a lot, peaking at 40.986 Eur./year.

Another variable that should be analysed is the household net income. The lower class earn up to 14.484 Eur./year (1st quartile), the middle class about 25.686 (mean) and the high class, 32.760 (the 3rd quartile). Finally, the household that earn the most, bring in 264.996.

As explained before socioeconomic variables of the main breadwinner are used. We are firstly going to consider their sex and age. The sample age ranges between 16 and 85 years old with an average of 56, and has a 67% of males. The 59% of this individuals are married, the 19% singles, the 12% widowers and the 10% separated/divorced. Regarding their citizenship, almost everyone has the Spanish one (93%), followed by the ones with foreign (5%) and dual citizenship (2%). The majority being highly educated, but followed closely with people that just studied until the 1st stage of secondary education.

Now, looking at their job/occupation, when they were surveyed they were mainly working, they have full time positions (96%) and their earnings come from a wage for being employed by someone else (47%, just 12% comes from self-employment). The other sources of income

variable includes contributory pensions, unemployment benefits, capital and property rents..., and represents the 41%.They work at the restaurant, security and retailing industry, in manufacturing and construction, in scientific roles and in elementary occupations (ordered from the one with the highest mean to the lowest). Even though the majority was working, it should be pointed out that the retired ones represent a large chunk, that also relates with why the other sources of income are that important.

Some remarks about the household should be done, the most common ones have between 2 and 4 members (the mean is 2,6) and its tenure regime is 52% owning its dwelling without a mortgage, the 28% are paying one and the 13% pays a rent. Almost half of the sample are located on a highly populated area, and the rest are divided almost equally between intermediately and depopulated areas. Lastly, the autonomous communities they live in have an average per capita GDP of 24.758 and an unemployment rate of the 16,7%.

The last part that is going to be discussed in this section are the check-ups done in order to ensure that the surveys are representative, and that is worth to study them. The first one being a comparison between the average yearly total spending from the INE's average yearly gambling spending (a variable that I called "MEDIA") and my computation of the average per year total spending of the survey (the yearly average of the variable "GASTO" with code "09430"). This comparison gave us that the total spending from the survey between the 2016-2020 period was 4%, 2,8%, 2,04%, 2% and 3,03% higher than the one from the INE averages respectively. That showed a not really significant difference between the dataset and the INE average corroborating that the survey has a pretty good degree of accuracy.

The last check-up that was done, was summing up all the "GASTO" (yearly spending of each code) codes of each household from the first document (mentioned at the beginning of the section), and compare it with the variable "GASTOT" (yearly total spending of the household) in order to see if they were the same, and they were.

Having all the data presented and explained above, now is time for a descriptive analysis for both a further and more graphical explanation of the sample and for having a first intuition of the results that are expected to be obtained on the empirical analysis that it is going to be conducted and discussed in the next section. All the graphs that are going to be presented have been made for this specific paper.

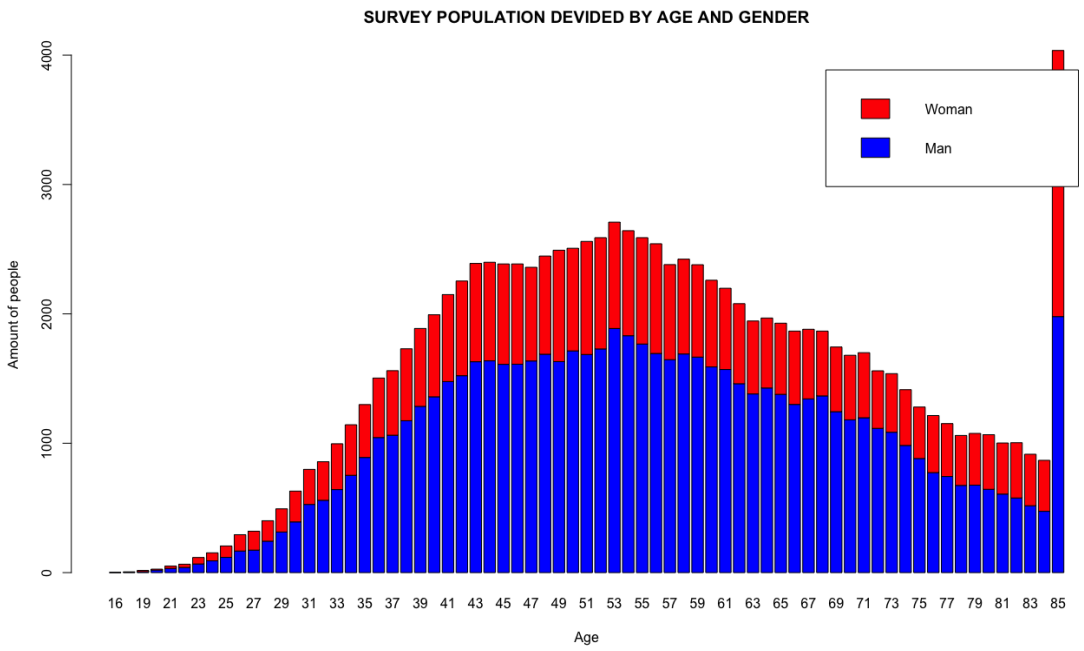

Figure 1: Distribution of the sample by the age and gender of the main breadwinner.

Figure 1 is a histogram that divides the sample by age and gender. It can be observed that households are headed mainly by men, with an age ranging between 16 and 85 years old (as mentioned before people aged 85 or older are categorized all at 85 years old), with a mean age between 56 and 57. If we divide it by sex, men have a mean age between 55 and 56 and women between 56 and 57. This reaffirms the information explained before from the variables age and male.

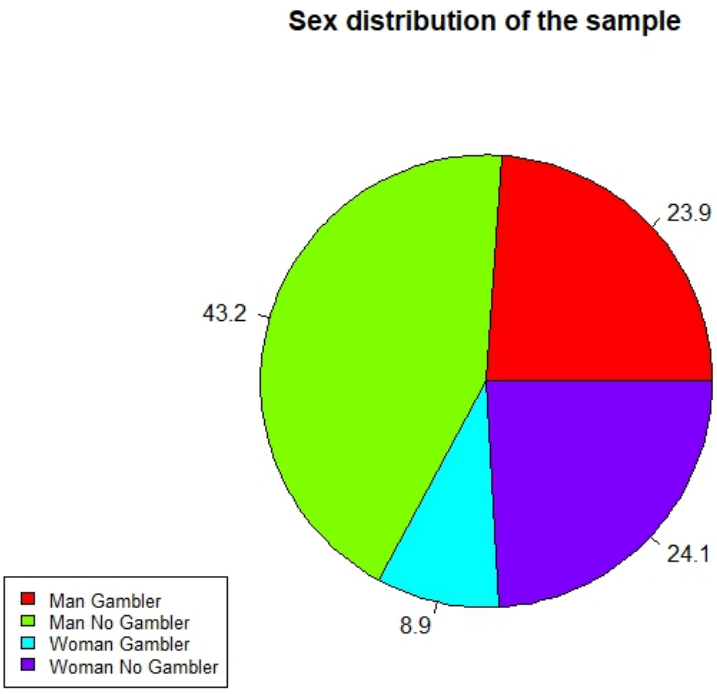

Figure 2: Distribution of the sample by gender of the main breadwinner and by if if its household gambles.

Figure 2 is a pie chart that also represents the gender division of the previous graph, but now is plotted against a dummy variable showing if they are gamblers or not. It can be observed that 32,7% of the sample gambles (and 67,3% does not), with 22,7 % of women and 35,62% of men doing it (this was obtained by dividing the number of men/women that gamble by total amount of men/women), so men are more prone to spend money on this goods.

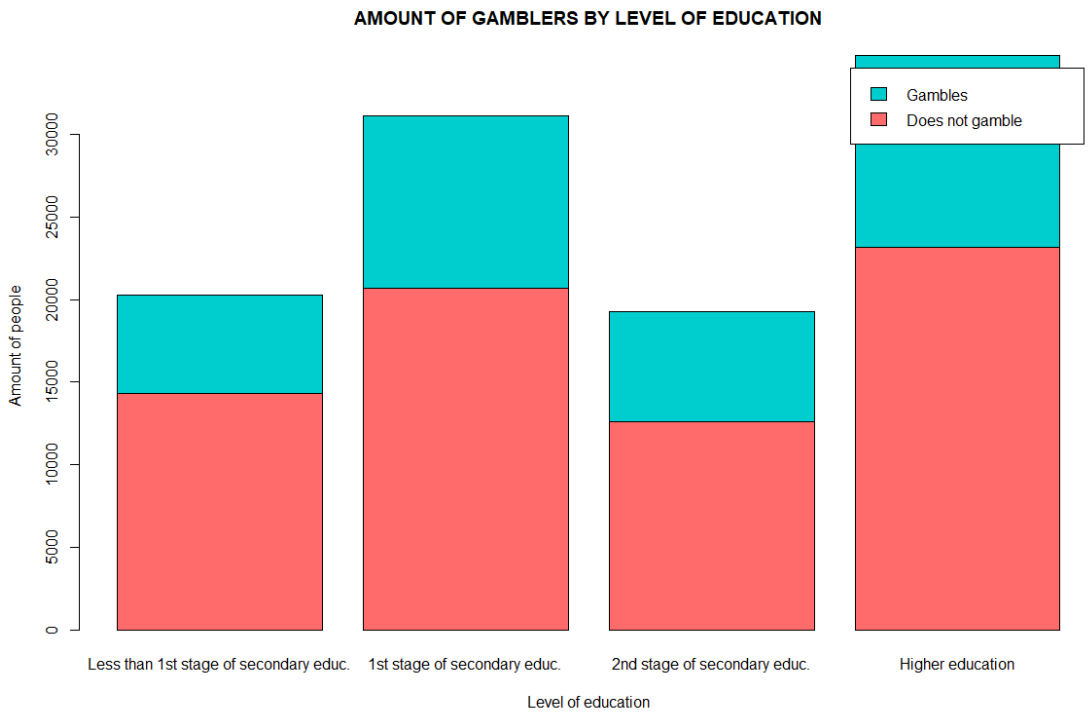

Figure 3: Division of households that gamble by the level of education of the main breadwinner.

Figure 3 is a barplot that divides the sample in 4 different levels of education (from the most basic to the most advance), this levels are the following: Less than secondary education, First stage of secondary education, Second stage of secondary education and Higher education. This is done in order to see how the amount spent on gambling increases/decreases as the level of education of the main breadwinner is higher/lower. It can be observed that the group with the lowest level of education is the one that gambles the least, with just a 29% of its members doing so. Then, it increases in the second and third level to 34% and 35% respectively. And finally, the group with the highest education gamble less than both of the middle education groups mentioned before, doing it the 33%.

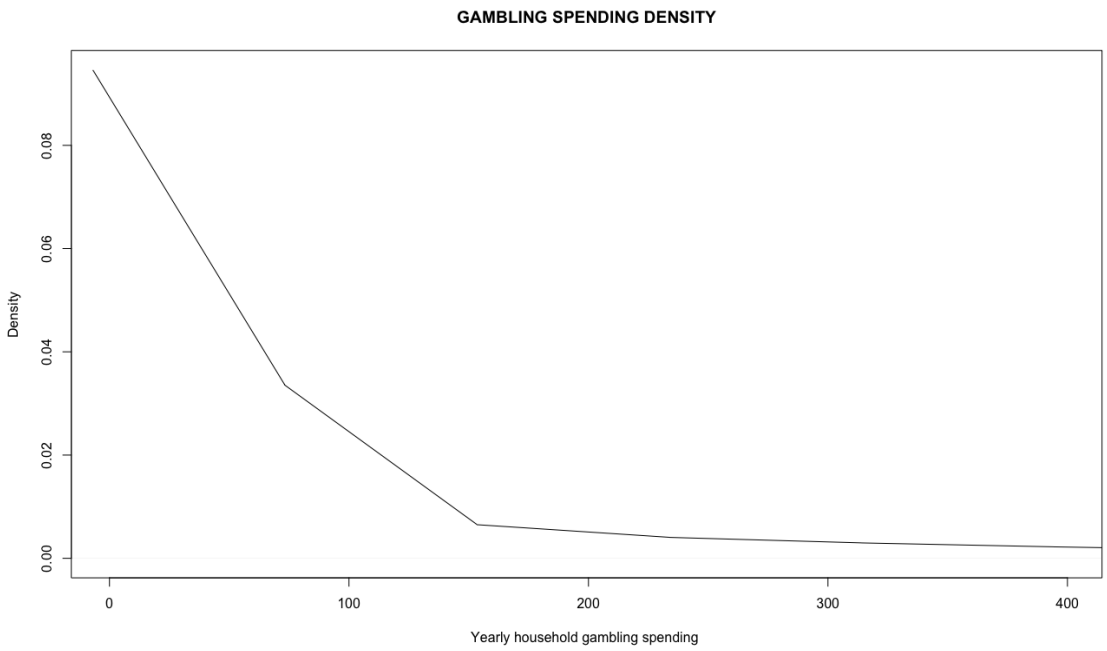

Figure 4: Density histogram of the yearly household gambling spending (percentile 0-90%).

Figure 4 is a density histogram that shows the density of people with each different amount of gambling spending. An important remark to be made, is that this histogram just takes into consideration the spending levels that range between the deciles 0% and 90%. This was done, as there were really high amounts spent in the last decile that made the x axis have a really large range, making it impossible to see the changes of density of the left side, that now can be clearly seen. What it can be observed is that the density of people that spends 0 is the highest, then this sharply decreases until at about a spending of 75 where it flattens a little bit, and finally, at about 175 it becomes almost constant with very little density.

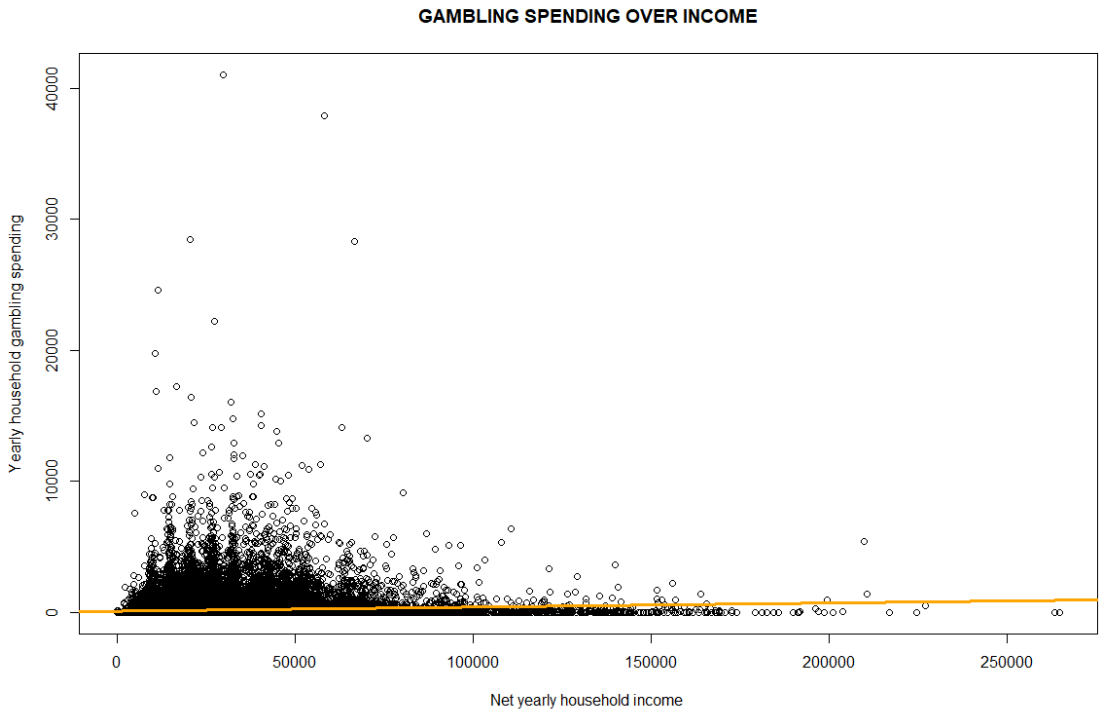

Figure 5: Yearly gambling spending over net yearly income of the household.

In Figure 5, you can observe a scatter plot of the yearly income of the household against the yearly spending on gambling. The result shows that those two variables have a strong positive correlation.

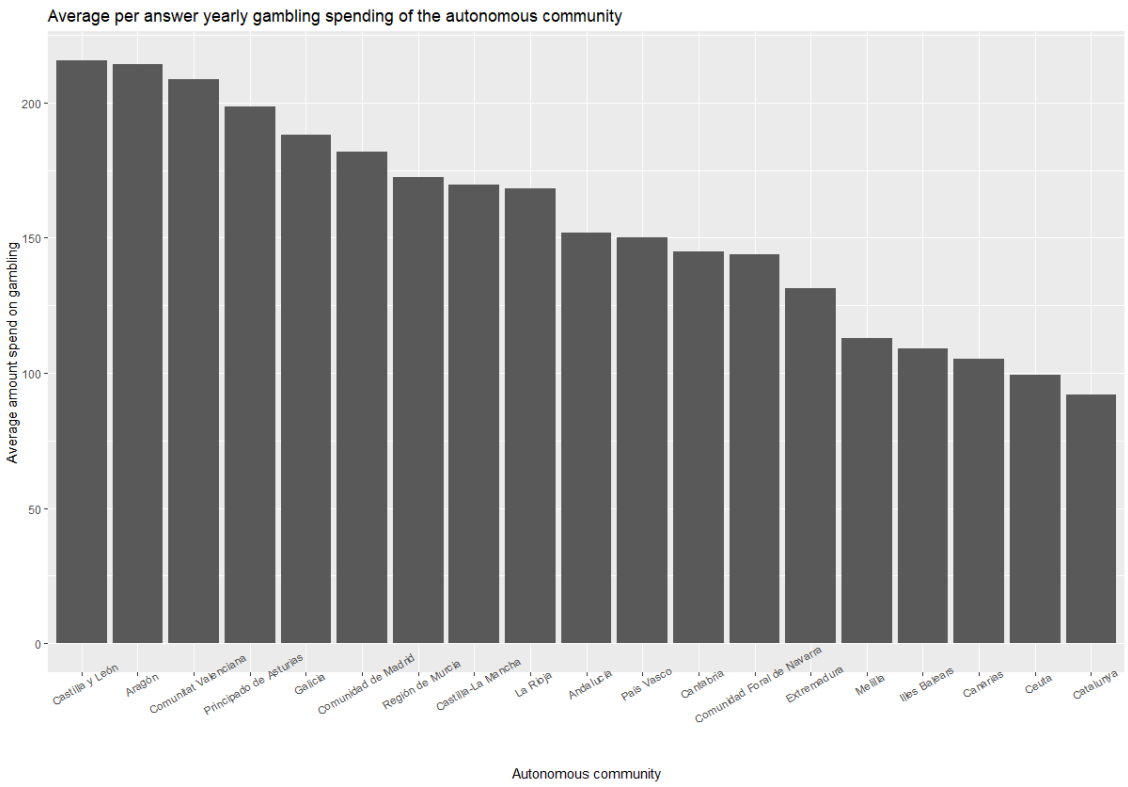

Figure 6: Average per respondent spending on gambling of each autonomous community.

In Figure 6, you can observe a barplot that represents the average amount that the respondents of each autonomous community spend on gambling (Total amount spent by a community/number of respondents of that particular community). The communities are ranked from the one that spends the most to the one that spends the least this being "Castilla y León" and "Catalunya" respectively, with the first more than doubling the other. The other communities are ranked as follows: "Aragón", "Comunitat Valenciana", "Principado de Asturias", "Galicia", "Comunidad de Madrid", "Región de Murcia", "Castilla La Mancha", "La Rioja", "Andalucía", "País Vasco", "Cantabria", "Comunidad Foral de Navarra", "Extremadura", "Melilla", "Illes Balears", "Canarias" and "Ceuta".

4.2 Empirical methodology

This analysis is going to be divided in two different stages, firstly it is going to be studied what makes someone gamble, this is going to be done by running an OLS regression. The second stage will try to define what shifts the amount gambled, this one will run a Tobit regression.

Both the OLS and Tobit models presented in this section where obtained from Wooldridge 2009.

4.2.1 The Ordinary Least Square Model

The Ordinary Least Square (OLS) regression estimates the relationship between the independent variable and the dependent variables of an econometric model. This estimation is obtained by minimizing the sum of squared residuals. This model can be estimated by a simple or multiple regression (as our case), depending on the number of regressors.

This model can be written as follows:

$$\mathrm{y} = \beta_0 + \beta_1 \mathrm{x}_1 + \beta_2 \mathrm{x}_2 + ... + \beta_k \mathrm{x}_k + \mathrm{u},\ \mathrm{u} \sim N(0, \sigma^2) \tag{1}$$

where y is the dependent variable, β_0 the intercept of the model, x are the k independent variables of the model and u is the error term, with expectation 0 and variance σ^2.

In the case of this study the model incorporates nonlinearities, this being logarithmic and quadratic explanatory variables.

4.2.2 The Tobit Model

The Tobit model, is an econometric model that was developed by James Tobit (1958) in order to handle cross sectional data samples whose dependent variable is a corner solution response. Such variable is zero for a nontrivial fraction of the population but is roughly continuously distributed over positive values.

Assuming that the error (u) is a random variable, the model takes into consideration a y that is continuous over strictly positive values (for the observations of the households that spend money on gambling) but that takes on a value of zero with positive probability (for the observations of the households that do not consider gambling spending as utility maximizing).

A Tobit estimation in this case is better than an OLS, as the last one would give us negative fitted values, implying negative predictions for y. The assumption of an explanatory variable appearing in level form having a constant partial effect on E(y|x) can be misleading. Apart from that, Var(y|x) would probably be heteroskedatic. As the distribution of y piles up at zero, y cannot be a conditional normal distribution. Making all the inference have only asymptotic justification.

Then, if we would like to just do an OLS regression but truncating the data for the first type of observations, we would have an omitted variables problem, making it biased and inconsistent. Apart from that, the error would have a mean different to zero and would

be heteroskedastic. We can explain the existence of the biasing due to the fact that by considering only the first type of observations, E(u) may be different to zero, so we cannot guarantee the estimator to be unbiased.

This model can be written as follows:

$$y^* = \beta_0 + \mathbf{x}\boldsymbol{\beta} + u,\ u|\mathbf{x} \sim N(0, \sigma^2) \tag{2}$$

$$y = \begin{cases} y^*\ if\ y^* > 0 \\ 0\ if\ y^* \leq 0 \end{cases}$$

where y^* is a latent variable, that depends on a set of exogenous variables x via a vector β, which determines the relationship between exogenous variables x and the latent variable y^*.

In this case the model assumes that the effect of the exogenous variables on the probability that an observation is censored and the effect on the conditional mean of the non-censored observation are the same. In other words, the income of the household affects equally the decision and the amount of gambling.

Both models are going to have different dependent variables, while the OLS regression uses a dummy variable that shows if the household spends money gambling or they do not. The Tobit model uses a variable that shows the yearly amount that the household spends on it.

Both models are going to have the same independent variables, and are going to be divided in three different stages (one regression will be run in each of them):

The first one, takes into consideration the fixed effects of the autonomous communities, the year, the monthly income of the household and its quadratic variable.

The second one, apart from the variables mentioned before, also includes some socioeconomic variables of the main breadwinner and its household. Like its age and gender, which is the main source of income, his/her citizenship, his/her level of studies, his/her activity situation, the type of position that it has, his/her type of working day, the tenure regime of the dwelling, both the amount of members and of kids of the household, the marital status and the amount of meals that the household does (biweekly).

Finally, the third adds some more macro-economic data regarding the autonomous communities, the location of the household and takes into consideration the different types of spots where people can gamble. These variables explain if the household is located in a city that is

province capital, the density of population of the city they live in, the GDP per capita and the unemployment rate of each autonomous community, and the amount of casinos, "salas de bingo", "salon de juegos" and "salas de apuestas" that it has.

The last remark that should be considered, is that all the variables that involve an amount of money, are being used in its logarithmic form. This are, the net yearly income of the household and its quadratic variable, the yearly total amount spent by the household and the per capita GDP of the autonomous community (In the tables representing the results of the regressions, can be clearly seen).

5 Empirical findings

In the table below, you can see the results of the Ordinary Least Square regressions (the table is just a reduced version, that can be fully seen in Table 5 of the appendices), that were previously explained and discussed in the empirical methodology section. This model is trying to determine what makes a household gamble, and does so by adding controls in three different stages (the types of variables added in each step, can be seen below the R^2 in the table).

Table 2: Results of the OLS models (summary)

	Dependent variable:		
	Gambler		
	(1)	(2)	(3)
Income (log)	0.370***	0.270***	0.250***
	(0.044)	(0.045)	(0.048)
Quadratic income (log)	−0.012***	−0.014***	−0.013***
	(0.002)	(0.002)	(0.002)
Observations	104,739	104,173	91,374
R^2	0.057	0.110	0.110
Main breadwinner socioeconomic data		Yes	Yes
Household socioeconomic data		Yes	Yes
Macro data of the area of residence			Yes

Note: Robust standard errors are in parentheses. * denotes significance at 10 percent, ** denotes significance at 5 percent, *** denotes significance at 1 percent or better. This represents a summary of the results obtained in Table 5 of the appendix. Sample period is 2016 to 2020 for 19 autonomous communities.

The first, and most important variable that is going to be discussed in this study is the income. Income is significant and positive, this shows that for any given increase in their disposable income, the household will become more prone to gamble. This is true for all the three regressions, but with the addition of new variables to the model, income has reduce its importance to shift if the household consumes this type of products or not.

In order to have a deeper knowledge of the income variable, the models were tested with and without the quadratic variable of the (log) of income (a summary of the results of this model, can be seen in Table 4 of the appendices), the result obtained where that the goodness of fit did not vary between using it or not. But the second and third model's income variable was not significant, and now it is. So we can conclude that the model works better with this addition.

Gender is not significant, there is no gender discrimination to the main breadwinner at the time the household decides if wants to gamble. At the same time, age is significant and positive, meaning that it affects the decision and that as older the individual is as more probable is that gambles.

Looking at its level of education we can see that heads with higher studies, have the households less expected to do so. But it cannot be concluded that as higher the level of education, as less the possibility to gamble, as the least educated individuals have the closest coefficient to the highest educated ones. Even if the two middle levels are really close, having a medium level of studies, maximizes the propensity.

Apart from that, both the source of income and the position undertaken are significant. With that last one, having coefficients as expected, do to its relation with the income and the educational level.

Main breadwinner citizenship is also significant, but in this case it is negative. So when comparing foreign and dual citizenship to the Spanish one, it can be observed that they are less related to play this games. This might be explained as people having this kinds of nationalities, mainly come from Latin America and Africa. And this type of immigration, is associated with lower levels of studies, that makes them work in sectors/positions with lower wages, and as it has been shown, lower levels of income increases the gambling aversion. Another important type of immigrants that come to Spain, are the ones from other Euro-

pean countries, this ones are associated with higher levels of education, which also increases aversion.

The amount of members in the household is positive and significant, I blame this to happen to the fact that the more people a household has the more probable is that one of them gambles, increasing this coefficient even if for the model is the same that a household of 10 people just one person gambles and in a household of two people, both of them do it. Another different story is the amount of children, that is significant but negative, as households need to depict money that could be use to buy this products on taking care of them, so they need to shift they consuming patterns.

Singles, widowers and separated, are significant and negative compared to married, but this can also be associate to happen because this last one is more probable that its household has more members than the ones with a head having another type of union. This results can be related to the previous findings on the amount of members of the household.

Finally, the location of the household should be interpreted. Per capita GDP and the unemployment rate of the autonomous community is not significant. What matters when trying to interpret it, is the density of population of the city they live in. The denser the area, the more prone they are to gamble. With that being said, and without expecting it, a household located in a province capital is less probable that consumes this goods. This may happen as there are many provinces in Spain with a small population and that those games are less accessible. But bigger capitals like Madrid, Barcelona..., are expect to follow the previous assumptions found on highly populated areas.

When the amount of gambling spots is studied, we find out that casinos and "bingos" are significant and negative, but at the same time, "salones de juegos" and "salas de apuestas" are significant and positive. This should be happening due to the ease there is to get through any of the last two spots, while finding casinos or "bingos" is way more difficult.

Table 3: Results of the Tobit models (summary)

	Dependent variable:		
	Gambling spending (log)		
	(1)	(2)	(3)
Income (log)	12.000***	8.800***	8.400***
	(0.770)	(0.780)	(0.850)
Quadratic income (log)	−0.510***	−0.440***	−0.420***
	(0.039)	(0.039)	(0.042)
Observations	104,739	104,173	91,374
Main breadwinner socioeconomic data		Yes	Yes
Household socioeconomic data		Yes	Yes
Macro data of the area of residence			Yes

Note: Robust standard errors are in parentheses. * denotes significance at 10 percent, ** denotes significance at 5 percent, *** denotes significance at 1 percent or better. This represents a summary of the results obtained in Table 6 of the appendix. Sample period is 2016 to 2020 for 19 autonomous communities.

Now its time to analyze the results of the Tobit models. Its main goal is to find out what makes a household increase the amount spend on gambling. This is done by censoring all the households that do not spend money in this products. And as mentioned in the empirical methodology, follows the assumption that the income of the household affects equally the decision and the amount of gambling (The whole results of the regressions can be found in Table 6 of the appendices).

With that being said, we find out an income elasticity of 12, 8,8 and 8,4. So an income increase of the 1%, increases spending by 12 %, 8,8% and 8,4% respectively for each model. With an income elasticity way higher than 1, gambling products are assumed to be luxury goods. This is consistent with the literature, where Humphreys and Perez 2011 and Perez and Muñiz-Fernández 2021 argued that spending (in that case of lotteries) increases with income increases. In the case of gambling, this is consistent with the assumptions made by Sarti and Triventi 2012 in the case of Italy. The Spanish taxation on gambling appears to be progressive, accomplishing vertical equity. That states that richer people, should pay proportionally more than poorer. The last remark that should be made, is that the result

obtained on this models have the highest income elasticity of all the literature.

To further explore the effects of the income, its quadratic variable has also been added in this case. Below you can see a graph showing the relationship between the gambling spending (log) and the income (log) of the third regression (the regression with all the added variables).

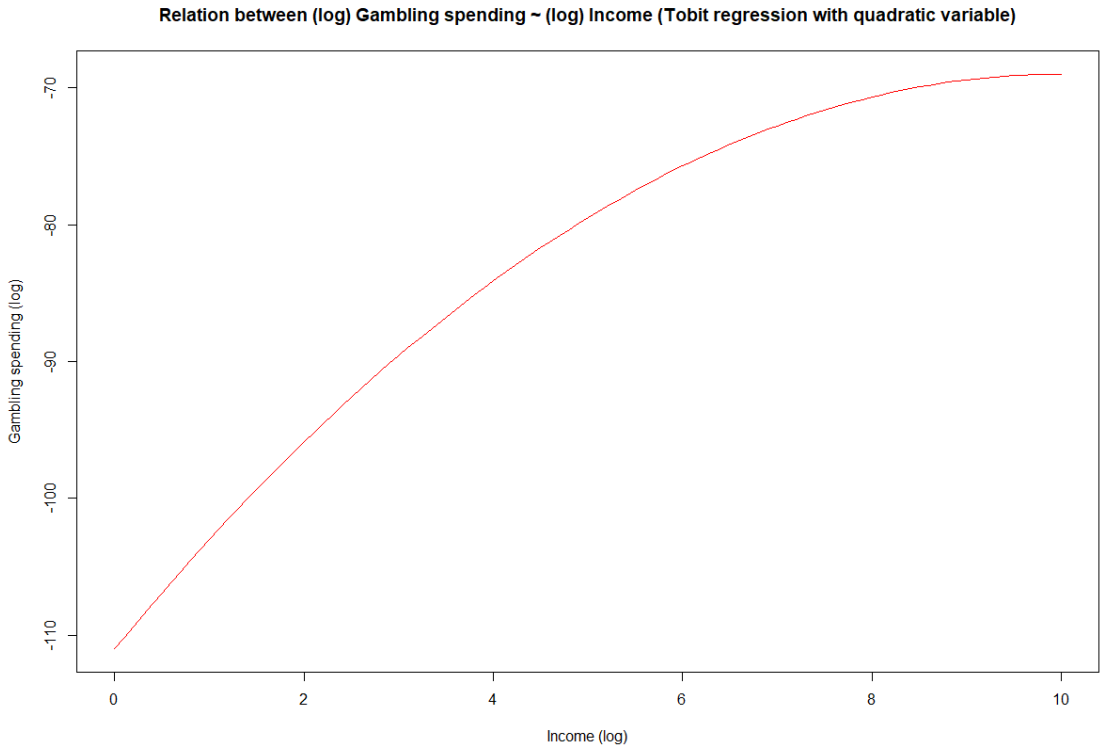

Figure 7: Relation between (log) Gambling spending and (log) Income with its quadratic variable, from the coefficients obtained in the regression of the third Tobit model (the one with all the variables). The line plotted came from the following equation: $y = \beta_0 + \beta_1 * x + \beta_2 * x^2$, where y is the (log) Gambling spending, x is (log) of Income and the betas are the coefficients obtained in the third model.

What we get is an upward slopping curve, that shows that gambling spending increases as income increases. This relation does not follow a linear trend but a logarithmic one, where at the early stages spending has a sharp increase with a small change in income. But up to a certain amount, when it increases the curve gets flatter and the change on spending is not as significant as it was before.

To end up with the income discussion, it is required to bring up something said and discussed by Garrett and Coughlin 2007 in the literature. They said that in order to determine the lottery spending, it should be taken into consideration the different sources from which the household obtains their income. Heads of the family with earnings related to self-employment

(freelance) are the ones spending the least. The ones that spend the most, are the ones that receive its earnings from being employed by someone else. And between both of them, the other kinds of income sources can be found. That includes: contributory pensions, unemployment benefits, capital and property rents....

The position they undertake is also significant, with coefficients following the same patterns that the income and the level of education that can be associated with their job have.

Age and gender are both positive and significant, that means that households headed by men spend significantly more, the same happen when people get older. The gender assumption differs with the previous findings about if the household spend money on gambling or not, so we can state that it does not matter if the household is headed by a men or a woman to take the decision to gamble, but it does at the time of determining how much to spend on it. Both, the age and gender arguments, are consistent with the previous findings in the literature.

Citizenship follows the same pattern seen before, relative to the ones that have the Spanish one, heads of the family with foreign or dual citizenship spend significantly less.

The number of members and children of the household also have the same sign on this coefficients as the one of previously mentioned in the OLS model, With spending increasing with the number of members and decreasing with this amount of children. The problem argued before, also is consistent with this cases, as households with more members may have a higher total income, and so more money to spend on this games (as the sum of their salaries may be higher than the sum of salaries from a household where each member earns more, but that have less people). Finally, as more children the household has, the more income needs to be shifted from gambling to them. As they are not able to generate any, and a lot of spending is required to raise them. Something also discussed by Worthington et al. 2007, that in Australia the number of dependent people decrease the amount intended to gamble.

Marriage apart from making households more prone to gamble, makes them spend the highest amount on it in comparison to other type of unions. Something previously shown by Humphreys and Perez 2011, for the Spanish lottery case.

In the case of education, the results differ from the ones found in the OLS models. In this case, households ruled by low levels of education spend more than the other ones. This is an assumption that agrees with all the literature (Leal, López-Laborda, and Rodrigo 2014; Sarti and Triventi 2012; Kohler 2016). With that being said, and with the previous findings in the other models, we can conclude that medium educational levels increase the facility

of the households to become gamblers, but that lower educational levels boost the amount intended for it.

Some changes have appeared with the place of residence, while depopulated areas are still negative and significant with respect to areas with a medium population, highly populated areas now lost their significance. Apart from that, households located in capitals still are significant and negative. Finally, some changes have also appeared regarding the amount of gambling spots. Casinos and "salas de apuestas" are no longer significant, "salones de juegos" are still positive and significant and "bingos" are significant but negative.

Last but not least, some discussion should be made on the per capita GDP and the unemployment rate. The unemployment rate has not changed its significance, something that contradicts what Leal, López-Laborda, and Rodrigo 2014 found. They argued that an increase on the unemployment rate reduces the amount spent on this goods. Garrett and Coughlin 2007 ratify this and further expand it. They say that due to the regressivity of lottery, a negative coefficient on the unemployment rate may reflect an increase on the upper distribution of income, reducing the amount sold. The fact that in this case gambling is found out to be progressive, as before implies a reduction of the lower distribution, which causes a further reduction of their amount expended. The difference now, is that this reduction is taken over by the richer group, whose income has increased, and so it has the amount they spend. Making the amount gambled unchanged, the only difference observable is a change on the type of consumer.

The case of the per capita GDP is completely different. It has become positive and significant, meaning that a 1% increase on this coefficient, increases by 3,6% the amount spent by the households. We can link this to the last part of the literature, where Chóliz and Manzón 2013 explained what happens with gambling in times of economic crises. The first assumption they make is that per capita GDP is positive and significant (Leal, López-Laborda, and Rodrigo 2014 also found that in their study), something that matches the finding of the model being interpreted. The second assumption they make is that gambling spending decreases with declines of the richness of the country, they associate this to happen to the fact that gambling is not an essential good. Something that can be extrapolated to this study, where a decline in the per capita GDP would reduce the disposable income of the households, further reducing sharply the amount spend on gambling due to the attributes of luxury goods (and gambling products are assumed to be so).

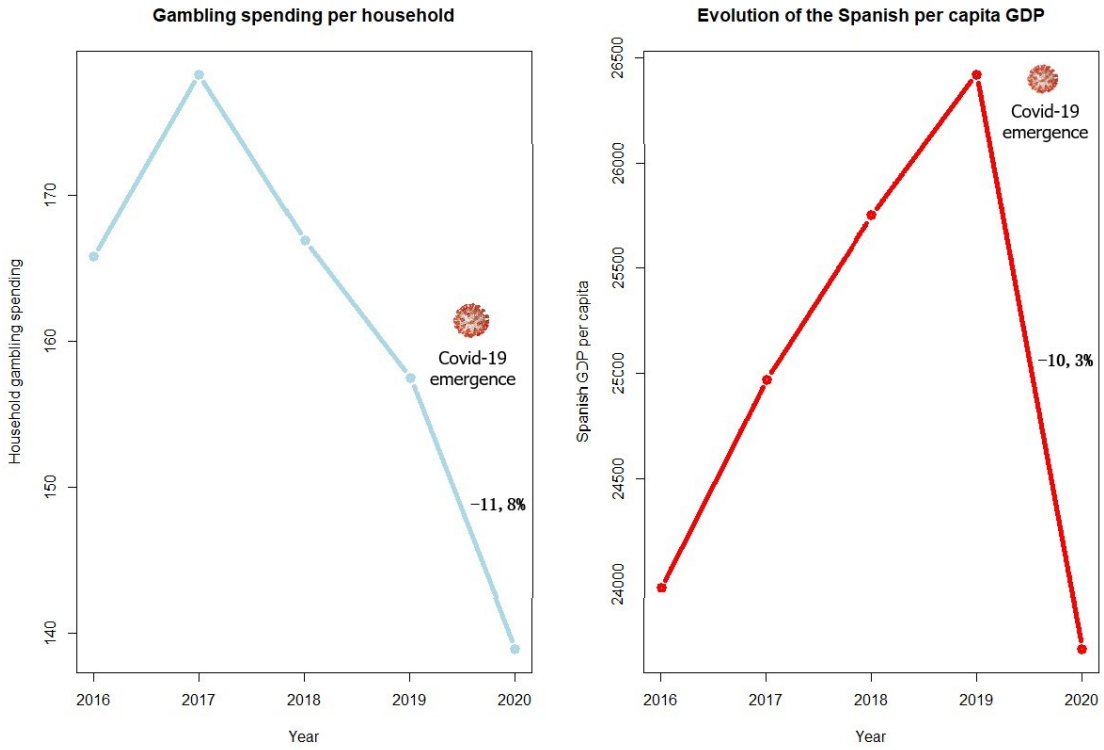

Figure 8: Comparison of the amount that households spend on gambling (yearly average), in the left. And the evolution of the Spanish per capita GDP, in the right. The regression models used in this study, just consider the different per capita GDP of each autonomous communities, and not the one from the whole country. Even though this is assumed to be still significant, due to the high degree of correlation expected between both variables.

This part should be terminated by analyzing the graphs seen above. As it has been elucidated, there is a relation between the amount spent on gambling and the per capita GDP. This relation can be clearly observed in this graphs, where between the beginning of the Covid-19 crises (2019) and 2020, a decline of the 10,3% of the per capita GDP shifted down by 11,8% the average yearly amount spent by the households of the sample. With this being said, per capita GDP is not the only variable that shifts spending, and that is why between 2017-2018 and 2018-2019, even if the per capita GDP surged, the spending did not. The same happens in this last year, that even if they both plummet, the effect that had was not as strong as it would have been if the only significant variable for determining the amount spent was the per capita GDP (If the per capita GDP of the country had the same coefficient as the one obtained in the Tobit model for the autonomous communities, this 10,3% decline should tumble down spending by 37.08%).

6 Concluding remarks

The present study uses regression analysis to investigate the determinants and incidence of both being a household that gambles and the gambling expenditures in Spanish households. The current study extends empirical work in this area in at least three ways.

Firstly, the study focuses on the relation that the net yearly household income has with the dependent variables mentioned above. The first relation resulted on income increasing the probability that the household gambles. The relation of the second model, gave an income elasticity of 12, 8,8 and 8,4 depending on the amount of controls added to the model (a model with more controls means a smaller income elasticity). Elasticities being that high, categorize gambling as a luxury good (a small change in income induces a bigger change in spending), making its taxation to be progressive (high income households contribute proportionally more than low income ones, accomplishing vertical equity). Finally, due to the introduction of its quadratic variable, it has been found that gambling spending positively increases with income, but up to a certain point the relation starts flattering (the relation does not maintain the same slope for all income levels, even if it is always upward-slopping).

Apart from that, the profile of the head of the family that maximizes the explained variables can be obtained. In the case of trying to maximizes the probability that a household gambles, the head of the family should have Spanish citizenship, live in a highly populated area, married but without children, with a level of education between 1st and 2nd stage of secondary education (medium level of education) and its source of income should come from being employed by someone else. When we do the same but trying to maximizes the amount spent, we find out that the household should be headed by a married men without children, that has Spanish citizenship and lives in a highly populated area, obtained the 1st stage of secondary education or less and that also receives its income from being employed by someone else. Being more aged also increases both the probability to gamble and the amount devoted.

It can be concluded that the main difference at the time of determining if a household gambles and the amount they spend on it, is that the decision of gambling is indifferent between the gender of its main breadwinner. But at the time of choosing the amount to spend, the ones having a man will spend significantly more. The other one is that households with heads that have a medium level of education are more prone to gamble, but the ones with lower levels devote more money on this games.

Last but not least, the per capita GDP is not significant to determine if a household gambles, but it is when determining how much of their income their shift to this products. This shows

us that in recession periods, when GDP decreases, the amount of households that gamble will remain unchanged (or may change insignificantly), but the quantity spent is the one that will plunge.

Future research should examine this more fully. One approach would be to use data did not represent households but individuals, as the information would be more precise and the profiles better characterized. Another one, would be differentiating the spending between games. Because, as seen in the literature in the studies undertaken by Martínez Martínez et al. 2013 and Chóliz and Manzón 2013, both the income elasticity and the profile of the player can widely differ.

References

Garrett, Thomas and Cletus Coughlin (Feb. 2007). "Inter-Temporal Differences in the Income Elasticity of Demand for Lottery Tickets". In: *National Tax Journal* 62. DOI: 10.2139/ssrn.1020737.

Worthington, Andrew et al. (May 2007). "Gambling Participation in Australia: Findings From the National Household Expenditure Survey". In: *Review of Economics of the Household* 5. DOI: 10.1007/s11150-007-9006-1.

Wooldridge, Jeffrey Marc (2009). *Introductory Econometrics: A Modern Approach*. South-Western.

Humphreys, Brad and Levi Perez (July 2011). "The Income Elasticity of Lottery: New Evidence from Micro Data". In: *Public Finance Review* 39, pp. 551–570. DOI: 10.1177/1091142111403620.

Sarti, Simone and Moris Triventi (Dec. 2012). "Gambling: The iniquity of a voluntary tax. The relationship between socio-economic position and propensity of gambling — Il gioco d'azzardo: l'iniquità di una tassa volontaria. La relazione tra posizione socio-economica e propensione al gioco". In: *Stato e Mercato* 96, pp. 503–534. DOI: 10.1425/38647.

Chóliz, Mariano and Miguel Manzón (Nov. 2013). "El efecto de la crisis económica sobre el gasto en juego de azar". In: *INFORMACIO PSICOLOGICA* 101, pp. 4–13. URL: https://www.informaciopsicologica.info/revista/article/view/77.

Martínez Martínez, Cristina et al. (2013). "Determinantes del gasto en juegos de lotería: la relativa importancia del nivel de renta". In.

Leal, Andrés, Julio López-Laborda, and Fernando Rodrigo (2014). "The Inside and Outside Revenue Impact of Regional Gambling Taxes in Spain". In: *Growth and Change* 45.1, pp. 79–97.

Vallejo Gabasa, Angel and Pilar Rivera Torres (2015). "Los Juegos de Azar en España: Un Análisis Descriptivo". In.

Kohler, Dimitri (2016). "On the regressivity of gambling taxes in Switzerland". In: *Swiss Journal of Economics and Statistics* 152.3, pp. 193–208.

Perez, Levi and Álvaro Muñiz-Fernández (Dec. 2021). "The income elasticity of lottery revisited: a worldwide perspective". In: *Economics and Business Letters* 10, pp. 403–407. DOI: 10.17811/ebl.10.4.2021.403-407.

Gómez, José A. and Carlos Lalanda (2022). *Auario del juego en España 2021*. Instituto de Política y Gobernanza de la Universidad Carlos III y CEJUEGO. URL: https://cejuego.com/wp-content/uploads/2021/11/2021-Anuario-Paginas.pdf.

Appendices

Table 4: Results of the linear OLS regressions (summary)

	Dependent variable:		
	Gambler		
	(1)	(2)	(3)
Income (log)	0.130***	−0.00001	−0.001
	(0.002)	(0.004)	(0.004)
Constant	−0.920***	−1.400***	−2.800**
	(0.030)	(0.045)	(1.200)
Observations	104,739	104,173	91,374
R^2	0.057	0.110	0.110
Main breadwinner socioeconomic data		Yes	Yes
Household socioeconomic data		Yes	Yes
Macro data of the area of residence			Yes

Note: Robust standard errors are in parentheses. * denotes significance at 10 percent, ** denotes significance at 5 percent, *** denotes significance at 1 percent or better. This represents a summary of the results obtained from the OLS regression without the usage of the quadratic variable of the income as a regressor. Sample period is 2016 to 2020 for 19 autonomous communities.

Table 5: Results of the OLS regressions

	Dependent variable:		
	Gambler		
	(1)	(2)	(3)
Income (log)	0.370***	0.270***	0.250***
	(0.044)	(0.045)	(0.048)
Quadratic income (log)	−0.012***	−0.014***	−0.013***
	(0.002)	(0.002)	(0.002)
Age		0.001***	0.001***
		(0.0002)	(0.0002)
Male		0.001	0.002
		(0.004)	(0.004)
Total spending (log)		0.140***	0.140***
		(0.004)	(0.004)
Freelance		−0.046***	−0.044***
		(0.005)	(0.005)
Other sources		−0.018***	−0.021***
		(0.007)	(0.007)
Foreign citizenship		−0.027***	−0.027***
		(0.007)	(0.008)
Dual citizenship		−0.025***	−0.025**
		(0.009)	(0.010)
Less 1st stage of sec. educ.		0.017***	0.014**
		(0.006)	(0.006)
1st stage of secondary educ.		0.022***	0.023***
		(0.005)	(0.005)
2nd stage of secondary educ.		0.018***	0.017***
		(0.005)	(0.005)
Working		-0.010	−0.012*
		(0.006)	(0.007)

Retired	-0.009	-0.006
	(0.006)	(0.006)
Studying	0.033	0.019
	(0.050)	(0.053)
Army	0.015	0.020*
	(0.011)	(0.012)
Scientific	0.002	0.006
	(0.008)	(0.008)
Supporting professionals	0.038***	0.042***
	(0.008)	(0.008)
Accountant/Administrative	0.055***	0.056***
	(0.009)	(0.009)
Restaurant/Security/Retailer	0.048***	0.052***
	(0.008)	(0.008)
Agriculture/Forestry/Fishing	0.016	0.026**
	(0.010)	(0.010)
Manufacturing/Construction	0.060***	0.064***
	(0.008)	(0.008)
Plant/Machinery operators	0.069***	0.068***
	(0.008)	(0.009)
Elementary occupations	0.058***	0.061***
	(0.009)	(0.009)
Full time	0.012	0.010
	(0.008)	(0.008)
Household members	0.130***	0.130***
	(0.003)	(0.004)
Household children	−0.084***	−0.083***
	(0.003)	(0.003)
Single	−0.018***	−0.016***

	(0.004)	(0.005)
Widower	-0.042^{***}	-0.039^{***}
	(0.006)	(0.006)
Separated	-0.023^{**}	-0.024^{**}
	(0.009)	(0.010)
Divorced	-0.007	-0.007
	(0.006)	(0.006)
Without mortgages	-0.005	-0.005
	(0.004)	(0.005)
Mortgage	-0.004	-0.006
	(0.004)	(0.005)
Meals	-0.002^{***}	-0.002^{***}
	(0.0001)	(0.0001)
Capital		-0.015^{***}
		(0.005)
Highly populated		0.011^{**}
		(0.005)
Depopulated area		-0.014^{***}
		(0.004)
Per capita GDP (log)		0.150
		(0.120)
Unemployment rate		0.002
		(0.002)
Casino		-0.009^{**}
		(0.005)
"Bingo"		-0.002^{**}
		(0.001)
"Salón de juegos"		0.0004^{***}
		(0.0001)

"Sala de apuestas"			0.0002*
			(0.0001)
Constant	−2.100***	−2.700***	−4.100***
	(0.220)	(0.220)	(1.200)
Observations	104,739	104,173	91,374
R^2	0.057	0.110	0.110

Note: Robust standard errors are in parentheses. * denotes significance at 10 percent, ** denotes significance at 5 percent, *** denotes significance at 1 percent or better. Autonomous communities and year dummy variables are included in each specification. The variables regarding main source of income are relative to Employee, the ones for citizenship are relative to Spanish citizenship, the ones for level of education are relative to Higher education, the ones for activity situation are relative to Other situation, the ones regarding working position are relative to Manager/Director, the ones regarding tenure regime are relative to Rent, the ones regarding marital status are relative to Marriage and the ones regarding density of population are relative to Intermediately populated. Sample period is 2016 to 2020 for 19 autonomous communities.

Table 6: Results of the Tobit regressions

	Dependent variable:		
	Gambling spending (log)		
	(1)	(2)	(3)
Income (log)	12.000***	8.800***	8.400***
	(0.770)	(0.780)	(0.850)
Quadratic income (log)	−0.510***	−0.440***	−0.420***
	(0.039)	(0.039)	(0.042)
Age		0.015***	0.014***
		(0.003)	(0.003)
Male		0.100*	0.110*
		(0.059)	(0.064)
Total spending (log)		2.700***	2.700***
		(0.060)	(0.065)
Freelance		−0.780***	−0.770***
		(0.074)	(0.080)
Other sources		−0.250**	−0.320***
		(0.110)	(0.120)

Foreign citizenship	-0.680^{***}	-0.720^{***}
	(0.120)	(0.130)
Dual citizenship	-0.650^{***}	-0.670^{***}
	(0.150)	(0.160)
Less 1st stage of sec. educ.	0.300^{***}	0.250^{**}
	(0.090)	(0.098)
1st stage of secondary educ.	0.350^{***}	0.360^{***}
	(0.071)	(0.077)
2nd stage of secondary educ.	0.270^{***}	0.240^{***}
	(0.071)	(0.077)
Working	-0.094	-0.120
	(0.100)	(0.110)
Retired	-0.05	0.056
	(0.097)	(0.110)
Studying	0.150	-0.068
	(0.950)	(1.000)
Army	0.140	0.240
	(0.180)	(0.200)
Scientific	-0.001	0.074
	(0.120)	(0.130)
Supporting professionals	0.640^{***}	0.720^{***}
	(0.120)	(0.130)
Accountant/Administrative	0.920^{***}	0.960^{***}
	(0.130)	(0.140)
Restaurant/Security/Retailer	0.820^{***}	0.880^{***}
	(0.120)	(0.130)
Agriculture/Forestry/Fishing	0.280^{*}	0.400^{**}
	(0.150)	(0.170)
Manufacturing/Construction	1.100^{***}	1.100^{***}

	(0.120)	(0.130)
Plant/Machinery operators	1.200***	1.200***
	(0.130)	(0.140)
Elementary occupations	1.000***	1.100***
	(0.130)	(0.150)
Full time	0.160	0.140
	(0.120)	(0.140)
Household members	1.600***	1.600***
	(0.051)	(0.056)
Household children	−1.200***	−1.200***
	(0.043)	(0.047)
Single	−0.360***	−0.330***
	(0.072)	(0.077)
Widower	−0.850***	−0.810***
	(0.098)	(0.110)
Separated	−0.540***	−0.550***
	(0.150)	(0.170)
Divorced	−0.270***	−0.270**
	(0.098)	(0.110)
Without mortgages	-0.023	-0.023
	(0.071)	(0.078)
Mortgage	0.014	-0.002
	(0.071)	(0.078)
Meals	−0.020***	−0.021***
	(0.001)	(0.002)
Capital		−0.250***
		(0.075)
Highly populated		0.065
		(0.077)

Depopulated area			-0.220^{***}
			(0.071)
Per capita GDP (log)			3.600^{*}
			(2.100)
Unemployment rate			0.036
			(0.032)
Casino			-0.029
			(0.074)
"Bingo"			-0.034^{**}
			(0.015)
"Salón de juegos"			0.008^{***}
			(0.001)
"Sala de apuestas"			0.001
			(0.002)
Constant	-74.000^{***}	-77.000^{***}	-111.000^{***}
	(3.900)	(3.900)	(21.000)
Observations	104,739	104,173	91,374

Note: Robust standard errors are in parentheses. * denotes significance at 10 percent, ** denotes significance at 5 percent, *** denotes significance at 1 percent or better. Autonomous communities and year dummy variables are included in each specification. The variables regarding main source of income are relative to Employee, the ones for citizenship are relative to Spanish citizenship, the ones for level of education are relative to Higher education, the ones for activity situation are relative to Other situation, the ones regarding working position are relative to Manager/Director, the ones regarding tenure regime are relative to Rent, the ones regarding marital status are relative to Marriage and the ones regarding density of population are relative to Intermediately populated. Sample period is 2016 to 2020 for 19 autonomous communities.