# Who Loves to Gamble? Socio-Economic Factors Determining Gambling Behaviour in Germany

Giebeler, Constanze and Rebeggiani, Luca

FOM University of Applied Sciences

July 2019

# Who loves to gamble? Socio-Economic Factors Determining Gambling Behaviour in Germany

**Constanze Giebeler**
**Luca Rebeggiani***

This Version: July 2019

Keywords: Gambling, Demand Analysis, Private Consumption, Risk and Uncertainty
JEL-Classification: D12, D81, L83, Z2

**Corresponding author*
*FOM University of Applied Sciences*
*Joseph-Schumpeter-Allee 23-25*
*D-53227 Bonn*
*luca.rebeggiani@fom.de*

## Abstract

The interest among academics and policy makers in the economics of gambling has risen substantially. Changes in gambling regulation, the relevance of gambling generated (tax) revenues and the usefulness to other fields of economics, increased the interest in understanding gambling demand. The focus of this paper is to provide one of the first comprehensive analyses of gambling demand in Germany, by studying the socio-demographic and socio-economic factors which influence gambling expenditures using one of the largest data sets available, the official income and consumption survey (Einkommens- und Verbrauchsstichprobe).

Applying models suitable for censored data, we identify the factors influencing gambling demand. Some findings are in line with the previous literature on gambling demand, others are quite surprising. They show that female household heads spend less on gambling than their male counterparts. The number of dependants in the household influences gambling behaviour negatively, while married couples have higher expenditures than single person households. We find that while gambling expenditures increase with the age of the household head, they do so at a diminishing rate. Income is also a strong determinant of gambling expenditure. Gambling expenditures rise with the household's income. We find no evidence that expenditures increase regressively. Furthermore, the education and occupation of the household head influences gambling behaviour. A higher education has a negative effect on the propensity to gamble. We also find that households with privately employed household heads tend to spend more on gambling than those with unemployed or self-employed household heads. On the other hand, a household head who is a civil servant or retired has a positive effect. In contrast to previous literature, we find a negative effect of urbanity on household gambling expenditure.

We discuss these and other results in the context of the theory of risk taking preferences and also with a focus on gambling regulation and taxation.

# 1 Introduction

In recent years, the interest among academics and policy makers in the economics of gambling has risen substantially. Changes in gambling regulation, the relevance of gambling generated (tax) revenues and the usefulness to other fields of economics, increased the interest in understanding gambling demand. The focus of this paper is to study the socio-demographic and socio-economic factors which influence gambling expenditures. The analysis of the determinants of gambling behaviour is of interest to researchers in the field of politics, economics as well as psychology.

The results of this paper hold implications for many policy issues related to gambling. First, socio-economic factors determining gambling behaviour of individuals are of importance for policy initiatives which deal with preventive measures to fight gambling addiction. Second, the question whether gambling is socially desirable or not is controversial in most countries. There exists a trade-off between the positive effects of tax revenues generated by gambling and the possible negative externalities gambling causes. Additionally, the results of this paper can shed light on the ongoing discussion of whether or not the taxation of gambling products might actually be a form of regressive taxation rather than a redistribution benefiting the socially disadvantaged.

Furthermore, some researchers have linked gambling behaviour to risk attitudes. Mishra et al. (2010) link gambling behaviour and an individual's propensity to take risks. If that is the case, the traits influencing gambling behaviour can be translated into the individual determinants of the willingness to take risks. One could extend the results of this paper to the examination of risk attitudes of individuals in the context of expected utility theory. In this case, the determinants of gambling attitudes are of interest for both behavioural and financial economics.

To analyse gambling behaviour in Germany, we use the official income and consumption sample (*Einkommens- und Verbrauchsstichprobe*). This official statistic about the living conditions of private households in Germany is the largest of its kind in the European Union. Among other things, it provides information about individual socio-demographic and socio-economic characteristics, household expenditure on consumer goods, income, as well as on the expenditure on gambling products. Since a high fraction of households do not consume gambling at all, the data on gambling expenditure is censored at zero. Thus, we have to apply suitable models that are robust to censored data. To identify the factors influencing gambling demand we employ the Tobit (Tobin 1958) and the Truncated Normal Hurdle model (proposed by Cragg 1971).

The findings are in line with most of the previous literature on gambling demand (see Section 3). They show that female household heads spend less on gambling than their male counterparts. The number of dependants in the household influences gambling behaviour negatively, while married couples have higher expenditures

than single person households. We find that while gambling expenditures increase with the age of the household head, they do so at a diminishing rate. Income is also a strong determinant of gambling expenditure. Gambling expenditures rise with the household's income. We find no evidence that expenditures increase regressively. Furthermore, the education and occupation of the household head influences gambling behaviour. A higher education has a negative effect on the propensity to gamble. We find that households with privately employed household heads tend to spend more on gambling than those with unemployed or self-employed household heads. On the other hand, a household head who is a civil servant or retired has a positive effect.[1] In contrast to previous literature, we find a negative effect of urbanity on household gambling expenditure.

The paper is structured as follows: First, Section 2 outlines the economic relevance of the gambling market in Germany and its structure. Section 3 provides an overview of the previous literature focussing on gambling demand. In the following section, a general identification strategy for censored data is presented. Since the data on gambling expenditure is constituted of a high fraction of zero observations, models that are robust to censored data need to be employed. Therefore, the Tobit and the Truncated Normal Hurdle model are introduced. Section 5 analyses the socio-demographic and socio-economic determinants of gambling expenditure. First, the model specification is discussed. Then, the regression results and several robustness measures are presented. Finally, Section 7 concludes the paper with a summary of the findings and discusses potential shortcomings and suggestions for future research.

# 2 Relevance and Structure of the Gambling Market in Germany

The importance of the German gambling market is often underestimated in its economic relevance. With total revenues of around €45 billion (DHS 2018) in 2016 it was about the size of total revenues of the pharmaceutical industry (IQVIA 2018) and more than one quarter of the size of total grocery revenues in Germany (Statistisches Bundesamt 2018). The gross gaming revenues (GGR)[2] added up to around €14.2 billion and the gambling market generated tax revenues of €5.3 billion in 2015 (Handelsblatt 2017). These numbers illustrate the economic significance of the

[1] The official income and consumption sample is subject to middle-class-bias (Rebeggiani 2007). Retirees and civil servants are overrepresented in the sample, while unemployed and self-employed individuals are under-represented. Therefore, the surprising results regarding the effects of the occupational status of the household head on gambling expenditures might partially be caused by a middle-class-bias. This topic is further discussed in Section 7.

[2] Gross Gaming Revenue is the amount wagered minus the winnings returned to players. It is the most common measure to evaluate the gambling sector's revenues.

German gambling sector.[3]

The German gambling market is divided into a regulated and an unregulated sector. The regulated market consists of state-run and social lotteries, commercial gambling machines (*Geldspielgeräte*),[4] casinos and horse racing (by commercial bookmakers). It holds the lion share of the overall market with gross gaming revenues of €10.4 billion in 2015, which accounted for 82% of market-wide GGR (Handelsblatt 2017). By far the most popular gambling product in the regulated market are the commercial gambling machines. They account for a market share of roughly more than 50% of the regulated segment (Rebeggiani 2010). The unregulated market for gambling products includes online providers (online-casinos, online-poker and online-lotteries) as well as sports betting. Nowadays, the major share of sports bets are placed online and the online sector gradually replaces commercial bookmakers (Rebeggiani 2012). This market consists of gambling providers who do not possess a gambling concession in Germany but hold a license in any other European country, making them illegal but tolerated in Germany (Rebeggiani and Breuer 2017). Since this market is unregulated, revenues can only be estimated. In 2015, GGR were estimated to gross €2.3 billion which accounted for 18% of total GGR in Germany (Handelsblatt 2017). Additionally, there exists a black market with estimated GGR of €1.5 billion (Handelsblatt 2017). Including the black market the total GGR in Germany in 2015 totalled €14.2 billion (Handelsblatt 2017). The GGR usually accounts for around one quarter of total revenues (Rebeggiani 2010).

Gambling is popular among the German population. In 2015 the Federal Center for Health Education (BZgA) conducted a representative survey of 11,438 individuals. The survey asked participants about their past gambling behaviour. It finds that the lifetime prevalence of taking part in any kind of commerical gambling was 77.6% for 16 to 70 year olds. Additionally, 37.3% reported that they had gambled during the past 12 months (Haß and Lang 2016). Thus, the gambling market is not only of economic significance but has to be considered from a social and political point of view.

To this note, policy makers are often irresolute between whether or not a high prevalence of gambling participation is a (socially) desirable outcome: It is a mixed blessing. On the one hand, gambling products are highly taxed. Since 2010 the tax revenue generated through the gambling sector continuously increased from €4.7 billion to €5.3 billion in 2015 (Handelsblatt 2017).[5] The money generated from these taxes is used to finance many charitable causes. Therefore, taxed gambling should benefit households with lower social standing in particular, as well as the common

[3]The data in this section refers to the (latest) available data. For that reason the figures stated are from 2015 as well as from 2016.

[4]There is a regulatory distinction between slot machines located in bars or commercial gambling halls (*Spielhallen*) and the ones located in official casinos, which allow greater stakes and therefore greater losses. The first category is labelled here "commercial gambling machines".

[5]These figures do not include trade taxes and corporation taxes.

good in general. On the other hand, negative externalities arise from gambling, especially when considering pathological gambling.[6] The negative social effects that arise from gambling are referred to as social costs. They can be both of direct and indirect nature. The direct effects include expenditures on measures to prevent addiction, mental health care, costs for divorces and acquisitive crimes among others. The indirect costs are made up of costs that arise from the loss of jobs, absenteeism due to illness or a decrease in productivity.[7] Although these negative externalities are impossible to be perfectly quantified, Becker estimated the direct and indirect social costs of gambling in Germany totalled up to €326 million in 2008 (Becker 2011, p.39). Therefore, there has been a long standing debate concerning the positive and negative externalities that arise from legal gambling. This debate likewise focuses on the socio-economic characteristics of gamblers. Many researchers hypothesize that the taxation of gambling products might actually be a form of regressive taxation (e.g. Beckert and Lutter 2008; Borg et al. 1988; Clotfelter and Cook 1987; Smith 2000). These studies find that low-income households spend proportionally more of their income on (taxed) gambling products than better-off households and therefore contribute relatively more to the tax revenue in relation to their income. Nonetheless, low-income households should benefit to a greater extent from the distribution of the tax towards public welfare. Thus, whether or not gambling incorporates a form of regressive taxation is mainly dependent on the way the government redistributes tax revenues generated through gambling. However, it is necessary to determine the relationship of income, socio-demographic characteristics and gambling expenditures in Germany to evaluate this potential trade-off.

# 3 Literature Review on the Determinants of Gambling Demand

In general, the previous literature on gambling demand agrees that socio-demographic as well as socio-economic characteristics influence both individual and household gambling behaviour. There are various studies that empirically examine the determinants of gambling consumption and expenditure (see Gandullia and Leporatti 2017 for a review).

The literature that studies the effects of socio-demographic variables on gambling behaviour typically focuses on the traits gender, age and ethnicity (e.g. Afifi et al. 2010; Casey et al. 2011; Castrén et al. 2018; Scott and Garen 1994; Worthington et al. 2003). Most studies agree that men are more likely to gamble and to spend

[6] Here, pathological gambling is defined as an addiction where individuals are unable to resist their urge to gamble despite harmful negative consequences or a desire to stop.

[7] For a comprehensive overview of the social costs of gambling see Collins et al. (2003) or Walker et al. (1999).

more money on gambling products than women (e.g. Gupta and Derevensky 1998; Kitchen and Powells 1991; Sawkins and Dickie 2002). This is in line with the more general result that men are more prone to risk and are more likely to engage in risky activities in all life domains (Dohmen et al. 2011; Harris et al. 2006). With respect to the influence of age the previous research is not that undivided. Although some studies find that the participation in gambling activities is particularly high for younger individuals (Albers and Hübl 1997; Gupta and Derevensky 1998), several other studies uncover that gambling expenditures increase with age (Worthington et al. 2003; Sawkins and Dickie 2002). Thus, there seems to exist a parabolic relationship between age and gambling behaviour (Castrén et al. 2018). Especially for individuals who reached the retirement age, there seems to be a noticeable increase. This effect might be due to increased free time as well as to preference changes and boredom initiated by retirement. McNeilly and Burke (2000) conducted a survey to investigate the motives for the retired population to gamble. The answers by the respondents most agreed upon when asked for their motives to gamble were "to have fun" and "to get away for the day". Concerning ethnicity, different studies show that minority racial groups are more likely to buy lottery tickets (e.g. Hansen 1995; Mikesell 1989; Scott and Garen 1994; Tan et al. 2010). However, no study clearly states that certain ethnic groups are more prone to gambling in general. The role of ethnicity seems to be dependent on the individual local environment, neighbourhood effects and cultural differences across countries (e.g. Casey et al. 2011; Kumar et al. 2011; Lin et al. 1998; Slutske et al. 2015; Worthington et al. 2003). Urban households have a higher tendency to gamble (Tan et al. 2010). This is presumably due to easier access to gambling venues due to the distance and accessibility to the closest venue is an important determinant of gambling prevalence (Storer et al. 2009; Young et al. 2012). Additionally, the composition of a household has an effect on gambling behaviour (Worthington et al. 2007). Married individuals tend to spend less on gambling than singles (Castrén et al. 2018) while married couples with children gamble more than lone parents (MacDonald et al. 2004). Worthington et al. (2003) find a negative effect of household size on lottery play suggesting that gambling expenditure decreases with the number of dependants within the household.

Other than socio-demographic variables, socio-economic determinants of gambling behaviour are of interest when analysing gambling behaviour. For instance, the level of education is negatively correlated with gambling participation and expenditure (Albers and Hübl 1997; Borg and Mason 1988; Hansen 1995; Kitchen and Powells 1991; Sawkins and Dickie 2002). Furthermore, the occupational status has an effect on gambling behaviour. Worthington et al. (2007) detect that the source of income rather than the income level determines gambling behaviour. For example, non-white collar households (Tan et al. 2010) and civil servants (Albers and Hübl

1997) are more likely to gamble than other occupational groups. This is in line with Layton's and Worthington's (1999) finding that managers and professionals spend relatively less on gambling. They also find that participants of pensions and veterans affairs payments have a greater probability to gamble and that receiving unemployment benefits has a positive impact on the probability to play lottery or poker machine gambling. Other studies confirm this result and find that participants in government income assistance programs are more likely to spend money on the lottery (Laitner 1999; Clotfelter and Cook 1989). Despite the mentioned relevance of other socio-economic characteristics, the literature commonly focuses on income as a measure of economic conditions (e.g. Castrén et al. 2018; Kitchen and Powells 1991; Layton and Worthington 1999; Tan et al. 2010). Gambling is considered to be a normal good with consumption increasing with income (Albers and Hübl 1997). Though, the previous literature finds an income elasticity of less than one (Clotfelter and Cook 1987 and 1989) which suggests a regressive nature of gambling. Low-income households spend proportionally more of their income on gambling products (Beckert and Lutter 2008; Castrén et al. 2018; Grote and Matheson 2011; Kitchen and Powells 1991; Rintoul et al. 2013). This general result supports the hypothesis that gambling might in fact be a form of regressive taxation. Generally, the impact of socio-economic influence is particularly important for policy makers. The evaluation of gambling regulations, legislation and addiction prevention is dependent on the knowledge about the socio-economic composition of the gambling population. Especially, the issue of a possible regressive nature of a gambling tax can only be objectively assessed if detailed information about the income of gamblers is available.

# 4 Data and Comparative Statistics

This section presents the data used in the analysis and presents some first impressions on the relationship between gambling expenditure and other outcomes. Section 4.1 discusses the data and its preparation. Section 4.2 hints at some first impressions on the drivers of gambling expenditures and their relationship to socio-demographic and economic characteristics.

## 4.1 Data

The income and consumption sample EVS (*Einkommens- und Verbrauchsstichprobe*) is an official statistic about the living conditions of private households in Germany. Among other things, it provides statistical information on the equipment with consumer goods, the income, asset and debt situation and consumption spending of

private households. The EVS is carried out in close coordination and cooperation between the Federal Statistical Office and the Federal States Statistical Offices. The data of the EVS is not publicly available, but anonymised scientific use files can be obtained upon request. The cross section sample includes households of all social groups, so that the EVS paints a representative picture of the German population.[8] The EVS is conducted every five years. All households participate on a voluntary basis. However, participating households receive a monetary award.[9] Typically, around 60,000 private households in Germany are surveyed within the framework of the EVS, including around 14,000 households in the new federal states and Berlin. The survey is the largest survey of its kind in the European Union. The EVS has been held in the former federal territory since 1962/63, as well as in the new federal states and Berlin since 1993. For our analysis we employ the latest available wave of the EVS which was conducted in 2013. The EVS 2013 contains data on 42,792 households.

The survey contains questions about demographic and socio-economic variables on the household level. The questions include information about the region and state the household resides in. Furthermore, participating households register all their income and expenditures within a three-month survey period.[10] The expenditure data includes the household spending on gambling products. Unfortunately, the spendings are not further diversified into different types of gambling products.[11] The reliability of self-reported data in gambling expenditures has to be addressed. Ladouceur and Walker (1996) find evidence that gamblers are subject to cognitive bias. Auer and Griffiths (2017) find that players with high losses are more likely to understate their losses. Additionally, the wording in the question about gambling expenditure in the EVS survey is vague.[12] The survey did merely ask the respondent

[8]The cross section data are usually extrapolated using weights obtained from the German microcensus to calculate aggregate figures. However, this was not done in this paper, since we are only interested in the individual determinants of gambling expenditures and not in a calculation of total gambling expenditures in Germany.

[9]The EVS sample is subject to sample bias. The monetary reward is credited against social and unemployment benefits. Thus, the incentive for those households to participate is diminished and unemployed individuals are underrepresented in the sample. Furthermore, compared to the German population self-employed individuals are underrepresented in the sample, while civil servants and retirees are overrepreseneted (Rebeggiani 2007).

[10]The participants of the EVS are divided into four groups with each group reporting in one quarter of the year under review. This ensures that possible cyclical economic variation is accounted for.

[11]This has relevance because it is expected that consumers of certain gambling products differ in their characteristics, e.g. if one compares lottery ticket buyers to slot machine gamblers. Nevertheless, previous studies generalize the results of the purchase of specific gambling products to gambling in general (e.g. Albers and Hübl 1997; Layton and Worthington 1999; Worthington et al. 2003). The topic will be further discussed in Section 7.

[12]The EVS questionnaire from 2013 included one column, where subjects should report their monthly gambling expenditures for each of the three months in the reporting period. The column was placed in the section on expenditure for recreation, entertainment and culture. The notes regarding what is meant by gambling expenditure included lotteries, bookmakers, casinos, slot machines, online sports betting and funfair tickets.

about monthly expenditure on gambling without further defining how to assess gambling expenditure. This leaves room for interpretation. Respondents could interpret gambling expenditures as actual expenditure or the difference between winnings and losses. There is also no information to what extent the respondents excluded or included reinvested winnings in their report. We proceed with the assumption that respondents reported their expenditures without incorporating wins or excluding reinvested winnings.

Additional to household data, the EVS also includes a wide range of variables for each household member individually. That is very rare among surveys on income and expenditure in such a great sample. Among those are information about birth year, sex, marital, occupational and social status, education, insurance, citizenship, taxation and income.[13]

The EVS proved to be an extraordinarily rich data set. Thanks to the extraordinarily careful editing procedure of the German Federal Statistical Office, there are, for example, no missing values. Nevertheless, some preparation of the data is needed for further analysis. First, all information on income and expenditures are given quarterly. Thus, these variables have to be averaged by month to make the analysis more generally comprehensible. Though, this leads to some inaccuracies due to the averaging process, it also yields an advantage to other data. Most other studies just include one data point, which is generated at the time of the sampling. Using only one point in time can be misleading though. For example, there might be the chance that this was a particularly bad month in terms of income or that this month due to an event gambling spending was extremely high. In the EVS we have information for an average month over a three-month period. Therefore, the data is more reliable because extraordinary events are averaged out.

Second, information on the age of participants is just given in terms of birth year. Thus, a new variable has to be created which captures the actual age of participants. Regrettably, this can create a small sample error because we do not know whether or not the participant's birthday has already taken place within the reporting year. Thus, the new variable containing age could understate a participant's age by one year.

Furthermore, a closer look at the data revealed that twelve households report a negative gross income. We dropped these twelve observations because the source of the negative income can not be identified. Additionally, the data on gambling expenditure showed suspicious numbers for the highest five observations. That is, the five observations with the highest gambling expenditure all equalled exactly €7,705 per household for the three-months survey period or an average of €2,568.33 per month. These figures were imputed by the Federal Statistical Office for reasons of confidentiality. Because of this, we dropped these five outliers. This leaves the dataset with

[13]Throughout this paper the main income earner is referred to as the household head. The participating households reported which household member is the main income earner.

Table 1: Summary statistics of average monthly household gambling expenditures

| | Mean | Sd | Min | Max | Variance | Skewness | Kurtosis | Obs. |
|---|---|---|---|---|---|---|---|---|
| (1) | 11.97 | 40.61 | 0 | 2,568.33 | 1,649.91 | 33.87 | 1,952.25 | 42,780 |
| (2) | 11.67 | 29.77 | 0 | 1,682.67 | 886.11 | 12.05 | 411.65 | 42,775 |

Row (1) represents the original statistics on av. monthly hh gambling expenditure. Row (2) represents the modified statistics after deleting the five highest observations.

Table 2: Distribution of average monthly household gambling expenditures

| Percentiles | 1% | 50% | 75% | 90% | 95% | 99% |
|---|---|---|---|---|---|---|
| (1) | 0 | 0 | 12 | 36.33 | 58 | 121.67 |
| (2) | 0 | 0 | 12 | 36.33 | 57.67 | 121.33 |

Row (1) represents the original distribution of av. monthly hh gambling expenditure. Row (2) represents the modified distribution after deleting the five highest observations.

a total of 42,775 observations. The new figures for monthly average gambling expenditures are displayed in Tables 1 and 2. Deleting these five observations neither changes the mean nor the distribution within the first 99 percentiles considerably but reduces the variance, skewness and kurtosis substantially. Therefore, deleting the observations eases any following estimation procedure without deteriorating the results.

Even after excluding the five highest observations the variance and kurtosis of the observed gambling expenditures remain high. The 420 highest observations range from €121.33 to €1,682.67. The highest observation equals a deviation from the mean of around 56 times its standard deviation. In practice a deviation from the mean of around four times its standard deviation is considered plausible. A closer look at the observations within the highest percentile reveals wide heterogeneity between observations. Especially, with respect to income and age, there seems to be no trend as to whether or how they influence gambling expenditure. It rather conveys the impression that the level of gambling expenditure within these observations is arbitrary. Given the high amounts of monthly gambling expenditures, it is plausible to assume that these data points stem from individuals who are either problem gamblers or have so much excess funds that they do not care about the amount they gamble with. Hence, these amount decisions were likely to be chosen irrationally and at random. Therefore, one should examine the extraordinary high observations in the highest percentile of observations separately and consider to exclude them from the regressions. This topic will be further discussed in Section 6.

Table 7 in the appendix displays the most important summary statistics for all variables used in the analysis of household gambling expenditure.

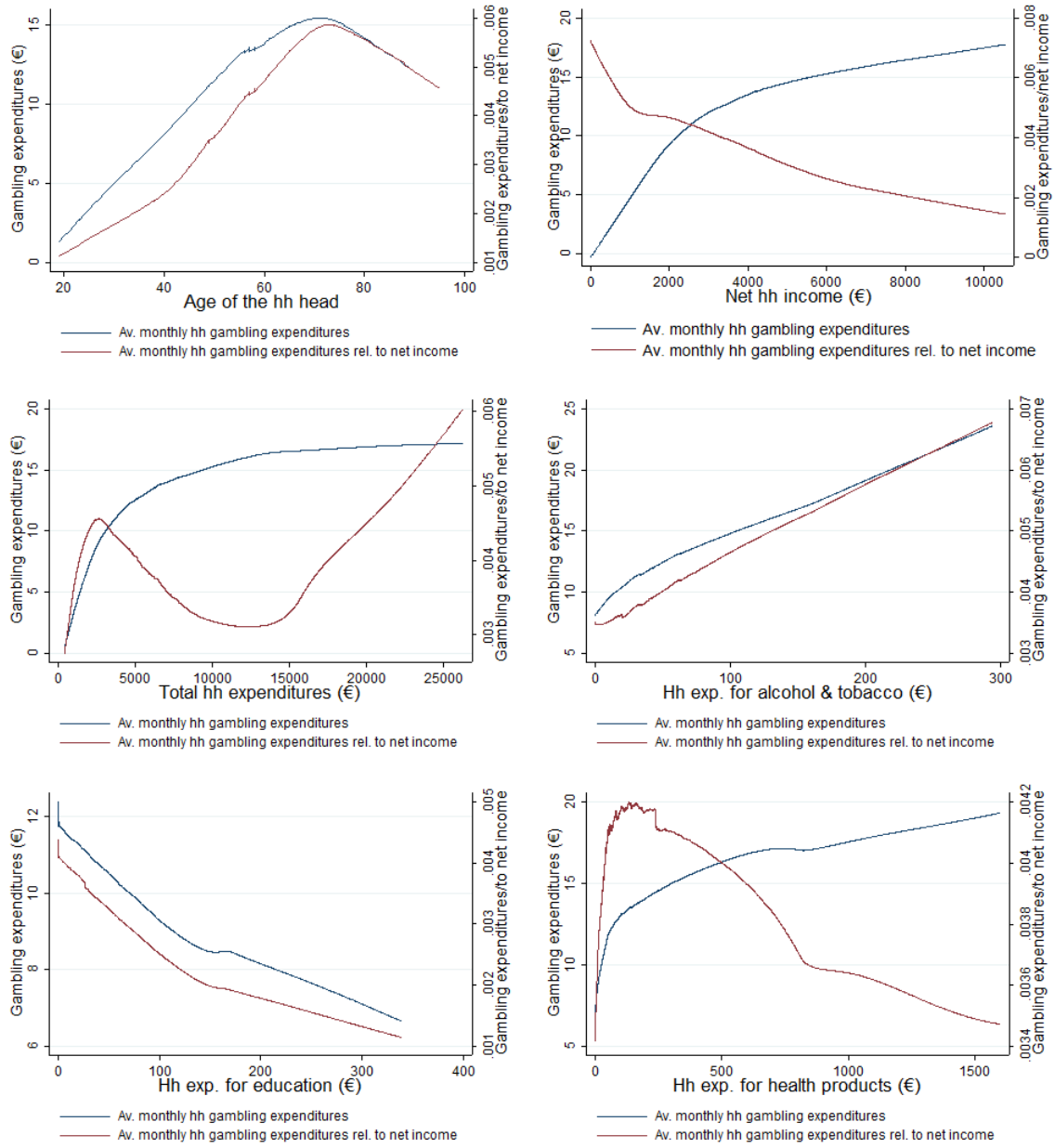
## 4.2 Comparative Statistics and First Impressions

This section further explores the data and yields some first impressions on the relationship of gambling expenditure and selected demographic and socio-economic characteristics. Figure 1 depicts the relationships between certain household figures and gambling expenditure as well as gambling expenditure relative to income (represented by the blue and red lines, respectively). The graphs present the relationship of absolute and relative gambling expenditures to the variables age, household net income, total household expenditures and expenditures on alcohol and tobacco, for educational purposes and health. We did not use scatter plots to picture the relationship of the variables.[14] Instead, we use locally weighted scatterplot smoothing (LOWESS) with bandwidth 0.8. Lowess carries out a locally weighted least squares regression of absolute and relative gambling expenditure on the mentioned variables. Further, we restricted the independent variables to the first 99th percentiles. This is due to the scarceness of data points for each value higher than the 99th percentile. Especially, for data on expenditures outliers in the highest percentile distorted the graphs to a great extent.

The first graph depicts the relationship of average total and relative expenditures of gambling and the age of the household head. Both figures continually increase up to an age of around 75 years. This suggests that gambling expenditures do rise with age and continue to increase with the average retirement age. At the turning point of an age of 75 this effect is reversed. The sudden decline could be explained by decreased mobility and higher ill-falling preventing any gambling participation. The graph depicting the relationship of gambling and income is particularly interesting. First, note that total expenditure does increase with income but at a diminishing rate. This supports previous findings that gambling is in fact a normal good with consumption rising with income. On the other hand, gambling expenditures relative to income are highest at low household income and tend to decline with income. This finding combined with diminishing total expenditures with income, facilitates the proposition that gambling is regressive. As total household expenditures are positively correlated with income, it is of no surprise that absolute gambling expenditures rise with them. Nevertheless, it is noteworthy that relative gambling expenditures first rise with household expenditures for low income households but then suddenly decline. It shows that low income households spend a larger fraction of overall expenditures on gambling than middle and high income households. Again, this supports the findings that gambling is regressive. With respect to expenditures on alcohol and tobacco, educational purposes and health products, the findings suggest both positive and negative relationships. In all cases absolute and relative gambling expenditures display the same trend. While the relationship of

[14] The sheer volume of the data combined with the concentration of gambling expenditure in the interval from €0 to €12 and large outlying observations, prevented the scatter plots to contribute any informational value.

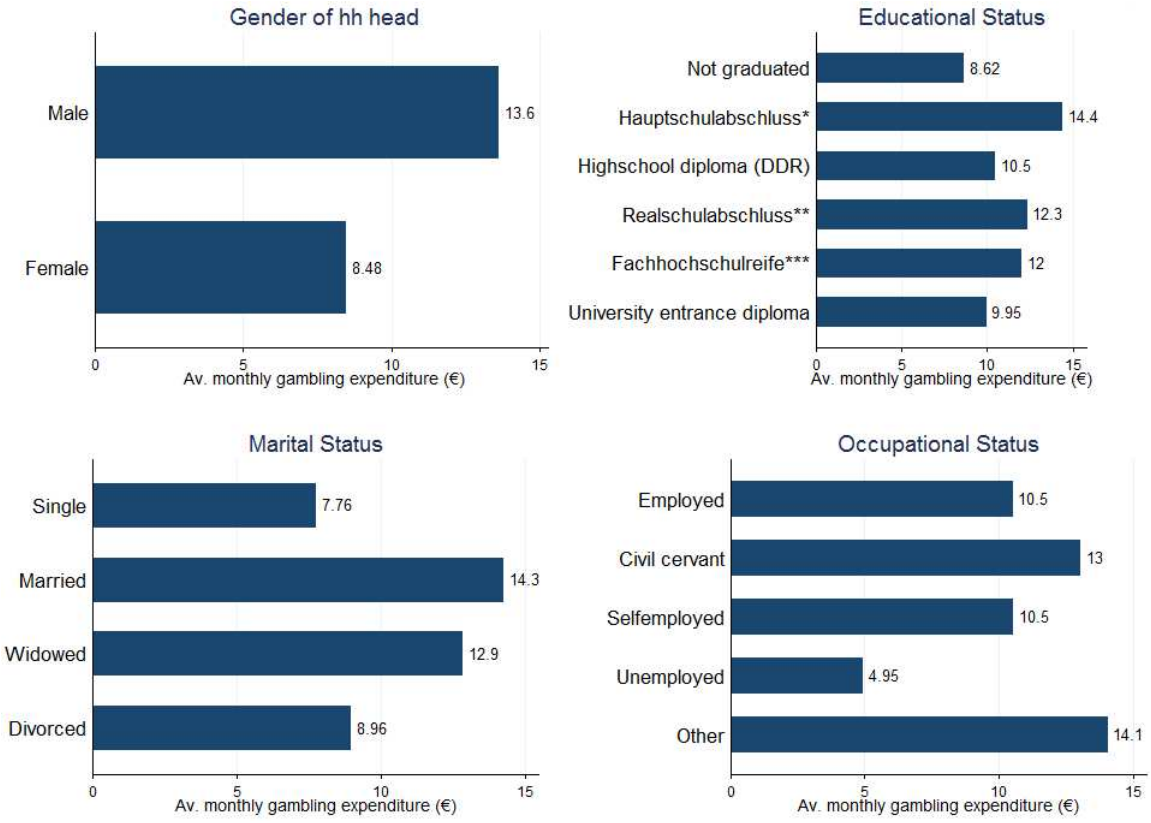
Figure 1: Comparative statistics of absolute and relative average monthly gambling expenditure and age, income and expenditure data




gambling and alcohol and tobacco consumption hints to a strong positive correlation, the reverse is true for the relationship of gambling to education. The absolute expenditure for health products is increasing with the average monthly gambling expenditure. However, the gambling expenditure relative to income decreases when the expenditure for health related products increases.

Figure 2 displays bar graphs on the average gambling expenditures by gender, educational, marital and occupational status of the household head. The first graph shows that male household heads spend on average about €5 more than their female counterparts. Possibly, this could be attributed to two factors. First, as established in Section 3, men tend to gamble more and are more prone to risk in general. Sec-

Figure 2: Average monthly gambling expenditure by gender, educational, marital and occupational status of the household head (in €)




* Certificate of secondary education
** General certificate of secondary education
*** Advanced technical college entrance qualification

ond, female household heads are more likely to be lone parents. As a lone parent an individual might be less likely to gamble because they have a greater sense of responsibility for the time and funds of the household and can presumed to have less excess money to spend on recreational activities. The bottom left bar graph depicts average household gambling expenditures by the marital status of the household head. On average, singles and divorcees spend the least on gambling. This is a rather unexpected result. One might assume that singles have more spare time to gamble as well as face less responsibility in how to allocate their time and money. On the other hand, it is plausible to assume that married households spend more on gambling simply because on average they have a higher combined income. Besides, one has to take into account that the data provides no information on the gambling products purchased. It seems likely for married couples to engage in gambling types such as the lottery and then spend more on each play combined than a single household would, solely due to their higher income. You can support this hypothesis by the number for widowed household heads. They spend on average around €5 and €4 more than single and divorced household heads, respectively. A widowed household

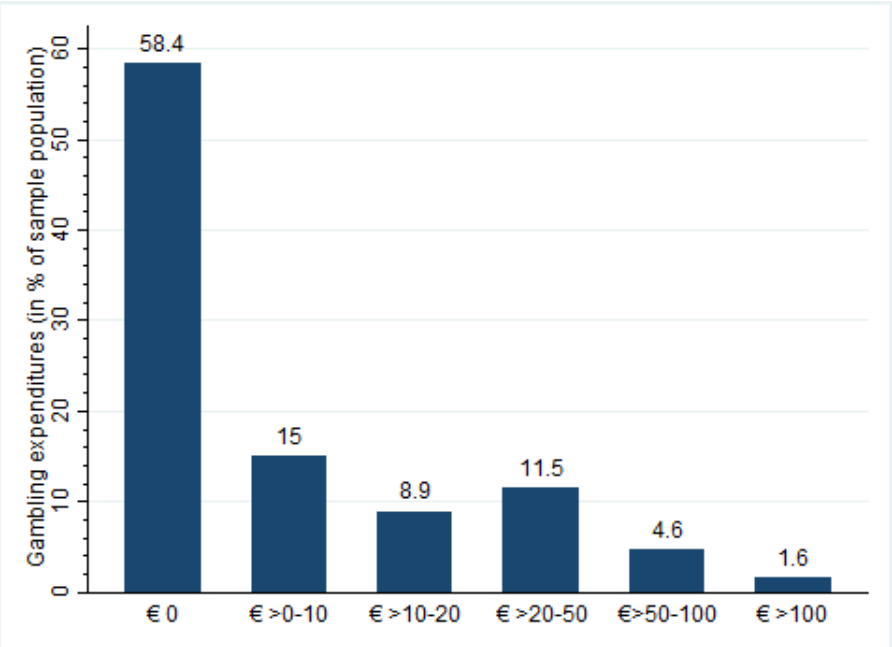
head is typically older than the average single or divorcee. Hence, it is probable that his/her income is higher and that the higher spending can be accounted for by that assuming he/she is not retired. With respect to the educational status the figures are not surprising. The average monthly expenditure on gambling decreases with the degree of education as found by Sawkins and Dickie (2002), Borg et al. (1991) and Hansen (1995) among others.[15] An exception to this pattern are individuals who did not graduate from any secondary education. However, these subjects only make up 0.4 % of the whole sample. Concerning occupational groups, civil servants spend on average more than any other identified occupational group on gambling. Our only notion on how to explain this, is that it could be the case that on average civil servants have a higher salary than the other depicted groups. This can be explained by the positive correlation of gambling expenditure and income. Furthermore, we find it surprising that self-employed individuals spend on average less than employed individuals. We suspected that self-employed individuals spend more on gambling because they have a higher willingness to take risks in general (Dohmen et al. 2011). It will be interesting to separate the effect of self-employment from a possible influence of income on gambling expenditure in the further regression analysis. Unemployed household heads spend on average the least on gambling products which is presumably due to a lack of excess funds for recreational activities.
Overall, the comparative statistics tend to support previous findings on gambling demand as were mentioned in Section 3. However, it will be interesting to examine the relationships of gambling expenditures and marital and occupational status closer and separate their effect from the influence of income.

# 5 Identification Strategy for Censored Data

This section provides the general estimation framework used for both the identification of effects determining gambling behaviour on the household as well as on the individual level. From Figure 3, it is obvious, that there is a high fraction of responding households for whom the observed outcome is zero (58.4 %). The dependent variable is constrained at zero and the observations cluster at this point.
This type of distribution is referred to as being censored (e.g. Greene 2003, p. 869). The distribution of gambling expenditure is censored to the left from below. While the data is censored at zero, the zeros in the sample represent a choice made by the agent. They neither represent missing values or non-response outcomes nor a decision out of control of the agent. Therefore, the zeros in the sample are the result of an optimal choice of the household (Humphreys 2013). That is, a zero outcome is the solution to a constrained utility maximization problem. It represents a corner

[15]The high school diploma obtained in the German Democratic Republic is comparable to an advanced technical college entrance qualification (Fachhochschulreife).

Figure 3: Frequencies of gambling expenditures by intervals




solution and we speak of corner solution outcomes (Wooldridge 2002, p. 517-520). Censored distributions are defined using latent variables. Suppose that the variable on gambling expenditure $y$ is censored at zero, then

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

where $y^*$ is a latent variable observed for values greater than zero and censored otherwise. In this setting an Ordinary Least Squares estimation (OLS) produces inconsistent estimates both on the complete sample as well as on the subsample for $y > 0$ (Wooldridge 2002, p. 524, 525). Instead, to account for the inconsistency we use the standard censored Tobit model or Type I Tobit model (after Tobin 1958).

**Tobit Model** .

Whenever the data presents a corner solution outcome for the dependent variables, a Type I Tobit Model is the most common approach to remedy the deficiencies of OLS. The model is described by the equations

$$y_i^* = x_i\beta \ + \ u_i, \qquad u_i \sim N(0, \sigma^2)$$

$$y_i = max\{0, y_i^*\},$$

with $i$ randomly drawn from the population. The corresponding log-likelihood func-

tion is given by

$$\ln \mathcal{L}(\beta, \sigma) = \sum_{i=1}^{N} \left\{ \underbrace{d_i \left[ \ln \phi \left( \frac{y_i - X_i\beta}{\sigma} \right) - \ln \sigma \right]}_{Uncensored} + \underbrace{(1 - d_i) \left[ \ln \Phi \left( \frac{X_i\beta}{\sigma} \right) \right]}_{Censored} \right\},$$

where $d_i$ is an indicator function ($d_i = 1$ if $y^* > 0$ and 0 otherwise), $\Phi$ is the standard normal cumulative distribution function and $\phi$ is the standard normal probability density function. The two parts of the log-likelihood function correspond to the classical regression of uncensored variables and the relative probability of an observation being censored, respectively. Thus, $\beta$ estimates the (linear) effect of $X_i$ on the latent variable $y_i^*$, not on the observed outcome $y_i$. The Tobit model makes the same assumptions about the distribution of the error term as OLS: $u_i \sim N(0, \sigma^2)$. Arabmazar and Schmidt (1981 and 1982) show among others that the Tobit estimator is inconsistent when the assumption of normally distributed errors is violated. Then the respective marginal effects when censoring at zero and given normally distributed errors are:

$$\frac{\partial E[y^*|x]}{\partial x_k} = \beta_k \tag{1}$$

$$\frac{\partial E[y|x]}{\partial x_k} = \beta_k \; \Phi \left( \frac{x\beta}{\sigma} \right) = \beta_k \; Prob \; [y > 0] \tag{2}$$

$$\frac{\partial E[y|x, y > 0]}{\partial x_k} = \beta_k + \beta_k \frac{\partial \lambda(c)}{\partial c}, \qquad \text{with } c = \frac{x\beta}{\sigma} \;\text{ and }\; \lambda(c) = \frac{\phi(c)}{\Phi(c)} \tag{3}$$

$$Prob \; [y > 0] = \Phi \left( \frac{x\beta}{\sigma} \right) \tag{4}$$

$$\frac{\partial Prob(y > 0)}{\partial x_k} = \phi \left( \frac{x\beta}{\sigma} \right) \frac{\beta_k}{\sigma} \tag{5}$$

Equation (1) yields the marginal effect of $x$ on the latent variable $y^*$ which is given by coefficients $\beta_k$ estimated in the Tobit regression. However, in a Tobit regression $\beta_k$ is of no informative value. We are mostly interested in the effect of the independent variables on the actual outcome $y$, rather than the constructed latent variable $y^*$. This effect is given by equation (2). It is the effect on the latent variable $\beta_k$ times the probability of an observation being uncensored given by equation (4). As you can see from equations (2) and (4), if the probability of an observation being greater than zero $\Phi(x\beta/\sigma)$ is equal to one, the marginal effect of $x$ on the observed outcome $y$ reduces to $\beta_k$ and the Tobit equation is equivalent to that from an OLS. Equation (3) corresponds to the marginal effect on the observed outcome conditional on being

uncensored, where $\lambda(c)$ is called the inverse Mills ratio formulated by Mills (1926). Equation (5) follows from (4) and gives the marginal effect on the probability that an observation is uncensored.

A drawback of the Type I Tobit Model is that it supposes that a single mechanism governs both the participation decision ($y = 0$ or $y > 0$) as well as the amount decision given $y > 0$. In the case of gambling expenditure, this means that the questions "Does someone choose to gamble?" and "How much would someone spend if she/he gambles?" are determined by the same underlying process. In a standard Tobit model the direction signs of partial effects will be the same for both the probability that $y > 0$ and the expectation of $y$ because they are different multipliers of $\beta$. Consequently, it is impossible for independent variables to have a negative effect on the participation decisions and a positive effect on the amount decision or vice versa. A remedy to this drawback are so called Hurdle models. Hurdle models are an alternative which allow both decisions to be determined by separate processes, where the hurdle is whether or not to choose $y > 0$.

**Cragg's Alternative to the Tobit Model**

Cragg (1971) proposed a Hurdle model which nests the usual Type I Tobit model. The first step is a probit model to determine the probability of $y > 0$, where the conditional independence assumption is assumed to hold. Second, $(y|x, y < 0)$ follows a truncated normal distribution. Therefore, this model is also often referred to as the Truncated Normal Hurdle (TNH) model. The density of $(y|x, y < 0)$ is equal to

$$f(y|x, y < 0) = \frac{\phi[(y - x\beta)/\sigma]/\sigma}{\Phi(x\beta/\sigma)},$$

where the term $[\Phi(x\beta/\sigma)]^{-1}$ ensures that the density integrates to unity over $y > 0$. Then, the density of $y$ conditional on $x$ becomes

$$f(y|x, \beta, \gamma) = [-1\Phi(x\gamma)]^{1[y=0]} \left[\Phi(x\gamma)\frac{\phi((y - x\beta)/\sigma)/\sigma}{\Phi(x\beta/\sigma)}\right]^{1[y>0]}.$$

This nests the Tobit density of $y$ for $\gamma = \beta/\sigma$. Thus, one can test the null hypothesis that the single underlying process determining the Tobit model is supported by the data. The log-likelihood function is given by:

$$\begin{aligned}\ln \mathcal{L}(\gamma, \beta, \sigma) = \sum_{i=1}^{N} \quad & 1_{[y_i=0]} \ln[1 - \Phi(x_i\gamma)] + 1_{[y_i>0]} \ln[\Phi(x_i\gamma)] \\ & + 1_{[y_i>0]} \left[-\ln\left[\Phi\left(\frac{x_i\beta}{\sigma}\right)\right] + \ln\left[\phi\left(\frac{y_i - x_i\beta}{\sigma}\right) - \ln(\sigma)\right]\right].\end{aligned}$$

From the log-likelihood function $\gamma$ can be estimated by a probit model and $\beta, \sigma$ by a truncated normal regression model.

# 6 Determinants of Household Gambling Expenditure

## 6.1 Model Specification and Sample Restrictions

In the following the included regressors and the functional form of the conducted regressions are presented. Furthermore, we discuss a sample restriction excluding the highest one percent of observations on gambling expenditure. In each of the following regressions monthly gambling expenditure serves as the dependent variable. On the right hand side we use variables that were established to have an impact on gambling behaviour in Section 3. The socio-demographic factors include the gender of the household head, a dummy variable for whether or not the household head is married[16], as well as the number of underaged children within the household. These variables should represent conclusive information on the household composition. Additionally, the age of the household head is included. Furthermore, due to the ambiguous findings on the impact of age on gambling expenditure, we include age squared to allow for a non-linear relationship. The data does not provide information on the ethnicity of the respondents. There is only data on whether a subject holds the German citizenship. However, this information provides little insight since 98.3 % of the sample hold a German citizenship. Additionally, the dataset provides no information about the religious beliefs of the individuals. One could derive some individual's religious affiliation from tax data but this is only possible for individuals which are Roman-Catholic or Protestant. From the data we cannot conclude whether the remaining subjects are irreligious or affiliated with any other religion. Nevertheless, we include a dummy variable for Christianity as a control to avoid omitted variable bias.[17] Furthermore, we include a dummy variable for whether or not the household resides in a municipality of a population of 100,000 or above.[18] The socio-economic variables include information about the education, the household head's occupational status and income of the households. Education is measured by a dummy variable that is equal to one if the household head completed the German university entrance diploma (Abitur) or the advanced technical college entrance qualification (Fachhochschulreife).[19] Occupational status is measured

[16] Civil unions are subsumed under the term marriage in this context.

[17] The problem with this approach is that unemployed individuals do not pay church taxes. Thus, the approach probably understates the number of individuals with Christian beliefs.

[18] In the following we refer to a municipality of a population of 100,000 or above to as a city.

[19] We attribute the advanced technical college entrance qualification to an university entrance qualification because both graduation certificates qualify for a university-type tertiary education.

by dummy variables for unemployment, self-employment, retirement and whether or not the household head is a civil servant (including professional military). The dummy variable for unemployment includes housewives and househusbands. The influence of income is measured by household's net income in logs to account for the high skewness of income within the sample population. From Figures 4 and 5 in the appendix it becomes clear that the log transformation reduces the skewness of income and the distribution approximates a normal distribution more closely. We do not have to edit the data to avoid missing values that arise from ln(0) as we do not observe any net income equal to zero.

To control for systematic differences between the quarter of the year the survey was taken in, quarterly dummies are used. To account for geographic differences we include a dummy variable for whether or not the household resides in the new federal territories (including Berlin).

As stated in Section 4, the gambling expenditures in the top one percent of the distribution are considerably higher than within the first 99 percentiles. We suspect that these values are not chosen rationally as the optimal household choice but are chosen at random. One could assume that the high gambling expenditures are due to a higher income of the respective households. It is true that the mean income of these households is on average higher than the mean income of the full sample. Nevertheless, when investigating a scatterplot of gambling expenditures and income, there seems to be no clear positive relationship between the amount of gambling expenditures and income. We conclude that income is not the only driver of those extreme values. Rather, we suspect that the amounts of gambling expenditure are mostly chosen arbitrarily and some of these households comprise problem or pathological gamblers.

Dickerson (1993) declares that gambling addiction impairs the control over gambling. He classifies impaired controls as behaviour where the gambler has lost or spent more than intended or found it hard to give up or stop. If we assume that the observations within the top percentile of gambling expenditures are subject to impaired control, it is advisable to analyse them separately from the rest of the sample. The amount spent on gambling by those households is therefore likely to have been chosen arbitrarily. Including these observations would bias the results of the following regressions.

Furthermore, Tourangeau's (2000) results suggest that frequent gamblers report their gambling behaviour less accurately. Given this finding, the reporting error is likely to be higher in the top percentile of gambling expenditures than in the rest of the sample.

Thus, we conduct regressions on both the full sample and the sample restricted to the bottom 99 percentiles., to see whether and to what extent the regression results

are affected.[20] Table 8 in the appendix displays the means for the regressors in the Tobit estimation for the full sample, the first 99 percentiles and the highest one percent of the sample. As one can see, the characteristics of the average household in the top percent differ from the average household within the full sample. The average household head in the highest percentile is by far less likely to be female, is on average 6 years older, is more likely to be married and less likely to have a higher secondary education. Furthermore, the households average income is substantially higher. The household head is more likely to be retired or a civil servant but less likely to be either unemployed or self-employed. The variation in income and retirement is likely to stem from the higher average age.

## 6.2 Regression Results and Robustness

The regression results of the Tobit regression on the full sample are given in column (1) in Table 3. The results of the regression on the sample excluding the highest percentile of gambling expenditures are presented with and without controls in column (2) and (3) of Table 3, respectively.[21] As one can see from Table 3, the results of the Tobit regression on the full and the restricted sample display the same signs for all coefficients but greatly differ in magnitude. Most coefficients of regression (1) are around one third bigger than those of the regression on the restricted sample (2). The directions of the greater magnitude are in line with the deviations from the means of the full sample from the highest percentile given in Table 8. Including the top one percent greatly affects the coefficients of the regressions in the direction of their means. As stated in the previous section, we assume that the highest observations on gambling expenditure are not generated by a optimal household choice and in contrast the lower observations might be subject to impaired control and a higher reporting error. For this reasons we proceed to evaluate the marginal effects only for the regression on the restricted sample.

The results of the Tobit regression show statistically significant coefficients for all variables but the marriage and retirement dummy variables. Note that the signs of the coefficients are in line with the previous findings on gambling behaviour (see Section 3). Being female, a higher education, the number of dependants (here: children) in the households and the dummy variable for home-ownership: all have a negative effect on gambling expenditure. The regression also confirms the previous findings that gambling expenditure increases with income. Additionally, the coefficient for being a civil servant is positive, as was suggested by Albers and Hübl (1997). A

[20] Another possible approach to exclude problem gamblers from the data would be to exclude any observations who spend a proportionally high amount of their income on gambling. However, we did not proceed with this method because we could not find a reliable threshold.

[21] For reference the results of an OLS estimation on the full as well as the restricted sample are given in Table 9 in the appendix.

Table 3: Tobit Regression Results on the Full and the Restricted Sample

| | (1) Full Sample | (2) Restricted Sample | (3) Restricted Sample |
|---|---|---|---|
| Hh head is female | -6.966*** | -4.408*** | -4.431*** |
| | (-9.84) | (-9.32) | (-9.33) |
| Age of hh head | 2.289*** | 1.595*** | 1.593*** |
| | (16.63) | (17.39) | (17.36) |
| Age of hh head squared | -0.0167*** | -0.0118*** | -0.0118*** |
| | (-12.58) | (-13.37) | (-13.34) |
| Hh head is married | 0.773 | 1.783*** | 1.810*** |
| | (0.99) | (3.41) | (3.44) |
| Children <18 years | -6.439*** | -4.507*** | -4.507*** |
| | (-13.96) | (-14.69) | (-14.68) |
| Hh lives in a city | -1.839** | -1.239** | -1.219** |
| | (-2.70) | (-2.73) | (-2.68) |
| Log(net income) | 18.14*** | 11.53*** | 11.41*** |
| | (25.20) | (23.96) | (23.19) |
| Hh head completed university entrance diploma | -8.444*** | -6.056*** | -6.063*** |
| | (-13.17) | (-14.15) | (-14.14) |
| Hh head is unemployed | -8.903*** | -7.079*** | -7.058*** |
| | (-5.11) | (-6.10) | (-6.05) |
| Hh head is self-employed | -5.620*** | -4.983*** | -4.926*** |
| | (-3.43) | (-4.55) | (-4.46) |
| Hh head is a civil servant | 3.380** | 2.245** | 2.275** |
| | (3.12) | (3.09) | (3.13) |
| Hh head is retired | 2.152 | 1.754* | 1.784* |
| | (1.88) | (2.30) | (2.31) |
| Hh owns house/apartment | -5.846*** | -2.694*** | -2.647*** |
| | (-7.91) | (-5.45) | (-5.34) |
| Controls | No | No | Yes |
| Constant | -221.8*** | -143.5*** | -142.7*** |
| | (-34.54) | (-33.49) | (-32.86) |
| Sigma | 52.24*** | 34.87*** | 34.87*** |
| | (174.44) | (167.92) | (167.92) |
| Observations | 42775 | 42351 | 42351 |
| Log-likelihood | -108,441.42 | -100,340.43 | -100,382.28 |

$t$ statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The table represents the regression results of a Tobit regression on the full sample without controls (1), the restricted sample without controls (2) and the restricted sample with controls (3). The controls include dummies for Christian belief, residing in the new territory (incl. Berlin) and dummies for the quarter of 2013 the individual reported in.

negative coefficient for self-employment is somewhat surprising. Although previous literature does not hint at how self-employment influences gambling behaviour, one would assume that being self-employed implies a higher willingness to take risks and in turn a higher affinity towards gambling. Also unemployment seems to decrease gambling expenditures even when separated from a possible income effect. Both coefficients for age and age squared, are statistically significant with different signs. The coefficients suggest that gambling expenditure increases with age but at a diminishing rate. A more thorough evaluation of the coefficients and the resulting marginal effects are given in Section 6.4.

In the following, we present several robustness checks to evaluate the results of regression (2) in Table 3. As a specification test for the model estimated in (2), we performed a link test as suggested by Pregibon (1980) based on Tukey (1949).[22] The link test reveals no problems with the specification.

The results of the Tobit regression (2) are complemented by a Tobit regression including the aforementioned controls (3) on the same restricted sample. Note that the inclusion of controls does not change the coefficients to a great extent. A likelihood ratio test rejects the hypothesis of a better model fit with controls at the 1 % level and reveals that adding the controls does not result in a statistically significant improvement in the model fit. Additionally, it is to mention that the coefficients for all control variables revealed themselves to be not significant at the 1 % level.[23] Therefore, we proceed with our analysis without adding controls (including squared logarithmic income).

Next, we check the validity of the Tobit assumption that a single mechanism governs both the participation decision ($y = 0$ or $y > 0$) as well as the amount decision given that $y > 0$. We do this by estimating a Truncated Normal Hurdle model as introduced in Section 5. We employ the same independent variables as in the Tobit estimation and conduct the regression on the sample restricted to the bottom 99 % of observations on gambling expenditures.[24] Table 4 displays the coefficients of

[22] The idea of the link test is that if a regression is correctly specified, one should not be able to find any additional independent variables that are significant (except by chance).

[23] Following Crowley et al. (2012) we took into consideration that income could be entering the estimation equation polynomially. Previous research has shown that gambling expenditure is regressive. While it increases with income, it is proposed to do so at a diminishing rate. The comparative statistics of the relationship of gambling expenditure and income as shown in the top right graph in Figure 1 also hint to the fact that a income might effect gambling behaviour non-linearly. To take this possibly regressive effect on income into account, we conducted another Tobit regression using the same dependent variables as in regression (2) of Table 3 but including the squared logarithm of net household income. We again conducted a link-test with the new model specification. The link-test rejected the hypothesis of a correctly specified model at the 0% level.

[24] Unfortunately, when we first tried to estimate the equation, the maximum likelihood estimator did not converge (after 100 iterations). A possible explanation for a lack of convergence is using the same independent variables in both "hurdles" of the estimation. We tried several logic combinations to assign the variables to the hurdles but convergence could still not be achieved. A lack of convergence can also be an indication for too many unexpected responses. In the regression equation at hand the inability to converge is most likely due to extreme values in the observations

the first and second tiers of the TNH regression. First, note that all statistically significant coefficients display the same sign for both tier 1 and tier 2. Only the signs of the coefficients for the dummy variables for self-employment differ. However, it is shown to be insignificant at the 5 % level. Second, one can observe that all of the statistically significant coefficients have the same direction as in the Tobit regression. Additionally, remember that for $\beta/\sigma = \gamma$ the TNH model reduces to the Tobit model. If we compare the figures for $\beta/\sigma$ to $\gamma$, the numbers for both values are similar in magnitude for most independent variables. Given this finding, it is likely that the TNH density function reduces to the Tobit density function for the given regression equation. We therefore do not reject the Tobit assumption that the same underlying process generates both the participation and the amount decision. Another problem could arise due to neglected simultaneity. As was established in Section 3, income has a significant effect on gambling behaviour. Note however, that it is plausible to suggest that the reverse is also true. First, consider the extreme case of a pathological gambler. Although gambling expenditures will rise with income, independent of the level of income the individual will gamble. On the other hand, if the gambling expenditures are relatively high to income, income will significantly decrease with gambling expenditure. On the contrary, gambling expenditure could also yield a positive effect on income. Consider an individual beating the odds and winning substantially. Then, his income is positively influenced by his gambling expenditure. Whereas, it might be that gambling expenditure in turn increase after a person wins. In these cases, simultaneity can be an issue. Therefore, we need to check if this is the case in the EVS data. In order to do so, we employ a model which fits a Tobit model where the covariate of income is endogenously determined. The estimation results are given in Table 10. A Wald-test of exogeneity rejects the null hypothesis of endogeneity at the 5% level. Therefore, we include log of income in the specification.

## 6.3 Heteroskedasticity

The consistency of the Tobit model depends on the assumption that the latent variable $y*$ has a homoskedastic normal distribution. Heteroskedasticity and non-normality result in the Tobit estimator $\hat{\beta}$ being inconsistent for $\beta$ (Wooldridge 2002, p. 533). However, the magnitude of this bias is unknown. Consequently, we test for

and heterogeneity in the observations of the top 1 %. All the regressors used in the model are either dummy variables or display a close representation of a normal distribution with small variances; age, age squared and log(net income). Thus, we concluded that the non-convergence must be due to the distribution of household gambling expenditure. Suggested remedies to the problem are either a transformation of the respective variable or rescaling it. To facilitate convergence we tried a logarithmic transformation of household gambling expenditures as well as rescaling them by 1/100 and 1/1000. None of these three approaches were fruitful. Thus, we cannot produce regression results on the full sample.

Table 4: Regression Results of the TNH regression

| | TNH | |
|---|---|---|
| | Tier 1 | Tier 2 |
| Hh head is female | -0.101*** | -40.01*** |
| | (-6.81) | (-6.04) |
| Age of hh head | 0.0435*** | 13.91*** |
| | (15.42) | (7.75) |
| Age of hh head squared | -0.000330*** | -0.0980*** |
| | (-12.04) | (-6.85) |
| Hh head is married | 0.0659*** | -6.012 |
| | (3.99) | (-1.06) |
| Children <18 years | -0.102*** | -41.12*** |
| | (-10.76) | (-7.27) |
| Hh lives in a city | -0.0583*** | 8.356 |
| | (-4.08) | (1.68) |
| Log(net income) | 0.275*** | 93.05*** |
| | (18.31) | (9.40) |
| Hh head completed university | -0.193*** | -18.496*** |
| entrance diploma | (-14.27) | ( 4.902) |
| Hh head is unemployed | -0.239*** | -45.23* |
| | (-6.82) | (-2.37) |
| Hh head is self-employed | -0.190*** | 9.211 |
| | (-5.58) | (0.74) |
| Hh head is a civil servant | 0.0706** | 9.931 |
| | (3.06) | (1.29) |
| Hh head is retired | 0.0232 | 15.04 |
| | (0.95) | (1.94) |
| Hh owns house/apartment | -0.0661*** | -18.38*** |
| | (-4.21) | (-3.35) |
| Constant | -3.540*** | -1397.0*** |
| | (-26.97) | (-9.63) |
| Sigma | 79.37*** | |
| | (19.28) | |
| Observations | 42,351 | |
| Log-likelihood | -99,407.474 | |

$t$ statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The table displays the results of a Truncated Normal Hurdle (TNH) regression. The dependent variable is average monthly household gambling expenditure. Tier 1 corresponds to the first hurdle, where the participation decision ($y > 0$) is determined via probit. Tier 2 corresponds to the second hurdle, where the amount decision (how high is $y$ given $y > 0$) is determined via a truncated normal regression.

non-normality and heteroskedasticity. We compute a Langrance Multiplier statistic for testing the Tobit specification against the alternative of a model that is non-linear in the regressors and contains an error term that can be both heteroskedastic and non-normally distributed. The test rejects the null hypothesis. This suggests that the error term in the estimation of the Tobit model in (2) is subject to heteroskedasticity and non-normality.[25]

To approach this problem, we estimate a Tobit model allowing for multiplicative heteroskedasticity as proposed by Petersen and Waldman (1981). Column (1) of Table 5 displays the results of a regression allowing for multiplicative heteroskedasticity in the Tobit model. Here, the variance is assumed to be of the form $\sigma_i^2 = exp(x_i\alpha)$. A test of the hypothesis that $\alpha$ equals zero can be based on the likelihood ratio statistic (Greene 2003, p. 768). For the two regressions the corresponding likelihood ratio statistic is $-2(100,272.41 - 100,340.43) = 136.0337$. The statistic has a limiting $\chi^2$ distribution with 1 degree of freedom. The hypothesis that $\alpha = 0$ is rejected. When compared to the initial Tobit regression assuming homoskedasticity, the coefficients for $\beta$ of both regressions are very similar. Also the signs of the coefficients are equal for all of the statistically significant coefficients. This hints to the conclusion that the coefficients of the Tobit regression assuming homoskedastic and normally distributed errors are close to the true values. Additionally, one has to be careful when comparing the estimates of $\beta$ of both models. Although, the $\beta$ coefficients may differ, the marginal effects $\partial E(y|y > 0)/\partial x$ and $\partial E(y)/\partial x$ can be very similar.

As a second measure to test the robustness of the results given the found heteroskedasticity and non-normality we employ the censored least absolute deviations estimator (CLAD). The CLAD estimator was introduced by Powell (1984). Given a data set censored at zero, the CLAD estimator is based on the minimization of the sum of absolute residuals. Unlike the Tobit model, Powell's CLAD estimator is consistent and asymptotically normal for a wide range of error distributions. The estimator considers alternative conditional moments that are less altered by censoring. It assumes an error term with a conditional median equal to zero. The CLAD model is described by the equation

$$y^* = x\beta + u \quad \text{with} \quad Median(u|x) = 0,$$

which implies that $Median(y^*|x) = x\beta$. Then $\beta$ can be estimated without distributional assumptions for the error term and without assuming that $x$ and $u$ are independent.

The use of the CLAD model is problematic if the majority of observations is within the censored range of the data. The CLAD model exclusively yields distinct regres-

[25] The heteroskedasticity as well as non-normality test statistics are by far smaller for the Tobit regression on the sample restricted to the bottom 99 percentiles than the regression on the whole sample.

Table 5: Regression results of a Tobit estimation allowing for multiplicative heteroskedasticity and a CLAD estimation

| | (1) Tobit Heteroskedastic | (2) CLAD |
|---|---|---|
| Hh head is female | -4.546*** | -4.767*** |
| | (-9.83) | (-9.98) |
| Age of hh head | 1.549*** | 1.064*** |
| | (17.49) | (9.41) |
| Age of hh head squared | -0.0114*** | -0.00747*** |
| | (-13.41) | (-7.50) |
| Hh head is married | 1.718*** | 0.253 |
| | (3.33) | (0.51) |
| Number of children <18 years | -4.709*** | -2.832*** |
| | (-14.96) | (-9.01) |
| Hh lives in a city | -1.134* | -1.888*** |
| | (-2.54) | (-4.38) |
| Log(net income) | 9.797*** | 9.355*** |
| | (19.45) | (19.39) |
| Hh head completed university entrance diploma | -5.955*** | -5.922*** |
| | (-14.09) | (-14.12) |
| Hh head is unemployed | -6.581*** | -6.349** |
| | (-6.06) | (-3.14) |
| Hh head is selfemployed | -4.472*** | -5.406*** |
| | (-4.06) | (-4.78) |
| Hh head is a civil servant | 2.189** | 1.304 |
| | (2.93) | (1.89) |
| Hh head is retired | 1.612* | 3.401*** |
| | (2.17) | (4.82) |
| Hh owns house/apartment | -2.722*** | -3.053*** |
| | (-5.61) | (-6.56) |
| Constant | -127.8*** | -94.41*** |
| | (-28.79) | (-17.84) |
| Sigma | 30.69*** | |
| | (80.48) | |
| $\alpha^b$ | 0.0000331*** | |
| | (11.35) | |
| Observations | 42,351 | 7,606 |
| Log-likelihood | -100,272.41 | |

$t$ statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The table displays the results of a Tobit regression allowing for multiplicative heteroskedasticity (1) and a Censored Least Absolute Deviations (CLAD) regression (2). The dependent variable is monthly average household gambling expenditure. Regression (1) is conducted on the sample restricted to the bottom 99 percentiles of the distribution for average monthly household gambling expenditure. Regression (2) is conducted on a random subsample of 10,000 observations, where 2,394 observations where dropped by the iterative linear programming algorithm.

sion results if the median of the regression function is positive and not within the censored part of the observations. This is not the case here (see Table 2). However, Powell extended his model in a way such that any other quantile can be used to conduct the estimation (1986). Horowitz (1988) showed that given a highly censored sample the expanded model using quantiles above the median yields satisfying results. In the performed CLAD regression we chose the 70th percentile instead of the median. We follow Rogers (1993) for the CLAD estimator and use bootstrap estimates of the standard errors. Due to the bootstrapping technique and the volume of observations the CLAD estimation is computationally highly intensive.[26] In order to produce results faster, we use a random subsample of 10,000 observations. This subsample consist of around one quarter of all observations. We argue that, given the volume of observations still in the subsample, an estimation on this subsample is still viable to be compared to the Tobit results. The results of the the CLAD estimation are given in column (2) of Table 5.[27] As one can see in Table 5 the results of the CLAD estimation are very similar to that of the Tobit under homoskedasticity as well as the Tobit model allowing for heteroskedasticity.

Unfortunately, neither the results of the Tobit regression allowing for heteroskedasticity nor the CLAD regression are informative. The CLAD estimation is uninformative about $E(y|x)$, $E(y|x, y > 0)$ and $P(y > 0|x)$. Further, the statistical software used to estimate the model (Stata) allowing for multiplicative heteroskedasticity does not provide the necessary coefficients to compute the marginal effects $\partial E(y|y > 0)/\partial x$ and $\partial E(y)/\partial x$, which would be of interest. Hence, we only use the two models to check the robustness of the Tobit estimation and to confirm its coefficients. In order to analyse the marginal effects of the regressors on gambling expenditure we have to rely on the estimation results of the Tobit model.

We conclude from the similarity of the results of the Tobit estimation on the restricted sample from Table 3 and the estimation results from regression (1) and (2) in Table 5, that even under heterosedasticity the Tobit results are valid. In addition, the OLS coefficients should approximate the marginal effect on the observed variable $y$, $\partial E(y)/\partial x$. In the next section it will be shown that this is the case here. Nevertheless, the magnitude of divergence to the true marginal effects remains unknown and the results are a mere approximation of the true effects. The results should be interpreted in this spirit.

[26] The estimation on the full sample as well as the restricted sample could not complete to produce results after running the estimation for more than four hours (after 100 bootstraps and allowing 16,000 iterations to find a solution).

[27] The estimation technique used is Buchinsky's (1991) iterative linear programming algorithm (ILPA). The ILPA first estimates a regression on the full sample and then deletes observations for which the predicted value of the dependent variable is less than zero. Then another quantile regression is estimated on the reduced sample, where negative predicted values are dropped again. That is, observations are dropped if the predicted value of the dependent variable is less than the censoring value. The estimator converges when there are no negative predicted values after two consecutive iterations. Fot that reason the table displays 7,606 observations rather than the full 10,000.

## 6.4 Marginal Effects

The coefficients for $\beta$ given by the Tobit regression cannot be interpreted straightforward as OLS coefficients. As shown by equation (1), $\beta$ only displays the effect on the latent variable rather than the observed outcome. Although we are estimating the effect of the independent variables $x$ on the latent variable $y^*$, $y^*$ does not have a meaningful interpretation. Instead, we are interested in the effects of $x$ on the unconditional expected value of the dependent variable, the expected value of the dependent variable conditional on being uncensored and the probability of being uncensored. These are given by equations (2), (3) and (5) respectively. The results are shown in Table 6.

The Tobit regression produces marginal effects conditional on the independent variables. The marginal effects displayed in Table 6 are given for a discrete change from 0 to 1 for the dummy variables. The marginal effects of the discrete independent variables age of the household head, number of children within the household and income are evaluated at their means. Therefore, the marginal effects have to be displayed for a reference household. Here, the marginal effects given in Table 6 correspond to a a household whose head is an unmarried male who is about 53 years old. He does neither have children living within his household nor does he live in an urban area or in his own property. He did not complete the university entrance diploma (or the advanced technical college entrance qualification) and is in employment. The average net household income corresponds to the mean of the (logarithmic) net household income, which is around €2,900. Therefore, these exact marginal effects are applicable to this individual and may differ for an individual with other characteristics. Further on, we will present the deviations in the effects for different ages of the household head and household income.

This said, an individual with the same characteristics as described, would spend around €2 less on gambling and would be 5% less likely to gamble at all, if he was female. If he was already gambling, being female would reduce his average spendings by around €1.52. Having one child would decrease the average monthly expenditure on gambling by €2 while the probability to gamble at all would decline by 5% per child. This confirms the finding that a higher number of dependants in the household decreases gambling expenditure. It might be that this decrease is due to a reduced willingness to take risks in general and an increase in responsibility due to the dependants.

Considering the socio-economic dummy variables on occupation and education, having a higher secondary education (university entrance diploma or advanced technical college entrance qualification) has the greatest effect on the average gambling expenditure. The reference individual spends around €2.8 less on gambling than a

Table 6: Marginal Effects of the Tobit regression

| | $\beta = \frac{\partial E(y^*)}{\partial x}$ | $\frac{\partial E(y)}{\partial x}$ | $\frac{\partial E(y\|y>0)}{\partial x}$ | $\frac{\partial Prob(y>0)}{\partial x}$ |
|---|---|---|---|---|
| Hh head is female | -4.408*** | -2.048*** | -1.519*** | -0.0502*** |
| Age of hh head | 1.595*** | 0.741*** | 0.550*** | 0.0182*** |
| Age of hh head squared | -0.0118*** | -0.00548*** | -0.00407*** | -0.000134*** |
| Hh head is married | 1.783*** | 0.829*** | 0.614*** | 0.0203*** |
| Children <18 years | -4.507*** | -2.094*** | -1.553*** | -0.0514*** |
| Hh lives in a city | -1.239** | -0.576** | -0.427** | -0.0141** |
| Log(net income) | 11.53*** | 5.358*** | 3.973*** | 0.131*** |
| Hh head completed university entrance diploma | -6.056*** | -2.814*** | -2.086*** | -0.0690*** |
| Hh head is unemployed | -7.079*** | -3.289*** | -2.439*** | -0.0807*** |
| Hh head is self-employed | -4.983*** | -2.315*** | -1.717*** | -0.0568*** |
| Hh head is a civil servant | 2.245** | 1.043** | 0.773** | 0.0256** |
| Hh head is retired | 1.754* | 0.815* | 0.604* | 0.0200* |
| Hh owns house/apartment | -2.694*** | -1.252*** | -0.928*** | -0.0307*** |

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

42,351 observations

The table displays the marginal effects of a Tobit regression of the restricted sample of the bottom 99 percentiles of average monthly household gambling expenditures. The dependent variable is average monthly household gambling expenditure. The marginal effects are given for a discrete change of the dummy variable from 0 to 1 (incl. number of children). The discrete variables are evaluated at their means 53.33 (age), 3,091.14 (age$^2$) and 7.97 (log income).

less educated individual with the same characteristics. This is in line with every previous finding on gambling behaviour. Furthermore, it is noteworthy that while education has the greatest effect on actual spending, being unemployed decreases the probability to gamble at all to the greatest extent, namely by 8 %. Even when the income effect of unemployment is separated, unemployed individuals spend less and are less likely to gamble. We find this result unexpected since previous studies found that the recipients of government payments generally have a higher tendency to gamble (Laitner 1999; Clotfelter and Cook 1989). Furthermore, the finding that self-employment decreases gambling expenditures and the probability to gamble is interesting. Although the effect of self-employment on gambling expenditures is not very large, being self-employed reduces the probability of $y > 0$ by nearly as much as having a child. As previously stated, we would have assumed that self-employed

household heads inhabit riskier attitudes and are thus more likely to gamble.
Another remarkable result, although already confirmed by Albers et al (1997), is that civil servants gamble more than the other occupational groups. The reference individual spends on average €1 more on gambling products and is 2.5 % more likely to gamble than if he was employed in the private sector. It would be interesting to find out more about the cause for this finding. Also striking is the insignificance of the retirement dummy. As suggested by Mcneilly and Burke (2000) retirees in the USA devote more time and funds to gambling. The discrepancy could be explained by cultural and behavioural differences of US and German retirees.
Surprisingly, living in an urban area reduces the gambling expenditure of the reference individual. As stated in Section 3, the previous findings suggest a higher gambling participation in urban areas due to closer gambling venues. In the EVS data, urbanity seems to have a reverse effect. This could be explained by fewer entertainment venues in rural areas. It could be that individuals there turn to gambling as a recreational activity[28] while urban households use other forms of recreation only available in bigger cities. Finally, a strong driver of gambling expenditure is income. This suggests that income is a normal good with consumption increasing with income. Concerning the aforementioned regressiveness of gambling expenditure, our regression results do not support this claim. The results rather show that the effect of income on gambling expenditure increases with higher incomes (see Table 12 in the appendix).
As stated previously, the marginal effects given by the Tobit regression are conditional on $x$. Tables 11, 12 and 13 in the appendix show the marginal effects of the Tobit regression on the observed outcome $y$ at different values of age, the household's net income and by gender of the household head. The magnitude of the marginal effects all increase with age. The same applies to an increase in household net income. With respect to income, the marginal effects are all greater for male household heads than they are for female.
Nevertheless, keep in mind that the Tobit regression is subject to heteroskedastic and non-normally distributed errors. Although the robustness measures point to the validity of the results, they should still be interpreted as a close approximation of the true effects.

[28] This includes gambling which is commonly also available in rural areas such as online gambling or commercial gambling machines.

# 7 Discussion and Conclusion

## 7.1 Discussion of the Results

Overall, our findings on the determinants of gambling expenditure confirm most results of previous international literature on gambling demand presented in Section 3. We find that a female household head spends on average less on gambling than a male. The level of education also has a strong negative effect on gambling expenditure. While the previous literature is divided about the effect of age on gambling behaviour, we find that gambling expenditure increases with the age of the household head but at a diminishing rate. The household composition influences the amount spent on gambling. Married couples tend to spend more than single households, while gambling expenditure decreases with the number of children in the household. Contrary to the results of Tan et al. (2010) and Beckert and Lutter (2008), we find that households in rural areas have higher gambling expenditures than urban households.

One of the strongest drivers of gambling expenditure is income. This suggests that gambling is a normal good with consumption increasing with income. Concerning the aforementioned regressiveness of gambling expenditure, our regression results cannot verify this claim. Our results rather show that the effect of income on gambling expenditure continually increases with higher incomes (see Table 12 in the appendix). On the other hand, the comparative statistics support the notion that, while gambling expenditure increases with income, it does so recessively.

Moreover, the occupation of the household head also influences gambling behaviour. Households with unemployed and self-employed household heads spend less than those with privately employed and household heads who are civil servants. The finding that a self-employment household head has a negative effect on gambling expenditure is surprising. Dohmen et al (2011) find that someone who is self-employed is more willing to take risks in general. Considering the correlation between risk and gambling attitudes as suggested by Mishra et al. (2010), these results need further investigation.

A possible explanation for the surprising results of the effects the occupational status has on gambling expenditures is the sample composition of the EVS. The EVS data suffers from a middle-class-bias (Rebeggiani 2007). Compared to the actual German population the fraction of household heads who are retirees or civil servants is overrepresented in the EVS, while self-employed and unemployed household heads are greatly underrepresented. Additionally, the mean household income in the EVS exceeds mean household incomes in other German data sets, such as the German microcensus.[29] This imbalance is partly due to the fact that the monetary reward for taking part in the survey is credited against social and unemployment benefits.

[29] The microcensus is a representative household survey in Germany.

Thus, the incentive for recipients of governmental transfers to participate in the EVS is weakened. Another potential explanation for the bias is that taking part in the EVS consumes a considerable amount of time. For example, retirees might be more likely to take part in the survey because they have more free time than individuals who are still in the workforce. Additionally, self-employed individuals might not have the spare time to participate. Therefore, the especially surprising results regarding the effects of the occupational status of the household head on gambling expenditures might partially be caused by a middle-class-bias.

The previous literature on gambling demand mainly considers U.S. or Canadian data. Although gambling attitudes in Northern America might be similar in Western Europe, we cannot be sure whether there might be systematic differences due to geographical and cultural disparities. Only a small strand of literature analysing European data sets is available (e.g. Beckert and Lutter 2008; Castrén et al. 2018; Gandullia and Leporatti 2017; Volberg et al. 2004). However, there exist only two empirical studies investigating determinants of gambling behaviour in Germany. The more recent study by Beckert and Lutter (2008) examines the socio-demographic and economic factors which influence lottery purchases using a representative survey. Their findings are similar to ours, stating that income and marriage both have positive effects on gambling while being female and the level of education negatively influence the purchase of lottery tickets. Different is that they find a positive relationship between urbanity and lottery participation. However, their results are obviously not directly comparable to ours. First, they only investigate the determinants of lottery participation and lottery ticket purchases. The determinants of purchasing lottery tickets as a specific form of gambling might differ from those of other gambling products (or gambling expenditure in general, as in our case). Second, they estimate the expenditure on lottery tickets via OLS. Since they face censored data, their results are probably subject to bias. Nevertheless, although the marginal effects might be biased, their findings essentially coincide with ours.

Another study by Albers and Hübl (1997) estimates individual patterns of gambling in Germany. In their study, the authors differentiate between the effects of independent variables on different forms of gambling. They find gambling to be a normal good and reject the hypothesis of gambling being a social demerit. Additionally, their results again confirm most of our findings. They show that the propensity to gamble is smaller for females and higher educated individuals but higher for civil servants than other occupational groups. Our findings differ in that they observe a negative coefficient for home-ownership. However, they only estimate the probability of participation (by a probit model). Thus, the study provides no information about the outcome of the amount decision. Therefore, our findings contribute to research on gambling behaviour in Germany as we analyse gambling in general and use techniques which incorporate censored data issues while obtaining the marginal

effects on actual expenditure rather than just participation.

Furthermore, this paper also contributes to the empirical literature on consumption in a more general sense. Empirical demand and consumption analyses often face censored data. Other than gambling, there are numerous non-indispensable goods such as tobacco, alcoholic beverages, health care, luxury goods or recreational activities which a high fraction of individuals not consuming at all.[30] All of these studies face the same issue of how to deal with censored data as we do with the data on gambling expenditure. Some papers bypass the issue altogether by estimating binary choice models of consumption.[31] These studies produce marginal effects on the participation decision but fail to derive marginal effects on the amount spent. For most goods though, the amount of the expenditure is of interest, rather than just the mere purchasing decision. For example, concerning gambling expenditure one should distinguish between small and high stake players and take the frequency of playing into account. Powell et al. (1999) find that the traits of high stake and excessive gamblers differ to those from occasional or one-time players. Thus, it is of interest to examine what influences the frequency and amount decision.

In order to produce marginal effects for both the participation and the amount decision, typically Tobit or Hurdle models are used with censored data. The calculated marginal effects provide information on both $Prob(y > 0|x)$ and $E(y|x)$. These models rely on the assumption of homoskedastic and normally distributed errors. However, it is fairly common to observe heteroskedasticity in any estimation on a sample as large as the EVS. Thus, many researchers face the problem of heteroskedasticty. To overcome this problem, a robust estimator, such as the CLAD estimator, can be employed. Though, while it produces robust estimation results, the coefficients are of no informative value regarding the marginal effects on the observed outcome. This constitutes a trade-off between retrieving the marginal effects of interest and estimating a robust model. This paper provides an approach of how to proceed if heteroskedasticity is found when applying Tobit models. Although the regression of the Tobit model suffers from heteroskedasticity and non-normality, the similarity of the results of heteroskedasticity robust estimations suggest their validity. This allows us to interpret the marginal effects calculated from the Tobit model, albeit just in an approximative quality.

[30] An overview of empirical studies of consumer demand is provided in Rebeggiani 2007, Chapter 4. On leisure demand, see e.g Pawlowski and Breuer 2012. The demand for tobacco is modelled e.g. by Aristei and Pieroni 2008.

[31] See, among many others, the already cited paper by Albers and Hübl 1997, or Fu and Florkowski 2016 on tobacco and alcohol demand.

## 7.2 Suggestions for Further Research

Empirical literature on gambling demand typically focuses on household determinants of gambling behaviour (e.g. Gandullia and Leporatti 2017, Kitchen and Powells 1991, Tan et al. 2010 or Worthington et al. 2003 and 2007, among others). This is mostly due to the fact that data on (gambling) expenditure is commonly only available at the household level. When analysing the influence of individual traits, the household head serves as the base for individual characteristics. Accordingly, it is implicitly assumed that neither another household member gambles nor are the characteristics of other household members incorporated into the analysis. This harbours inaccuracies. In addition, the analysis on the individual level is important since it grants more insights into the distinctive motivation and determinants of gambling behaviour than an analysis of the whole household does. In the case of gambling, the implicit assumption that only the household head gambles flaws the results especially for married couples. For once, it is assumed that spouses of the household head do not participate in any gambling activity at all. Assuming that most household heads in samples are male, this systematically understates the gambling participation of women. Further, if household heads are female, these women often are single parents. As MacDonald et al. (2004) find, lone parents gamble less than married individuals. This advances the underestimation of female gambling expenditure. As a consequence, the approach of attributing all expenditures solely to the household head leads to inaccuracies in the analysis. This problem persists with nearly all empirical demand studies concerning the consumption of divisible goods. For example, if one considers expenditures on clothes, there are likely to be substantial differences in attitudes and willingness to pay of individual household members.

Nevertheless, as previously mentioned, representative data on individual expenditure is scarcely available. Commonly, microeconomic data on expenditures is only accessible on the household level. In most of these samples, individual socio-economic and demographic information exists only for the household head. Thus, studies of demand behaviour are usually conducted at the household level, where the household serves as a consumption entity. In theoretical literature on demand this is called the unitary model. Doss (1996) rejects the unitary model and proposes more flexible models that allow for intra-household dynamics. Therefore, we propose that further research is needed to analyse individual rather than household (gambling) expenditure.

Another point of concern is the lack of differentiation of gambling expenditures in the EVS data. As findings by Albers and Hübl (1997) suggest, the demographic and socio-economic determinants of gambling participation differ for specific gambling products. For example, they find that being male has a negative effect on the propensity to play TV-lottery while it has a positive effect on on the participation

decision to use gambling machines or visit casinos. Furthermore, most of the coefficients in their regression show the same signs but the magnitudes of the estimated coefficients differ. Thus, there is a difference in the determinants of gambling behaviour for different gambling products. For policy-makers it is of importance to be able to differentiate between the drivers of gambling demand of specific products. Some gambling products might be more socially desirable while others create more severe negative externalities. In order to decide on policies specific to a certain gambling product, findings on the determinants of gambling expenditure on each individual gambling product are required.

Furthermore, the lack of differentiation of the types of gambling in the EVS data makes it difficult to compare the results obtained here to the literature about risk attitudes. There are differences in the risk attitudes of gamblers who choose specific games (Albers and Hübl 1997). The riskiness of a stake varies greatly with the chosen product. For example, compare a 6-out-of-49 lottery to roulette. While the lottery provides a low winning probability, the stakes are low and the potential winnings very high. On the other hand, the winning probability in a game of roulette is considerably higher, while the potential winning is lower. Thus, the (perceived) risk of gambling is dependent on the game played. However, the EVS yields no information on the riskiness of the gambles the individuals participated in. Therefore, one cannot extend the marginal effects found on gambling expenditure to the willingness to take risks or risk attitudes in general. Nonetheless, it would be of interest to retrieve the effects of socio-demographic and socio-economic variables on differentiated gambling products for which the risk is known.

# Bibliography


Afifi, Tracie O., et al. "Demographic and social variables associated with problem gambling among men and women in Canada." *Psychiatry research* 178.2 (2010): 395-400.

Albers, Norman, and Lothar Hübl. "Gambling market and individual patterns of gambling in Germany." *Journal of Gambling Studies* 13.2 (1997): 125-144.

Arabmazar, Abbas, and Peter Schmidt. "Further evidence on the robustness of the Tobit estimator to heteroskedasticity." *Journal of Econometrics* 17.2 (1981): 253-258.

Arabmazar, Abbas, and Peter Schmidt. "An investigation of the robustness of the Tobit estimator to non-normality." *Econometrica: Journal of the Econometric Society* (1982): 1055-1063.

Aristei, David, and Pieroni, Luca. "A double-hurdle approach to modelling tobacco consumption in Italy." *Applied Economics*, 40.19 (2008): 2463-2476.

Auer, Michael, and Mark D. Griffiths. "Self-reported losses versus actual losses in online gambling: An empirical study." *Journal of Gambling Studies* 33.3 (2017): 795-806.

Beckert, Jens, and Mark Lutter. "The inequality of fair play: Lottery gambling and social stratification in Germany." *European Sociological Review* 25.4 (2008): 475-488.

Becker, Tilman. *Soziale Kosten des Glücksspiels in Deutschland.* Peter Lang-Verlag, Frankfurt (2011).

Borg, Mary O., Paul M. Mason, and Stephen I. Shapiro. "The incidence of taxes on casino gambling: Exploiting the tired and poor." *American Journal of Economics and Sociology* 50.3 (1991): 323-333.

Borg, Mary O., and Paul M. Mason. "The budgetary incidence of a lottery to support education." *National Tax Journal* (1988): 75-85.

Buchinsky, Moshe. "Methodological issues in quantile regression." *The theory and practice of quantile regression.* Ph.D. dissertation, Harvard University (1991).

Casey, David M., et al. "The role of family, religiosity, and behavior in adolescent gambling." *Journal of Adolescence* 34.5 (2011): 841-851.

Castrén, Sari, et al. "The relationship between gambling expenditure, socio-demographics, health-related correlates and gambling behaviour— a cross-sectional population-based survey in Finland." *Addiction* 113.1 (2018): 91-106.

Clotfelter, Charles T., and Philip J. Cook. "Implicit Taxation in Lottery Finance." *National Tax Journal* 40.4 (1987): 533.

Clotfelter, Charles T., and Philip J. Cook. "The Demand for Lottery Products." *Working Paper, National Bureau of Economic Research (1989).*

Collins, David, and Helen Lapsley. "The social costs and benefits of gambling: An introduction to the economic issues." *Journal of Gambling Studies* 19.2 (2003): 123-148.

Cragg, John G. "Some statistical models for limited dependent variables with application to the demand for durable goods." *Econometrica: Journal of the Econometric Society* (1971): 829-844.

Crowley, Frank, John Eakins, and Declan Jordan. "Participation, expenditure and regressivity in the Irish lottery: Evidence from Irish household budget survey 2004/2005." *The Economic and Social Review* 43.2, Summer (2012): 199-225.

DHS. "Umsätze auf dem Glücksspiel-Markt in Deutschland von 2005 bis 2016 (in Milliarden Euro)." *Statista - Das Statistik-Portal*, Statista, de.statista.com/ statistik/ daten/ studie/ 5417/umfrage/ umsaetze-auf-dem-gluecksspiel-markt-seit-2005/, Accessed 14. Juli 2018.

Dickerson, Mark. "Internal and external determinants of persistent gambling: Problems in generalising from one form of gambling to another." *Journal of Gambling Studies* 9.3 (1993): 225-245.

Dohmen, Thomas, et al. "Individual risk attitudes: Measurement, determinants, and behavioral consequences." *Journal of the European Economic Association* 9.3 (2011): 522-550.

Doss, Cheryl R. "Testing among models of intrahousehold resource allocation." *World Development* 24.10 (1996): 1597-1609.

Eisen Nong Lin, Seth A., et al. "Familial influences on gambling behavior: an analysis of 3359 twin pairs." *Addiction* 93.9 (1998): 1375-1384.

Fu, Shengfei and Florkowski, Wojciech J. "Polish Household Consumption of Tobacco and Alcohol: A Censored System" Paper prepared for the Southern Agricultural Economics Association Annual Meeting, 2016.

Gandullia, Luca, and Lucia Leporatti. "The demand for gambling in Italian regions and its distributional consequences." *Papers in Regional Science* (2017).

Gupta, Rina, and Jeffrey L. Derevensky. "Adolescent gambling behavior: A prevalence study and examination of the correlates associated with problem gambling." *Journal of Gambling Studies* 14.4 (1998): 319-345.

Greene, William H. *Econometric analysis*. Pearson Education (2003).

Grote, Kent, and Victor A. Matheson. "The Economics Of Lotteries: A Survey Of The Literature." *Working Paper No.* 11-09, *College of the Holy Cross* (2011).

Handelsblatt (Handelsblatt Research Institute). "Bruttospielerträge im gesamten Glücksspielmarkt inklusive Schwarzmarkt in Deutschland im Jahr 2015 (in Milliarden Euro)." *Statista - Das Statistik-Portal*, Statista, de.statista.com/ statistik/ daten/ studie/ 753272/ umfrage/ bruttospielertraege-im-deutschen-gluecksspielmarkt/, Accessed 14. Juli 2018.

Hansen, Ann. "The tax incidence of the Colorado state lottery instant game." *Public Finance Quarterly* 23.3 (1995): 385-398.

Harris, Christine R., Michael Jenkins, and Dale Glaser. "Gender Differences in Risk Assessment: Why do Women Take Fewer Risks than Men?" *Judgment and Decision Making* 1.1 (2006): 48-63.

Haß, W., and P. Lang. *Glücksspielverhalten und Glücksspielsucht in Deutschland. Ergebnisse des Surveys 2015 und Trends.* Forschungsbericht der BZgA. Bundeszentrale für gesundheitliche Aufklärung, Köln (2016).

Horowitz, Joel L. "Semiparametric M-estimation of censored linear regression models." *Advances in Econometrics* 7 (1988): 45-83.

Humphreys, Brad R. "Dealing with zeros in economic data." *Department of Economics, University of Alberta* (2013).

IQVIA. "Umsatzentwicklung auf dem deutschen Pharma-Gesamtmarkt von 2006 bis Q1 2018 (in Milliarden Euro)*." *Statista - Das Statistik-Portal*, Statista, de.statista.com/statistik/daten/studie/158096/umfrage/pharma-gesamtmarkt-umsatzentwicklung-seit-2006/, Accessed 14. Juli 2018.

Kitchen, Harry, and Scott Powells. "Lottery expenditures in Canada: A regional analysis of determinants and incidence." *Applied Economics* 23.12 (1991): 1845-1852.

Kumar, Alok, Jeremy K. Page, and Oliver G. Spalt. "Religious beliefs, gambling attitudes, and financial market outcomes." *Journal of Financial Economics* 102.3 (2011): 671-708.

Ladouceur, Robert, and Michael Walker. "A cognitive perspective on gambling." *Trends in Cognitive and Behavioural Therapies* (1996): 89-120.

Laitner, John. "Means-tested public assistance and the demand for state lottery tickets." *Review of Economic Dynamics* 2.1 (1999): 273-290.

Layton, Allan, and Andrew Worthington. "The impact of socio-economic factors on gambling expenditure." *International Journal of Social Economics* 26.1/2/3 (1999): 430-440.

MacDonald, Martha, John L. McMullan, and David C. Perrier. "Gambling households in Canada." *Journal of Gambling Studies* 20.3 (2004): 187-236.

McNeilly, Dennis P., and William J. Burke. "Late life gambling: The attitudes and behaviors of older adults." *Journal of Gambling Studies* 16.4 (2000): 393-415.

Mikesell, John L. "A note on the changing incidence of state lottery finance." *Social Science Quarterly* 70.2 (1989): 513.

Mills, John P. "Table of the ratio: area to bounding ordinate, for any portion of normal curve." *Biometrika* (1926): 395-400.

Mishra, Sandeep, Martin L. Lalumière, and Robert J. Williams. "Gambling as a form of risk-taking: Individual differences in personality, risk-accepting atti-

tudes, and behavioral preferences for risk." *Personality and Individual Differences* 49.6 (2010): 616-621.

Pawlowski, Tim, and Breuer, Christoph. "Expenditure Elasticities of the Demand for Leisure Services." *Applied Economics* 44.26 (2012): 3461-3477.

Petersen, Dana, and Donald M. Waldman. "The treatment of heteroskedasticity in the limited dependent variable model." *Working Paper, University of North Carolina* (1981).

Powell, James L. "Least absolute deviations estimation for the censored regression model." *Journal of Econometrics* 25.3 (1984): 303-325.

Powell, James L. "Censored regression quantiles." *Journal of Econometrics* 32.1 (1986): 143-155.

Powell, Jeff, et al. "Gambling and risk-taking behavior among university students." *Substance Use & Misuse* 34.8 (1999): 1167-1184.

Pregibon, Daryl. "Goodness of link tests for generalized linear models." *Applied Statistics* (1980): 15-14.

Rebeggiani, Luca. *Personelle Einkommensverteilung, privater Konsum und Wachstum.* Metropolis Verlag, Marburg (2007).

Rebeggiani, Luca. *Deutschland im Jahr Drei des GlüStV. Reformvorschläge zur Regulierung des deutschen Glücksspielmarktes*, Report for the German Private Lottery Association, Hannover (2010).

Rebeggiani, Luca. "Regulierung des deutschen Sportwettenmarktes in komparativer Perspektive: Glücksspielgesetze in der Europäischen Union." *Sport und Sportgroßveranstaltungen in Europa - Zwischen Zentralstaat und Regionen*, Hamburg University Press (2012): 51-78.

Rebeggiani, Luca, and Markus Breuer. "Neue Ordnung, neues Glück? Ordnungs- und fiskalpolitische Aspekte des deutschen Sportwettenmarkts." *Wirtschaftsdienst* 97.9 (2017): 655-663.

Rintoul, Angela C., et al. "Modelling vulnerability to gambling related harm: How disadvantage predicts gambling losses." *Addiction Research & Theory* 21.4 (2013): 329-338.

Rogers, William. "Quantile regression standard errors." *Stata Technical Bulletin* 2.9 (1993).

Sawkins, John W., and Valerie A. Dickie. "National lottery participation and expenditure: preliminary results using a two stage modelling approach." *Applied Economics Letters* 9.12 (2002): 769-773.

Scott, Frank, and John Garen. "Probability of purchase, amount of purchase, and the demographic incidence of the lottery tax." *Journal of Public Economics* 54.1 (1994): 121-143.

Slutske, Wendy S., et al. "Local area disadvantage and gambling involvement and disorder: Evidence for gene-environment correlation and interaction." *Journal of Abnormal Psychology* 124.3 (2015): 606.

Smith, Julie. "Gambling taxation: Public equity in the gambling business." *Australian Economic Review* 33.2 (2000): 120-144.

Statistisches Bundesamt. "Umsatz der Lebensmittelindustrie in Deutschland in den Jahren 2008 bis 2017 (in Milliarden Euro)." Statista - *Das Statistik-Portal*, Statista, de.statista.com/statistik/daten/studie/164959/umfrage/umsatz-der-nahrungsmittelindustrie-in-deutschland-seit-2005/, Accessed 14. Juli 2018.

Storer, John, Max Abbott, and Judith Stubbs. "Access or adaptation? A meta-analysis of surveys of problem gambling prevalence in Australia and New Zealand with respect to concentration of electronic gaming machines." *International Gambling Studies* 9.3 (2009): 225-244.

Tan, Andrew, Steven T. Yen, and Rodolfo M. Nayga. "Socio-demographic determinants of gambling participation and expenditures: evidence from Malaysia." *International Journal of Consumer Studies* 34.3 (2010): 316-325.

Tobin, James. "Estimation of relationships for limited dependent variables." *Econometrica: Journal of the Econometric Society* (1958): 24-36.

Tourangeau, Roger, Lance J. Rips, and Kenneth Rasinski. *The psychology of survey response.* Cambridge University Press, Cambridge (2000).

Tukey, J. W. One degree of freedom for non-additivity. *Biometrics* 5 (1949): 232–242.

Volberg, Rachel A., et al. "Prevalence and risks of pathological gambling in Sweden." *Acta Psychiatrica Scandinavica* 104.4 (2001): 250-256.

Walker, Douglas M., and Andy H. Barnett. "The social costs of gambling: An economic perspective." *Journal of Gambling Studies* 15.3 (1999): 181-212.

Wooldridge, Jeffrey. *Econometric Analysis of Cross Section and Panel Data*. MIT Press, Cambridge (2002).

Worthington, Andrew, et al. "Gambling participation in Australia: findings from the national Household Expenditure Survey." *Review of Economics of the Household* 5.2 (2007): 209-221.

Worthington, Andrew C., et al. "Socioeconomic and demographic determinants of household gambling in Australia." *Discussion Paper No. 156, School of Economics and Finance, Queensland University of Technology* (2003).

Young, Martin, Francis Markham, and Bruce Doran. "Too close to home? The relationships between residential distance to venue and gambling outcomes." *International Gambling Studies* 12.2 (2012): 257-273.

# Appendices

## A Summary Statistics

### A.1 Summary Statistics of the EVS 2013

Table 7: Summary statistics of the EVS 2013

| | Mean | Sd | Min | Max |
|---|---|---|---|---|
| *Demographic characteristics* | | | | |
| Hh head is female | .38 | .49 | | |
| Age of hh head | 53.32 | 15.73 | 19 | 95 |
| Hh head is married* | .52 | .50 | | |
| Number of children <18 years | .35 | .76 | 0 | 7 |
| Hh head is Christian** | .31 | .46 | | |
| Hh owns house/apartment | .51 | .50 | | |
| Hh living in a city*** | .31 | .46 | | |
| Hh lives in the new federal territory† | .24 | .43 | | |
| *Socio-economic characteristics* | | | | |
| Hh head completed Abitur‡ | .48 | .50 | | |
| Hh head is unemployed | .05 | .21 | | |
| Hh head is self-employed | .04 | .19 | | |
| Hh head is a civil servant | .09 | .28 | | |
| Hh head is retired | .31 | .46 | | |
| *Household income* (in €) | | | | |
| Gross hh income | 4,496.5 | 2,967.5 | 56.7 | 29,832.0 |
| Net hh income | 3,461.2 | 2,111.4 | 7.3 | 17,892.7 |
| *Household expenditures* (in €) | | | | |
| Total | 5,642.8 | 6,283.2 | 407.3 | 176,254.7 |
| Gambling | 11.7 | 29.8 | 0 | 1,682.7 |
| Groceries | 318.9 | 175.0 | 0 | 2,061.0 |
| Alcoholic beverages and tobacco | 44.8 | 63.0 | 0 | 809.7 |
| Health products | 129.3 | 352.6 | 0 | 11,179.33 |
| Private consumption | 2,690.8 | 1,706.7 | 269.7 | 25,998.3 |
| Educational purposes | 22.8 | 89.2 | 0 | 3,748.7 |
| Observations | 42,775 | | | |

The table displays the summary statistics for selected variables of the EVS 2013. Information on individual characteristics is only displayed for the household head.
* Including civil unions
** Equal to one if hh head pays church tax
*** Classified as a municipality with a population of 100,000 or more
† Including Berlin
‡ Includes both the German university entrance diploma (Abitur) and the advanced technical college entrance qualification (Fachhochschulreife)

Table 7 displays the most important summary statistics of the EVS 2013 wave. Information on individual characteristics is only displayed for the household head. Around 38 % of all household heads are female and the average age of the household head is around 53 years. About half of the sample households are married with

an average of 0.35 children per household. Around half of the households live in their own residential property. 31 % of the households live in municipalities with a populations of 100,000 or more and about one quarter live in the new federal territories (including Berlin). Most of the household heads are employed (51 %) while 31 % are retired, 9 % are civil servants, 4 % are self-employed and 5 % are unemployed (including housewives and househusbands). The household heads in the sample seem to be rather educated, with 48 % who completed the university entrance diploma (Abitur) or the the advanced technical college entrance qualification (Fachhochschulreife). The income and expenditure values differ greatly within the sample. On average households earned a gross income of €5,642.8 and a net income of €3,461.2, with (standard deviations of €2,967.5 and €2,111.4, respectively). The figures on total household expenditure exceed those of income with an average of €5,642.8. Figures on expenditures on other specific goods can also be seen in Table 7.

## A.2 Distribution of Net Income

Figures 4 and 5 depict the distribution of net household income and the logarithm of net household income. The blue lines represent the normal distribution. The distribution of net household income is skewed and not normally distributed. One can clearly see that by taking the logarithm of net income the distribution approximates a normal distribution to a much grater extent.

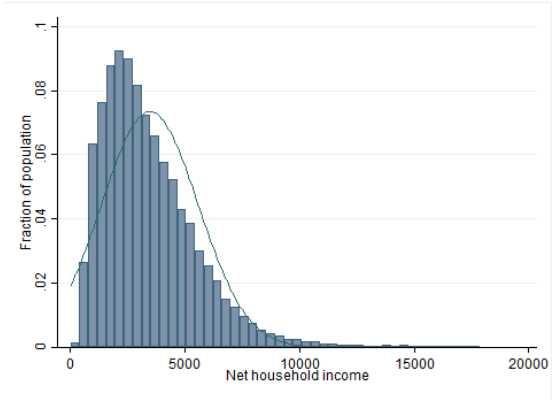



Figure 4: Distribution of net income across the sample

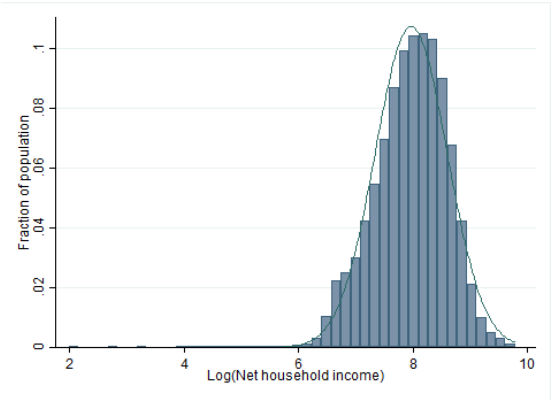



Figure 5: Distribution of logarithmic net income across sample

## A.3 Means of Household (Head) Characteristics for Different Subsamples

Table 8: Means of household (head) characteristics of the full and restricted sample

| | (1) Full Sample | (2) Bottom 99 % | (3) Top 1 % |
|---|---|---|---|
| Hh head is female | 0.38 | 0.38 | 0.25 |
| Age of hh head | 53.32 | 53.26 | 59.92 |
| Hh head is married | 0.52 | 0.52 | 0.63 |
| Children <18 years | 0.35 | 0.35 | 0.15 |
| Net hh income (in €) | 3461 | 3452 | 4339 |
| Hh head completed Abitur | 0.48 | 0.44 | 0.48 |
| Hh head is unemployed | 0.045 | 0.045 | 0.019 |
| HH head is self-employed | 0.038 | 0.028 | 0.038 |
| Hh head is a civil servant | 0.088 | 0.088 | 0.11 |
| HH head is retired | 0.311 | 0.42 | 0.31 |

The table displays the means of household (head) characteristics of observations for the full sample (1), the first 99 percentiles of gambling expenditure (2) and the highest one percent of gambling expenditures (3).

# B Additional Regression Results

## B.1 OLS Regression Results

Table 9: OLS regression results

| | Full Sample | Restricted Sample |
|---|---|---|
| Hh head is female | -2.948*** | -1.924*** |
| | (-8.87) | (-9.10) |
| Age of hh head | 0.595*** | 0.477*** |
| | (9.63) | (12.16) |
| Age of hh head squared | -0.00399*** | -0.00337*** |
| | (-6.58) | (-8.73) |
| Hh head is married | 0.133 | 1.043*** |
| | (0.36) | (4.38) |
| Number of children <18 years | -2.968*** | -2.276*** |
| | (-13.99) | (-16.90) |
| Hh lives in a city | -0.0799 | -0.0742 |
| | (-0.25) | (-0.36) |
| Log net income | 7.084*** | 4.656*** |
| | (21.47) | (22.17) |
| Hh head completed university entrance diploma | -2.336*** | -2.016*** |
| | (-7.69) | (-10.44) |
| Hh head is unemployed | -0.613 | -1.228** |
| | (-0.82) | (-2.59) |
| Hh head is selfemployed | -0.427 | -1.271** |
| | (-0.56) | (-2.64) |
| Hh head is a civil servant | 1.048* | 0.764* |
| | (2.00) | (2.29) |
| Hh head is retired | 1.238* | 1.098** |
| | (2.23) | (3.11) |
| Hh owns house/apartment | -2.678*** | -1.039*** |
| | (-7.54) | (-4.60) |
| Constant | -59.98*** | -40.09*** |
| | (-21.24) | (-22.31) |
| Observations | 42,775 | 42,351 |

$t$ statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The table displays the results of an OLS regression on the full sample (1) and on the sample restricted to the first 99 percentiles of average average monthly gambling expenditure (2). The dependent variable is average monthly gambling expenditure.

## B.2 Tobit Model with Continuous Endogenous Covariates

Table 10: Regression results of a Tobit model with instrumental variables

| Dep. variable: Av. monthly hh gambling expenditure | |
|---|---|
| | Tobit IV |
| Monthly net hh income | 0.00172*** |
| | (7.14) |
| Hh head is female | -4.994*** |
| | (-10.20) |
| Age of hh head | 1.621*** |
| | (17.63) |
| Age of hh head squared | -0.0119*** |
| | (-13.30) |
| Hh head is married | 4.020*** |
| | (7.01) |
| Children <18 years | -4.192*** |
| | (-13.57) |
| Hh lives in a city | -1.199** |
| | (-2.61) |
| Hh head completed university entrance diploma | -5.153*** |
| | (-11.03) |
| Hh head is unemployed | -12.68*** |
| | (-11.29) |
| Hh head is self-employed | -6.218*** |
| | (-5.45) |
| Hh head is a civil servant | 3.025*** |
| | (4.05) |
| Hh head is retired | -0.488 |
| | (-0.62) |
| Hh owns house/apartment | -0.100 |
| | (-0.18) |
| Constant | -60.53*** |
| | (-25.94) |
| $\alpha$ | 0.000377 |
| | (1.32) |
| Observations | 42,351 |

$t$ statistics in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
The table displays the results of a Tobit regression with continuous endogenous regressors on the restricted sample of the bottom 99 percentiles of the distribution of average monthly household gambling expenditures. The dependent variable is average monthly household gambling expenditures. Income tax is used as an instrument for net income.

# C Marginal Effects on Average Monthly Household Gambling Expenditure

## C.1 Marginal Effects of the Tobit Regression at Different Values of Age

Table 11: Marginal effects of the Tobit regression at different values of age

| | Age of the household head | | | | | |
|---|---|---|---|---|---|---|
| | 20 | 30 | 40 | 50 | 60 | 70 |
| Hh head is female | -1.258 | -2.012 | -2.804 | -3.482 | -3.953 | -4.220 |
| Age of hh head | 0.455 | 0.728 | 1.015 | 1.260 | 1.431 | 1.527 |
| Age of hh head squared | -0.00337 | -0.00539 | -0.00751 | -0.00932 | -0.0106 | -0.0113 |
| Hh head is married | 0.509 | 0.814 | 1.135 | 1.409 | 1.599 | 1.707 |
| Children <18 years | -1.287 | -2.058 | -2.868 | -3.560 | -4.042 | -4.315 |
| Log(net income) | 3.292 | 5.265 | 7.337 | 9.108 | 10.34 | 11.04 |
| Hh head completed the Abitur | -1.729 | -2.765 | -3.853 | -4.783 | -5.431 | -5.798 |
| Hh head is unemployed | -2.021 | -3.232 | -4.504 | -5.591 | -6.348 | -6.777 |
| Hh head is self-employed | -1.423 | -2.275 | -3.170 | -3.936 | -4.469 | -4.771 |
| Hh head is a civil servant | 0.641 | 1.025 | 1.428 | 1.773 | 2.013 | 2.149 |
| HH head is retired | 0.501 | 0.801 | 1.116 | 1.385 | 1.573 | 1.679 |

The table displays the marginal effects of the regressors $x$ on the observed outcome conditional on $x$, $(y|x)$ at different values of the age of the household head. The marginal effects are calculated from Tobit regression (2) in Table 3. The marginal effects are given for a discrete change of the dummy variable from 0 to 1 (incl. number of children). The discrete variables are evaluated at their means 3,091.14 ($age^2$) and 7.97 (log income). All marginal effects are significant (at least) at the 5 % level.

## C.2 Marginal Effects of the Tobit Regression at Different Values of Household Net Income

Table 12: Marginal effects of the Tobit regression at different values of household income

| | Average net household net income (in €) | | | | |
|---|---|---|---|---|---|
| | 1000 | 2000 | 3000 | 4000 | 5000 |
| Hh head is female | -1.458 | -1.840 | -2.073 | -2.240 | -2.370 |
| Age of hh head | 0.528 | 0.666 | 0.750 | 0.811 | 0.858 |
| Age of hh head squared | -0.00390 | -0.00492 | -0.00555 | -0.00600 | -0.00634 |
| Hh head is married | 0.590 | 0.744 | 0.839 | 0.906 | 0.959 |
| Children <18 years | -1.491 | -1.881 | -2.120 | -2.290 | -2.423 |
| Hh lives in a city | -0.410 | -0.517 | -0.583 | -0.629 | -0.666 |
| Log(net income) | 3.814 | 4.813 | 5.423 | 5.860 | 6.199 |
| Hh head completed the Abitur | -2.003 | -2.527 | -2.848 | -3.077 | -3.255 |
| Hh head is unemployed | -2.341 | -2.954 | -3.329 | -3.597 | -3.805 |
| Hh head is self-employed | -1.648 | -2.080 | -2.343 | -2.532 | -2.679 |
| Hh head is a civil servant | 0.742 | 0.937 | 1.055 | 1.141 | 1.207 |
| Hh head is retired | 0.580 | 0.732 | 0.825 | 0.891 | 0.943 |

The table displays the marginal effects of the regressors $x$ on the observed outcome conditional on $x$, $(y|x)$ at different values of net household income. The marginal effects are calculated from Tobit regression (2) in Table 3. The marginal effects are given for a discrete change of the dummy variable from 0 to 1 (incl. number of children). Ahe of the household head is evaluated at its mean 53.33. All marginal effects are significant (at least) at the 5 % level.

## C.3 Marginal Effects of the Tobit Regression by Gender of the Household Head

Table 13: Marginal effects of the Tobit regression by gender of the household heads

| | Gender of the household head | |
|---|---|---|
| | Male | Female |
| Gender of hh head | -2.048 | -1.829 |
| Age of hh head | 0.741 | 0.662 |
| Age of hh head squared | -0.00548 | -0.00486 |
| Hh head is married | 0.829 | 0.740 |
| Children <18 years | -2.094 | -1.870 |
| Hh lives in a city | -0.576 | -0.514 |
| Log(net income) | 5.358 | 4.784 |
| Hh head completed the Abitur | -2.814 | -2.512 |
| HH head is unemployed | -3.289 | -2.936 |
| Hh head is self-employed | -2.315 | -2.067 |
| Hh head is a civil servant | 1.043 | 0.931 |
| HH head is retired | 0.815 | 0.727 |

The table displays the marginal effects of the regressors $x$ on the observed outcome conditional on $x$, $(y|x)$ for whether the household head is male of female. The marginal effects are calculated from Tobit regression (2) in Table 3. The marginal effects are given for a discrete change of the dummy variable from 0 to 1 (incl. number of children). The discrete variables are evaluated at their means 53.33 (age), 3,091.14 (age$^2$) and 7.97 (log income). All marginal effects are significant (at least) at the 5 % level.