The Economics of Gambling

A Collection of Essays

by

Rhys Wheeler

This thesis is submitted for the degree of Doctor of Philosophy

Submitted January 2018

Declaration

I declare that this thesis is my own work and has not been submitted for the award of a higher degree elsewhere.

Acknowledgements

I am incredibly grateful to the many people who have supported me throughout the course of my PhD, without whom I most certainly would not have been able to complete it.

Firstly, I am grateful to my supervisors Ian Walker and David Peel without whom this PhD would not have even started. I am particularly grateful to Ian who – regardless of where he is in the world – has always made himself available and willing to help with any problem and to discuss ideas with me. I must also thank Ian for accompanying me to several conferences over the course of my PhD and for introducing me to numerous researchers with whom I have had many interesting discussions and their comments have undoubtedly improved the work in this thesis. I am also grateful to members of the staff in the department who have helped my work with insightful questions and comments in seminars and annual NWDTC conferences. I consider myself incredibly lucky to have found myself working towards a PhD in such a supportive department. I am also thankful to participants of the International Conference on Gambling and Risk Taking 2016 and the Royal Economics Society Annual Conference 2017 whose comments helped shape my work. Also, I am thankful to Lisa Farrell for taking the time and effort necessary to go through this thesis and the useful comments provided.

I am grateful for financial assistance from the Economic and Social Research Council, who funded my PhD, and the Gambling Commission who supported the application. I am thankful to Ben Hader and James Holdaway, both of the Gambling Commission, who helped organise several opportunities to present work there and the resulting comments helped improve the policy relevance of my work.

Fellow PhD students, past and present, have provided both academic help and friendship during my PhD. I am especially grateful to Robert Pryce, Simon Spavound, Emma Gorman, Muhammad Salihu, Zhenxiong Li, and Xingzhi Yao for their friendship over the past 3 years. The laughs and the camaraderie amongst us have played an important role in my PhD experience, and it has been made all the better for it. I am also grateful to my friends outside of academia. In particular Samuel Taylor-Woods, whose favourite pastime appears to be reminding me that any work-related problems are not so bad and who has always given me a healthy perspective on whatever it is I am doing.

Last but not least I am grateful to my family for their seemingly limitless love and support, not just during my PhD but for my entire life. I am especially grateful to my parents, Shirley and David Wheeler, who have supported me in every way imaginable. Finally, I am grateful to my grandfathers, Ronald Peacock and Michael Wheeler. Their love, support, and interest in everything I have done has been invaluable, yet they will not have chance to read this one.

Abstract

Chapter 2 examines the harm associated with being a problem gambler. Problem gambling is conventionally determined by having a score in a questionnaire screen that exceeds some critical value. The UK is fortunate in having large representative sample surveys that embed such questions, and our estimate from the 2010 survey is that several hundred thousand people in the UK could be afflicted by PG. However, existing literature has not evaluated the size of the harm associated with being a problem gambler and this chapter uses this individual level survey data to evaluate the effect of problem gambling on self-reported well-being. Together with a corresponding effect of income on well-being a money-metric of the harm associated with being a problem gambler is derived. An important methodological challenge is that well-being and the harm experienced may be simultaneously determined. Nonetheless, instrumental variable estimates suggest that problem gambling imposes an even larger reduction in well-being than least squares would suggest. The role of gambling expenditures in the transmission between problem gambling and well-being is considered, distinguishing between draw-based games, such as lotto, from scratchcards, and from other forms of gambling.

Chapter 3 investigates the price elasticity of demand for the UK National Lottery – a state-licensed, draw-based lotto game. Little is known about the price elasticity of demand for gambling products because the "price" is typically hard to define. The exception is "lotto" where an economics literature has focused on the response of sales to variations in the prize distribution. Existing literature has used these responses make inferences about the price elasticity of demand, where price is defined as the cost of entry minus the expected winnings. In particular, the variation in the value of the jackpot prize pool, due to rollovers that are a feature of lotto, has been used as an instrument for price. This chapter argues that rollovers do not make valid instruments, because of their correlation with lagged sales, and propose an alternative identification strategy which exploits two arcane features of lotto. Finally, this chapter evaluates whether changes to the design of the UK National Lottery in 2013 and 2015 had a positive effect on the sales figures.

Chapter 4 investigates the extent to which the large, flat-rate tax imposed on the UK National Lottery is regressive. This chapter evaluates a Working-Leser demand model for lotto tickets using both Heckman's selection model and Cragg's double hurdle estimator using

household-level data. A unique strategy is employed to identify these two-stage routines by exploiting exogenous differences in consumer preference arising from religious practice. The income elasticity of lottery tickets is found to be significantly lower than previous estimates, suggesting that lottery tickets are inferior goods and that the (high) flat-rate tax imposed on lotto tickets is more regressive than previously thought.

Whilst the three chapters are stand-alone essays, they are linked by the use of modern statistical techniques and the use of the best possible data. Together, they address key issues on the economics of gambling and the results are new to their respective literatures and of interest to academics and policy makers alike.

Notes

This PhD has been supported by a CASE award from the Economic and Social Research Council.

References are collated at the end of this thesis, rather than at the end of each chapter.

All three chapters have been presented at various stages of completion at the International Conference on Gambling and Risk Taking 2016 in Las Vegas. Chapter 2 has also been presented at the Royal Economics Society Annual Conference 2017 in Bristol, the International Health Economics Association Conference 2017 in Boston, USA, and the North West Social Science Doctoral Training Partnership (NWSSDTP) Economics Conference 2017 in Manchester, UK, as well as at numerous seminars. Chapter 3 has also been presented at the North West Doctoral Training Centre (NWDTC) Economics Conference 2016 in Lancaster, UK. Comments and discussions at these conferences have helped shape the ideas and, ultimately, the chapters of this thesis. Chapter 2 has also been completed with the help of Ian Walker and Robert Pryce who are co-authors to this chapter. Chapter 3 was completed under the supervision of Ian Walker. Chapter 4 was aided by help from Robert Pryce who has expertise in using the dataset in this chapter and using religious belief as an identification strategy for demand models using micro-level data.

An earlier version of Chapter 2 appears as a working paper in the Lancaster University Economics Working Paper Series as No. 2017/011.

Data for Chapter 2 were collected by the National Centre for Social Research (NATCEN) on behalf of the Gambling Commission and provided to us by the UK Data Service (UKDS) at Essex University. It is readily available to other researchers at www.ukdataservice.ac.uk. Data for Chapter 3 were collected by Richard K. Lloyd and are used here with written permission. The data are readily available at his website, www.lottery.merseyworld.com. Data for Chapter 4 were collected by the Office for National Statistics (ONS) and the Department for Environment, Food and Rural Affairs (DEFRA) and provided by the UKDS. It is also readily available to other researchers via the UKDS website. The merging of annual data for Chapter 4 was done by Robert Pryce.

Table of Contents

List of Figures

List of Tables

1 Introduction

1.1 Background

The premise of any gamble is simple. It involves a minimum of two parties who are each willing and able to wager something which the other party values on different outcomes of some future, uncertain event. The event in question can be anything for which the outcome is not known – from the cards drawn from a shuffled deck to the outcome of a national election and the number of points scored by any given sports team to today's closing value of the FTSE100.

Written evidence of gambling activity in some form or another can be found as far back as the ancient Egyptians and Assyrians. In fact, it has likely existed for as long as mankind has been able to form opposing predictions on the outcome of some future event. Dice have been found dating back to 3000 BC in Mesopotamia, gambling was enjoyed by the Greek and Roman elite as after-dinner entertainment, lots were drawn by soldiers to lay claim to Jesus' belongings, and Queen Elizabeth I used a lottery to fund defence against the Spanish Armada. Moreover, the relationship between gambling, morality, and the law enriches the history of one of mankind's oldest pastimes – drawing interest from philosophers, psychologists, religious leaders, and, of course, economists alike.

There are many aspects of gambling that make the activity of particular interest to economists. Its very nature provides an environment to examine how individuals behave under risk and uncertainty. With gambling opportunities almost never available at fair odds, participation in gambling poses the question of why those same individuals will pay to bear risk *and* pay premiums to avoid risk with insurance (Friedman and Savage, 1948). Viewing bets as simple financial contracts gives economists a setting to examine the implications of the efficient market hypothesis without the complexities of traditional financial markets (Vaughan-Williams, 2005). As with any other good or service, economists naturally ask questions about supply, demand, surpluses and tax, and the use of gambling by governments – most often in the form of lotteries – to raise public finance makes gambling a particularly interesting topic. A final concern amongst economists with regards to gambling one of externalities. In addition to the association between gambling and crime, problem gambling is receiving increased attention from both policy makers and academics as a public health issue.

1.1.1 Gambling as a global industry

Throughout history gambling activity has been subject to various degrees of regulation. Today, this regulation is determined at various levels of government and exists in some form in almost all countries around the world.

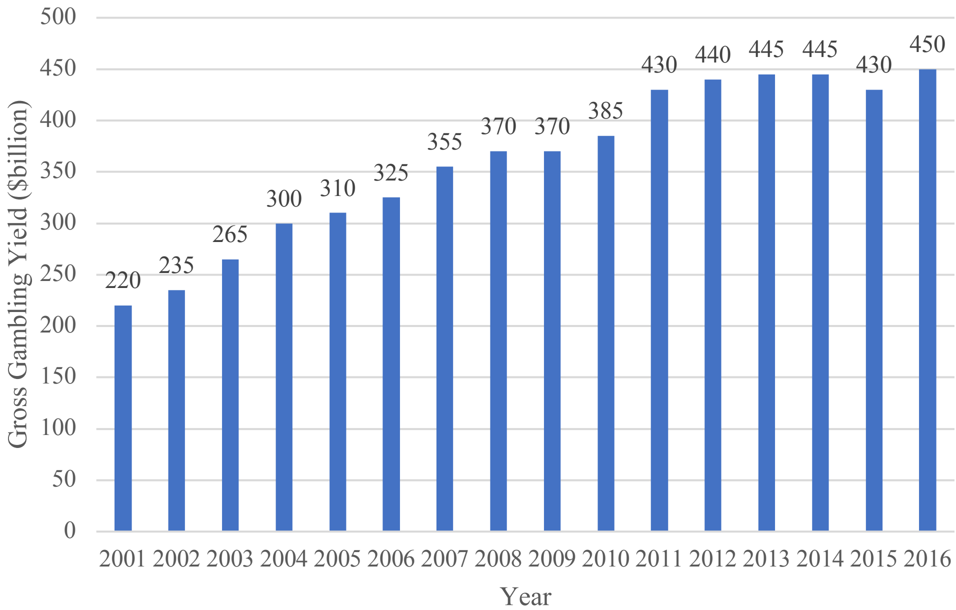
In the US, gambling is regulated at the state level and the legality of gambling varies from state to state and from one form of gambling to another. Lotteries, for example, are permitted (and operated) in all but 6 states[1], whereas only 4 states permit gambling online[2]. Australia and Canada both have legalised gambling nationwide, but specific regulations and licencing occurs at the province level. In the EU, gambling in-person is regulated and licensed by individual member states, though legal rules vary only slightly between nations. Recent developments in technology and the rapid rise of online gambling led to a cooperation agreement being signed in 2015 by EU members to allow gambling regulators to share their practices. In China, gambling is officially illegal, however citizens are able to participate in state-run lotteries and restricted gambling activity is permitted in Hong Kong and Macau as an added attraction for tourists.

Partly due to increasingly liberal attitudes of governments around the world, and partly to technological developments allowing individuals to gamble remotely and online, the gambling sector expanded to become a significant global industry over the past 5 decades. Figure 1.1 shows the global annual "gross gambling yield" (GGY) – the industry term for wagers placed minus winnings paid – from 2001 to 2016. In this 15-year period alone, annual takings for gambling operators doubled from $220b to $450b. This rapid rise in global revenues can be seen in each country where legal gambling occurs. A report by the Canadian Gaming Association (2011) shows that from 1995 the legalised gambling industry tripled in size and has become the largest of the entertainment sector by revenue. The Canadian gambling market contributed $31b CAD to the economy, employing over 128,000 people in 2010. In Australia, a report by the Australian Productivity Commission (2010) showed the GGY in 2008-2009 was $19b AUD, or 3.1% of household consumption. The American Gaming Association (2016) reports the GGY from casinos alone in the US was $38.4b USD in 2015 and the North

[1] The 6 states which prohibit lotteries are Alabama, Alaska, Hawaii, Mississippi, Nevada, and Delaware. American Samoa, an unincorporated territory of the United States also prohibits lotteries.
[2] The 4 states which permit gambling online are Delaware, Nevada, New Jersey, and Pennsylvania.

American Association of State and Provincial Lotteries reports a GGY from lotteries of an additional $35b USD in 2016.

Figure 1.1: Global gross gambling yield (GGY) in $USD (billions) 2001-2016



Source: Statista (2017)

1.1.2 Gambling in the UK

In the UK, gambling is regulated by the Gambling Commission and gambling firms must possess a license to operate legally. Many forms of gambling are permitted under UK law both in person and remotely via the internet including lotteries, gambling at casinos, betting on sports and other events, and on virtual games but age restrictions are enforced[3].

1.1.2.1 Size of the UK gambling market

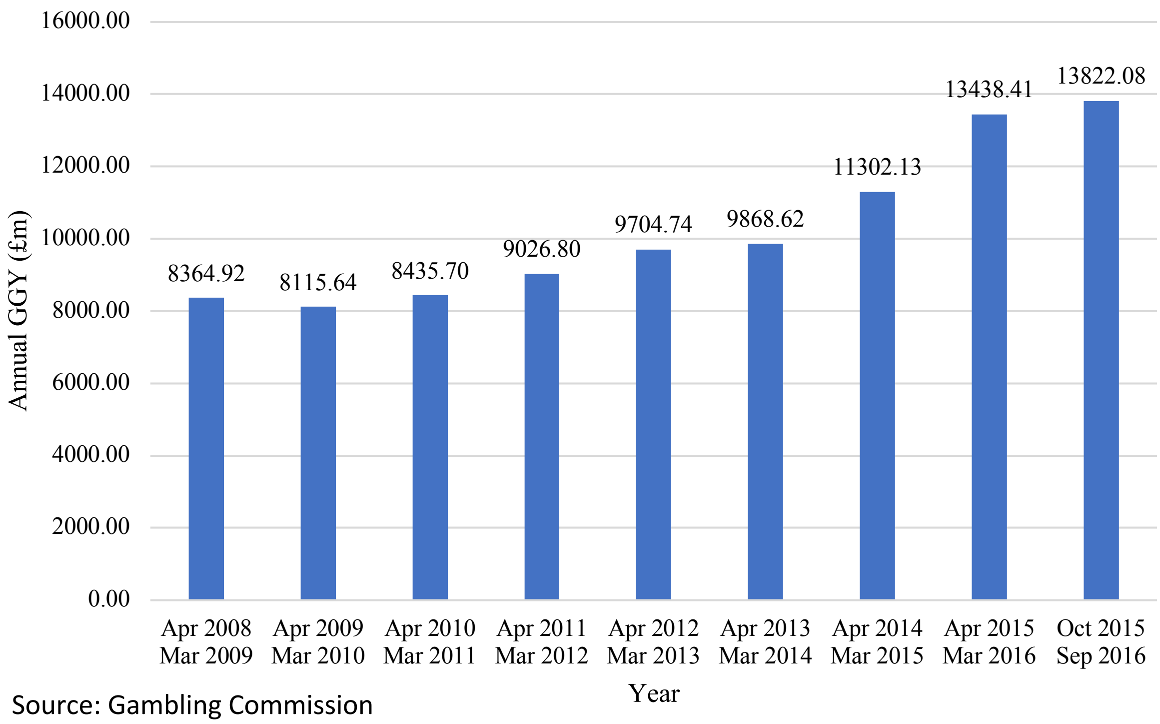
In the year to September 2016, the GGY of the UK gambling market totalled £13.8b and employed 106,678 people (Gambling Commission, 2016). Figure 1.2 shows the increase in the size of the regulated UK gambling market by GGY from April 2008 to September 2010. Official figures show that the market has grown by almost £5.5b over this time period. A large portion of this reported increase is due to a legal change in October 2014 which expanded the scope of the UK regulator to include all remote bets placed by UK customers, regardless of the registered country of the betting company used. Prior to this, the regulator was limited to reporting figures from UK registered companies and, unfortunately, no GGY figures

[3] In the UK, an individual must be 18 or over to participate in any commercial gambling with the exception of lotto, scratchcards, and football pools, for which an individual must be 16 or over.

comparable to those post-2014 are available. Nonetheless, comparing annual GGY figures before this date shows steady growth in the UK market from April 2009 to March 2014, increasing by around £1.8b over this time period. The market appears to have continued this growth post-2014, increasing by around £400m in GGY from 2015 to 2016.

Gambling in the UK not only generates these significant revenues for the gambling companies themselves, but also makes substantial contributions to the UK tax budget. In 2016-2017 the industry generated £2.7b in tax receipts and, as Figure 1.3 shows, these tax revenues have increased year-on-year since 2006-2007. Since 2001, winnings from gambling in the UK are not considered income or capital gains and are therefore received tax-free but betting and gaming taxes in the UK are instead imposed on gambling operators themselves. Gambling duties are charged as a proportion of profits or GGY of individual gambling firms and the rate applied varies depending on both the size of the company and the type of gambling activity. For instance, the current duty levied on earnings from remote gambling, in-person betting, and small in-person gaming is 15% of profits from these activities, whereas bingo and lotteries are taxed at lower rates of 10% and 12%, respectively[4].

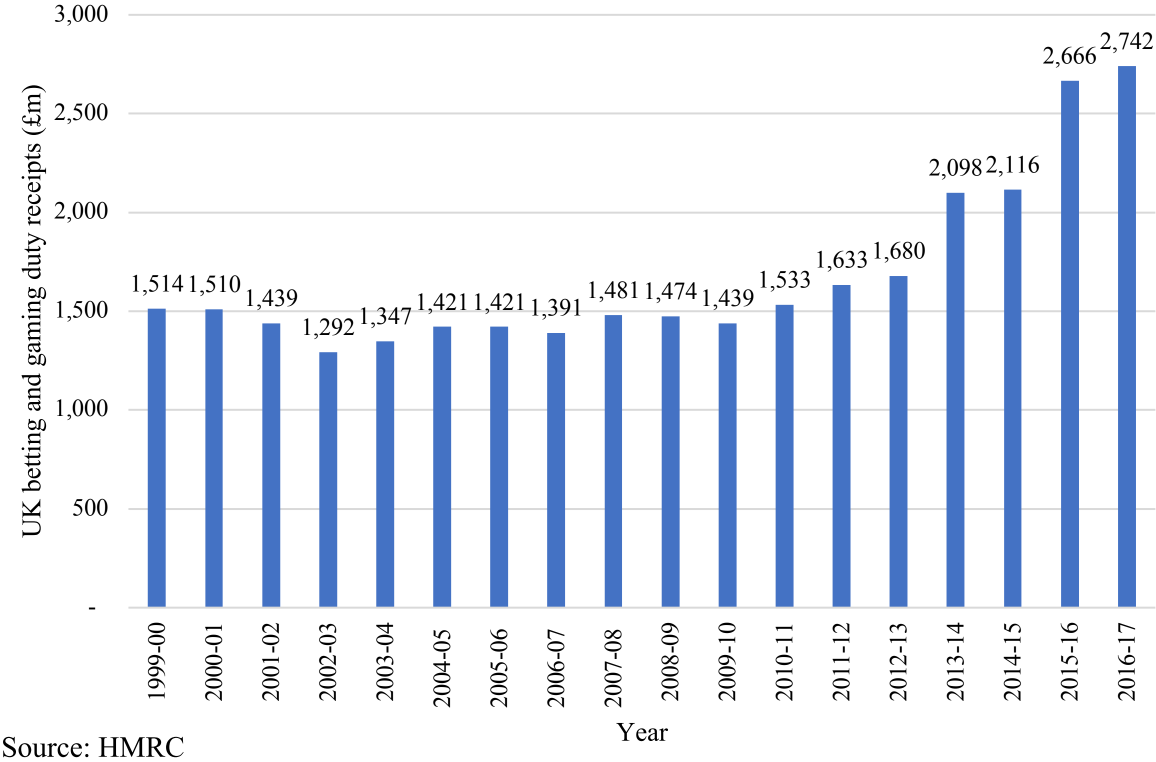
Figure 1.2: UK GGY in £millions April 2008-September 2016



Source: Gambling Commission

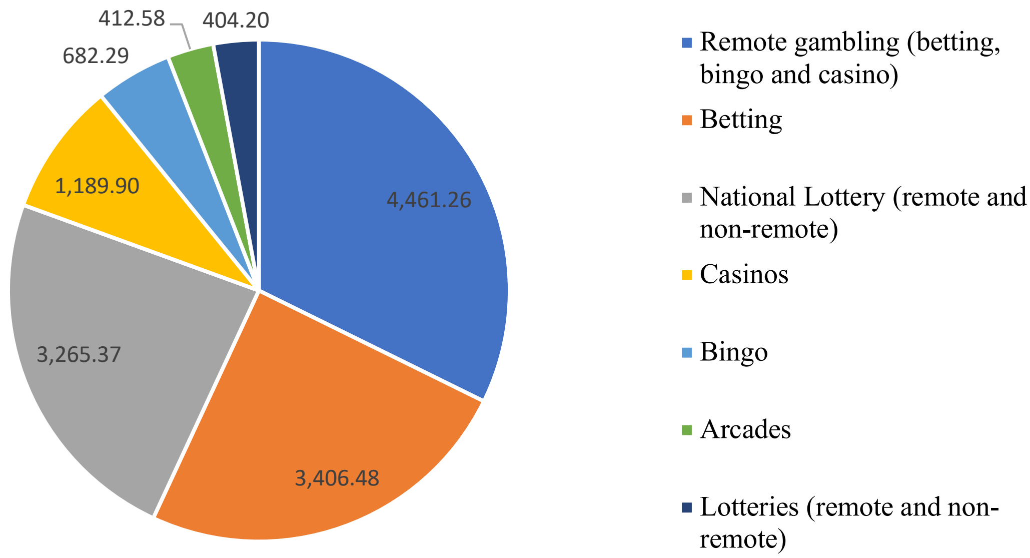
[4] A complete overview of UK gambling duty can be found on the UK government website at www.gov.uk.

Figure 1.3: UK betting and gaming tax receipts 1999-2017 (£m)



Source: HMRC

Figure 1.4: UK market share of each gambling category by GGY



Source: Gambling Commission (2016)

Figure 1.4 shows the composition of the UK market by the GGY of each activity. The year to 2016 was the first in which remote gambling was the largest UK gambling sector by

GGY, more than double the profits compared to 2015 (£2.2b)[5] and accounting for 33% of total gambling revenues. The second largest sector is "betting" which includes gambling in-person at a bookmaker (either placing bets on real-life events or spent on virtual gaming machines in-store), bets placed on site at sporting events, and entries to the football pools. Third is the National Lottery which operates a range of numbers-based lotteries and scratchcard games and makes up almost a quarter of the market by GGY. The UK National Lottery is of particular interest as its primary function is to raise revenues for "good causes" – public goods which are otherwise not financed using regular tax money. The remainder of the UK market is made up of revenues from casinos, bingo, arcades, and other small charitable lotteries.

1.1.2.2 The National Lottery

The largest gambling provider by GGY in the UK is Camelot, who is licensed by the Gambling Commission to operate the National Lottery and the games it provides feature in all of the following chapters. It currently offers four numbers-based lottery games[6], a range of in-person instant win games known as scratchcards, and several online instant win games. As with many lotteries around the world, a large share of the proceeds from the UK National Lottery are used for financing public goods[7]. Unlike most lotteries, however, these funds are allocated to public goods – termed "good causes" – which are not covered by regular tax money in the UK. The UK National Lottery generates an average of £30 million per week in good causes funding – a total of £37 billion since its first game in 1994. Investment of these funds is decided by 12 organisations chosen by the UK government and distributed across four broad categories: health, education, environment and charitable causes (40%), sport (20%), arts (20%), and heritage (20%). According to the National Lottery website[8], these funds have helped to finance over 500,000 projects in the UK and have previously been used to finance projects as varied as the London 2012 Summer Olympics and restoration work on Stonehenge and the Royal Opera House.

[5] Remote gambling here refers to any form of gambling not done in person at designated business premises. The vast majority of remote betting today is done via the internet, though it also includes gambling via telephone.
[6] Two of the four numbers games, Lotto and Euromillions, primarily offer "pari-mutuel" prizes – in which players aim to win shares of prize pools funded by entry fees – and the other two are fixed-prize games.
[7] Currently, 28% of lottery sales revenue is earmarked for spending on good causes. 12% of revenues are paid in lottery duty to HMRC, 50% is returned in prizes, with the remaining 10% of revenues covering operating costs (9%) and profit (1%).
[8] www.national-lottery.co.uk

Dwindling ticket sales for the National Lottery's flagship game – Lotto – has recently prompted two overhauls of the game's design. Changes in October 2013 saw the sticker price of tickets double from £1 to £2, the introduction of 50 raffle prizes of £20,000 to each draw, and changes made the share of revenues allocated to each prize pool. The final change to the game saw the fixed prize paid to players who matched 3 of the 6 winning numbers increase from £10 to £25. A second redesign occurred in October 2015 which introduced a further raffle prize of £1m each to each draw and expanded the set of numbers from which players choose (and which make up the winning combination) from 49 to 59, decreasing the probability of winning the jackpot prize from around 1/14m to 1/45m. This later reform to the game also saw the introduction of a free ticket to the next draw being awarded for matching 2 of the 6 winning numbers and lifted a cap on the number of times the game could "roll-over" before the jackpot prize money would be shared amongst winners of lower prize tiers.[9]

1.1.3 Problem gambling

Whilst for many individuals gambling is merely a source of harmless fun, there are a number for whom it can be a very real and serious problem. This problem gambling has received growing attention in recent years from the media, academics, and politicians. However, the literature in economics, medicine, and psychology on problem gambling is still in its relative infancy, with the American Psychiatric Association only classifying the condition as an addictive disorder as recently as 2010.

Problem gambling is typically diagnosed by an individual exceeding some critical value on a screen of scored questions. Use of such screens in representative sample surveys has allowed researchers to estimate the prevalence of problem gambling for entire populations. Estimates of this prevalence rate varies across the UK from 0.5% according to the Health Survey England in 2012, to 1.1% from the Welsh Problem Gambling Survey in 2015. Calado and Griffiths (2016) present a review of the literature surrounding prevalence rates around the world. Their data suggests that the UK's problem gambling prevalence rate is comparable with that of Australia (0.4-0.6%), and relatively low compared those of Canada (2-3.5%) and the US (3.5-4.6%), where the same diagnostic screens have been used. Whilst problem gambling only affects a small *proportion* of the population, the *number* of people afflicted with the

[9] Prior to the 2015 game changes, the jackpot would be shared amongst winners of the next highest prize tier if it was not won for 4 consecutive draws. After the changes, the game was free to roll-over until the jackpot reached £50m; after which, if it had still not been won, would be shared amongst winners of the next highest prize tier. This cap was reduced in 2016 to £22m.

condition is still large. In the UK, the estimated prevalence rate of 0.5% indicates over 300,000 people are affected, and in the US a prevalence rate of 4.6% implies there are almost 15 million problem gamblers.

1.2 Aims and structure

The aim of this thesis is to make three novel contributions to the current understanding of issues surrounding the economics of gambling. The following chapters can be read as stand-alone essays on three different aspects of the economics of gambling but are also connected by the application of modern statistical techniques to the best data available. Collectively, these chapters contribute to our understanding of the level of harm experienced by those afflicted with problem gambling, of how to model lotto demand, and the regressivity of taxes imposed on the UK lotto.

1.2.1 Overview of Chapter 2

Whilst the prevalence of problem gambling is well documented in the literature, and there is some evidence on the externalities caused by gambling, there is little known about the level of harm experienced by afflicted individuals. Chapter 2 aims to fill this gap in the literature using a well-being methodology in which self-reported happiness is the dependent variable and problem gambling and income are the independent variables of interest. Dividing the coefficients on the two independent variables gives a money-metric of the loss in well-being associated with being a problem gambler and acts as a catch-all measure of the harm experienced by those affected. By aggregating this money-metric of the harm associated with well-being using prevalence rates, average income, and the UK population figure, the implied loss in happiness associated with problem gambling is equivalent to increasing national incomes by approximately £30b. To obtain causal estimates, an instrumental variables strategy is employed using parental problem gambling and aggregate estimates from this are even larger. The results are found to be robust to both measurement error in the independent variables and ordinality assumptions about well-being. The transmission mechanism of lost well-being from problem gambling via gambling expenditure is also considered. Although measurement error in the expenditure renders much of this analysis ineffective, there is tentative evidence that expenditure on scratchcards plays a mediating role.

1.2.2 Overview of Chapter 3

This chapter uses aggregate-level data for each draw of UK lotto to investigate the price elasticity of demand for the game and the importance of prices as a determinant of demand.

Accurately modelling demand for the UK lotto is important because of its function as a means to raise finance for public good provision and the large tax revenues collected from its operation. Moreover, such a model is a necessity in determining whether these revenues are being maximised. Existing models of lotto demand use a unique feature of the game – rollovers – as an instrumental variable to overcome the simultaneity bias inherent to such analysis. Chapter 3 argues that rollovers make poor instruments as they are also correlated with sales and instead proposes exploiting behavioural biases in players to identify the model. The role of prize size is also investigated as a key determinant of demand for lotto games. Findings from these models and testing against models of demand that use price alone suggest that existing models of lotto demand are also deficient insofar as prizes also affect the demand for tickets beyond their influence on price. Finally, using data beyond the main sample period, this chapter evaluates whether the game redesigns of 2013 and 2015 increased revenue relative to what would have been received had the operator left the game unchanged. The evidence suggests that sales did indeed increase beyond what they would have been had the game design remained the same.

1.2.3 Overview of Chapter 4

Lotteries notoriously offer participants one of the worst returns of any form of gambling and the UK lotto is no exception. With just 50% of revenues being returned in prizes, the combined good causes contribution and lottery duty of the UK game is in effect taxed at 40% with the remainder covering operating costs and profit. Moreover, with the game being operated under license from the government, the question of the regressivity of these taxes is even more important. Due to the flat rate of tax imposed on the UK lotto, an analysis of this question trivialises to determining whether the income elasticity of demand is less than 1. Chapter 4 improves upon existing literature by estimating a Working-Leser demand function and finds income elasticity estimates significantly lower than previous estimates. Estimating this specification using micro-level data requires modern statistical techniques to overcome bias induced by the presence of zeroes in expenditure data. Both Heckman's selection model and double hurdle models are considered. The identification of both these two-stage procedures is done using exogenous differences in consumer preference arising from religious belief. Whilst there is little qualitative difference implied by the estimates of income elasticity, choosing between estimates from these two models has importance to the implied magnitude of the regressivity of lotto taxes. A Suits' index estimate is also calculated and suggests that the regressivity of taxation on UK lotto tickets is in the ballpark of alcohol and tobacco taxation.

2 How much of a problem is problem gambling?

2.1 Introduction

This chapter is concerned with evaluating the reduction in well-being caused by being a "problem gambler". Problem gambling (PG) is usually defined by aggregating responses across questions which are embedded within a screen designed by psychologists to capture such behaviour. When administered to large samples of individuals this facilitates an estimate of the prevalence of problem gambling. An individual is defined as a problem gambler (PG=1) if that individual's score on the screen exceeds some critical value. There are several such screens used in this literature and they each contain questions that are designed to detect behaviour associated with pathological gambling and gambling harms. Our estimate for Great Britain (i.e. UK minus Northern Ireland, who are a very small proportion of the UK population and were not surveyed in our data) is that there are approximately ⅓ million (about 0.7% of approximately 45m) adults who are assessed to be problem gamblers, and this is typical of the literature. A recent review of PG by Williams *et al* (2012) provided estimates for the USA that suggested the PG prevalence rate was 2.2% of approximately 240m US adults, or around 5m.

However, none of the extensive literature that attempts to quantify the prevalence of problem gambling has attempted to also quantify the costs that problem gambling imposes on the individuals afflicted by it, and this is the main contribution of this paper. Here, a well-being methodology is used which scales the effect of PG on self-reported well-being with the effect of income on well-being to produce a money metric (see, for example, the recent application to valuing the availability of health insurance using the Oregon experiment by Finklestein *et al*, 2012). The baseline estimate of the aggregate loss in well-being associated with PG is approximately £70 thousand ($90k at current exchange rates) per problem gambler, or over £25 billion per annum across the GB population as a whole – a figure that is the same order of magnitude as that often associated with alcohol abuse, exceeds the tax take on gambling products, and even exceeds overall gambling expenditures[10]. The US estimates of the

[10] HMRC reports gambling tax revenue of £2.1 billion for 2014/15, and the Gambling Commission reports further good causes revenue from the National Lottery portfolio of games of £3.8 billion. Alcohol expenditure in 2010 was approximately £16b and the Institute of Alcohol Studies (IAS) estimates of the harms associated with alcohol abuse in the UK is in the order of £21b pa. This estimate is not comparable with our well-being based measure, but rather aggregates effects associated with crime, absenteeism, and health. The IAS reports results that also include wider harms that aggregate to over £50b pa. Tax revenue from alcohol is approximately £10b p.a. Note that the measure here is *not*

prevalence of PG would suggest an overall loss in well-being that would be equivalent to $450 billion pa if the presented estimates were applicable to the US.

Gambling is an important part of many economies. UK expenditure net of winnings (sometimes referred to as Gross Gambling Yield, GGY) in 2014/5 was approximately £3 billion, or about 0.2% of GNP; and overall expenditure is close to £12b or more than 0.7% of GNP. The Economist (2017) cites a report by H2 Gambling Capital Consulting that estimates that US gambling losses amount to $116 billion in 2016, despite the much stricter regulatory controls on off-course/internet betting in the US than in most other countries.

Consumers in the UK do not directly pay tax on gambling - with the exception of the products sold by the National Lottery, sales of which are taxed at 12% plus a levy for "good causes" of approximately 28%. Suppliers of other products pay 15% on "profits", defined as revenue minus winnings and in the case of FOBTs (Fixed odds betting terminals) that are thought to be particularly likely to be associated with PG, the tax is 25%. Relative to other "sins", gambling in the UK is not highly taxed; and taxes have been driven downward as regulators and tax authorities have struggled with the increasingly footloose nature of the industry that is becoming more highly dominated by online, and often offshore, provision.

The motivation for sin taxes is driven by the notion that consumption causes "internalities" as well as externalities. The former are harms that are self-inflicted, sometimes uncertain, and often long-term, but which are not fully internalised by the consumer.[11] In which case, penal levels of tax can be motivated not just by Pigovian considerations but also by wider considerations associated with the behavioural deficiencies in individual preferences. Indeed, whether to legalise and tax the consumption of some commodities, as opposed to criminalise, has been analysed by Becker *et al* (2006) who show that the demand and supply elasticities as well as the nature of harms, play an important role in the optimal design of policy. Thus, the contribution here speaks to one of the critical parameters relevant to the design of public policy relevant to potentially harmful products.

---

one of social costs in the traditional sense (see Walker and Barnett (1999) for issues around the measurement of social costs in the context of gambling).

[11] See Gruber and Koszegi (2004) for an analysis of cigarette taxation in the context of a model with time-inconsistent behaviour. Also see Leicester and Levell (2016) for an analysis of the effects of a smoking policy on well-being which is consistent with a behavioural interpretation of addiction.

This analysis exploits the availability of data on well-being in a large household sample survey. We construct a financial measure of PG harm by estimating the relationship between subjective well-being, PG, and income. The methodology draws on the seminal work on "happiness" and a particularly good early exemplar is Clark and Oswald (2002). The methodology estimates well-being effects for a variety of life outcomes using large random samples of individuals, and the relative coefficients of income and of the life event in question are used to provide a financial compensating amount for that event. The method has been used to evaluate the effects of marital status, unemployment, health, and many other phenomena.[12] The present application also adopts this idea of scaling the effect of PG on well-being by the effect of income on well-being to monetize the estimate of the well-being effect of PG. This well-being approach is a catch-all one – it looks, not at the mediating mechanisms, but directly at the effect on the well-being of individuals, irrespective of how that comes about. The results are dramatic: the baseline estimate is that the harm associated with PG is over £70,000 pppa which, for ⅓ of a million PGs, amounts to an aggregate loss in well-being for the UK of over £25b pa. The size of consumer surplus enjoyed by responsible gamblers, and the revenue generated for the HMRC, is likely to be small in comparison to such huge welfare losses.

However, there are a number of threats to the legitimacy of the well-being methodology that are typically not addressed in the existing literature. In particular, measurement error and other sources of endogeneity are usually ignored by the simple regression method that is used to obtain the statistical estimates. Existing measures of PG are based on self-reports and are very likely to be measured with error in the data. Thus, OLS estimates will inevitably understate the effect of PG on well-being.[13] On the other hand, PG might be symptom of low well-being rather than the other way around. This reverse causality is likely to bias the estimate of PG on well-being upward. Since these two sources of bias counteract each other, it is unclear what the net effect would be – the true effect might be larger or smaller than baseline results obtained by OLS.

[12] Powdthavee and van den Berg (2011) use the well-being method to evaluate the effect of a variety of medical conditions on several measures of well-being. However, they do not consider the issues raised above which are likely to bias the results in different ways for different well-being measures. An excellent review of the issues around the use of subjective well-being measures can be found in Nikilova (2016), albeit in the context of development.

[13] To make things more complicated, measurement error in income is likely to understate the effect of income on well-being – and since it is used to scale the effect of PG on well-being this will tend to overstate the financial effect of the loss in well-being associated with PG.

This chapter attempts to tackle these endogeneity issues head-on using an instrumental variables (IV) approach. The reported coefficients suggest even larger losses in aggregate well-being than the simple headline results – nearly double what the OLS estimates suggest. However, a substantially larger dataset would be needed to come to any firm conclusions and, for the moment, the OLS estimates can be treated as a plausible lower bound.

A complementary approach to this investigation of well-being data would consider the effect of PG on a list of all relevant mediators – for example, in the case of PG, researchers may look at the effect on mental health, employment, wages (a measure of productivity in the labour market) conditional on employment, tax receipts, and welfare payments etc. The predicted effects would then need to be "valued" and aggregated[14]. A recent UK example is Thorley *et al* (2016) which focuses on those outcomes that affect some of the range of other people and agencies, apart from the PG. That study, from the Institute for Public Policy Research, focuses on only certain aspects of health, housing, crime, and welfare and employment, and provides estimates of a cost of just £1.2b pa. This alternative methodology is likely to miss many elements of the transmission mechanism for which data is not available. Moreover, this method is also more likely to miss true externalities – the effects of one person's PG on other people is not measured in the IPPR study, for example. The well-being method is less likely to be affected by this – since it will capture the effects of own PG on other people to the extent that the former feels altruistic towards the latter.[15] In any event, the well-being method seems likely to yield bigger estimates than the alternative to the extent that the latter embraces only a subset of possible mediators.

Nonetheless, the analysis in this chapter is deficient in that it has no policy implications over and above highlighting the magnitude of the PG problem, important though that might be. To address how to ameliorate the PG problem would require structurally modelling PG. While the former requires an understanding on the nature of the behavioural deficiencies underlying individual behaviour, the latter would require information on gambling expenditure that our data does contain; together with a sense of the behavioural deficiencies that inhibit restraint in

[14] An exhaustive report on the Australian gambling market by Delfabbro (2010) for the Australian Productivity Commission (APC) reviews the literature on a wide variety of harms associated with PG (ch 3) and comments on the APC's own attempt (Australian Productivity Commission, 2010) to aggregate harms and compare with consumer surplus benefits (ch 6).

[15] Attempts to estimate the effects of PG on the well-being of spouses yielded small and imprecise estimates. Of course, partner separation might be partly driven by PG and the sample of intact partnerships is probably not representative – these are surviving partnerships.

the longer term that would likely require panel data. At present, there are only two panel studies in the world, none for the UK, and, in any event, the two that do exist are far too small to support fixed effect estimation. And even if they were not, differencing self-reported gambling expenditure, income, and PG in panel data is inevitably likely to exacerbate the measurement error problems considerably, even if it were to eliminate simultaneity bias.

However, this chapter may be able to investigate the way in which PG affects W, conditional on PG status. The most obvious contenders as mediating variables are gambling expenditure and gambling losses and Section 2.5.4 attempts to quantify their role, once the magnitude of the PG problem has been established. However, gambling expenditure in the data is heavily underreported, with the exception of scratchcards and lotto. The finding of no effect of *overall* gambling spend on the impact that PG has on well-being is likely to be a manifestation of the measurement error in the data. However, since expenditures on National Lottery products (lotto and scratchcards) are comparatively well measured in the data, one might be prepared to place at least some faith in the estimates that scratchcard spending does have a sizeable and statistically significant mediating role, while spending on lotto does not.

2.2 Related literature

There is a considerable literature on problem gambling. All of the quantitative work uses one or more of a number of screens that consist of a set of questions that are thought to be indicative of PG. An overview of the problem gambling literature is provided by Orford *et al* (2003), which exploits the British Gambling Prevalence Surveys (BGPS) that pre-date the 2010 GPS used in our analysis. Griffiths provides an updated review of the British literature in Griffiths (2014), which includes analysis of the 2010 GPS data used here, as well as providing wider international comparisons. The BGPS is one of a small number of random sample surveys of populations that have been conducted in the world for this purpose – many samples elsewhere are drawn from specific subsets of the population. Indeed, Britain has had three such surveys - although the changes across years have been small and the samples are not large enough to have the power to reject stability of the prevalence of problem gambling across time. For the 2010 dataset used here, Griffiths argues that "…problem gambling in Great Britain is a minority problem that effects less than 1% of the British population…", and that "Problem gambling also appears to be less of a problem than many other potentially addictive behaviours”[16].

Related research does consider the public health consequences of gambling which looks at specific outcomes in a piecemeal fashion. An excellent early overview is by Shaffer and Korn (2002) which candidly confesses that the causal effect of gambling on adverse outcomes such as mental health, crime, domestic violence, etc. cannot be inferred from their correlations, even though some of these correlations are large. Establishing that an effect of gambling on *any* of these outcomes is causal is likely to be problematic. Establishing the causal effect on *all* possible outcomes is likely to be considerably harder. The well-being approach offers a practical and way of condensing the problem into a univariate outcome.

However, none of the extensive PG literature that focusses on measuring the *prevalence* of PG makes any attempt to uncover the *size* of the problem that PG generates for those people who are so afflicted. So, this literature is seriously incomplete, and this work attempts to fill that gap as its primary contribution. Only Forrest (2016) and Farrell (2017), who also use the 2010 BGPS draws any attention to this issue. While Forrest (2016) does not report the effect of problem gambling on well-being he does report that non-problematic gambling *increases* well-being by 0.2 points, on the 0 to 10 scale, perhaps reflecting the consumer surplus from being able to gamble. Attempts to replicate this finding here were, however, unsuccessful using the well-being specification and the same data. Farrell (2017) controls for income in her estimation of the relationship between well-being and PG but her interest is in improving the fit of the relationship between well-being and PG, not providing causal estimates, and she makes no attempt to derive welfare inferences from her analysis.

The precise question asked in the BGPS data was “Taking all things together, on a scale of 1 to 10, how happy would you say you are these days?” Deaton and Stone (2013) refer to measures of well-being such as that in the BGPS as “evaluative” and they report that there is a stable relationship in the literature between such evaluative measures of well-being and log

[16] Here, Griffiths is referring to Sussman *et al* (2011) who surveyed the prevalence of other addictions and found that addictions to alcohol, cigarette smoking, illicit drugs, work, and shopping appear to have a prevalence rate of around 5% to 15% of the population.

income, with a coefficient that is typically around ½ - implying that a 100% increase in income would raise well-being by ½.[17]

Forrest (2016) reports a large difference in mean well-being for PG vs non-PG individuals in the BGPS data[18]. He goes on to investigate the other correlates of well-being in the 2010 GPS data, including income intervals and many other control variables. He reports estimates that suggest that the effect of PG is strongly and significantly negative. While he does not report the implied effect of income because of the binned nature the data, he does demonstrate that the effect of PG on W is comparable with the effect of divorce and widowhood, relative to married. However, it seems likely that several of the control variables that Forrest includes represent "bad controls": variables that are themselves endogenous and whose presence results in biased coefficients of the PG variable and/or those on income intervals[19]. For example, education, employment, self-reported health, and marital status are all arguably endogenous; and they are likely correlated with PG as well as with income and well-being. It is unclear what the direction of bias would be on the PG or income coefficients associated with including such bad controls.

In addition to the bad controls problem, which is addressed here, there are strong grounds for thinking that PG itself is measured with error. It is, after all, self-reported and individuals may wish to conceal their problem from the interviewer if not from themselves, even though the interview is constructed in a such a way that interviewer is not able to see the subjects' responses to the PG screens. Moreover, PG itself, even if it is not subject to measurement error, is likely to be endogenous because both PG and well-being may be correlated with some unobservable factors that are not explicitly included in the modelling - for example, with non-cognitive traits such as self-control. Or, PG might cause low well-being at the same time as low well-being causes PG. Resolving this endogeneity issue is crucial for being able to put a causal interpretation to the estimated relationship between PG and well-being in observational data. A causal estimate of the effect of PG is needed, rather than a simple

[17] Interpreting the coefficient on log-income in the relationship between well-being and log income as the coefficient of a Constant Relative Risk Aversion (CRRA) expected utility function then this coefficient of ½ would be within the ballpark of estimates from other methodologies. For example, Hartley *et al* (2013) estimate the CRRA coefficient using a sample of gameshow players in an environment where players might win a wide range of amounts. Their well-determined estimate of risk aversion would be consistent with a coefficient on log income of 1.

[18] Since 2010 the DSM and PGSI screens used in the BGPS surveys have instead been incorporated in the Health Survey of England (HSE) in 2012, and the Scottish Health Survey (SHS) every year since 2012.

[19] The bad controls problem is discussed in section 3.2.3 of Angrist and Pischke (2009).

correlation, since the objective is to obtain estimates that will help policymakers understand the consequences of a policy-induced change in PG.

Moreover, a common weakness of the existing literature is that income is typically measured with error and this will tend to attenuate the coefficient on income in a well-being equation. Since the money metric associated with the event in question will vary inversely with the estimate of the effect of income on well-being this attenuation in the latter will inflate the money metric. Powdthavee (2010) appears to be the only paper to suggest how important this problem is. He corrects for the measurement error in income, in the relationship between income and well-being in the British Household Panel Study data, using information on whether the interviewer saw the payslip of household members. Using this as an instrumental variable for log income did indeed result in a large and statistically significant increase in the estimated effect of log income.

A further concern relates to the idea of "Rational Addiction" pioneered by Becker and Murphy (1988). They proposed a forward-looking model of addiction where agents respond to expected changes in future prices/costs as well as to current ones, and where current consumption affects the marginal utility of future consumption. If this model were a true description of behaviour, and the well-being measure was an accurate metric of lifecycle well-being, then one would expect to observe *no* well-being effect of PG. PG status would be optimally chosen so that although instantaneous *ex post* well-being of problem gamblers is low relative to non-problem gamblers, PG would not experience *ex ante* regret relative to non-problem gamblers at the onset of their path to addiction.

The RA theory has been widely criticised for not being able to explain the empirical observation of widespread ex-ante expression of regret by addicts. In fact, this is not a valid criticism of the theory – it is quite possible that addicts, once addicted feel *currently* worse off than they would have been had they not decided *originally* that the discounted lifetime benefits exceed the lifetime discounted costs. Moreover, extensions of the original theory, by Orphanides and Zervos (1995) and Gruber and Koszegi (2001), allows for the possibility that individuals experience either imperfect or subjectively biased perception of their odds, or time inconsistency over their potential to become addicted. Individuals, in these extensions to the theory, still optimally make forward looking decisions but are nonetheless allowed to ultimately regret those decisions because they may have overestimated their ability to win, or underestimated either the ease with which they become addicted, or the present value costs of that addiction.

Finally, the strong effect of PG on well-being begs the question of what mediating factors are involved in the underlying transmission mechanism. Most evaluation work focusses on the “total” effect of some “treatment”, rather than on the underlying “channels” that drive the effect. Evaluation work does not usually investigate the possibility that the total effect may be driven by specific channels that relate to “mediating” variables that affect the final outcome. Here, the role of gambling expenditures/losses is, albeit tentatively, explored, distinguishing between expenditure on draw-based lotto games and scratchcard–style games, from other forms of gambling. It is not surprising, in the light of the lacuna around the magnitude of the harm that PG implies, that the literature is again silent on the potential role that mediating factors might play in determining this unknown magnitude. However, if it is possible to establish some mediation effects then one would at least be able to say something about the likely size of the taxes that might be required to ameliorate the extent of self-harm since there is a (very small) literature on gambling price elasticities.[20] Moreover, since the marketplace for gambling products is highly regulated and far from being competitive it is difficult to resist the conclusion that the relatively concentrated nature of supply, and the low marginal cost of the products supplied, would yield price elasticities that are probably close to -1. Given this, there may be grounds for thinking that changes in the structure of gambling taxation might be used to change behaviour so as to reduce harms. Section 2.6 speculates what might be required.

2.3 Data

The primary dataset is the British Gambling Prevalence Survey (BGPS) 2010. An excellent overview of the content and construction of the BGPS is provided by Wardle *et al* (2011). The data surveys over seven thousand households and so is relatively large for its type but it is nonetheless underpowered for its primary purpose of measuring and comparing problem gambling since PG prevalence turns out to be relatively low. Moreover, the response rate is only 65% and clearly non-random since young single men, in particular, are under-represented and it is precisely this demographic that are most likely to be problem gamblers. Sample weights are available but no statistically significant differences are found from using the unweighted data, relative to the weighted data.

[20] See Frontier Economics (2014) for a recent survey.

BGPS contains two PG screens: DSM (the Diagnostic and Statistical Manual of Mental Disorders) IV and PGSI (the Problem Gambling Severity Index)[21]. The DSM IV is an application of the DSM framework – a guide for medical practitioners outlining diagnosis tools to be used in clinical settings for a wide range of mental disorders. Application of the DSM IV screen results in a dichotomous diagnosis for problem gambling, with no consideration for the fact that PG may present with varying degrees of severity[22]. The PGSI screen is a condensed version of the Canadian Problem Gambling Index – a 31-item screen specifically designed for use in population level surveys to elicit a broader measure of gambling problems (Ferris and Wynne, 2001). This is captured in the PGSI's low, medium and high categorisations for problem gambling[23].

Orford *et al* (2010) provides a comparison of the DSM and PGSI screens, basing their discussion of the merits and weaknesses of each screen on their respective psychometric properties. Their study finds that, whilst all PGSI and DSM IV items received higher endorsement from males than females, four of the nine PGSI items suffer from particularly high proportions of male endorsements[24], possibly understating the already small prevalence of problem gambling amongst females. Moreover, their study singles out item 5 – "do you feel you have a gambling problem?" – as a bad question to include. They explain that item 5's relative poor performance is likely because individuals are more reluctant to acknowledge that they have a problem than they are to admit known associated behaviours. Nonetheless, Orford *et al* (2010) find the PGSI outperforms the DSM IV in terms of internal consistency – how well individual items are correlated – and uni-dimensionality – how well the individual items are at measuring the same problem – suggesting the PGSI may be better suited to population surveys like the BGPS 2010 used here.

[21] This chapter assumes, throughout, that all non-gamblers are not problem gamblers so non-gamblers in the data are re-coded from missing to PG=0. This assumption makes no difference to the subsequent econometric results or the welfare inferences made, although the Forrest (2014) specification does find a small positive effect of non-problem gambling on well-being.
[22] DSM V, developed after the GPS 2010 survey was conducted, now also distinguishes between problem (scoring 3-4) and pathological (5+) gambling behaviour. Since well-being seems to be almost constant for scores above 2 this distinction in DSM V seems likely to be relatively unimportant.
[23] PGSI scores of 1-2 are classified as a low risk/level of problems with gambling, 3-7 is classified a being at moderate risk/ level of problem gambling and 8+ denotes high risk/level of problem gambling.
[24] These are items 5, 6, 7, and 8 which can be found in Appendix 2.8.1.

Table 2.1: Summary statistics

Panel (a):	DSM		
PGSI	No problem	Problem gambler	Total
No problem	6474	6	6480
Low risk	330	3	333
Moderate risk	78	13	91
Problem gambler	10	28	38
Total	6892	50	6942

Panel (b):	Non Gamblers	Non-Problem Gamblers	Problem Gamblers	All
Core variables:				
Well-being Score, 1-10	7.868	7.949	6.240	7.916
	(1.9915)	(1.8491)	(2.6462)	(1.8982)
Personal Income, £ pa	14,904.64	15,842.07	17,547.08	15,616.66
	(9,332.521)	(9,065.814)	(10,522.33)	(9,153.726)
Household Income, £ pa	28,781.91	29,842.58	28,020.13	29,560.56
	(16,868.25)	(15,862.56)	(15,037.04)	(16,121.37)
Gambling Spend, £ pm	0	21.979	308.07	18.467
		(76.0745)	(598.6111)	(86.6464)
Female=1	0.579	0.531	0.200	0.541
	(0.4939)	(0.4991)	(0.4041)	(0.4983)
Age, years	51.606	49.508	37.600	49.953
	(18.3987)	(16.9282)	(14.2628)	(17.3456)
Ethnicity:				
White	0.852	0.953	0.820	0.927
	(0.3545)	(0.2107)	(0.3881)	(0.2604)
Mixed Ethnicity	0.014	0.006	.	0.008
	(0.1184)	(0.0799)		(0.0910)
Asian/Asian British	0.082	0.019	0.120	0.036
	(0.2742)	(0.1382)	(0.3283)	(0.1863)
Black/Black British	0.043	0.017	0.060	0.024
	(0.2034)	(0.1276)	(0.2399)	(0.1523)
Chinese/Other	0.008	0.004	.	0.005
	(0.0889)	(0.0638)		(0.0708)
Marital Status:				
Married	0.654	0.670	0.520	0.665
	(0.4759)	(0.4701)	(0.5047)	(0.4720)
Separated/Divorced	0.095	0.098	0.120	0.097
	(0.2932)	(0.2968)	(0.3283)	(0.2961)
Single	0.173	0.163	0.360	0.167
	(0.3782)	(0.3690)	(0.4849)	(0.3726)
Widowed	0.079	0.070	.	0.071
	(0.2690)	(0.2544)		(0.2574)
Observations	1,760	5,132	50	6,941

Notes: Std dev in parentheses. No observations is recorded as "." Personal and household income is reported as fitted values from the interval regression reported in Table A2.8. Gambling spend is the mid-points of the binned data.

The decision to use the DSM screen as the focus in this chapter is based largely on the fact that it has makes a very clear distinction between PG and non-PG – while the PGSI screen distinguishes between degrees of problem gambling. However, all analysis in Section 2.5 has been repeated using PGSI as diagnosis tool (instrumented with DSM score where appropriate) and these results can be found in the chapter appendix. Despite the differences in the performance of the two screens as highlighted in Orford *et al* (2010), there is no substantive difference in the results when using PGSI as the PG measure, with reported CVs being of a similar magnitude.

Panel (a) of Table 2.1 describes the sample sizes for the two screens. It is clear that PG afflicts a small minority of the population: just 0.55% according to PGSI and 0.72% according to DSM. The British adult population (16+ for lotto gambling purposes) is approximately 45 million so these percentages correspond to approximately ¼ and ⅓ million people respectively. The two PG screens overlap but contain different numbers of questions and they are scored differently: PG=1 is defined as a score on DSM>2, or as a score on PGSI>7. More detail on the screens is provided in the appendix, Section 2.8.1. Panel (b) of Table 2.1 gives a breakdown of the data: gamblers who are PG, gamblers who are not PG, and non-gamblers. It is clear that problem gamblers are much more likely to be male, young, single (never married), and Asian or Black (although cell sizes are tiny here). While problem gamblers have 10% higher personal income, they live in households that are 12% poorer – because they are much more likely to be single. Moreover, they spend over *fourteen* times as much as the gambling non-PG group and they experience a level of well-being that is approximately a whole standard deviation lower.

Figure 2.1 very clearly shows the highly left skewed distributions of scores for both screens, and Figure 2.2 shows a scatterplot of the two (where, for clarity, we omit the 93% of individuals who score zero on both screens). Only 28 individuals are classified as PG=1 for both screens, and only 60 individuals are classified as PG=1 for at least one screen. Precision is always an empirical issue, but such is the low proportion of PG in the data that even the BGPS sample is unlikely to be sufficiently large to provide the power to reliably estimate the effects of PG on well-being in the population if they are small. Of course, if the effects were small then there would be no need to worry about them.

A near-unique feature of the 2010 BGPS survey is it contains information, at the individual level, on well-being. This information is not available in almost all of the similar surveys conducted elsewhere in the world. As reported above, well-being is recorded on a 0 to 10 integer scale and the distribution is shown in Figure 2.3 for those who are designated as problem gamblers (PG=1) by the DSM screen, and those not (PG=0). There is more than twice the proportion of PG=1 individuals who have well-being below 7 as there are PG=0 individuals.

The relationship between well-being and the DSM score for defining problem gambling is presented in

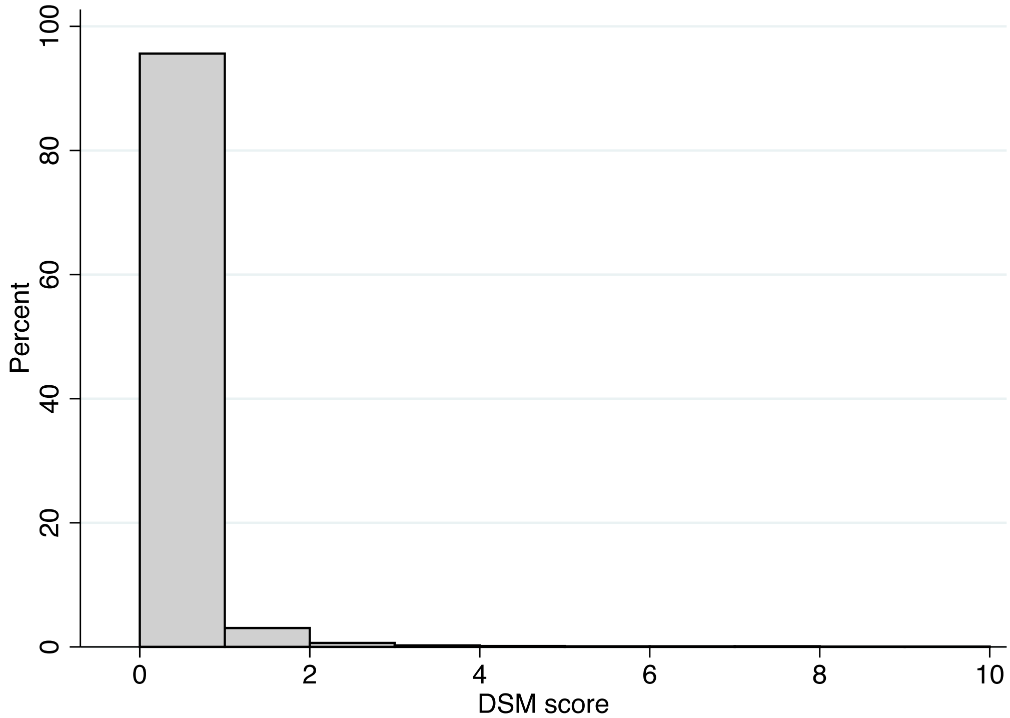
Figure 2.4. This is drawn with "jitter" to show more clearly how the data is distributed across these two integer-valued variables – more observations at intercises are indicated by denser clouds. The solid line shows a local polynomial regression and the grey area is the 95% confidence interval around this. It is clear that a score above 2 in DSM is indeed indicative of lower well-being, but there is no step-change at a score of 3, and the well-being gradient above 3 is not significantly different from zero. However, even scores of 1 or 2 also have a pronounced effect on well-being.[25] The graph suggests that a score of 3 compared to 0 reduces well-being by close to 2 units – approximately a whole standard deviation.

Income in BGPS is recorded at both the individual level, and at the household level, as net annual income and is coded into £5k bins. Since a continuous measure of income is required to be able to implement the well-being method, this chapter eschews the alternative of using a set of dummy variables to indicate which bin each individual belongs to. Moreover, because the income data is likely to be the object of measurement error, estimation of the relationship between (log) net household annual income and a number of explanatory variables is done using an interval regression.[26] This estimation methodology respects the grouped nature of the dependent variable and relies on the assumed log Normality of the income residuals to form a continuous relationship with the explanatory variables.

[25] Dummy variables for DSM=3, 4 ….10 in the well-being equations are jointly insignificantly different from that for DSM=2.

[26] Another alternative is to use the midpoints of the income bins as the measure of income. Doing so makes no substantive difference to the findings – the money metric of lost well-being from problem gambling remains large and significant. However, since measurement error is not accounted for using this approach, the coefficient on household income simply becomes smaller, making estimates of the cost of PG, even larger.

Figure 2.1: Distribution of PG scores: DSM and PGSI



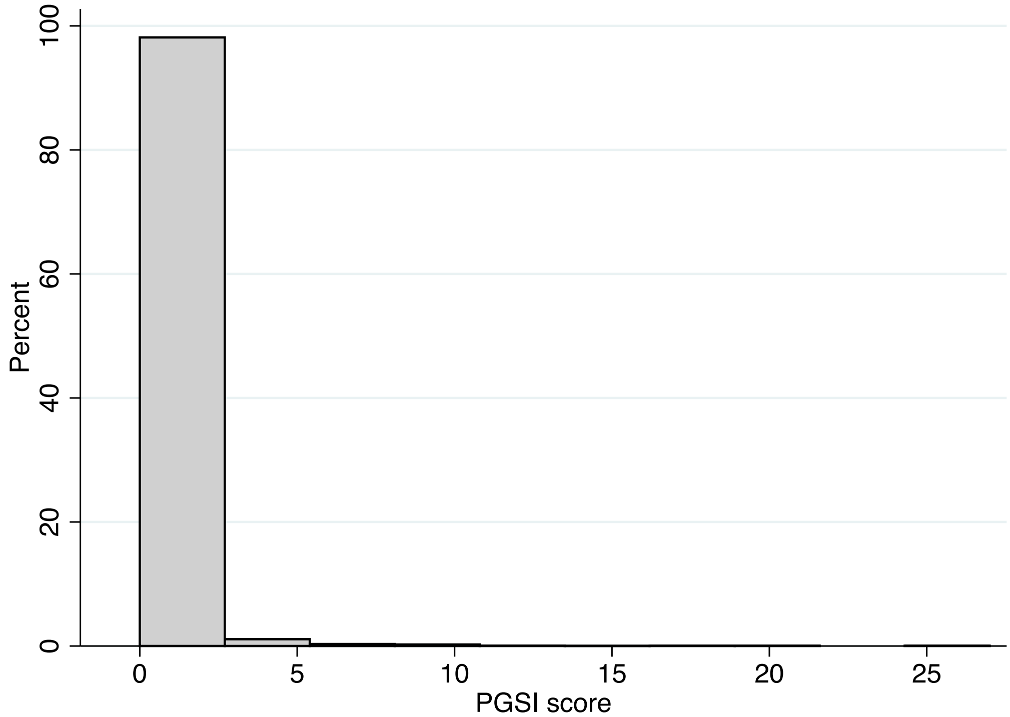
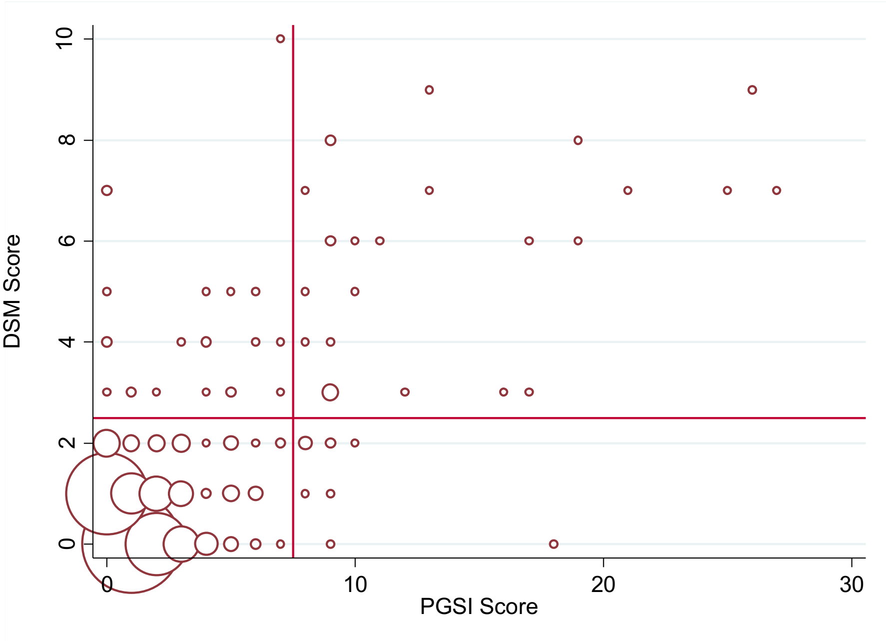
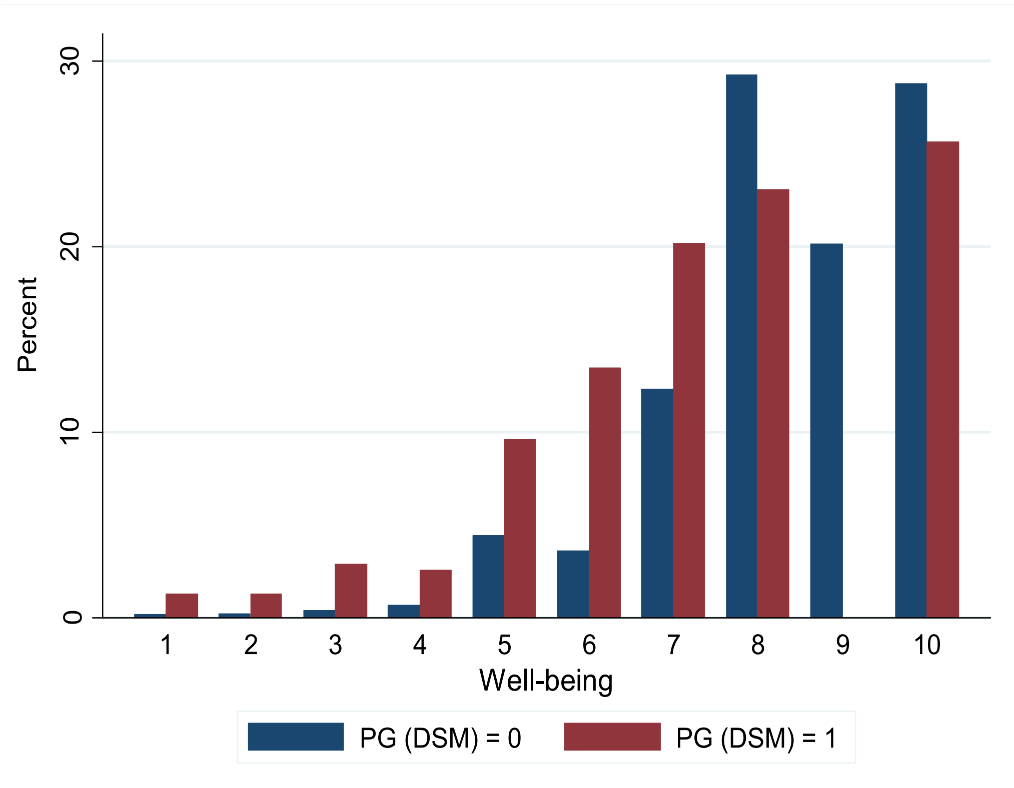


Figure 2.2: Scatterplot of scores using DSM and PGSI screens



Note: PG defined by DSM score >2 or PGSI>7. Bubble size shows proportion of sample.

Figure 2.3: Distribution of well-being in BGPS 2010



Note: No PG=1 observations record a well-being level of 9.

The selection of explanatory variables, and the use of log household income as the measure of income, is deliberately parsimonious and is inspired by the theory of human capital due to Becker (1964) and its empirical implementation in Mincer (1974), and in many hundreds of subsequent analyses of earnings[27]. Thus, the log of the recorded grouped net income is replaced by its prediction from this interval regression. Appendix Table A2.8 presents the empirical estimates of this interval regression. The estimates are very conventional with a large negative effect of female, a strong quadratic effect of age, and a strong positive effect of better educational qualifications and better health. The great virtue of this interval regression is that it simultaneously provides a continuous measure of net income, and it resolves the measurement error problem in self-reported income. Figure 2.5 shows the scatterplot (with "jitter" applied to allow us to see how many individuals are at each integer value in the graph from the density of the clouds) of predicted log net household income against well-being and superimposes a local polynomial of the relationship between these two variables. This relationship is relatively flat but does show a significant monotonically increasing relationship that is concave over most of its range.

Thinking of this well-being vs income relationship as an (expected) utility function then the concave nature would be consistent with diminishing marginal utility of income. While there has been considerable controversy over inferring the marginal utility of income, a cardinal concept, from observed decisions made under uncertainty there is ample evidence in the literature that one can. For example, Hartley *et al* (2013) estimate a model of decision-making behaviour under uncertainty in a gameshow setting, which features a range of stakes from low to very high during the course of play. They find that behaviour *is* consistent with a utility of income relationship that is log linear. In their analysis, using a very different methodology from that used below, the coefficient on log income is found to be, very precisely, 1. Here, Figure 2.5 suggests that doubling income from, say, a log income level of 9.5 to 10.5 would indicate a rise in well-being of around ½ - which is close to what we estimate in the next section using conventional statistical analysis.

[27] The interval regression exploits the detailed information on education included in BGPS, as well as age, age-squared, gender, self-assessed ethnicity, marital status, and an indicator of self-assessed health. Since the purpose in modelling income is to obtain a consistent prediction of the true level of income from the observed interval data there is little concern about the exogeneity of the variables in this auxiliary equation. Variations in the precise specification of this interval regression make very little difference to the subsequent analysis.

Figure 2.4: Relationship between well-being and DSM screen score

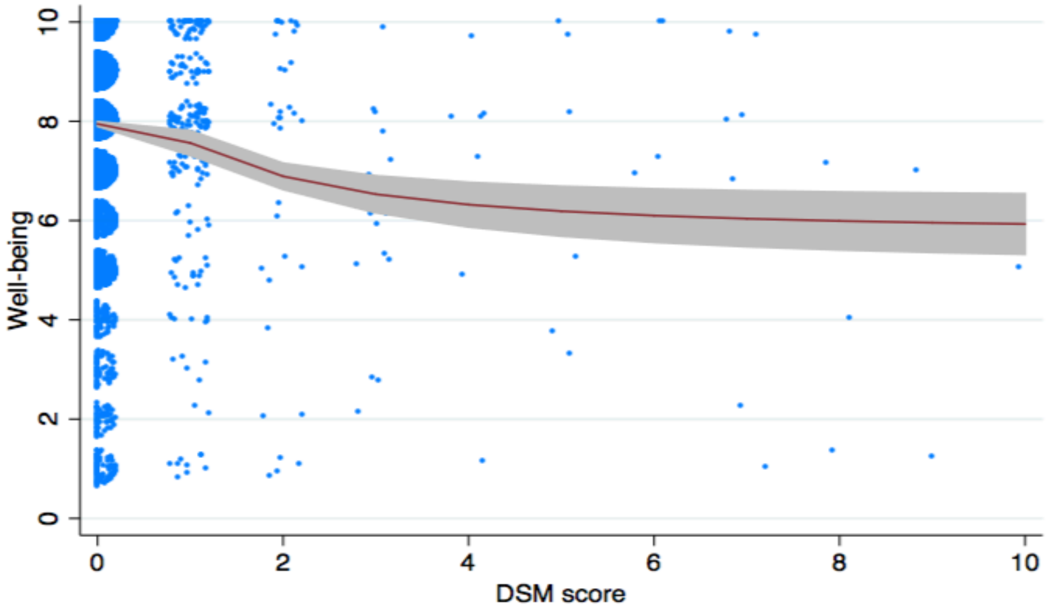
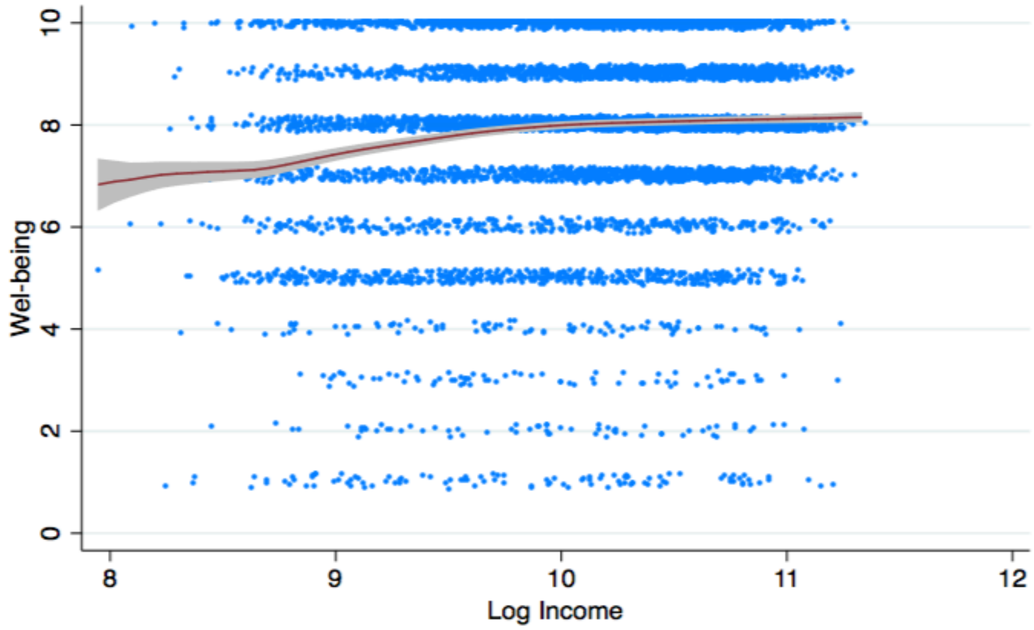
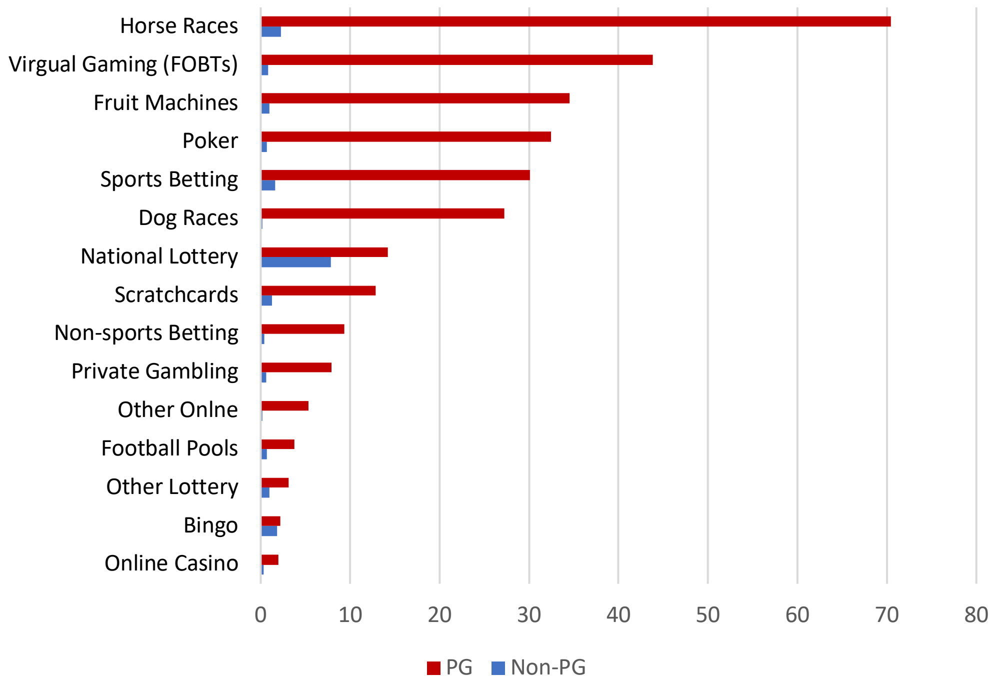


Figure 2.5: Well-being and (predicted) net annual household income



Finally, the gambling expenditure data referred to in Table 2.1 is aggregated up from information about expenditures on each type of gambling. Figure 2.6 compares the mean spend by type of gambling for gamblers who are PG with those who have positive spend on overall gambling but are not PG. It is clear that PGs spend more on all products than non-PGs and this is especially the case for most forms with the exception of the National Lottery draw games and bingo. This is explored in more detail in the mediation analysis in Section 2.5.4.

Figure 2.6: Monthly expenditure (£) on gambling products by gamblers



Note: The raw data is reported in bands and here aggregates imputed expenditure using the mid-points of the reported bins. The red bars are for the PG=1 group and the blue bars are for the PG=0 individuals who report positive levels of overall gambling expenditure.

2.4 Estimating the determinants of well-being

There is a well-established well-being methodology for estimating the consequences of life events and an outline of the method and a seminal application of it can be found in Clark and Oswald (2002). In its simplest incarnation, it starts from the presumption that there is some parametric relationship between individual well-being (W_i), where the i subscript indicates individual i, income (Y_i), the (discrete) event in question, which in this case is PG_i, and a vector of control variables (X_i). This relationship may be summarized by $W_i = W(\mathbf{X}_i, Y_i\ PG_i)$. When taken to the data, this is often specified as W_i being linear in $\mathbf{X}_i$ and PG_i, but often *log* linear in Y_i, and this is typically estimated by OLS. Thus, a typical model is

$$W_i = \boldsymbol{X}_i'\boldsymbol{\beta} + \gamma lnY_i + \delta PG_i + u_i \tag{2.1}$$

where u_i is the residual that captures variation in W_i that is not captured by the included variables. Thus, the difference in well-being associated with $PG_i = 1$ rather than 0 is simply $\Delta W_i = \delta$. Since γ is the effect of a unit change in lnY_i on well-being it follows that the same difference in W could be achieved by changing lnY_i by an amount equal to δ/γ. This implies that δ/γ is the (percentage) change in Y_i required to hold well-being at the level associated with not being a problem gambler for i. This proportionate difference in Y_i is commonly referred to as a measure of compensating variation (CV) – the percentage change in Y_i required to "compensate" i for being PG_i=1.[28]

The decision to model PG as a dummy variable for scoring above some prescribed threshold in a given screen raises questions about the behavioural assumptions underlying the model presented in (2.1). In particular, it is tempting to view the coefficient δ as a measure of changing state from non-PG to PG. This view implies that changing states occurs instantaneously and without the individual anticipating the change which is unsupported at best and patently unrealistic at worst. This interpretation would certainly be valid if it were applied to panel data and especially so if we were measuring changes from PG=0 to PG=1 for individuals between waves. However, in the cross-sectional context of the BGPS, modelling PG in this merely compares the average difference in well-being of those who are PG=1 to PG=0 at a single point in time. It is somewhat irrelevant to the model in this context of whether or not potential PGs anticipate the affliction or how fast shifting from PG=0 to PG=1 occurs. It is indeed likely that becoming PG occurs gradually and can be anticipated, potentially quite far in advance, by the individual. Assuming that this anticipation has a negative impact on well-being implies that the obtained coefficient δ would be biased towards zero, since there are PG=0 respondents in the survey whose well-being is lower in anticipation of becoming PG=1 at a later date. Similarly, any PG=1 respondents who are optimistic about becoming non-PGs in the future would have the same effect of biasing δ towards zero. This adds further weight to this chapter's conclusion that the OLS estimates presented are a lower bound of the cost of PG in the UK.

The proposed methodology is not without its critics. The first criticism stems from the fact that it is not at all clear that the scale of W_i from 1 to 10 can be given a cardinal interpretation. That is, the restriction that moving from W_i=1 to 2 is as good (bad) as a move from 4 to 5 is a strong one, impossible to verify, and hence difficult to rely on in such data. It is just as plausible that the move from 1 to 2, thereby doubling W_i, can only be achieved for

[28] Estimating separate equations for PG=0 and PG=1 would be considerably more flexible but, in the present case, sample size for the PG=1 group precludes this. Interacting PG with lnY would allow the marginal utility of income to differ between the two groups – something that might be expected since this marginal utility of income is related to the degree of risk aversion. However, this interaction proved to be statistically insignificant from zero.

someone who has a W_i of 4 by moving to, say, 8. It seems plausible that a W_i measure could well be monotonically increasing, but non-linear, transformation of some true underlying metric of well-being, so that only the ordinality of measured W_i can be relied upon. To investigate the robustness of the conclusions on the impact of PG, in addition to estimating the conventional model where W is estimated by linear regression, Section 2.5.1 also provides estimates under the assumption that lnW_i is linear in the explanatory variables, and further uses Box-Cox estimation. The latter procedure transforms the dependent variable in a way that nests both linear and log linear and enables testing of these special cases. In particular, the following is also estimated

$$\frac{W_i^{\lambda-1}}{\lambda} = \boldsymbol{X_i'\beta} + \gamma lnY_i + \delta PG_i + u_i. \tag{2.2}$$

Here λ=1 corresponds to the linear special case and λ=0 to the log case. Estimates in Section 2.5.1 show that these alternative specifications make little difference to the PG money metric estimate.

A critical contribution to the well-being methodology by Bond and Lang (2010) raises this ordinality issue. They note that since the W data is categorical, where the categories represent intervals along some continuous distribution, the implied CDFs of these distributions are likely to cross when estimated using large samples. Therefore, some monotonic transformation of the utility function, $F(W_i)$, can always reverse the ranking of overall well-being: for example, between the PG=1 group and the PG =0 group. Of course, more categories will help resolve this problem – the issue would not even arise if W were continuous – but there is nothing to say that 10 categories is enough to ensure the reliability of the method. A popular solution to this problem is to adopt a specification that *only* relies on the ordinal nature of such well-being data. The simplest case is where one is prepared to assume that the unobserved component of well-being is normally distributed so that we can easily fix the cut-points between values of W_i of 1, 2, 3 etc. in which case one can estimate the means and variances of each group using ordered Probit estimation. In particular, the method estimates the determinants of the latent variable W_i^*,

$$W_i^* = \boldsymbol{X_i'\beta} + \gamma lnY_i + \delta PG_i + u_i \tag{2.3}$$

here $W_i = j$ if $\mu_{j-1} < W_i^* < \mu_j$, for j = 1, 2, 3,... 10, u_i is assumed to be Normal, and μ_j are the unknown cut-points that are estimated by exploiting the assumed Normality of u_i. In order to compute the compensating variation in this ordered Probit case requires a transformation of the coefficients into marginal effects, to make them comparable to the OLS coefficients, and then cumulate the predicted probabilities across each of the levels of W, using the proportions reporting each level of W as weights.

The second criticism of the method is a practical one – that, in practice, PG, is measured with error. This will be true, not least, because PG is defined using self-reported responses to the questions in the screen that is employed; it is noticeable in the data that the different screens produce different results (although insubstantially so). OLS estimates of δ, when PG is subject to measurement error (ME), will be attenuated – i.e. biased towards zero. The solution to a ME problem is to instrument with another measure (even one that is also measured with error). When the ME is classical, IV then produces consistent estimates of δ. In practice, GMM estimation may be used to ensure consistency even if ME is non-classical (see Kane *et al*, 1999, and Light and Flores-Lagunes, 2006).

Fortunately, the BGPS data provides not just one screen for PG but two. Thus, it is possible to instrument the PG variable, computed from one screen according to whether the score exceeds the critical value, with the score on the other screen. Indeed, the scores for each question within the alterative screen might be used to form many instruments; although for the moment a parsimonious approach using the overall score from the alternate screen is adopted.

A third criticism of the methodology is that of assuming well-being is linear in the log of income. The relationship between well-being and log household income in Figure 2.5 is, after all, non-linear in places and more or less flat over large sections of the income distribution. Given the role of γ in the denominator of the CV calculation, a small coefficient here, as might be expected from the relationship depicted in Figure 2.5, will inevitably lead to large CV estimates. Nonetheless, experiments with the specification of income in the well-being formula – including linear, quadratic, and cubic expansions of both level and log household income – had no material effect on the resultant CV estimates. Therefore, this chapter proceeds with the above specification in which well-being is linear in the log of household income[29].

An alternative method for modelling the relationship between income and well-being is to suggest that it is *relative*, rather than absolute, income which matters. For example, Clark

[29] Estimates using log personal, rather than household, income were also considered and, whilst smaller, these made little material impact to the CV estimates presented in this chapter. Estimates using polynomial expansions of log household income can be found in the appendix **Error! Reference source n ot found.**.

et al (2007) argue that a strong correlation between *relative* income and happiness can be used to explain the coexistence of the 'Easterlin Paradox' – the common observation that GDP per capita does not correlate with self-reported happiness across countries – and the body of evidence showing that income is a (positive) predictor of subjective well-being in micro-level studies. However, to use relative income requires knowledge of exactly who individuals may be comparing themselves to, information on which is unsurprisingly absent in the data used here. If the whole population is assumed, this would involve subtracting average income in the sample from individual income – but this would have no effect on the income coefficient. One might be tempted to approximate relative income as deviations from the average income for individuals with the same socio-economic characteristics. However, it is impossible to know which of these characteristics form the relevant groups which individuals in the sample compare themselves to. As such, any assumptions made here would be entirely arbitrary and this approach is eschewed in favour of the absolute income approach of equation (2.1).

There remains one further criticism of this well-being methodology: one that is generally ignored in the well-being literature, but was, nonetheless, a question that was raised by Forrest (2104, 2016) in the present context of problem gambling. This criticism is that PG_i is itself endogenous - that is, it is correlated with both u_i and W_i perhaps because there are missing variables that confound the relationship. For example, in this context, individuals with low W_i, for reasons that are not observed and controlled for, may be more likely to have PG_i=1. OLS estimation of δ will then be biased – upwards (downwards) if cov(PG_i,u_i) > 0 (<0) since δ will capture both the effect of PG_i, and the effect of the unobservables that are correlated with both PG_i and W_i. One might expect that low well-being types of people to be *more* likely to develop problem gambling (i.e. cov(PG_i,u_i) < 0) so OLS estimates of the PG coefficient would be biased *upwards*. The solution to this problem is again found through instrumental variables. That is, we need to find some variable, call it Z_i, that affects PG_i but only affects W_i through its effect on PG_i – so there is no direct effect of Z_i on W_i.

Forrest chooses not to pursue this on the grounds that the fact that PG and W are strongly correlated is sufficient to make even the OLS estimates of policy interest. This is a legitimate view – the strong correlation suggests that people who are PG=1 *and* have low W are worthy of the attention of policymakers. This is the '*where there's smoke there's fire*' view that is often adopted in the epidemiology literature and is enshrined in the commonly referenced Hill's Criteria (see Hill, 1965). However, if the policy objective is to raise W, at least for those with low *W*, there are likely to be much better ways of profiling for this than using PG– not least because PG=1 is so scarce that it is likely to miss the overwhelming majority of low *W* cases. Thus, the case for relying on OLS results without further investigation is weak[30]. However, since PG is itself scare it is difficult to find some *Z* that is strongly correlated with it. The candidate for *Z* used here is parental PG - sample members are asked whether one or both parents gambled regularly and, if so, whether they now regard their parent's gambling when they were young as problematic. While this might well be highly correlated with own PG, it is more difficult assess whether one can be confident that parental PG has no direct effect on *W* - that is, can we be sure that the effect of parental PG on W is mediated only through its effect on own current PG? There may, for example, be a case for thinking that individuals with parent(s) who were thought to have been PG might positively affect the PG status of the child through some heritable transmission mechanism.[31]

The definition of PG pays no regard to the financial transactions that underlie PG: that large amounts of money are, on average, lost through heavy gambling expenditure that Table 2.1 showed was many times higher for PG cases than for non-PG cases. While this chapter addresses the effect of PG on well-being, it is silent on how this effect happens. The methodology is not well-adapted to generate policy implications if it cannot tell us what the transmission mechanism through which PG impacts *W*. Therefore, Section 2.5.4 augments this standard well-being method with "mediation analysis". This facilitates the decomposition of the effect of PG on well-being into a direct effect, and an effect that is mediated through the indirect channel of associated gambling expenditures. The mediation approach is to estimate a pair of linear equations – one for the mediator, that depends on the treatment (and covariates) and one for the outcome which depends on the mediator and the treatment (and covariates). Then the "direct effect" is computed as the partial effect of treatment on the outcome (holding the mediator fixed), while the "indirect effect" is the product of the partial effect of treatment on the mediator, and the partial effect of the mediator on the outcome. Assuming a linear specification for both equations, (and no interaction between treatment and mediator) then a numerically equivalent strategy is to add the mediator to the model in (2.1) and measure the extent to which the treatment effect decreases relative to the estimate when the mediator is

[30] Hill's Criteria was appealed to in recent submissions to the Department for Digital, Culture, Media and Sport. See Hofler (2005) for many of the reasons why this is inappropriate.

[31] Studies of twins are commonly used to examine heritability where a within twin pair PG correlation for MZ twins that exceeds that for DZ twins may be indicative of a genetic (or epi-genetic) mechanism at work. Slutske et al (2010) uses a large sample of Australian twins and does find such evidence.

excluded (the "difference in coefficients" method).[32] It was clear from Figure 2.6 that non-problem gamblers spend very little on gambling products apart from lotto and other lottery draw games, and scratchcards, and bingo. Thus, if denoting potential mediators as the vector of gambling expenditures, $\boldsymbol{G}$, then the essence of mediation analysis is to estimate

$$W_i = \boldsymbol{X_i'\beta} + \gamma lnY_i + \delta PG_i + \boldsymbol{G_i'\theta} + u_i. \tag{2.4}$$

Here, the direct effect is the estimate of δ from (2.4) and the indirect effect is found by subtracting this direct effect coefficient from the estimate of the same in (2.1).

2.5 Results

The empirical estimation deliberately adopts a parsimonious specification of the well-being equation for fear that including bad controls may bias the estimates of the coefficients on the variables of interest – problem gambling and log income. Only those variables that one can be reasonably confident are themselves exogenous are included: age, age^2, gender, and indicators for marital status, and ethnicity.[33]

2.5.1 Correlation analysis

Headline estimates of the parameters of interest, using OLS, are presented in Table 2.2, which sets out to investigate the issue raised by Bond and Lang concerning how the ordinal nature of the data might be cardinalised. Table 2.2 compares a conventional specification, where the W_i ranking is used as linear cardinalisation (so that W_i=4 is assumed to mean twice as good as W_i=2, which is twice as good as W_i=1) with a log-linear model, where W_i is replaced by $W_i'=lnW_i$ (so that W_i'=2 is assumed to mean twice as good as W_i'=1, but a value of 3 is twice as good as a value of 2). The second column corresponds to such a log-linear cardinalisation.

The final column is known as the Box-Cox specification that nests the earlier two as specials cases. Note that the specifications differ only in the transformation of the dependent variable and so the *ratio* of any pair of coefficients will have the same interpretation across specifications. Thus, the interpretation of the estimated parameter, δ/γ, remains legitimate across the specifications – and the fact that they are very similar, and certainly not statistically different from each other, suggests that the overall welfare consequences are captured

[32] See the simple exposition in Huber (2016) and references therein.
[33] Experiments that also included or excluded marital status, ethnicity, region, and even education made no substantive difference to the estimates of δ/γ so these are not considered "bad controls". However, the results are sensitive to the inclusion of self-assessed health and is omitted to avoid a bad controls problem.

reasonably well by the simplest specification.[34] It is not very surprising that the ratio of coefficients on two explanatory variables is not greatly affected when the dependent variable alone is subjected to a monotonic transformation. The headline welfare loss calculation multiplies the estimated δ/γ (of -2.5) by average annual household income, of almost £24k, to yield an annual loss for a household of average income of approximately £60k. Since the interpretation of this is the average effect, the aggregate welfare loss can be found by multiplying by the number of people with PG (approximately ⅓m) to yield a figure that is close to £25b.[35]

Table 2.2: OLS estimated parameters of interest across cardinalisations

Dependent variable	W	Ln(W)	$\frac{W_i^{\lambda}-1}{\lambda}$
Ln Y (γ)	0.534***	0.103***	5.275***
	(0.0583)	(0.0120)	(0.6553)
PG (δ)	-1.382***	-0.289***	-13.444***
	(0.3719)	(0.0890)	(3.6509)
δ/γ	-2.589***	-2.811***	-2.549***
	(0.7605)	(0.9338)	(0.7707)
CV (£b, pa)	25.4	27.5	25.0

Notes: Estimated robust standard errors are in parentheses. ***/**/* indicates statistically significant at 1%/5%/10%. Female, age, age^2, marital status, ethnicity, and government office region are included as control variables. Full results can be seen in Appendix Table A2.9. Corresponding results where PG is defined using the PGSI score are provided in Appendix Table A2.13. Table A2.14 replicates this table using the weighted data, but the results are largely unchanged.

The PG losses estimated above rely on an income coefficient that itself relies on the predictions from the interval regression from the binned income data. This should, in principle, resolve any bias from measurement error in the raw data. An alternative way to sidestep this issue is to impose an extraneous estimate of the parameter of interest from the existing literature. Thus, Table 2.2 was re-estimated imposing the coefficient on log income to be either unity, as suggested by the Hartley *et al* (2014) gameshow study, or ½, as suggested by the Deaton and Stone review. The PG coefficient changed little under these constraints and the estimated implied financial losses from PG are slightly smaller in the case of ½ and less so in the case of assuming 1. Appendix Table A2.10 provides estimates using equivalised income

[34] The estimated value of λ is 2.27 with a standard error of 0.040 – which rejects *both* extremes in columns 1 and 2.

[35] When dropping non-gambler observations, the results remain almost identical to those in Table 2.2. When including a dummy variable for non-gamblers in Table 2.2, the coefficient is found to be statistically insignificant and the remaining coefficients do not change.

which adjust for household size, n, whereby income is deflated by $(1+n)^{0.7}$ and shows that the income coefficient fell by around 20% but the PG coefficient remained almost unchanged implying larger CV estimates. Appendix Table A2.11 uses mid-points of the income bins rather than the interval regression predictions from the estimates in Table A2.8. Table A2.12 uses personal income rather than household income. None of these changes make any substantive difference to the conclusion from Table 2.2 – that the CV is very large.

2.5.2 Causal analysis

However, these results are contingent on the exogeneity of PG_i which is relaxed in Table 2.3. The first column corrects for measurement error in PG_i by exploiting the correlation between the DSM definition of PG_i and the score from the alternative screen. The Staiger and Stock (1997) rule of thumb that the first stage F-statistic should exceed 10 is easily satisfied for the results in the first column of Table 2.3, suggesting that the PGSI score is a very strong instrument. The second column also includes parental PG in the instrument set in an attempt to deal with the second potential source of endogeneity. The F-statistic still satisfies the rule of thumb. In each case presented in Table 2.3, the PG_i coefficient is substantially larger than that in Table 2.2. The crucial welfare effect, δ/γ, remains the same irrespective of whether Parental PG is added to the instrument set. However, column 3 which, uses $ParentalPG_i$ alone as an IV, does not produce as large an F-statistic in the first stage as the other cases. Not surprisingly, because it is common for the IV estimate to be even more biased than the OLS estimate in such cases, the estimate of δ becomes much larger because of this relatively weak instrument.

Forrest (2016) is rightly suspicious of the ability of this data to yield a valid instrument for PG. Wardle *et al* (2011) note (in their Table 6.3) that individuals with parents who were problem gamblers were themselves five times more likely to be problem gamblers (as defined by DSM) than those who were not PG. But it is not sufficient that the instrument be correlated with the endogenous variable. It must also be the case that the *only* transmission route by which $ParentalPG_i$ affects W_i is through its effect on PG. In the just identified case, whether or not the instrument has a direct effect on the dependent variable of interest rather than just via the endogenous variable, is not something that can be readily inferred, so the validity of the instrument(s) remains an article of faith. One might argue that $ParentalPG_i$ in the past has an effect on current own well-being apart than through its effect through own PG_i which, if true, would undermine its use as an instrument. This is clearly a very credible criticism. However, the difference in W for those with $ParentalPG = 1$ compared to 0 is statistically insignificant in the raw data.

Table 2.3: IV estimated parameters of interest

Dependent variable: W_i

Instruments:	PGSI score	Parental PG and PGSI score	Parental PG
Ln Y ($\boldsymbol{\gamma}$)	0.529***	0.529***	0.458***
	(0.0589)	(0.0567)	(0.0873)
PG ($\boldsymbol{\delta}$)	-2.482***	-2.498***	-18.072*
	(0.6079)	(0.5972)	(9.5863)
$\boldsymbol{\delta/\gamma}$	-4.695***	-4.724***	-39.462*
	(1.2869)	(1.2512)	(25.0554)
CV (£b, pa)	46.0	46.3	386
First stage F-statistic	4890.97***	2488.75***	28.41***

Notes: Estimated standard errors, obtained from bootstrapping, are in parentheses. ***/**/* indicates statistically significant at 1%/5%/10%. F-statistic is the Stock-Yogo definition – using the Windemeijer definition for multiple instruments in column 2 produces very similar results. Female, age, age^2, marital status, and ethnicity are included as control variables and full results are presented in Appendix Table A2.15. The first stage estimates for this specification, and for the alternative PGSI definition of PG, are presented in Appendix Table A2.16. Corresponding results where PG is defined using the PGSI score and instrumented using the DSM score are provided Appendix Table A2.17.

It has been argued that in the over-identified case, where there already exists one or more valid instruments, it is possible to test validity of a *second* instrument, conditional on the validity of a first instrument, using the Sargan–Hansen test (see Sargan, 1958, and Hansen, 1982). It seems likely that the PGSI score *is* a valid IV for PG, as defined by the DSM screen, since both screens have been designed with the objective of assigning PG status and inspection of the questions in the appendix below suggest a lot of overlap across the two screens. Indeed, the overlap is so great that the conclusions of this chapter do not depend greatly on which screen is used (see Appendix Table A2.17 for PGSI results). Moreover, even though parental PG *is* found to be a statistically valid IV, and so yields a consistent estimate of the effect of own PG on W, there is still the question of how one interprets the resulting estimate. In a model with heterogeneous effects, while OLS estimation yields a biased estimate of the average effect of PG on W, this is not the case with IV. However, while an IV estimate is unbiased, IV does not necessarily yield an estimate of an average effect in the same way as OLS does. In particular, IV estimates are best interpreted as the causal effect of the treatment (in this case, PG) on individuals who are treated *by virtue of the instrument*. This is referred to as a Local Average Treatment Effect (LATE) in the literature. In the present PG case, exogenous variation in PG occurs *only* for the group of individuals who are PG by virtue of parental PG – the *complier* group.

Some applied econometricians have argued that, while the IV analysis here does not necessarily obtain an estimate of the average effect of the treatment in question on a readily identifiable population, it nonetheless estimates something that is still relevant for policy. In contrast, others argue that a LATE estimate is not useful and must be augmented with something else to produce economically meaningful parameters (e.g. a structural econometric model). Unlike the case of a schooling reform instrument used in Harmon and Walker (1995) in the context of the causal effect of education on wages, it is difficult to argue that the adults were so affected by their parents' PG that they became PG themselves, especially because this group is so small. In particular, it is quite conceivable that Parental PG makes *some* people more *likely* to be PG (compliers) through some common environment or even genes, while others are *defiers* – people who observe their parents were PG and were determined not to become like that. Formally, IV LATE estimates are the weighted average of the defier and complier estimates.

In the specification using both PGSI score and parental problem gambling as instruments tests of the validity of the instruments using Hansen's J-test for over-identification fails to reject the null hypothesis that the use of parental PG as a first-stage instrument is valid, conditional on the validity of PGSI. However, the Hansen test (and earlier Sargan test) are not generally applicable in the context of a model where there are heterogeneous effects. Fortunately, it is not necessary to rely on this test because Table 2.3 suggests that nothing much hangs on the case for using Parental PG as an IV –the same results are obtained when Parental PG is omitted and the alternative PGSI score is the *only* instrument. The welfare relevant parameter, δ/γ, is virtually the same in both columns. The suggestion is that it is the measurement error in PG that accounts for much of the bias in the OLS estimate in column 1 of Table 2.2. This is fortunate since it implies that we *can* extrapolate from the IV estimates. Thus, if taking δ/γ to be -4.7 then this implies an average welfare effect of around £110k pppa, which aggregates to approximately £37b pa.[36]

A possible alternative to the instrumental variables procedure for causal identification as described above can be found in Lewbel (2012) who builds upon earlier work by Klein and

[36] Dropping non-gamblers, or including a dummy for them, makes no difference to the results in Table 2.3.

Vella (2006). This method uses a subset of the regressors from the model which are uncorrelated with the covariance of heteroscedastic errors to construct instruments in a two-stage routine to identify coefficients on the endogenous variable(s). Exploiting the fact that these constructed instruments are, by construction, uncorrelated with the error terms, this approach also facilitates the use of the Sargan-Hansen test discussed above when there are not enough instrument candidates to achieve over-identification. As such, a secondary benefit of this approach is to provide evidence in support of otherwise suspect instruments.

However, when the endogenous variable is binary, as in the present case, the utility of using heteroscedasticity per Lewbel (2012) to achieve identification has come under scrutiny (see, for example, Emran, Robano and Smith, 2012). Lewbel (2016) demonstrates that exploiting heteroscedasticity to identify the model is possible when the endogenous regressor is binary but notes that it requires a very strong distribution restriction on the error term. Moreover, the required assumption is not testable and is as much an article of faith as regular IV techniques. As such, even as a means to conduct the aforementioned Sargan-Hansen test of instrument validity will likely yield misleading estimates. Nonetheless, the null hypothesis of this test is again not rejected. The resultant CV estimates, though smaller than those obtained using PGSI and parental PG as instruments alone, are still greater than those obtained via OLS. This provides further evidence towards the conclusion of this chapter that the OLS estimates provide a lower bound on the cost of PG. The results from using Lewbel's identification methodology therefore makes only a marginal contribution to the evidence already presented and estimates can be found in Appendix Table A2.18.

2.5.3 Robustness

Table 2.3 was also re-estimated including exactly the same set of exogenous variables in both first and second stages. The second stage results of the relevant parameters were almost identical to the ones provided above. The sensitivity of the results in column 1 of Table 2.2, to how income is defined was also explored. There are two choices in the BGPS data – individual net income, or household net income. The tables here use (predicted log) household level income. Replacing this by (predicted log) individual level income (again from an interval regression) produces a slightly smaller income effect. However, household income is approximately double the level of individual income and so the corresponding welfare loss measures, using the household definition, are somewhat larger than those found with individual income.[37] For example, using predicted log individual income the aggregate CV for the basic OLS specification is £21b rather than £31b. Using the (log of) the midpoint of the bin which each individual reports (whether at individual or at household level) yields much smaller estimates than when use the predicted log from the interval regression. This is to be expected and the large difference is indicative of considerable measurement error in the raw binned data, as well as the inappropriateness of using the midpoint when the raw income data is highly left skewed. So the estimates reported in Table 2.2 and Table 2.3 of this section are strongly preferred, and the results referred to above are confined to Table A2.11 and Table A2.12 in the appendix.[38]

Finally, Table 2.4 presents Ordered Probit results using both un-instrumented *PG* (defined by the DSM screen) in the first column and PG instrumented with PGSI score and Parental PG in the second column. Ordered Probit treats the dependent variable as ordinal so the transformation of W is no longer relevant to the estimates, apart from the cut-points. However, interpretation is more difficult in the context of Ordered Probit since the marginal effects of LnY and PG rely on the parameters in Table 2.4, and on the cut point estimates which are not reported. For any cardinalisation of W, one would expect PG to lower the probability of high W and increase the probability of low W. Using Ordered Probit estimates to infer the annual CV is slightly more involved.[39] The marginal effects reported in the bottom half of Table 2.4 describe the effects of PG on the probability that W=1, 2…..,10. There are positive significant effects of PG=1 (as opposed to 0) on the probability of having low W, and negative effects on the probability that W is high. Figure 2.7 plots the probability of W = 1, 2, 3…,10,

[37] A minor problem with household income is that there is a coding mistake in the raw data that cannot now be fixed – two different income bins shared the same code on the income show-card. One solution to this problem is to recode the observations that chose *either* of these two codes as missing. Two alternatives would be to check by adding up the (midpoint of the binned) individual level data and replace accordingly; or, similarly, use STATAs missing values routine to construct a replacement exploiting the relationship between household and individual incomes in the data. None of these methods made any effective difference to the estimates.

[38] The pursuit of a matching methodology would be ideal to support the IV estimates. However, since the PG group is so very small it is not likely that one would ever be able to get good matches. Similarly, one might like to pursue the bounding idea in Altonji, Elder and Taber (2006) and its extensions in Oster (2016). However, the fact that the treatment group is so small will inevitably imply that the bounds would be large.

[39] Similar to the OLS and IV cases, CV is calculated by taking the ratio of estimated marginal effects of ln(Y) and PG on W and multiply by the proportion of afflicted individuals, using their respective mean incomes. These must be calculated using the marginal effects and mean income for each value of W to correctly estimate the aggregate cost. The cost estimates provided in
Figure 2.4, of £65b, are the summation of this calculation for W=1,…,10, weighted by the proportion of our sample who report each well-being level.

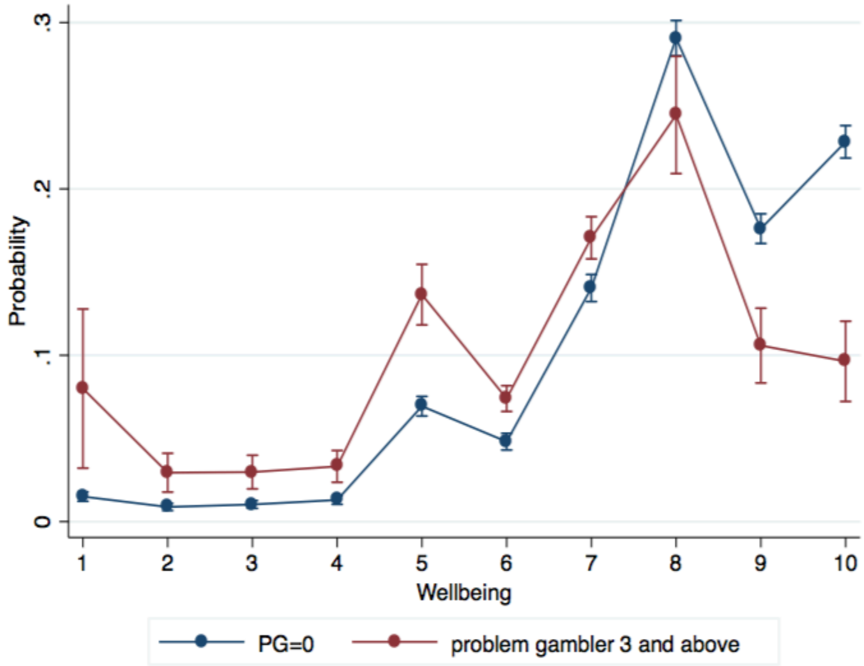
at the averages of the other explanatory variables, for the PG = 1 and 0 groups, using the estimates from the bottom half of column 2 of Table 2.4. The figure also shows the confidence intervals around these predicted probabilities of being at each level of W by PG. The PG = 0 group is much larger so the confidence intervals are much tighter. Nonetheless, the estimates separate the groups very well.

Table 2.4: Ordered probit estimated parameters of interest

Dependent Variable Instrument set	W PGSI	W PGSI and Parental PG
Ln Y ($\boldsymbol{\gamma}$)	0.236***	0.236***
	(0.0313)	(0.0313)
PG ($\boldsymbol{\delta}$)	-1.221***	-1.219***
	(0.2939)	(0.2930)
$\boldsymbol{\delta/\gamma}$	-5.171***	-5.165***
	(1.4538)	(1.4501)
Marginal Effects, $\boldsymbol{\Delta\,\mathrm{Pr}(W=1,\ldots,10)\,/\Delta PG}$		
W=1	0.045***	0.045***
	(0.0116)	(0.0116)
2	0.021***	0.021***
	(0.0057)	(0.0057)
3	0.023***	0.023***
	(0.0061)	(0.0060)
4	0.027***	0.027***
	(0.0069)	(0.0069)
5	0.114***	0.114***
	(0.0275)	(0.0275)
6	0.061***	0.061***
	(0.0149)	(0.0148)
7	0.120***	0.120***
	(0.0291)	(0.0290)
8	0.044***	0.043***
	(0.0115)	(0.0114)
9	-0.100***	-0.099***
	(0.0242)	(0.0241)
10	-0.355***	-0.354***
	(0.0854)	(0.0851)
CV (£b, pa)	50.64	50.59

Notes: Estimated standard errors are in parentheses. ***/**/* indicates significance at 1%/5%/10%. Female, age, age^2, and ethnicity are included as control variables and their coefficients are reported in Appendix Table A2.19. The first stage results from Table A2.16 are re-used here. Corresponding estimates where PG is defined using the PGSI screen, and then instrumented by the DSM score, is provided in the Table A2.20.

Figure 2.7: Predicted probabilities of W=1,2,3,...,10 for PG=0 and PG=1 at means



There are statistically significantly larger probabilities of PG = 1 individuals having values of *W* below 7 compared to PG = 0 individuals; and significantly larger probabilities of PG = 0 individuals having values of *W* above 7 compared to PG=1 individuals[40].

A further issue is the appropriateness of the definition of PG itself. The DSM (and PGSI) score comes from simply adding up the responses to each question, thus attributing them with equal weight in terms of their effect on well-being. A simple alternative would be to allow the data to decide by including controls for each of the 10 (9) questions. Doing this (see Appendix Table A2.21 for DSM and Appendix Table A2.22 for PGSI) suggests that only *one* of the 10 DSM questions has a statistically significant effect on well-being: Q5 which, not surprisingly, asks "Have you gambled to escape from problems or when you are feeling depressed, anxious or bad about yourself?" and a test of the joint insignificance of the remaining questions fails to reject. Q5 of PGSI asks a related but different question: "have you felt that you might have a problem with gambling?" and this question is insignificant when using controls for all PGSI questions. Dropping controls for all the other questions in the screen (i.e. separate models defining PG using Q5 of *either* the PGSI or DSM screen) yields a δ

[40] Extending the IV analysis to the ordered Probit case is not straightforward. Chesher and Smolinski (2012) show that control function methods in this case impose unrealistic restrictions, and are set, rather than point, identified. However, they show that this problem becomes less severe the less discrete the dependent variable is.

coefficient of -2.01*** in the DSM Q5 case and, -0.82 in the PGSI Q5 case when using OLS. Instrumenting PG when defined by these questions using the cross score (without also including Parental PG) gives a δ coefficient of -4.04*** in the DSM Q5 case and, -5.827*** in the PGSI case. It seems these Q5s are quite different. In the DSM case the implication is that PG is very clearly endogenous – PG is caused by other problems. In the PGSI case it seems less clear that Q5 is endogenous.

2.5.4 Mediation analysis

Finally, this section returns to the baseline OLS specification from column 1 of Table 2.2 to consider mediation. The results of this are presented in Table 2.5 where the baseline column is taken from Table 2.2. In specification 1, the aggregate of all gambling expenditure forms the mediator; while specifications 2 and 3 consider expenditure on two of the most popular gambling products – Lotto and scratchcards – individually as mediators. In each case, expenditures are measured in £/month. Only in specification 3 do the results tentatively suggest that expenditure on scratchcards plays a mediating role in the loss in well-being associated with problem gambling. The direct effect of PG on W is £27.8b in aggregate. The indirect effect is only significant at the 10% level, but the magnitude of the effect, -0.172, represents almost 13% of the total effect (δ in column 1). Thus, there is a potential in these results to support policy that decreases scratchcard sales, perhaps through regulatory actions that raise the price of such games relative to others that seem benign, such as lotto. However, we know almost nothing is known about the cross-price demand elasticities between scratchcards and other gambling products. It is therefore unclear whether a reduction in the demand for scratchcards would increase the demand for other gambling, which may not be as benign as the estimates suggest because the coefficients are likely to be heavily attenuated towards zero because of measurement error in expenditure data.

The reservations about the causal interpretation of the PG effect on W applies to this mediation analysis too. In particular, gambling spending and its composition might be a result of low well-being. Given the difficulty in resolving the problem of reverse causation in the earlier work, it is likely to be impossible to resolve this additional concern with the existing data. Moreover, while the BGPS data on monthly lottery expenditure, and on monthly scratchcard expenditure compares closely to the sales reported by the private sector company that is licensed to sell such products, the same is not true for the other components of gambling overall expenditure. It seems that there is considerable under-reporting on other forms of gambling in BGPS relative to what is known from industry sources.

The implication of this is that the coefficient on *other gambling* in Table 2.5 is likely to be considerably biased towards zero. The true value of this coefficient may be substantially larger. But since the scratchcard and lottery spending data appears to be reasonably accurate, at least in aggregate, there is a case for thinking that the statistical significance of scratchcard spending suggests that this product may play a role in the transmission mechanism, while Lotto spending does not.

Table 2.5: Mediation analysis

	Baseline	Mediation 1	Mediation 2	Mediation 3
Ln Y ($\boldsymbol{\gamma}$)	0.534***	0.535***	0.527***	0.525***
	(0.0583)	(0.0524)	(0.0525)	(0.0524)
PG ($\boldsymbol{\delta}$)	-1.382***	-1.415***	-1.254***	-1.210***
	(0.3719)	(0.2726)	(0.2780)	(0.2689)
$\boldsymbol{\delta/\gamma}$	-2.589***	-3.247***	-2.927***	-2.838***
	(0.7605)	(0.9613)	(0.9715)	(0.9410)
All gambling expenditure	-	1.15×10^{-4}	-	-
		(0.0003)		
Other gambling expenditure	-	-	0.0002	-
			(0.0003)	
Lotto expenditure	-	-	0.001	-
			(0.0018)	
Scratchcard expenditure	-	-	-0.016***	-0.015**
			(0.0054)	(0.0067)
Indirect Effect	-	0.033	-0.128	-0.172*
		(0.0644)	(0.1144)	(0.0930)
CV (£b, pa)	25.4	25.9	23.3	22.6

2.6 Comparison with HSE and SHS data

A major concern with the BGPS 2010 data used in the analysis above is the response rate of 65%. Perhaps even more troubling is that single, young males – the most likely demographic to be problem gamblers – appear to be under-represented in the sample according to the summary statistics in Table 2.1. This raises concerns about non-response bias in the preceding estimates and is certainly a weakness of the data as a whole. With the particular dataset used here, one may be suspicious, for instance, about the ability of a survey conducted with the express purpose to evaluate the prevalence of gambling (and problem gambling) to elicit responses from problem gamblers who are concealing their affliction from others or in denial about it with themselves. Moreover, the external validity of the findings above is suspect even if the proportion of problem gamblers is the same for respondents and non-respondents since the distribution of other characteristics – most notably well-being and income – may differ between responding and non-responding problem gamblers.

The fact that using sample weights makes only a negligible difference the preceding estimates only partially relieves concerns about non-response bias since weights only adjust for the distribution of well-known population characteristics – such as age, sex, employment status and income – but not for the unknown distributions of characteristics like problem gambling or well-being. An alternative approach to examining the extent to which non-response bias may be an issue is to find and use other data which is less likely to suffer from the same non-response issues. Fortunately, the Health Survey England 2012 (HSE) and Scottish Health Survey 2012-2015 (SHS) – both of which are large, individual-level, representative sample surveys administered in the UK – contain the DSM and PGSI problem gambling screens used here as well as information on well-being, income and demographic characteristics[41]. Importantly, PG questions feature only very briefly in these surveys and identifying the prevalence of problem gambling is only a minor concern of both the SHS and HSE which are designed to make observations on public health more generally. Thus, non-participants in the HSE and SHS surveys, particularly those who are problem gamblers, are less likely to do so because of the same reasons as those who refuse the BGPS 2010 questionnaire.

The response rates of the HSE and SHS surveys are similar to that of the BGPS, at 65% for the former in 2012 and 59% (2015) to 66% (2012) for the latter. Table 2.6 compares descriptive statistics from the BGPS 2010 data and the pooled HSE 2012 and SHS 2012-2015 datasets. There is little difference in the average age and the proportion of males between the datasets, but the proportion of single (never married) individuals are slightly better represented. The number of problem gamblers in the HSE and SHS data is almost double that of the BGPS data, though the sample size is around three times as large[42]. The proportion of single PGs is 3.4 percentage points higher in the HSE/SHS dataset than in the BGPS and the proportion of

[41] Unfortunately, the HSE and SHS data do not contain information on gambling expenditure or parental gambling behavior. Thus, the BGPS 2010 data is the only dataset available to conduct the entire main analysis of this chapter.
[42] Despite the much larger pool of problem gamblers in the HSE and SHS combined data, the number was still too small for matching methodologies to be properly determined.

Table 2.6: Comparison of BGPS and HSE/SHS descriptive statistics

	BGPS 2010			HSE and SHS		
	Non-PG	PG	All	Non-PG	PG	All
Well-being Score, 1-10	7.928	6.240	7.916	.	.	.
	(1.8865)	(2.6462)	(1.8982)			
Well-being Score, 0-10*	.	.	.	7.791	6.920	7.758
				(1.7994)	(2.3810)	(1.8324)
WEMWBS⁺	.	.	.	50.596	41.650	50.568
				(8.6832)	(10.3619)	(8.7349)
GHQ♦	.	.	.	1.434	3.463	1.439
				(2.6899)	(3.9189)	(2.6987)
Household Income, £ pa	29571.74	28,020.13	29,560.56	37,686.30	29,171.44	36,852.86
	(16,129.46)	(15,037.04)	(16,121.37)	(16,102.89)	(16,151.75)	(16,269.38)
Female=1	0.544	0.200	0.541	0.559	0.172	0.557
	(0.4981)	(0.4041)	(0.4983)	(0.4965)	(0.3791)	(0.4968)
Age, years	50.043	37.600	49.953	50.302	41.273	50.991
	(17.3347)	(14.2628)	(17.3456)	(18.2139)	(16.6647)	(18.5327)
Ethnicity:						
White	0.928	0.820	0.927	0.954	0.848	0.947
	(0.2592)	(0.3881)	(0.2604)	(0.2093)	(0.3604)	(0.2247)
Mixed Ethnicity	0.008	.	0.008	0.004	0.010	0.004
	(0.0914)		(0.0910)	(0.0651)	(0.1005)	(0.0661)
Asian/Asian British	0.035	0.120	0.036	0.025	0.081	0.029
	(0.1848)	(0.3283)	(0.1863)	(0.1556)	(0.2739)	(0.1674)
Black/Black British	0.024	0.060	0.024	0.008	0.040	0.009
	(0.1515)	(0.2399)	(0.1523)	(0.0896)	(0.1979)	(0.0942)
Chinese/Other	0.005	.	0.005	0.008	0.020	0.009
	(0.0711)		(0.0708)	(0.0886)	(0.1414)	(0.0929)
Marital Status:						
Married	0.666	0.520	0.665	0.622	0.464	0.618
	(0.4716)	(0.5047)	(0.4720)	(0.4848)	(0.5013)	(0.4858)
Separated/Divorced	0.097	0.120	0.097	0.104	0.111	0.102
	(0.2959)	(0.3283)	(0.2961)	(0.3047)	(0.3159)	(0.3029)
Single	0.165	0.360	0.167	0.194	0.394	0.192
	(0.3713)	(0.4849)	(0.3726)	(0.3951)	(0.4911)	(0.3939)
Widowed	0.072	.	0.071	0.080	0.030	0.087
	(0.2582)		(0.2574)	(0.2717)	(0.1723)	(0.2823)
Observations	6,891	50	6,941	23,250	99	23,349

Notes: Std dev in parentheses. No observations recorded as "." Personal and household income is reported as fitted values from interval regressions. Gambling spend is the mid-points of the binned data. *Only 16,522 (75 PG) respondents were asked the 0-10 well-being question in the SHS dataset. ⁺Only 20,451 (80 PG) respondents completed the WEMWBS screen. ♦Only 22,878 (95 PG) respondents completed the GHQ screen.

males is 3.8 percentage points higher, though the average age of PGs in the HSE/SHS data is almost 4 years older. Average income is also notably higher in HSE/SHS compared to the BGPS and can be at least partially explained by the time difference between the surveys[43].

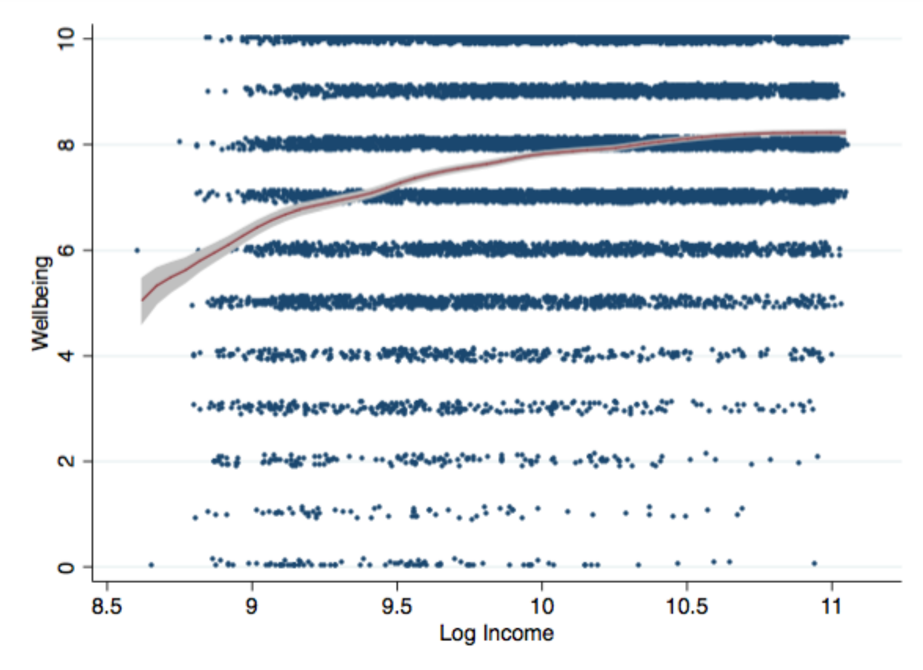
Both the HSE and SHS surveys contain two measures of well-being; the Warwick and Edinburgh Mental Well-being Scale (WEMWBS) and the General Health Questionnaire 12-item screen (GHQ). WEMWBS is a mental well-being screen comprised of 14 positively framed questions about mental health and well-being, each of which receive a scaled score from 1-5 and responses are aggregated to achieve some mental well-being score – where higher scores indicate better well-being. The GHQ is a 12 question, negatively worded screen often used to detect depression. Affirmative answers to each question score 1 point and the summation of these scores indicate mental health – where higher scores indicate worse well-being. The SHS also contains an extra well-being question; "How satisfied are you with life as a whole nowadays?" with responses ranging from 0 (extremely dissatisfied) to 10 (extremely satisfied) and is the closest question to the one asked in BGPS.

In all three surveys, well-being questions are asked at the end of the questionnaire. This raises potential problems for comparing across the surveys since responses may be influenced by the preceding questions when well-being items are asked. That is, there is scope for the different preceding questions to have a material impact on individuals' evaluation of their well-being at that point in time – even if "these days" is included in the question and meant to be evaluative in nature. However, though not identical, the average score from the 0-10 well-being question in SHS is very similar to that from the BGPS for the whole sample, suggesting any difference the preceding questions may have had on well-being is not present. There is, though, a difference in the average well-being score between non-PG and PG individuals across the surveys. It is less than 1 in the SHS data, compared to over 1.5 from the BGPS, which may be explained by the preceding questions causing a non-monotonic change in well-being for respondents, but there is no evidence to support or refute such a conclusion. The coefficient on problem gambling is therefore likely to be correspondingly smaller in a well-being equation when using the SHS and HSE data compared to the estimates above. Moreover, this would yield a smaller estimate of compensating variation associated with being PG, ceteris paribus.

[43] Income in HSE and SHS is again recorded in bins, thus interval regression is employed as with the BGPS data to obtain a continuous measure. The resultant estimates of this interval regression are also in line with the literature.

Whether this follows for the WEMWBS and GHQ measures of well-being is, however, unclear from the raw data.

Figure 2.8: Well-being and net annual household income in the SHS



The remaining component of the compensating variation calculation, the relationship between well-being and income, is illustrated in Figure 2.8 which plots a local polynomial fit of the SHS well-being question (0-10) against fitted log household income and is directly comparable with Figure 2.5. There appears to be a much more pronounced relationship between income and well-being compared to the BGPS, with a unit increase in log income from 9.5 to 10.5 associated with an increase in well-being of around 1 – double what is indicated by Figure 2.5. This implies that one should expect a higher coefficient on log income in well-being regressions using the HSE/SHS data, further reducing the resultant CV estimate. Using the WEMWBS measure of well-being yields a similar picture and the GHQ measure is equally pronounced but with a negative gradient.

Table 2.7 compares estimates of the well-being methodology across the BGPS and the three measures of well-being in the pooled HSE and SHS data. Having established in Section 2.5.1 that monotonic transformations to the well-being measure makes no practical difference to the resultant CV estimates, the well-being measures are normalised by subtracting the mean and dividing by the standard deviation to allow direct comparisons to be made between regressions using the four measures. Note that for all variables, the sign on the coefficients in the GHQ regression are reversed, which is expected since higher GHQ scores correspond to a *lower* mental well-being.

Table 2.7: OLS well-being regression comparison between BGPS and HSE/SHS data

	BGPS	SHS/HSE		
Dependent Variable:	Well-being 1-10	Well-being 0-10	WEMWBS	GHQ 12
Ln(Y) (δ)	0.233***	0.722***	0.772***	-0.640***
	(0.0287)	(0.0273)	(0.0229)	(0.0242)
DSM PG (γ)	-0.742***	-0.187	-0.797***	0.651***
	(0.1928)	(0.1746)	(0.1767)	(0.1729)
Constant	-2.072***	-6.725***	-7.649***	6.432***
	(0.3127)	(0.2904)	(0.2454)	(0.2546)
δ/γ	-3.186***	-0.260	-1.033***	-1.017***
	(0.9285)	(0.2423)	(0.2320)	(0.2746)
Observations	6,942	13,335	17,312	19,711
R-squared	0.065	0.122	0.106	0.076

Robust standard errors in parentheses. Well-being measures all normalised by subtracting mean and dividing by standard deviation. ***/**/* denotes statistical significance at the 1%/5%/10% level. Full estimates including demographic controls are presented in Appendix Table A2.23.

The coefficient on log income is significantly larger in magnitude for all well-being measures in the HSE and SHS data compared to estimates from the BGPS. It is this difference which primarily drives the much smaller δ/γ metric, indicating a lower CV estimate than those presented so far, though the reason for why income correlates with well-being so much better in the SHS and HSE data is unclear. The coefficient on problem gambling when using the 0-10 scale from the SHS is significantly smaller than that from the BGPS. However, using WEMWBS as the relevant measure of well-being in the HSE and SHS data yields a coefficient very similar to that in the BGPS. As a result, the CV estimates from HSE and SHS data are significantly smaller than from BGPS; though, the implied aggregate compensating variation from WEMWBS and GHQ estimates are still large – around £10b per annum. The estimates from the SHS 0-10 measure are smaller still, and are in fact insignificantly different from zero, driven largely by a statistically insignificant coefficient on PG. The poorly determined coefficient on PG in this regression is likely due to the smaller sample of PGs which were asked this question, relative to the other well-being measures in the SHS and HSE data.

The significant difference in estimates between the BGPS used in the main analysis of this chapter and the HSE/SHS data may well be indicative of a non-response bias in the BGPS data which the use of sample weights is unable to correct. For all well-being measures in the HSE/SHS data the resultant compensating variation is significantly smaller, around one third of the estimate from BGPS. This is driven by non-PGs reporting, on average, a larger difference in well-being to their PG counterparts in the BGPS than in the HSE/SHS and by income having a greater influence on reported well-being in the latter dataset. However, the response rate of the HSE and SHS is not dissimilar to that of the BGPS and may well face non-response biases of their own. The difference in the objectives of the SHS and HSE surveys compared to the BGPS make it likely that single, young and male PGs are better represented as is evidenced in the summary statistics – and it is these characteristics that the BGPS appears to under-represent the most.

2.7 Conclusion

Problem gambling is thought to affect a small proportion of the adult population but the contribution of this chapter is to quantify how much of a problem PG is for those who are afflicted. Baseline OLS estimates are in excess of £31 billion pa – around 1.7% of GNP. Further estimates, allowing for measurement error and the endogeneity of PG, suggest this cost to be more than twice as large as the baseline case. Various robustness checks failed to indicate that the overall harm would be less than the headline estimate. Using a nonlinear Ordered Probit model generates even (slightly) higher estimates - although any of these estimates would imply that the PG problem swamps the tax revenues from gambling, by an order of magnitude. The baseline estimates appear to be the lower bound of the likely range of self-harm costs, and even this lower bound suggests that PG would be an enormous social problem. Unfortunately, it is not possible to provide any direct comparisons with the previous literature or with alternative methodologies.[44]

The estimated size of these internality costs suggest that PG may be susceptible to policy that addresses its behavioural origins (see Chetty (2015)), which might be uncovered by testing PG subjects in the laboratory compared to matched non-PG cases. Moreover, given the potential size of this problem it would be important to design the policy response to it by exploring the transmission mechanism by which low well-being develops. Only if the transmission of PG to well-being was mediated via gambling expenditure may it then make

[44] However, Forrest (2016), while not addressing the range of econometric issues that are of concern here, supports this conclusion that the PG problem is a very serious one by showing that his estimates of PG on well-being for men was of a similar order of magnitude as that of being a widower, relative to being married. Indeed, for women, he found that, the effect of PG on W was even larger than that of being a widow on W.

sense to use gambling tax policy to reduce the extent of transmission. Even if this were the case this would not necessarily imply that gambling *should* be made illegal, or even taxed more heavily.

There is the additional consideration of the consumer surplus enjoyed by players. Although the results here suggest that there are significant self-harms associated with DSM scores of 1 or 2 and these people constitute a much larger proportion of the population that the traditional PG (DSM>2) group, the overwhelming majority of the adult gambling population appear to experience little or no problem with this activity. Interpreting the estimates of the loss in well-being associated with PG in the spirit of Orphanides and Zervos (1995), then the average estimate already incorporates the positive well-being enjoyed by those who are lucky enough to gamble without regret.

If one were not prepared to accept this interpretation, then the cross-section work in Farrell and Walker (2000) and the time series work reported Chapter 0 provide estimates of the price elasticity of demand for lotto that would imply that the consumer surplus enjoyed by lotto players is in the order of £1 billion pa. Extrapolating from this to the rest of gambling spending, the aggregate consumer surplus across all forms of gambling would still be very small in comparison with the welfare losses reported here.[45]

Nonetheless, taxing gambling more heavily would only be part of the solution to the PG problem if gambling expenditure was an important mediator for PG and there was an elasticity to exploit by taxation. The lottery draw and scratchcard spending in the BGPS data does suggest a correlation with PG. The data – even with the aforementioned measurement error issues – suggests that PG=1 individuals spend almost two and a half times as much on lottery draw games as do PG=0, and almost fourteen times as much on lottery scratchcards[46].

[45] However, the take-out for non-lottery forms of gambling are subject to take out rates that would typically be less than 10% so this might be underestimating the consumer surplus enjoyed through these other sources of gambling.

[46] GPS contains data on monthly spending, reported in intervals, for every type of gambling for those that say they engage in each type. The lottery draw spending grosses up, using the bin mid-points, to £3.20b, almost exactly matching the official annual sales (of £3.16b), but scratchcard spending underestimates official sales (of £1.34b) by approximately 14%. Respondents are allowed to say that they would prefer not to say how much they spend and this is not an insubstantial proportion of those that report buying scratchcards in the last year. It is likely that these refuseniks are larger than average spenders and this is not captured in the calculations here which are therefore best thought of as a lower bound. There are clear differences in the gambling spending of PG=1 and PG=0 individuals. PG=1 spend £308 per month and lotto is about 5% of this. While PG=0 spend an average of £16 per month and lotto accounts for around 35% of this. Investigating the extent to which expenditure, and its

However, as Forrest (2014) points out, one would need to be able to argue that the elasticity of gambling demand was high, especially for those with PG.[47] The absence of evidence that taxation would reduce expenditure and that expenditure matters causally for PG is an important priority for future work. In any event, given the highly skewed nature of PG, there may be a case for trying to profile problem gamblers and apply treatments that do not rely on financial incentives solely to those that have a high probability of PG, rather than imposing some policy intervention onto the population as a whole, especially when doing so would harm the overwhelming majority of non-PG cases. Sadly, the reduced-form work shows that the few variables that are significantly indicative of PG (age and gender) might not be very helpful in profiling PG since there are many single, young men who are not problem gamblers.

The exploratory mediation results suggest that scratchcards may play a mediating role in the impact of PG on well-being, while other products do not, which suggests a role for policy that generates substitution effects towards more benign products. Regrettably, even less is known about the cross-price elasticities than about own price elasticities.

composition, is an important part of the transmission mechanism that determines PG is a topic for future research if, an only if, data that included PG, Y and W *and* gambling expenditures exists.

[47] Intuitively, one would expect addicted consumers to exhibit less price elastic demand. Taxation, therefore, might have little effect on the behaviour of addicts but nonetheless cause a large deadweight loss on those who are not addicted. While estimates of the average elasticity for various types of gambling do exist (see the report by Frontier Economics, 2014), there appear to none that allow for heterogeneity across the distribution of gambling. See Hollingsworth *et al* (2016) for the case of alcohol demand.

2.8 Appendix

2.8.1 DSM and PGSI screens

In DSM-IV respondents are asked the following 10 questions to determine whether or not they show signs of problem gambling. Respondents are asked how often they exhibit the behavior in each question (with options for never, some of the time, most times, every time). Answers ‘never’ and ‘some of the time’ score 0, whilst ‘most times’ and ‘every time’ are scored as 1. A cumulative score of 3 or more from the following questions indicate a problem gambler. In the past 12 months:

1. How often do you go back another day to win back money you lost?
2. How often have you found yourself thinking about gambling?
3. Have you needed to gamble with more and more money to get the excitement you are looking for?
4. Have you felt restless or irritable when trying to cut down on gambling?
5. Have you gambled to escape from problems or when you are feeling depressed, anxious or bad about yourself?
6. Have you lied to family, or others, to hide the extent of your gambling?
7. Have you made unsuccessful attempts to control, cut back or stop gambling?
8. Have you committed a crime in order to finance gambling or to pay gambling debts?
9. Have you risked or lost an important relationship, job, educational or work opportunity because of gambling?
10. Have you asked others to provide money to help with a desperate financial situation caused by gambling?

In PGSI respondents answer: never, sometimes, most of the time, or almost always (scoring as 0, 1, 2, 3 respectively) to the following 9 questions. In the past 12 months, how often:

1. Have you bet more than you could really afford?
2. Have you needed to gamble with larger amounts of money to get the same excitement?
3. Have you gone back to try to win back the money you’d lost?
4. Have you borrowed money or sold anything to get money to gamble?
5. Have you felt that you might have a problem with gambling?
6. Have you felt that gambling has caused you any health problems, including stress or anxiety?
7. Have people criticized your betting, or told you that you have a gambling problem, whether or not you thought it is true?
8. Have you felt your gambling has caused financial problems for you or your household?
9. Have you felt guilty about the way you gamble or what happens when you gamble?

The PGSI screen differs from DSM in that it attempts to assess the severity of the individual’s problem gambling. A score of 0 indicates non-problem gambling, 1-2 is assessed as a low-level risk of problem gambling, 3-7 would indicate a moderate risk of becoming a problem gambler, and a score of 8+ is typically used to define problem gambling.

2.8.2 Auxiliary estimates

Table A2.8: Interval regression estimates of log household income and log personal income

Dependent variable	Ln(Household Income)	Ln(Personal Income)
Age	0.0210*** (0.00422)	0.035*** (0.0038)
Age^2	-2.01e-04*** (4.40e-05)	-3.02e-04*** (3.92e-05)
Female	-0.302*** (0.0974)	-2.837*** (0.0956)
Education Qualifications: NQF 1	0.474*** (0.0339)	0.396*** (0.0307)
NQF 2	0.309*** (0.0387)	0.237*** (0.0351)
NQF 3	0.690*** (0.0379)	0.606*** (0.0339)
NQF 5	0.472*** (0.0433)	0.472*** (0.0390)
NQF 6	0.650*** (0.0553)	0.649*** (0.0505)
NQF 7	0.734*** (0.0665)	0.819*** (0.0604)
Mixed Ethnicity	-0.394*** (0.122)	-0.142 (0.1095)
Asian/Asian British	-0.386*** (0.0589)	-0.214*** (0.0535)
Black/Black British	-0.595*** (0.0773)	-0.163** (0.0688)
Chinese/Other	-0.174 (0.137)	-0.394*** (0.1290)
Unemployed	-0.957*** (0.0618)	-1.353*** (0.0576)
Long-term Disability	-0.932*** (0.0634)	-0.895*** (0.0587)
Family Carer	-0.490*** (0.0405)	-0.999*** (0.0368)
Retired	-0.475*** (0.0425)	-0.587*** (0.0381)
Good Health	-0.129*** (0.0254)	-0.067*** (0.0228)
Fair Health	-0.246*** (0.0338)	-0.165*** (0.0228)
Bad Health	-0.246***(0.0589)	-0.197*** (0.0301)
Very Bad Health	-0.209* (0.112)	-0.105 (0.1031)
Separated/Divorced	-0.395*** (0.0599)	0.293*** (0.0406)
Single	-0.276*** (0.0441)	0.232*** (0.0676)
Widowed	-0.287*** (0.0872)	0.321*** (0.0910)
Constant	9.979*** (0.153)	8.818*** (0.1385)
Observations	4,540	5,667

Notes: Omitted categories: Male, education level 0 (NQF0=1), white, paid work, very good health, married. ***/**/* indicates statistically significant at 1%/5%/10%. Government office region and marital*gender interaction omitted from reporting.

Table A2.9: Full OLS estimated coefficients (PG defined by DSM>2)

Dependent Variable	W	Ln(W)	$\frac{W_i^{\lambda}-1}{\lambda}$
Ln Y (γ)	0.534***	0.103***	5.275***
	(0.0583)	(0.0120)	(0.6553)
PG (δ)	-1.382***	-0.289***	-13.444***
	(0.3719)	(0.0890)	(3.6509)
Age	-0.040***	-0.006***	-0.507***
	(0.0086)	(0.0017)	(0.0993)
Age^2	0.001***	0.000***	0.007***
	(0.0001)	(0.0000)	(0.0010)
Female	0.230***	0.038***	2.697***
	(0.0454)	(0.0088)	(0.5352)
Mixed Ethnicity	-0.611**	-0.105*	-7.622***
	(0.2630)	(0.0620)	(2.7549)
Asian/Asian British	-0.099	-0.019	-0.839
	(0.1379)	(0.0271)	(1.6039)
Black/Black British	0.227	0.038	2.476
	(0.1718)	(0.0361)	(1.9657)
Chinese/Other	-0.020	0.021	-1.303
	(0.2760)	(0.0409)	(3.4672)
Married	0.358***	0.044***	5.208***
	(0.0790)	(0.0157)	(0.8956)
Separated/Divorced	-0.022	-0.000	-0.671
	(0.1023)	(0.0207)	(1.1498)
Widowed	-0.547***	-0.085***	-7.016***
	(0.1294)	(0.0265)	(1.4647)
Constant	2.755***	1.016***	1.261
	(0.5894)	(0.1208)	(6.6333)
$\boldsymbol{\delta/\gamma}$	-2.589***	-2.811***	-2.549***
	(0.7605)	(0.9338)	(0.7707)
CV (£b, pa)	25.4	27.5	25.0
Observations	6,942	6,942	6,942
R-squared	0.071	0.054	0.073

Notes: Estimated standard errors are in parentheses. ***/**/* indicates statistically significant at 1%/5%/10%. Omitted categories: Male, white, single, North East. Government office region omitted from reporting. CV is computed by multiplying the estimated δ/γ by 0.0072*46m*$\bar{Y}$: where $\bar{Y}$ = £29,560; 0.0072 is the proportion with PG=1, 46m is the adult population, and CV is recorded in £ billion pa. The estimated value of λ in the final column is 2.27 and this rejects λ=1 (W) and λ=0 (Ln W).

Table A2.10: OLS estimated coefficients using log equivalised income

Dependent Variable	W	Ln(W)	Box-Cox
Ln (Equivalised Income)	0.445***	0.084***	4.510***
	(0.0519)	(0.0105)	(0.5911)
DSM PG	-1.413***	-0.295***	-13.849***
	(0.3738)	(0.0897)	(3.6871)
Age	-0.041***	-0.006***	-0.520***
	(0.0086)	(0.0017)	(0.1004)
Age Squared	0.001***	0.000***	0.007***
	(0.0001)	(0.0000)	(0.0010)
Female	0.225***	0.037***	2.685***
	(0.0454)	(0.0088)	(0.5401)
Mixed Ethnicity	-0.644**	-0.112*	-7.972***
	(0.2640)	(0.0620)	(2.7904)
Asian/Asian British	-0.112	-0.022	-0.926
	(0.1382)	(0.0273)	(1.6155)
Black/Black British	0.176	0.028	2.050
	(0.1706)	(0.0358)	(1.9681)
Chinese/Other	-0.005	0.024	-1.165
	(0.2764)	(0.0409)	(3.4980)
Married	0.527***	0.077***	6.905***
	(0.0739)	(0.0148)	(0.8390)
Separated/Divorced	-0.034	-0.003	-0.779
	(0.1027)	(0.0208)	(1.1626)
Widowed	-0.533***	-0.082***	-6.936***
	(0.1297)	(0.0266)	(1.4792)
Constant	4.033***	1.272***	13.255**
	(0.4888)	(0.0989)	(5.5736)
Observations	6,942	6,942	6,942
R-squared	0.069	0.051	0.072

Notes: Estimated standard errors are in parentheses. ***/**/* indicates statistically significant at 1%/5%/10%. Omitted categories: Male, white, single, North East. Government office region omitted from reporting. Equivalised income calculated as the fitted household income obtained from Table A2.8 divided by $(1+n)^{0.7}$, where n is the number of individuals in the household.

Table A2.11: OLS regressions using midpoints of household and personal income

Income measure:	Household Income	Personal Income
Ln(Midpoint Y)	0.311***	0.127***
	(0.0442)	(0.0275)
DSM PG	-1.618***	-1.463***
	(0.4438)	(0.3972)
Age	-0.041***	-0.037***
	(0.0109)	(0.0093)
Age^2	0.001***	0.000***
	(0.0001)	(0.0001)
Female	0.223***	0.231***
	(0.0551)	(0.0514)
Mixed Ethnicity	-0.809***	-0.766***
	(0.3122)	(0.2846)
Asian/Asian British	-0.299*	-0.359**
	(0.1615)	(0.1494)
Black/Black British	0.113	-0.075
	(0.2136)	(0.1792)
Chinese/Other	0.165	-0.022
	(0.2938)	(0.2869)
Married	0.536***	0.644***
	(0.0877)	(0.0757)
Separated/Divorced	-0.041	-0.101
	(0.1262)	(0.1096)
Widowed	-0.596***	-0.666***
	(0.1732)	(0.1441)
Constant	4.914***	6.692***
	(0.5019)	(0.3373)
CV (£b, pa)	51.0	59.5
Observations	4,540	5,957
R-squared	0.073	0.062

Notes: Robust standard errors are in parentheses. ***/**/* indicates statistically significant at 1%/5%/10%. Omitted categories: Male, white, single, North East. Government office region omitted from reporting. CV is computed by multiplying the estimated δ/γ by 0.0072*46m*$\bar{Y}$: where $\bar{Y} \approx$ £29,560 for household income and $\bar{Y} \approx$ £15,500 for personal income; 0.0072 is the proportion with PG=1; 46m is the adult population: and CV is recorded in £ billion pa.

Table A2.12: OLS estimated coefficients using personal income and PG defined as DSM>2

Dependent Variable	W	Ln(W)	Box-Cox
Ln(Personal Income)	0.356***	0.072***	3.306***
	(0.0482)	(0.0099)	(0.5474)
DSM PG	-1.380***	-0.288***	-13.573***
	(0.3693)	(0.0887)	(3.6577)
Age	-0.042***	-0.007***	-0.519***
	(0.0087)	(0.0017)	(0.1016)
Age^2	0.001***	0.000***	0.007***
	(0.0001)	(0.0000)	(0.0010)
Female	0.375***	0.068***	4.021***
	(0.0543)	(0.0108)	(0.6314)
Mixed Ethnicity	-0.780***	-0.137**	-9.432***
	(0.2654)	(0.0628)	(2.7899)
Asian/Asian British	-0.232*	-0.043	-2.244
	(0.1373)	(0.0269)	(1.6109)
Black/Black British	-0.033	-0.011	-0.175
	(0.1690)	(0.0357)	(1.9499)
Chinese/Other	0.038	0.033	-0.783
	(0.2803)	(0.0417)	(3.5399)
Married	0.605***	0.090***	7.764***
	(0.0725)	(0.0145)	(0.8256)
Separated/Divorced	-0.176*	-0.031	-2.146*
	(0.1042)	(0.0212)	(1.1775)
Widowed	-0.667***	-0.109***	-8.188***
	(0.1310)	(0.0269)	(1.4950)
Constant	4.705***	1.362***	22.741***
	(0.4746)	(0.0977)	(5.3821)
δ/γ	-3.876***	-4.002***	-4.106***
	(1.1842)	(1.3666)	(1.3271)
CV (£b, pa)	20.1	20.7	21.2
Observations	6,942	6,942	6,942
R-squared	0.066	0.049	0.069

Notes: Estimated standard errors are in parentheses. ***/**/* indicates statistically significant at 1%/5%/10%. Omitted categories: Male, white, single, North East. Government office region omitted from reporting. CV is computed by multiplying the estimated δ/γ by 0.0072*46m*$\bar{Y}$: where $\bar{Y} \approx$ £15,500; 0.0072 is the proportion with PG=1; 46m is the adult population: and CV is recorded in £ billion pa. The estimated value of λ in the final column is 2.27 and this rejects λ=1 (W) and λ=0 (Ln W).

Table A2.13: OLS estimated coefficients (PG defined by PGSI>7)

Dependent variable	W	Ln(W)	$\frac{W_i^{\lambda}-1}{\lambda}$
Ln Y (γ)	0.533***	0.102***	5.281***
	(0.0582)	(0.0120)	(0.6574)
PG (δ)	-1.305***	-0.293***	-12.417***
	(0.4342)	(0.1090)	(4.1227)
Age	-0.039***	-0.006***	-0.500***
	(0.0086)	(0.0017)	(0.0997)
Age^2	0.001***	0.000***	0.007***
	(0.0001)	(0.0000)	(0.0010)
Female	0.235***	0.039***	2.758***
	(0.0454)	(0.0088)	(0.5364)
Mixed Ethnicity	-0.609**	-0.105*	-7.623***
	(0.2632)	(0.0621)	(2.7655)
Asian/Asian British	-0.106	-0.020	-0.913
	(0.1384)	(0.0271)	(1.6146)
Black/Black British	0.229	0.039	2.498
	(0.1717)	(0.0361)	(1.9708)
Chinese/Other	-0.015	0.022	-1.253
	(0.2759)	(0.0408)	(3.4774)
Married	0.356***	0.043***	5.208***
	(0.0791)	(0.0157)	(0.8994)
Separated/Divorced	-0.031	-0.002	-0.755
	(0.1024)	(0.0208)	(1.1553)
Widowed	-0.549***	-0.085***	-7.056***
	(0.1294)	(0.0265)	(1.4702)
Constant	2.747***	1.016***	1.154
	(0.5877)	(0.1201)	(6.6493)
δ/γ	-2.450***	-2.860**	-2.351***
	(0.8625)	(1.1151)	(0.8403)
CV (£b, pa)	18.2	21.3	17.5
Observations	6,942	6,942	6,942
R-squared	0.070	0.053	0.073

Notes: Estimated standard errors are in parentheses. ***/**/* indicates statistically significant at 1%/5%/10%. Omitted categories: Male, white, single, North East. Government office region omitted from reporting. CV is computed by multiplying the estimated δ/γ by 0.0055*46m*$\bar{Y}$: where $\bar{Y} \approx$ £29,560; 0.0055 is the proportion with PG=1; 46m is the adult population: and CV is recorded in £ billion pa. The estimated value of λ in the final column is 2.27 and this rejects λ=1 (W) and λ=0 (Ln W).

Table A2.14: Level and log well-being regression using sample weights and PG defined as DSM>2

Dependent Variable	W	Ln(W)
Ln(Household Income)	0.529***	0.100***
	(0.0597)	(0.0121)
DSM PG	-1.438***	-0.296***
	(0.3704)	(0.0886)
Age	-0.037***	-0.005***
	(0.0088)	(0.0017)
Age^2	0.001***	0.000***
	(0.0001)	(0.0000)
Female	0.232***	0.038***
	(0.0467)	(0.0090)
Mixed Ethnicity	-0.470*	-0.073
	(0.2483)	(0.0522)
Asian/British Asian	-0.061	-0.015
	(0.1410)	(0.0284)
Black/Black British	0.130	0.025
	(0.1822)	(0.0350)
Chinese/Other	-0.314	-0.023
	(0.3592)	(0.0543)
Married	0.322***	0.036**
	(0.0810)	(0.0159)
Separated/Divorced	-0.043	-0.007
	(0.1039)	(0.0207)
Widowed	-0.546***	-0.086***
	(0.1325)	(0.0262)
Constant	2.789***	1.042***
	(0.5994)	(0.1212)
CV (£b, pa)	26.6	29.1
Observations	6,942	6,942
R-squared	0.067	0.050

Notes: Bootstrapped standard errors are in parentheses. ***/**/* indicates statistically significant at 1%/5%/10%. Omitted categories: Male, white, single, North East. Government office region omitted from reporting. CV is computed by multiplying the estimated δ/γ by 0.0055*46m*$\bar{Y}$: where $\bar{Y} \approx$ £29,560; 0.0055 is the proportion with PG=1; 46m is the adult population: and CV is recorded in £ billion pa.

Table A2.15: Full IV estimated coefficients (PG defined by DSM>2)

Instrument:	PGSI Score	PGSI Score and Parental PG	Parental PG
Dependent variable	W	W	W
Ln Y (γ)	0.529***	0.529***	0.458***
	(0.0589)	(0.0567)	(0.0873)
PG (δ)	-2.482***	-2.498***	-18.072*
	(0.6079)	(0.5972)	(9.5863)
Age	-0.041***	-0.041***	-0.051***
	(0.0090)	(0.0086)	(0.0124)
Age^2	0.001***	0.001***	0.001***
	(0.0001)	(0.0001)	(0.0001)
Female	0.218***	0.218***	0.054
	(0.0453)	(0.0460)	(0.1159)
Mixed Ethnicity	-0.624**	-0.624**	-0.798***
	(0.2642)	(0.2701)	(0.2801)
Asian/Asian British	-0.088	-0.088	0.068
	(0.1403)	(0.1395)	(0.2070)
Black/Black British	0.231	0.231	0.284
	(0.1694)	(0.1595)	(0.2573)
Chinese/Other	-0.034	-0.034	-0.224
	(0.2678)	(0.2795)	(0.3006)
Married	0.357***	0.357***	0.329***
	(0.0787)	(0.0758)	(0.1042)
Separated/Divorced	-0.022	-0.022	-0.010
	(0.0939)	(0.1025)	(0.1363)
Widowed	-0.551***	-0.551***	-0.611***
	(0.1349)	(0.1263)	(0.1401)
Constant	2.841***	2.842***	4.022***
	(0.6127)	(0.5623)	(1.0491)
δ/γ	-4.695***	-4.724***	-39.462*
	(1.2869)	(1.2512)	(25.0554)
CV (£b, pa)	46.0	46.3	386
1st stage F	4890.97***	2488.75***	28.41***
Observations	6,942	6,942	6,942
R-squared	0.066	0.066	0.066

Notes: Estimated standard errors, obtained from bootsrapping, are in parentheses.. ***/**/* indicates statistically significant at 1%/5%/10%. CV is computed by multiplying the estimated δ/γ by 0.072 * 46m * $\bar{Y}$, where $\bar{Y}$ = £29,560, 0.072 is the proportion with PG=1, 46m is the adult population, and CV is recorded in £ billion pa. Omitted categories: Male, white, single, North East. Government office region omitted from reporting.

Table A2.16: First-stage coefficient estimates of PG (opposite score and parental gambling)

Dependent Variable	DSM>2	DSM>2	DSM>2	PGSI>7	PGSI>7	PGSI>7
PGSI Score	0.048***		0.048***			
	(0.0039)		(0.0041)			
DSM Score				0.092***		0.092***
				(0.0107)		(0.0105)
Parents did not		-0.001	0.001		-0.002	0.001
gamble		(0.0022)	(0.0019)		(0.0020)	(0.0018)
Parents were PG		0.031**	0.014*		0.024**	0.007
		(0.0127)	(0.0078)		(0.0108)	(0.0066)
Age	-0.000	-0.001*	-0.000	0.000*	-0.000	0.000*
	(0.0003)	(0.0004)	(0.0003)	(0.0003)	(0.0003)	(0.0003)
Age2	0.000	0.000	0.000	-0.000	0.000	-0.000
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Female	0.000	-0.009***	0.000	-0.001	-0.006***	-0.001
	(0.0016)	(0.0020)	(0.0015)	(0.0014)	(0.0018)	(0.0015)
Mixed Ethnicity	-0.011	-0.010***	-0.011	-0.006	-0.008***	-0.006
	(0.0071)	(0.0032)	(0.0071)	(0.0044)	(0.0029)	(0.0043)
Asian/British Asian	0.005	0.011	0.004	-0.001	0.007	-0.002
	(0.0065)	(0.0095)	(0.0065)	(0.0046)	(0.0077)	(0.0047)
Black/British Black	-0.001	0.006	-0.000	0.003	0.008	0.003
	(0.0066)	(0.0111)	(0.0068)	(0.0066)	(0.0108)	(0.0069)
Chinese/Other	-0.010**	-0.013***	-0.010**	-0.006	-0.008***	-0.006
	(0.0049)	(0.0036)	(0.0048)	(0.0044)	(0.0028)	(0.0047)
Married	0.000	-0.003	0.001	-0.005*	-0.005	-0.005*
	(0.0031)	(0.0042)	(0.0030)	(0.0029)	(0.0041)	(0.0028)
Separated/Divorced	0.004	0.001	0.004	-0.005	-0.005	-0.005
	(0.0039)	(0.0059)	(0.0040)	(0.0035)	(0.0049)	(0.0034)
Widowed	-0.001	-0.003	-0.001	-0.006**	-0.005	-0.006**
	(0.0029)	(0.0039)	(0.0028)	(0.0029)	(0.0037)	(0.0028)
NQF1	-0.001	-0.003	-0.001	0.004**	0.002	0.004*
	(0.0025)	(0.0031)	(0.0025)	(0.0021)	(0.0026)	(0.0023)
NQF2	-0.002	-0.002	-0.002	0.002	0.002	0.003
	(0.0030)	(0.0036)	(0.0030)	(0.0025)	(0.0031)	(0.0026)
NQF3	0.002	-0.001	0.002	0.003	0.001	0.003
	(0.0031)	(0.0037)	(0.0030)	(0.0025)	(0.0029)	(0.0024)
NQF5	0.003	0.000	0.003	0.003	0.003	0.003
	(0.0030)	(0.0038)	(0.0030)	(0.0027)	(0.0033)	(0.0027)
NQF6	-0.000	0.002	-0.000	0.001	0.003	0.001
	(0.0042)	(0.0058)	(0.0041)	(0.0039)	(0.0051)	(0.0038)
NQF7	0.002	0.002	0.002	0.001	0.004	0.001
	(0.0055)	(0.0071)	(0.0055)	(0.0041)	(0.0053)	(0.0038)
Unemployed	0.003	0.017	0.003	0.008	0.021*	0.008
	(0.0078)	(0.0119)	(0.0086)	(0.0086)	(0.0123)	(0.0086)
Long-term disability	0.005	0.002	0.005	-0.008**	-0.007	-0.008*
	(0.0050)	(0.0072)	(0.0048)	(0.0039)	(0.0055)	(0.0041)

Contd.

Table A2.14 Contd.: First-stage coefficient estimates of PG (opposite score and parental gambling)

Dependent Variable	DSM>2	DSM>2	DSM>2	PGSI>7	PGSI>7	PGSI>7
Caring for family	0.001	-0.002	0.001	0.000	-0.000	-0.000
	(0.0023)	(0.0029)	(0.0024)	(0.0019)	(0.0025)	(0.0019)
Retired	0.001	-0.001	0.001	-0.003	-0.002	-0.003
	(0.0020)	(0.0025)	(0.0020)	(0.0017)	(0.0020)	(0.0017)
Good health	0.002	0.006**	0.002	0.001	0.004**	0.001
	(0.0019)	(0.0024)	(0.0019)	(0.0014)	(0.0019)	(0.0014)
Fair health	-0.002	0.002	-0.002	0.005**	0.007**	0.005**
	(0.0020)	(0.0027)	(0.0019)	(0.0021)	(0.0027)	(0.0023)
Bad health	-0.001	0.008	-0.002	0.013***	0.016**	0.012***
	(0.0036)	(0.0062)	(0.0035)	(0.0045)	(0.0067)	(0.0045)
Very bad health	-0.010	-0.003	-0.011*	0.018	0.016	0.018
	(0.0067)	(0.0040)	(0.0064)	(0.0128)	(0.0138)	(0.0126)
Constant	-0.007	0.028***	-0.009	-0.014*	0.013	-0.015*
	(0.0091)	(0.0107)	(0.0089)	(0.0079)	(0.0097)	(0.0080)
Observations	6,942	6,942	6,942	6,942	6,942	6,942
R-squared	0.423	0.019	0.424	0.410	0.018	0.411

Notes: Estimated standard errors, obtained by bootstrapping, are in parentheses. ***/**/* indicates statistically significant at 1%/5%/10%. Omitted categories: Parents gambled but were not PG, male, white, single, NQF level 0, employed, very good health, North East. Government office region omitted from reporting.

Table A2.17: IV estimated coefficients (PG defined by PGSI>7)

Instruments	DSM Score	DSM Score and Parental PG	Parental PG
Ln Y (γ)	0.523***	0.523***	0.362***
	(0.0600)	(0.0628)	(0.1127)
PG (δ)	-3.544***	-3.552***	-36.896***
	(0.8455)	(0.8990)	(12.1641)
Age	-0.040***	-0.040***	-0.041***
	(0.0091)	(0.0086)	(0.0142)
Age^2	0.001***	0.001***	0.000***
	(0.0001)	(0.0001)	(0.0001)
Female	0.219***	0.219***	-0.018
	(0.0476)	(0.0475)	(0.1037)
Mixed Ethnicity	-0.630**	-0.630**	-0.954***
	(0.2739)	(0.2661)	(0.2892)
Asian/Asian British	-0.094	-0.093	0.092
	(0.1419)	(0.1412)	(0.2344)
Black/Black British	0.242	0.242	0.424
	(0.1780)	(0.1840)	(0.3977)
Chinese/Other	-0.035	-0.035	-0.328
	(0.2811)	(0.2842)	(0.3050)
Married	0.347***	0.347***	0.216
	(0.0821)	(0.0845)	(0.1572)
Separated/Divorced	-0.043	-0.044	-0.232
	(0.1032)	(0.1064)	(0.1817)
Widowed	-0.561***	-0.561***	-0.744***
	(0.1303)	(0.1334)	(0.1763)
Constant	2.882***	2.883***	5.066***
	(0.6083)	(0.6423)	(1.3024)
δ/γ	-6.775***	-6.791***	-101.863*
	(1.8102)	(1.9335)	(55.8167)
CV (£b, pa)	50.4	50.6	758
Observations	6,942	6,942	6,942
R-squared	0.075	0.075	0.085

Notes: In the first column PG is instrumented using the DSM score, while in the second column it is instrumented with DSM score and parental problem gambling using GMM estimation. In the final column PG is instrumented with parental problem gambling only. Female, age, age^2, marital status, ethnicity, and government office region are included as control variables. Estimated standard errors are in parentheses. ***/**/* indicates statistically significant at 1%/5%/10%. CV is computed by multiplying the estimated δ/γ by 0.0055*46m*$\bar{Y}$, where $\bar{Y}$ = £29,560, 0.0055 is the proportion with PG=1, and is recorded in £ b pa.

Table A2.18: Heteroskedastic IV Estimated Coefficients

Instrument:	PGSI Score and Parental PG	Synthetic	PGSI Score, Parental PG and Synthetic
Dependent variable	W	W	W
Ln Y (γ)	0.529***	0.533***	0.533***
	(0.0567)	(0.0583)	(0.0583)
PG (δ)	-2.498***	-1.541***	-1.577***
	(0.5972)	(0.3790)	(0.3814)
Age	-0.041***	-0.041***	-0.041***
	(0.0086)	(0.0086)	(0.0086)
Age^2	0.001***	0.001***	0.001***
	(0.0001)	(0.0001)	(0.0001)
Female	0.218***	0.228***	0.227***
	(0.0460)	(0.0454)	(0.0454)
Mixed Ethnicity	-0.624**	-0.613**	-0.613**
	(0.2701)	(0.2625)	(0.2625)
Asian/Asian British	-0.088	-0.098	-0.097
	(0.1395)	(0.1377)	(0.1377)
Black/Black British	0.231	0.228	0.228
	(0.1595)	(0.1717)	(0.1718)
Chinese/Other	-0.034	-0.022	-0.023
	(0.2795)	(0.2756)	(0.2756)
Married	0.357***	0.358***	0.358***
	(0.0758)	(0.0789)	(0.0789)
Separated/Divorced	-0.022	-0.022	-0.022
	(0.1025)	(0.1022)	(0.1022)
Widowed	-0.551***	-0.547***	-0.547***
	(0.1263)	(0.1292)	(0.1292)
Constant	2.842***	2.766***	2.768***
	(0.5623)	(0.5888)	(0.5889)
CV (£b, pa)	46.3	28.3	29.0
1st stage F	2488.75***	1418.10***	1306.521***
Observations	6,942	6,942	6,942
R-squared	0.066	0.071	0.071

Notes: Estimated standard errors, obtained from bootsrapping, are in parentheses.. ***/**/* indicates statistically significant at 1%/5%/10%. CV is computed by multiplying the estimated δ/γ by 0.072 * 46m * $\bar{Y}$, where $\bar{Y} = £29{,}560$, 0.072 is the proportion with PG=1, 46m is the adult population, and CV is recorded in £ billion pa. Omitted categories: Male, white, single, North East. Government office region omitted from reporting.

Table A2.19: Full ordered probit estimated parameters

Instruments:	PGSI	PGSI and Parental PG
Ln Y	0.236***	0.236***
	(0.0313)	(0.0313)
DSM PG	-1.221***	-1.219***
	(0.2939)	(0.2930)
Age	-0.024***	-0.024***
	(0.0048)	(0.0048)
Age Squared	0.000***	0.000***
	(0.0000)	(0.0000)
Female	0.120***	0.120***
	(0.0260)	(0.0260)
Mixed Ethnicity	-0.367***	-0.367***
	(0.1291)	(0.1291)
Asian/Asian British	-0.033	-0.033
	(0.0755)	(0.0755)
Black/Black British	0.115	0.115
	(0.0920)	(0.0919)
Chinese/Other	-0.059	-0.059
	(0.1585)	(0.1586)
Married	0.247***	0.247***
	(0.0418)	(0.0418)
Separated/Divorced	-0.034	-0.034
	(0.0531)	(0.0531)
Widowed	-0.345***	-0.345***
	(0.0693)	(0.0693)
Observations	6,942	6,942

Notes: In the first column PG is instrumented using the PGSI score, while in the second column it is instrumented with PGSI score and parental problem gambling. Female, age, age^2, marital status, ethnicity, and government office region are included as control variables. Omitted categories are male, single, white, and North East. Government office region and cut-points are omitted from reporting Estimated standard errors are in parentheses. ***/**/* indicates statistically significant at 1%/5%/10%.

Table A2.20: Ordered probit estimated parameters of interest (PG defined by PGSI>7)

Dependent Variable	W	W
Ln Y (γ)	0.234***	0.234***
	(0.0313)	(0.0313)
PG (δ)	-1.614***	-1.613***
	(0.3171)	(0.3170)
δ/γ	-6.887***	-6.883***
	(1.6587)	(1.6585)
Marginal Effects, $\Delta\ \Pr(W = 1, \ldots, 10)\ /\Delta PG$		
W=1	0.060***	0.060***
	(0.0128)	(0.0128)
2	0.028***	0.028***
	(0.0065)	(0.0065)
3	0.030***	0.030***
	(0.0068)	(0.0068)
4	0.035***	0.035***
	(0.0077)	(0.0077)
5	0.151***	0.151***
	(0.0299)	(0.0299)
6	0.080***	0.080***
	(0.0162)	(0.0162)
7	0.159***	0.158***
	(0.0314)	(0.0314)
8	0.057***	0.057***
	(0.0128)	(0.0127)
9	-0.131***	-0.131***
	(0.0262)	(0.0262)
10	-0.469***	-0.469***
	(0.0920)	(0.0920)
CV (£b, pa)	51.3	51.2

Notes: In the first column PG is instrumented with DSM score. PG in the second column is instrumented with parental problem gambling as well as DSM score. Female, age, age^2, ethnicity, and government office region are included as control variables and their coefficients are not reported here. Estimated standard errors are in parentheses. ***/**/* indicates statistically significant at 1%/5%/10%.

Table A2.21: Well-being OLS regressions using responses to individual DSM questions

Dependent Variable	W	W	W
Ln(Household Income)	0.535***	0.536***	0.534***
	(0.0580)	(0.0581)	(0.05814)
DSM1=1	-0.382		
	(0.2341)		
DSM2=1	-0.245		
	(0.2101)		
DSM3=1	0.403		
	(0.3879)		
DSM4=1	0.023		
	(0.4811)		
DSM5=1	-1.461***	-2.012***	
	(0.5317)	(0.4259)	
$\widehat{\mathrm{DSM}\,5} = 1$			-4.039***
			(1.05927)
DSM6=1	-0.730		
	(0.6599)		
DSM7=1	-0.396		
	(0.4999)		
DSM8=1	-0.967		
	(0.9192)		
DSM9=1	0.245		
	(0.5955)		
DSM10=1	0.375		
	(0.4932)		
Constant	2.791***	2.719***	2.776***
	(0.5853)	(0.5868)	(0.58765)
δ/γ	-5.865***	-3.751***	-7.569***
	(1.7207)	(0.8941)	(2.1609)
CV (£b, pa)	53.7	27.2	54.9
Observations	6,936	6,942	6,942
R-squared	0.075	0.073	0.067

Notes: Robust standard errors are parentheses. ***/**/* indicates statistically significant at 1%/5%/10%. Omitted categories: Male, white, single, North East. Controls omitted from reporting. DSM Q5 is instrumented with PGSI score in column 3. CV calculated as the sum of individual question coefficients divided by the log income coefficient, weighted by the proportion of affirmative answers, and multiplied by the adult population and mean household income. δ/γ is computed as the sum of individual question coefficients divided by the log income coefficient.

Table A2.22: Well-being OLS regressions using responses to individual PGSI questions

Dependent Variable	W	W	W
Ln Y	0.539***	0.538***	0.533***
	(0.0582)	(0.0584)	(0.05935)
PGSI 1=1	-0.612		
	(0.6732)		
PGSI 2=1	-1.552		
	(1.2505)		
PGSI 3=1	0.421		
	(0.5332)		
PGSI 4=1	2.745***		
	(1.0582)		
PGSI 5=1	1.303	-0.817	
	(1.2001)	(0.6541)	
$\widehat{PGSI\ 5} = 1$			-5.827***
			(2.07485)
PGSI 6=1	-1.877		
	(1.2107)		
PGSI 7=1	0.289		
	(0.7368)		
PGSI 8=1	-1.418		
	(1.6303)		
PGSI 9=1	-1.244		
	(0.7883)		
Constant	2.702***	2.675***	2.756***
	(0.5872)	(0.5898)	(0.60277)
δ/γ	-3.608	-1.518	-10.939***
	(2.4530)	(1.2289)	(4.0522)
CV (£b, pa)	16.4	5.1	36.4
Observations	6,942	6,942	6,942
R-squared	0.072	0.067	0.050

Notes: Robust standard errors are parentheses. ***/**/* indicates statistically significant at 1%/5%/10%. Omitted categories: Male, white, single. Controls omitted from reporting. PGSI Q5 is instrumented with DSM score in column 3. CV calculated as the sum of individual question coefficients divided by the log income coefficient, weighted by the proportion of affirmative answers, and multiplied by the adult population and mean household income. δ/γ is computed as the sum of individual question coefficients divided by the log income coefficient.

Table A2.23: OLS well-being regression comparison between BGPS and HSE/SHS

	BGPS	SHS/HSE		
Dependent Variable:	Well-being 1-10	Well-being 0-10	WEMWBS	GHQ 12
Ln(Y) (δ)	0.233***	0.722***	0.772***	-0.640***
	(0.0287)	(0.0273)	(0.0229)	(0.0242)
DSM PG (γ)	-0.742***	-0.187	-0.797***	0.651***
	(0.1928)	(0.1746)	(0.1767)	(0.1729)
Age	-0.020***	-0.042***	-0.026***	0.018***
	(0.0045)	(0.0029)	(0.0026)	(0.0025)
Age^2	0.000***	0.001***	0.000***	-0.000***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Female	0.115***	0.046***	0.035**	0.090***
	(0.0239)	(0.0162)	(0.0146)	(0.0138)
Separated/Divorced	-0.233***	-0.023	0.181***	-0.117***
	(0.0455)	(0.0374)	(0.0305)	(0.0314)
Single	-0.220***	0.022	0.136***	-0.197***
	(0.0411)	(0.0299)	(0.0264)	(0.0261)
Widowed	-0.499***	-0.009	0.134***	-0.182***
	(0.0614)	(0.0385)	(0.0335)	(0.0302)
Mixed Ethnicity	-0.402***	0.296**	0.166	-0.114
	(0.1372)	(0.1455)	(0.1318)	(0.0890)
Asian/Asian British	-0.134*	0.137**	0.410***	-0.128***
	(0.0721)	(0.0698)	(0.0585)	(0.0450)
Black/Black British	-0.026	-0.122	0.386***	-0.034
	(0.0846)	(0.2445)	(0.1011)	(0.0814)
Chinese/Other	-0.089	0.117	0.288***	-0.094
	(0.1450)	(0.0889)	(0.0829)	(0.0767)
Constant	-2.072***	-6.725***	-7.649***	6.432***
	(0.3127)	(0.2904)	(0.2454)	(0.2546)
δ/γ	-3.186***	-0.260	-1.033***	-1.017***
	(0.9285)	(0.2423)	(0.2320)	(0.2746)
Observations	6,942	13,335	17,312	19,711
R-squared	0.065	0.122	0.106	0.076

Robust standard errors in parentheses. Well-being measures all normalised by subtracting mean and dividing by standard deviation. ***/**/* denotes statistical significance at the 1%/5%/10% level.

Table A2.24: OLS estimates using polynomial expansions of log household income

	(1)	(2)	(3)
Ln(Y) Polynomial Expansion:	Linear	Quadratic	Cubic
Ln(Y)	0.534***	4.365***	-7.924
	(0.0583)	(1.1373)	(22.4224)
Ln(Y)^2		-0.193***	1.054
		(0.0567)	(2.2587)
Ln(Y)^3			-0.042
			(0.0757)
DSM PG (δ)	-1.382***	-1.378***	-1.378***
	(0.3719)	(0.3710)	(0.3710)
Age	-0.040***	-0.040***	-0.040***
	(0.0086)	(0.0086)	(0.0086)
Age^2	0.001***	0.001***	0.001***
	(0.0001)	(0.0001)	(0.0001)
Female	0.230***	0.232***	0.232***
	(0.0454)	(0.0454)	(0.0454)
Mixed Ethnicity	-0.611**	-0.627**	-0.631**
	(0.2630)	(0.2612)	(0.2616)
Asian/Asian Briish	-0.099	-0.134	-0.138
	(0.1379)	(0.1383)	(0.1382)
Black/Black British	0.227	0.229	0.218
	(0.1718)	(0.1720)	(0.1728)
Chinese/Other	-0.020	-0.062	-0.066
	(0.2760)	(0.2744)	(0.2748)
Married	0.358***	0.384***	0.386***
	(0.0790)	(0.0796)	(0.0796)
Separated/Divorced	-0.022	0.013	0.012
	(0.1023)	(0.1028)	(0.1028)
Widowed	-0.547***	-0.452***	-0.449***
	(0.1294)	(0.1343)	(0.1342)
Constant	2.755***	-16.224***	24.027
	(0.5894)	(5.6972)	(74.0037)
Log Income Marginal Effect at means (γ)	0.534***	0.466***	0.447***
	(0.0583)	(0.0586)	(0.0654)
δ/γ	-2.589***	-2.957***	-3.083***
	(0.7605)	(0.8886)	(0.9552)
Observations	6,942	6,942	6,942
R-squared	0.071	0.073	0.073

Notes: Robust standard errors are parentheses. ***/**/* indicates statistically significant at 1%/5%/10%. Omitted categories: Male, white, single. Controls omitted from reporting. δ/γ is computed as the PG coefficient divided by the marginal effect of log household income at means.

3 The Economics of Lotto Design

3.1 Introduction

Lotto is the most popular form of lottery game and typically used by governments and charitable organisations to raise funds for goods and services which may not be easy to fund through other means, such as raising taxes or voluntary contributions. One of the main objectives for lottery operators is to maximise net revenues from the games they offer by appropriate design of the game. As is often the case when estimating demand models, price is endogenous to sales and care must be taken in statistical modelling to infer the causal relationship. This is a particularly tricky issue in the case of lotto draws since prizes, and thus their expected value and the 'effective price' of a ticket, are funded using sales revenue for that draw. Early literature (see Clotfelter and Cook, 1993, and Walker, 1998) pointed out that lotto features "peculiar economies of scale" which imply that price is endogenous in a way unique to lotto, and that modelling demand needed to recognise how sales respond to short term variation in the distribution of prizes. However, the empirical models of sales revenue in more recent research has not been based on a sound approach to identification. To resolve the issue instrumental variable methods are typically applied, and almost all of the existing literature relies on rollovers to fulfil the role of the instrument. However, rollovers themselves depend on sales in the previous draw and current sales are correlated with sales in the previous draw, so rollovers do not satisfy the requirements of a valid instrument. This paper aims to correct this shortcoming though a novel identification strategy and the application of semi-parametric analysis.

Section 3.2 provides an overview of the lotto literature. Section 3.3 argues that rollovers make unsuitable instruments for price since they themselves are also inherently endogenous due to way in which lotto is designed. Section 3.4 proposes that a novel yet powerful alternative instrument candidate is available by exploiting systematic non-random number selection by lotto players – a phenomenon known as 'conscious selection'. Section 3.5 describes the data used in Section 3.6 to provide empirical evidence by estimating demand models for lotto using OLS and instrumental variables techniques using this conscious selection phenomenon.

More recent literature suggests that the expected price is not a sufficient statistic for determining sales of lotto tickets. In particular, prizes may affect sales apart than via the expected price. Walker and Young (2001) allow for higher order moments of the prize distribution so that variance and skewness play a role in determining sales. In contrast, Forrest *et al* (2002) uses the jackpot size itself as the relevant explanatory variable. Therefore, Section 3.7 presents estimates of a reduced form of lottery demand, where the jackpot prize included in the model directly and is identified using the same strategy as for the price model. Both Walker and Young (2001) and Forrest *et al* (2002) impose specific but arbitrary functional form restrictions which we have no more reason to believe than the functional form imposed by the price. So, an additional contribution in Section 3.7 is to extend their work by adopting a totally flexible approach in a semi-parametric estimation and test this against a parameterised polynomial expansion of prizes.

Re-designs to the UK National Lottery's flagship product occurred in 2013 and later in 2015 with a view to rejuvenate dwindling sales and, by extension, good causes funding. The final contribution of this chapter, in Section 3.8 uses data beyond the main sample to examine whether these changes were successful in this objective.

3.2 Background

Commercial gambling products are rarely available at favourable odds and the widespread prevalence of gambling at unfair odds has long been a puzzle for economists. One argument that has been used to rationalise gambling with Expected Utility (EU) Theory is based on the idea that stakes are usually small, so that any downside losses are also small, relative to wealth, and the likely gains are either small or at such long odds that they can be neglected. This argument suggests that agents will act in a way that, at least locally, is risk neutral. But this explanation also demands that the expected loss is small relative to the non-pecuniary gains associated with participating. In the case of lotto, 'take-out' rates – the proportion of sales revenue not returned in prizes – are typically large, often in excess of 50%, so the potential for non-pecuniary gains has to be large enough to outweigh this. Lotto operators emphasise the, albeit remote, possibility of life changing gains and, while the expected value of such unlikely prizes is small, the fact that participation might offer the ability to dream about such prospects might be real and important.

Friedman and Savage (1948) embed this possibility in a utility of wealth function that they use to rationalise the coexistence of insurance and gambling, although Markowitz (1952) and Hartley and Farrell (2002) marshal convincing arguments against this idea as a plausible explanation. In addition, in the case of lotto, the rationale for the large take-out rate is that the revenue is used to fund public goods and most operators dedicate a large proportion of the take-out to "good causes". Morgan (2000) showed that, within an EU framework, games with fixed prizes, such as a raffle, can in theory come closer to efficient public good provision than reliance on voluntary contributions. At the very least, lotteries with a fixed prize component will yield levels of contributions above those obtained from reliance only on voluntary contribution, and large – albeit, fixed – prizes could raise sufficient revenue to provide public goods close to, but not exceeding, an optimum level. Morgan and Sefton (2000) provide empirical support. Thus, the expectation of losing might be offset by the warm glow that one is losing to a good cause.

If risk aversion is locally close to neutral, and there is sufficient warm glow, then the relevant determinant of lottery demand will be the expected value of the gamble since the mean of the prize distribution (including the loss of the stake as a negative prize) is then a sufficient statistic for demand. This motivates the specification that appears most commonly in the literature and which is followed below.

The 2010 British Gambling Prevalence Survey estimated that 72% of the adult population – approximately 34 million individuals – had engaged in some form of gambling activity in the previous 12 months (Wardle *et al,* 2011). Of all the gambling products available, lotteries have proven to be the most popular with the UK National Lottery, the sole licensed distributor of the UK lotto game, achieving sales of £4.6 billion in the year to March 2015 (Gambling Commission, 2015). This accounts for approximately 0.7% of household expenditure (Office for National Statistics, 2014). In the US, lotteries are available in 43 states, each providing state-level tax dollars, and they collectively accumulated $70.1 billion worth of ticket sales in 2014. Since their implementation in New Jersey in 1971, state-run lotteries in the US have raised a total of $300 billion in revenue for state spending.

Lotto is a pari-mutuel, low-cost game which offers large prizes with small win probabilities. Prior to reforms in October 2013 and October 2015, the UK lotto cost £1 to enter and regularly offered top prizes (jackpots) in excess of £8 million[48]. The game involved players paying the entry fee and choosing six numbers from 1-49 and participants were free to choose their own numbers or, from 1996 onwards, have the vending terminal randomly make the selection for them. Tickets are valid only for specific draws, which typically occur on Wednesday and Saturday, in which a mechanised device chooses six numbers from 1-49. Players were rewarded with cash prizes if their chosen numbers matched between three and six

[48] Game redesigns in the UK and US have driven large spikes in sales and prize money with recent record jackpots of £66 million in the UK and $1.5 billion in the US Powerball.

of the randomly drawn numbers. A fixed prize of £10 was awarded to those who matched three numbers, and players won equal shares of prize pools for matching 4-6 numbers, with the jackpot being shared between those who matched all six. The '6 from 49' design means that the probability of any given ticket winning this prize is extremely unlikely at approximately 1/14 million. For this reason, it is not unusual for there to be no winners of the jackpot, in which case the prize money originally allocated to this prize is added to the jackpot prize pool in the following draw in an event known as a 'rollover'.

Historically, lotteries have been used to finance public good provision especially when alternative funding is hard to raise. For example, the US confederate states made extensive use of lottery funding in the US civil war, the English used them to finance the defence of the realm against the Spanish Armada, and US elite universities (Yale, Harvard, and Princeton) used them to fund infrastructure long before they had wealthy alumni to draw upon. Because of their success in raising public finance, lotteries today are often operated either directly by governments or a private sector licensee under strict regulation (Morgan, 2000). The UK lottery was introduced partly with a view to funding the renovation of the Royal Opera House at a time when it would have been politically impossible to use regular tax dollars for this purpose. Lotteries are also often used in federated countries with constitutional constraints on their powers of taxation. For this reason, one of the main objectives of lotto design is to maximise tax revenues for public good provision (Clotfelter and Cook, 1990). In the economics literature, attention has been focused on modelling consumer demand for lotto tickets and eliciting a price elasticity of demand to evaluate whether the objective of revenue maximisation is being achieved.

Clotfelter and Cook (1990) note that the definition of price in the case of lottery tickets requires some clarification since consumers face two prices when making their purchase decision: the 'sticker' price and the 'effective' price – which is simply the sticker price minus the expected value of winnings. The former is fixed whereas the latter varies from draw to draw due to changes in the size of prize pools which are dependent on sales for that particular draw. For this reason, it is the effective price which is favoured in the literature to estimate the price elasticity (Walker, 1998; Farrell and Walker, 1990; Forrest *et al* 2000).

3.3 An analytical model of lotto supply

Early models of lotto sales choose strong parametric restrictions (for example in Farrell *et al*, 1999; and Forrest *et al*, 2002) which involve modelling current sales, S_t, as a function of past sales, S_{t-j}, and the effective price, P_t, along with controls for time trends and exogenous demand shocks.

The simplest lotto games are designed so that players choose n integers from a possible N with pari-mutuel prizes available for matching $k \leq n$ of the numbers drawn randomly by a mechanical device. More complex designs involve multiple devices and allow for fine gradation in the prize pools. Typically, fixed proportions of sales revenue are allocated to the pari-mutuel prize 'pools' of which successful players win equal shares. The operator retains a proportion of sales, $\tau \in [0,1)$, for paying tax, funding public goods and to cover operating costs and this proportion is known in the industry as the 'take-out rate'. The shares of sales assigned to specific pools are chosen by the operator and, in practice, their sum is typically regulated to ensure that a fixed proportion, $1-\tau$, is returned in prizes.

Let $\rho_k \in [0,1]$ be the proportion of total prize money, $(1-\tau)S_t$, allocated to the prize pool associated with matching k numbers. Because the prize pools are pari-mutuel and funded using proportions of sales, the *expected* prize paid to any winning individual of that prize tier is constant when weighted by the likelihood of winning. Normalising the price of an entry to 1, if all prize pools were won by at least one player then the effective price of a ticket would trivially be τ. Real lotto games are often designed so that there is a non-trivial probability of there being *no* winner of the prize pool associated with *k*=*n*, known as the "jackpot" pool, in which case the money in that pool is added to the same prize pool in the following draw[49]. This chapter refers to a draw *following* the draw where there are no winners of this prize as a 'rollover draw'. If the prize for a particular draw is not won, it *reduces* the value of prizes shared amongst players for that draw and *increases* the price. Since the jackpot is transferred to the next draw it then also raises the expected value of prizes for that draw, thus *decreasing* the price for that draw. Walker and Young (2001) show that the possibility of rollovers has a significant impact on the expected value of prizes, and hence also on the effective price.

Scoggins (1995) emphasises that the probability of a rollover occurring is a function of both the level of sales and the statistical difficulty of the game. Let π_n be the probability of winning a share of the jackpot prize, awarded for matching all winning numbers. In addition

[49] Whilst any prize pool could theoretically fail to have at least one winner, in practice sales are sufficiently large that only the jackpot prize pool is sufficiently difficult to win to induce a rollover, hence only rollovers of this prize are considered in the subsequent analysis for determining the effective price.

to setting the shares of sales allocated to each prize pool, the operator is also able to determine the statistical difficulty of the game by appropriate choice of n and N to influence π_n as follows[50]:

$$\pi_n = \frac{n!\,(N-n)!}{N!} \tag{3.1}$$

From this it is possible to determine the likelihood of a rollover occurring, $p_{R,t}$. If there is only one ticket sold, the likelihood that the prize pool will roll over is simply $1-\pi_n$, if two tickets are sold the probability is $(1-\pi_n)^2$. Hence, if S_t tickets are sold the probability of a rollover occurring is:

$$p_{R,t} = (1-\pi_n)^{S_t}. \tag{3.2}$$

The simplest possible model of lotto is a game with only one prize pool ($\rho_n = 1$), which is shared among all players who match all n winning numbers. If the previous draw had at least one winner, this draw is a non-rollover draw and the prize pool is only determined by τ and the number of tickets sold, S_t. The effective price of a ticket in this draw is also influenced by the probability of there being no winners as follows:

$$P_t = \left(1-p_{R,t}\right)\tau + p_{R,t}. \tag{3.3}$$

That is, the effective price is the probability weighted average of the ticket price if the jackpot is won and the ticket price if there are no winners. For rollover draws where the previous jackpot was not won, the expected value of prize shares if the jackpot is won is increased by R_t/W_t, where R_t is the amount of money added to the draw (the jackpot in $t-1$) and $W_{n,t}$ is the number of winners of the jackpot prize in draw t. Since $W_{n,t}$ is simply $\pi_n S_t$ in expectation, the effective price for a rollover draw can be expressed as

$$P_t = \left(1-p_{R,t}\right)\left(\tau - \frac{R_t}{S_t}\right) + p_{R,t}. \tag{3.4}$$

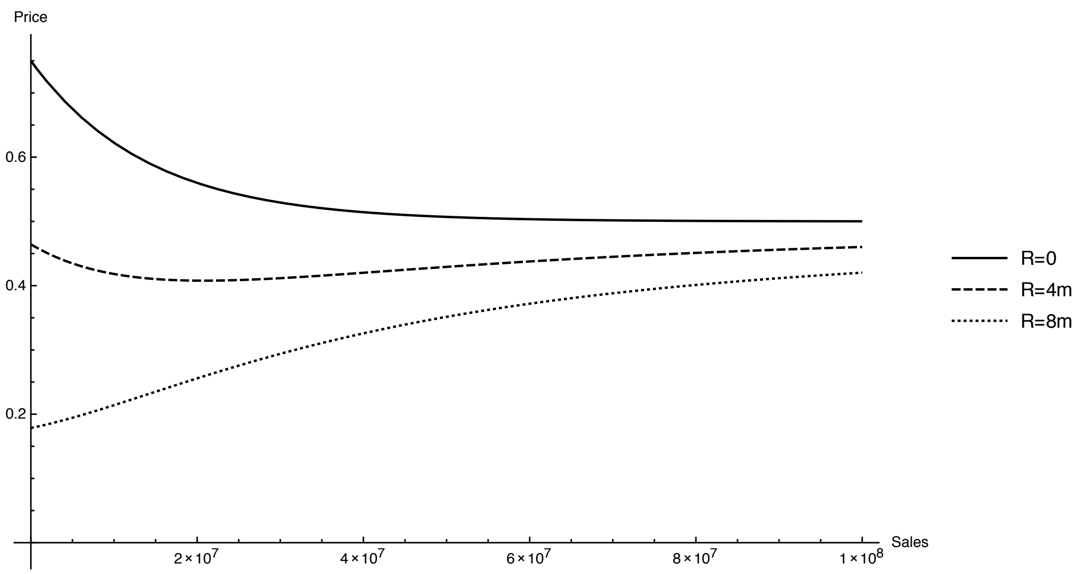
This can be thought of as the inverse supply function of lotto tickets, where $R_t = 0$ corresponds to the non-rollover case. Several features of lotto game designs are clear from the definition of price. Trivially, increasing the proportion of sales retained by the operator

[50] This is a special case of the hypergeometric function which can be used to determine the probability of matching any $k \leq n$ numbers and can be written as $\pi_k = \binom{n}{k}\binom{N-n}{n-k}/\binom{N}{n}$, where parentheses denote binomial choice functions.

increases the price, $\delta P_t/\delta\tau > 0$, which makes the game less attractive to players. As in Cook and Clotfelter (1993), price has a clear inverse relationship with rollover size, $\delta P_t/\delta R_t < 0$, since rollovers increase the expected value of prizes available. Figure 3.1 illustrates how the inverse supply function asymptotes towards τ from above as more tickets are sold when there is no rollover present due to the increasing likelihood that the prize pool will be won. Cook and Clotfelter (1993) termed this the 'peculiar economies of scale of lotto'.

In the event of a rollover draw, however, the effective price of a ticket *increases* as sales increase because higher sales increases the likelihood that the fixed rollover component of the jackpot prize will be shared amongst more winners, who then each receive a smaller share. This causes price to asymptote towards τ from below when the jackpot prize in enhanced by a rollover.

Figure 3.1: Inverse supply function for a 6/49 lottery with $\tau = 0.5$ *and rollovers of £0, £4m, and £8m.*



The possibility of a rollover in lotto has two important implications. Firstly, for any expected level of sales, the operator is able to alter π_n (by choosing *n* and *N*) to influence price through adjusting the likelihood of a rollover occurring. Secondly, players will form expectations about the likelihood of a rollover occurring, and thus their expectation of price, by estimating the level of sales and adjust their consumption decision accordingly. In particular, potential players may decide to defer the purchase of a ticket if they expect low sales, and a high probability of a rollover, in favour of likely higher prizes in the following draw. The problem for the operator is to balance the gain in sales when rollovers occur with the suppression of sales because of the probability of one occurring.

Prior to October 2015 the UK lotto was characterised by $n = 6$ and $N = 49$ – a design that has proved very popular in the industry. This translates to a rollover probability of approximately 1 in 14 million. There were also pari-mutuel prizes available for correctly predicting four and five numbers, and five plus a bonus number (b) drawn from the same set of 49 after the six main numbers were drawn. The prizes for matching 4, 5, 5+b and 6 of the winning set were funded using fixed proportions of the total prize money available *after* the additional fixed prizes of £10 were paid to those who matched 3 numbers. Thus, the prize pool in draw *t* for matching $k = 4, 5, 5 + b, 6$ numbers, J_{kt}, can be determined using[51]:

$$J_{kt} = \rho_k\left[(1 - \tau)S_t - 10W_{3,t}\right]. \tag{3.5}$$

A final feature of the UK lotto is the twice-weekly drawing. In the sample period analysed below, UK lotto draws occurred on Wednesdays and Saturdays of the same week[52]. The two draws are intrinsically linked by rollovers in the sense that money from jackpots not won on Wednesday is added to the corresponding prize pool for the following Saturday and vice versa. Despite the link between draws and being identical by design, in the following analysis demand for Wednesday and Saturday are estimated draws separately. Doing so does not then restrict slope coefficients to be identical for both games and allows for the possibility that players on Wednesday and Saturday may have different risk preferences and responses to price variation.

This assumption is apparently justified simply by examining the descriptive statistics in Section 3.5 with sales for Wednesday draws being approximately half that of Saturday draws. Consequently, the average likelihood of a given draw being enhanced by a rollover is much larger for Saturday games than for Wednesday. The difference in price for non-rollover draws on Saturdays is subsequently much smaller than it is for Wednesdays - as would be expected from Figure 3.1.

The important empirical lesson to be taken from the simple structure of the game is that the expected value, and hence price, in any draw depends on the size of the rollover jackpot,

[51] For the UK lotto, the share of the pari-mutuel prize fund allocated to each prize tier is parameterised as follows; $\rho_6 = 0.52$, $\rho_{5+b} = 0.16$, $\rho_5 = 0.10$, and $\rho_4 = 0.22$.
[52] Between November 1994, the introduction of the game, and February 1997 draws only occurred every Saturday.

which depends linearly on the level of sales in the previous draw, and on the (expected) level of sales in the current draw. Thus, the price is endogenous because of its dependence on S_t. This is obvious from Figure 3.1 and will be empirically important if sales are centred around a relatively steep part of the inverse supply curve. In practice, games typically are designed so that they operate at such levels that engenders a high probability of rollovers occurring – events which tend to increase sales in the next draw (and, to a lesser extent, in subsequent draws).

3.4 Identifying the demand for lotto

Using effective price as a determinant of lotto sales is a useful model specification since it allows for a price elasticity to be easily obtained. Walker (1998) suggests that lotteries will maximise sales, and therefore be "efficient", when marginal revenue from additional sales is zero (since the marginal cost of production is close to zero), implying an optimal price elasticity of demand of -1. Allowing for a lagged dependent variable, the long-run price elasticity of demand was estimated to be -1.07 for the UK National Lottery and Walker concluded that the game was indeed appropriately designed. This estimate for the UK game is supported by Forrest *et al* (2000) who estimated a value for the price elasticity of demand of -1.03, and Farrell *et al* (2000) with a value of -1.06.

However, models using effective price suffer from the endogeneity issue highlighted above. Consider the following myopic model as estimated in Scoggins (1995), Farrell *et al* (1999), Forrest *et al* (2000), and Forrest *et al* (2002), amongst others, with sales at t dependent on lagged sales, S_{t-i}, current price, P_t, and controls for seasonality and structural events, X_t, to be estimated via OLS:

$$S_t = \alpha + \sum_{i=1}^{I} \beta_i S_{t-i} + \eta P_t + \gamma X_t + \varepsilon_t. \tag{3.6}$$

Since the game is drawn twice-weekly and Wednesday and Saturday are treated as different games here, sales in $t-1, t-3$ ... are 'cross' lags of sales and $t-2, t-4$ … are 'own' lags of sales. It is possible to retrieve a long run price elasticity of demand, ϵ_{LR}, evaluated at mean sales, from this model using the following formula:

$$\epsilon_{LR} = \frac{\partial S_t}{\partial P_t} \frac{P}{\bar{S}\left(1 - \sum_{i=2,4,\ldots} \beta_i\right)} \tag{3.7}$$

Recalling the definition of price in equation (3.4), it is clear that it is unrealistic to assume that $E[\varepsilon_t|P_t] = 0$, therefore OLS estimation will be biased and an instrumental variable approach is necessary. This issue is raised in Walker and Young (2000) who, along with subsequent work by Forrest *et al* (2000) and Forrest *et al* (2002), use rollover size as their exclusion restriction in the first stage of their modelling and find estimates of price elasticity which are not statistically different from -1, concluding that the game is taxed efficiently.

Recall, however, that in the simple game outlined above $R_t = \tau S_{t-1}$ which, if sales are serially correlated as suggested in equation (3.6), would imply that the size of the rollover is itself not orthogonal to sales since it is also correlated with previous sales which then mechanically determine the rollover size. This casts doubt on the validity of relying on rollover size as an instrument.

The proposed solution to the endogeneity issue here involves extending the work of Farrell *et al* (2000) who appealed to 'conscious selection' as an explanation of why rollovers occurred far more frequently than the theoretical rollover probability predicts. This term is used within the industry to refer to systematic non-random number choice by players. For the average level of sales reported in Section 3.5 and the design parameters of the UK game, the supply theory above suggests that only 5.7% of Saturday draws should roll-over, whereas the actual proportion of Saturday draws with no jackpot winners is 13.4%. Farrell *et al* (2000) attribute this discrepancy between theoretical and realised rollover proportions to the fact that players, who are able to choose their own numbers, tend to select some numbers more frequently than others. By comparing the actual distribution of prize winners of each prize pool with the hypothetical distributions under the assumption of random selection by players, it is possible to estimate the likelihood of each of the 49 available numbers being chosen by a randomly selected ticket purchased for the game. Moreover, so long as each number has appeared at least once, it is possible to exploit the variation in the number of prize winners, conditional on the number of tickets sold, to estimate the proportion of tickets containing a specific number. Farrell *et al* (2000) construct a likelihood function, with 48 independent parameters, and estimate the probability of each number between 1-49 appearing on a randomly selected ticket and find this varies from as low as 1.2% (number 46) to 2.9% (number 7) compared to a probability of 2.4% that would have been expected if numbers were chosen randomly. Overall, they find that numbers 1-12 prove to be most popular and 32-49 being the least popular.

Conscious selection has two relevant impacts on the effective price of a ticket. Firstly, there will be an increase (decrease) in the likelihood of a rollover if unpopular (popular) numbers are drawn among the winning numbers. Secondly, if a rollover does occur when unpopular (popular) numbers are drawn then the rollover size will be unusually large (small) relative to the number of tickets sold as there will be fewer (more) winners of the three-ball prize which reduce the pari-mutuel prize fund[53].

The definition of price derived in Section 3.3 implicitly assumes that players choose numbers randomly. The implication that the likelihood of a rollover occurring is inversely related to the number of tickets sold is certainly plausible, but equation (3.2) only accurately models this probability if each of the tickets sold are unique (and only approximates the true probability if numbers are randomly selected by players). Rather, the true probability is dependent on the number of unique selections bought by players which should increase with sales but not necessarily in a linear fashion. As more tickets are sold, and with systematic non-random number selection by players, it becomes increasingly likely that the same combination of numbers appears on more than one ticket. To capture this, the rollover probability can be better described as:

$$p_{R,t} = (1-\pi_n)^{f(S_t,\delta_{j,t})} \tag{3.8}$$

where $\delta_{j,t}$ is some indicator of the popularity of each of the $j = 1, \dots 49$ available numbers in draw t. f denotes some function relating the level of sales and the popularity of each of the 49 numbers to the number of *unique* combinations sold. Including this refined definition of rollover probability implies re-writing effective price as:

$$P_t = \left[1-(1-\pi_n)^{f(S_t,\delta_{i,t})}\right]\left(\tau - \frac{R_t}{S_t}\right) + (1-\pi_n)^{f(S_t,\delta_{i,t})}. \tag{3.9}$$

The identification strategy used here approximates the popularity of numbers by using variables indicating the number of small (1-12), medium (13-31) and large (32-49) numbers which appear in the winning configuration. The more popular small and medium numbers are therefore expected to be negatively related to the probability of a rollover occurring and large numbers to positively related. The set of numbers to be represented by the 'small', 'medium' and 'large' instruments used here reflects their occurrence in calendars. This deliberately

[53] Recall that the jackpot in draw *t* is defined by $J_{kt} = \rho_k\left[(1-\tau)S_t - 10W_{3,t}\right]$.

exploits just one heuristic employed by players to choose birthdays or other memorable as a 'strategy' to choose numbers which is discussed, for example, in Forrest *et al* (2002). Specifically, the set of numbers in 'small' appear as both days and months, 'medium' numbers can only be used to represent days of the month, and is adjusted to account for the variable number of days in each month (and leap years), and 'large' numbers do not appear as a day or month in the calendar. Specifically, these instruments are constructed according to:

$$\begin{aligned} \Delta_{small,t} &= \sum_{i=1}^{12} \delta_{i,t} \\ \Delta_{medium,t} &= \sum_{i=1}^{28} \delta_{i,t} + \frac{45}{48}\delta_{29,t} + \frac{11}{12}\delta_{30,t} + \frac{7}{12}\delta_{31,t} \\ \Delta_{large,t} &= \sum_{i=32}^{49} \delta_{i,t} \end{aligned} \tag{3.10}$$

Where

$$\delta_{i,t} = \begin{cases} 1, if\ i\ is\ among\ the\ winning\ numbers\ in\ draw\ t, \\ 0, otherwise \end{cases} \tag{3.11}$$

This approach is preferable to the more obvious solution of using 49 number dummies for two reasons. Firstly, it overcomes the problem that, individually, so many dummies would make weak instruments of rollovers because of the limited frequency each number appears in the winning configuration. Secondly, though this approach is less precise, the quantity of small, medium, and large numbers, as defined in equation (3.10), in the winning configuration must still be exogenous to sales since this winning combination is unknown *ex ante* by players, and Figure 3.3 shows that they sufficiently capture the effects of conscious selection on the probability of a rollover occurring. That is, these variables are both exogenous to the level of sales and sufficiently correlated with price, making them useful as instruments.

Figure 3.2 shows the theoretical and actual proportion of draws with 1-6 numbers greater than 31 among the winning combination. This figure highlights how the quantity of 'large' winning numbers for any given draw is indeed random by comparing the actual distribution of how many large numbers make up the winning set with the theoretical. This randomness is important to illustrate since it adds weight to the argument that players have no information *ex ante* about which numbers will form the winning set for any given draw. With no information available about the winning numbers in any given draw, the decision by players to purchase a ticket is then clearly uncorrelated with the winning numbers. Similar graphics can easily be drawn for small and medium numbers which highlight the same principle. However, it is possible to argue that price is dependent on the winning numbers because of the effect they have on the rollover probability arising from conscious selection. This is demonstrated in Figure 3.3 and Figure 3.4. Figure 3.3 shows that as more numbers greater than 31 are drawn, the likelihood of there being no winner (i.e. a rollover) increases, thus by the mechanics of equation (3.9) price must also increase. Similarly, the most popular numbers, 1-12, have the opposite effect on the rollover probability as can be seen in Figure 3.4.

To model conscious selection in this way is, admittedly, rather arbitrary and a niche literature exists which presents a broader discussion of number choice by players in lotto games. Some of these would likely only add to the strength of the method used here, whilst others would not be detected in the present approach.

It is well documented that individuals choose numbers which are, on average, smaller than one would expect from a random and not just because they form calendar dates. Boland and Pawitan (1999) is just one example in which 234 university students are explicitly asked to generate 6 *random* numbers from 1 to 42 (this is the design of the Irish Lotto). The average of all selections is found to be 2 less than the average expected from true random selection, and this bias would only improve the power of the instruments considered here.

Boland and Pawitan (1999) also find the separation between numbers chosen by the participants is larger than what would be expected from random and there is a clear reluctance to choose consecutive numbers in particular. Conscious selection of this form is entirely separate to the approach used here and would not be captured by using ‘small’, ‘medium’ and ‘large’ indicators as defined above. Wang *et al* (2016) present evidence of both of the patterns identified in Boland and Pawitan (1999) using individual-level data for players of the Dutch Lotto and roulette data from Holland Casino. They also find that the layout of numbers presented to players influences their choices, with a preference for numbers located towards the centre of the choice form as well as special patterns. Again, neither of these behavioural biases would be detected in the model of conscious selection used here.

Figure 3.2: Theoretical and actual proportion of draws with n numbers > 31 drawn

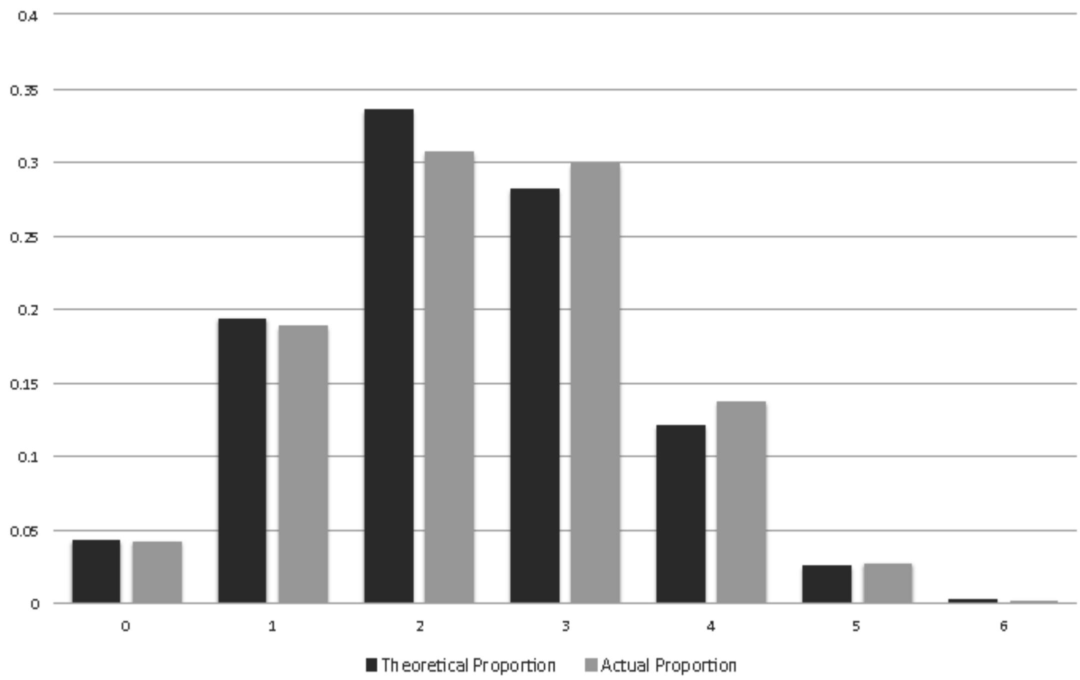


Figure 3.3: Proportion of draws which roll-over by number of balls > 31 drawn

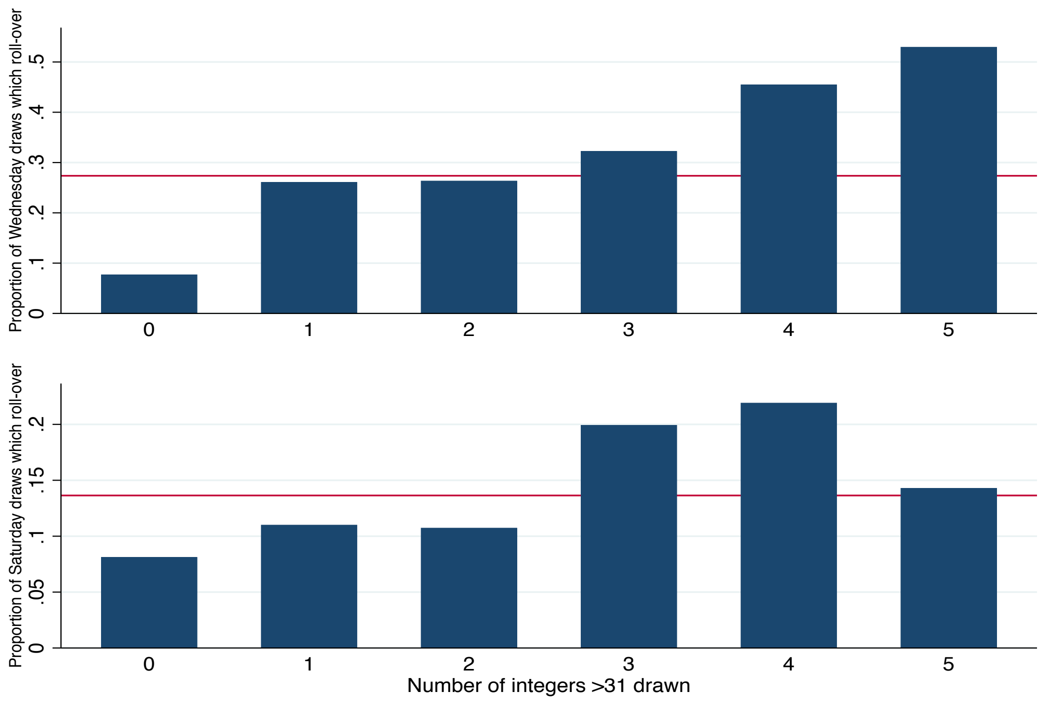
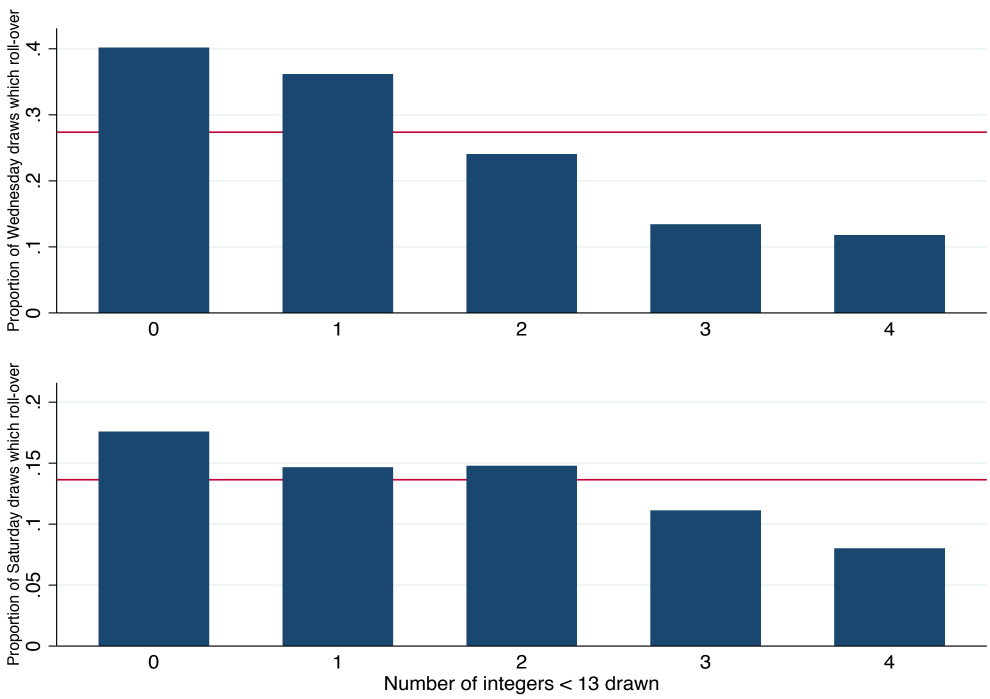


Figure 3.4: Proportion of rollover draws by number of balls 1-12 drawn



Suetens *et al* (2016) expands the set of possible conscious selection heuristics in a recent study of individual-level Danish lotto data who benefit from being able to observe directly the precise proportion of players who select each number (1-36 in the Danish game) over 28 draws[54]. They find that players are susceptible both to the "gambler's fallacy" (that players believe in frequent reversals) and the "hot hand fallacy" (the belief in the continuation of a sequence). They show that amongst those players who select their own numbers (it is possible to have the sales terminal choose the numbers randomly), but who vary the choices they make, are less likely to select a number which was drawn in the immediately preceding draw, but more likely to select a number which has appeared repeatedly in recent draws. Only a third of their full sample (and less than half of the subset of frequent players) change their number selection, with the remainder choosing the same numbers each week.

Experiments using indicators for consecutive numbers following the findings of Boland and Pawitan (1999) and Wang *et al* (2016), in conjunction with the small, medium and large indicators, made no qualitative difference to the estimates presented in Section 4.6. The method of modelling consecutive numbers using birthdays appears sufficient to capture the effects of

[54] Similar to Farrell *et al* (2000), Suetens *et al* (2016) also find that the most popular balls are numbered less than 13, with monotonically decreasing popularity of ball as their number increases.

conscious selection on the likelihood of a rollover occurring. Therefore, it is this strategy which forms the focus of the remainder of this chapter. First and second-stage regressions of the price and rollover model using consecutive numbers as an additional instrument can be found in the Appendix 3.10.5.

Instrumenting rollover size, which is also endogenous since it is determined by autocorrelated ticket sales, is done by exploiting the mechanism through which fixed prizes deduct from the pari-mutuel prize fund. A variable which contains the exogenous variation in the number of players who match three numbers of the winning set is used. The rationale here hinges on the fact that these prizes are paid before money is allocated to the jackpot prize pool which would ultimately form the rollover size if there are no winners. An unexpectedly high number of winners of the £10 fixed prize would reduce the amount of money in a rollover, should one occur, so this variable is expected to be negatively correlated to rollover size.

3.5 Data

With the exception of Section 3.8, the analysis in this chapter uses data which contains information on ticket sales, prize pools, the number of winners of each pool, rollover sizes, the date, and the winning combinations drawn for 1,739 draws of the UK National Lottery between the 5th February 1997 and the 2nd October 2013[55]. Prior to the 5th February 1997, the UK National Lottery was drawn only once per week – on Saturdays – with the introduction of Wednesday draws on this date. Consequently, this date was chosen as the start of the sample to avoid complications in the time series analysis caused by a change in the frequency of draws. On the 5th October 2013, the operator redesigned the UK game by changing the sticker price of a ticket to £2 and restructured the shares of sales allocated to the individual prize pools. Thus, the draw immediately prior to this date offers a natural termination point for our data. As such the first 117 draws from the dataset, and all of the draws since the 2013 game redesign are omitted from the sample in the main analysis.

Table 3.1 presents summary statistics for this dataset. There are 870 Wednesday draws and 869 Saturday draws, of which 369 were rollovers – a proportion of 21.22%. As expected from Figure 3.1, price is lower for rollover draws than for non-rollover draws. This lower price for Saturday non-rollover draws than similar draws on Wednesday can be explained by sales

[55] Section 3.8 uses more data on draws made after the game re-design in 2013 to evaluate whether the changes in design had a positive effect on sales

on Saturday being significantly larger, resulting in the likelihood of a rollover occurring being much smaller.

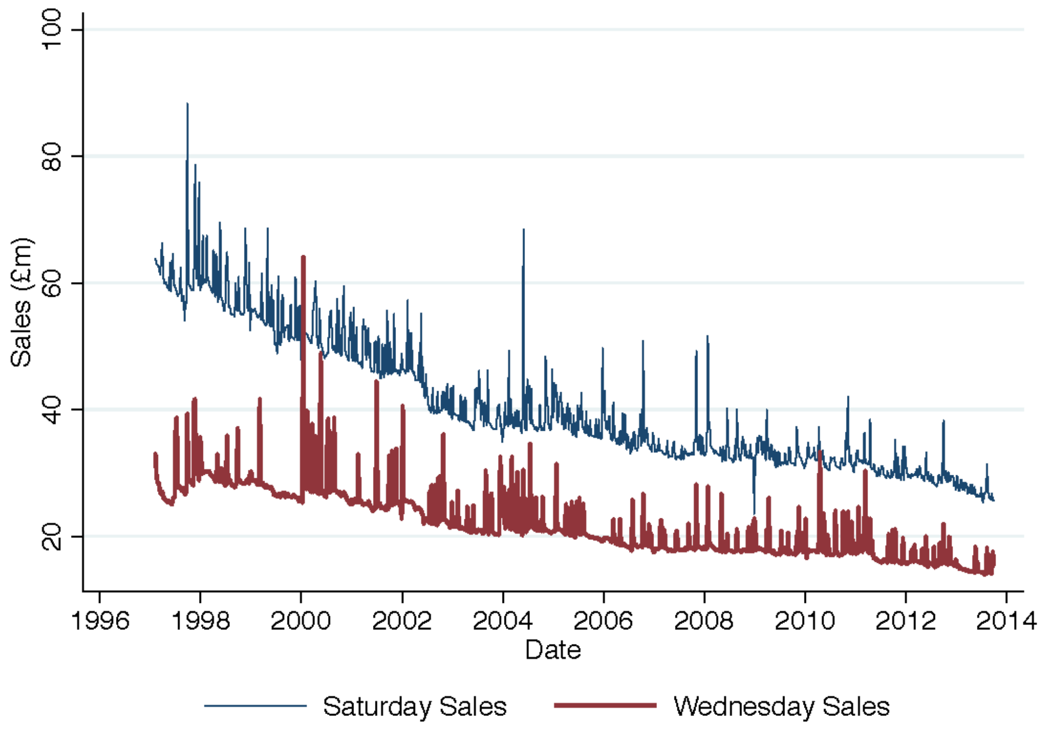
Table 3.1: Summary statistics - weekly averages

Saturday	Non-Rollover	Rollover	All
No. Draws	617	252	869
Sales (millions)	40.673	41.152	40.812
Price (£)	0.5100	0.4272	0.4860
Three-ball winners	716,748	716,383	716,643
Three-ball winners (proportion)	0.0175	0.0173	0.0175
Rollover size (£m)	0	3.622	1.051
Wednesday			
No. Draws	752	117	870
Sales (millions)	21.389	26.129	22.032
Price (£)	0.5329	0.3124	0.5030
Three-ball winners	380,648	463,059	391,864
Three-ball winners (proportion)	0.0178	0.0176	0.0178
Rollover size (£m)	0	6.713	0.912

The effect of rollovers on average Saturday sales is noticeably small, just 1%, while the effective price falls by 16% which, taken at face value, would imply a price elasticity of just -0.06. In contrast, the effect of a rollover-enhanced jackpot on a Wednesday induces a 22% rise in sales from a 41% fall in the effective price implying an elasticity of -0.5. There are two explanations for this difference. Firstly, Saturday rollover draws are enhanced by a proportion of sales in the previous Wednesday which are, on average, around half of usual Saturday sales, so the rollover size in a Saturday draw is correspondingly smaller. Secondly, as can be seen in Figure 3.5, sales fall much faster over time in the Saturday draw game compared to the Wednesday draw game. Thus, the difference in average sales for rollover and non-rollover draws over the entire sample is masked by falling sales overall.

Figure 3.5 illustrates the declining trend of sales in both Saturday and Wednesday draws over the 16-year period covered by the data from around 90 million tickets sold per week in 1997 to just over 40 million in 2013. These declining sales figures no doubt contributed to the decision to redesign the game in October 2013 and second design change occurred in 2015. Peaks in the graph highlight the impact that rollovers (and double, triple and even quadruple rollovers) have on ticket sales. It is the effect of rollovers on sales (either directly or via the influence on price) that are the focus of our attention in the model used in this chapter. Nonetheless, the time series nature of the data may raise concerns about stationarity. Augmented Dickey-Fuller tests (Dickey and Fuller, 1979) for both Wednesday and Saturday sales series strongly rejects the presence of a unit root[56].

Figure 3.5: Draw-by-draw lotto sales from February 1997 to September 2013



3.6 Price model estimates

This section presents estimates of the demand model outlined above which assumes that it is rollover induced variation in price that drives sales variation. The theory outlined in Section 3.4 suggests that the effective price is endogenous to sales such that least squares will produce biased estimates. The identification strategy used here relies on the effect conscious selection by players has on the likelihood of a rollover occurring and the effect of random variation in fixed-prize winners on the rollover size – both of which impact price (via rollovers)

[56] ADF unit root (with trend) test statistics for Saturday and Wednesday sales series are -16.259 and -22.937, respectively, with 1% / 5% / 10% critical values of -3.960 / -3.410 / -3.120. ADF unit root (without trend) test statistics for Saturday and Wednesday sales series are -7.665 and -12.763, respectively, with 1% / 5% / 10% critical values of -3.430 / -2.860/ -2.570.

but are not correlated with sales. Fitted values of both rollover probability and size from a model including these effects are then used to obtain an instrumented price, rather than instrumenting price directly. This circumvents the issue of both rollover size and probability entering the definition of price in a clearly nonlinear way in equation (3.9). Moreover, this allows the separation of the effects of the instruments on the size and probability of rollovers.

Heckman's (1979) two-step selection procedure offers a neat solution to instrumenting the rollover size and probability simultaneously and is what is used here. Moreover, the Heckman selection model is preferred over a Tobit estimation since we expect different effects of our first-stage covariates on the frequency of rollovers and rollover size. In particular, lagged sales are expected to have a negative effect on the rollover probability but a positive effect on rollover size, should one occur, which would not evident in a Tobit model. The selection equation of Heckman's procedure involves using the number indicator variables of equation (3.10) from the previous draw, along with all exogenous variables from the second stage, as determinants of a dummy variable for draw *t* being a rollover draw. Fitted values from this are used as the rollover probability in draw *t-1*. The second stage of this application of Heckman's selection model is used to obtain an estimate of the effect of unexpected variation in 3-ball prize winners on rollover size in draw *t*. Specifically, the following model is estimated:

$$R_t = \begin{cases} R_t & if\ R_t^* > 0, \\ 0 & otherwise \end{cases} \tag{3.12}$$

where R_t is the observed rollover, R_t^* is the latent rollover size and is determined by,

$$R_t^* = \gamma_0 + \gamma_1 S_{t-1} + \gamma_2 S_{t-2} + \gamma_3 \left(W_{3,t-1} - \pi_3 S_{t-1}\right) + \eta_t \tag{3.13}$$

where γ_j $(J = 0, \dots, 3)$ are parameters to be estimated. The rollover probability is modelled as,

$$\Pr(R_t > 0) = \alpha_0 + \alpha_1 S_{t-1} + \alpha_2 S_{t-2} + \alpha_3 \Delta_{small,t-1} + \alpha_4 \Delta_{medium,t-1} + \alpha_5 \Delta_{large,t-1} + \nu_t. \tag{3.14}$$

where α_j $(J = 0, \dots, 5)$ are again parameters to be estimated and R, S and Δ's are as defined previously.

The transformed variable, $W_{3,t-1} - \pi_3 S_{t-1}$, in the rollover size equation is the exogenous component of the variation in the number of 3-ball prize winners (the only fixed prize winners in the UK game during our sample period) from the previous draw. The number of 3-ball winners is itself dependent on the level of sales, which is autocorrelated, insofar as the expected number of winners is simply the probability of any given ticket winning the prize

(1 in 57) multiplied by the number of tickets sold. Thus, the number of 3-ball winners is itself invalid as instrument. However, random variation in the number of 3-ball winners in the previous draw, $W_{3,t-1} - \pi_3 S_{t-1}$, is purged of the relationship with lagged sales and, appealing to the work of Conley *et al* (2012), here assume that this transformed variable is "plausibly exogenous" to facilitate the use of this random variation in prize winners as an instrument. This application of Heckman's selection model is identified by the role that exogenous variation in 3-ball winners has in determining the size of a rollover should one occur, but that it has no bearing on the probability of there being no winners which only depends on the level of sales and winning numbers in each draw due to conscious selection.

The fitted values of $\Pr(R_t^* > 0)$ are used as the rollover probability in draw *t-1*, and R_t^* from the second stage of Heckman's selection model is used in place of R_t in the price equation to obtain an instrumented price variable, $\widehat{P_t}$. This instrumented price is then used in the following sales model:

$$S_t = \beta_0 + \beta_1 S_{t-1} + \beta_2 S_{t-2} + \beta_3 \widehat{P_t} + \psi X_t + \varepsilon_t. \tag{3.15}$$

Table 3.2 presents the resultant first-stage Heckman model estimates. The bottom panel of Table 3.2 refers to the selection equation of the Heckman model. The estimates indicate that drawing one extra medium ball in the winning combination, rather than a small number, increases the rollover probability. Drawing one more 'large' number instead of a number between 1-12 in the winning combination increases the likelihood of a rollover even further. These findings are consistent with the those of Farrell *et al* (2000) that numbers below 13 tend to be the most popular amongst players and that numbers greater than 31 are chosen least of all. These effects are highly significant for the often rollover-enhanced Saturday draws. Drawing an extra 'medium' or 'large' number, instead of an additional 'small number' on a Wednesday increases the likelihood that Saturday's draw will have a rollover-enhanced jackpot. For Wednesday the effect is less pronounced, but still significant for 'large' numbers. This can be explained by Saturday ticket sales being sufficiently high that even unpopular combinations are chosen by at least one player, thus making conscious selection more difficult to detect. Nonetheless, the direction of the coefficients and their increase in magnitude from medium to large dummies is encouraging. Moreover, the significance of the Saturday estimates, and of the large numbers dummy in the Wednesday estimates, encourages the use of this idea as a powerful instrument.

The top panel of Table 3.2 reports estimates of rollover size conditional on a rollover occurring. For both Wednesday and Saturday draws, exogenous variation in the number of three-ball winners have negative and significant coefficients. An extra "unexpected" winner of the 3-ball prize implies a reduction in rollover size of $10\rho_6$ (i.e. £6.66). The coefficient estimates for both Wednesday and Saturday are not statistically different from this theoretical value. Coefficients on lagged sales are also consistent with what one would expect. Extra ticket sales in the preceding draw (S_{t-1}) increases the size of the rollover conditional on one occurring. Extra sales in the corresponding draw from the preceding week (S_{t-2}) would have reduced the likelihood of a rollover occurring in draw $t-1$, and thus sales (and any rollover from that draw) for that draw would have been somewhat lower than usual on average.

Table 3.2: Heckman selection model estimates for rollover size and probability

Dependent Variable	Saturday $R_t^{Sat} \mid R_t^{Sat} > 0$	Wednesday $R_t^{Wed} \mid R_t^{Wed} > 0$
$W_{3,t-1} - \pi_3 S_{t-1}$	-4.911***	-6.695***
	(1.2904)	(1.0356)
S_{t-1}	1.029***	0.722***
	(0.0330)	(0.0770)
S_{t-2}	-0.091***	-0.106*
	(0.0209)	(0.0571)
Constant	-12.539***	-41.083***
	(1.946x10^6)	(5.7091)
Selection Equation	Pr ($R_t^{Sat} > 0$)	Pr ($R_t^{Wed} > 0$)
$\Delta_{medium,t-1}$	0.189***	0.026
	(0.0553)	(0.0414)
$\Delta_{large,t-1}$	0.338***	0.109**
	(0.0543)	(0.0529)
S_{t-1}	-0.079***	-0.0290
	(0.0216)	(0.0215)
S_{t-2}	-0.000	-0.010
	(0.0151)	(0.0207)
Constant	0.223	-0.803
	(1.4474)	(1.8896)
Observations	868	868
Censored observations	616	750
λ	-0.0561	2.4478***
$\chi^2_{(1)}$ test of $\lambda = 0$	0.06	17.01
P-value	0.8131	0.000

Notes: Robust standard errors in parentheses. ***/**/* denotes statistical significance at 1%,5%, 10% confidence. Trend and seasonality controls omitted from reporting. Sales variables in millions.

Finally, the inverse Mill's ratio, λ, provides a test of the correlation between fitted values of ν_t and η_t. If the coefficient on λ is 0 then rollover size would be uncorrelated with rollover probability. The estimate for Wednesday draws rejects the hypothesis that the two are uncorrelated, thus also rejects OLS and Tobit specifications. Whilst the Saturday estimate of λ fails to reject the null that the two equations are correlated, appealing to the theory laid out above and the encouraging rejection for Wednesday draws, this procedure still appears to be the best available.

Table 3.3 reports estimates of the demand equation using OLS (columns 1 and 2) and using instrumented price (columns 3 and 4) per the fitted values from the estimates in Table 3.2. OLS estimates show a significant, positive relationship between current sales and sales for the preceding two draws. This evidence verifies the hypothesis that sales are autocorrelated and that rollovers are indeed invalid instruments since their frequency and size are simply functions of past sales. OLS estimates suggest that a £0.10 decrease in price (which is not uncommon for rollovers of around £4m) would yield a £4.610m (£3.077m) increase in sales for Saturday (Wednesday) draws. These are consistent with estimates of long-run price elasticity of -0.556 for Saturday and -0.750 for Wednesday games. Both of these price elasticity estimates are significantly different from -1, indicating that the design of the game in the sample period could increase sales revenue by making the game more expensive to play[57].

Elasticity estimates derived from using instrumented price, however, suggest different conclusion. With direction and magnitude of coefficients on lagged sales broadly similar, estimates of the price coefficients suggest those obtained from OLS are attenuated towards 0. The price elasticity estimate for Saturday remains qualitatively the same at -0.636, suggesting that revenue could be increased by making the game more expensive. However, for the Wednesday game a price elasticity estimate of -1.472 suggests that the game is over-priced and revenues could be increased by making the game more attractive to play.[58] Assuming a price elasticity of -1 would maximise revenues for the lotto monopolist, using the parameter estimates of Table 3.3 and rearranging the elasticity equation (3.7) suggests, ceteris paribus, that increasing the price of Saturday draws from £0.4860 to £0.7711 and reducing the price of Wednesday tickets from £0.5030 to £0.3625 would increase revenues from the games. The most obvious method of achieving this would be to increase the take-out rate of Saturday games and, using some of the extra net revenues, to subsidise Wednesday draws. A static set of equations indicate that removing £6.2m per draw from the Saturday prize pool and using £3.1m

---

[57] χ^2 tests of price elasticity OLS estimates' statistical difference from -1 yields values of 85.48 (p=0.000) for Saturday and 27.47 (p=0.000) for Wednesday.

[58] χ^2 tests of price elasticity IV estimates' statistical difference from -1 yields values of 57.97 (p=0.000) for Saturday and 17.37 (p=0.000) for Wednesday.

of this in Wednesday prize pools would increase the game's weekly net revenues by the difference – £3.1m.

Table 3.3: Second-stage estimates of lotto demand

	OLS		**Control Function using Heckman First-Stage**	
Dependent variable	S_t^{Sat} **(1)**	S_t^{Wed} **(2)**	S_t^{Sat} **(3)**	S_t^{Wed} **(4)**
S_{t-1}	0.114**	0.119***	0.133**	0.0643*
	(0.0576)	(0.0391)	(0.0546)	(0.0332)
S_{t-2}	0.107***	0.0657***	0.102***	0.0796**
	(0.0288)	(0.0318)	(0.0284)	(0.0321)
P_t	-41.677***	-30.768***	-	-
	(3.6558)	(2.1264)	-	-
$\widehat{P}_t$	-	-	-47.283***	-55.856***
	-	-	(3.5934)	(4.6827)
Constant	77.755***	28.128***	79.771***	45.237***
	(3.5936)	(2.3921)	(3.5393)	(2.5605)
LR Elasticity	-0.5561***	-0.7499***	-0.6416***	-1.4718***
	(0.0481)	(0.0475)	(0.0475)	(0.1172)
Durbin-Watson d-stat	1.983	1.895	2.030	1.951
ARCH LM test (H_0: no ARCH effects)	0.087	0.590	0.097	0.153
ARCH LM test (p-value)	0.7682	0.4424	0.7557	0.6955
AIC	4109.37	3468.07	4060.98	3699.79
Observations	868	868	867	868
R^2	0.943	0.898	0.946	0.867

Notes: Robust standard errors in parentheses. ***/**/* denotes statistical significance at 1%,5%, 10% confidence. Trend and seasonality controls omitted from reporting. Sales variables in millions.

In addition to the stationarity of sales data for both Saturday and Wednesday discussed briefly in Section 3.5, extra time series diagnostics are displayed in the footer of Table 3.3. Durbin-Watson statistics (Durbin and Watson, 1950) for Saturday and Wednesday models estimated with OLS and instrumented price show no evidence of autocorrelation in the error terms. Furthermore, Engel's (1982) Lagrange Multiplier safely fails to reject the null hypothesis that no ARCH effects are present in the residuals, even at the 10% level.

3.7 Reduced form model of the effect of rollovers

A frequent criticism of modelling lotto demand using effective price models is that their foundations lie in expected utility theory which is notoriously ineffective at explaining why individuals simultaneously take risks (via gambling) and hold insurance (Forrest *et al*, 2002). Moreover, these models assume that prize sizes – particularly jackpot prizes – only affect demand for lotto games through their effect on the expected value of winnings, which is simply an average of the prize distribution. However, there is a further thread to the literature that suggests that gambling demand responds to higher moments of the prize distribution. In particular, the idea that gamblers are positively motivated by skewness in the prize distribution is commonplace (see, for example, Golec and Tamarkin, 1998). Cain *et al* (2002) show that the Golec and Tamarkin racetrack results fail to control for the collinearity between moments. Walker and Young (2001) is not subject to this criticism in their analysis of lotto sales which show that higher moments of the prize distribution have a significant effect on sales.

A theoretical rationale of a preference for skewness is implicit in Prospect Theory where the values associated with risky prospects are multiplied by decision weights which "… measure the impact of events on the desirability of prospects, and not merely on the perceived likelihood of these events" (Kahneman and Tversky, 1979, p.280). The theory suggests that individuals tend to overweight low probability events and underweight high probability events when making decisions in the face of uncertainty. This tendency of players to overestimate the chance of low probability events (longshots) occurring may be sufficient to make unfair gambles attractive. Quiggin (1991) uses a rank-dependent utility function to explain why risk-averse people might play unfair gambles if such games comprise a large number of smaller prizes and a few large prizes, which is how most lottery games are structured. Under this argument, large prizes might be particularly relevant in determining sales. This motivates a specification where variation in the largest prize is the proximate determinant of sales variation. Such a specification has been used by Forrest *et al* (2002) so this section replicates their work but in the context of using conscious selection as the identification strategy. Moreover, their work is extended to include a fully flexible semi-parametric specification as well as higher order polynomial transformations of the rollover size to account for the fact that players may be motivated by prizes beyond the effect of the average size of the prize distribution. In particular, models which are cubic in rollover size are considered.

Given the large size of jackpots relative to other prizes, variations in top-tier prize pools have a much larger effect on the higher moments than the average (expected) value of a lottery ticket. This may explain why Cook and Clotfelter (1993) observe that "bettor's evaluation of a lotto bet tends to be more sensitive to the size of the jackpot than the objective probability of winning" (p. 638) and that 'bigger is better' when it comes to the demand for lotto prizes.

Forrest *et al* (2002) evaluate the extent to which jackpot models are able to explain ticket sales relative to effective price models. Their jackpot model predicts that an additional £1 million in jackpot size would increase sales between £22,000 (Wednesday) and £53,000

(Saturday) conditional on the total prize pool remaining the same. Using rollover size in their identification strategy, they find that jackpot models yield significantly higher adjusted R^2 statistics for both draw days compared to the corresponding price models. Whilst this would tentatively suggest that their jackpot model may be superior to existing effective price models, non-nested tests were inconclusive in determining which model should be preferred.

This section is concerned with estimating the following:

$$S_t = \beta_0 + \beta_1 S_{t-1} + \beta_2 S_{t-2} + \beta_3 \widehat{R_t} + \beta_4 \widehat{R_t^2} + \beta_5 \widehat{R_t^3} + \gamma X_t + \varepsilon_t. \tag{3.16}$$

Since rollover size is endogenous to sales, causal estimation of equation (3.16) is again reliant on IV techniques. Fortunately, it is possible to recycle the identification strategy outlined in Section 3.4 and the Heckman selection model estimates presented in Section 3.6. Rather than imposing the strict functional form of expected price, estimation of this rollover model simply includes the predicted values (and their square and cube transformations) of rollover size, $\widehat{R_t}$, from the Heckman selection model in place of R_t.

Table 3.4 presents OLS and instrumented rollover estimates of this model for both Wednesday and Saturday games[59]. Similar to the price model estimates, OLS shows a significant, positive relationship between current and lagged sales in both the Saturday and Wednesday games. For Wednesday draws the higher order expansions of rollover size are statistically significant, with sales responding positively to R_t^2 and negatively to R_t^3. An F-test for joint significance suggest that coefficient estimates for Saturday games suggest that R_t^2 and R_t^3 are, however, insignificant. These OLS estimates suggest that an £1m increase in rollover size would induce an increase in sales by £1.1m (£0.5m) for a given Saturday (Wednesday) draw. Since the operator returns approximately half of these extra sales in prizes, it would appear that augmenting the jackpot prize would be an ineffective way of increasing revenue for good causes.

When controlling for endogeneity in the size of the rollover, the relationship between current sales and sales for the immediately preceding draw is negative and significant for both Wednesday and Saturday games. This suggests that the positive relationship between current

[59] Testing revealed that semiparametric estimation can be approximated by a parametric estimation of a regression including cubic expansion of rollover size. Hence, only regression estimates of this specification are reported here. Statistics from these tests are reported in Table 3.6. Estimates of models which are linear and quadratic in rollover size are reported in Appendix Section 3.10.

and past sales in the price model is erroneous and was not being detected because of the strict functional form. Similarly, the positive coefficients from OLS estimates of the rollover model may be due to not controlling for the correlation between rollover size and lagged sales. F-tests of the joint significance of $\widehat{R_t^2}$ and $\widehat{R_t^3}$ suggest that the price model was not detecting the effects of higher moments of prizes, which were instead being captured by higher sales in the previous draw. The instrumented estimates suggest an increase in rollover size of £1m would increase Saturday sales by approximately £1.8m and Wednesday sales by £0.5m. In both cases, OLS again appears to be downward biased, though the magnitude of this bias seems greater for Saturday draws than it is for Wednesday draws.

Table 3.4: OLS and Heckman-instrumented estimates of rollover induced ticket demand

	OLS		**Heckman First-Stage**	
Dependent variable	S_t^{Sat}	S_t^{Wed}	S_t^{Sat}	S_t^{Wed}
	(1)	**(2)**	**(3)**	**(4)**
S_{t-1}	0.107**	0.068***	-1.222***	-0.361***
	(0.0444)	(0.0212)	(0.3656)	(0.1110)
S_{t-2}	0.119***	0.096***	0.247***	0.120***
	(0.0262)	(0.0270)	(0.0498)	(0.0461)
R_t	1.234***	0.400*	-	-
	(0.3165)	(0.2199)	-	-
R_t^2	-0.096	0.095**	-	-
	(0.1236)	(0.0429)	-	-
R_t^3	0.008	-0.002	-	-
	(0.0092)	(0.0014)	-	-
$\widehat{R_t}$	-	-	1.853***	0.519**
	-	-	(0.4007)	(0.2123)
$\widehat{R_t^2}$	-	-	-0.033	0.029
	-	-	(0.0200)	(0.0387)
$\widehat{R_t^3}$	-	-	0.000	-0.001
	-	-	(0.0004)	(0.0012)
Constant	55.127***	14.779***	71.86***	39.54***
	(2.4441)	(1.6803)	(5.6343)	(7.2570)
Joint F-test $R_t^2, R_t^3 = 0$	0.43	38.04***	5.96***	4.22**
Prob>F	0.6503	0.0000	0.0027	0.0151
Durbin-Watson d-stat	1.914	1.605	1.959	1.988
ARCH LM test (H0: no ARCH effects)	0.032	2.120	0.040	0.024
ARCH LM test (p-value)	0.858	0.1454	0.8406	0.8777
AIC	3941.94	2858.93	4455.42	4373.87
Observations	868	868	868	868
R^2	0.953	0.9496	0.916	0.711

Notes: Robust standard errors in parentheses. ***/**/* denotes statistical significance at 1%,5%, 10% confidence. Trend and seasonality controls omitted from reporting. Sales and rollover size variables in millions.

Similar to the price model estimates in the preceding section, Durbin-Watson statistics (Durbin and Watson, 1950) for Saturday and Wednesday models estimated with both OLS and instrumented rollovers show no evidence of autocorrelation in the error terms. Engel's (1982) Lagrange Multiplier test again safely fails to reject the null hypothesis, even at the 10% level, that ARCH effects are present in the residuals, though OLS estimates for Wednesday sales do come noticeably close to 10% significance (p-value = 0.1454).

In order to evaluate whether this reduced form model is preferred to the price model of Section 3.6, non-nested hypothesis testing developed by Cox (1961), Cox (1962), and later by Pesaran (1974) is used. This test compares two models, M_1 and M_2, tests two hypotheses of the form:

H_0: M_1 superior to M_2

H_0: M_1 not superior to M_2.

The test is then repeated where the null hypothesis is re-defined as M_1 superior to M_2. Determining a superior model using non-nested hypothesis testing requires both rejecting the null of the inferior model *and* not rejecting the null hypothesis of the superior model. As such, there are four possible outcomes: a rejection of M_1, a rejection of M_2, failure to reject both, and a rejection of both. The latter two outcomes would be inconclusive whilst the former would suffice for determining model superiority.

Table 3.5 presents the results from the Cox-Pesaran test. This test is repeated for both Saturday and Wednesday draws and for both OLS and instrumented estimation regimes.

Table 3.5: Cox-Pesaran non-nested hypothesis testing of model preference

	OLS		**Heckman First-Stage**	
Saturday	**z-statistic**	$p > \|t\|$	**z-statistic**	$p > \|t\|$
H_0: Rollover model preferred	1.88	0.030	-108.75	0.000
H_0: Price model preferred	-13.95	0.000	-2.46	0.007
	OLS		**Heckman First-Stage**	
Wednesday	**z-statistic**	$p > \|t\|$	**z-statistic**	$p > \|t\|$
H_0: Rollover model preferred	-6.51	0.000	-482.09	0.000
H_0: Price model preferred	-55.48	0.000	-2.53	0.006

For Saturday games, OLS estimates suggest, albeit only at the 1% level, that the rollover model is superior to the price model. When controlling for endogeneity, the test is inconclusive in testing the competing models of Saturday ticket sales. For Wednesday sales, non-nested testing of both OLS and instrumented estimates are inconclusive. The rejection of both of the competing models mirrors the results found in Forrest *et al* (2002) who suggest that an improvement to both models would to allow for a more flexible specification of the role of jackpot prizes, which encourages proceeding with a semi-parametric estimation routine.

This section has so far assumed that g, from equation (3.17) below, is somewhat arbitrarily cubic in R_t, whilst the price model estimated in Section 3.6 assumed that g mediates the effect of rollover size solely through price. Allowing R_t to enter non-parametrically overcomes the limitations inherent to imposing such arbitrary constraints, regardless of whether those constraints are guided by theory. Specifically, the semi-parametric technique used here estimates directly the following specification:

$$S_t = \beta' \boldsymbol{X}_t + g(R_t) + \varepsilon_t. \tag{3.17}$$

This model of sales is estimated using methods developed by Robinson (1988). Following contributions made by Blundell *et al* (1998) and Blundell and Powell (2003) to Robinson's estimation method in the presence of endogenous regressors, rollover size is instrumented per the Heckman first-stage in Section 3.6[60], and the residual of this first stage is included in the parametric component, $\boldsymbol{X}_t$. Other controls which form the parametric component of this specification and include own and cross lagged sales and variables to control for trend.

Table 3.6 reports the coefficient estimates of the parametric component. As with the majority of estimates above, coefficients on both own and cross lagged sales are positive and highly significant for both Wednesday and Saturday draws. Moreover, the coefficient on the residual from the first stage, ρ, is also significant indicating the assumption that rollover size is endogenous is justified. Using a test developed by Hardle and Mammen (1993) reveals that both Wednesday and Saturday models are statistically different from a parametric fit which is linear in rollover size, and Saturday modelling is statistically different from parametric fits which are quadratic and cubic in rollover size at the 5% level. Estimates for Wednesday, however, are not statistically different from either the quadratic and cubic parametric rollover

[60] Details of this estimation procedure can be found in Appendix Section 3.10.

models. The difference in these tests between Wednesday and Saturday models are not surprising upon examination of Figure 3.6 and Figure 3.7 which are plots of the nonparametric estimates of the function g for Saturday and Wednesday draws, respectively. For Saturday, the relationship is highly non-linear, with modest increases in sales for single and double rollovers in the £0-10m range, but large increases in sales for treble rollovers, almost all of which are over £10m. For the Wednesday draw it can be seen that the non-parametric fit indeed looks close to quadratic, with an increasing gradient for larger rollover sizes.

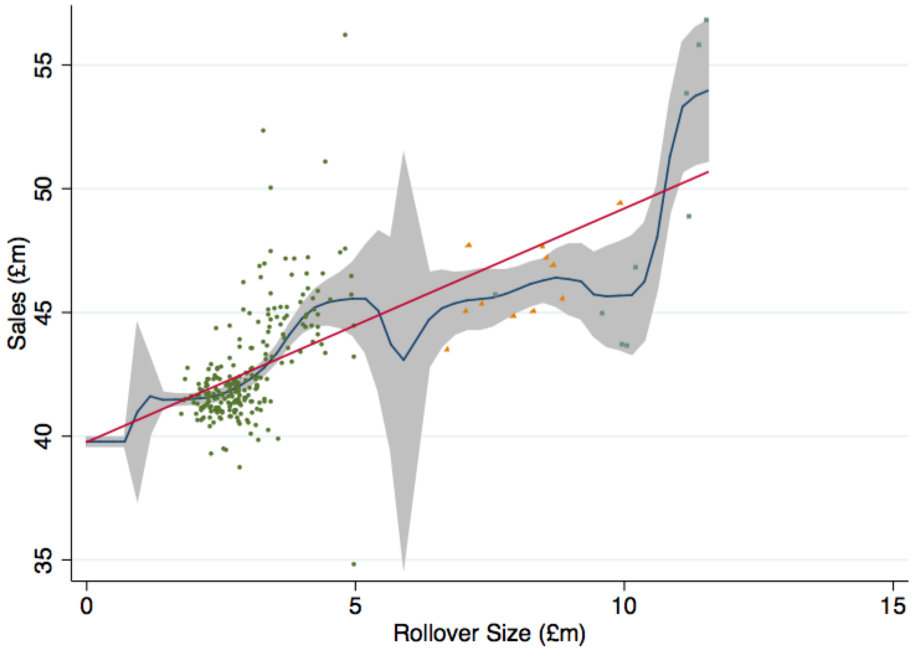
Differing estimates suggest that players in Wednesday and Saturday draws respond differently to increases in rollover sizes. Moreover, the non-linear relationship between rollover size and sales for both Wednesday and Saturday draws indicates that previous literature reliant solely on price variation is not capturing the true response of players' demand to changes in the prize distribution.

Table 3.6: Parametric coefficient estimates for semi-parametric models of sales

Dependent Variable	S_t^{Sat}	S_t^{Wed}
	(1)	(2)
S_{t-1}	0.186***	0.243***
	(0.0616)	(0.0567)
S_{t-2}	0.153***	0.220***
	(0.0248)	(0.0409)
ρ	-4.225***	-4.243***
	(1.6615)	(1.1715)
t-test vs parametric fit:		
(H_0: Models are not different)		
Linear	3.592***	4.069***
p-value	0.00	0.00
Quadratic	2.595**	0.524
p-value	0.04	0.63
Cubic	1.983**	0.335
p-value	0.04	0.96
R^2	0.9472	0.9477
Observations	845	852

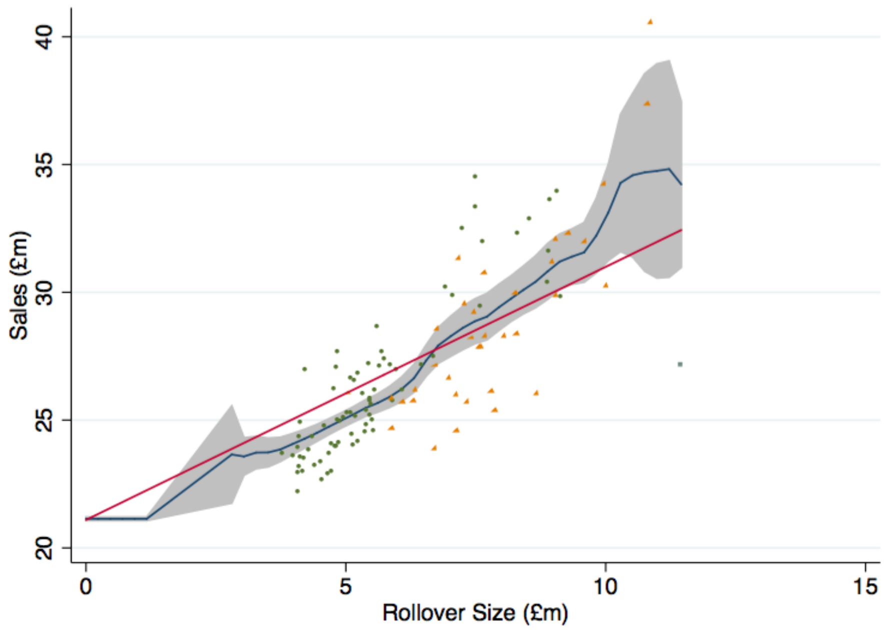
Notes: Robust standard errors in parentheses. ***/**/* denotes statistical significance at 1%,5%, 10% confidence. Trend and seasonality controls omitted from reporting. Sales variables in millions.

Figure 3.6: Non-parametric estimate of rollover size and sales for Saturday draws



Notes: 95% confidence interval. Red line is the linear fit. Green dots / yellow triangles / blue squares denote single / double / treble rollovers.

Figure 3.7: Non-parametric estimate of rollover size and sales for Wednesday draws



Notes: 95% confidence interval. Red line is the linear fit. Green dots / yellow triangles / blue squares denote single / double / treble rollovers.

3.8 Evaluating lotto re-designs

Changes made to the UK lotto game in October 2013 and October 2015 were likely in response to the dwindling sales which can be seen in Figure 3.5. The 2013 changes saw the sticker price increase from £1 to £2, a change in the share of prize money allocated to each prize tier which saw a larger share of the pari-mutuel prize fund being allocated to the jackpot, an increase in the fixed-prize awarded for matching 3 of the 6 winning numbers from £10 to £25, and 50 fixed raffle prizes of £20,000 for each draw. In 2015, the set of numbers from which winning combinations were drawn (and from which players could choose) was increased to 59, a prize of one free ticket to the following draw for matching two numbers was introduced, the number of £20,000 raffle prizes available each draw was reduced to 20 and a £1m raffle prize was added. Also in 2015, the cap on the number of consecutive draws for which the jackpot prize could have no winners was removed, with the operator instead allowing the prize to "roll-over" until it reached £50m; after which, if it had not been won again, would be shared between winners of the next-highest prize tier. This cap was lowered in August 2016 to £22m. The natural question to ask is whether the changes effective in rejuvenating sales and, if so, how effective were they?

The price model estimates from Section 3.6 suggested lotto sales could be increased if the Saturday draw, where demand is price inelastic, were made more expensive to play relative to the Wednesday draw, where demand is price elastic. Introducing raffle prizes from 2013 onwards could, at least in theory, have achieved precisely this. The 50 raffle prizes of £20,000 for each draw, which any given ticket is equally likely to win, are paid before prize money is shared amongst the pari-mutuel prize funds. Effectively, this change reduces the value of money made available to pari-mutuel prizes by £1m. This then influences the effective price by reducing the size of a rollover, should one occur.

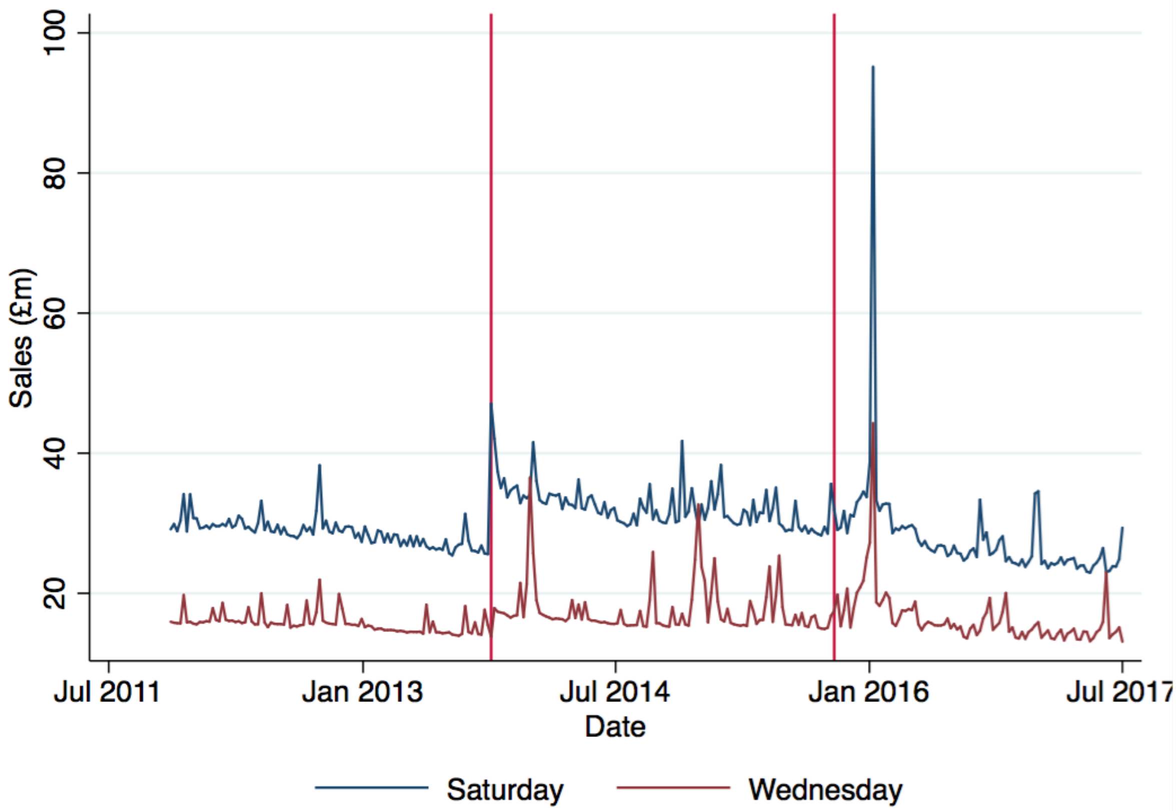
The implication of this change for effective price is two-fold. Firstly, should there be no winners of the jackpot prize in the current draw, the increase in the effective price (caused by money lost to the rollover) is smaller than it would otherwise have been if there were no raffle prizes awarded. Having no winners of the jackpot prize occurs more frequently on Wednesday draws than Saturday due to lower sales. Therefore, this smaller increase in effective price will be experienced more frequently in Wednesday draws and should decrease the average price of Wednesday draws relative to that of Saturday, compared to the old design. Secondly, the decrease in price for the subsequent draw is smaller than it would have otherwise been because the size of the rollover has been reduced. Again, this decreases the average price of Wednesday draws relative to Saturday because Wednesday have fewer jackpots enhanced by rollovers from Saturday than vice versa. Thus, this smaller decrease in price is experienced less frequently on Wednesdays than on Saturdays and will decrease the relative price of Wednesday draws on average. These two effects are amplified by the fact that the £1m taken from the pari-mutuel prize pool is a larger share of the pool for Wednesday draws than it is for Saturday draws. Moreover, the changes in 2015 saw even more money (£1.4m) being awarded in raffle prizes.

However, other changes implemented simultaneously in 2013 have an opposing effect on Wednesday prices relative to Saturdays. In particular, the share of the pari-mutuel prize pool allocated to the jackpot prize increased from around 50% to over 80% at the expense of prize money allocated to the lower pari-mutuel prize tiers. This increase in the share of prize money allocated to the jackpot means that the raffle prize money deducted from the pari-mutuel prize fund was unlikely to affect the average size of the jackpot and, by extension, the size of rollovers and their effect on effective price when they occur. Moreover, an increased frequency of rollovers arising from doubling the sticker price in 2013 (also doubling the effective price for both Wednesday and Saturday draws and fewer tickets being bought overall), and increasing the difficulty of the game in 2015, will likely lead to an increased effective price for *both* games.

Whilst theory can predict the direction of the effect of each of the design changes discussed above, little is known about the magnitude of the effect changing game-specific parameters has on the demand for tickets, which is not surprising given the unique way in which the price of lotto, rollovers, and sales are all endogenous to one another. Thus, given the conflicting effects that the simultaneous design changes are predicted by theory to have on the price of lotto tickets, whether or not they are effective in reviving lotto sales revenue is unclear. A naïve answer can be found by simply comparing sales figures for each draw before and after the changes.

Figure 3.8 illustrates the sales figures for Wednesday and Saturday draws of the UK lotto from November 2011 to July 2017 – allowing some comparison to be made between ticket sales for the pre-2013, post-2013 and post-2015 designs. The red vertical lines correspond to the dates on which the first draws took place of the new designs.

Figure 3.8: Draw sales for the UK lotto (Nov 2011-Jul 2017)



The 2013 reform appears to have had an immediate impact on sales figures, particularly for Saturday draws, causing an apparent parallel shift upwards in both Wednesday and Saturday revenues. The 2015 re-design does not display such an obvious increase in sales, relative to the 2013 design, but does produce a small subset of draws in early 2016 where sales reached in excess of £90m for a Saturday draw and over £40m for a Wednesday draw. These spikes correspond to an unusually large rollover which occurred due to the operator removing the cap on the number of consecutive draws a rollover was allowed to happen – instead opting to terminate consecutive rollovers when the jackpot prize reached £50m in value[61]. After the immediate shift in sales following the 2013 design changes, sales – particularly on Saturday – appear to continue to follow the downward trend visible in Figure 3.4. Moreover, the sales depicted towards the end of the sample in

[61] In 2016, the cap on the value which the rollover prize could reach was lowered to £22m. If the rollover reaches this amount, and there are no winners in the following draw, it is shared between winners of the next highest prize tier.

Figure 3.8 seem to return to the ballpark figures seen before October 2013 – suggesting that the changes made to the game offered only a temporary boost in sales.

Table 3.7 presents supporting evidence by comparing average sales revenue per draw for the two-year period before the first game changes, the post-2013 design and the post-2015 design. Following the 2013 change, average weekly revenues rose by around £5m, an increase of just under £2m in revenues from Wednesday draws and around £3m for Saturday draws. The higher standard deviation in revenues compared to the pre-2013 design reflects the fact that rollovers – which induce increases in sales – occur much more frequently because of the fall in the quantity of tickets sold.

Table 3.7: Comparison of sales revenues between game designs (£m)

	Pre-2013	Post-2013	Post-2015
Wednesday	15.772	17.532	16.404
	(1.4209)	(3.4835)	(3.9801)
Saturday	28.776	32.283	28.033
	(1.9675)	(2.7633)	(7.9474)
Weekly	44.548	49.815	44.437
	(3.0683)	(4.8408)	(11.5762)

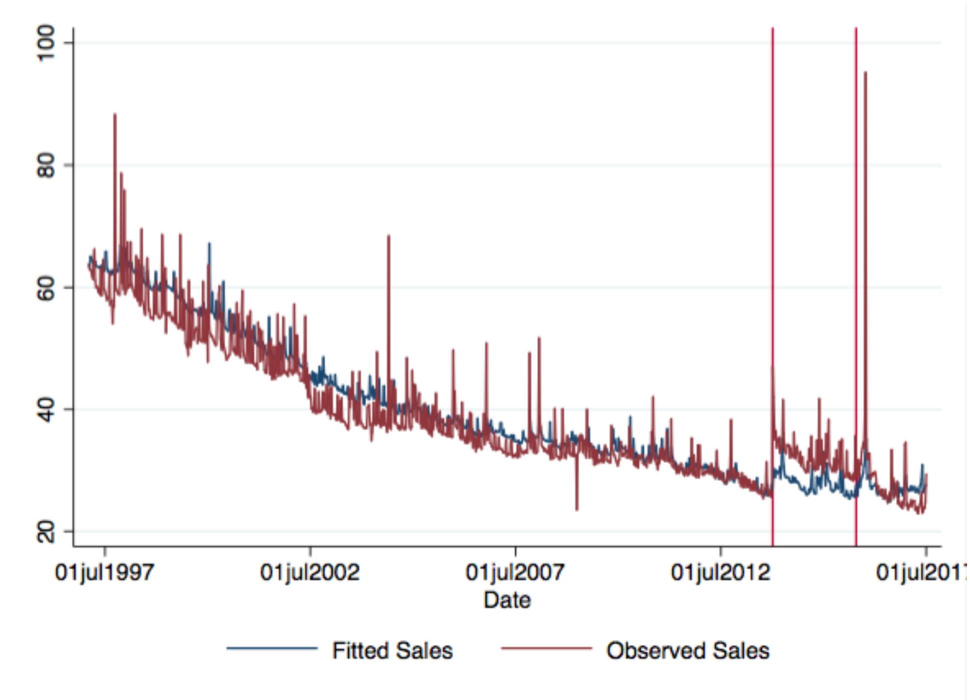
Notes: Pre-2013 sales only considered for the two years (104 weeks) before the original 2013 game re-design. Standard deviations in parentheses.

Whilst informative, the story told by

Figure 3.8 and Table 3.7, that sales increased by roughly £5m per week following the 2013 reform before falling back to pre-2013 levels, are only indicative of the success of the changes. To properly evaluate the effectiveness of the changes, one would need to estimate the level of sales under the original game design over the same time period – for which, clearly, there is no data. The most obvious approach to this is produce out of sample forecasts using, say, estimates of the models presented above in Section 3.6 and Section 3.7 and the mathematical identities which determine price in Section 3.3 as an approximation to this missing counterfactual.

Given the autocorrelation of sales, and the endogeneity of both price and rollovers, this would then require a dynamic forecasting approach, in which fitted values of past sales are used in all the equations of the model, where appropriate to fit subsequent values. However, this too is fraught with difficulty since, without observed data, the variation in price driven by both the size and, particularly, the frequency of rollovers is not known. The endogeneity of rollovers to sales makes deriving these from the equations of the model intractable, thus such an approach would either require using observed frequencies from the new designs as an approximation or simulating rollover occurrence for each draw using probabilities derived from equation (3.2).

Figure 3.9: Dynamic forecast fit of Saturday lotto sales 1997-2017



Attempts at this procedure[62], such as what is depicted in Figure 3.9, generally continue along the underlying time trend variables in the out of sample period. Whilst it is likely that without design changes to the game in 2013 and 2015 sales would have continued to fall, the outputs like that of Figure 3.9 are highly sensitive to the specification of the time trend – with experiments using varying degrees of polynomial expansion in the time trends yielding substantially different forecasts – and estimates based on these outputs should be treated with caution without further supporting evidence on the marginal effect of each of the changes implemented on lotto demand and price. Moreover, the relatively flat sales in the two years before the 2013 design changes suggest that rollover frequency and size had, at least locally,

[62] This was done using the dynamic forecasting suite of commands in Stata 14, which forecasts out of sample using the structural model of supply and demand for lotto tickets outlined in Sections 3.3 and 3.4 and parameterised in Section 3.6. Identities are used to enforce the mechanical nature in which sales determine price, both directly and via rollover size and probability in the supply equation.

been unchanged over that period. In the absence of a more complete model of lotto demand, the declining trend evident in Figure 3.5 and, though unlikely to be useful, in Figure 3.9 suggest a comparison of the average sales figures in Table 3.7 is likely to be sufficient in determining the (lack of) success of the design changes.

3.9 Conclusion

This chapter has modelled the sales of lotto in the UK under the assumption that sales are driven by the effect of rollovers assuming that the transmission mechanism is only through the mean of the prize distribution. By exploiting two of the institutional aspects of the game, a novel strategy to identify the causal effect of effective price of a ticket has been employed. The resultant estimates suggest that the Wednesday draw should be made more attractive relative to the Saturday draw – a novel finding for this literature.

Models of lotto demand in which the mean of the prize distribution is the key dependent variable are often criticised on the grounds that they are based on expected utility theory which is notoriously ineffective at explaining why gambling occurs amongst otherwise risk-averse individuals. This chapter further extends the existing literature by adopting a more pragmatic reduced form model that is cubic in the rollover size, rather than assuming that the effect on sales is only via the expected price. The same instrumental variables strategy used to identify the price model is employed to overcome the endogeneity of rollover size, and estimates suggest £1 million increases in the jackpot prize from rollovers increases Saturday sales by around £1.8 million and Wednesday sales by £0.5 million. A semi-parametric model that allows the effect of rollover size on sales to be fully flexible is also estimated. Whilst testing formally rejects the parametric reduced form of modelling, the semi-parametric model is effectively economically equivalent to the parametric model in its effects.

Finally, major re-designs to the UK's main lotto game in 2013 and 2015 which were likely implemented in response to declining sales figures are assessed. Simple comparison of pre- and post-redesign revenues shows that the 2013 changes were a relative success, increasing sales from an average of £44.5 million per week to £49.8 million – equivalent to over £275 million per year in extra revenues. The 2015 renovation of the game proved less successful, seeing sales return to their pre-2013 levels. Thorough analysis of the effectiveness would require knowledge of the level of sales had the game re-designs never occurred. Attempts to forecast sales out of sample from the model developed in Section 3.6 mirror the declining trend in sales pre-2013, indicating that changes to the game's design were more successful than simply comparing sales figures implies. However, these forecasts could be improved with a more complete simulation of lotto design and the effects of changing the game parameters in future research.

3.10 Appendix

3.10.1 Estimates with rollovers as instruments

Table A3.8: Comparison of second stage estimates of lotto demand using rollover size and conscious selection as instruments

Instrument:	**Rollover Size**		**Conscious Selection**	
Dependent Variable	S_t^{Sat}	S_t^{Wed}	S_t^{Sat}	S_t^{Wed}
	(1)	**(2)**	**(3)**	**(4)**
S_{t-1}	0.137***	0.119***	0.133**	0.0643*
	(0.0478)	(0.0258)	(0.0546)	(0.0332)
S_{t-2}	0.109***	0.028**	0.102***	0.0796**
	(0.0264)	(0.0276)	(0.0284)	(0.0321)
$\widehat{P}_t$	-48.717***	-34.904***	-47.283***	-55.856***
	(3.9276)	(1.8253)	(3.5934)	(4.6827)
Constant	81.245***	30.285***	79.771***	45.237***
	(3.4573)	(1.8068)	(3.5393)	(2.5605)
LR Elasticity	-0.6510***	-0.8545***	-0.6416***	-1.4718***
	(0.0524)	(0.0464)	(0.0475)	(0.1172)
Observations	868	868	867	868
R^2	0.9526	0.9403	0.946	0.867

Notes: Robust standard errors in parentheses. ***/**/* denotes statistical significance at 1%,5%, 10% confidence. Trend and seasonality controls omitted from reporting. Sales variables in millions.

3.10.2 Consumer Surplus, Tax, and Deadweight Loss

Assuming a linear demand function for lotto tickets, it is possible to provide estimates of long-run tax revenues, consumer surplus and, due to the existence of tax and good causes receipts, deadweight loss using estimates from Table 3.3.

Figure A3.10: Linear demand for Saturday lotto tickets

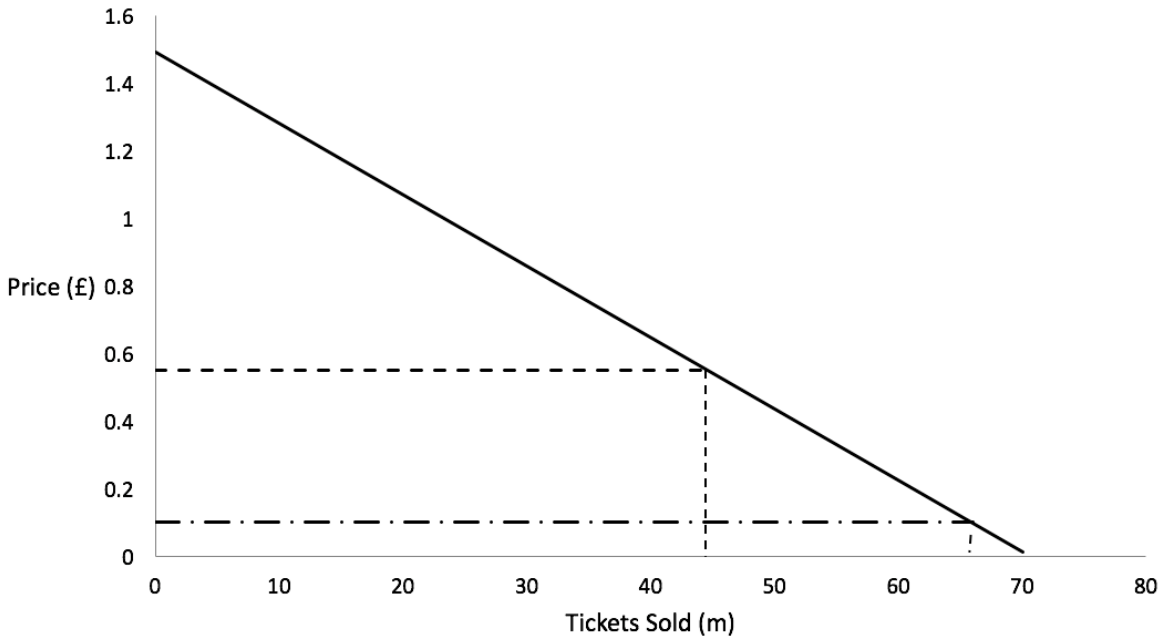


Figure A3.10 shows such a demand curve, for Saturday tickets, parameterised using column 5 of Table 3.3, as well as indicating average price (dashed line, $P = £0.51$) and marginal cost (dot-dashed line, £0.10) which is the amount per £1 spent on National Lottery tickets shared between the operator and the vendor.

This suggests consumer surplus is in the region of £20m per draw (£1b per annum) for Saturday games and £3m per draw (£150m per annum) for Wednesday draws. Tax and good causes revenues are approximately £16m per draw for Saturday and £8.8m per draw for Wednesday (approximately £1.3b per annum combined) and deadweight losses are estimated to be £7.8m for Saturday and £3.4m for Wednesday (approximately £582m per annum combined).

3.10.3 Semi-parametric Estimation Routine

The semi-parametric approach used here is the partially linear model and is estimated using the double residual method developed by Robinson (1988). Partially linear models contain an assumed linear parametric component of variables, $\boldsymbol{X}_t$, with unknown parameters, β, and an unknown function, g, of a variable, say R_t, for which we cannot make assumptions of the functional relationship between itself and the dependent variable, say S_t. Specifically, the model can be written as,

$$S_t = \beta' \boldsymbol{X}_t + g(R_t) + \varepsilon_t \tag{3.18}$$

Consistent estimation of (3.18) requires that $E(\varepsilon_t | \boldsymbol{X}_t, R_t) = 0$. Robinson's estimator for the unknown parameter vector, β, and unknown function, g, can then be obtained by transforming the model,

$$S_t - E(S_t|R_t) = \beta'\big(\boldsymbol{X}_t - E(\boldsymbol{X}_t|R_t)\big) + \varepsilon_t \tag{3.19}$$

and replacing $E(S_t|R_t)$ and $E(\boldsymbol{X}_t|R_t)$ by their respective non-parametric estimators, $\widehat{m}_h^S(R_t)$ and $\widehat{m}_h^{\boldsymbol{X}}(R_t)$, which are found using kernel density estimates with bandwidths h. Robinson showed that using OLS to estimate this transformed model yields $\hat{\beta}$ coefficients which converge at a rate of $\sqrt{n}$. After obtaining estimates for β, it is then possible to estimate the unknown function g using the following,

$$\hat{g}_h(R_t) = \widehat{m}_h^S(R_t) - \beta' \widehat{m}_h^{\boldsymbol{X}}(R_t). \tag{3.20}$$

If R_t is endogenous, such that $E(\varepsilon_t|R_t) \neq 0$ or $E(S_t|R_t) \neq 0$, this approach will yield inconsistent estimators since $E(\varepsilon_t|R_t)$ would be non-zero. Blundell *et al* (1998) develops a method for estimating such partially linear models when the non-parametric variables are endogenous. Their approach relies on the existence of some instrumental variable, Z_t, such that

$$R_t = \gamma Z_t + \nu_t \tag{3.21}$$

with $E(\nu_t|Z_t) = 0$ and $E(\varepsilon_t|R_t, \nu_t) = \rho\nu_t$. Then $\varepsilon_t = \rho\nu_t + \eta_t$ and an analogue of Robinson's partially linear model holds as follows,

$$S_t = \beta'\boldsymbol{X}_t + g(R_t) + \rho\nu_t + \eta_t \tag{3.22}$$

which we can then re-write and estimate as per Robinson's methodology using:

$$S_t - E(S_t|R_t) = \beta'\big(\boldsymbol{X}_t - E(\boldsymbol{X}_t|R_t)\big) + \rho\big(\nu_t - E(\nu_t|R_t)\big) + \eta_t. \tag{3.33}$$

3.10.4 Auxiliary Rollover Model Estimates

Estimates in Table A3.9 are auxiliary to those found in Table 3.4 in which demand for lotto tickets dependent on a function that is cubic in rollover size.

Table A3.9: Heckman-instrumented estimates of liner and quadratic rollover induced ticket demand

	Heckman First-Stage			
Dependent Variable	S_t^{Sat} **(1)**	S_t^{Wed} **(2)**	S_t^{Sat} **(3)**	S_t^{Wed} **(4)**
S_{t-1}	-1.539***	-0.357***	-1.253***	-0.333***
	(0.3587)	(0.0989)	(0.3722)	(0.1031)
S_{t-2}	0.255***	0.117***	0.247***	0.122***
	(0.0492)	(0.0434)	(0.0500)	(0.0459)
$\widehat{R_t}$	1.684***	0.654***	1.641***	0.704***
	(0.3396)	(0.1434)	(0.3330)	(0.1438)
$\widehat{R_t^2}$	-	-	-0.011***	-0.006
	-	-	(0.0039)	(0.0080)
Constant	78.386***	38.885***	72.913***	37.247***
	(5.3897)	(6.2732)	(5.4966)	(6.5729)
Durbin-Watson d-stat	1.9685	1.9905	1.9625	1.9868
ARCH LM test (H0: no ARCH effects)	0.051	0.034	0.062	0.027
ARCH LM test (p-value)	0.8219	0.8529	0.8038	0.8706
AIC	4464.907	4374.156	4455.880	4375.155
Observations	868	868	868	868
R^2	0.9146	0.7099	0.9156	0.7103

Notes: Robust standard errors in parentheses. ***/**/* denotes statistical significance at 1%,5%, 10% confidence. Trend and seasonality controls omitted from reporting. Sales and rollover size variables in millions.

3.10.5 Estimates Using Consecutive Numbers in the Identification Strategy

Appendix Table 3.10: Heckman selection model estimates for rollover size and probability using consecutive numbers in the selection equation

	Saturday	Wednesday
Dependent Variable	$\boldsymbol{R_t^{Sat} \mid R_t^{Sat} > 0}$	$\boldsymbol{R_t^{Wed} \mid R_t^{Wed} > 0}$
$W_{3,t-1} - \pi_3 S_{t-1}$	-4.825***	-7.021***
	(1.2555)	(1.0962)
S_{t-1}	1.031***	0.722***
	(0.0326)	(0.0814)
S_{t-2}	-0.091***	-0.105*
	(0.0209)	(0.0574)
Constant	-12.574***	-41.265***
	(1.9464)	(5.9041)
Selection Equation	$\mathbf{Pr}\ (\boldsymbol{R_t^{Sat} > 0})$	$\mathbf{Pr}\ (\boldsymbol{R_t^{Wed} > 0})$
$\Delta_{medium,t-1}$	0.183***	0.021
	(0.0552)	(0.0427)
$\Delta_{large,t-1}$	0.336***	0.119**
	(0.0543)	(0.0552)
$\delta_{consecutive,t-1}$	0.310***	0.068
	(0.0940)	(0.0636)
S_{t-1}	-0.081 ***	-0.029
	(0.0218)	(0.0216)
S_{t-2}	-0.000	-0.011
	(0.0151)	(0.0208)
Constant	0.143	-0.782
	(1.4543)	(1.8938)
Observations	868	868
Censored observations	616	750
λ	-0.0845	2.4165***
$\chi^2_{(1)}$ test of $\lambda = 0$	0.15	15.94
P-value	0.6944	0.0001

Notes: Robust standard errors in parentheses. ***/**/* denotes statistical significance at 1%,5%, 10% confidence. Trend and seasonality controls omitted from reporting. Sales variables in millions.

Appendix Table 3.11: Second-stage price model estimates with consecutive numbers in the identification strategy

	Control Function using Heckman First-Stage	
Dependent variable	S_t^{Sat} **(1)**	S_t^{Wed} **(2)**
S_{t-1}	0.133**	0.071**
	(0.0546)	(0.0347)
S_{t-2}	0.101***	0.073**
	(0.0283)	(0.0328)
$\widehat{P_t}$	-46.906***	-51.058***
	(3.5658)	(4.6143)
Constant	79.610***	42.254***
	(3.5295)	(2.5916)
LR Elasticity	-0.636***	-1.337***
	(0.0470)	(0.1137)
Durbin-Watson d-stat	2.0331	1.9418
ARCH LM test (H_0: no ARCH effects)	0.088	0.060
ARCH LM test (p-value)	0.7673	0.8058
AIC	4061.838	3777.626
Observations	868	868
R^2	0.9460	0.8541

Notes: Robust standard errors in parentheses. ***/**/* denotes statistical significance at 1%,5%, 10% confidence. Trend and seasonality controls omitted from reporting. Sales variables in millions.

Appendix Table 3.12: Second-stage rollover model estimates with consecutive numbers in the identification strategy

	Heckman First-Stage	
Dependent variable	$\boldsymbol{S_t^{Sat}}$	$\boldsymbol{S_t^{Wed}}$
	(3)	**(4)**
S_{t-1}	-1.254***	-0.340***
	(0.3721)	(0.1070)
S_{t-2}	0.249***	0.116**
	(0.0502)	(0.0459)
$\widehat{R_t}$	1.884***	0.483**
	(0.4070)	(0.2104)
$\widehat{R_t^2}$	-0.033	0.030
	(0.0199)	(0.0385)
$\widehat{R_t^3}$	0.000	-0.001
	(0.0004)	(0.0012)
Constant	72.269 ***	38.497***
	(5.6937)	(7.0793)
Joint F-test $R_t^2, R_t^3 = 0$	5.97***	4.20**
Prob>F	0.0027	0.0155
Durbin-Watson d-stat	1.959	1.988
ARCH LM test (H0: no ARCH effects)	0.040	0.024
ARCH LM test (p-value)	0.8407	0.8776
AIC	4455.41	4373.82
Observations	868	868
R^2	0.9159	0.711

Notes: Robust standard errors in parentheses. ***/**/* denotes statistical significance at 1%,5%, 10% confidence. Trend and seasonality controls omitted from reporting. Sales and rollover size variables in millions.

4 The regressivity of lotto taxation

4.1 Introduction

This chapter is concerned with estimating the extent to which high rates of tax on UK lotto games are disproportionately borne by low-income individuals – i.e. the regressivity of such taxes. The main analysis focuses on parameterising Engel curves from a Working-Leser specification in which a proxy for log income, log total expenditure, is the key independent variable in determining the budget share of lottery tickets. Since lotteries are taxed at a constant rate, this enables us to infer on whether lottery taxes are regressive, progressive, or proportional directly from estimates of income elasticity when its value is less than, more than, or equal to one, respectively. The household-level nature of the data used necessitates non-standard statistical techniques to deal with the potential for bias associated with households that do not, at least in the survey period, participate in lotto[63] – namely Heckman's (1979) selection model and Cragg's (1971) double hurdle routine. A novel strategy is used to identify these two-step models by exploiting exogenous differences in consumer preference arising from religious practices; namely, abstaining from alcohol, pork, and gambling products by practicing Muslims.

The results classify the UK lotto as an inferior good and headline estimates of the game's income elasticity are between -1.4 and -0.7, indicating that taxes imposed on lotto tickets are significantly more regressive than the previous literature suggests. Advances on previous literature are also made by employing non- and semi-parametric techniques to examine directly the relationship between lottery expenditure and income over the entire distribution. To complement the income elasticity estimates from the Working-Leser parametric specification, a second non-parametric approach is used in which an index of regressivity, known as the Suits' index, is calculated in an attempt to quantify how regressive lotto taxes are.

Lotteries are typically state-owned or state-licensed enterprises, the proceeds of which are earmarked for financing public projects or services; in particular, on those that are not typically financed by government. Lotteries, in countries where they are available, are frequently the most popular gambling product both in terms of revenue and prevalence. In the

[63] Only games which feature draws that occur at regular intervals are considered here and not scratchcards.

UK, government-licensed lottery games achieve annual sales around £4.5b (2.6% of household expenditure) and the British Gambling Prevalence Survey 2010 reveals that 80% of those who had gambled in the previous 12 months had participated in the UK National Lottery in the same period. Moreover, state-run lotteries are common across the globe. In the US, for instance, lotteries are available in 43 states with each contributing to state-level finances some of which are earmarked for particular forms of expenditure – although none impose the *additionality* constraint that the UK game operates under. US lotteries achieved sales of $70.1b in 2014 and have raised over $300b towards state revenues since the first modern lottery in New Jersey, 1971.

Goods which are subject to high rates of taxation naturally raise concerns about regressivity but the popularity, state provision, and earmarking of funds for public goods make lotteries particularly interesting candidates for examination. In the UK only around 50% of ticket sales are returned in prizes – making lotteries one of, if not the, worst-return gambling products available. The remainder of ticket revenues are divided between 'good causes'[64] (28%), tax (12%), retailer commission (5%), and operating costs and profit (5%). Accounting for only the consumers' and supplier's interests, such an allocation of revenues would then appear as an effective tax rate of 40% – substantially higher than tax rates paid on other gambling products[65]. The lotto tax rate is analogous to 'sin taxes' such as tobacco, alcohol, and sugar duties which are popular fiscal policies designed to discourage the consumption of goods which are potentially damaging to the consumer. Implementing such punitive taxes rarely faces opposition when the state can appeal to a paternalistic 'moral high ground' and claim to be either internalising the social cost of consumption or acting to deter consumption of otherwise harmful goods. However, whilst problem gambling is costly both to the individual and to the state (see Thorley, Stirling, and Huynh, 2016; and Pryce, Walker and Wheeler, 2017), there is no evidence that playing lotto games are connected to problem gambling or is linked to any other harms[66].

---

[64] 'Good causes' funds are distributed by government-appointed organizations and in the year ending March 2016 were invested as follows: Health, education, environment and charitable causes – 40%, sport – 20%, arts – 20%, heritage – 20%.

[65] At the time of writing, tax rates on other gambling products range from 3% (spread betting) to 25% (high-stakes fixed odds betting terminals). A full list of UK gambling duty rates can be found on the UK Government webpages via www.gov.uk.

[66] Although Pryce et al (2017) finds that there is a statistically significant effect of scratchcards on the loss in well-being associated with being a problem gambler.

Though beyond the scope of this research, it is important to consider the *overall* regressivity of lotteries by additionally addressing the question who benefits, and by how much, from the proceeds of taxes collected from players. Clearly, this would require an analysis of the income distribution of both the players, which is the subject of this chapter, and that of the beneficiaries of good causes funding and the sales tax proceeds. If the proceeds from a lottery whose taxation is regressive are spent entirely on projects which exclusively benefit the lowest income individuals then, arguably, one would be less concerned about the regressive tax than about the state's role in how low-income individuals allocate their limited budgets. At the opposite extreme, a lottery which finances projects primarily for the benefit of individuals in the right tail of the income distribution, would place even more emphasis on the question of the extent to which lotto taxation is regressive. Though this concern is not dealt with in the analysis here, Borg and Mason (1988), Rubenstein and Scafidi (2002), and more recently Feehan and Forrest (2007) all present evidence that suggests that the latter is likely to be the case.

4.2 Related literature

Since the introduction of the New Jersey lottery in 1971 many aspects of lotteries have received attention in economics literature. An early contribution was Clotfelter and Cook (1991), and Perez and Humphreys (2013) provide a thorough overview of the literature. There are several studies on the incidence of lottery taxes in the literature and a select few are reviewed here.

Brinner and Clotfelter (1975) provide early evidence on the incidence of lottery taxation using individual-level survey data made up from separate random-sample surveys in Connecticut and Massachusetts, and from a sample of lottery winners from Pennsylvania. Their evidence suggests that average nominal expenditure on lotteries peaks in the middle of the income distribution before quickly falling away at higher income levels. Moreover, since taxes are a constant proportion of sales revenue they are able to conclude the tax is regressive. Their paper also shows that mean lottery expenditure, as a percentage of total income, falls continuously from low to high income brackets. Whilst informative, these findings are merely descriptive so Brinner and Clotfelter (1975) also provide basic regression evidence from county-level income and demographic data, together with the geographic distribution of small prize winners as a noisy measure of ticket sales. Their regression analysis corroborates the conclusions from the extensive descriptive statistics and shows that expenditure as a percentage of income declines as incomes rise. Other early evidence on the regressivity of US lottery taxation follows the methods of Brinner and Clotfelter (1975) by using aggregate data at both the zip code (e.g. Clotfelter, 1979, and Clotfelter and Cook, 1987) and county (e.g. Mikesell, 1989) levels. Only Mikesell (1989) is inconclusive on whether lottery taxes are regressive or not by estimating the income elasticity of lotteries in Illinois between 1985-1987 to be statistically indistinguishable from 1. These findings are the exception, however, with all other studies of the time finding evidence of regressivity.

Studies of the income distribution of lottery players using aggregate level data are often criticised for assumptions made about regional homogeneity and the inherent measurement error arising from individuals purchasing tickets outside of their home district. When modern state lotteries were in their infancy, the popularity of approaching the question in this way was due to limited individual-level data. Moreover, early research using individual or household level data were susceptible to either small samples, sample selection issues, or both. One such example is that of Spiro (1974), who analyses a mailed survey of 1,250 winners of the Pennsylvania state lottery. With only 271 usable responses, the external validity of the study is severely limited by both a small sample size and selection bias resulting from only surveying winners and a low response rate. As a result, Sprio (1974) argues that an income elasticity estimate of 0.22 is an upper bound, since any correlation between the probability of response and income is likely to be negatively correlated. However, this ignores the sample selection issues inherent to only surveying winners. Even if the income distribution and number of tickets purchased by each of the 271 respondents is representative of all players, one would then be also required to rely on aggregate level assumptions about the income distribution to formulate conclusions about the regressivity of lottery taxation for all players. Finally, only surveying winners who have participated *for sure* in the lottery prohibits any insight into the behaviour and characteristics of non-lottery players. Of specific interest here would be whether changes in income, or indeed any other factor, affects propensity of individuals to purchase a ticket, or if changes in these variables merely encourages existing players to alter the number of tickets they purchase.

As the global lottery market has continued to grow over the past four decades, there is an increasing number of large population surveys being made available to researchers which also circumvent many of the issues which arise with early individual-level and aggregate-level studies. Kitchen and Powells (1991) use the Canadian 1986 Survey of Family Expenditure – a household-level study – to obtain income elasticities for lotteries played in all six major Canadian regions. Their data benefits from over 10,000 observations – considerably more than early studies such as Spiro (1974) – and by randomly sampling from the entire population. Their estimation approach involves a regression of the amount spent on lottery tickets on household income and wealth, as well as controls for demographic characteristics. A common issue with such data is the presence of a zero lower bound in expenditure data which would bias coefficient estimates from OLS regression. Kitchen and Powells (1991) use Tobin's (1958) estimation routine – the Tobit model – which corrects for this bias. The Tobit model is often preferred to OLS in this instance since it effectively accounts for the effect of regressors on both the decision to purchase and on the amount to spend. Kitchen and Powells (1991) find income elasticities in all regions to be less than 1, ranging from 0.70 (Quebec) to 0.92 (Alberta). They also conclude that taxes on Canadian lotteries are regressive, but less regressive than previous estimates both in the US and Canada would suggest.

Farrell and Walker (1999) compare coefficient and income elasticity estimates from OLS and Tobit models with a third estimation procedure developed by Heckman (1979). They echo the arguments of Scott and Garen (1994) that a Tobit model may be invalid in estimating lottery expenditures for two reasons. First, Tobit estimates are sensitive to an assumption of normality in unobserved heterogeneity, which is particularly concerning when using micro-level data where heteroscedasticity is often expected. Second, the Tobit model imposes the strong restriction that the influence of regressors on the decision to participate are proportional to the influence they have on the amount purchased, conditional on participating. The Heckman selection model generalises the Tobit routine to allow for the dependent variables to have differing effects on the decision to purchase and the amount to spend but, as Farrell and Walker (1999) point out, it requires an exclusion restriction for identification. The routine requires at least one variable to influence participation, but not the quantity purchased conditional on participation, in order for the model to be identified. Farrell and Walker (1999) use car ownership – a factor which plausibly reduces the transaction cost of participation but does not influence the number of tickets to buy – to overcome this problem, though admit this is not ideal since "there may well be grounds for thinking that car ownership is correlated with unobservables that affect lotto demand" (p. 112). Nonetheless, they find that OLS and Tobit understate the regressivity of lotto taxation relative to Heckman's selection model with income elasticities of 0.27, 0.45, and 0.13, respectively.

More recently, the use of the Heckman selection model in estimating lottery ticket purchases has been criticised. In a study using data from Alberta in Canada, Humphreys, Lee and Soebbing (2010) argue that the Heckman model would only be appropriate if individuals who gamble will always be observed to gamble and that nonparticipants will never gamble. Such an assumption ignores the draw-by-draw variation in lottery ticket sales driven by rollovers which is well documented in the literature. Instead, Humphreys, Lee and Soebbing (2010) propose using a double hurdle routine which further relaxes the assumptions of the Tobit model and compare differences in coefficient estimates of explanatory variables on lottery spending between the two models. They note the double hurdle routine is particularly appealing since it allows for abstention as well as non-participation driven by low prizes in any draw. Their results reveal a further deficiency of the Tobit estimation routine. The Tobit model for their data simply shows that spending on lotto increases with income, whereas double hurdle estimates reveal that income has no bearing on whether an individual participates, only their expenditure conditional on participation. This result might imply that the lotto taxation in Alberta may not be regressive, although Humphreys, Lee and Soebbing (2010) stop short of formally analysing the this issue. Moreover, they do not present Heckman model estimates for comparison, so there remains a question about which model would be preferential in practice for micro-level data.

A final strand of the literature exists in relation to evaluating the regressivity of taxation. Suits (1977a) proposes a method of investigating the issue using an index of tax progressivity which is analogous to the Gini coefficient of income inequality. The index, *S*, ranges from +1, extreme *pro*gressivity, to -1, extreme *re*gressivity, where 0 would indicate a perfectly proportional tax. The index is constructed by plotting the accumulated percent of the total tax burden against the accumulated percent of total income (with individuals sorted by income). The value of the index for a given tax is then computed by subtracting the ratio of the area under the curve and the area under the diagonal from 1. Suits' index has become a popular tool in evaluating tax policy since it provides a normalised measure of regressivity which can be compared to other taxes. Suits (1977b) estimates this index for a variety of gambling products in the US and shows that only casino games (outside of Nevada) and sports betting are taxed progressively with index values of 0.26 and 0.29, respectively. Taxes on state lotteries were found to be the third most regressive of all gambling products with an index of -0.31. Many of the studies reviewed in this section also provide estimates of this index for lotteries, unanimously finding lottery taxation to be regressive, ranging from -0.13 in Quebec (Kitchen and Powells, 1991) to -0.46 in Massachusetts (Brinner and Clotfelter, 1975). Table 3 of Perez and Humphreys (2013) provides a more complete list of the reported Suits' index values from the literature.

4.3 The data

The data used here are taken from the UK Expenditure and Food Survey (EFS) – later re-named the Living Costs and Food Survey (LCF) – which is a large, individual-level survey of individual and household expenditure. Collection of expenditure data is done via a two-week expenditure diary that is self-completed by all survey participants within each household. An interview is also conducted to obtain some demographic and socio-economic information and details of both infrequent expenditure on goods and services – for instance, cars or vacations – and regular expenditures such as water, energy, and rent/mortgage which may fall outside the diary period. The survey is conducted year-round and the dataset used here contains responses from 2001-2013. Over this time period, the design of the UK National Lottery was unchanged, though it was revised in October 2013 when our dataset ends, and the design of the survey was constant. Over the whole sample period, the sticker price of tickets also remained constant at £1 and average returns were around 50%. This means that this product can be largely considered homogenous for the whole sample period, conveniently providing little concern for structural breaks in the demand analysis here.[67]

Since labour and expenditure decisions are often made at the household level, expenditure and income is aggregated to the household level for the analysis in this chapter.[68] In total, there are 79,065 households sampled evenly across each year of the available dataset. The survey is conducted year-round and across the whole of the UK with samples for each year drawn from postcode data and designed to be representative of the UK population.

The EFS/LCF data contains information on expenditure on individual items, total expenditure, and gross incomes and all values which relate to currency amounts are converted to weekly averages. Total expenditure is a derived variable in this dataset and differs from

[67] The design of lotto games means that the expected value of tickets – and therefore the expected cost – varies with the number of players because of the possibility of prizes not being won (for further explanation see Walker and Young (2001), for example). With declining participation over the sample period, the expected cost of playing increases. In this analysis, month fixed-effects capture this trend.

[68] It seems likely that many households will collectively own tickets bought by separate household members so analysing behaviour at the individual level is probably unwise, as well as impossible because of restrictions on access to the individual diary data.

aggregating expenditure on individual items because it excludes council tax, water, and sewerage charges in order to better maintain the anonymity of the respondents since these taxes vary across Local Authorities. This is therefore the preferred measure of total expenditure rather than manually aggregating expenditures for each household. Income is recorded as gross, rather than net, and two options are available in the data – 'current' and 'normal' gross incomes. Differences between the two measures are that the former includes any recent social security benefit payments, actual current pay (based on the last pay slip), and sick pay, whilst the latter only consists of current pre-tax pay provided the respondent has not been out of work due to illness for more than 13 weeks. Where appropriate, this chapter exclusively uses the 'normal income' measure in order to better capture demand which arises as a result of regular income.

The preferred analysis here involves regressing the budget share of lottery expenditure on log income and other relevant controls. To calculate the budget share of lotteries, the recorded expenditure on lotto tickets over the two-week period is simply divided by the total expenditure derived variable for each household. In addition to income, estimates presented in Section 4.6 also include controls for government region and the age, sex, employment status and years of education of the household representative person (HRP). As noted in Kitchen and Powells (1991), there is no obvious prior for the direction or significance of many of these variables, but their inclusion and the resultant estimates may be of use to the relevant stakeholders who have an interest in the household characteristics of lotto players beyond just their income. Given the likely correlation between some of these controls and income, these estimates will, however, provide an estimate of income elasticity which is conditional on the control variables. For completeness, all the analysis in Section 4.6 was repeated without the extra control variables to obtain an unconditional estimate of income elasticity, which may arguably be of more interest to the regulator, and these estimates can be found in the Appendix Section 4.8.3. For all models, this income elasticity was even smaller than those that form the body of this chapter, suggesting lotto taxation is even more regressive than the main analysis implies.

Means and standard deviations of the relevant variables to our analysis are presented in Table 4.1. There appears be little difference between purchasing and non-purchasing households in income and the age of the HRP. However, differences are noticeable in means of sex, employment status, and education level of the HRP. Purchasing households are more likely – relative to their non-purchasing counterparts – to have a male HRP, to have an employed HRP, and have a HRP who left education at or below the age of 16. Such a profile of purchasing households is consistent with previous studies on the demographics of gamblers conducted on individual-level surveys such as in Wardle *et al* (2010).

Table 4.1: Mean and standard deviation of variables used in the analysis by observed household participation in lotto over the 2-week diary period

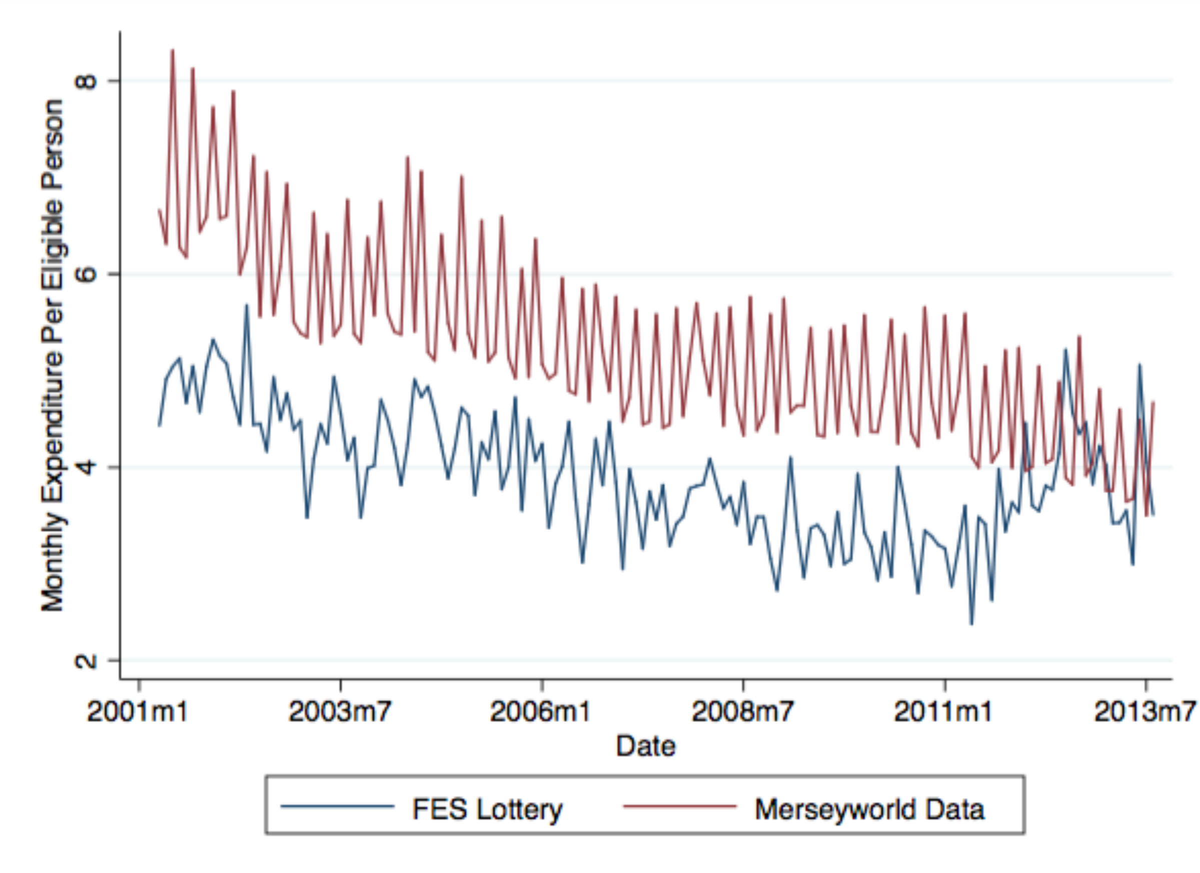
	Non-purchasing Households	**Purchasing Households**	**All**
Weekly Gross Normal Income	609.063	604.605	607.144
	(616.3904)	(437.5022)	(546.6059)
Weekly Total Expenditure	438.733	454.394	445.475
	(386.6743)	(334.3482)	(365.2696)
Weekly Lottery Expenditure	-	4.289	1.846
	-	(4.6843)	(3.7357)
Age HRP	51.239	53.040	52.014
	(17.5049)	(15.1596)	(16.5600)
Male HRP = 1	0.585	0.664	0.619
	(0.4928)	(0.4722)	(0.4856)
Employment Status of HRP:			
Employed = 1	0.568	0.621	0.591
	(0.4954)	(0.4851)	(0.4917)
Unemployed = 1	0.028	0.019	0.024
	(0.1641)	(0.1352)	(0.1524)
Retired = 1	0.285	0.270	0.279
	(0.4516)	(0.4437)	(0.4483)
Sick/Unoccupied = 1	0.119	0.091	0.107
	(0.3239)	(0.2871)	(0.3089)
Age HRP left education:			
At or before 16 = 1	0.552	0.706	0.618
	(0.4973)	(0.4557)	(0.4858)
17-18 = 1	0.177	0.159	0.169
	(0.3815)	(0.3654)	(0.3748)
18+ = 1	0.269	0.134	0.211
	(0.4432)	(0.3412)	(0.4079)
n	45,028	34,037	79,065

One might be concerned, as with studies into alcohol expenditure for example, about the possibility of under-reported lotto expenditures in the EFS/LCF. To investigate this potential problem, Figure 4.1 presents a comparison of per-person monthly expenditure on lotto as recorded in the EFS/LCF with per-person monthly expenditure with the aggregate-level dataset used in the previous chapter[69]. Clearly, from 2001-2011, the EFS/LCF under-reports lotto expenditure relative to the aggregate level data, but the magnitude of under-reporting appears fairly constant over time. From 2011 to the end of the data, there is a

[69] Computation of a per-person monthly expenditure from the Merseyworld dataset assumes a constant 46m adult population.

reduction in the level of under-reporting of lotto expenditures, though reasons for this are unclear with no evident change to the categorisation of lotto expenditures in the data documentation.

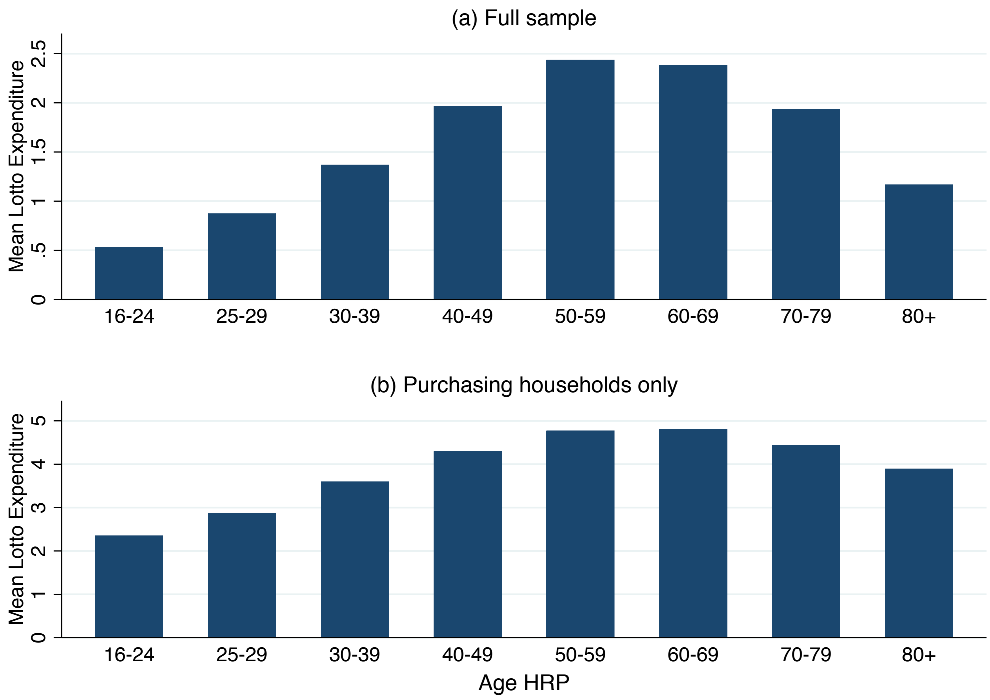
Figure 4.1: Comparison of EFS/LCF lotto spend with Merseyworld data



The under-reporting of lotto expenditures in the EFS/LCF data would be particularly concerning in the analysis later in this chapter *if* the level of under-reporting varies with income. This is, of course, something which is impossible to know since this data is collected on the population of lotto players. Moreover, if the level of under-reporting does *not* vary with income, then this is simply a measurement error issue, and one that is in the dependent variable which presents no problem for the statistical analysis in this chapter.

A comparison of means in Table 4.1 only provides an insight into the characteristics of players versus non-players, and not the correlations between those characteristics and the level of spending on tickets. Figure 4.2 illustrates that whilst there is little difference in mean age of the HRP between purchasing and non-purchasing households, there is relatively substantial variation in expenditure over the age distribution. Panel (a) gives average weekly lottery expenditure by age of HRP for the whole sample and there is a clear correlation between age of HRP and spending, rising through the age ranges before peaking at 60-69 year olds. Panel (b) also shows a similar, albeit less pronounced, pattern with the sample restricted to ticket purchasers only. Differences in expenditure across other demographic variables, which are also more moderate when restricting the sample to ticket purchasers, are also evident and similar graphs to Figure 4.2 for the additional control variables can be found in the appendix.

Figure 4.2: Mean weekly spend on lotto by age of HRP

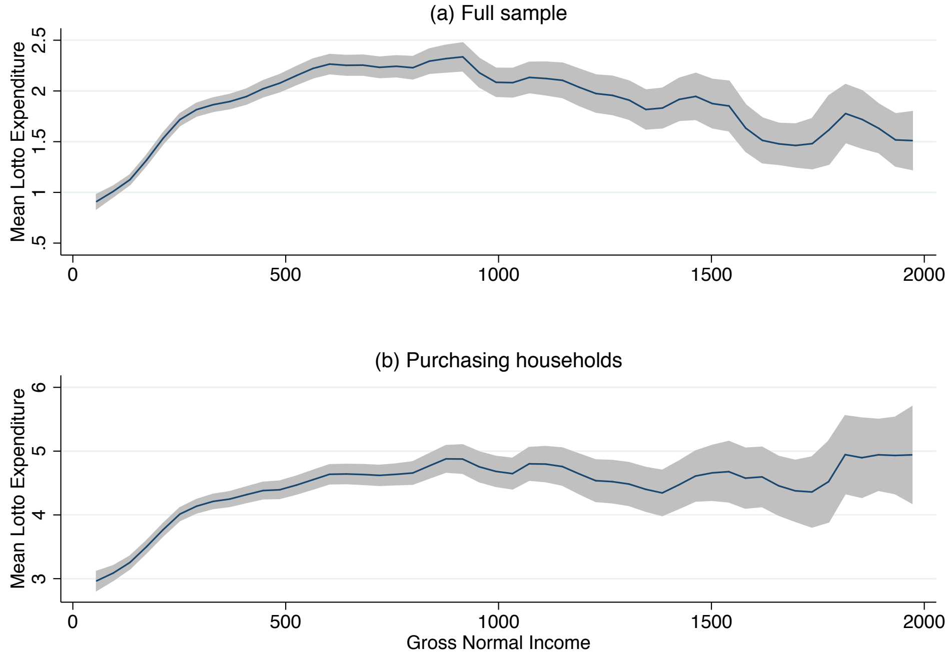


Crucial to the analysis in this paper is the variation in lotto participation and expenditure across the income distribution. Figure 4.3 provides a local polynomial fit of lottery expenditure against income, with the data trimmed to exclude the upper and lower 1% of observations by weekly income. Naturally, the polynomial fit is higher for purchasing households versus the whole sample, however a similar shape can be observed from both panels of the figure. Mean lotto expenditure increases with household income at the left of the distribution before peaking at around an income of £900 per week (a gross household income of £46,800 p.a.) and then declining. Across the whole sample, the downward trend broadly applies over the rest of the income distribution, but amongst players there is a slight increase starting at very high levels of income around £1,500 per week. The percentage increase in lottery expenditure appears to fall behind percentage increases in gross normal income, except perhaps for incomes below around £250 per week in both panels – tentatively indicating a small, or even negative, income elasticity.

Figure 4.4 illustrates how the expenditure share on lotto declines with total expenditure. The natural logarithm of total expenditure is used on the x-axis for clarity. Since the sticker price of each ticket is fixed and equal to £1 in our sample period, the apparent “lines” declining from left to right coincide with those households who purchase a fixed number of tickets per week. This scatter plot gives a clearer indication that one should expect low, or perhaps negative, income elasticity estimates with the average expenditure share on lottery tickets declining, on average, everywhere across the total expenditure range.

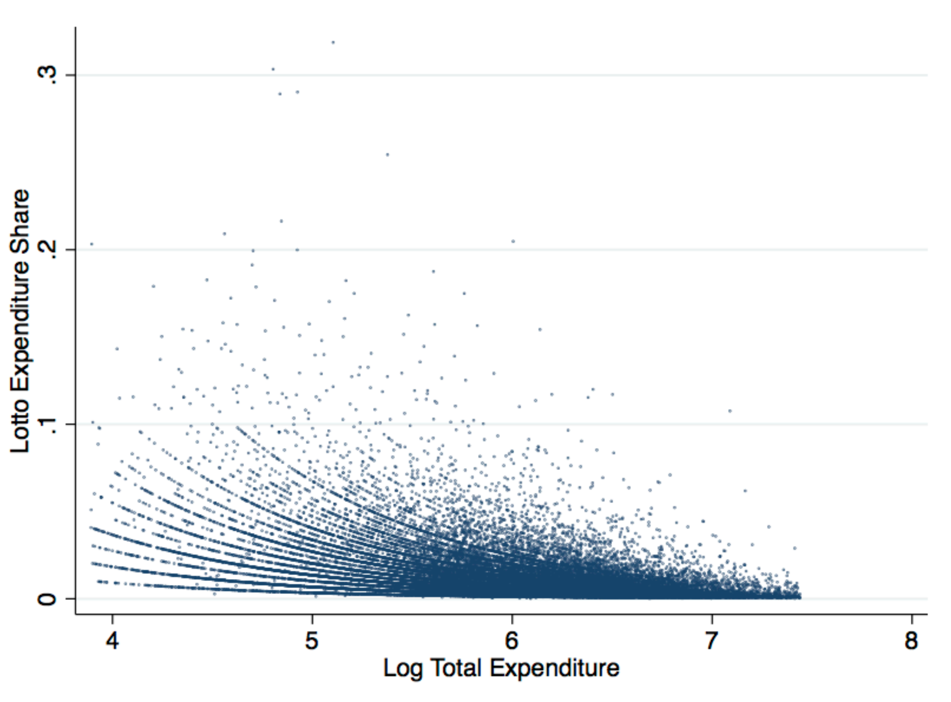
As with any research using data from population sampling, a final concern with the data is that of the response rate and possible non-response bias. Over the entire sample the response rate is 55% at the household level, though this declines over the sample period as is common with repeated socioeconomic surveys. Table 4.2 tabulates the number of households in the data and the household-level response rate for each year of data available. Figure 4.5 illustrates the sample size in each quarter of data available and that responses are distributed fairly uniformly throughout each year. The decreasing response rate evidenced in Table 4.2 is also apparent from the declining frequency of usable observations. Concerns about non-response bias are, however, allayed by there being no statistically significant difference in estimates when using sample weights included with the EFS/LCF datasets.

Figure 4.3: Local polynomial fit (95% CI) of weekly lotto expenditure versus weekly gross normal income for the whole sample and purchasing households



Note: data trimmed to exclude the upper and lower 1% of households by gross normal income

Figure 4.4: Lottery expenditure versus log total expenditure (purchasing households only)

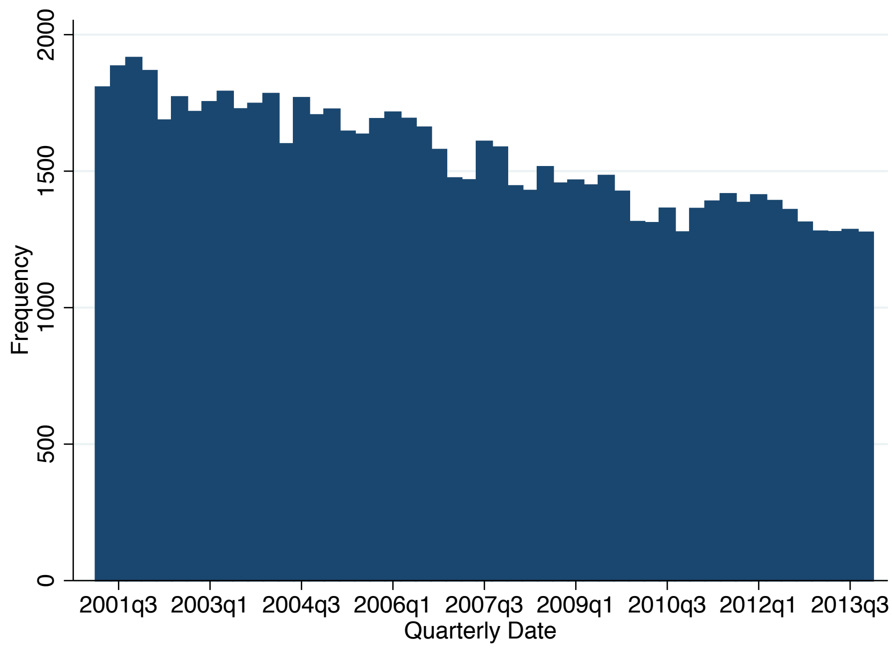


Note: data trimmed to exclude the upper and lower 1% of households by total expenditure

Table 4.2: Number of household-level observations and response rate for each year of EFS/LCF data available

Year	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**
n	5,606	7,041	7,018	6,855	6,696	6,645	6,136
Response rate (%)	62	58	58	57	57	55	53
Year	**2008**	**2009**	**2010**	**2011**	**2012**	**2013**	-
n	5,843	5,822	5,263	5,551	5,473	5,116	-
Response Rate (%)	51	50	50	54	52	48	-

Figure 4.5: Number of household observations by quarter of data sample



4.4 Estimation approach

Due to the flat rate of tax imposed on lottery tickets in the UK, analysis of the tax incidence over the income distribution is equivalent to estimating the income elasticity of demand. This follows simply from the fact that the tax contributed by any given consumer is directly proportional to their expenditure. By far the most common approach in previous studies of the regressivity of lottery products using micro-level data is to estimate Engel curves in which the number of tickets purchased by an individual or household is linear in the level income or expenditure. Only minor deviations from this specification have been considered when using micro-data – for example, Farrell and Walker (1999) include a squared income term – and aggregate-level studies only deviate by taking logs of the regressand. For comparative purposes, Section 4.6.1 briefly replicates this approach by estimating the functional form:

$$E_i = X_i'\beta + \gamma Y_i + \varepsilon_i \tag{4.1}$$

where E_i is expenditure by individual i on lottery tickets, Y_i is some measure of income, and X_i are controls for various demographic characteristics. Though clearly convenient, this naïve specification is deficient in both theoretical and empirical aspects.

Although linear Engel curves as described in equation (4.1) will satisfy the additivity criterion – that income elasticities sum to 1 – when estimated as a set of all commodity groups, such functional forms are unnecessarily restrictive and only result from specific, and unlikely, utility functions (Pollak, 1971). Moreover, Leser (1963) notes that such a specification often produces negative expenditure values within the observed range of incomes and that changes in elasticity across the distribution are not consistent with economic theory (e.g. that the implied income elasticity of inferior goods *increases* with income). As a result, these models are typically 'rejected' by the data since they produce poor fit statistics and large estimated standard errors.

The preferred model of this chapter is one in which the budget share of lottery tickets for a given individual, $w_i = E_i/Y_i$, is linear in log total expenditure, $\ln Y_i$, and estimates are presented in Section 4.6.2. This model was proposed by Working (1943) and later discussed by Leser (1963). The Working-Leser model is formally described as,

$$w_i = X_i'\beta + \gamma \ln Y_i + \varepsilon_i \tag{4.2}$$

where X_i' are again controls for various demographic characteristics[70], β is a vector of parameters to be estimated, and γ is a parameter to be estimated describing the effect of log income, $\ln Y_i$. Since w_i is simply lottery expenditure divided by income, eliciting income elasticity of lottery tickets is trivial and given by $\eta = 1 + \frac{\gamma}{w}$. This model underpins the Almost

[70]Experiments with the specification to include controls for the number of people in the household made very little difference to the resultant estimates of income elasticity. Estimates obtained using Heckman selection and double hurdle routines with household size controls are therefore relegated to Appendix Table A4.19. Accounting for household size with equivalised income resulted in smaller income elasticities than those already reported, some unbelievably so.

Ideal Demand System (AIDS) developed by Deaton and Muellbauer (1980) which is often used with macro-level data to determine the income elasticities of broad commodity groups. The AIDS model is popular in the wider economics literature due, in part, to its derivation from consumer theory and the simple functional form.

The model used here deviates from the AIDS specification only by omitting price information for three reasons. First, the inclusion of price is only indirectly relevant for investigating the relationship between expenditure on lottery tickets and income. Second, due the repeated cross-sectional nature of the data any draw-by-draw price variation, such as that arising from rollovers, for all observations in each 2-week diary window is the same. Thus, variation in observed expenditure resulting from a change in price can be easily captured using year-month fixed effects, which are employed throughout the analysis in Section 4.6. Third, this strategy also circumvents issues posed by the anonymization processes of the data which results in only the month of the diary being included. That is, even if there were some advantage to be gained from including draw-by-draw price as an explanatory variable in the model, doing so is impossible since the data are censored to protect the identity of respondents in such a way that it is not clear which are the relevant draws in the diary window. Using fixed effects for each month within the sample should sufficiently capture variation in expenditure caused by prices, and is unlikely to be improved upon by, say, including average price of lotto within each month which would hardly vary at all. By adopting the Working-Leser specification here, the literature on the income elasticity of lotto and the regressivity of its taxation is brought up to date with the broader economic studies of other goods such as carbon taxes (Brannlund and Nordstrom, 2004), alcohol (Gil and Molina, 2009), and comprehensive studies over several commodity groups (e.g. Banks *et al* 1997).

The use of total expenditure, rather than income, on the right-hand side of equation (4.2) is common throughout the literature and respects a plausible two-stage budgeting process in which utility-maximising households first allocate disposable income between consumption and saving, then decide on the quantities of specific goods – in this case lottery tickets – to be consumed. Total expenditure then reflects the disposable income available to households in the current period. Moreover, reported gross or net incomes – particularly for those who are self-employed – can vary significantly over the short and medium term. Total expenditure, meanwhile, can be seen as a measure of average (or expected) levels of income, thus providing a better measure of income in the long-run.

In Section 4.6.4 the assumption that budget share of lottery tickets is linear in log income is relaxed by estimating the following equation using semi-parametric estimation as described in Robinson (1988),

$$w_i = X_i'\beta + f(\ln Y_i) + \varepsilon_i \tag{4.3}$$

in which f is some unknown function. This agnostic methodology allows full flexibility in the possible shape of f. Moreover, following Hardle and Mammen (1993) it is possible to test whether the function f can be sufficiently approximated using polynomial expansions of ln Y_i. This essentially becomes a test of whether the functional form imposed by the Working-Leser specification in equation (4.2) is rejected by the data.

4.5 Econometric issues

Two principal empirical issues arise in the application of a Working-Leser model as defined in equation (4.2) to household-level data. The first is the presence of zeroes in expenditure data which would bias estimates obtained using OLS. The second is the endogeneity of total expenditure with budget shares of lottery tickets.

4.5.1 The presence of zeroes in expenditure data

As is common with demand studies which use micro-level data, the analysis here is complicated by non-purchasers of lottery tickets having their observed expenditure – and, by extension, their budget share – truncated at zero. For the present sample only 43% of households purchased a ticket for the UK National Lottery in the two-week diary window. Though significantly smaller than headline figures for annual participation (e.g. BGPS, 2010, estimates 59% of adults purchased tickets in the previous 12 months), the observed proportion of ticket buyers is broadly in-line with short-term estimates (e.g. in the past month) which ranges from 36% to 48% for participation in the previous four weeks (Gambling Commission, 2017).

Several empirical models exist to deal with such censored data. A simple solution is presented in Tobin (1958) – commonly known as the Tobit model – which estimates,

$$w_i^* = X_i'\beta + \gamma \ln Y_i + \varepsilon_i \tag{4.4}$$

where w_i^* is some latent variable (related to expenditure share of lottery tickets), $X_i, Y_i, \beta,$ and γ are defined as before and,

$$w_i = \begin{cases} w_i^* \; if \; w_i^* > 0 \\ 0, \text{otherwise} \end{cases}. \tag{4.5}$$

Such a model can then be estimated using maximum likelihood and is easily implemented in most statistical software packages. However, when taken to expenditure data a potentially unpalatable implication of the Tobit model is that the regressors have the same effect on both the decision to purchase and the quantity purchased conditional on the former being true. For instance, the Tobit model implies that a change for a given level of income would have the same effect on both the probability of purchasing a lottery ticket and the number of lottery tickets to buy – which may not necessarily be the case.

Section 4.6.2 presents estimates from two alternatives to the Tobit model to allow for the possibility that the regressors may have differing effects on the binary decision to purchase lotto tickets and, conditional on purchase, the quantity of tickets to buy. Heckman (1979) generalises the Tobit model by modelling selection in participation separately to the amount spent on lotto tickets. Specifically, the two-step model in the context of purchasing lottery tickets can be written as,

$$w_i = X_{1i}'\beta + \gamma \ln Y_i + \varepsilon_i \tag{4.6}$$

where w_i, X_{1i}, Y_i, β and γ are as previously defined and,

$$\Pr\,(w_i > 0) = X_{2i}'\psi + \nu_i \tag{4.7}$$

where ψ is a vector of parameters to be estimated, X_{2i} are variables which determine whether household i purchased a lotto ticket. Heckman's selection model further assumes the errors of the two equations in the model are jointly normally distributed such that,

$$\varepsilon \sim N(0, \sigma)$$

$$\nu \sim N(0,1)$$

$$corr(\varepsilon, \nu) = \rho.$$

Equation (4.6) is again the budget share equation for lotto tickets and equation (4.7), estimated via probit, is the selection equation which determines whether w_i is observed (i.e. whether $w_i > 0$). Heckman's selection model corrects for the bias, which arises because of unobserved

w_i being recorded as 0, by including the estimated inverse Mill's ratio[71], λ, of the selection equation in X_{1i}. Heckman (1979) shows that the coefficient on the inverse Mill's ratio is $\rho\sigma$ and testing whether $\rho\sigma = 0$ amounts to a test of whether the model could have been estimated using either the Tobit specification in equation (4.4) or via OLS as in equation (4.2). Implicit in Heckman's methodology for correcting zero-induced bias is that when the dependent variable, w_i, is recorded as zero, it is because of an inability to measure its true value and *not* because individual i chose not to purchase any lotto tickets. The resulting estimates of β and γ can therefore be interpreted as the coefficients which would have been obtained had we been able to observe the true budget share of lottery tickets for all households.

Identification of the Heckman selection model can be achieved in one of two ways. If X_2 is comprised of the set of all variables from the right-hand side of equation (4.6), identification relies on the functional form imposed by the Heckman model being correctly specified. Whilst convenient, relying on the functional form being correct for identification requires an appeal to theory for support and must, in essence, remain an unverifiable assumption. The second approach to identification – and the standard in the literature – calls for an exclusion restriction (Heckman, 1990). The model can be identified when X_2 also contains a variable which affects the probability of purchasing a lottery ticket, but not the budget share conditional on an affirmative response to the former. Importantly, the issue of identification here is purely a statistical one and, in some ways, the theoretical justifications for choosing one instrument, or set of instruments, over another are arguably less important than, say, for causal identification. Rather, it is the statistical properties – that the instrument is uncorrelated with the level of spend conditional on play but correlated with participation – that are most important.

Nonetheless, the hypothesis underpinning the identification strategy employed here is that it exploits exogenous differences in individuals' consumption which arise from religious practice. The obvious approach would be to include indicator variables for religion directly, however this is not recorded in the data. Instead, this chapter exploits the effects of religious practices on consumption – which *is* observable in expenditure data – to identify the Heckman selection model similar to Pryce (2016).

[71] The inverse Mill's ratio for some random variable X is defined as $\lambda(X) = \phi(X)/\Phi(X)$ where ϕ and Φ are the standard normal density and cumulative distribution functions, respectively.

To implement this idea, dummy variables are included which indicate whether household i does *not* purchase any alcohol and/or pork in the two-week diary window in the selection equation. Islam – the second largest religion in England and Wales according to the 2011 Census – prohibits gambling and the consumption of both alcohol and pork products. A non-purchasing household of lottery tickets who does so because the occupants practice Islam, for instance, would also not purchase alcohol or pork. Thus, if the identification strategy is correct, the coefficient on both these dummies should be negative when estimating equation (4.7). The need to use dummies for non-purchasers of both alcohol and pork, rather than just one, arises because of the effect on consumption preferences for followers of other religions. Judaism, for instance, which also prohibits the consumption of pork does not prohibit either gambling or alcohol consumption. This would then clearly increase one's prior on the value of the coefficient on non-pork consumption in the selection equation, but it is unclear whether the increase would be large enough to cause the estimate to become positive. The confounding effect of multiple faiths being practiced in the population reduces the effectiveness of pork consumption alone as a variable with which to identify the Heckman selection model. It is possible, however, to capture the predicted differences in consumption from observing religious practice by including an interaction between alcohol and pork consumption. Therefore, if the proposed exclusion restriction is justified, the expectation is for a negative coefficient on the interaction term. If this is true, then it seems likely that this effect is driven primarily by households who do not purchase lottery tickets because of their religious belief.

Table 4.3 compares the proportion of households who purchase lotto tickets between the purchasing and non-purchasing households of alcohol and pork. The proportions provide encouraging evidence that both alcohol and pork are correlated with the decision to purchase lotto tickets. Only 27.5% of households who did not purchase pork and alcohol purchased lotto tickets, compared with 51.6% of households that bought both pork and alcohol. Examination of the totals columns reveals that individually purchasing of pork or alcohol are themselves correlated with the purchase of lotto tickets as the identification strategy outlined above requires – albeit to a smaller extent. Only 35.6% of non-purchasers of pork bought a lotto ticket compared to 48.7% for those who did purchase pork, and 33.5% of non-alcohol households participated in lotto compared to 47.5% for alcohol purchasing households.

Table 4.3: Proportion of households who purchased lotto by whether they also purchased alcohol and pork

	No Alcohol	Bought Alcohol	Total
No Pork	0.275	0.409	0.356
	n=13,432	*n=20,752*	*n=34,184*
Bought Pork	0.405	0.516	0.487
	n=11,588	*n=33,285*	*n=44,873*
Total	0.335	0.475	0.431
	n=25,020	*n=54,037*	*n=79,057*

Analogous to the assumptions required for an instrumental variables approach, the usefulness of religious belief, which is (imprecisely) observed here through the non-purchase of alcohol and pork, as an exclusion restriction to identify the Heckman selection model also relies on the assumption that it is uncorrelated with the level of play. That is, being Muslim should have no effect on the observed budget share of lottery tickets beyond the effect it has on the selection equation. If being Muslim meant a decreased (or increased) level of play, relative to non-Muslims, *if* they participated in lotto would preclude its usefulness as an exclusion restriction. Whilst there is no way to examine this directly – non-purchasers of lotto who also do not purchase pork or alcohol because of their religious belief necessarily do not reveal their expenditure if they were to play – Table 4.4 provides some evidence that the non-purchase of pork and alcohol has no significant impact on lotto expenditure amongst those who do participate. Whilst expenditures on lotto are lowest for those households who purchase no pork or alcohol, the mean difference is small – only around £0.50 per week less than the sample average – and not statistically different from alcohol and/or pork purchasing households.

Table 4.4: Mean expenditure on lotto by alcohol and pork consumption conditional on participation (s.d. in parentheses)

	No Alcohol	Bought Alcohol	Total
No Pork	3.757	4.115	4.006
	(4.2977)	(4.9637)	(4.7742)
	n=3,697	*n=8,489*	*n=12,186*
Bought Pork	4.069	4.550	4.446
	(4.3524)	(4.6930)	(4.6262)
	n=4,689	*n=17,160*	*n=21,849*
Total	3.931	4.406	4.289
	(4.3309)	(4.7886)	(4.6844)
	n=8,386	*n=25,649*	*n=34,035*

There are, of course, other reasons why participation and expenditure on lotto may vary with alcohol and pork consumption. The most obvious is income, which is controlled for in the selection equation anyway, but others may well exist. Again, whilst the issue here is purely a statistical one, there may remain doubts over whether alcohol and pork consumption are sufficiently exogenous to consumption conditional on lotto participation to be used as an exclusion restriction. To help address this concern, Table 4.5 and Table 4.6 present average lottery participation and expenditure conditional on play, respectively, by the purchase of two innocuous goods: poultry and potatoes[72]. The choice to use these two goods is entirely arbitrary; there are plenty of candidates which could have been used instead, and which produce the similar conclusions, since most items purchased are unlikely to correlate with lottery consumption on religious grounds.

Table 4.5: Proportion of households who purchased lotto by whether they also purchased potato and poultry

	No Potato	Bought Potato	Total
No Poultry	0.379	0.408	0.394
	n=19, 435	*n=18,895*	*n=38,330*
Bought Poultry	0.478	0.460	0.465
	n=12,334	*n=28,401*	*n=40,735*
Total	0.417	0.439	0.430
	n=31,769	*n=47,296*	*n=79065*

Table 4.6: Mean expenditure on lotto by potato and poultry consumption conditional on participation (s.d. in parentheses)

	No Potato	Bought Potato	Total
No Poultry	4.158	4.214	4.187
	(4.4666)	(4.6647)	(4.5689)
	n=7,371	n=7,718	n=15,089
Bought Poultry	4.393	4.358	4.369
	(5.0427)	(4.6459)	(4.7727)
	n=5,890	n=13,058	n=18,948
Total	4.262	4.305	4.288
	(4.7324)	(4.6533)	(4.6842)
	n=13,261	n=20,776	n=34,037

[72] The need to use two innocuous goods here is due to the fact that comparing with alcohol or pork would still have effects on participation and expenditure induced by religious belief.

The relationship seen between lottery participation and pork and alcohol purchases is not repeated here. Whilst there is a slight reduction in the participation rate (8 percentage points) for households who purchase neither poultry nor potatoes versus those who buy both, this is far from the difference seen with pork and alcohol non-purchasers (24 percentage points). Similarly, there appears to be no correlation between poultry and pork purchasing and lottery expenditure, conditional on play. As such, the choice to use pork and alcohol purchase as the exclusion restriction in estimating the Heckman selection model seems reasonably robust based on the religious arguments above and these descriptive statistics.

Alternative candidates to the proposed identification strategy may consider exploiting the, perhaps more obvious, correlations between lotto play and other risky behaviours such as other gambling expenditure, smoking, alcohol consumption. Such a simple argument would almost certainly be correlated with lottery participation and such relationships are well documented in the economics literature. Griffiths *et al* (2010) is one such example using BGPS 2007 data that finds positive correlations between gambling behaviour and smoking and alcohol consumption. Griffiths and Sutherland (1998) highlights the correlation between (underage) gambling on the UK National Lottery or scratchcards and smoking, drug use, alcohol consumption, and other 'undesirable' behaviour[73] in adolescents. Indeed, correlations in such consumption data fit with the notion of 'risky' or 'addictive' personality traits. An alternative to using consumption altogether, if such data existed alongside lotto expenditure and income, might well be to include any manner of risk-seeking indicators.

However, whilst the correlation between all of these candidate identification strategies and lotto participation is well founded, this rationale would undoubtedly not satisfy the second criteria for statistical identification: that consumption of these goods (by any measure) on the grounds that lotto players' exhibit an increased propensity for risk seeking behaviour more generally would be uncorrelated with the level of lotto play conditional on participation.

Whilst the use of alcohol alone is ill-advised for the reasons set out above, the approach employed here relies on its interaction with pork consumption, and that this effect is induced by exogenous differences in preferences arising from religions practice. Therefore, in the absence of a more obvious and robust identification candidate (or set of candidates), and encouraged by the evidence in Tables 4.3 to 4.6, this chapter pursues the use of (non-

[73] Contact with the police, suspension from school and failing at school.

)consumption of pork and alcohol for identification of the Heckman and double-hurdle models outlined above.

The final econometric model considered in this paper to handle zeroes in the expenditure data is the double-hurdle model proposed by Cragg (1971). Comparing estimates from the Heckman selection model with Cragg's double-hurdle model addresses directly the concerns raised in Humphreys, Lee and Soebbing (2010) who argue that the Heckman selection model is inappropriate with micro-level data if zeroes are observed, not because of an inability to observe the true value of w_i, but because the purchase of lottery tickets is infrequent. Cragg's double-hurdle model used in the present context can be written as,

$$w_i = \delta_i w_i^* \tag{4.8}$$

where w_i is the observed expenditure share of lottery tickets for household i, w_i^* is the latent variable related to the budget share of lottery tickets which behaves as a conventional Tobit,

$$w_i^* = X_{1i}'\beta + \gamma \ln Y_i + \varepsilon_i \tag{4.9}$$

and the participation variable, δ_i, is defined as,

$$\delta_i = \begin{cases} 1, \text{ if } X_{2i}\psi + \nu > 0 \\ 0, \text{ otherwise} \end{cases}. \tag{4.10}$$

Again, the double-hurdle model presents a generalisation of the Tobit model but differs in its approach in dealing with the presence of zeroes in the data compared to the Heckman selection model. Whereas Heckman's selection model first estimates the selection equation and uses the estimated inverse Mill's ratio to correct for bias in the second-stage participation equation, Cragg's double hurdle estimation routine estimates equations (4.9) and (4.10) simultaneously. Moreover, whilst the Heckman selection model implies that zeroes are present because of an unobservable response, the double-hurdle model instead assumes that w_i is only observed if both $w_i^* > 0$ and $\delta_i = 0$. This means another way to express equation (4.8) would be to say,

$$w_i = \begin{cases} w_i^*, if \;\; w_i^* > 0 \;\text{ and } \delta_i > 0, \\ 0, \text{otherwise.} \end{cases} \tag{4.11}$$

Therefore, for some positive value of the budget share of lotto tickets to be observed requires a positive outcome of both the participation decision (equation 4.10) *and* the consumption decision (equation 4.9), giving rise to the name "double-hurdle". This implies that observing zeroes occurs a result of household i *choosing* not to consume lotto tickets, rather than their choice of budget share being unobservable (Perez and Humphreys, 2013).

Finally, as a two-stage estimation routine, the use of a double hurdle model experiences the same identification issues as the Heckman selection model. Humphreys, Lee and Soebbing (2010) note that whilst not explicitly required, an exclusion restriction is common practice in the existing literature in the application of double hurdle models. This identification issue can be overcome in the same way as for the Heckman selection model – by including dummy variables for the non-purchase of alcohol and pork to the vector of variables X_2 in the selection equation defined in equation (4.10).

4.5.2 Endogeneity of total expenditure

The final econometric issue to be dealt with is the endogeneity between (log) total expenditure, Y_i, and the budget share of lottery tickets which is necessarily defined as $w_i = E_i/Y_i$. This specification clearly induces a simultaneity bias since the expenditure share of lotto tickets is itself a function of total expenditure and precludes a causal interpretation on the effect of income on lottery expenditure from estimates obtained via basic methodologies. A common solution to this endogeneity problem is to employ an instrumental variables approach and this paper follows the standard set in previous literature (e.g. De Agostini, 2014, Farrell and Shields, 2007, Banks *et al*, 1997) by using gross total income as an instrument for log total expenditure.

The validity of total income as an instrument relies on two conditions. First, the instrument must be correlated with the endogenous regressor, a condition which is easily verified using the Staiger and Stock (1997) rule of thumb that an F-test statistic of the significance of the instrument coefficients in the first-stage exceeds 10. Second, the instrument must also only affect the dependent variable of interest – the budget share of lottery tickets – via its effect on the endogenous regressor. However, in the just identified case – where the number of instrumental variables equals the number of endogenous regressors as there are here – it is not a testable condition and must remain be an article of faith. The validity of gross income as an instrument, in particular satiating this second condition, requires an appeal to theory and the two-stage budgeting process described in Gorman (1959) in which households first divide their gross income between broad categories of goods (including savings), before deciding on shares of those budgets to be allocated on individual products in a second stage.

As explained in De Agostini (2014) total income will therefore be highly correlated with total expenditure as part of the first-stage of this budgeting process but exogenous to the specific good (in this case, lottery tickets) in the second stage.

4.6 Results

The estimation results are presented as follows. Section 4.6.1 covers estimates of the naïve linear model presented in equation (4.1). Though this model is largely inappropriate when taken to the data for reasons outlined above, it is included here to allow for comparisons to be made with results from previous studies. Section 4.6.2 presents estimates of the Working-Leser specification as presented in equation (4.2) onwards. Section 4.6.3 presents Suits' (1977a) index for the regressivity of lotto taxation and compares this index for each year of data available to asses any possible changes in the regressivity of these taxes over time. Finally, Section 4.6.4 presents non- and semi-parametric explorations of the appropriateness of the Working-Leser functional form.

4.6.1 Linear expenditure model estimates

In order to directly compare income elasticity estimates – and therefore whether lottery duty is regressive – with the existing literature, Table 4.7 presents estimates from a linear expenditure model as described in equation (4.1). Here, weekly expenditure on lotto tickets is the dependent variable and is assumed to be linear in weekly total expenditure as the measure of income, controls for the sex, age, employment status and education level of the HRP, region, and month-year fixed effects.

Estimates from OLS and Tobit estimation of the linear expenditure model in Table 4.7 are strikingly similar to those in Farrell and Walker (1999) whose study of the UK game was taken before there was sufficient data available from the Family Expenditure Survey – the predecessor to the Expenditure and Food Survey used here. Their estimates using survey data gathered on behalf of the Office of the National Lottery (OFLOT) produced income elasticity estimates of 0.27 and 0.45 using OLS and Tobit estimation, respectively, suggesting that lottery tickets are a normal good, and that the flat-rate tax is indeed regressive. The estimates here present an income elasticity of 0.22 when using OLS. When using Tobit estimation to correct for bias induced by zeroes in the data, the income elasticity estimate doubles to 0.46, still indicative of a regressive tax. In both cases, a positive coefficient on total expenditure suggest that expenditure on lotto tickets increases with income – though it is unclear whether this is

Table 4.7: OLS and Tobit estimates of equation (4.1)

	(1)	(2)
	OLS	Tobit
Dependent Variable	E_i	E_i
Total Expenditure (/1000)	0.921***	1.924 ***
	(0.0633)	(0.1269)
Female HRP	-0.457***	-1.043***
	(0.0272)	(0.0594)
Age HRP	0.030***	0.067***
	(0.0012)	(0.0028)
Self-employed	-0.332***	-1.144***
	(0.0557)	(0.1126)
Unemp. seeking work	-0.588***	-1.831***
	(0.0646)	(0.1845)
Unemp about to work	-0.184	-1.269
	(0.4507)	(1.0668)
Sick	-0.533***	-1.540***
	(0.0610)	(0.1283)
Retired	-1.122***	-2.803***
	(0.0530)	(0.1074)
Unoccupied	-0.544***	-2.027***
	(0.0513)	(0.1398)
Left Education 17-18	-0.676***	-1.579***
	(0.0362)	(0.0791)
Left Education 18+	-1.364***	-3.902***
	(0.0336)	(0.0905)
Constant	1.829***	-0.722**
	(0.1580)	(0.3238)
Income Elasticity	0.222***	0.464***
	(0.0149)	(0.0306)
Month-Year Fixed Effects	Yes	Yes
Observations	78,867	78,867
R-squared	0.053	-

Robust standard errors in parentheses. */**/*** denotes statistical significance at the 1%/5%/10% level, respectively. Omitted categories: Male HRP, employee, left education before 17, North East. Regional controls omitted for brevity. Income elasticity calculated as $\eta = \gamma * \frac{\bar{Y}}{\bar{E}}$ where γ is the coefficient on log total expenditure, $\bar{Y}$ is mean total expenditure in thousands, and $\bar{E}$ is mean expenditure on lottery tickets. Full estimates are presented in Appendix Table A4.12.

due to increased participation or higher expenditures amongst those who would participate at lower incomes. The OLS estimate of income elasticity also matches the upper bound of Spiro (1974) from a study of the Pennsylvania state lottery in the 1970s. The Tobit estimate from Table 4.7 is in-line with the implied estimate from Humphreys, Lee and Soebbing (2010) of

0.40, but is significantly smaller than the elasticities obtained by Kitchen and Powells (1991), 0.70-0.92, in their respective studies of Canadian lotteries.

There is little difference in the coefficients of demographic controls when comparing between OLS and Tobit estimates of the linear expenditure model. Households with a Female HRP spend significantly less than their Male HRP counterparts, as do those whose HRP's employment status is anything other than an employee or about to start work. Increasing levels of education for the HRP – as determined by years of schooling – also correlates with lower spending on lotto tickets. These estimates are consistent with the existing literature on the demographic profile of gamblers, even though the demographics used here are only those of the HRP.

As discussed in Section 4.4, however, these results should be treated with caution due to the deficiencies associated with the assumed functional form. Notably, the estimated R-squared resulting from this model is characteristically low and concurrent with previous studies using such a specification – making the estimates unreliable and effectively rejected by the data. Moreover, naively estimating Equation (4.1) via OLS and Tobit ignores the endogeneity which necessarily arises from lottery expenditure being a component of total expenditure, and the questionable assumption of Tobit estimation that the regressors have the same effect on expenditure and participation. The use of instrumental variables and alternative statistical models such as Heckman selection and double-hurdle routines have been employed by previous literature in this setting, as discussed in Section 4.2. Results from these methodologies typically estimate income elasticity to be smaller than those obtained from OLS or Tobit, indicating that the latter are understating the regressivity of taxation on lotto.

4.6.2 Working-Leser model estimates

Columns 1 and 2 of Table 4.8 present OLS and Tobit estimates of equation (4.2) and act as baseline results for which direct comparisons can be made with the income elasticity estimates of the linear expenditure model in Table 4.7. Using the naïve OLS and Tobit methods, we find income elasticities significantly larger than those under the linear expenditure model at 0.68 and 1.03, respectively. Similar to the estimates presented in the previous section, the Tobit coefficient on log total expenditure (0.158) is significantly larger than OLS (-1.859), and is not statistically different from 0. The OLS estimates imply that as income increases, the expenditure share on lotto decreases and would unambiguously render a flat-rate tax on lotto as regressive – as is supported by an income elasticity of 0.675. The coefficient estimate obtained by Tobit (0.158) is not statistically different from 0, however, and implies that the budget share of lotto tickets is constant across the income distribution suggesting a flat rate of tax on lotto is proportional. This is again supported by the estimated income elasticity of 1.03 and not being statistically different to 1. Under the Working-Leser model, the control coefficients yield similar conclusions about the profile of lottery players. Having a female, non-employee and better educated HRP are all correlated with a lower expenditure share of lottery tickets in all specifications presented in Table 4.8.

However, the estimates in columns 1 and 2 are still unreliable for making inference about the income elasticity – and therefore the regressivity of taxation – of lottery tickets. Column 3 re-estimates equation (2) using two-stage least-squares (2SLS) in order to correct for the simultaneity bias that arises from the functional form of the Working-Leser model. Log total expenditure is instrumented using gross income and the results of the first stage of this procedure are contained in Appendix Table A4.14[74]. The resulting coefficient on log income is significantly larger in magnitude at -2.35 and the resultant income elasticity of 0.589 suggests a higher level of regressivity from taxation than the estimates obtained via OLS. The IV Tobit estimates in column 4 highlight even further the need to account for simultaneity bias than do the regular 2SLS estimates. Similar to the 2SLS comparison with OLS, the IV Tobit estimate of the coefficient on log total expenditure (-3.358) is significantly smaller than the un-instrumented version in column 2, and much more so than with 2SLS. Moreover, the corresponding estimate of income elasticity is smaller still at 0.412.

The final deficiency of the above estimates is addressed with the results in Table 4.9 which presents estimates of the Working-Leser specification using both Heckman's selection routine and Cragg's double-hurdle estimator. Both have advantages over the Tobit estimates in Table 4.8 since, in both models, the regressors are free to have differing effects on both the decision to purchase decision and the expenditure decision when a consumer considers the purchase of a lottery ticket. Throughout the remaining analysis log total expenditure is

[74] As discussed in Section 4.5.2, in order for the instrumental variables technique to be valid two conditions must be satisfied: gross income must be correlated with the endogenous regressor, log total expenditure, and uncorrelated with the budget share of lotto. Whilst the latter condition must remain an article of faith in this just-identified case, the former is clearly satisfied following the Staiger and Stock (1997) rule that the first-stage F-statistic be greater than 10 is easily satisfied with the reported F-statistic of 68.67. The coefficient on total income in the first stage is both positive and significant at the 1% level and the R-squared is predictably large at 0.47 thus there is no concern about a possible weak instrument problem.

instrumented using gross normal income with the estimates from Table A4.14 as is done in the IV columns of Table 4.8.

Table 4.8: OLS and Tobit estimates of equation (4.2) using un-instrumented and instrumented log total expenditure

	(1)	(2)	(3)	(4)
	OLS	Tobit	IV (2SLS)	IV Tobit
Dependent Variable	w_i	w_i	w_i	w_i
Ln(Total Expenditure) (/1000)	-1.859***	0.158	-2.350***	-3.358***
	(0.0848)	(0.1558)	(0.0955)	(0.6485)
Female HRP	-0.002***	-0.003***	-0.002***	-0.004***
	(0.0001)	(0.0002)	(0.0001)	(0.0003)
Age HRP	0.000***	0.000***	0.000***	0.000***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Self-employed	-0.001***	-0.004***	-0.001***	-0.004***
	(0.0001)	(0.0003)	(0.0001)	(0.0003)
Unemp. seeking work	-0.000	-0.002***	-0.001***	-0.005***
	(0.0003)	(0.0007)	(0.0003)	(0.0009)
Unemp. about to work	-0.000	-0.002	-0.000	-0.004
	(0.0013)	(0.0034)	(0.0013)	(0.0035)
Sick	0.000	-0.001**	-0.000	-0.004***
	(0.0003)	(0.0005)	(0.0003)	(0.0007)
Retired	-0.002***	-0.006***	-0.002***	-0.008***
	(0.0002)	(0.0003)	(0.0002)	(0.0005)
Unoccupied	-0.001***	-0.005***	-0.001***	-0.006***
	(0.0002)	(0.0005)	(0.0002)	(0.0006)
Left Education 17-18	-0.002***	-0.005***	-0.002***	-0.005***
	(0.0001)	(0.0002)	(0.0001)	(0.0003)
Left Education 18+	-0.003***	-0.012***	-0.003***	-0.011***
	(0.0001)	(0.0003)	(0.0001)	(0.0003)
Constant	0.018***	-0.002	0.020***	0.019***
	(0.0008)	(0.0014)	(0.0008)	(0.0040)
Income Elasticity	0.675***	1.028***	0.589***	0.412***
	(0.0149)	(0.0273)	(0.0167)	(0.1135)
Demographic Controls	Yes	Yes	Yes	Yes
Month-Year Effects	Yes	Yes	Yes	Yes
1st Stage F-Statistic	-	-	65.264***	65.264***
Observations	78,867	78,867	78,867	78,867
R-squared	0.073	-	0.072	-

Robust standard errors in parentheses. */**/*** denotes statistical significance at the 1%/5%/10% level, respectively. Demographic controls omitted from reporting for brevity. For full estimates see Appendix Table A4.13. In all columns ln(total expenditure) is divided by 1000 to allow coefficients to be legible. Income elasticity is calculated as $\eta = 1 + [\gamma/(1000 * 0.00571)]$ where γ is the coefficient on log total expenditure and 0.00571 is the mean expenditure share of lottery tickets. Columns 3 and 4 have ln(total expenditure)/1000 instrumented by gross income. The first stage is available in Appendix Table A4.14. Columns 2 and 4 are estimated by maximum likelihood. Omitted categories: Male HRP, employee, left education before 17, North East.

Since they are identical, column 1 of Table 4.9 presents points estimates from the selection equation for both the Heckman selection and double hurdle routines. The dependent variable is a dummy equal to 1 if household i purchased a lottery ticket and 0 otherwise and is estimated for both the Heckman selection and double hurdle models using probit. Column 2 presents the marginal effects of the second stage equation for the Heckman selection estimation – the parameters of interest from equation (4.6). Column 3 concludes with the marginal effects of from the second stage of the double hurdle model – equation (4.9).

In the selection equation, the inclusion of dummies for no alcohol and no pork consumption and an interaction term form the identification strategy appears justified since, as expected, the coefficients are all negative and highly significant. This is consistent with the identification strategy set out in Section 4.5.1 and can be explained using the hypothesis of exogenous differences in consumption behaviour arising from religious practice. These dummies and the interaction term are omitted from the second-stages in columns 2 and 3 as the identifying exclusion restriction. Though negative, the point estimate of the income coefficient in the selection equation is insignificant suggesting that income plays no role in the purchase decision of households. This finding is consistent with that of Humphreys, Lee, and Soebbing (2010). Interestingly, point estimates of the coefficients on the demographic controls in the selection equation mirror those presented in Table 4.7 and Table 4.8. The significance and direction of coefficients indicate households with female, educated, and non-employee HRPs are less likely to even participate in the UK lotto than their counterparts in the omitted categories.

Column 2 reports the marginal effects of the variables of interest from equation (4.6) of the Heckman selection model. There is a negative and highly significant relationship between income and the budget share of lottery tickets – suggesting that separating the estimation of purchase decision and expenditure decision is justified. The implied income elasticity of -0.629, is considerably lower than existing estimates. These Heckman estimates imply that lottery tickets are inferior goods, a novel finding for the literature, and would suggest that the regressivity of lotto taxation is more severe than previously thought. Demographic effects are all of the same direction as in the selection equation suggesting that individuals in the demographics identified above are not only less likely to purchase lottery tickets but also to spend less conditional on participation. The significant and positive coefficient on the

Table 4.9: Heckman selection and double hurdle selection equation estimates and consumption equation marginal effects of the Working-Leser model

	(1) Selection Equation	(2) Heckman	(3) Double Hurdle
No Alcohol	-0.298***	-	-
	(0.0143)		
No Pork	-0.198***	-	-
	(0.0115)		
No Alcohol * No Pork	-0.077***	-	-
	(0.0204)		
Ln(Total Expenditure) /1000	-24.832	-9.640***	-32.165***
	(18.1572)	(0.4238)	(8.6294)
Female HRP	-0.123***	-0.003***	-0.062***
	(0.0110)	(0.0002)	(0.0052)
Age HRP	0.008***	0.000***	0.004***
	(0.0005)	(0.0000)	(0.0002)
Self Employed	-0.236***	-0.002***	-0.114***
	(0.0179)	(0.0004)	(0.0085)
Unemp. (Seeking work)	-0.313***	-0.004***	-0.156***
	(0.0362)	(0.0008)	(0.0166)
Unemp. (About to work)	-0.287*	0.000	-0.137*
	(0.1607)	(0.0035)	(0.0750)
Sick	-0.274***	-0.003***	-0.137***
	(0.0258)	(0.0005)	(0.0120)
Retired	-0.451***	-0.005***	-0.217***
	(0.0192)	(0.0004)	(0.0087)
Unoccupied	-0.375***	-0.004***	-0.181***
	(0.0249)	(0.0006)	(0.0113)
Left Education 17-18	-0.241***	-0.003***	-0.117***
	(0.0136)	(0.0003)	(0.0064)
Left Education 18+	-0.601***	-0.006***	-0.275***
	(0.0142)	(0.0004)	(0.0060)
Constant	0.625***	-	-
	(0.1258)		
Income Elasticity	-	-0.687***	-1.424**
		(0.0742)	(0.6503)
λ	-	0.011***	-
		(0.0007)	
Demographic Controls	Yes	Yes	Yes
Month-Year Effects	Yes	Yes	Yes
Observations	78,867	78,867	78,867

Robust standard errors in parentheses. */**/*** denotes statistical significance at the 1%/5%/10% level, respectively. Regional controls omitted from reporting for brevity. For full estimates see Appendix Table A4.15. In all columns ln(total expenditure) is divided by 1000 to allow coefficients to be legible. Income elasticity is calculated as $\eta = 1 + [\gamma/(1000 * 0.00571)]$, where γ is the coefficient on log total expenditure and 0.00571 is the mean expenditure share of lottery tickets. The first stage is available in Appendix Table A4.14. Columns 2 and 4 are estimated by maximum likelihood. Omitted categories: Male HRP, employee, left education before 17, North East.

inverse Mill's ratio, λ, is further evidence that these estimates should be preferred over those obtained via Tobit and that the purchase equation and expenditure equation are positively correlated, as one would expect.

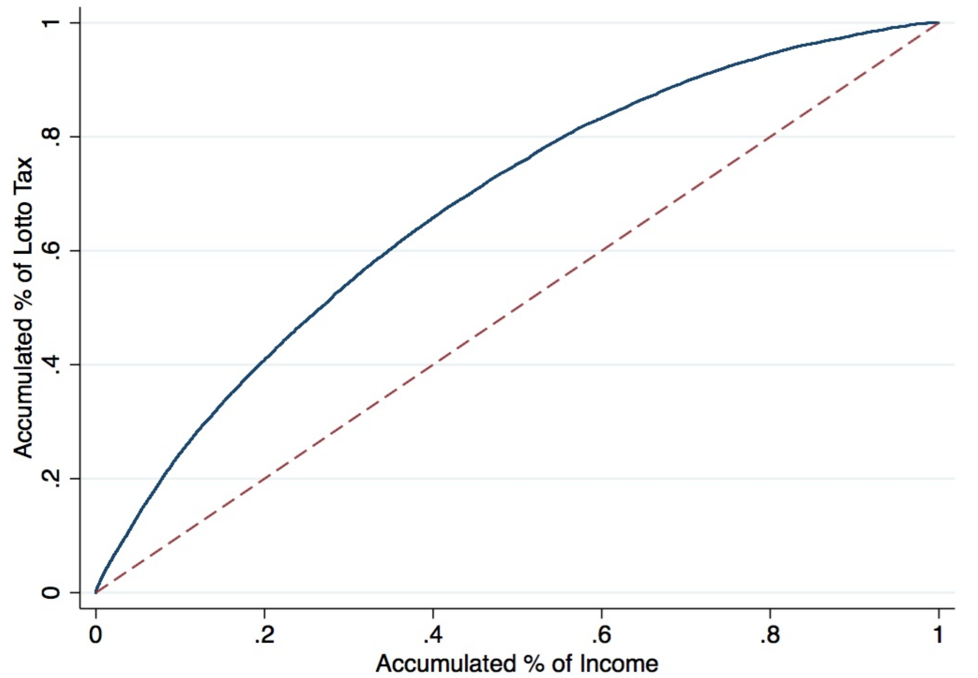
To allow examination of the arguments of Humphreys, Lee, and Soebbing (2010), that the Heckman selection model is inappropriate in the context of lotteries because of infrequent participation, column 3 presents the marginal effects of the second stage from Cragg's (1971) double hurdle procedure. When compared with estimates on column 2, there is little *qualitative* difference between the two methods for all covariates. Income is again negatively related to expenditure on lottery tickets, demographic controls broadly have the same effects, and the resultant income elasticity estimate again suggests lotteries are in fact inferior goods and that flat-rate taxes are highly regressive. Quantitatively, however, the marginal effect of income on the expenditure share of lottery tickets is over three times larger at -32.165 and the estimated income elasticity is over twice as large at -1.424 and, though not as well determined as in column 2, is still significant at the 5% level. The theoretical superiority of the double hurdle model to handle infrequent purchases – as is likely the case with lottery tickets – makes -1.424 the preferred income elasticity estimate and suggests that the estimate of -0.629 from Heckman estimation is an upper bound.

4.6.3 Suits' index

With all estimates presented indicating that the flat-rate of taxation of lottery tickets is regressive, the natural question is to ask *how* regressive are such taxes? As described in Section 4.2, Suits (1977a) proposes a simple, nonparametric index, S, ranging from -1 for extreme regressivity to +1 for extreme progressivity of taxes, with 0 denoting a perfectly proportional tax. Graphically, Suits' index for lotteries using the present data can be derived from Figure 4.6. This plots the accumulated percent of income, sorted by the income of each household, on the x-axis against accumulated percent of lotto tax paid on the y-axis. Then, defining the area beneath the solid line which plots this relationship from the data as L, and the area beneath the 45-degree dashed line as K – which represents a perfectly proportional tax – Suits' index is simply $S = 1 - \frac{L}{K}$. When $L > K$, as is the case here, this index is clearly less than 0 and the lottery tax is defined as regressive. The area K is trivially 0.5, and the estimate of L using numerical integration with the present data is 0.679, giving a Suits' of -0.357 which lies towards the lower end of the range of previous estimates. Only Brinner and Clotfelter (1975) and

Clotfelter and Cook (1987) estimated a lower Suits' index than the one reported here at -0.46 and -0.48 in their respective studies of US lotteries.

Figure 4.6: Graphical representation of Suits' index of tax incidence for the UK lotto



The long time span which is covered by the data gives the unique opportunity to examine whether lotteries have become more regressive over time. Table 4.10 presents Suits' index estimates for each year of the available data. There appears to be little difference between the estimates for each year with the least regressive year occurring in 2002 ($S = -0.3319$) and the most regressive year in 2005 ($S = -0.3739$).

Table 4.10: Suits' index estimates for each year of available data

Year	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**
Suits'	-0.3599	-0.3319	-0.3594	-0.3369	-0.3739	-0.3610	-0.3456
Year	**2008**	**2009**	**2010**	**2011**	**2012**	**2013**	**All years**
Suits'	-0.3681	-0.3299	-0.3561	-0.3424	-0.3076	-0.3367	-0.3574

4.6.4 Semi-parametric estimation of the Working-Leser model

Finally, this section returns to the Working-Leser model considered above but relaxes the assumption that the budget share of lotto tickets is log linear in total expenditure. Instead, a completely agnostic approach is taken towards the functional form by allowing the data to dictate the shape of the Engel curve using semi-parametric estimation. This is accomplished by using Robinson's (1988) semi-parametric routine[75] to estimate equation (4.3),

$$w_i = X_i'\beta + f(\ln Y_i) + \varepsilon_i.$$

Robinson's estimator ensures consistent estimation of β and estimates the unknown function, f, using kernel density estimation. Blundell *et al* (1998) and Blundell and Powell (2003) show that when the non-parametric component is endogenous, instrumental variables techniques can be employed and the residuals of this first-stage are to be included in the parametric component of equation (4.3). Therefore, the endogeneity of log total expenditure is controlled for by instrumenting with gross normal income and using the residuals from the estimates presented in Appendix Table A4.14. Following the standard notation for this procedure, these residuals, v_i from the first stage are included in the semi-parametric regression as,

$$w_i = X_i'\beta + f(\ln Y_i) + \rho v_i + \varepsilon_i \tag{4.12}$$

and ρ, the coefficient on these residuals, is estimated as part of the parametric component. Table 4.11 presents estimates of the parametric component of equation (4.12), β and ρ. The effects of the control variables on the budget share of lotto using the Robinson's semi-parametric estimator are considerably smaller than all parametric models estimated above.

Though the sex, education level, and some employment statuses of the HRP affect the budget share allocated to lotto in the same direction as previous models, these estimates suggest that having an unemployed or sick HRP increases the budget share for lotto, albeit by a small amount. The significant, though also small, coefficient on the residuals from the first-stage regression indicates that estimating equation (4.3) without instrumenting log total expenditure would yield biased estimates, thus the instrumental variables process and the inclusion of these residuals here is justified. The solid line in Figure 4.7 is the estimate of the non-parametric component, f, of equation (4.12) and the shaded area shows the 95% confidence interval. Due to the large dataset used here, this confidence interval is unsurprisingly small for much of the total expenditure range. The figure is trimmed to exclude the top and bottom 1% of households by total expenditure to allow the graph to be legible – though these observations were included

[75] This routine is identical to that used in Chapter 0 and a more detailed description of the procedure is available in Section 3.10.3.

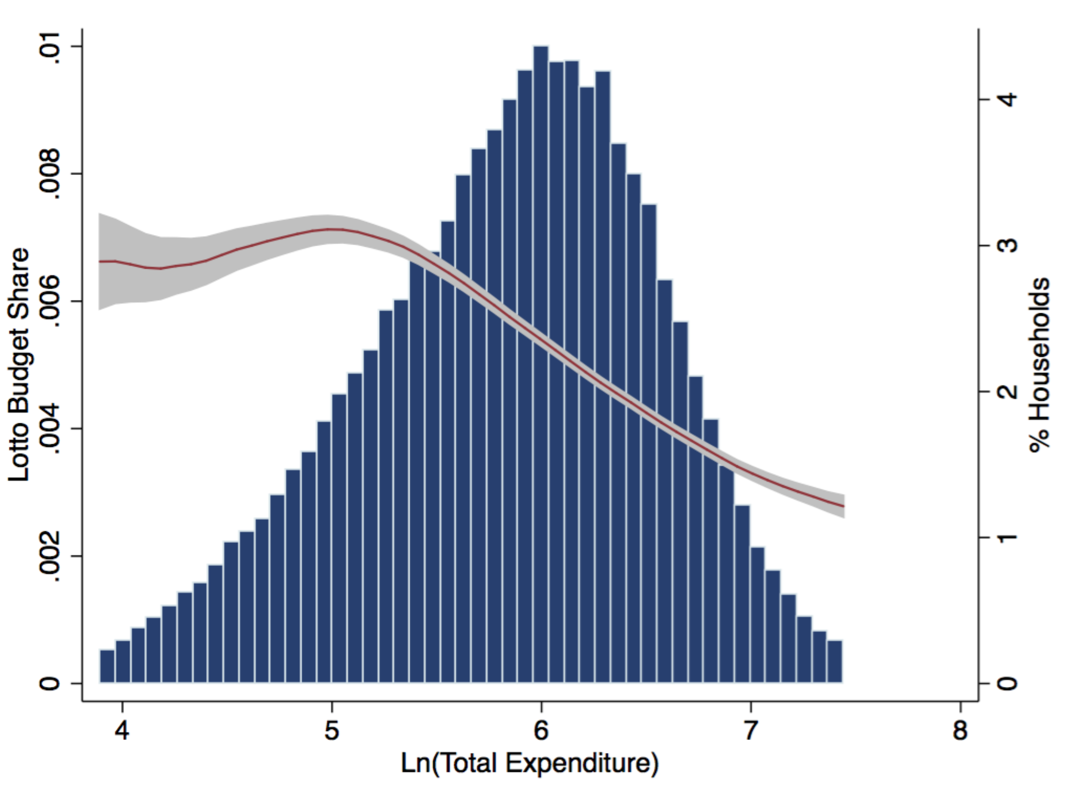
in the estimation process. The estimated budget share of lotto over the distribution is indicated on the left-hand y-axis. Superimposed in Figure 4.7 is a histogram of household weekly total expenditure which corresponds to the right-hand y-axis.

Table 4.11: Parametric component of semi-parametric estimation of the Working-Leser model for lotto demand

	(1)
Dependent Variable	w_i
Female HRP	-0.001***
	(0.0001)
Age HRP	0.000***
	(0.0000)
Self Employed	-0.001***
	(0.0001)
Unemp. (seeking work)	0.001***
	(0.0003)
Unemp. (about to start work)	0.001
	(0.0013)
Sick	0.002***
	(0.0003)
Retired	-0.001***
	(0.0001)
Unoccupied	-0.000
	(0.0001)
Left Education 17-18	-0.003***
	(0.0001)
Left Education 18+	-0.004***
	(0.0001)
Constant	0.006***
	(0.0006)
ρ	-0.041***
	(0.0027)
Month-Year Effects	Yes
Observations	78,867
R-squared	0.044

Robust standard errors in parentheses. */**/*** denotes statistical significance at the 1%/5%/10% level, respectively. Omitted categories: Male HRP, employee, left education before 17, North East. ρ is the coefficient on residuals from instrumenting log total expenditure with gross income as in Appendix Table A4.14.

Figure 4.7: Semi-parametric estimate of lotto budget share against log total expenditure with distribution of households by log total expenditure



Moving from left to right across the total expenditure distribution, the estimated non-parametric component, *f*, rises until a log weekly expenditure of around 5 (which translates to £148 per week). The peak in lotto budget share occurs at a level of total expenditure which is considerably lower than the average as shown by the superimposed distribution of total expenditure. The fitted relationship between lotto budget share and log total expenditure falls rapidly after this peak over the remainder of the distribution – furthering the conclusions drawn from the parametric estimates that the budget share of lotto declines with higher levels of income and the lotto taxes are regressive.

Hardle and Mammen (1993) provide a test of the fitted relationship between the budget share of lotto tickets and log total expenditure against a parametric fit of any degree polynomial of the latter. This test evaluates whether the fitted budget share of lotto tickets from the semi-parametric routine is statistically different from that of a given polynomial expansion, i.e. whether the parametric fit is a sufficiently close approximation of the underlying relationship. It is not surprising upon examination of the fitted function in Figure 4.7 that this test rejects the linear specification of the Working-Leser model, and even specifications which are quadratic and cubic in log total expenditure. Nonetheless, the Working-Leser model which forms the basis of the parametric estimates presented in Section 4.6.2 has a more substantial theoretical underpinning than the linear expenditure model favoured by the existing lotto literature, indicating that the estimates contained in this chapter are still both an improvement and important.

4.7 Conclusion

Despite their notoriously low return rates, lotteries are one of the most popular forms of gambling enjoyed not just in the UK but around the world. Their usefulness as a means of raising public finance when it is otherwise difficult to do so has meant that many games are either operated directly, or have their operation licensed, by the state. However, unlike many other goods subjected to so-called “sin taxes”, there is little-to-no supporting evidence that playing lotto games generates any externality or is in some way otherwise harmful to the individual. Without a supporting paternalistic ‘moral high ground’ argument – and assuming one of the objectives of government is maintain a progressive tax structure – the question of whether, and to what extent, a high rate of tax imposed on such games is regressive becomes an important one.

This chapter has sought to answer this question by estimating a Working-Leser demand function for lottery tickets using household-level data. Since taxes on the UK lotto are constant, the question of whether taxes are regressive, proportional, or progressive can be answered by determining whether the income elasticity of demand for lotto tickets is less than, equal to, or greater than one, respectively. The preferred estimates of this chapter yield headline estimates of the income elasticity for lotto between -1.4 and -0.7. These estimates are significantly lower than those in the existing literature, and well below the critical value of +1 for proportionality in taxes. To the best of the author’s knowledge, this is the only estimate in the literature that indicates lotteries are actually inferior goods. Such low income elasticity estimates therefore suggest that lotto taxes are significantly more regressive than previously thought. This conclusion is supported by a Suits’ index estimate of -0.36, lower than any other study that also calculates an income elasticity and only exceeded by early estimates for lotteries in the US.

Using a Working-Leser demand specification provides an improvement upon the existing literature surrounding the income elasticity of lotto and the regressivity of lotto taxation. Previous studies exclusively use naïve models of demand in which the level of expenditure is some function – typically linear – of income. In the wider economics literature, this model has been criticised for its lack of foundation in microeconomic theory and is, at best, merely an approximation of the true relationship between income and expenditure. Section 4.6.2 briefly replicates this approach and the resultant estimates are, unsurprisingly, very similar to those of the existing literature, estimating income elasticity to be between 0.22 and 0.46. Whilst they would still imply that lotto taxes are regressive, these naïve estimates understate the regressivity of such taxes and contrary to the Working-Leser estimates, class lotto as normal goods, rather than inferior.

Household-level expenditure data presents a number of econometric issues, primarily the presence of zeroes in lotto expenditure which would bias coefficient estimates from OLS. Moreover, the simplest approach to correcting for this bias, using a Tobit model, is undesirable because the independent variables are likely to have different effects on whether a given household participates in lotto and the level of play conditional on participation. The use of Heckman's selection model and Cragg's double-hurdle model overcomes this issue with Tobit by allowing the independent variables to affect the participation and consumption decisions separately, but their two-stage nature present identification issues of their own. This chapter deals with the identification of these models by exploiting differences in consumption behaviour which arises because of religious belief – namely, the identification strategy uses alcohol and pork consumption as an exclusion restriction. It is from these models which the headline estimates of income elasticity are obtained.

Such a large difference between the income elasticity estimates presented here and those in the existing literature highlights the importance of correctly specifying the model used when investigating the relationship between lottery expenditure and income. Moreover, despite its popularity in economics literature and theoretical foundations, the Working-Leser model favoured here still imposes a strict (log-linear) functional form on the relationship between income and the budget share of lotto tickets. Thus, as the final contribution of this chapter, Robinson's semi-parametric estimator has been employed in which the budget share of lotto is fully flexible in its relationship to income – allowing the data to dictate the shape of this function. Thanks to the large dataset used here, this estimator is well determined and shows clearly that the effect of income on lotto consumption varies substantially over the income distribution. Testing against this fully flexible specification reveals that parametric fits of polynomial expansions up to cubic in log income are statistically different. Nonetheless, the Working-Leser model estimates presented here offer a novel finding for the literature and indicate that lotto taxes are even more regressive than previously thought. This is supported with an estimate of Suits' regressivity index of -0.36, far lower than any previous estimate for the UK lotto.

4.8 Appendix

4.8.1 Lottery expenditure by demographic characteristics

Figure A4.8: Mean weekly household lotto expenditure by HRP employment status for the full sample and purchasing households

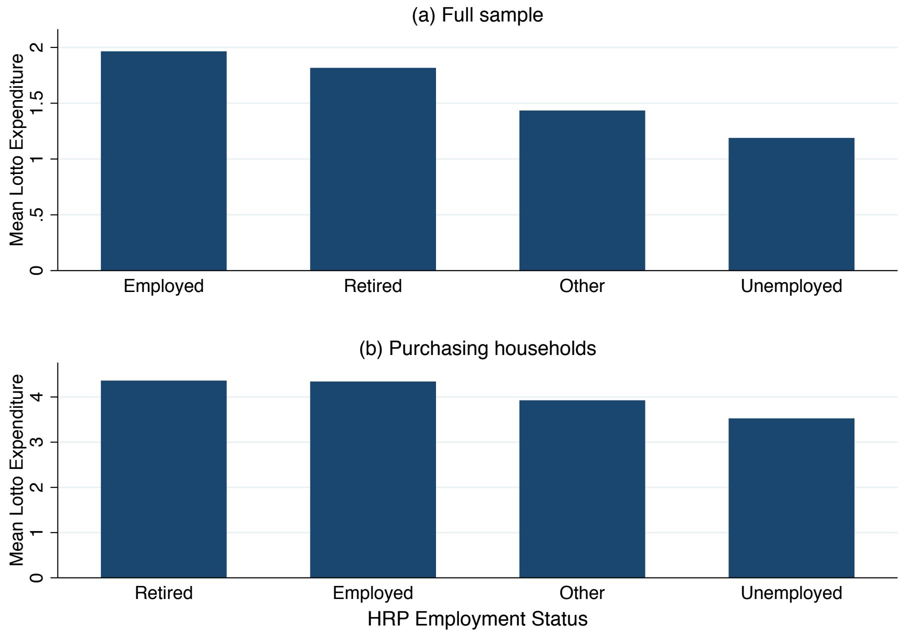


Figure A4.9: Mean weekly household lotto expenditure by region for the whole sample and purchasing households

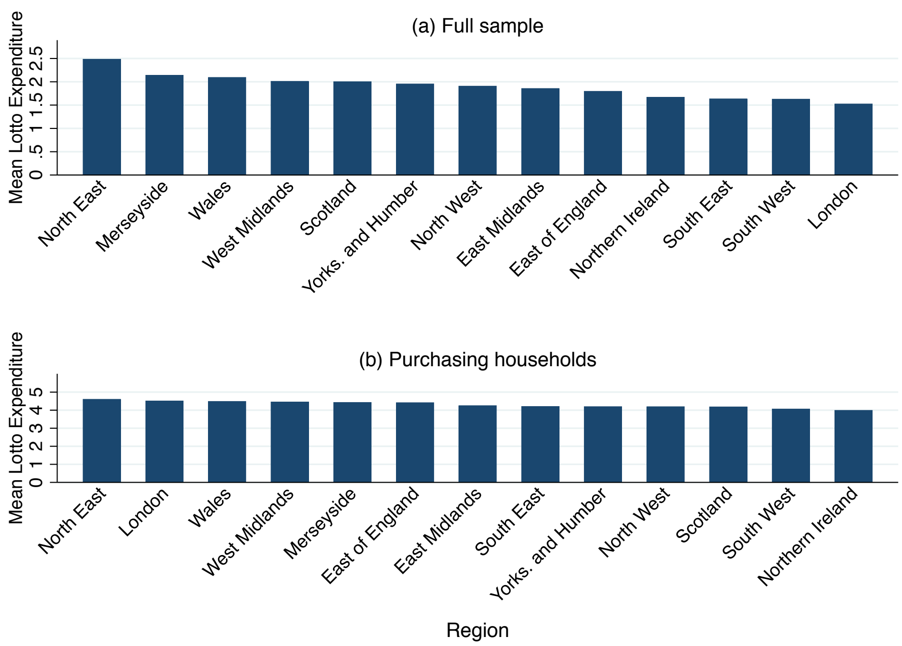
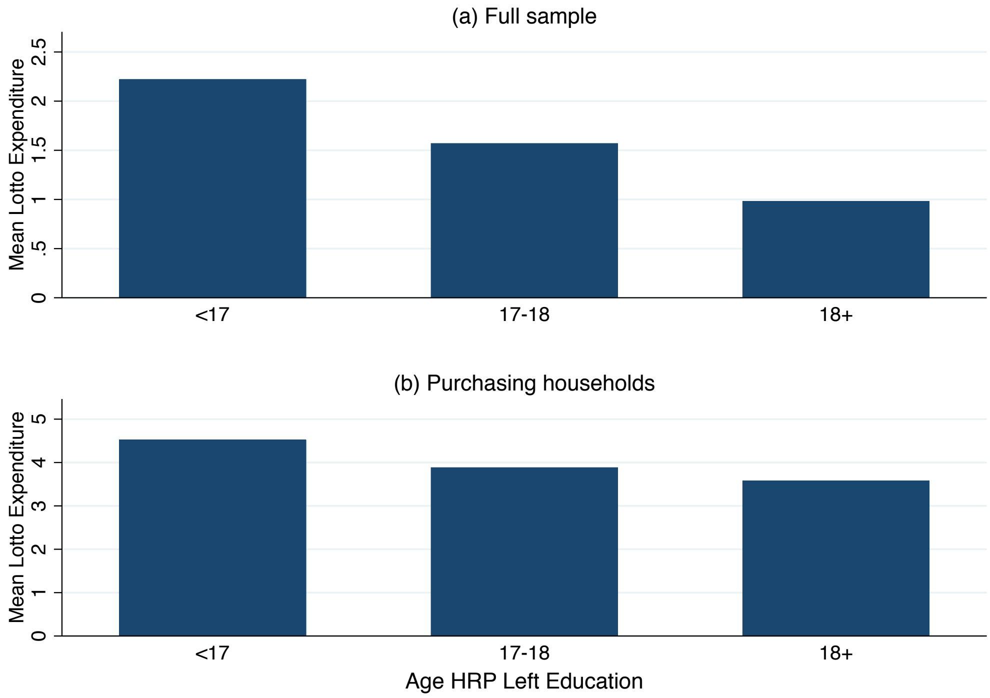


Figure A4.10: Mean weekly household lotto expenditure by age HRP left education for the whole sample and purchasing households



4.8.2 Full coefficient estimates

Table A4.12: Full OLS and Tobit estimates of equation (4.1)

	(1)	(2)
Dependent Variable:	E	E_i
Total Expenditure	0.921***	1.924***
	(0.0633)	(0.1269)
Female HRP	-0.457***	-1.043***
	(0.0272)	(0.0594)
Age HRP	0.030***	0.067***
	(0.0012)	(0.0028)
Self-employed	-0.332***	-1.144***
	(0.0557)	(0.1126)
Unemp. seeking work	-0.588***	-1.831***
	(0.0646)	(0.1845)
Unemp about to work	-0.184	-1.269
	(0.4507)	(1.0668)
Sick	-0.533***	-1.540***
	(0.0610)	(0.1283)
Retired	-1.122***	-2.803***
	(0.0530)	(0.1074)
Unoccupied	-0.544***	-2.027***
	(0.0513)	(0.1398)
Left Education 17-18	-0.676***	-1.579***
	(0.0362)	(0.0791)
Left Education 18+	-1.364***	-3.902***
	(0.0336)	(0.0905)
North West	-0.567***	-1.258***
	(0.0809)	(0.1497)
Merseyside	-0.361***	-0.746***
	(0.1058)	(0.1953)
Yorks. and Humber	-0.549***	-1.230***
	(0.0813)	(0.1502)
East Midlands	-0.689***	-1.651***
	(0.0829)	(0.1557)
West Midlands	-0.525***	-1.316***
	(0.0855)	(0.1539)
East of England	-0.727***	-1.899***
	(0.0833)	(0.1553)
London	-0.794***	-2.296***
	(0.0819)	(0.1589)
South East	-0.926***	-2.365***

	(0.0778)	(0.1472)
South West	-0.915***	-2.237***
	(0.0820)	(0.1565)
Wales	-0.378***	-0.893***
	(0.0913)	(0.1682)
Scotland	-0.446***	-0.957***
	(0.0819)	(0.1511)
Northern Ireland	-0.763***	-1.767***
	(0.0827)	(0.1610)
Constant	1.829***	-0.722**
	(0.1580)	(0.3238)
Income Elasticity	0.222***	0.464***
	(0.0149)	(0.0306)
Month-Year Fixed Effects	Yes	Yes
Observations	78,867	78,867
R-squared	0.053	-

Robust standard errors in parentheses. */**/*** denotes statistical significance at the 1%/5%/10% level, respectively. Omitted categories: Male HRP, employee, left education before 17, North East.

Table A4.13: Full OLS and Tobit estimates of equation (4.2) with ln(total expenditure) uninstrumented and instrumented by gross normal income

	(1)	(2)	(3)	(4)
	OLS	Tobit	IV (2SLS)	IV Tobit
Dependent Variable	w_i	w_i	w_i	w_i
Ln(Total Expenditure) (/1000)	-1.859***	0.158	-2.350***	-3.358***
	(0.0848)	(0.1558)	(0.0955)	(0.6485)
Female HRP	-0.002***	-0.003***	-0.002***	-0.004***
	(0.0001)	(0.0002)	(0.0001)	(0.0003)
Age HRP	0.000***	0.000***	0.000***	0.000***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Self-employed	-0.001***	-0.004***	-0.001***	-0.004***
	(0.0001)	(0.0003)	(0.0001)	(0.0003)
Unemp. seeking work	-0.000	-0.002***	-0.001***	-0.005***
	(0.0003)	(0.0007)	(0.0003)	(0.0009)
Unemp. about to work	-0.000	-0.002	-0.000	-0.004
	(0.0013)	(0.0034)	(0.0013)	(0.0035)
Sick	0.000	-0.001**	-0.000	-0.004***
	(0.0003)	(0.0005)	(0.0003)	(0.0007)
Retired	-0.002***	-0.006***	-0.002***	-0.008***
	(0.0002)	(0.0003)	(0.0002)	(0.0005)
Unoccupied	-0.001***	-0.005***	-0.001***	-0.006***
	(0.0002)	(0.0005)	(0.0002)	(0.0006)
Left Education 17-18	-0.002***	-0.005***	-0.002***	-0.005***
	(0.0001)	(0.0002)	(0.0001)	(0.0003)
Left Education 18+	-0.003***	-0.012***	-0.003***	-0.011***
	(0.0001)	(0.0003)	(0.0001)	(0.0003)
North West	-0.003***	-0.005***	-0.003***	-0.005***
	(0.0003)	(0.0005)	(0.0003)	(0.0005)
Merseyside	-0.001***	-0.002***	-0.001***	-0.002***
	(0.0004)	(0.0007)	(0.0004)	(0.0007)
Yorks. and Humber	-0.002***	-0.004***	-0.002***	-0.004***
	(0.0003)	(0.0006)	(0.0003)	(0.0006)
East Midlands	-0.003***	-0.006***	-0.003***	-0.006***
	(0.0003)	(0.0006)	(0.0003)	(0.0006)
West Midlands	-0.002***	-0.005***	-0.002***	-0.005***
	(0.0003)	(0.0006)	(0.0003)	(0.0006)
East of England	-0.003***	-0.008***	-0.003***	-0.007***
	(0.0003)	(0.0006)	(0.0003)	(0.0006)
London	-0.003***	-0.008***	-0.003***	-0.008***
	(0.0003)	(0.0006)	(0.0003)	(0.0006)
South East	-0.004***	-0.009***	-0.004***	-0.008***

	(0.0003)	(0.0005)	(0.0003)	(0.0005)
South West	-0.004***	-0.009***	-0.004***	-0.008***
	(0.0003)	(0.0006)	(0.0003)	(0.0006)
Wales	-0.002***	-0.004***	-0.002***	-0.004***
	(0.0003)	(0.0006)	(0.0003)	(0.0006)
Scotland	-0.002***	-0.004***	-0.002***	-0.004***
	(0.0003)	(0.0006)	(0.0003)	(0.0006)
Northern Ireland	-0.004***	-0.007***	-0.004***	-0.007***
	(0.0003)	(0.0006)	(0.0003)	(0.0006)
Constant	0.018***	-0.002	0.020***	0.019***
	(0.0008)	(0.0014)	(0.0008)	(0.0040)
Income Elasticity	0.675***	1.028***	0.589***	0.412***
	(0.0149)	(0.0273)	(0.0167)	(0.1135)
Month-Year Effects	Yes	Yes	Yes	Yes
1st Stage F-Statistic	-	-	65.264***	65.264***
Observations	78,867	78,867	78,867	78,867
R-squared	0.073	-	0.072	-

Robust standard errors in parentheses. */**/*** denotes statistical significance at the 1%/5%/10% level, respectively. Log total expenditure is divided by 1000 to allow coefficients to be legible. Omitted categories: Male HRP, employee, left education before 17, North East.

Table A4.14: First-stage IV estimates

	(1)
Dependent Variable	ln(Y)/1000
Gross Normal Income	0.609***
	(0.0734)
Self Employed	0.090***
	(0.0087)
Unemp. (seeking work)	-0.586***
	(0.0436)
Unemp. (about to start work)	-0.191**
	(0.0908)
Sick	-0.557***
	(0.0363)
Retired	-0.418***
	(0.0283)
Unoccupied	-0.367***
	(0.0369)
Left Education 17-18	0.123***
	(0.0086)
Left Education 18+	0.152***
	(0.0209)
North West	0.049***
	(0.0115)
Merseyside	0.002
	(0.0146)
Yorks. and Humber	0.037***
	(0.0113)
East Midlands	0.058***
	(0.0118)
West Midlands	0.048***
	(0.0117)
East of England	0.097***
	(0.0128)
London	0.046***
	(0.0136)
South East	0.125***
	(0.0139)
South West	0.094***
	(0.0116)
Wales	0.018
	(0.0125)
Scotland	0.035***

	(0.0116)
Northern Ireland	0.106***
	(0.0124)
Constant	5.418***
	(0.0495)
Month-Year Fixed Effects	Yes
F-Statistic (1, 78693)	68.67***
Observations	78,867
R-squared	0.470

Robust standard errors in parentheses. */**/*** denotes statistical significance at the 1%/5%/10% level, respectively. Omitted categories: Male HRP, employee, left education before 17, North East.

Table A4.15: Full Heckman and double hurdle estimates of the Working-Leser model

	(1) Selection Equation	(2) Heckman	(3) Double Hurdle
No Alcohol	-0.298***		
	(0.0143)		
No Pork	-0.198***		
	(0.0115)		
No Alcohol * No Pork	-0.077***		
	(0.0204)		
Ln(Total Expenditure) /1000	-24.832	-9.640***	-32.165***
	(18.1572)	(0.4238)	(8.6294)
Female HRP	-0.123***	-0.003***	-0.062***
	(0.0110)	(0.0002)	(0.0052)
Age HRP	0.008***	0.000***	0.004***
	(0.0005)	(0.0000)	(0.0002)
Self Employed	-0.236***	-0.002***	-0.114***
	(0.0179)	(0.0004)	(0.0085)
Unemp. (Seeking work)	-0.313***	-0.004***	-0.156***
	(0.0362)	(0.0008)	(0.0166)
Unemp. (About to work)	-0.287*	0.000	-0.137*
	(0.1607)	(0.0035)	(0.0750)
Sick	-0.274***	-0.003***	-0.137***
	(0.0258)	(0.0005)	(0.0120)
Retired	-0.451***	-0.005***	-0.217***
	(0.0192)	(0.0004)	(0.0087)
Unoccupied	-0.375***	-0.004***	-0.181***
	(0.0249)	(0.0006)	(0.0113)
Left Education 17-18	-0.241***	-0.003***	-0.117***
	(0.0136)	(0.0003)	(0.0064)
Left Education 18+	-0.601***	-0.006***	-0.275***
	(0.0142)	(0.0004)	(0.0060)
North West	-0.210***	-0.004***	-0.104***
	(0.0270)	(0.0005)	(0.0130)
Merseyside	-0.114***	-0.002***	-0.056***
	(0.0349)	(0.0006)	(0.0169)
Yorks. And Humber	-0.202***	-0.003***	-0.099***
	(0.0273)	(0.0005)	(0.0132)
East Midlands	-0.281***	-0.004***	-0.137***
	(0.0279)	(0.0005)	(0.0134)
West Midlands	-0.243***	-0.003***	-0.119***
	(0.0272)	(0.0005)	(0.0131)

East of England	-0.321***	-0.004***	-0.156***
	(0.0272)	(0.0005)	(0.0130)
London	-0.376***	-0.004***	-0.181***
	(0.0273)	(0.0006)	(0.0130)
South East	-0.388***	-0.005***	-0.187***
	(0.0261)	(0.0005)	(0.0125)
South West	-0.381***	-0.006***	-0.185***
	(0.0275)	(0.0005)	(0.0131)
Wales	-0.177***	-0.003***	-0.087***
	(0.0298)	(0.0005)	(0.0144)
Scotland	-0.147***	-0.003***	-0.072***
	(0.0277)	(0.0005)	(0.0134)
Northern Ireland	-0.299***	-0.005***	-0.146***
	(0.0287)	(0.0005)	(0.0137)
Constant	0.625***	-	-
	(0.1258)		
Income Elasticity	-	-0.687***	-1.424**
		(0.0742)	(0.6503)
lambda	-	0.011***	-
		(0.0007)	
Month-Year Fixed Effects	Yes	Yes	Yes
Observations	78,867	78,867	78,867

Robust standard errors in parentheses. */**/*** denotes statistical significance at the 1%/5%/10% level, respectively. Demographic controls omitted from reporting for brevity. In all columns ln(total expenditure) is divided by 1000 to allow coefficients to be legible. Income elasticity is calculated as $\eta = 1 + \gamma/(1000 * 0.00571)$ where γ is the coefficient on log total expenditure and 0.00571 is the mean expenditure share of lottery tickets. All columns have log total expenditure instrumented by gross normal income. The first stage is available in Appendix Table A4.14. Omitted categories: Male HRP, employee, left education before 17, North East.

4.8.3 Working-Leser model estimates with no demographic controls

Table A4.16: First-stage with no demographic controls

Dependent Variable	(1) ln(Y)
Gross Normal Income	0.001***
	(0.0001)
Constant	0.005***
	(0.0000)
Month-Year Fixed Effects	Yes
F-Statistic (1,78693)	112.96***
Observations	79,057
R-squared	0.376

Robust standard errors in parentheses. */**/*** denotes statistical significance at the 1%/5%/10% level, respectively.

Table A4.17: OLS and Tobit estimates of equation (4.2) with ln(total expenditure) uninstrumented and instrumented by gross normal income

Dependent Variable	(1) OLS w_i	(2) Tobit w_i	(3) IV w_i	(4) IV Tobit w_i
Ln(Total expenditure)/1000	-2.751***	-2.075***	-3.264***	-5.285***
	(0.0764)	(0.1365)	(0.0705)	(0.5158)
Constant	0.023***	0.010***	0.026***	0.028***
	(0.0007)	(0.0012)	(0.0007)	(0.0031)
Income Elasticity	0.519***	0.637***	0.429***	0.075***
	(0.0134)	(0.0239)	(0.0123)	(0.0903)
Month-Year Fixed Effects	Yes	Yes	Yes	Yes
1st-Stage F-Statistic	-	-	112.96***	112.96***
Observations	79,057	79,057	79,057	79,057
R-squared	0.036	-	0.035	-

Robust standard errors in parentheses. */**/*** denotes statistical significance at the 1%/5%/10% level, respectively. Income elasticity is calculated as $\eta = 1 + \gamma/(1000 * 0.00571)$ where γ is the coefficient on log total expenditure and 0.00571 is the mean expenditure share of lottery tickets. Log total expenditure is divided by 1000 to allow coefficients to be legible. Columns 3 and 4 have log total expenditure instrumented by gross normal income. The first stage is available in Appendix Table A4.16.

Table A4.18: Heckman selection and double hurdle estimates of the Working-Leser model without demographic controls

	(2) Selection Equation	(3) Heckman	(8) Double Hurdle
No Alcohol	-0.306***		
	(0.0140)		
No Pork	-0.269***		
	(0.0112)		
No Alcohol * No Pork	-0.080***		
	(0.0200)		
Ln(Total Expenditure)/1000	-100.826***	-12.049***	-73.280***
	(11.5683)	(0.2468)	(5.6302)
Constant	0.848***	-	-
	(0.0851)	-	-
λ	-	0.009***	-
	-	(0.0006)	-
Income Elasticity	-	-1.109***	-4.522***
	-	(0.0432)	(0.4242)
Month-Year Fixed Effects	Yes	Yes	Yes
Observations	79,057	79,057	78,988

Robust standard errors in parentheses. */**/*** denotes statistical significance at the 1%/5%/10% level, respectively. Demographic controls omitted from reporting for brevity. In all columns ln(total expenditure) is divided by 1000 to allow coefficients to be legible. Income elasticity is calculated as $\eta = 1 + \gamma/(1000 * 0.00571)$ where γ is the coefficient on log total expenditure and 0.00571 is the mean expenditure share of lottery tickets. All columns have log total expenditure instrumented by gross normal income. The first stage is available in Appendix Table A4.16.

Table A4.19: Heckman and double-hurdle estimates of the Working-Leser model with controls for household size

	(1) Selection Equation	(2) Heckman A	(3) Double Hurdle Margins
No Alcohol	-0.261***	-	-
	(0.0145)		
No Pork	-0.148***	-	-
	(0.0117)		
No Alcohol * No Pork	-0.101***	-	-
	(0.0205)		
Ln(Tota Expenditure)/1000	-333.525***	- 12.578***	-7.712***
	(28.9541)	(0.6217)	(0.2252)
Household Size	0.345***	0.004***	0.003***
	(0.0133)	(0.0003)	(0.0001)
Female HRP	-0.073***	-0.002***	-0.001***
	(0.0107)	(0.0002)	(0.0001)
Age HRP	0.006***	0.000***	0.000***
	(0.0005)	(0.0000)	(0.0000)
Self Employed	-0.228***	-0.002***	-0.001***
	(0.0180)	(0.0004)	(0.0002)
Unemp. (Seeking work)	-0.469***	-0.005***	-0.004***
	(0.0391)	(0.0008)	(0.0002)
Unemp (About to Work)	-0.342**	-0.000	-0.002
	(0.1612)	(0.0035)	(0.0013)
Sick	-0.428***	-0.004***	-0.004***
	(0.0291)	(0.0006)	(0.0002)
Retired	-0.487***	-0.005***	-0.004***
	(0.0200)	(0.0005)	(0.0001)
Unoccupied	-0.493***	-0.005***	-0.004***
	(0.0268)	(0.0006)	(0.0002)
Left Education 17-18	-0.178***	-0.002***	-0.001***
	(0.0144)	(0.0003)	(0.0001)
Left Education 18+	-0.508***	-0.005***	-0.003***
	(0.0159)	(0.0005)	(0.0001)
North West	-0.185***	-0.004***	-0.001***
	(0.0272)	(0.0005)	(0.0002)
Merseyside	-0.107***	-0.002***	-0.001***
	(0.0350)	(0.0006)	(0.0003)
Yorks. and Humber	-0.181***	-0.003***	-0.001***
	(0.0274)	(0.0005)	(0.0002)
East Midlands	-0.259***	-0.004***	-0.002***
	(0.0281)	(0.0005)	(0.0002)
West Midlands	-0.228***	-0.003***	-0.001***
	(0.0274)	(0.0005)	(0.0002)
East of England	-0.276***	-0.004***	-0.002***
	(0.0276)	(0.0005)	(0.0002)

London	-0.345***	-0.004***	-0.002***
	(0.0276)	(0.0006)	(0.0002)
South East	-0.325***	-0.004***	-0.002***
	(0.0266)	(0.0005)	(0.0002)
South West	-0.342***	-0.005***	-0.002***
	(0.0278)	(0.0005)	(0.0002)
Wales	-0.168***	-0.003***	-0.001***
	(0.0300)	(0.0006)	(0.0002)
Scotland	-0.119***	-0.003***	-0.001***
	(0.0279)	(0.0005)	(0.0002)
Northern Ireland	-0.296***	-0.005***	-0.002***
	(0.0289)	(0.0006)	(0.0002)
Income Elasticity	-	-1.202***	-0.350***
		(0.1088)	(0.0394)
Lambda	-	0.011***	-
		(0.0008)	
Constant	1.771***	-	-
	(0.1646)		
Observations	78,867	78,867	78,867

Robust standard errors in parentheses. */**/*** denotes statistical significance at the 1%/5%/10% level, respectively. Demographic controls omitted from reporting for brevity. In all columns ln(total expenditure) is divided by 1000 to allow coefficients to be legible. Income elasticity is calculated as $\eta = 1 + \gamma/(1000 * 0.00571)$ where γ is the coefficient on log total expenditure and 0.00571 is the mean expenditure share of lottery tickets. All columns have log total expenditure instrumented by gross normal income. Omitted categories: Male HRP, employee, left education before 17, North East.

5 Conclusion

The purpose of this thesis was to address three topics surrounding the economics of gambling. Beyond the unifying theme of gambling, the three main chapters are also linked by in their application of modern statistical techniques to the best available data in order to approach their respective research questions. In doing so, all three chapters present results which are both novel and important to the literature and which are not only of interest to academics, but to policy makers and the industry as a whole. Chapter 2 addresses the extent to which individuals who are defined as problem gamblers are harmed by their affliction. Chapters 3 and 4 investigate issues surrounding the UK's most popular gambling product – lotto. In Chapter 3 the game is found to be poorly designed when considering its primary objective of maximising revenues for good causes and tax – and even more so since changes to its design in 2015. Chapter 4 uses modern econometric techniques to evaluate the extent to which the incidence of high rates of tax levied on lotto are borne disproportionately by the poor. The results suggest that lotto is even more regressive than previously thought and that lotto is not a normal good as evidence in the previous literature suggests; instead, lottery tickets should be classified as inferior.

5.1 Summary of Chapter 2

Chapter 2 attempts to quantify the extent to which problem gambling causes harm to afflicted individuals. To answer this question, an increasingly popular well-being methodology is employed in which well-being is estimated as a function of problem gambling and income, plus other controls. This methodology enables, by dividing the resultant coefficient on the former by the coefficient on the latter, a money metric of the loss in well-being associated with being a problem gambler. Baseline estimates place this value – also known as the compensating variation – at £90,000 per annum for the average problem gambler. Whilst the prevalence of problem gambling is small, less than 1% of the population, there are an estimated 300,000 problem gamblers in the UK, implying the aggregate well-being loss is in excess of £31 billion per year.

Evaluating the effect of problem gambling on well-being is a useful methodology since it can, in essence, be viewed as a catch-all measure of the welfare effects of the condition. This gives the methodology, at least in principle, superiority over studies which evaluate the cost of problem gambling by aggregating the associated externalities such as crime and labour market effects. Analysis of externalities necessarily avoids the cost of problem gambling to the afflicted individual, which the analysis here finds to be significant, and is susceptible to missing some costs if all externalities are not known or accounted for. However, the well-being methodology is still in its relative infancy and criticisms of the approach are addressed directly in Chapter 2. Treating self-reported well-being (on a 1-10 scale) as cardinal, as is the case with OLS, is unlikely to be a valid assumption. Bond and Lang (2010) argue that these well-being scores are categorical in nature and the reported values correspond to some interval along a continuous distribution. They argue that monotonic transformations of the underlying utility function can reverse the ranking of these intervals and instead the measure can only be safely treated as ordinal. Several transformations in well-being are considered but have no significant effect on the headline estimate of compensating variation. Using an ordered probit model, which treats the reported well-being scores as ordinal rather than cardinal per the recommendation in Bond and Lang (2010), yields higher estimates of the cost of problem gambling (£50 billion per annum) than OLS, suggesting that the effect measured using OLS may be a lower bound on the cost of well-being.

The money metric derived using the well-being methodology is susceptible to measurement error in both income and problem gambling. The presence of measurement error in either of these covariates will bias their respective coefficients towards zero, increasing (decreasing) the estimated compensating variation when it is present in the former (latter). Income in the dataset used is recorded in intervals, which are undesirable for the methodology, so correction for measurement error and to obtain a continuous income variable is done using interval regression, the results of which are very conventional. Fortunately, the data contains two measures of problem gambling and measurement error in one screen is corrected by using the score in the other as an instrumental variable. Again, this yields even higher estimates of the cost of problem gambling than the baseline OLS, precisely because of measurement error attenuating the coefficient on problem gambling towards zero.

OLS estimates of the compensating variation, whilst novel to the literature and important in their own right, do not provide a causal interpretation on the cost of problem gambling. Attempting to find a causal estimate of this effect is done using parental gambling behaviour as an instrumental variable. Finding an instrument is particularly difficult for problem gambling since the proportion of problem gamblers is so small it will almost certainly lead to a weak instrument problem and is evident in the estimates in Chapter 2. Nonetheless, the importance of finding a causal estimate warrants its inclusion and indicates again much higher costs than the baseline estimates.

Chapter 2 concludes by investigating the mediating role of gambling expenditure on reduction in well-being associated with being a problem gambler. Uncovering such effects is important since it would provide some evidence to policy makers on the effectiveness of taxation in mitigating the harm of problem gambling. Mediation analysis here is, however, severely hampered by the presence of measurement error in these expenditures which is self-reported and again recorded in intervals in the data. Moreover, there is no obvious candidate variable in the data with which these expenditures could be instrumented to correct for this measurement error. Thus, for most gambling types, the expenditure has no effect on well-being and it is unclear whether this is due to measurement error or because there is simply no effect. Nonetheless, expenditures on two of the most popular gambling products – lotto and scratchcards – are measured fairly well when compared to aggregate-level data and are considered as potential mediators. Spending on scratchcards are found to have a statistically significant mediating role on the impact of being a problem gambler, whereas lotto expenditure is not.

It would, however, be unwise to conclude from these estimates that tax policy targeted at scratchcards in the hope it would generate substitution effects towards the benign lotto game would be effective in mitigating the harm associated with problem gambling. Future work would need to incorporate own- and cross-price elasticities of all gambling products to evaluate whether taxation of the different goods would generate substitution effects towards benign products. Moreover, future work could improve upon the estimates here with better expenditure data being recorded alongside well-being, problem gambling and income to investigate more accurately and across all gambling products whether expenditure has any mediating effects. Moreover, policies which target behavioural origins of problem gambling may be more effective in reducing the associated costs and future research on this would also be beneficial.

5.2 Summary of Chapter 3

Chapter 3 addresses the role which the design of the UK's most popular gambling product – lotto – influences demand for the game. Lotto is a particularly important product to analyse not only because of its popularity amongst players but also because of its popularity with governments and charitable organisations as a means of raising tax dollars and financing public good provision. In this chapter, the demand for lotto is first modelled under the assumption that sales are driven by the effect of rollovers on the mean of the prize distribution – i.e. via the expected value of a lotto ticket. This approach is popular in the literature as it facilitates the estimation of a price elasticity of demand and tests of this against -1 are often used to determine whether the current design of the game is maximising the revenues used towards public good provision.

Much of the existing literature identifies such models of lotto demand by instrumenting price with either the occurrence or size of rollovers or both. However, the pari-mutuel design of lotto means that rollover size is simply a pre-determined proportion of sales in the previous draw, and if sales are autocorrelated then rollover size is not exogenous to sales and is an invalid instrument. Rollover occurrence is also an invalid instrument since the likelihood of a rollover occurring is dependent on the number of unique tickets sold in the previous draw. A primary contribution of Chapter 3 is then to provide a novel identification strategy for the price model of lotto demand. Rollover occurrence is instrumented using the numbers which form part of the winning configuration and the rationale is based on exploiting systematic non-random number selection by lotto players – termed "conscious selection" by the industry. A peculiar feature of the UK lotto is the award of a fixed prize for matching 3 of the 6 numbers which make the winning configuration and the fact that these prizes are paid before the remaining prize money is allocated to the pari-mutuel prize pools. As such, unexpected variation in the number of winners of the 3-ball prize exogenously affects the size of the pari-mutuel prizes available, including the jackpot prize which becomes the rollover size if there are no winners. To instrument rollover size, then, this random variation in fixed prize winners is used.

The resultant estimates of price elasticity of demand for lotto are -0.6 for the more popular Saturday draws and -1.5 for Wednesday draws, indicating inelastic demand for the former and elastic demand for the latter. These estimates suggest that the game is not revenue maximising and could improve its tax and good causes funding by increasing the price of Saturday draws relative to Wednesday draws – a new finding for the literature.

A common criticism of models of lotto demand in which price is the key dependent variable is that it is based on expected utility theory which is notoriously ineffective at explaining why gambling occurs amongst otherwise risk-averse individuals. Moreover, the definition of the price of lotto tickets is little more than an arbitrary functional form of rollovers. Thus, Chapter 3 further extends the research base by estimating models in which rollovers directly affect demand. The same instrumental variables strategy is employed to overcome the endogeneity of rollover size, and estimates suggest £1 million increases in the jackpot prize from rollovers increases Saturday sales by around £1.8 million and Wednesday sales by £0.5 million. Non-nested testing of this reduced form model of rollover size against the price model proved inconclusive suggesting future research could improve upon the estimates presented here by assuming some other functional form of the prize distribution. Semi-parametric estimation, in which rollover size is fully flexible in its role in determining lotto ticket sales is not statistically different from quadratic or cubic parameterisations for Wednesday draws, but is statistically different for Saturdays for the same specifications.

Finally, Chapter 3 examines major re-designs to the UK's main lotto game in 2013 and 2015, which were likely implemented in response to declining sales figures. Simple comparison of pre- and post-redesign revenues shows that the 2013 changes were a relative success, increasing sales from an average of £44.5 million per week to £49.8 million – equivalent to over £275 million per year in extra revenues. The 2015 renovation of the game proved less successful, seeing sales return to their pre-2013 levels. Thorough analysis of the effectiveness would require knowledge of the level of sales had the game re-designs never occurred. Attempts to forecast sales out of sample from the models developed in Chapter 3 mirror the declining trend in sales pre-2013, indicating the game changes were more successful than simply comparing sales figures suggests. However, these forecasts could be improved with a more complete simulation of lotto design and the effects of changing the game parameters in future research.

5.3 Summary of Chapter 4

Lotteries' usefulness as a means of raising public finance means many of the games around the world are either operated directly by, or under strict license from, the state and often with large take-out rates. In addition to deadweight losses generated by such high taxes there is an additional strand to the literature surrounding the extent to which the incidence of these taxes is disproportionately borne by the poor. Since taxes imposed on lotteries are constant, analysis of whether they are regressive, progressive or proportional can be achieved by determining whether the income elasticity for the game is less than, greater than or equal to 1, respectively. Chapter 4 contributes to the literature surrounding the regressivity of lotto taxation by estimating a Working-Leser model of household demand for lottery tickets using a large household-level survey – the Expenditure and Food Survey (EFS).

Much of the existing literature approaches the question of regressivity of lotto taxation by estimating a linear expenditure model in which income is linear in the level of tickets purchased. Replicating this approach using the aforementioned data yields very similar estimates to those in the existing literature of between 0.22 and 0.46. These estimates would indicate that lotteries are classed as a necessity good and, since they are below the critical value of 1 for proportionality, indicate that taxes on lotto are regressive. The linear expenditure model is, however, often criticised in the wider economics literature for its lack of microeconomic foundation and poor fit statistics when taken to such micro-level data.

Instead, Chapter 4 contributes to the existing literature by estimating a Working-Leser model of household demand for lottery tickets to obtain income elasticity estimates. Rather than modelling the level of tickets demanded as a function of income, the budget share of lottery tickets is instead estimated as a function of log income. The preferred estimates of this model suggest that the income elasticity for lotto tickets is between -1.4 and -0.7, which is significantly lower than previous estimates and well below the critical value of 1 for proportional taxation. Furthermore, these estimates suggest taxes imposed on lotteries are even more regressive than previously thought and that lotteries are not a necessity good as is the conventional wisdom, rather they should be classed as inferior.

The large difference in the income elasticity estimates from the two models considered in Chapter 4 highlights the importance of correctly specifying the model used when model the relationship between lottery expenditure and income. The preferred Working-Leser model still imposes a strict (log-linear) functional form on the relationship between income and the budget share of lotto tickets. Thus, as an additional contribution of the chapter, a semi-parametric model is pursued to allow the data to dictate the shape of this function. Due to the large dataset used here, this estimator is well determined and shows that the budget share of lotto tickets is positively related to income at the very lower end of the distribution, before peaking well below median income and then falling rapidly over the rest of the distribution. Tests of fully flexible specification against the parametric estimates reveals that polynomial expansions up to cubic in log income are statistically different. Nonetheless, the Working-Leser model estimates presented here offer a novel finding for the literature and indicate that lotto taxes are even more regressive than previously thought.

Using household-level expenditure data presents a number of econometric issues. Primarily, the presence of zeroes in lotto expenditure leads to biased coefficients when estimating via OLS. The simplest approach to correcting for this bias, using a Tobit model, is undesirable because it imposes the strict assumption that the independent variables have the same effect on the household's participation and consumption decisions. The analysis in Chapter 4 uses Heckman's selection model and Cragg's double-hurdle routine to overcome this issue, both of which allow the independent variables to affect the participation and consumption decisions separately. However, the two-stage nature of these routines present identification issues of their own. In order to overcome this identification issue, Chapter 4 exploits exogenous differences in consumption behaviour which arises because of religious belief as an exclusion restriction.

The final contribution of Chapter 4 is to estimate Suits' index of tax regressivity – a Gini-style metric which allows comparison of the regressivity of different taxes. The key determinant for this index is whether the computed index is less than, greater than, or equal to 0 which indicates regressive, progressive and proportional taxes, respectively. This index is calculated at -0.36 using the present data – considerably smaller than much of the existing literature – and supports the findings from income elasticity estimates that lotto taxes are much more regressive than previously believed.

6 References

American Gaming Association. (2016). *State of the States.* American Gaming Association.

Angrist, J.D. and J-S. Pischke (2009), Mostly Harmless Econometrics: An Empiricist's Companion, *Princeton University Press.*

Altonji, J.G., T.E. Elder and C.R. Taber (2005), "Selection on Observed and Unobserved Variables:", *Journal of Political Economy,* 113, 51-184.

Australian Productivity Commission (2010), Gambling, *Australian Productivity Commission Inquiry Report 50*, Commonwealth of Australia.

Banks, J., Blundell, R., and Lewbel, A. (1997). "Quadratic Engel curves and consumer demand", *Review of Economics and Statistics*, 79(4), 527-539.

Becker, G.S. (1964), Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education*,* Columbia University Press for NBER, New York.

Becker, G.S. and K. M. Murphy (1988), "A Theory of Rational Addiction", *Journal of Political Economy*, 96, 675-700.

Becker, G.S., M. Grossman, and K. M. Murhpy, (2006), "The Market for Illegal Goods: The Case of Drugs", *Journal of Political Economy,* 114, 38-60.

Black, D.A., M.C. Berger, and F.A. Scott, 2000, "Bounding Parameter Estimates with Nonclassical Measurement Error", *Journal of the American Statistical Association*, 95, 739-748.

Blundell, R., & Powell, J. L. (2003). Endogeneity in nonparametric and semiparametric regression models. In M. Dewatripont, L. P. Hansen, & S. J. Turnovsky, *Advances in Economics and Econometrics* (Vol. 2, pp. 312-357). Cambridge: Cambridge University Press.

Blundell, R., Duncan, A., & Pendakur, K. (1998). Semiparametric estimation and consumer demand. *Journal of Applied Econometrics, 13*, 435-461.

Bond, T.A. and K. Lang (2010), "The sad truth about happiness scales", *NBER WP19950.*

Borg, M., and Mason, P. (1988). “The budgetary incidence of a lottery to support education”, *National Tax Journal*, 41(1), 75-85.

Boland, P. J., and Pawitan, Y. (1999). “Trying to be random in selecting numbers for lotto”, *Journal of Statistics Education*, 7(3).

Brannlund, R., and Nordstrom, J. (2004). “Carbon tax simulations using a household demand model”, *European Economic Review*, 48(1), 211-233.

Brinner, R. E., and Clotfelter, C. T. (1986). “An economic appraisal of state lotteries”, *National Tax Journal*, 28(4), 395-404.

Cain, M., Peel, D., & Law, D. (2002). Skewness as an explanation of gambling by locally risk averse agents. *Applied economics letters, 9*(15), 1025-1028.

Calado, F., & Griffiths, M. D. (2016). Problem gambling worldwide: An update and systematic review of empirical research (2000-2015). *Journal of Behavioural Addictions, 5*(4), 592-613.

Canadian Gaming Association. (2011). *Economic Impact of the Canadian Gaming Industry.* Toronto: Canadian Gaming Association.

Chesher, A. and K. Smolinski (2012), “IV models of ordered choice”, *Journal of Econometrics*, 166, 33-48.

Chetty, R. (2015), “Behavioral Economics and Public Policy: A Pragmatic Perspective”, *American Economic Review: Papers and Proceedings*, 105, 1–33.

Clark, A. E., Frijters, P., and Shields, M. (2007). “Relative income, happiness and utility: An explanation for the Easterlin Paradox and other puzzles”, *IZA Discussion Paper 2840*.

Clark, A. and A. Oswald (2002), “A simple statistical method for measuring how life events affect happiness”, *International Journal of Epidemiology*, 31, 1139-1144.

Clotfelter, C. T. (1979). “On the regressivity of state-operated ‘numbers’ games”, *National Tax Journal*, 32(4), 543-548.

Clotfelter, C. T., and Cook, P. J. (1987). “Implicit taxation in lottery finance”, *National Tax Journal*, 40(4), 533-546.

Clotfelter, C. T., & Cook, P. J. (1990). On the economics of state lotteries. *Journal of Economic Perspectives, 4*(4), 105-119.

Clotfelter, C. T., and Cook, P. J. (1991). "Selling hope: State lotteries in America", *Harvard University Press, Cambridge, MA.*

Conley, T. G., Hansen, C. B., & Rossi, P. E. (2012). Plausibly exogenous. *Review of Economics and Statistics, 94*(1), 260-272.

Cook, P. J., & Clotfelter, C. T. (1993). The peculiar scale economies of lotto. *The American Economic Review, 83*(3), 634-643.

Cox, D. R. (1961). Tests of separate families of hypotheses. *Proceedings of the fourth Berkeley symposium on mathematical statistics and probability*, *1*, pp. 105-123.

Cox, D. R. (1962). Further results on tests of separate families of hypotheses. *Journal of the Royal Statistical Society. Series B (Methodological), 24*(2), 406-424.

Cragg, J. (1971). "Some statistical models for limited dependent variables with an application to the demand for durable goods", *Econometrica*, 39(5), 829-844.

De Agostini, P. (2014). "The effect of food prices and household income on the British diet", *ISER Working Paper Series No. 2014/10.*

Deaton, A., and Muellbauer, J. (1980). "An almost ideal demand system", *The American Economic Review*, 70(3), 312-326.

Deaton, A. and Stone, A. A. (2013), "Two happiness puzzles", *American Economic Review (Papers and Proceedings)*, 103, 591–597.

Delfabbro, D. (2010), A Review of Australian Gambling Research: Implications for Inter-Jurisdictional Public Policy and Regulation: Update Report, http://www.gamblingresearch.org.au/home/research/gra+research+reports/a+review+of+australian+gambling+research+2010.

Dickey, D. A., and Fuller, W. A. (1979), "Distribution of the estimators for autorgressive time series with a unit root", *Journal of the American Statistical Association*, 74, 427-431.

Downs, C. (2014), "Gambling on Debt: Problem Gamblers Strategies for Concealing Deviant Identities", *Presentation to the AGRI Conference, Alberta, 2014*. http://dspace.ucalgary.ca/bitstream/1880/49991/7/Downs_AGRI_Conference_2014.pdf.

Durbin, J. and Watson, G. S. (1950), "Testing for serial correlation in least squares regression. I.", *Biometrika*. 37, 409-428.

Emran, M.S., Robano, V., and Smith, S. C. (2014), "Assessing the Frontiers of Ultrapoverty Reduction: Evidence from Challenging the Frontiers of Poverty Reduction/Targeting the Ultra-poor, an Innovative Program in Bangladesh", *Economic Development and Cultural Change,* 62(2), 339-380.

Engle, R. F. (1982), "Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation", *Econometrica*, 50, 987-1007.

Farrell, L. (2017), "Understanding the Relationship Between Subjective Wellbeing and Gambling Behavior", *Journal of Gambling Studies*, forthcoming.

Farrell, L., Hartley, R., Lanot, G., and Walker, I. (2000). The demand for lotto: The role of conscious selection. *Journal of Business & Economic Statistics, 18*(2), 228-241.

Farrell, L., Morgenroth, E., and Walker, I. (1999). A time series analysis of U.K. lottery sales: Long and short run price elasticities. *Oxford Bulletin of Economics and Statistics, 61*(4), 513-526.

Farrell, L., and Shields, M. A. (2007). "Children as consumers: Investigating child diary expenditure data", *Canadian Journal of Economics*, 40(2), 445-467.

Farrell, L. and I. Walker (1999), "The Welfare Effects of Lotto", *Journal of Public Economics,* 72, 99-120.

Feehan, P., and Forrest, D. (2007). "Distribution of UK National Lottery grants across local authority areas", *Applied Economics Letters*, 14(5), 361-365.

Ferris, J., and Wynne, H. (2001). "The Canadian Problem Gambling Index: Final report". *Report to the Canadian Centre on Substance Abuse.*

Forrest, D. (2014), “An Economic and Social Review of Gambling in Great Britain”, *Journal of Gambling Business and Economics*, 7, 1-33.

Forrest, D. (2016), “The Unhappiness of Problem Gamblers”, *mimeo*.

Forrest, D., Gulley, O. D., and Simmons, R. (2000). Testing for rational expectations in the UK national lottery. *Applied Economics, 32*(3), 315-326.

Forrest, D., Simmons, R., and Chesters, N. (2002). Buying a dream: Alternative models of demand for lotto. *Economic Inquiry, 40*(3), 485-496.

Friedman, M., & Savage, L. J. (1948). The utility analysis of choices involving risk. *Journal of Political Economy, 56*(4), 279-304.

Frontier Economics (2014), “The UK betting and gaming market: estimating price elasticities of demand and understanding the use of promotions”, *Report to HMRC*.

Gambling Commission. (2015). *Industry statistics April 2010 to September 2014.* Gambling Commission.

Gambling Commission. (2016). *Industry Statistics.* Gambling Commission.

Gambling Commission (2017), “Gambling participation in 2016: Behaviour, awareness and attitudes”, *Annual report.* URL: http://www.gamblingcommission.gov.uk/PDF/survey-data/Gambling-participation-in-2016-behaviour-awareness-and-attitudes.pdf.

Gil, A., and Molina, J. (2009). “Alcohol demand among young people in Spain: An addictive QUADS”, *Empirical Economics*, 36(3), 515-530.

Golec, J., and Tamarkin, M. (1998). Bettors love skewness, not risk, at the horse track. *Journal of Political Economy, 106*(1), 205-225.

Gorman, W. M. (1959). “Separable utility and aggregation”, *Econometrica*, 21, 63-80.

Griffiths, M. and Sutherland, I. (1998). “Adolescent gambling and drug use”, *Journal of Community and Applied Social Psychology*, 8, 423-427.

Griffiths, M., Wardle, H., Orford, J., Sprotson, K., and Erens, B. (2010). "Gambling, alcohol consumption, cigarette smoking, and health: Findings from the 2007 British Gambling Prevalence Survey", *Addiction Research and Theory*, 18(2), 208-223.

Griffiths, M. (2014), "Problem gambling in Great Britain: A brief review", Report for the Association of British Bookmakers. Available at http://abb.uk.com/wp-content/uploads/2014/09/Problem-Gambling-in-Great-Britain-A-Brief-Review.pdf.

Gruber, J. and B. Koszegi, (2001), "Is Addiction "Rational"? Theory and Evidence," *Quarterly Journal of Economics*, 116,1261-1303.

Gruber, J. and B. Koszegi, (2004), "Tax incidence when individuals are time-inconsistent: the case of cigarette taxes", *Journal of Public Economics*, 88, 1959-1987.

H2 Gambling Capital (2016), cited in *The Economist*, "The World's Biggest Gamblers", February 9, 2017.

Hansen, L.P. (1982), "Large Sample Properties of Generalized Method of Moments Estimators", *Econometrica, 50,* 1029–1054.

Hardle, W., and Mammen, E. (1993). Comparing nonparametric versus parametric regression fits. *Annals of Statistics, 21*(4), 1926-1947.

Harmon, C. and I. Walker (1995), "Estimates of the Economic Return to Schooling: Evidence from the UK", *American Economic Review*, 85, 1278-1286.

Hartley, R., and Farrell, L. (2002). Can expected utility theory explain gambling? *The American Economic review, 92*(3), 613-624.

Heckman, J. (1979). "Sample selection bias as a specification error", *Econometrica*, 47(1), 153-161.

Heckman, J. (1990). "Varieties of selection bias", *The American Economic Review*, 80(2), 313-318.

Hill, A.B. (1965). "The Environment and Disease: Association or Causation?", Proceedings of the Royal Society of Medicine. 58, 295–300.

Höfler, M. (2005), “The Bradford Hill considerations on causality: a counterfactual perspective”, *Emerging Themes in Epidemiology”,* 2005**2**:11

Hollingsworth, B., R. Pryce, and I. Walker (2016), “The Differing Response to Alcohol Price Changes: Quantity and Quality”, *mimeo*.

Huber, M. (2016), “Disentangling policy effects into causal channels”, *IZA World of Labor*, doi: 10.15185/izawol.259.

Humphreys, B., Lee, Y., and Soebbing, B. (2010). “Consumer behaviour in lottery: The double hurdle approach and zeroes in gambling survey data”, *International Gambling Studies*, 10(2), 165-176.

Institute of Alcohol Studies (2013), “Economic Impacts of Alcohol - Factsheet”, http://www.ias.org.uk/uploads/pdf/Factsheets/Economic%20impacts%20of%20alcohol%20factsheet%20August%202013.pdf

Kane, T.J., Rouse, C., and Staiger, D (1999), “Estimating returns to schooling when schooling is misreported”, *National Bureau of Economic Research WP 7235*.

Kitchen, H., and Powells, S. (1991). “Lottery expenditures in Canada: A regional analysis of determinants and incidence”, *Applied Economics*, 23(12), 1845-1852.

Klein, R. and Vella, F. (2006). “Estimating the return to endogenous schooling decisions for Australian workers via conditional second moments”, *IZA Discussion Paper 2407.*

Layard, R. (2005), Happiness: Lessons from a New Science, London, Penguin.

Leicester, A. and P. Levell, (2016), “Anti-Smoking Policies and Smoker Well-Being: Evidence for Britain”, *Fiscal Studies*, 37, 224-257.

Light, A. and A. Flores-Lagunes (2006), “Measurement error in schooling: Evidence from siblings and identical twins”, *B.E. Journal of Economic Analysis and Policy*, 5, 1.

Leser, C. E. V. (1963). “Forms of Engel functions”, *Econometrica*, 31(4), 694-703.

Lewbel, A. (2012). “Using Heteroscedasticity to Identify and Estimate Mismeasured and Endogenous Regressor Models”, *Journal of Business and Economic Statistics*, 30(1), 67-80.

Lewbel, A. (2016). “Identification and Estimation Using Heteroscedasticity Without Instruments: The Binary Endogenous Regressor Case”, *Boston College Department of Economics,* Working Paper No. 927.

Markowitz, H. (1952). The utility of wealth. *Journal of Political Economy, 60*(2), 151-158.

Mikesell, J. (1989). “A note on the changing incidence of state lottery finance”, *Social Science Quarterly*, 70(2), 513-521.

Mincer, J.A. (1974), “The Human Capital Earnings Function”, in J.A. Mincer (ed.) *Schooling, Experience, and Earnings*, NBER.

Morgan, J. (2000). Financing public goods by means of lotteries. *Review of Economic Studies, 67*(4), 785-810.

Morgan, J., and Sefton, M. (2000). Funding public goods with lotteries: Experimental evidence. *The Review of Economic Studies, 67*(4), 785-810.

Nikolova, M. (2106), “Happiness and Development”, *IZA Discussion Paper 10088*.

Office for National Statistics. (2014). *Consumer trends, Q4.* Office for National Statistics.

Orford, J., K. Sproston, B. Evans, C. White, and L. Mitchell (2003), Gambling and Problem Gambling in Britain, Brunner-Routledge.

Orford, J., Wardle, H., Griffiths, M., Sprotson, K., and Erens, B. (2010). “PGSI and DSM-IV in the 2007 British Gambling Prevalence Survey: Reliability, item response, factor structure and inter-scale agreement”, *International Gambling Studies,* 10(1), 31-44.

Orford, J., H. Wardle, and M. Griffiths (2013), “What proportion of gambling is problem gambling? Estimates from the 2010 British Gambling Prevalence Survey”, *International Gambling Studies*, 13, 4-18.

Oster, E. (2016), “Unobservable Selection and Coefficient Stability: Theory and Evidence”, *Journal of Business and Economic Statistics*, available online at http://dx.doi.org/10.1080/07350015.2016.1227711

Orphanides, A. and D. Zervos (1995), “Rational Addiction with Learning and Regret”, *Journal of Political Economy*, 103, 739-758.

Perez, L., and Humphreys, B. (2013). “The who and why of lottery: Empirical highlights from the seminal economic literature”, *Journal of Economic Surveys*, 27(5), 915-940.

Pesaran, M. H. (1974). On the general problem of model selection. *The Review of Economic Studies, 41*(2), 153-171.

Pollak, R. A. (1971). “Additive utility functions and linear Engel curves”, *The Review of Economic Studies*, 38(4), 401-414.

Powdthavee, N. (2010), "How much does money really matter? Estimating the causal effects of income on happiness", *Empirical Economics*, 39, 77-92.

Powdthavee, N. and B. van den Berg (2011), “Putting different price tags on the same health condition: Re-evaluating the well-being valuation approach”, *Journal of Health Economics*, 30, 1032–1043.

Pryce, R. (2016). “The economics of alcohol: A collection of essays”, *Lancaster University, Lancaster, UK.*

Pryce, R., Walker, I., and Wheeler, R. (2017). “How much of a problem is problem gambling?”, *Lancaster University Economics Working Paper Series No. 2017/01.*

Quiggin, J. (1991). On the optimal design of lotteries. *Economica, 58*(229), 1-16.

Robinson, P. M. (1988). Root-N-consistent semiparametric regression. *Econometrica, 56*(4), 931-954.

Rubenstein, R., and Scafidi, B. (2002). “Who pays and who benefits? Examining the distributional consequences of the Georgia lottery for education”, *National Tax Journal*, 2(2), 223-238.

Sargan, J. D. (1958), “The Estimation of Economic Relationships Using Instrumental Variables”, Econometrica, 26, 393–415.

Scoggins, J. F. (1995). The lotto and expected net revenue. *National Tax Journal, 48*(1), 61-70.

Scott, F., and Garen, J. (1994). “Probability of purchase, amount of purchase, and the demographic incidence of the lottery tax”, *Journal of Public Economics*, 54(1), 121-143.

Shaffer and D.A. Korn (2002), “Gambling and related mental disorders: a public health analysis”, *Annual Review of Public Health*, 23, 171–212.

Slutske, S., G. Zhu, M.H. Meier, and N.G. Martin (2010), “Genetic and Environmental Influences in Disordered Gambling in Men and Women”, *Archives of General Psychiatry*, 67.6, 624-630.

Spiro, M. (1974). “On the tax incidence of the Pennsylvania lottery”, *National Tax Journal*, 27(1), 57-71.

Staiger, D., and J. H. Stock (1997): “Instrumental Variables Regression with Weak Instruments,” *Econometrica*, 65, 557–86.

Stock, J. H., J. H. Wright, and M. Yogo (2002), “A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments,” *Journal of Business and Economic Statistics*, 20, 518–29.

Suetens, S., Galbo-Jorgensen, C. B., & Tyran, J.-R. (2016). Predicting lotto numbers: A natrual experiment on the gambler's fallacy and the hot-hand fallacy. *Jourenal of the European Economic Association, 14*(3), 584-607.

Suits, D. B. (1977a). “Measurement of tax progressivity”, *The American Economic Review*, 67(4), 747-752.

Suits, D. B. (1977b). “Gambling taxes: Regressivity and revenue potential”, *National Tax Journal*, 30(1), 19-36.

Sussman, S., N. Lisha, and M. Griffiths (2011), "Prevalence of the addictions: a problem of the majority or the minority?" *Evaluation and the Health Professions,* 34, 3-56.

Thorley C., A. Stirling, and E. Huynh (2016), Cards on the table - The cost to government associated with people who are problem gamblers in Britain, IPPR Report.

Tobin, J. (1958). “Estimation of relationships for limited dependent variables”, *Econometrica*, 26(1), 24-36.

Vaughan-Williams, L. (2005). *Information Efficiency in Finacnial and Betting Markets.* Cambridge, UK: Cambridge University Press.

Walker, I. (1998). The economic analysis of lotteries. *Economic Policy, 13*(27), 359-401.

Walker, I. and R. Wheeler (2016), “The Economics of Lotto Design”, *mimeo*.

Walker, I., & Young, J. (2001). An economist's guide to lottery design. *The Economic Journal, 111*, 700-722.

Wang, T. V., Potter van Loon, R. J. D., Assem, M. J., and van Dolder, D. (2016), “Number preferences in lotteries”, *Judgement and Decision Making*, 11(3), 243-259.

Wardle, H., Moody, A., Spence, S., Orford, J., Jotangia, D., Volberg, R., Griffiths, M., Hussey, D., and Dobbie, F. (2011), “British Gambling Prevalence Survey 2010”. *London: The Stationery Office*.

Williams, R.J., Volberg, R.A. and Stevens, R.M.G. (2012), The Population Prevalence of Problem Gambling: Methodological Influences, Standardized Rates, Jurisdictional Differences, and Worldwide Trends, Ontario Problem Gambling Research Centre and the Ontario Ministry of Health and Long Term Care.

Woodhouse, J. (2017), *Fixed odds betting terminals*, Houses of Parliament Briefing Paper 06946.Working, H. (1963). “Statistical laws of family expenditure”, *Journal of the American Statistical Association*, 38(221), 43-56.