Economics Working Paper Series

2018/022

The Decline and Fall of UK Lotto

Ian Walker and Rhys Wheeler

The Department of Economics
Lancaster University Management School
Lancaster LA1 4YX
UK

The Decline and Fall of UK Lotto

Ian Walker[+] and **Rhys Wheeler**[‡]

[+] Lancaster University Management School, and IZA Bonn

[‡] Lancaster University Management School

September 2018

Keywords: gambling, lotto, problem gambling, mediation, well-being

JEL codes: I3, I31, D8

Abstract

UK lotto sales have fallen by 60% from its heyday in 1996 (just two years after the introduction of the game) when weekly sales were close to £100m, to £40m per week in nominal terms by 2013 (and, with inflation averaging 2.9%, the real fall in sales revenue was approximately 75%). The aim of this paper is to estimate the demand for lotto - so as to understand the fall in lotto sales revenue, and to evaluate attempts since 2013 to arrest the fall.

Acknowledgements: We are grateful to the Gambling Commission, that regulates the UK industry, especially Ben Hader and James Holdaway, for advice and encouragement but the authors are solely responsible for the analysis, and the work in no way reflects the views of the Gambling Commission. The work was supported by an EU Framework 7 grant for the ALICE-RAP project - see http://www.alicerap.eu/, and by an Economic and Social Research Council (ESRC) Collaborative Award in Science and Economics (CASE) PhD studentship. We would be happy to assist subsequent researchers and our own code is available from the corresponding author. We are also grateful for valuable comments from Leverhulme Trust Visiting Professor Ron Harstad (grant VP2-2014-068). We are also grateful to participants at the 2016 International Conference on Gambling and Risk Taking in Las Vegas.

Corresponding author: Ian Walker, Professor of Economics, Lancaster University Management School, Lancaster LA1 4YX, UK. Ian.Walker@Lancaster.ac.uk.

1. Introduction

Lotto is the most popular form of lottery game and such games are typically used by governments and charitable organisations to raise funds. Its distinctive feature is that the jackpot prize is often not won and "rolls over" to the following draw – multiple rollovers are possible and large jackpots can accumulate.

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. 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. However, UK lotto sales have fallen by 60% from its heyday in 1996 (just two years after the introduction of the game) when weekly sales were close to £100m, to £40m per week in nominal terns by 2013 (and, with inflation averaging 2.9%, the real fall in sales revenue was approximately 75%).[1] The aim of this paper is to estimate the demand for lotto - so as to understand the fall in lotto sales revenue, and to evaluate recent attempts to arrest the fall.

The sticker price of lotto is usually fixed in nominal terms for fairly long periods. Lotto is very much a convenience product and it is usually priced in some convenient unit depending on the currency.[2] Thus, the effect of short-term price variation is usually inferred from changes in the overall prize pool since the expected value of a lotto ticket varies from draw to draw because of variation in the jackpot prize pool. As is often the case in demand models, the effective price (i.e. net of the expected value of the prize pools) is endogenous to sales and care

[1] The National Lottery introduced a variety of other products over this period which offset the lotto fall. Nevertheless, the overall revenue from the sales of the portfolio of products in 2013 was approximately £67m per week in nominal terms corresponding to approximately a one third nominal fall and approximately a 50% real fall for 1996/7.

[2] An exception was the Irish lotto game which experienced a fall in sales associated with the introduction of the euro currency. Legislation in Ireland required that prices be converted into euros exactly. In anticipation of this the price of a ticket was increased from £0.50 by 50% and, on the instruction of the euro the price was fixed, for a short period until the game could be tweaked, to a rather inconvenient figure of €0.95.

must be taken in the statistical modelling to enable the causal effect of price to be inferred. This is a particular 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. Lotto is a pari-mutuel, low-cost game which offers the possibility of large prizes with small win probabilities. Prior to a reform in October 2013, the UK lotto cost £1 to enter and regularly offered jackpots prizes of more than £8 million[3]. The game involved players choosing six numbers from 1 to 49 and participants were free to choose their own numbers.[4] A ticket is valid only for a specific draw, which typically occur on Wednesday and Saturday, in which a mechanised device chooses six numbers from 1 to 49 (59 after the second reform that occurred in late 2015). Players were rewarded with cash prizes if their chosen numbers matched between three and six of the randomly drawn combinations of numbers. A fixed prize of £10 (later £25) was awarded to those who matched three numbers, and players could win equal shares of prize pools for matching 4-6 numbers, with the jackpot being shared between those who matched all six. The 6/49 design means that the probability of any given ticket winning this jackpot prize is extremely unlikely - approximately 1/14 million (1 in 45m in the case of 6/59 design). Multiple sequential rollovers are much more likely in the 6/59 design.

Early literature (see Clotfelter and Cook, 1993, and Walker, 1998) pointed out that lotto features a "peculiar economies of scale" which imply that price is endogenous in a way, explained later, that is associated with rollovers. There have been three strands of empirical models of sales revenue – and both ignore the peculiar economies of scale feature of lotto. The first strand is based on strong functional form assumptions, that sales are linear in price, and inferences from such models have assumed that game design affects sales only through price. The second, recognises the endogeneity of price and takes a reduced form approach that assumes that rollovers are exogenous, and inferences from such modelling has assumed that game design affects sales only through. The third strand literature has assumed linearity in price and instrumented price with rollovers. There are two objections to existing work: it has all been based on arbitrary functional form assumptions that sales depend linearly on the effective price or; and it has all either assumed that effective price variation arising from rollover is exogenous, or rollovers can provide an instrument for the effective price.

[3] 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.

[4] From 1996 onwards, the vending terminal could instead randomly make the selection of numbers for them

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. Thus, the contribution of this paper, relative to the existing literature, is to correct this shortcoming though a novel identification strategy and to apply semi-parametric methods.

Section 2 provides an overview on the lotto literature. Section 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 4 proposes that a novel yet powerful alternative instrument candidate is available by exploiting systematic non-random number selection by lotto player – a phenomenon known as 'conscious selection'. Section 5 describes the data used in Section 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 suggest 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 6 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 7, is to extend their work by adopting a totally flexible approach in a semi-parametric estimation and test this against a parametric 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 rejuvenating dwindling sales and, by extension, good causes funding. The final contribution of this paper, in Section 8, uses data beyond the main sample to examine whether these changes were likely to be successful in this objective.

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.

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 and hence funds 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 choosing 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. 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 the parametric modelling of current sales (or its log), 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 finer gradation in the prize pools that a single device. 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 overall sales revenue, $\tau \in [0,1)$, to pay for operating costs and tax liabilities, and to also fund public goods in many cases. 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. 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[5]. This paper refers to a draw *following* the draw where there are no jackpot winners 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. Since the jackpot is transferred to the next draw it raises the expected value of prizes for that draw, but *decreases* the ex ante expected value of a ticker for the previous 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 to setting the shares of sales allocated to each prize pool, the operator is also able to determine

[5] 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.

the statistical difficulty of the game by appropriate choice of n and N to influence π_n as[6]: $\pi_n = n!\,(N-n)!/N!$ 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 $1-\pi_n$. Hence, if S_t tickets are sold the probability of a rollover occurring is:

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

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

$$P_t = (1-p_{R,t})\left(\tau - \frac{R_t}{S_t}\right) + p_{R,t}. \tag{2}$$

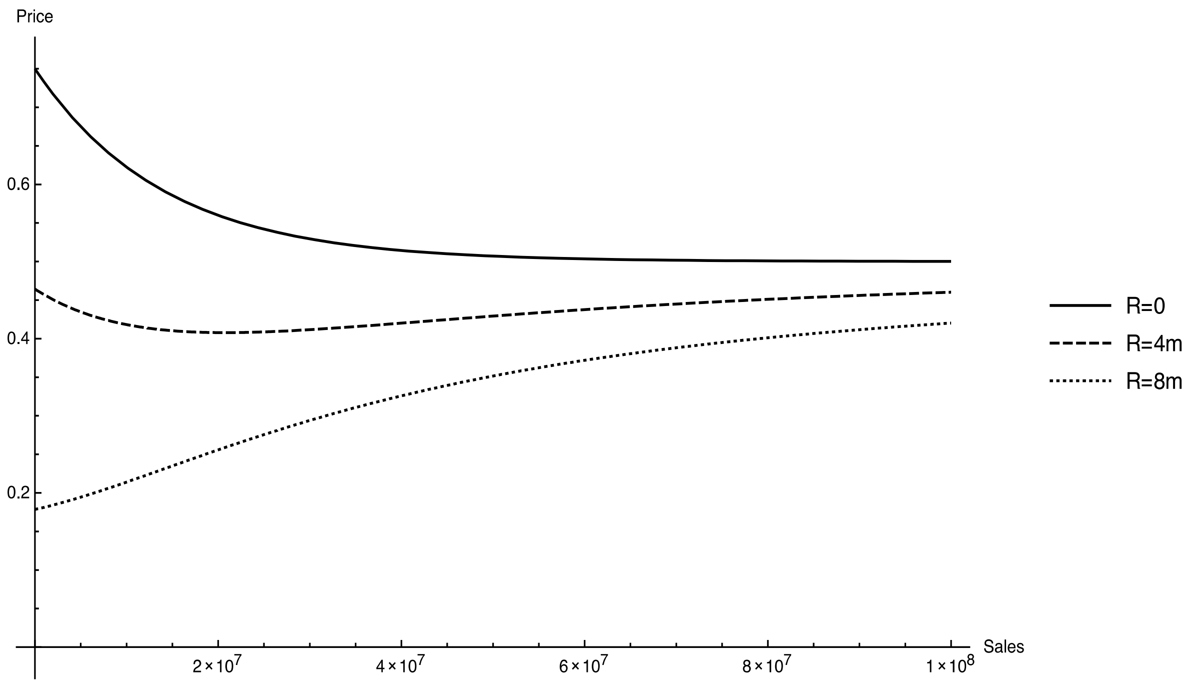
This can be thought of as the inverse supply function of lotto tickets, where $R_t = 0$ corresponds to the non-rollover case. Several features of lotto game designs are clear from the definition of price. Trivially, increasing the proportion of sales retained by the operator increases the price, $\delta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

[6] 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.

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.

Figure 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, since higher sales makes it more likely 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 1: Inverse supply function for a 6/49 lottery with $\tau = 0.5$ *and rollovers of £0, £4m, £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[7]: $J_{kt} = \rho_k\left[(1-\tau)S_t - 10W_{3,t}\right]$ – that is, the k^{th} share of the overall prize pool stripped of the non-parimutuel £10 prizes.

A final feature of the UK lotto is the twice-weekly draw. In the sample period analysed below, UK lotto draws occurred on Wednesdays and Saturdays of the same week[8]. 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 draws are estimated 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.

[7] 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 $\rho_4 = 0.22$.

[8] Between November 1994, the introduction of the game, and February 1997 draws only occurred every Saturday.

This assumption is apparently justified simply by examining the descriptive statistics in Section 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 between rollover and for non-rollover draws on Saturdays is subsequently much smaller than it is for Wednesdays - as would be expected from Figure 1.

The possibility of a rollover in lotto has two important implications. Firstly, the operator is able to choose π_n (by choosing n and N) to best suit the likely size of the market and so influence price variation through 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 own purchase 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 choose the design to balance the gain in sales when rollovers occur with the suppression of sales because of the probability of one occurring.

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, 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 1 and the importance of this endogeneity will be empirically important if sales are distributed around a relative steep part of the inverse supply curve. In practice, games typically are designed so that they operate below such level. This engenders a high enough probability of rollovers occurring – events which tend to increase sales in the next draw (and, to a lesser extent, in subsequent draws through habituation). But not so high as to encourage high levels of intertemporal substitution that prevents large jackpots building up quickly.

4. Identifying the demand for lotto

As usual, the solution to identifying a demand curve is to rely on exogenous variation in the (inverse) supply curve. This is the essence of what we do here. While rollovers are random events and therefore might be thought of as the source of a solution to the endogeneity of price, the product is more complicated than this intuition , as we make clear below.

Using the variation in the effective price as a determinant of lotto sales allows the price elasticity to be inferred. Walker (1998) suggests that, because marginal revenue is close to zero and intertemporal substitution is likely to be limited, the estimated long-run price elasticity of demand of -1.07 suggested that the recently introduced UK lotto game was appropriately designed to maximise revenue. 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 simple 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 shocks, X_t, to be estimated via OLS:

$$S_t = \alpha + \sum_{i=1}^{I} \beta_i S_{t-i} + \eta P_t + \gamma X_t + \varepsilon_t. \tag{3}$$

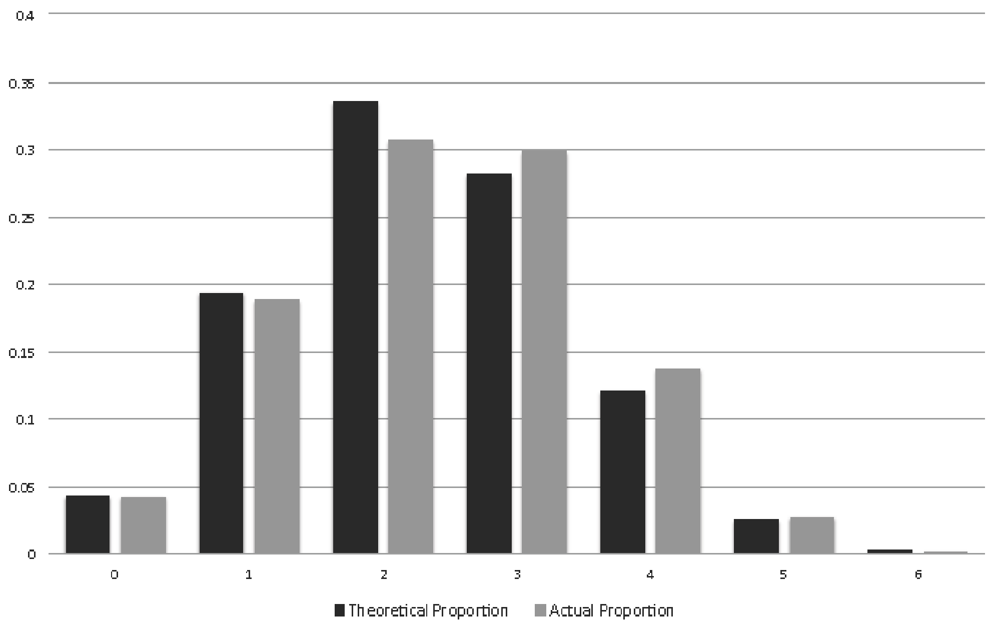
Since the game is now drawn twice-weekly and Wednesday and Saturday are treated as different games here, sales in $t-1, t-3$ ... are 'cross' lags of sales and $t-2, t-4$ … are 'own' lags of sales. It is possible to retrieve a long run price elasticity of demand, ϵ_{LR}, evaluated at the mean sales, from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

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

Recalling the definition of price in equation (2), 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 will result in the size of the rollover being correlated with current sales. This casts doubt on the validity of relying on rollover size as an instrument for the current jackpot, price, and rollover.

Figure 2 establishes that the machinery is unbiased. This highlights how the number of ‘large’ (i.e 32+) winning numbers that appear in any given draw is indeed random by comparing the actual distribution of how many large numbers make up the winning set with the theoretically expected number – this is confirmed by a KS test that the distributions are identical. That is, Figure 2 shows that higher numbered balls are just as likely to be picked by the machinery as lower numbered ones.

Figure 2: Theoretical and actual proportion of draws with n numbers > 31 drawn



The proposed solution to the endogeneity issue here, relies on the machinery being unbiased, and 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 our data, and the design parameters of the UK game, the supply-side theory above suggests that only 6% of Saturday draws should roll-over, whereas the actual proportion of Saturday draws with no jackpot winners is 13%. 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 players for their own tickets. Moreover, so long as each number has appeared reasonably frequently, 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 sold 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 2.04% (i.e. 1/49) that would have been expected if numbers had been chosen randomly. Overall, they find that numbers 1-12 prove to be most popular and the numbers 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 reduces the pari-mutuel prize fund[9].

The definition of price derived in (2) implicitly assumed 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 (2) only accurately captures this probability if each of the tickets sold are unique (and it 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: $p_{R,t} = (1-\pi_n)^{f(S_t,\delta_{j,t})}$ 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:

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

$$P_t = \left[1 - (1 - \pi_n)^{f(S_t,\delta_{i,t})}\right]\left(\tau - \frac{R_t}{S_t}\right) + (1 - \pi_n)^{f(S_t,\delta_{i,t})}. \tag{5}$$

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 of numbers. This reflects the propensity for players to favour birth dates in their number selection and follows from the findings of Farrell *et al* (2000). 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 be positively related. Specifically, the following variables are constructed to instrument rollover probability, where the complexity of the middle expression arises for the variable number of days in each month:

$$\begin{aligned}
\Delta_{small,t} &= \sum_{i=1}^{12} \delta_{i,t} \\
\Delta_{medium,t} &= \sum_{i=1}^{28} \delta_{i,t} + \frac{45}{48}\delta_{29,t} + \frac{11}{12}\delta_{30,t} + \frac{7}{12}\delta_{31,t} \\
\Delta_{large,t} &= \sum_{i=32}^{49} \delta_{i,t}
\end{aligned} \tag{6}$$

where $\delta_{i,t}$ is 1 if the number *i* is among the winning numbers in draw *t*, and 0 otherwise. Defining just three instruments in this way is preferable to the more obvious solution of using all but one number dummies because it overcomes the problem that so many dummies would make them individually weak instruments of rollovers because of the limited frequency that each number appears in the winning configuration.

On average, 14% (28%) of Saturday (Wednesday) draws roll over - there are more on Wednesday draws because sales are much lower and so the probability that *no* ticket with the winning combination has been sold is larger. Figure 3 shows that, in fact, when more large numbers (32+) feature in the winning combination there is a higher proportion of rollovers than average. the effect of conscious selection by players affects the probability of a rollover occurring. Figure 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 (2), 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 4.

Figure 3: *Proportion of draws which roll-over by number of balls > 31 drawn: Wednesday and Saturday draws*

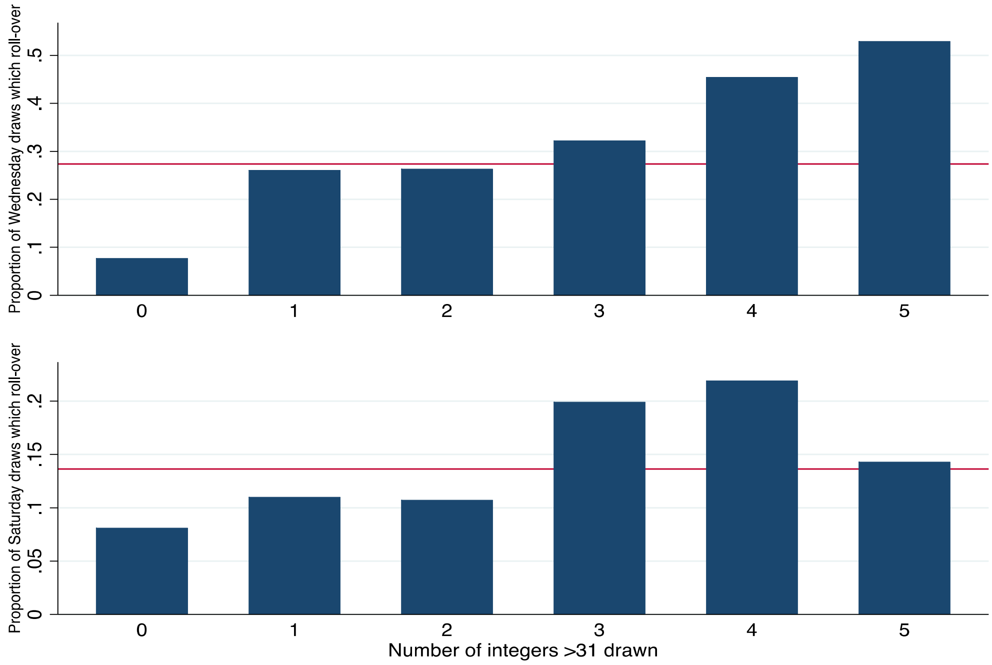
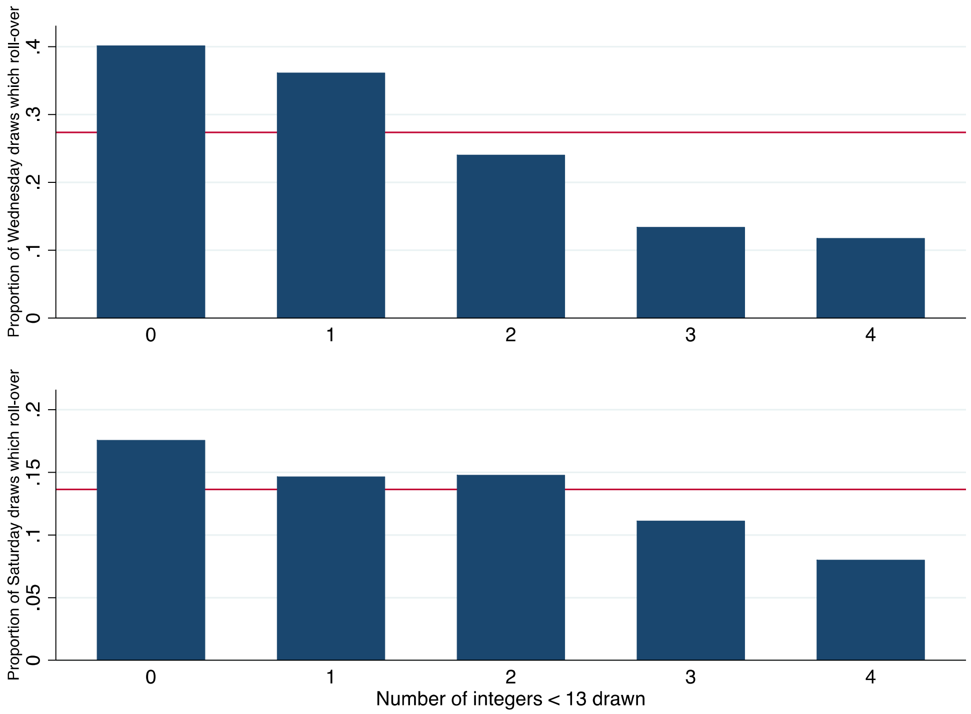


Figure 4: *Proportion of rollover draws by number of balls 1-12 drawn: Wednesday and Saturday draws*



5. Data

The analysis in this paper 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[10]. 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. We return to the later observations below for out-of-sample analysis.

Table 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%. As expected from 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.

Note that the effective price is lower for rollover draws than for non-rollover draws at any level of sales. This lower price for Saturday non-rollover draws than similar draws on Wednesday can be explained by sales being significantly larger, resulting in the likelihood of a rollover occurring being much smaller.

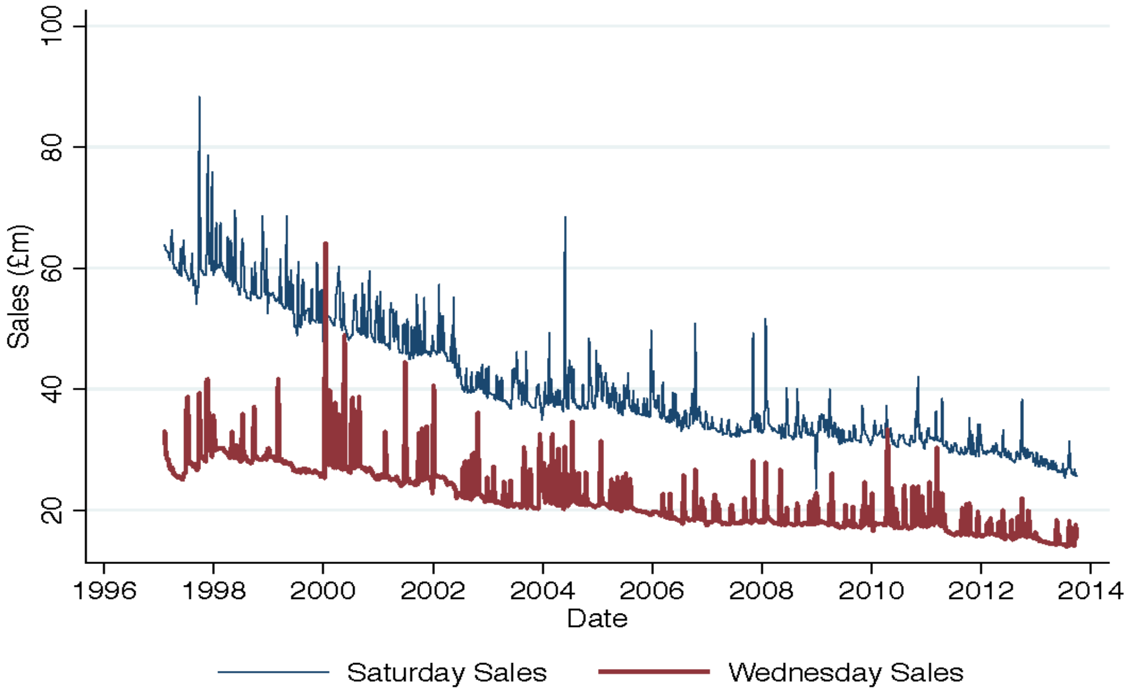
[10] Section 7 uses more data on draws by including those made after the game re-design in 2013 to evaluate whether the changes in design had a positive effect on sales.

The effect of rollovers on 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 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 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. Peaks in the graph highlight the impact that rollovers (and double, triple and even quadruple rollovers) have on ticket sales. This trend in sales figures are somewhat irrelevant to our econometric analysis, although it is very relevant for the operator whose job it is to maximise sales revenue. Rrather it is the effect of rollovers on sales (either directly or via the influence on price) that are the focus of our attention. Nonetheless, these declining sales figures no doubt contributed to the decision to redesign the game in October 2013 and second design change that occurred in 2015

Table 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

Figure 5: Draw-by-draw lotto sales from February 1997 to September 2013



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 suggests that the effective price is endogenous to sales and so least squares will produce biased estimates of the causal effect of price – that is necessary to infer the causal effect of changes to price that are driven by changes in the design parameters of the game. 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) 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 one solution to modelling the rollover size and probability simultaneously, and this is what is used here. Moreover, this 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 be possible under a Tobit model. The selection equation of Heckman's procedure involves using the number indicator variables of equation (7), along with all exogenous variables from the second stage, from the previous draw 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:

$$R_t = \begin{cases} R_t & if\ R_t^* > 0, \\ 0 & otherwise \end{cases} \tag{8}$$

where $R_t^* = \gamma_0 + \gamma_1 S_{t-1} + \gamma_2 S_{t-2} + \gamma_3 \left(W_{3,t-1} - \pi_3 S_{t-1}\right) + \eta_t$. The rollover probability is given by:

$$\Pr(R_t^* > 0) = \alpha_0 + \alpha_1 S_{t-1} + \alpha_2 S_{t-2} + \Delta_{small,t-1} + \Delta_{medium,t-1} + \Delta_{large,t-1} + \nu_t \tag{9}$$

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, i.e. not pari-mutuel, prize 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 (approximately 1 in 57) multiplied by the number of tickets sold. Thus, the number of 3-ball winners is itself NOT a valid 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), we 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 the exogenous variation in 3-ball winners has in determining the size of a rollover should one occur, but this 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:

$$S_t = \beta_0 + \beta_1 S_{t-1} + \beta_2 S_{t-2} + \beta_3 \widehat{P_t} + \psi_t X_t + \varepsilon_t. \tag{10}$$

Table 2 presents the first-stage Heckman model estimates. The bottom panel 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. However, 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 often chosen by at least one player, thus making conscious selection more difficult to detect. Nonetheless, the sign 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 our use of this idea as the basis of an instrument.

The top panel of Table 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 money 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.

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 were 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. The Saturday estimate of λ fails to reject the null, perhaps because the typical level of sales is sufficiently high that the inverse supply survey is sufficiently close to horizontal. In contrast, Wednesday sales are typically much lower and this leads to rejection of exogeneity for Wednesday draws because the slope of the inverse supply curve at that lower level of sales is steeper.

Table 2: Heckman selection model estimates for rollover size and probability

	Saturday	**Wednesday**
Dependent Variable	$\boldsymbol{R_t^{Sat} \mid R_t^{Sat} > 0}$	$\boldsymbol{R_t^{Wed} \mid R_t^{Wed} > 0}$
$W_{3,t-1} - \pi_3 S_{t-1}$	-4.911***	-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	$\mathbf{Pr}(\boldsymbol{R_t^{Sat}} > \mathbf{0})$	$\mathbf{Pr}(\boldsymbol{R_t^{Wed}} > \mathbf{0})$
$\Delta_{medium,t-1}$	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.

Long-run elasticity estimates in Table 3, derived from using instrumented price, support this. The OLS and H2S elasticities are not significantly different from each other, while the Wednesday estimates are significantly different. The price elasticity estimate for Saturday remains qualitatively at -0.6 suggests that revenue could be increased by making the game more expensive. However, for the Wednesday game a price elasticity estimate of -1.5 suggests that the game is over-priced and revenues could be increased by making the game more attractive to play.[11] Assuming a price elasticity of -1 would maximise revenues for the lotto monopolist, then the estimates from Table 2 and rearranging the elasticity equation (7) suggests, ceteris paribus, that increasing the price of Saturday draws from £0.49 to £0.77 and reducing the price of Wednesday tickets from £0.50 to £0.36 would increase revenues from the games.

Table 3: Second-stage estimates of lotto demand

	OLS		Control Function using Heckman First-Stage	
Dependent variable	$\boldsymbol{S_t^{Sat}}$	$\boldsymbol{S_t^{Wed}}$	$\boldsymbol{S_t^{Sat}}$	$\boldsymbol{S_t^{Wed}}$
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	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 (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.

[11] 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.

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 rationalising unfair bets (Forrest *et al*, 2002). Moreover, these models assume that the effect of 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 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. In the context of lotto, Walker and Young (2001) 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. Thus, 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) and so this section replicates their work, but using conscious selection as the identification strategy. Moreover, we extend our work to include a fully flexible semi-parametric specification as well as one with 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.

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 lotto prizes. Forrest *et al* (2002) evaluate the extent to which jackpot models are better able to explain ticket sales relative to effective price models. Their jackpot model predicts that an additional £1 million in the jackpot size would increase sales by £22,000 (Wednesday) and £53,000 (Saturday), conditional on the total prize pool remaining the same. Using rollover size as 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. However, non-nested tests were inconclusive.

This section is concerned with estimating the following:

$$S_t = \beta_0 + \beta_1 S_{t-1} + \beta_2 S_{t-2} + \beta_3 \widehat{R_t} + \beta_4 \widehat{R_t^2} + \beta_5 \widehat{R_t^3} + \gamma X_t + \varepsilon_t. \qquad (12)$$

Since rollover size is endogenous to sales, causal estimation of equation (12) is again reliant on IV techniques. Fortunately, it is possible to recycle our identification strategy and the Heckman selection model estimates presented in Section 6. Rather than imposing the strict functional form that sales if linear in the expected price whose variation is driven by equation (2), estimation of this rollover model simply includes the (cibic in) predicted values of rollover size, $\widehat{R_t}$, from the Heckman selection model in place of R_t.

Table 4 presents OLS and instrumented rollover estimates of this model for both Wednesday and Saturday games[12]. OLS continues to show 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. However, an F-test suggests that, for Saturday draws, R_t^2 and R_t^3 are jointly 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.

[12] 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 5. Estimates of models which are linear and quadratic in rollover size are reported in the Appendix.

Table.2: OLS and Heckman-instrumented estimates of rollover induced ticket demand

Dependent variable	**OLS** S_t^{Sat}	S_t^{Wed}	**Heckman First-Stage** S_t^{Sat}	S_t^{Wed}
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)
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.

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 and past sales in the price model is erroneous and was not being detected because of the restrictive nature of the 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 order to evaluate whether this reduced form model is preferable 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 and $H_0: M_1$ 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 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. 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, both non-nested testing of both OLS and instrumented estimates are inconclusive. The rejection of both 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. This encourages proceeding with a semi-parametric estimation routine.

Table 3: 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

This section has so far assumed that g, from equation (12) below, is somewhat arbitrarily cubic in R_t, whilst the price model estimated in Section 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 or not. Specifically, the semi-parametric technique used here estimates directly the following specification:

$$S_t = \beta' \boldsymbol{X}_t + g(R_t) + \varepsilon_t. \tag{12}$$

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 this estimation method in the presence of endogenous regressors, rollover size is instrumented as per the Heckman first-stage in Section 6[13], and the residual of this first stage is then included in the non-parametric component. Other controls, $\boldsymbol{X}_t$, form the parametric component of this specification and include own and cross lagged sales and variables to control for trend.

Table 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 is also significant indicating the assumption that rollover size is endogenous is justified. 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 the 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 models. The difference in these tests between Wednesday and Saturday models are not surprising upon examination of Figure 6 which are plots of the nonparametric estimates of the function g for Wednesday and Saturday 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

[13] Details of this estimation procedure can be found in the Appendix.

non-parametric fit indeed looks close to quadratic, with what seems to be an increasing gradient for larger rollover sizes.

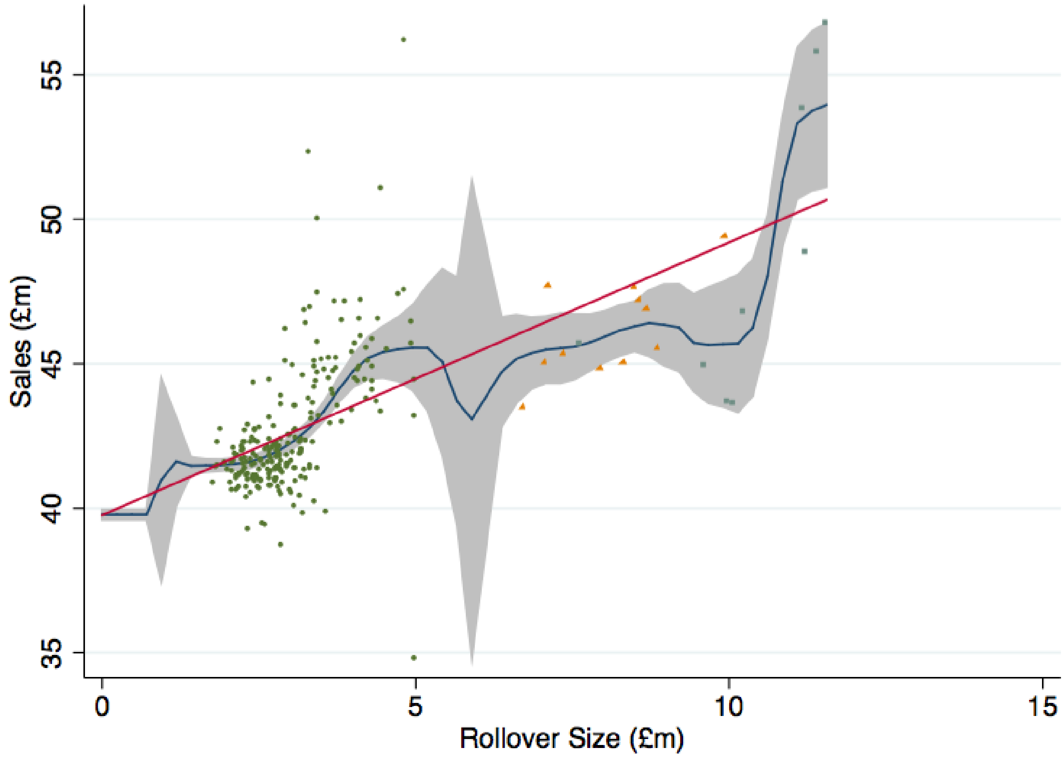
Differing estimates suggest that players in Wednesday and Saturday draws respond differently to increases in rollover sizes. Moreover, the non-linear relationship between rollover size and sales for both Wednesday and Saturday draws does indicate that previous literature using price variation alone is not sufficient to accurately capture the tresponse of players to changes in the prize distribution.

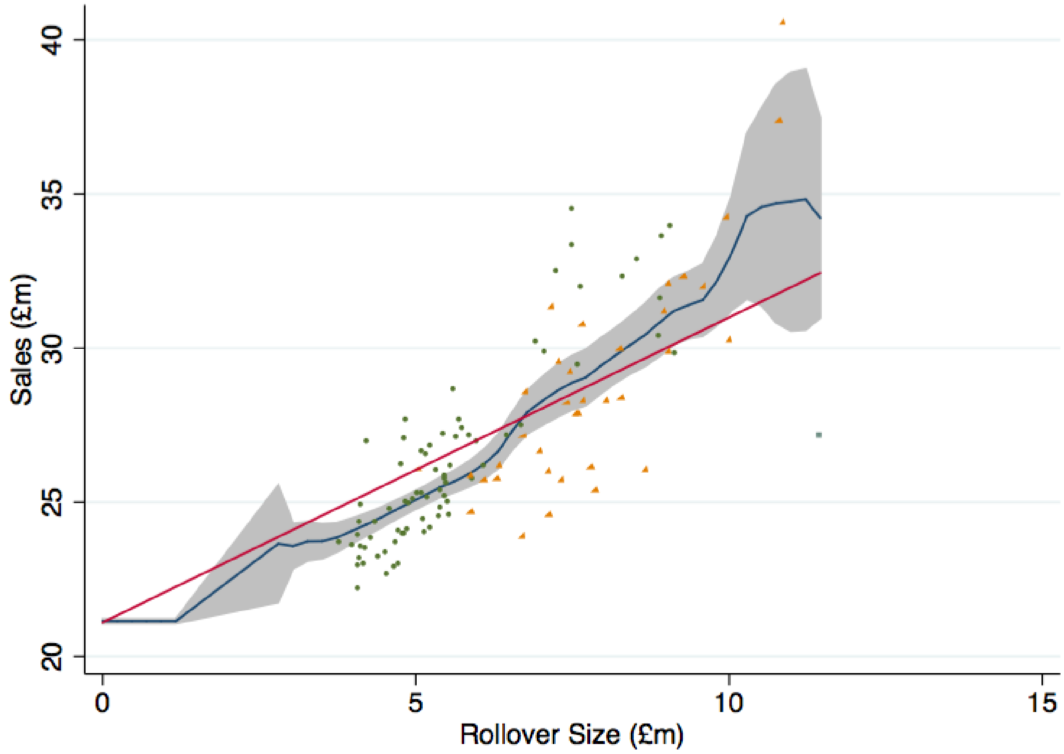
Table 4: Parametric coefficient estimates from the semi-parametric models of sales

Dependent Variable	$\boldsymbol{S_t^{Sat}}$	$\boldsymbol{S_t^{Wed}}$
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 of the non-parametric vs parametric fitted models:		
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 6: Non-parametric estimates of rollover size effects for Saturday and Wednesday





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

8. Evaluating lotto reforms

Changes made to the UK lotto game in October 2013 and October 2015 were likely to have been in response to the dwindling sales which can be seen in Figure 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 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 were effective in rejuvenating sales and, if so, how effective were they?

The price model estimates from Section 3 tentatively suggested that 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.

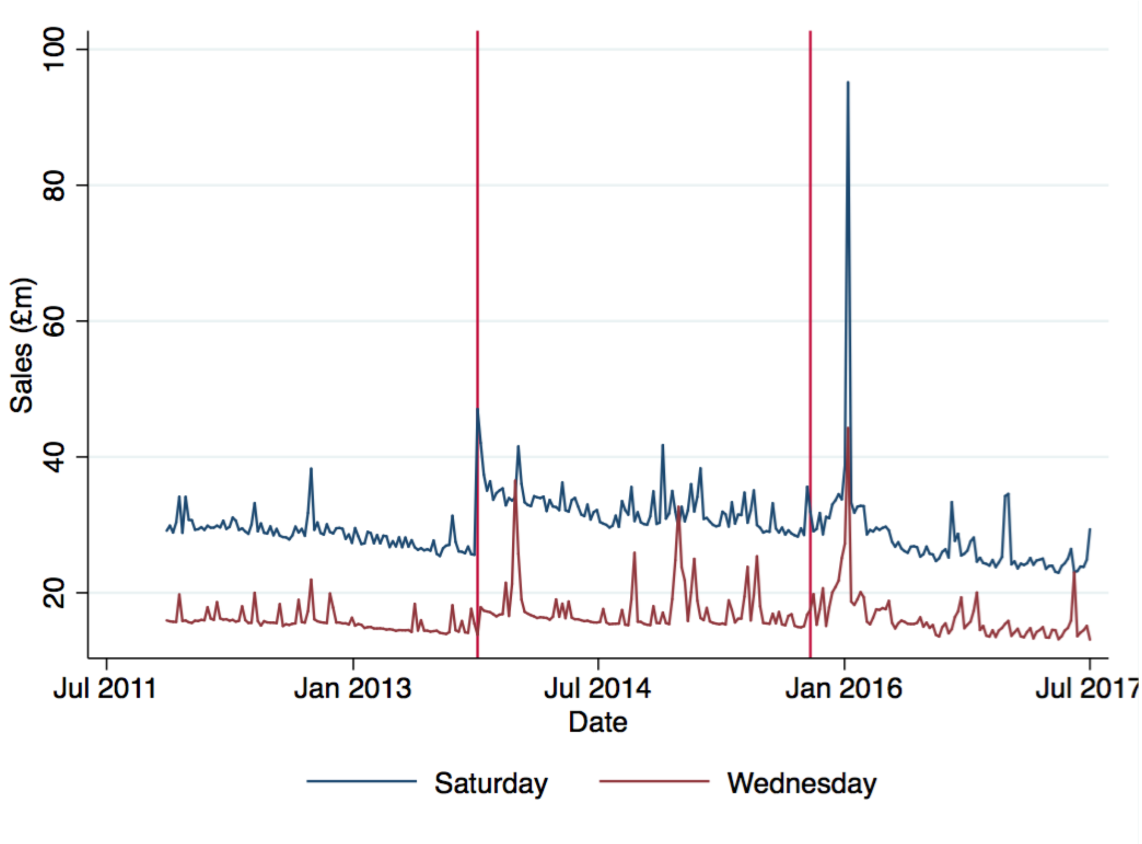
Secondly, the decrease in price for the subsequent draw is smaller than it would have otherwise been because the size of the rollover has been reduced. Again, this decreases the average price of Wednesday draws relative to Saturday because Wednesday have fewer jackpots enhanced by rollovers from Saturday 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 per draw) being awarded in raffle prizes, partly at the expense of the jackpot.

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 of 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 collectively 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 7 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. 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 revenue numbers. The 2015 re-design does not obviously display such an 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 a cap on the number of consecutive draws that a rollover was allowed to happen – instead opting to terminate consecutive rollovers when the jackpot prize reached £50m in value[14]. 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 . Moreover, the sales depicted towards the end of the sample in Figure 3 seem to return to similar figures seen before October 2013 – suggesting that the changes made to the game offered only a temporary boost in sales.

Figure 7: Draw sales for the UK lotto (Nov 2011-Jul 2017)



[14] 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, then it is shared between winners of the next highest prize tier.

Table 7 compares sales revenue for the two-year period before the 2013 reform with the post-2013 design and the post-2015 design. Following the 2013 change, average weekly revenues rose by around £5m (over 10%), an increase of just under £2m in revenues from Wednesday draws and around £3m for Saturday draws. The larger proportionate rise on Wednesday is consistent with the raffle prizes being more highly valued when sales are low. 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. In contrast the late 2015 reform appears to have had an approximately 10% negative effect on sales – which was proportionately twice as large on Saturday as Wednesday, perhaps reflect the further shift towards raffle prizes that are more valuable for Wednesdays with its lower level of sales.

Table 7: Comparison of sales revenues between game designs (£m per week)

	Pre-2013 game	**Post-2013 game**	**Post-2015 game**
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 7 and Table 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 (lack of) success of these reforms. 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 constructing this missing counterfactual is to use out of sample forecasts from the estimates of the models presented earlier. To give a flavour of the implications we use the "structural" model (columns 3 and 4 of Table 3) where all aspects of the reform are assumed to be mediated through the expected value of tickets.

Given the autocorrelation of sales, and the endogeneity of both price and rollovers this approach requires dynamic forecasting. This is fraught with difficulty since, without the counterfactual data, the variation in "price" is driven by both the counterfactual size and, particularly, the counterfactual frequency of rollovers, which are not known. The endogeneity of rollovers to sales makes deriving these from the equations of the model difficult. Our approach is to anchor the predictions at the last week of the original design and use these as the lagged values in the following week. On the basis of the last pre-reform sales figures for Wednesday and Saturday we compute the expected rollover probability and size in the following draw.

The one parameter we lack is the effect of the sticker price on demand. If tickets were perfectly divisible then it might be reasonable to assume that the sticker price has no effect: doubling the sticker price will also double the prize pool and the game remains effectively the same as before. However, there is no facility for buying half a ticket - although one might agree with someone else to share the costs (and benefits) of a single ticket. Someone who is only on the margin of buying a £1 ticket are likely to drop out rather than buy a £2 ticket, given that half a £2 ticket can no longer be obtained.

But we have no way of inferring what the effect of the sticker price reform has been from the pre-reform time series data. Instead we use the available microdata and assume that tickets are perfectly indivisible together with estimates of the probability of buying none, one, two, three etc. We observe the distribution of (household) purchases using cross section data in the Living Costs and Food Surveys corresponding to Nov 2012 - Oct 2013 when tickets were available at £1 each; and those from Oct 2013 to Sept 2014 when tickets cost £2. LCF collects lottery spending through a diary record for each household. We find that there is under-reporting in the data, by comparing with the time series sales data, of approximately 31% (both before and after the reform). The data records fortnightly spend, and less than 5% of participating households record non-integer amounts. However, 17% of participating household house record odd (i.e not even) integer amounts indicating that they do not buy the same number of £1 tickets in both weeks. Thus, we round the data into the nearest *even* integer to get an approximation of weekly tickets bought.

It is not possible to say where in the distribution of spending this shortfall comes from. For simplicity, we overweight all those households who purchase at least one ticket in each of the two years of data separately to ensure that the average weighted spending in each year matches the average official sales revenue figures. We underweight the game non-participitants by the same degree to ensure that the weighted data yields the same unweighted sample size. That is, we assume that the under-recoding is entirely due to players misreporting zero expenditure when in fact they have bought tickets. The participation rate is 40% in the diary fortnight averaged over the year immediately prior to the £2 reform and 29% in the year immediately after the £2 reform. We also assume that such forgetfulness is random, so we inflate the proportion recording £1 by the same proportionate rise as the proportion recording £2 etc.

We treat demand as a latent variable and estimate a zero-inflated negative binomial model to estimate the probability of buying 0, 1, 2, 3, etc in the diary fortnight in the 12 months *before* the reform as a function of gross income (see Appendix section E for the distribution of lotto expenditure across the income distribution), a quadratic function of the age of the household head, a dummy for male household head, dummies for each fortnight of the year, dummies for each region, and dummies for one, three, four or more adult households. We then predict how many would spend £2, £4, £6 etc in a week if 1, 3, 5 etc are no longer available. We do this by predicting the probabilities that those observed to buy a £2 ticket in the postreform year would be more likely to buy 0 rather than 2, and the same for those who would buy or 2 over 0 (and 2 vs the most likely of 1 and 3, etc). We then gross up these probabilities proportionately to match the overall official sales in that post reform year. We then compare the proportions predicted to spend nothing, £2, £4, etc. with the actual proportions in the following year of LCF data (again corrected for under-reporting). We find we underpredict the actual proportions of £2 by 38%, £4 by 30%, and £6 by 25%, and the estimates beyond that are very imprecise. We aggregate these to find an overall over-prediction of 35%. Thus, our best guess is that doubling the sticker price was responsible for a drop in the number of tickets bought by about one third across the year. Unfortunately, we are not able to do this separately for Wednesday and Saturday, which is a pity since there is appears to be some difference in the aggregate data of the size of the drop in the ticket sales figures.

We then proceed to simulate the aggregate sales. For each game we used the lagged sales figures to forecast the level of sales in the next draw and the probability that there will be no winners, using the observed winning numbers and the prior level of sales. We assume that the time trend and the pattern of autocorrelation remains the same. We model the probability of matching 3 numbers and deduct these (now £25) fixed prizes from the pari-mutuel pool. This rollover probability could be used to weight the forecast for the following draw – that is, we could make two forecasts for if a rollover occurs or not and weight the two predicted sales together. We could then use these two predictions to include as the lagged dependent variables in the following draw, which would also have two outcomes. This geometric series would quickly result in billions of calculations being required. Our alterative is to simulate only one outcome from each draw and at each iteration take a draw from the binomial distribution that determines whether a rollover occurs of not. We apply the appropriate cap to rollovers. We cumulate sales until the end of 2015 when the second reform kicks-in, and average the result.

The second reform skews the pari-mutuel prizes towards the jackpot, changes the rollover cap so this switches on after the jackpot reaches £50m, and changes the number of numbers by adding 50-59. We continue to apply the same adjustment for the £2 ticket as before. We assume that these are chosen with the same probability that 32 to 49 were chosen prior to this second reform. We adopt the same prediction process as with the first reform to prevent the number of predictions we have to make from exploding. We now use this to also simulate the probability of matching 2 balls since this entitles such winners to a free ticket in the next draw. We assume that all such 2-ball prizes are claimed and that these are financed from reducing the prize pool for the pari-mutuel prizes. Again we cumulate the results, for both draws, until the end of the data series and average.

We finally need to simulate the counterfactual level of sales that would have occurred had the game design not be changed. We assume the econometric parameters still apply and that the game design remains the same. However, because we assume that there will continue to be the same trend decrease in sales, we now have to predict how the rollover probability changes over time. Thus, we ignore the post 2012 reform data and project sales forward one week at a time allowing for the downward trend. Each week we compute the rollover probability from the predicted sales in the previous week. And each week we proceed as before and compute the new level of sales assuming that the rollover had occurred with the calculated probability. We repeat for subsequent weeks and cumulate the results. In the case of the second reform period we no longer know what the numbers on the winning ticket would have been since there are now ten additional numbers. Instead, we generate pseudo winning numbers from the 49 original numbers assuming that the degree of conscious selection continued to remain the same.

Table 8 shows the actual and forecast sales, per week, in nominal terms (i.e not adjusting for inflation over the post reform years). Sales revenue, if the original game design have been retained, would have continued to fall to £43.9m per week in the first reform period and then down to £40.2m in the second reform period. Thus, the first reform resulted in a large rise in sales of £5.9m per week. However, the second reform returned sales to pre-2013 levels. Note, however, that this is not as bad as might be thought because the counterfactual level of sales would also have fallen, to £40.7m. Overall, the second reform had relatively modest adverse effects of revenue, relative to the first reform design, and this is likely to be due to the longer odds coming on top of the lower numbers if tickets sold due to the £2 sticker price.

Table 8: Comparison of sales revenues across game designs (£m per week)

	Pre-2013 game	**Post-2013 game**	**Post-2015 game**
Actual sales revenue in the three periods	44.5	49.8	44.4
Predicted sales revenue under old design	44.5	43.9	40.7
Difference between actual and predicted sales revenue	0	5.9	3.7

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

9. Conclusion

This paper 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 paper 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 but when compared with the counterfactual level of sales the damage relative to the 2013 reformed design is modest. However, our crude attempts to forecast sales out of sample from the model developed from the original game design mirror suggests that the decline in sales seems likely to continue, although we will need a longer run of post re-design sales to have the power to reject the hypothesis that the trend has changed.

References

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.

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.

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.

Farrell, L., Hartley, R., Lanot, G., & 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., & 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.

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

Forrest, D., Simmons, R., & 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.

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

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

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

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

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

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

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

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.

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

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.

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

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

Appendix

A Estimates with rollovers as instruments

Table A1 Comparison of second stage estimates of lotto demand using rollover size and conscious selection as instruments

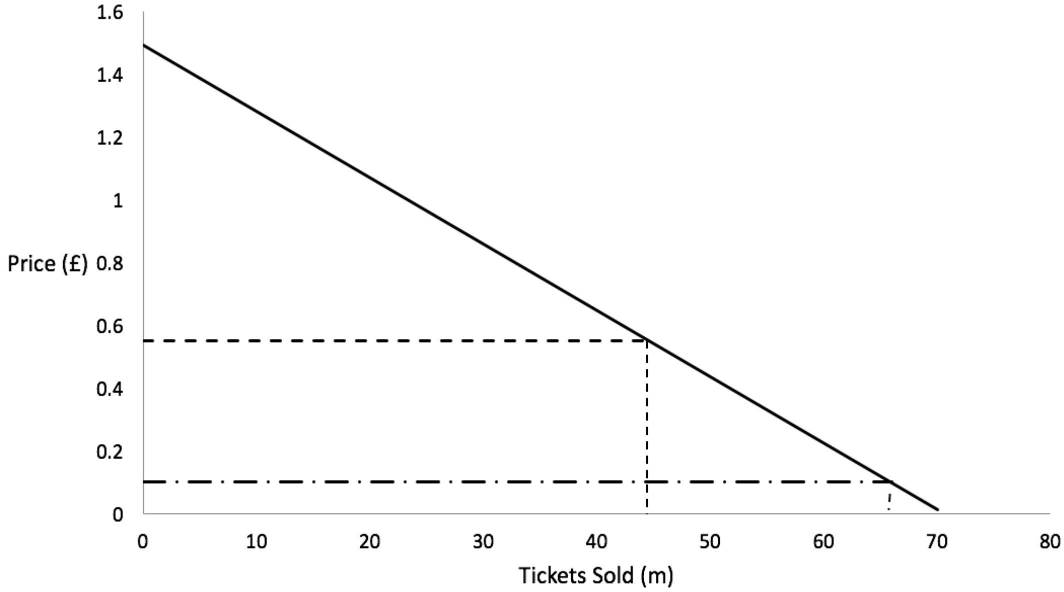
Instrument:	**Rollover Size**		**Conscious Selection**	
Dependent Variable	$\boldsymbol{S_t^{Sat}}$	$\boldsymbol{S_t^{Wed}}$	$\boldsymbol{S_t^{Sat}}$	$\boldsymbol{S_t^{Wed}}$
	(1)	**(2)**	**(3)**	**(4)**
S_{t-1}	0.137***	0.119***	0.133**	0.0643*
	(0.0478)	(0.0258)	(0.0546)	(0.0332)
S_{t-2}	0.109***	0.028**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

B. 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. Figure A1 shows such a demand curve, for Saturday tickets, parameterised using column 5 of Table 3, as well as indicating average price (dashed line) and marginal cost (dot-dashed line) 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). Figure A1shows such a demand curve, for Saturday tickets, parameterised using column 5 of Table 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).

Figure A1: Linear demand for Saturday lotto tickets



C. Semi-parametric Estimation

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, $S_t = \beta'\boldsymbol{X}_t + g(R_t) + \varepsilon_t$. Consistent estimation of this requires that $E(\varepsilon_t|\boldsymbol{X}_t, R_t) = 0$. Robinson's estimator for the parameter vector, β, and unknown function, g, can then be obtained by transforming the model to become $S_t - E(S_t|R_t) = \beta'\big(\boldsymbol{X}_t - E(\boldsymbol{X}_t|R_t)\big) + \varepsilon_t$ 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 estimating β, it is then possible to estimate the function g using the following, $\hat{g}_h(R_t) = \widehat{m}_h^S(R_t) - \beta'^{\widehat{m}_h^X}(R_t)$. 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 $R_t = \gamma Z_t + \nu_t$ 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: $S_t = \beta'\boldsymbol{X}_t + g(R_t) + \rho\nu_t + \eta_t$ which we can then re-write and then estimate, as per Robinson's methodology,

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

D. Auxiliary Rollover Model Estimates

Estimates in Table A are auxiliary to those found in order to evaluate whether this reduced form model is preferable 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: $\mathrm{H}_0\colon \mathrm{M}_1$ superior to M_2 and $\mathrm{H}_0\colon \mathrm{M}_1$ 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 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. 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, both non-nested testing of both OLS and instrumented estimates are inconclusive. The rejection of both 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. This encourages proceeding with a semi-parametric estimation routine.

Table A2 in which demand for lotto tickets dependent on a function that is cubic in rollover size.

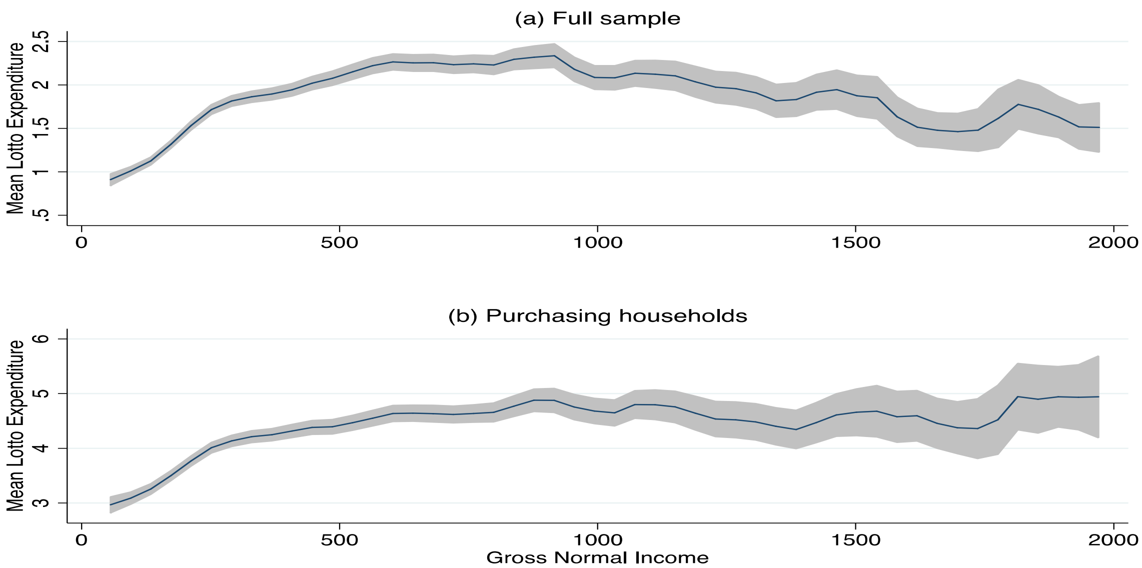
Table A2: Heckman-instrumented estimates of linear and quadratic rollover induced demand

Dependent Variable	Heckman First-Stage 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.

E . Lottery spending data in Living Standards and Food Survey

Figure A3: Local polynomial fit (95% CI) of weekly lotto expenditure versus gross income



Note: Source Wheeler (2018). Data is trimmed to exclude the upper and lower 1% of households by gross normal income