EC331: RESEARCH IN APPLIED ECONOMICS

The differential impacts of socioeconomic characteristics on skills-based and pure-chance gambling

Department of Economics

Dissertation Tutor: Elizabeth Jones

1004749

Word Count: 4,973

Abstract

This paper determines the socioeconomic drivers of gambling activity in the UK and contributes to existing literature by investigating the differential impacts of these drivers on two distinct and mutually exclusive categories of gambling; skills-based and pure-chance. The study finds that characteristics such as age, gender, income, ethnicity, parental gambling and smoking status have significant impacts on gambling expenditure. Furthermore, the direction and size of these impacts differs significantly for the skills-based and pure-chance categories, suggesting the need for more detailed and targeted policies.

Table of Contents

1. Introduction

Over the past few decades, the British gambling industry has evolved to offer a diverse range of gambling activities including casino games, sports-betting, bingo and gaming machines. The British Gambling Commission, an institution which regulates operators within the industry, found that of the adult population (over 16s) surveyed in the year to March 2012, approximately 58% had participated in some form of gambling in the four weeks preceding the survey. Since the full legalisation of gambling in 2007, forms such as government lotteries have been promoted in order to support public initiatives like healthcare and education. Whilst the extent and stringency of government regulation differs significantly across activities, a rise in gambling participation is almost always accompanied by added risk of addiction and social costs; in the year to March 2012 alone, more than £40bn was spent on high-speed, high-stakes gambling machines. This trend is likely to intensify further as the mobile gambling sector continues to flourish. In 2012, 43% of gamblers were either current or potential future users of smartphones for gaming/betting online. As several UK gambling laws are still in their infancy, authorities rely on existing research in order to implement legislation which can protect vulnerable segments of the population and prevent "problem-gambling" behaviour.

While economists have explored a number of rationales for gambling (e.g. non-concave utility functions and differing risk profiles), other academics have examined the individual characteristic traits displayed by gamblers (e.g. age, gender, education and alcohol/tobacco consumption). This paper combines both lines of enquiry and considers the impact of various socioeconomic factors on gambling tendencies. The results are evaluated in the context of economic theory and used to establish policy recommendations for the gambling sector. As the industry houses a spectrum of gambling activities which differ in wager sizes and gambling environments however, it is intuitive that the impact of socioeconomic characteristics would vary with the type of activity under consideration. This study contributes to the existing literature by accounting for these differences and examining the differing impacts of socioeconomic characteristics on two broad categories of gambling; "skills-based" and "pure-chance".[1]

The paper is structured as follows: Section 2 discusses the existing theory and empirical literature on this topic, Section 3 outlines the data and methodology, Section 4 includes the results and interpretation of the econometric analysis and finally, Section 5 summarises the findings and outlines policy recommendations.

[1] "Skills-based" and "Pure-chance" activities are formally defined in the Data and Methodology section.

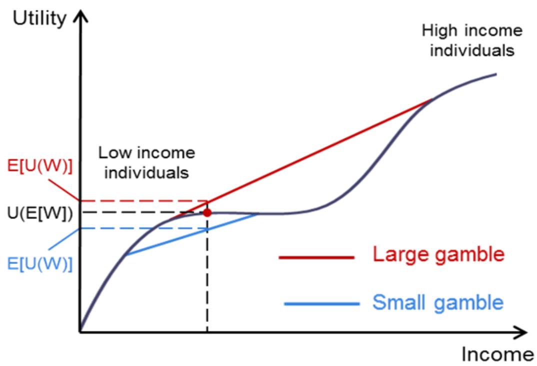
2. Literature Review

2.1 Theoretical Frameworks

Micro-economists have long puzzled over the rationale underlying decisions made by gamblers to partake in negative expected utility gambles. A wealth of literature employs concave utility functions and traditional utility maximisation theory to analyse decision-making under risk (Ingersoll, 1987; Mette et al., 2004). Several forms of gambling such as national lotteries however do not support risk aversion literature as, in theory, a risk averse individual would not engage in an actuarially fair gamble, let alone lotteries which yield negative expected returns.

Friedman and Savage (1948) adapt the utility model to explain simultaneous risk seeking and risk averse behaviour by positing that the utility function should incorporate two concave segments reflecting different wealth levels (low and high) and a convex segment demarking a transition between the two.

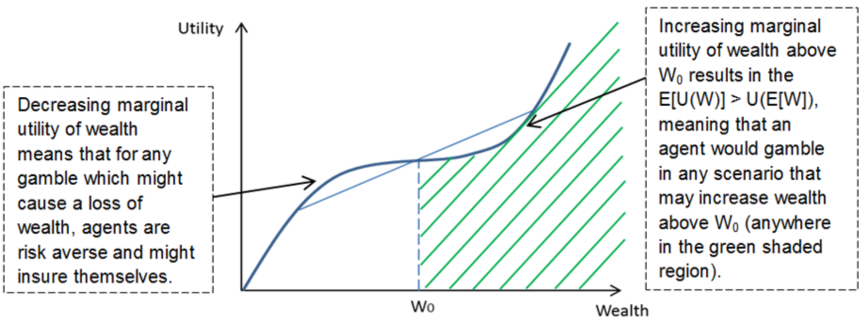
Figure 1: Friedman and Savage (1948) Modified Utility Theory



Under this framework, individuals with a low level of wealth are more likely to partake in a gamble offering a low probability of winning a large amount (W_0+L) and a high probability of losing a small amount (W_0-L), as this could increase their economic status. These individuals experience a higher expected utility of wealth than the certain level of wealth for the gamble: E[U(W)] > U(E[W]). Meanwhile, wealthy individuals would be more intent on avoiding risky prospects.

Bhattacharyya and Garrett (2008) observe Friedman's (1948) proposition and note further that Quiggin (1991) proves the invalidity of the third, concave segment. The utility function would therefore consist of just a concave segment followed by a convex one:

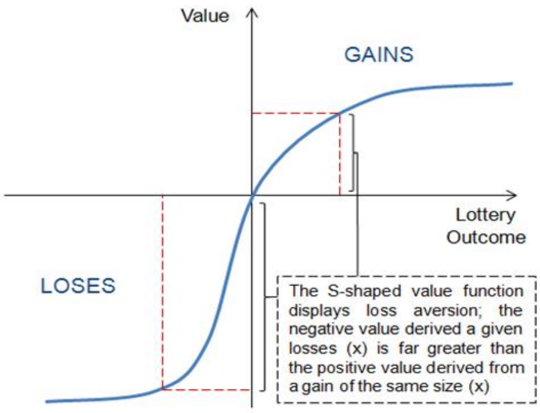
Figure 2: Bhattacharya (2003) – Alternative Utility Function of Wealth



Bhattacharyya and Garrett (2008) test their proposition empirically and prove the hypothesis that "expected return from lottery tickets is a declining and convex function of the skewness of prize distributions". In simple terms, Bhattacharyya's hypothesis implies that as the prize amount for a gamble increases, the player is increasingly happy to accept a smaller chance of winning.

The Prospect Theory framework proposed by Khaneman and Tversky (1978) uses a similar theoretical underpinning to model gambling behaviour with risk aversion.

Figure 3: Khaneman and Tversky (2003) Prospect Theory



The value function shown in the diagram is concave for gains and convex for losses, thereby displaying loss aversion. Agents assess possible outcomes with weights which are based on probabilistic beliefs, but do not represent the actual probabilities. Individuals often misperceive actual probabilities associated with outcomes, thereby overweighting a low probability of winning and underweighting high probabilities of losing.

The outcomes are also assessed in relative terms, rather than in absolute terms. Prospect theory posits that a combination of these misconceptions by agents lead to systematic errors in rationality, thereby explaining decisions to partake in gambles with negative expected returns.

2.2 Empirical Literature

Brenner and Brenner (1990) use Friedman and Savage's theoretical underpinning in their empirical book "Speculation and Gambling: A theory, a history and a future of some human decisions". They focus on lotteries and suggest that certain ethnicities are motivated to a greater extent by the urge to become wealthier as a gamble may be the only opportunity to significantly increase their wealth. This explains why the Productivity Commission Report (1993) and Raylu and Oei (2002) observe high rates of gambling among certain ethnicities such as Black-Americans and Chinese. The assumption that agents play only for wealth however seems premature given that several gamblers are addicts and hence are unlikely to be able to consider the expected change in their wealth before engaging in risk-based activities. A limitation of both these studies is that they ignore a major determinant of gambling; the risk attitudes of individuals.

A psychology review by J.Lloyd et al. (2011) concludes that increasing age is associated with gambling for enjoyment as opposed to monetary gain. Furby and Beyth-Marom (1992) attempt to explain this phenomenon by suggesting that young adults are more attracted to risky activities due to deficiencies in their cognitive skills (which determine awareness, problem-solving and analytical abilities). This notion is supported in a study by Spear (2000), which observes that young adults may experience a significant increase in reward salience; a phenomenon where the individual pays more attention to the magnitude of potential rewards. Whilst research by Nyman et al. (2008) and Sprosten et al. (2000) either fails to take into account the effect of age on gambling or concludes that it is insignificant, Winters et al. (2002) find that as people age, their preference changes from skills-based games (e.g. poker and betting) to pure-chance games (e.g. bingo and lotteries). This supports the conclusion by Lloyd et. al (2011), indicating that the agent's motivation to gamble changes over his/her lifetime. An equally likely and more intuitive explanation is that an average person becomes busier during his/her middle ages (due to family commitments) and hence engages less in time consuming activities. Unfortunately there is a lack of literature which explores this topic in greater depth.

Interesting extensions to gambling literature explore the interactive effects of age and gender on gambling behaviour. Several studies have concluded that women, unlike men, in their middle ages face increased odds of problem-gambling. As indicated by Afifi et-al. (2010), these findings are consistent with a well-acknowledged phenomenon known as the telescoping effect; this states that women tend to engage in gambling later on in life but are likely to progress faster towards problem-gambling than men (Grant and Kim, 2002; Ladd and Petry, 2002; Ibanez et al., 2003; Tavares et al., 2003). Furthermore, studies such as Berry et al (2002), Crisp et al. (2004) and Falenchuk (2007) observe that women often resort to gambling in order to overcome depression or escape from their problems.

Substances such as alcohol and cigarettes are also commonly consumed by individuals for stress relief purposes. Research by Grant et. al. (2003) indicates that problem gambling is more imminent in individuals with an alcohol addiction. Similarly, Rodda et al. (2004) discover a robust linear relationship between electronic machine gaming and a high smoking rate. Whilst their sample size of 81 gamers could be deemed insufficient for drawing representative conclusions, other studies such as McGrath and Barrett (2009) have also found strong links between the two.

The aforementioned results confirm that decisions to gamble are not driven solely by monetary considerations but are also the product of several socioeconomic factors such as ethnicity, age and education. Additionally, the gambling environment itself can have a range of sensory and physical influences on the decisions made by individuals. It is likely therefore that internet gamblers may undergo very different gambling experiences to those who gamble in person.

Cotte and Latour (2009) investigate the characteristics observed in Internet and Non-internet gamblers and find that internet gamblers tend towards a more competitive and less social engagement. As indicated by Gainsbury et. al. (2012) however, Cotte and Latour's findings are based on a sample of 30 gamblers from Las Vegas and are unlikely to be representative of a wider set of internet gamblers. Wood and Williams (2011) provide a comprehensive investigation based on data from gamblers in 105 countries and conclude that characteristics which increase the likelihood of internet gambling include single status, male gender, a younger age and alcohol/tobacco use. Gainsbury et al. (2012) shed further light on the differences by deducing that internet gamblers are more versatile in terms of the activities they engage in; they participate in 4.96 activities on average as opposed to the mean of 3.27 for non-internet gamblers. Furthermore, the paper also suggests that internet gamblers possess a more positive outlook on gambling as a whole.

3. Data and Methodology

3.1 Research Strategy

This paper investigates the differing impacts of socioeconomic factors on skills based and pure chance gambling expenditures by building a separate empirical model for each category and comparing the results. "Skills-based" activities (SBAs) are defined as forms of gambling where the participant is able to influence (at least partly) the outcome of the gamble after the initial bet is placed. "Pure-chance" activities (PCAs), on the other hand, are where the participant has no influence over the outcome of the gamble after the initial bet is placed. A list of the activities included within each category can be found in Appendix I.

3.2 Data and Preliminary Analysis

This study uses data from the British Gambling Prevalence Survey (2010), a large-scale nationally representative survey of gambling in Great Britain. The survey contains observations from 7,756 individuals and is the first to be conducted after the full implementation of the Gambling Act 2005. A plethora of previously unrecorded variables pertaining to participation in all forms of gambling, the prevalence of problem gambling and attitudes to gambling are therefore available for analysis in this cross-sectional study.

The dataset contains a large number of zero expenditure observations, meaning that the issue of potential bias needs to be considered. A cluster of zero expenditure observations could result in an OLS estimation yielding biased coefficients (Appendix II)[2]. As this study investigates drivers of consumption decisions (i.e. factors affecting how much individuals spend on gambling) rather than participation decisions, selection bias is irrelevant and therefore circumvented by conducting a conditional OLS analysis which assumes expenditure>0[3]. This is achieved by limiting the dataset to a definitive core of 3,986 pure-chance and 636 skills-based observations.

The expenditure variables are derived from existing monthly expenditure figures for individual gambling activities:

Pure-chance monthly spend = ∑ (National lottery, Other lotteries, Scratch cards, Bingo & Online Bingo)

Skills-based monthly spend= ∑ (Poker, Virtual Gaming & Betting – Football, Horses, Dogs, Spread and Other)

A brief summary of the resulting expenditure distributions confirms that participation and average monthly spend is significantly higher and more variable in the case of skills-based gambling; a potential indication that the risk profiles and decision making processes of individuals differ across SBAs and PCAs

Variable	**Obvs.**	**Mean**	**Std. Dev.**	**Min**	**Max**
Pure-chance monthly spend (£)	3986	17.13	26.79	5.5	541.5
Skills-based monthly spend (£)	636	64.28	174.81	5	2328

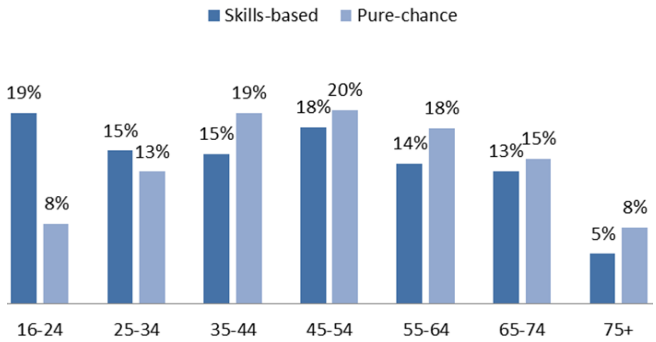
A preliminary analysis of the composition of the dataset (Appendix IV) shows that of the PCA participants, the majority are of white ethnicity (95%), married (65%), engage in full-time employment (65%) and gamble in-person only at venues (79%). The educational backgrounds of individuals span the entire spectrum, including those with Degrees (24%), A-levels (11%), GCSEs (27%) and no qualifications (30%). Similarly, personal and household income levels are distributed evenly across five quintiles. Furthermore, the sample consists of approximately equal numbers of male (48%) and female (52%) participants.

[2] This is sometimes accounted for by adopting a "Tobit" approach. The Tobit approach is not a good alternative in this case as the choice of small positive numbers to replace the zero values would be arbitrary.
[3] This means that inferences based on the results of this study can be only be made for the sub-sample of the population who display positive expenditure (i.e. those who gamble). A Heckman-Two Stage Model may be used to examine the participation and consumption decisions separately, but this is beyond the scope of this paper.

An analysis of the skills-based gambling participants shows a majority to be males (78%), of white ethnicity (95%) and engaging in-person at casinos/betting venues (69%). As with the pure-chance participants, the skills-based participants are also evenly distributed across each of the education and income categories.

Additionally, the age groups of participants appear broadly similar, with the majority of observations evenly distributed around the 45-54 age band. The main difference appears to be in the 16-24 category which contains a large number of observations for skills-based.

Figure 4: Distribution of individuals by age bands



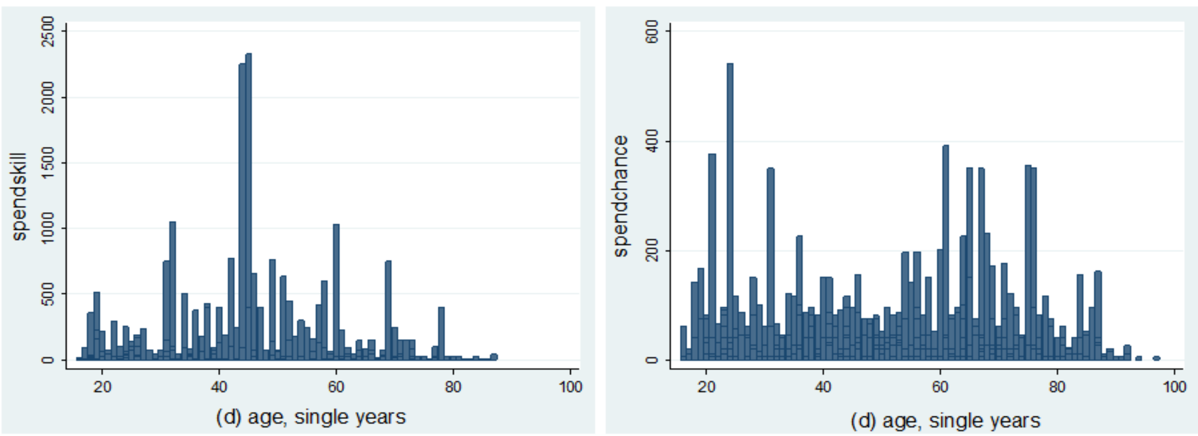
A detailed univariate analysis of the data is also conducted in order to identify any underlying expenditure trends and determine the structure of the empirical models:

Gender: On average, males spend more than females on both types of gambling. More importantly, the difference is significantly larger for skills-based activities, suggesting that males may be more prone to skills-based gambling expenditure:

		Chance (£)	Skill (£)	Total (£)
Gender	Male	17.6	75.5	93.1
	Female	16.7	24.5	41.2

Age: The expenditure on pure-chance activities follows a quadratic (U-shape) relationship with respect to age. The skills-based expenditure however, peaks in the middle-ages, with a maximum in the 43-45 age band. This is most likely due to the higher minimum bets/participation costs in activities such as horse racing, dog racing and poker.

Figure 5: Graphs of gambling expenditure by age



Income: An increase in income results in decreased pure-chance gambling. This may suggest that unlike wealthy individuals, lower income households purchase lotteries and scratch-cards in order to boost their quality of life. This seems especially likely given the large drop in "pure-chance" expenditure after the 4th quintile (as opposed to a gradual decline). Skills-based expenditure however, increases gradually with income until the highest (5th) quintile, where a significant increase is observed. Whilst the gradual increase is most likely to be symptomatic of an increased disposable income, the doubling of average expenditure in the 5th quintile suggests that there may be a possible wealth/status effect which results in individuals spending greater amounts on "prestigious" sports such as horse racing and dog racing.

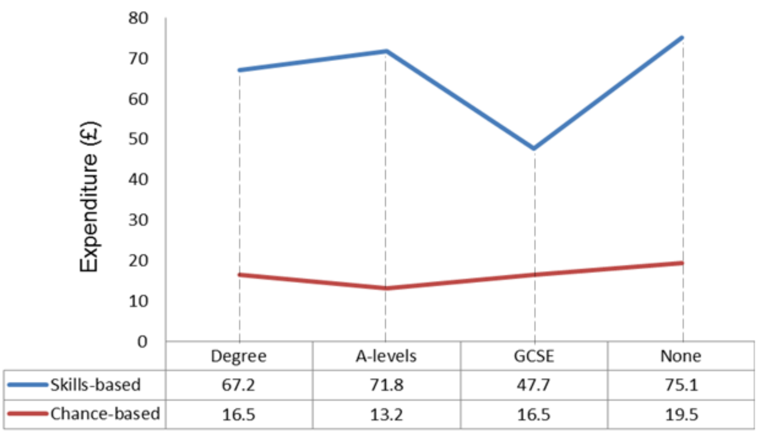
		Chance (£)	Skill (£)	Total (£)
Annual Personal Income Quintiles	1st (lowest)	17.4	35.1	52.5
	2nd	17.8	41.1	58.9
	3rd	17.3	46.0	63.3
	4th	15.2	64.5	79.7
	5th (highest)	15.0	127.6	142.6

Smoking: On average, smokers appear to spend more on both pure-chance and skills-based activities, but the difference is much larger in the case of skills-based activities.

		Chance (£)	Skill (£)	Total (£)
Smoking	Smoke	20.2	84.8	105.0
	Do not smoke	16.0	52.8	68.8

Control variables & Other variables of interest: Socioeconomic factors such as marital status, alcohol consumption and parental gambling show very slight or no trends with respect to both types of gambling expenditure. Surprisingly, the level of education also displays no clear trend. This seems particularly unusual for SBAs, where one would expect more educated individuals to be endowed with numerical and analytical abilities which might attract them towards SBAs more so than their less qualified counterparts.

Figure 6: Average expenditure levels for different educational backgrounds



A plausible explanation for this is that "skills-based" is too broad a definition; it encompasses activities such as horse and dog betting, which do not require the same degree of numerical and analytical skills. An analysis of poker and spread-betting expenditure is therefore carried out to ascertain any potential relationship between education and gambling. However, this also displays no clear trend.

3.3 Econometric Model Specification and Diagnostics

The preceding univariate analysis and a review of models presented in existing research (particularly Nyman et. al, 2008) serve as guidelines for the structure of the empirical model:

$$\begin{aligned} ln(expenditure) = \; & \alpha + \beta_1 Age + \beta_2 Gender + \beta_3 Education_i + \beta_4 P.Income_i \\ & + \beta_5 H.Income + \beta_6 Marital\ Status + \beta_7 ParentalGambling_i \\ & + \beta_8 Smoke + \beta_9 Online_i + \beta_{10} Control\ variables_i + \varepsilon_i \end{aligned}$$

This model is applied to both skills-based and pure-chance observations in order to estimate two separate regressions. Logarithmic transformations of the respective expenditures are used as dependent variable in both models to account for skewness and ensure a more linear relationship between variables.

The empirical validity of the estimations is determined by scrutinising the models for issues of omitted relevant variable bias, non-normality, multi-collinearity, heteroscedasticity and endogeneity:

	Test(s)	Model implication(s) and Mitigant(s)	Appendix Ref.
Omitted Relevant Variables	**Ramsey RESET Test (low power):** H0: Model has no omitted relevant variables *Result:* Omitted Variables	***Implications*** – Relevant omitted variables such as risk attitudes, proximity to gambling venues, probabilistic beliefs and household expenditure potentially bias the coefficients of related variables in the regression as follows: $\underbrace{\widehat{\beta_1}}_{\text{Coeff. estimate of related variable}} = \beta_1 + \underbrace{\beta_2 \frac{\text{cov}(x_i, z_i)}{\text{var}(x_i)}}_{\text{Omitted relevant variable bias}}$ – A biased coefficient alters standard errors, t-ratios and hypothesis test outcomes, rendering conclusions invalid ***Mitigants*** – DSM and Gambling Attitude scores are added as proxies for risk attitude and probabilistic beliefs. Additionally, the log of the number of people in the household is added as a proxy for household expenditure. These variables capture the bias and refine the coefficients on variables of interest – Both models have also been estimated with different forms and variables to ensure that coefficients remain broadly similar	Appendix V

Non-normality	**Residuals Plots:** *Result:* Non-normal residuals	***Implications*** – Non-normally distributed population errors lead to the invalidity of the *t* and *F* distributions ***Mitigants*** – Basu, Chen and Oh (2011), amongst other researchers, confirm that assumptions of asymptotic normality can be used to circumvent potential problems arising from non-normality	Appendix VI
Heteroscedasticity	**Breusch-Pagan/ Cook-Weisberg Test** H0: Error term variance is heteroscedastic *Result:* Heteroscedastic error variance	***Implications*** – OLS estimates no longer remain Best Linear Unbiased Estimators (BLUE) as OLS no longer minimises variance – Standard errors can also be biased in the presence of heteroscedasticity, leading to biased test statistics ***Mitigants*** – Various functional forms are investigated for both models, incorporating transformations of numerical variables – White's Heteroscedasticity consistent robust standard errors are used to relax the assumptions that errors are independent and identically distributed. This yields more reliable test statistics and p-values	Appendix VII
Endogeneity & Reverse Causality	**Instrumental Variable (IV) Estimation** *Result:* Weak potential endogeneity and reverse causality	***Implications*** – Positive endogeneity bias stemming from the risk profile/tolerance of an individual potentially affects the variable for smoking status. This may result in the inconsistency of OLS estimates: $\ln(expenditure) = \alpha_1 + \beta_1 Smoking + \beta_i Other\ Variables + \varepsilon_1$ $Smoking = \alpha_1 + \beta_2 Risk\ affinity + \beta_i Other\ Variables + \varepsilon_2$ $cov(\varepsilon_1, \varepsilon_2) > 0 \Rightarrow cov(expenditure, \varepsilon_2) \neq 0 \therefore$ OLS is inconsistent – The relationship between gambling expenditure and smoking status also potentially suffers from reverse causality: it is plausible that a rise in gambling expenditure and debt may incentivise an individual to smoke in order to reduce stress ***Mitigants*** – IV estimation is conducted using the Two Stage Least Squares (2SLS) method, but a test of the only available instrument (Attitude score) shows it to be a weak instrument. As IV regressions with weak instruments can cause further inconsistency and larger standard errors, OLS estimation is preferred to 2SLS. The attitude score is added in the OLS regression as a proxy to reduce bias – As reverse causality does not affect coefficient estimates or standard errors, this issue has been set aside as a topic for further research. A panel dataset created with future BGPS surveys can investigate causality by taking lags of the respective variables	n/a
Multi-collinearity	**Variance Inflation Factor (VIF) Test & Correlation Matrices** *Result:* No multi-collinearity	***Implications*** – Whilst multi-collinearity does not violate any OLS assumptions, it increases standard errors, meaning that coefficients need to be larger to be significant ***Mitigants*** – Variables that are highly correlated with key variables of interest are omitted from the models (e.g. the omitted interactive variables in the regression)	Appendix VIII

4. Results and Analysis

The final results in the table below confirm the differential impacts of several socioeconomic characteristics such as age, gender, income, ethnicity, smoking status and online gambling.

		Skills-based			Pure-chance		
		Coefficient	Robust Std. Error	Significance	Coefficient	Robust Std. Error	Significance
Age	Age	0.037	(.0159)	**	0.003	(.0010)	***
	Age^2	-0.0004	(.0002)	**	n/a		
Gender (default=Female)	Male	0.476	(.1107)	***	-0.005	(.0271)	
Education (default= no qualifications)	Degree	-0.226	(.1471)		-0.171	(.0406)	***
	Professional Qualification	-0.219	(.2069)		-0.110	(.0519)	**
	A-Levels	-0.377	(.1605)	**	-0.205	(.0466)	***
	GCSE	-0.416	(.1256)	***	-0.058	(.0377)	
Personal Income (default=1st quintile)	5th Quintile	0.437	(.1760)	**	0.077	(.0446)	*
	4th Quintile	0.331	(.1774)	*	0.017	(.0431)	
	3rd Quintile	0.197	(.1346)		0.036	(.0389)	
	2nd Quintile	0.209	(.1577)		0.030	(.0419)	
Household Income	Household income band	0.012	(.0040)	***	0.0003	(.0011)	
Household expenditure proxy	ln(no. of people in household)	-0.140	(.1133)		0.010	(.0335)	
Marital Status (default= Single, Divorced, Widowed)	Married	-0.202	(.1169)	*	0.032	(.0326)	
Ethnicity (default=white)	Black/Black British	-0.777	(.3447)	**	0.320	(.0888)	***
	Asian/Asian British	-0.235	(.4138)		0.052	(.0858)	
	Chinese	-0.359	(.6819)		-0.216	(.2061)	
	Mixed	0.209	(.5496)		-0.017	(.1502)	
Parental Gambling (default=parents did not gamble)	Regular	0.098	(.0984)		0.139	(.0314)	***
	Problem Gamblers	-0.691	(.2146)	***	0.065	(.0644)	
Smoking Status (default=does not smoke)	Smoke	0.207	(.1071)	*	0.101	(.0305)	***
Method of Gambling (default=in-person only)	Online only	0.672	(.1976)	***	-0.067	(.0764)	
	Both	0.420	(.1194)	***	0.236	(.0357)	***
DSM	DSM score	0.133	(.0166)	***	0.076	(.0109)	***
Risk profile proxy	Attitude score	0.040	(.0110)	***	0.020	(.0031)	***
Region (default=england and wales)	Scotland	n/a	n/a		0.196	(.0447)	***

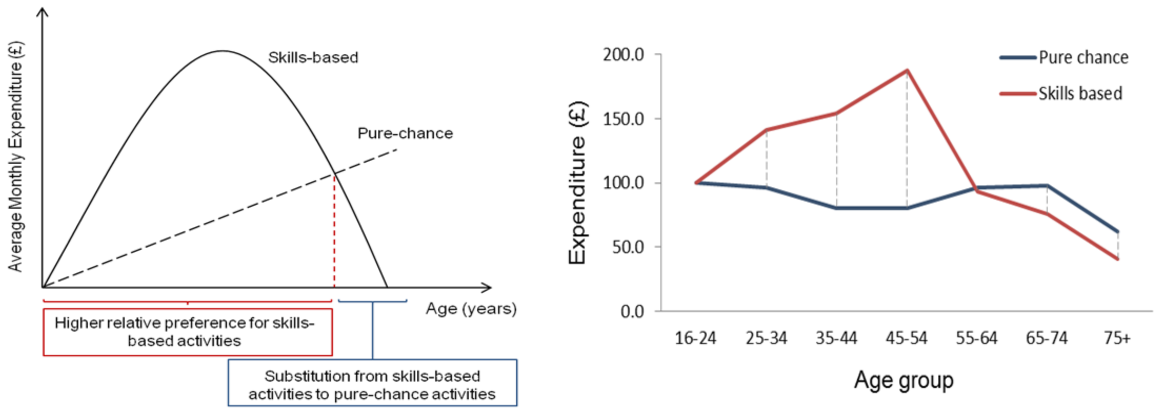
*Statistical significance of individual coefficients is recorded at the *10%, **5% or ***1% level*

	Skills-based		Pure-chance	
Constant	-0.113	(.5280)	1.583	(.1323)
No. of observations	636		3980	
R-squared	0.278		0.095	
F-stat	0.000		12.80	
Probability>F	0.000		0.000	

4.1 Age

The results display a quadratic (inverted-U shape) relationship between age and skills-based expenditure, implying that a 1-year increase in the age of an individual leads to a 3.7% increase in monthly skills-based expenditure, until the age of 46[4], after which the monthly expenditure falls with age. Contrastingly, the pure-chance regression displays a positive linear association; a 1-year increase in age is associated with a 3% increase in monthly pure-chance expenditure. A further analysis shows that these trends partially support the conclusion by Winters et al. (2002) and suggest a substitution of skills-based gambling for pure-chance gambling over an individual's lifetime. Data from the 513 individuals who engaged in both SBAs and PCAs (and were therefore in both regressions) is used to generate an index of the average expenditure for a range of age groups:

Figure 7: Theoretical substitution relationship and corresponding average expenditure indices for 513 common observations



The indices demonstrate clearly the theoretical substitution relationship implied by the regression coefficients. Furthermore, they suggest that the results of Winters et. al. (2002) are only partially supported; initially, an increase in age is associated with a relative preference for SBAs and only after a certain point (around 55-64 years in this case) do PCAs replace skills-based activities.

The trend indicated by the data is supported by the findings of Lloyd et. al. (2011); an increasing age is associated with gambling for recreation as opposed to monetary gain. It is unsurprising therefore that until age of 46, people tend to display a relative preference for SBAs like poker, spread betting and horse/dog race betting, as these are far more engaging and provide a more intense experience. Meanwhile, at higher ages, the value of this experience may be offset by the high potential losses in these activities. This is increasingly the case during the transition from middle-age to old-age because the financial impact of a loss is likely to increase. In contrast, the low participation costs and potential loses in PCAs

[4] The age at which the maximum expenditure occurs is calculated as follows: 0.037/(2*0.0004)=46.25

attract the elderly. This is especially true for Bingo, which provides low-cost entertainment for a prolonged period.

4.2 Gender

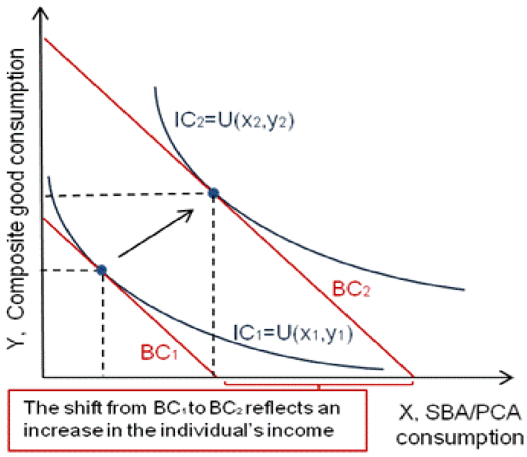
Gender has a significant impact on SBA expenditure, with males spending 47.6% more than females. This finding is consistent with Jianakoplos and Bernasek (1998) and Eckel and Grossman (2008); women display more risk aversion in financial decision making than men. The lower risk propensity of women is therefore likely to deter them from SBAs where the size of potential losses tends to be larger. This is not the case in PCAs, where the cost of participation /losses is significantly lower, explaining why pure-chance expenditure shows no significant difference between males and females.

4.3 Income

An individual within the 4th and 5th (top two) income quintiles spends 43.7% and 33.1% more on SBAs relative to an individual in the bottom quintile. Similarly, the 5th quintile is associated with a 7.7% increase in monthly PCA expenditure. Whilst these observations are inconsistent with the Friedman and Savage (1948) hypothesis, they are in alignment with other economic propositions. Firstly, as income increases, assuming that the costs of living remain constant (approximated in the regression by the number of people in the household), the increase in disposable income could result in additional gambling expenditure.

This implies that both SBAs and PCAs are normal consumption goods, whilst arguably; a lottery represents a special case of inferior goods. The assumption does seem valid however as the PCA category also contains other activities like Bingo, which agents engage in primarily for entertainment and not for monetary gain. Therefore an increase in income would not incentivise the agent to reduce their bingo participation in the same way as it might for lotteries.

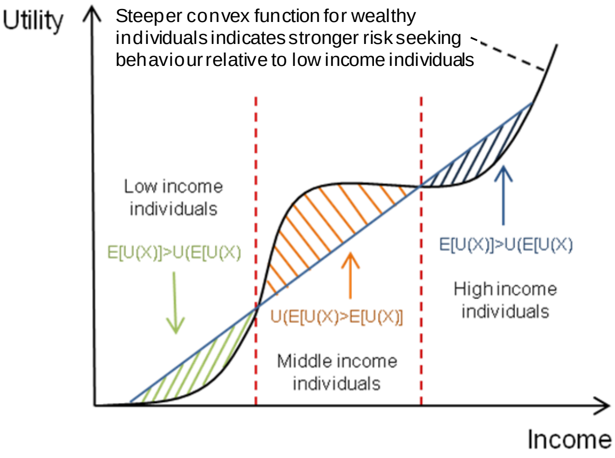
Figure 8: Income effect on gambling expenditure/consumption



An alternative utility theory proposition also provides an explanation for the observed trend with respect to income. Friedman and Savage's (1948) utility curve can be adapted to incorporate a third "middle" income segment in order to better reflect the income groups that exist in today's society.

Whilst the predictions of Friedman and Savage (1948) still hold true for low incomes, the middle and high income categories display new trends.

Figure 9: Alternative Utility Theory:
An adaptation of Friedman and Savage (1948)



The middle income group, who typically earn a comfortable living, are less likely to be attracted to lotteries such as the one described above. Meanwhile, the high potential losses in SBAs also act as a deterrent, resulting in overall risk aversion for middle income groups. As the proportion of potential losses relative to income is significantly lower for high income individuals, they are more likely to engage in risky SBAs and certain PCAs like Bingo for entertainment.

This theory demonstrates clearly the rationale that leads to high overall gambling expenditure in high income groups, but as the middle and lower income variables are insignificant in the estimated model, the validity of corresponding segments cannot be established without further research.

4.4 Education

Individuals with A-levels and GCSEs spend 37.7% and 41.6% less on SBAs respectively than individuals with no qualifications. Meanwhile those with Degrees, Professional qualifications and A-levels spend 17.1%, 11% and 21% less on PCAs than people with no qualifications. Whilst, the differences between the various qualification groups are fairly small, the general trend, contrary to the initial conjecture, indicates that educated individuals are likely to spend less on all gambling activities.

Literature suggests that inaccurate cognitions such as the beliefs that outcomes are not random, that luck is a stable characteristic, that superstitious behaviours are effective in altering outcomes, and that persistent play is the key to recovering losses are the main contributors of gambling expenditure (Ladouceur et al., 2001; Raylu & Oei, 2002; Lambos & Delfabbro, 2007). As these erroneous cognitions are likely to decrease with education (even if it is only up to GCSE level), it seems logical that a rise in education is associated with decreasing overall gambling expenditure.

4.5 Smoking

Smokers on average spend 20.7% and 10.1% more on SBAs and PCAs respectively. These results are in alignment with Rodda et. al (2004) and potentially support the conclusions of Prospect Theory (Khaneman and Tversky, 1979). Evidence suggests that most smokers underestimate the likelihood of contracting smoking related health issues; this attitude to/perception of risk may also cause them to overweight low probabilities of winning and underweight high probabilities of losing in gambling activities. The impact is greater on SBA expenditure (as indicated by the larger coefficient) due to the repetitive decision making involved in most SBAs. A repeated misjudgement of probability and derived utility would result in relatively large expenditure on SBAs.

4.6 Online gambling

As expected, the gambling method/environment (online/in-person) also impacts skills-based and pure-chance expenditure: individuals who participate in online SBAs (whether exclusively or in addition to in-person SBAs) spend significantly more than those who gamble in-person only; 67.2% more for online only and 42% more for both. The same trend applies for PCAs; individuals gambling both online and in-person spend 23.6% more.

Figure 11 highlights the factors which attract people towards online gambling. The greater impact of online gambling on SBA expenditure is justified by McMillen et al. (2005), who observe that several SBA gamblers report that easy and immediate access to cash enables to exceed pre-decided expenditure limits.

Figure 10: Key drivers of online gambling participation

4.6 Family background

Other social background characteristics such as ethnicity, marital status and parental gambling also display significant differential impacts on monthly skills-based and pure-chance spend. Individuals who are married, of black ethnicity and have parents with gambling problems are likely to spend 20.2%, 77.7% and 69.1% respectively less on SBAs. Meanwhile, black ethnicity and regular parental gambling are associated with a 32% and 13.9% respective increase in PCA expenditure. Whilst the significant increase in PCA spend for black ethnicities supports previous literature such as Raylu et. Oei (2002), the only potential explanation for the large decrease in SBA spend is historical context. Activities such as horse and dog racing were historically participated in by predominantly white ethnicities; this trend may have translated into a cultural preference for other gambling activities within black ethnic communities.

5. Conclusion

5.1 Conclusions and Policy Implications

The aim of this paper was to determine the socioeconomic characteristics which affect gambling expenditure and to investigate whether the impact of these differed for skills-based and pure-chance activities.

The preceding analysis of the British Gambling Prevalence Survey (2010) data identifies drivers of gambling expenditure and confirms that the impact of socioeconomic characteristics such as age, gender, income, ethnicity, smoking and online gambling differs significantly for skills-based and pure-chance activities. Although some of the emerging trends are only significant at the 10% level, the study contributes to existing literature by bridging the gap between theoretical and empirical investigations of gambling.

The inferable implications for gambling policy within the UK are twofold; macro (affecting the industry as a whole) and micro (affecting only certain socioeconomic groups). On a macro level, data suggests that the skills-based and pure-chance categories should be considered separately and policy should be implemented based on the traits displayed by individuals in each category. Key policy issues on a micro level appear to be high online expenditures, gambling prevalence amongst substance addicts and an insufficient cognitive awareness amongst individuals without qualifications. McMillen et. al. (2005) find that easy access to cash is the primary cause of high online expenditure. Furthermore, research suggests that a delay of just 5-minutes can reduce the impact of an initial gambling impulse. A method that increases the payment time for online gambling portals may therefore reduce excess expenditure on gambling activities. High gambling spend amongst smokers could also be curtailed by introducing a gambling screening process amongst individuals receiving substance misuse treatment. Finally, the cognitive ability of individuals with no qualifications can be enhanced by introducing gambling awareness sessions into the lower secondary school national curriculum. This would reduce incorrect cognitions such as the belief that outcomes are not random, that superstitious behaviours are effective in altering outcomes, and that persistent play is the key to recovering losses.

5.2 Research Limitations and Extensions

The most obvious limitation of this study is that it is based on cross sectional data, meaning thereby that whilst a relationship can be observed, causality cannot strictly be inferred. Unfortunately as this was the first survey after the full implementation of the British Gambling Act 2005 and variables of interest were not available in previous BGP surveys, a dynamic empirical model could not be built. Subsequent surveys could be used to create a panel dataset and investigate dynamic impacts. Data collected over a period of time would also alleviate the previously identified problems of endogeneity and reverse causality, as lags can be used to establish causes and effects.

Another criticism of BGPS data is that it is ill-equipped for a separate analysis of the two categories: several key factors affecting the expenditure levels are not present in the data. The most significant example of this in the case of pure-chance expenditure is the absence of a variable indicating an individual's probabilistic beliefs. Whilst some individuals believe that if they are to be lucky enough, they could win with just one lottery ticket, others believe in maximising their chances by buying a large number of tickets. The nature of these beliefs is clearly a large determinant of expenditure on activities such as scratch cards and lotteries. Other unrecorded variables such as household expenditure, number of kids and proximity to gambling venues also present sources of potential bias. Despite the proxies used to minimise the bias, the effect of these omitted variables is reflected in the low R-squared value for the pure-chance regression.

The interactive impact of socioeconomic factors on gambling expenditure is another aspect of this investigation that could not be explored in further depth, either because interactive variables were insignificant or displayed multi-collinearity with the variables they were derived from. Given the social developments within UK in the last few decades, especially with regards to gender equality, it may be the case that gender impacts vary for different age groups and educational levels. The shift in social attitudes over time has seen an increase in the qualifications held by most females and a decrease in social stigma surrounding female gambling, alcohol and tobacco consumption. The expenditure gap between males and females could therefore be larger for older age groups. Further studies in this area could potentially investigate the gender impacts within different age groups and substance consumption impacts for different gender groups by incorporating interactive variables into the specified model.

Overall, this paper serves as a building block for in-depth studies of gambling expenditure. Exploring and understanding the observed relationships in greater detail will enable policy makers to better serve the interests of the public.

Bibliography

Afifi et al. (2010), *Psychiatry Research 178, 395–400*

Berry, R.E. et. al. (2002), *Northwestern Ontario Women's Experiences of Gambling and Problem Gambling, University of Manitoba*

Bhattacharya, N., & Garrett, T. A. (2008), *Why people choose negative expected return assets – an empirical examination of a utility theoretic explanation. Applied Economics 40, 27–34*

Bolen, D. W., & Boyd, W. H. (1968), *Gambling and the gambler: A review and preliminary findings. Archives of General Psychiatry,18(5), 617−629*

Brenner, R. & Brenner, G. (1990), *Gambling and Speculation: A Theory, a History, and a Future of some Human Decisions, Cambridge University Press*

Cotte, J., & Latour, K.(2009), *Blackjack in the kitchen: Understanding online versus casino gambling, Journal of Consumer Research, 35, 742–758*

Crisp, B.R et al. (2004), *Not the same: a comparison of female and male clients seeking treatment from problem gambling counselling services, Journal of Gambling Studies, 20, 283–299*

D.Kahneman, and A.Tversky (1979), *Prospect Theory: An Analysis of Decision under Risk, Econometrica, XLVII, pages 263-291*

Eckel, C.C. & Grossman, P.J. (2008), *Men, women, and risk aversion: experimental evidence. In: Plott, C., Smith, V. (Eds.), Handbook of Experimental Economics Results. Elsevier, New York, 1061–1073*

Falenchuk, O. (2007), *Vulnerability and co-morbidity factors of female problem gambling, Journal of Gambling Studies 23, 323–334*

Furby, L., & Beyth-Marom R. (1992), *Risk taking in adolescence: A decision-making perspective, Developmental Review, 12(1), 1–44*

Grant, J.E., Kim, S.W. (2002), *Gender differences in pathological gamblers seeking medication treatment, Comprehensive Psychiatry 43, 56–62*

Ibanez, A. et. al. (2003), *Gender differences in pathological gambling, Journal of Clinical Psychiatry 64, 295–301*

Ingersoll, J. E. (1987), *Theory of Financial Decision Making, Rowman and Littlefield, Totowa, New Jersey*

J.A. Nyman et al. ***(2008)****, The Journal of Socio-Economics 3, 2492–2504*

Jianakoplos, N.A. & Bernasek, A. (1998), *Are women more risk averse?, Economic Inquiry 36, Page 620–630*

J. Lloyd ***et al. (2010)****, How Psychological Symptoms Relate to Different Motivations for Gambling: An Online Study of Internet Gamblers, Bio Psychiatry; 68:733–740*

Ladouceur, R. et al (2001), *Cognitive treatment of pathological gambling. Journal of Nervous and Mental Disease, **189**, 774–780*

Ladd, G.T., Petry, N.M. (2002), *Gender differences among pathological gamblers seeking treatment. Experimental and Clinical Psychopharmacology 10, 302–309*

Lambos, C. & Delfabbro, P.H. (2007), *Numerical reasoning ability and irrational beliefs in problem gambling, International Gambling Studies, 7, 157-172.*

***M. Friedman and L. Savage (1948)**, "The utility analysis of choices involving risk", Journal of Political Economy 56 (4), pp. 279–304*

***Mette, W. et. al.(2004)**, On the measurement of risk aversion from experimental data, Applied Economics, 36, 2443–51*

N. Raylu, T.P.S. Oei (2002), *Pathological Gambling - Clinical Psychology Review 22, 1009–1061*

Quiggin, J. (1991), *On the optimal design of lotteries, Economica, 58, 1–16*

Rodda, S. et al. (2004), *The Relationship Between Anxiety, Smoking, and Gambling in Electronic Gaming Machine Players, Journal of Gambling Studies, Volume 20, Number, Page 71-81*

S. Gainsbury et al. (2012), *Computers in Human Behavior 28, 1388–1398*

Spear P. (2000), *The adolescent brain and age-related behavioural manifestations. Neuroscience and Biobehavioral Reviews, 24:417–463*

Sproston, K. et. al. (2000), *Gambling Behaviour in Britain: Results from the British Gambling Prevalence Survey, National Centre for Social Research*

Steinberg, L. (2007), *Risk taking in adolescence: New perspectives from brain and behavioral science. Current Directions in Psychological Science, 16(2), 55–59*

Tavares, H. et. al. (2003), *Factors at play in faster progression for female pathological gamblers: an exploratory analysis. Journal of Clinical Psychiatry 64, 433–438*

Winters, K. et. al. (2002), *A prospective study of youth gambling behaviours, Psychology of Addictive Behaviors 16(1), 3-9*

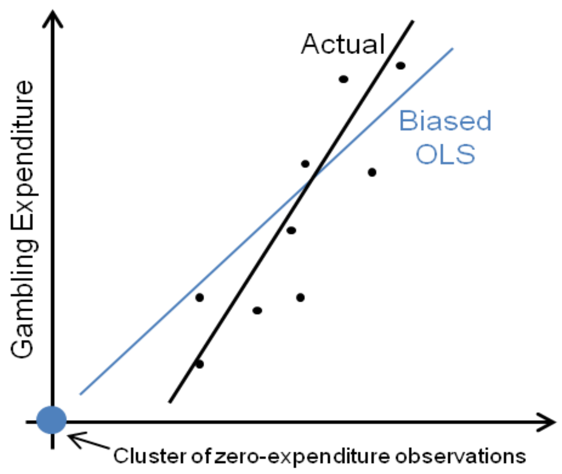
Wood, R. & Williams, R. (2011), *A comparative profile of the Internet gambler: Demographic characteristics, game play patterns, and problem gambling status, New Media & Society, 13, 1123–1141*

Appendices

Appendix I – “Skills-based” and “Pure-chance” activity categorisation

Skills-based	Betting: Horse races, Dog races, Football, Spread and Other; Poker; Virtual gaming
Pure-chance	National lottery; Other lottery; Scratch cards; Bingo (online & in-person)

Appendix II – Biased OLS estimation with zero-expenditure clusters



Appendix III – Variable list

Variable	Type	Description
Age	C	Individual’s age at the time of survey (in years)
Age^2	C	Squared value of the individual’s age at the time of survey
Gender	D	1 for male and 0 for female
Ethnicity	D	Black/Black British – 1 for Black ethnicity, 0 otherwise Asian/Asian British – 1 for Asian ethnicity, 0 otherwise Chinese – 1 for Chinese ethnicity, 0 otherwise Mixed – 1 for Mixed ethnicity, 0 otherwise
Personal Income	D	5th quintile – 1 for 5th (highest) quintile, 0 otherwise 4th quintile – 1 for 4th quintile (second highest) , 0 otherwise 3rd quintile – 1 for 3th quintile (median) , 0 otherwise 2nd quintile – 1 for 2th quintile (second lowest) , 0 otherwise

Parental Gambling	D	Regular – 1 for individuals with parents who gambled regularly but did not suffer from any gambling problems, 0 otherwise Problem gamblers– 1 for parents with gambling problems , 0 otherwise
Smoking	D	Smoke – 1 for current smokers, 0 otherwise
Method of gambling	D	Online only – 1 for individuals who gamble online only, 0 otherwise Both – 1 for individuals who gamble online and in-person, 0 otherwise
DSM	C	Numerical DSM score (1-10) with a score of 10 representing an addiction to gambling and 1 representing a normal participation
Region	D	Scotland – 1 for individuals who gamble online only, 0 otherwise
Attitude score	Deriv.	Numerical variable score (1-10) derived from responses to questions regarding views on gambling; 1 being most averse to gambling
Married	D	1 for individuals who are married, 0 otherwise
Household Income	Deriv.	Numerical variable derived from household income bands
Household Expenditure proxy	Deriv.	Logarithmic transformation of the number of people in the respondent's household; ln(number of people in household)

Appendix IV – Summary statistics (Dataset composition)

		Pure-chance		Skills-based	
		Number of individuals	%	Number of individuals	%
Education	Degree	947	24	125	20
	Professional Qualification	298	8	41	7
	A-Levels	424	11	86	14
	GCSE	1,049	27	176	28
	None	1,181	30	199	32
Gender	Male	1,912	48	496	78
	Female	2,074	52	140	22
Weekly Household Income Quintiles	1st (lowest)	669	20	144	27
	2nd	709	21	134	25
	3rd	666	20	82	15
	4th	661	20	99	18
	5th (highest)	599	18	79	15
Annual Personal Income Quintiles	1st (lowest)	549	16	101	18
	2nd	689	20	100	18
	3rd	898	26	165	29
	4th	600	17	88	16
	5th (highest)	730	21	107	19

Ethnicity	White	3,777	95	603	95
	Mixed	33	1	7	1
	Asian/Asian British	82	2	9	1
	Black/Black British	75	2	13	2
	Chinese	13	0	3	0
Parental Gambling	Never gambled	2,861	72	397	63
	Regular gamblers	945	24	205	32
	Problem gamblers	161	4	32	5
Smoking Status	Smoke	1,135	28	228	36
	Do not smoke	2,851	72	408	64
Alcohol	Drinks	3,037	76	519	82
	Does not drink	949	24	117	18
Marital Status	Married	2,592	65	349	55
	Not married	1,394	35	287	82
Method	In-person only	3,118	79	437	69
	Online only	93	2	1	0.2
	Both	760	19	195	31

Appendix V – Omitted Relevant Variable Diagnostic Tests

Pure-chance:

1) Ramsey RESET test using powers of the fitted values of ln(spendchance):

   Ho: model has no omitted variables

   F (3, 3980) = 10.96

   Prob > F = 0.0000

   Conclusion: Reject H_0 at the 5% significance level, thus, model potentially has omitted relevant variables

2) Model specification link test for single equation models:

```
      Source |       SS       df       MS              Number of obs =    3980
-------------+------------------------------           F(  2,  3977) =  209.70
       Model |  260.813824     2  130.406912           Prob > F      =  0.0000
    Residual |  2473.13652  3977  .621859823           R-squared     =  0.0954
-------------+------------------------------           Adj R-squared =  0.0949
       Total |  2733.95034  3979  .687094833           Root MSE      =  .78858

------------------------------------------------------------------------------
lnspendcha~e |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         hat |   1.353797    .390413     3.47   0.001      .588369    2.119226
      _hatsq |  -.0678774    .074313    -0.91   0.361    -.2135725    .0778176
       _cons |  -.4535849   .5103642    -0.89   0.374    -1.454185    .5470151
------------------------------------------------------------------------------
```

   Conclusion: As _hatsq is insignificant, this test fails to reject the hypothesis that the model is specified correctly, thereby implying no omitted relevant variables.

Skills-based:

1) Ramsey RESET test using powers of the fitted values of ln(spendskill):

   Ho: model has no omitted variables

   F(3, 636) = 2.84

   Prob > F = 0.0370

   <u>Conclusion:</u> Reject H_0 at the 5% significance level, thus, model potentially has omitted relevant variables

2) Model specification link test for single equation models:

Source	SS	df	MS
Model	315.580029	2	157.790015
Residual	810.560561	633	1.28050642
Total	1126.14059	635	1.77344975

Number of obs	=	636
F(2, 633)	=	123.22
Prob > F	=	0.0000
R-squared	=	0.2802
Adj R-squared	=	0.2780
Root MSE	=	1.1316

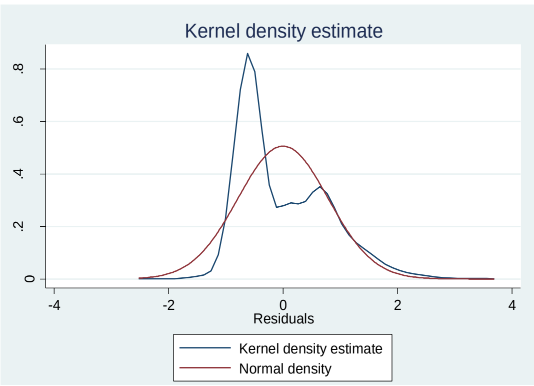
lnspendskill	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
_hat	.4916734	.3565123	1.38	0.168	-.2084165	1.191763
_hatsq	.0764899	.0527751	1.45	0.148	-.0271455	.1801253
_cons	.801728	.588544	1.36	0.174	-.3540068	1.957463

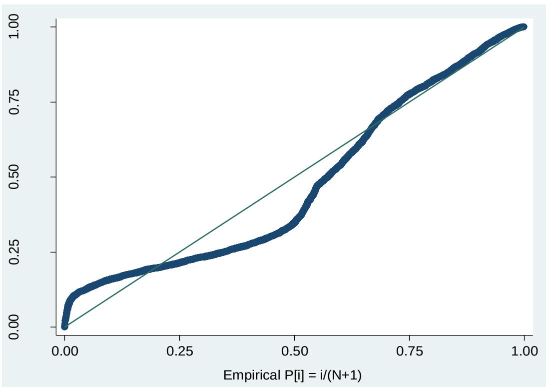
   <u>Conclusion:</u> As _hatsq is insignificant, this test fails to reject the hypothesis that the model is specified correctly, thereby implying no omitted relevant variables.

<u>Appendix VI – Normality Diagnostic Tests</u>

Pure-chance:

1) Residuals density plot and Standardised normality probability plot:





   <u>Conclusion:</u> Both tests indicate non-normal residuals

2) Shapiro-Wilk test for normality:

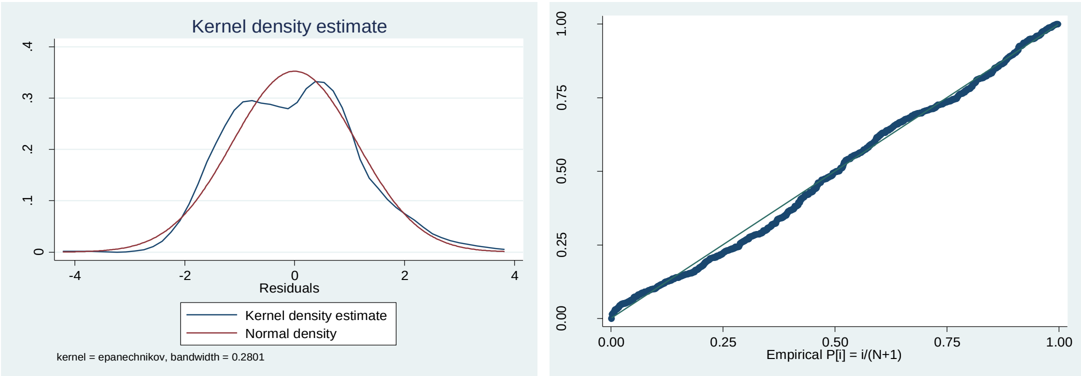
Shapiro-Wilk W test for normal data

Variable	Obs	W	V	z	Prob>z
resid	3980	0.92228	171.736	13.405	0.00000

Conclusion: Low Probability>Z suggests that the residuals display non-normality.

Skills-based

1) Residuals density plot and Standardised normality probability plot:



Conclusion: Both tests indicate non-normal residuals

2) Shapiro-Wilk test for normality:

Shapiro-Wilk W test for normal data

Variable	Obs	W	V	z	Prob>z
resid	636	0.98845	4.828	3.825	0.00007

Conclusion: Low Probability>Z suggests that the residuals display non-normality.

Appendix VII – Heteroscedasticity Diagnostic Tests

Pure-chance

Breusch-Pagan / Cook-Weisberg test for heteroscedasticity:

Ho: Constant variance

Variables: fitted values of lnspendchance

chi2(1) = 168.97

Prob > chi2 = 0.0000

Conclusion: Reject Ho at 5% significance level, thus, errors are heteroscedastic.

Skills-based

Breusch-Pagan / Cook-Weisberg test for heteroscedasticity:

Ho: Constant variance

Variables: fitted values of lnspendskill

chi2(1) = 26.68

Prob > chi2 = 0.0000

Conclusion: Reject Ho at 5% significance level, thus, errors are heteroscedastic.

Appendix VIII – Multicollinearity Diagnostic Tests

Pure-chance

1) Correlation matrix for detected multicollinearity in interactive variables:

	age	male	inponl	edqprof	age_in~l	male_e~f
age	1.0000					
male	-0.0113	1.0000				
inponl	-0.2483	0.0504	1.0000			
edqprof	0.0410	0.0211	0.0223	1.0000		
age_inponl	-0.1127	0.0453	0.9378	0.0277	1.0000	
male_edqprof	0.0389	0.2088	0.0187	0.7052	0.0210	1.0000

2) VIF test for final model:

   Mean VIF value = 1.33

   Conclusion: No multicollinearity in final model.

Skills-based

1) Correlation matrix for detected multicollinearity in interactive variables:

	age	male	married	online	male_m~d	age_on~e
age	1.0000					
male	-0.0185	1.0000				
married	0.2487	0.0139	1.0000			
online	0.0046	0.0211	-0.0438	1.0000		
male_married	0.2263	0.4622	0.7889	-0.0345	1.0000	
age_online	0.0046	0.0211	-0.0438	1.0000	-0.0345	1.0000

2) VIF test for final model:

   Mean VIF value = 4.46

   Conclusion: No multicollinearity in final model.

Appendix IX – Initial Regression (Skills-based)

The initial skills-based estimation includes all the variables of interest along with other control variables. Some of the insignificant controls are dropped in the final regression. However, the model is estimated with different combinations of controls to ensure that the dropping of these variables does not significantly impact the coefficients on the variables of interest.

```
Linear regression                                      Number of obs =     636
                                                       F( 29,   605) =       .
                                                       Prob > F      =       .
                                                       R-squared     =  0.2843
                                                       Root MSE      =  1.1542
```

lnspendskill	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
age	.0352247	.0161528	2.18	0.030	.0035022	.0669471
agesq	-.0003775	.0001619	-2.33	0.020	-.0006954	-.0000596
male	.4785752	.1101906	4.34	0.000	.2621727	.6949777
married	-.1998287	.1172868	-1.70	0.089	-.4301675	.0305101
edqdeg	-.2241521	.1453036	-1.54	0.123	-.5095128	.0612086
edqprof	-.1973863	.2131318	-0.93	0.355	-.6159543	.2211818
edqal	-.3909385	.1607226	-2.43	0.015	-.7065804	-.0752965
edqgcse	-.4151466	.1255973	-3.31	0.001	-.6618062	-.1684869
lnhhinc	.9108936	.321916	2.83	0.005	.278685	1.543102
anpinc2	.2220255	.1557926	1.43	0.155	-.0839344	.5279855
anpinc3	.212376	.1371551	1.55	0.122	-.056982	.481734
anpinc4	.3245734	.1792904	1.81	0.071	-.0275339	.6766806
anpinc5	.4590219	.1759469	2.61	0.009	.1134811	.8045627
lnnumpeep	-.1311025	.1138798	-1.15	0.250	-.3547501	.0925452
ethmix	.1880366	.5661927	0.33	0.740	-.9239052	1.299978
ethas	-.2541738	.4330719	-0.59	0.557	-1.104681	.596333
ethbla	-.7866589	.3488741	-2.25	0.024	-1.47181	-.1015076
ethchi	-.397485	.6715054	-0.59	0.554	-1.71625	.9212796
pargambreg	.1010142	.0987653	1.02	0.307	-.0929503	.2949787
pargambprob	-.6715235	.2131471	-3.15	0.002	-1.090122	-.2529254
smoke	.2060429	.1070735	1.92	0.055	-.004238	.4163238
drinks	-.122421	.1234009	-0.99	0.322	-.3647671	.1199252
online	.6911194	.2085145	3.31	0.001	.2816192	1.10062
inponl	.3907274	.1203627	3.25	0.001	.1543479	.6271069
nnorth	.0645439	.1120182	0.58	0.565	-.1554479	.2845357
wales	.3992926	.2482437	1.61	0.108	-.0882315	.8868167
scotland	-.039959	.1365228	-0.29	0.770	-.3080751	.228157
dsmtotsc	.1328137	.0163096	8.14	0.000	.1007834	.164844
atgs8sc	.0421671	.0112169	3.76	0.000	.0201384	.0641959
adv	-.1276272	.1333086	-0.96	0.339	-.389431	.1341767
_cons	-2.940465	1.425136	-2.06	0.040	-5.73928	-.1416497

Appendix IX – Initial Regression (Pure-chance)

A similar initial pure-chance estimation is presented below. Again, some of the insignificant controls are dropped in the final regression but the omission of these controls does not impact the coefficients on the variables of interest

.

Linear regression		
	Number of obs =	3980
	F(25, 3954) =	12.80
	Prob > F =	0.0000
	R-squared =	0.0952
	Root MSE =	.79095

lnspendcha~e	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
age	.0032097	.0010136	3.17	0.002	.0012225	.0051969
edqdeg	-.1713543	.040633	-4.22	0.000	-.2510178	-.0916907
edqprof	-.1097499	.0519094	-2.11	0.035	-.2115217	-.0079781
edqal	-.2049147	.0465704	-4.40	0.000	-.296219	-.1136104
anpinc5	.0770054	.0444956	1.73	0.084	-.010231	.1642418
ethbla	.3201551	.0887922	3.61	0.000	.1460724	.4942379
pargambreg	.1388154	.0313873	4.42	0.000	.0772786	.2003522
smoke	.1013558	.0304529	3.33	0.001	.0416509	.1610607
inponl	.2362025	.0357255	6.61	0.000	.1661605	.3062446
scotland	.196219	.0446598	4.39	0.000	.1086606	.2837773
dsmtotsc	.0760046	.0108909	6.98	0.000	.0546523	.0973569
atgs8sc	.0204857	.0030556	6.70	0.000	.014495	.0264765
male	-.0047459	.0270599	-0.18	0.861	-.0577986	.0483069
married	.0318238	.0325921	0.98	0.329	-.0320751	.0957227
edqgcse	-.0582161	.0376763	-1.55	0.122	-.1320828	.0156506
lnhhinc	.0253885	.085857	0.30	0.767	-.1429396	.1937167
anpinc2	.0301435	.0417764	0.72	0.471	-.0517618	.1120488
anpinc3	.0359873	.03886	0.93	0.354	-.0402002	.1121748
anpinc4	.0173043	.0431526	0.40	0.688	-.0672991	.1019077
lnnumpeep	.0097954	.0335052	0.29	0.770	-.0558938	.0754845
ethmix	-.0165201	.1502426	-0.11	0.912	-.3110803	.2780401
ethas	.0517312	.0857948	0.60	0.547	-.116475	.2199374
ethchi	-.2161431	.2061505	-1.05	0.294	-.6203144	.1880282
pargambprob	.0647513	.0644306	1.00	0.315	-.0615689	.1910716
online	-.0666802	.0764017	-0.87	0.383	-.2164707	.0831104
_cons	1.498904	.3830553	3.91	0.000	.7478997	2.249909

Appendix IX – Intermediate Regressions

The intermediate regression below shows the skills-based model with interactive variables. Several of these variables are either omitted or highly insignificant in the regression below due to high correlation with the original variables they are derived from. For example; the variable "online" becomes insignificant and is omitted as a result of including the age_online multiplicative dummy variable. The same problem was found in the pure-chance estimation.

```
Linear regression                                      Number of obs =     636
                                                       F( 31,   603) =       .
                                                       Prob > F      =       .
                                                       R-squared     =  0.2921
                                                       Root MSE      =  1.1498
```

lnspendskill	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
age	.0402192	.0169608	2.37	0.018	.0069098	.0735286
agesq	-.0003701	.0001627	-2.28	0.023	-.0006896	-.0000506
male	.9382123	.3459916	2.71	0.007	.2587174	1.617707
married	.1119311	.1922317	0.58	0.561	-.2655938	.4894561
edqal	-.2376326	.2906514	-0.82	0.414	-.8084447	.3331794
edqgcse	-.6301779	.2347476	-2.68	0.007	-1.0912	-.1691557
lnhhinc	.9869533	.3176106	3.11	0.002	.3631961	1.610711
anpinc4	.4271648	.179212	2.38	0.017	.0752093	.7791203
anpinc5	.5116193	.1771353	2.89	0.004	.1637423	.8594963
ethbla	1.365218	1.472558	0.93	0.354	-1.526748	4.257184
pargambprob	-.7161395	.2142172	-3.34	0.001	-1.136842	-.2954372
smoke	.2218607	.1070183	2.07	0.039	.0116868	.4320346
online	0	(omitted)				
inponl	.2883196	.2990317	0.96	0.335	-.2989505	.8755898
dsmtotsc	.1386086	.0155293	8.93	0.000	.1081105	.1691066
atgs8sc	.0412672	.0110707	3.73	0.000	.0195254	.063009
edqdeg	-.2204822	.1476717	-1.49	0.136	-.5104956	.0695312
edqprof	-.2301467	.20027	-1.15	0.251	-.6234582	.1631648
anpinc2	.2599827	.1577476	1.65	0.100	-.0498187	.569784
anpinc3	.2211338	.1363068	1.62	0.105	-.04656	.4888275
lnnumpeep	-.1301833	.1123029	-1.16	0.247	-.3507356	.0903689
ethmix	.2232654	.521741	0.43	0.669	-.8013849	1.247916
ethas	-.2154102	.4325632	-0.50	0.619	-1.064924	.6341032
ethchi	-.340187	.6905318	-0.49	0.622	-1.696326	1.015953
pargambreg	.104777	.1003388	1.04	0.297	-.0922789	.301833
age_male	-.0071403	.0059619	-1.20	0.232	-.018849	.0045683
male_married	-.3904626	.2197051	-1.78	0.076	-.8219427	.0410175
male_edqal	-.1685278	.3333546	-0.51	0.613	-.8232048	.4861492
male_edqgcse	.2518945	.2643922	0.95	0.341	-.2673468	.7711358
ethbla_hhinc	-.0264532	.0198902	-1.33	0.184	-.0655158	.0126093
ethbla_anpinc4	-2.177371	.7384129	-2.95	0.003	-3.627545	-.7271978
ethbla_anpinc5	0	(omitted)				
age_inponl	.0030716	.0072821	0.42	0.673	-.0112298	.0173729
age_online	.0129119	.0044294	2.92	0.004	.0042131	.0216108
_cons	-3.845281	1.381793	-2.78	0.006	-6.558992	-1.131569