Institute for Fiscal Studies

Tax design in the alcohol market

IFS Working Paper W17/28

Rachel Griffith
Martin O'Connell
Kate Smith

E·S·R·C
ECONOMIC & SOCIAL RESEARCH COUNCIL

Tax design in the alcohol market

Rachel Griffith, Martin O'Connell and Kate Smith*

September 8, 2017

Abstract

We study optimal corrective taxation in the alcohol market. Consumption generates negative externalities that are non-linear in the total amount of alcohol consumed. If tastes for products are heterogeneous and correlated with marginal externalities, then varying tax rates on different products can lead to welfare gains. We study this problem in an optimal tax framework and empirically for the UK alcohol market. Welfare gains from optimally varying rates are higher the more concentrated externalities are amongst heavy drinkers. A sufficient statistics approach is informative about the direction of reform, but not about optimal rates when externalities are highly concentrated.

Keywords: externality, corrective taxes, alcohol
JEL classification: D12, D62, H21, H23
Acknowledgements: The authors gratefully acknowledge financial support from the European Research Council (ERC) under ERC-2015-AdG-694822, the Economic and Social Research Council (ESRC) under the Centre for the Microeconomic Analysis of Public Policy (CPP), RES-544-28-0001, and the Open Research Area, ES/N011562/1, and the British Academy under pf160093. Data supplied by TNS UK Limited. The use of TNS UK Ltd. data in this work does not imply the endorsement of TNS UK Ltd. in relation to the interpretation or analysis of the data. All errors and omissions remained the responsibility of the authors.

*Griffith is at the Institute for Fiscal Studies and University of Manchester, O'Connell is at the Institute for Fiscal Studies and Smith is at the Institute for Fiscal Studies and University College London. **Correspondence:** rgriffith@ifs.org.uk, martin_o@ifs.org.uk and kate_s@ifs.org.uk.

1 Introduction

Alcohol consumption is associated with costs to society from anti-social behaviour, crime and public costs of policing and health care. These externalities are non-linear in alcohol consumption, with a small number of heavy drinkers creating the majority of the costs. Governments attempt to reduce problematic alcohol consumption through restricting availability (Seim and Waldfogel (2013) provide a recent analysis) and with policies that aim to increase prices. We study the design of alcohol taxes.

In this paper we make two contributions. First, we characterise optimal corrective taxes in the alcohol market. If consumers' tastes for different alcohol products and their price responsiveness are correlated with the marginal externalities that their alcohol consumption creates, then it is optimal to levy different tax rates across products. In general, the optimal tax rates depend on consumers' substitution across sets of products. We derive approximations that use a small number of sufficient statistics by placing restrictions on switching patterns; this approach is similar to that used for labour taxes (e.g. Saez (2001)), social insurance (e.g. Chetty (2008)), and more recently applied to environmental policy (e.g. Jacobsen et al. (2016)). Second, we empirically implement our approach in the UK market for alcoholic beverages. We solve for optimal tax rates and show how the welfare gains from varying tax rates across different types of alcohol depend on how concentrated the creation of alcohol externalities are among heavy drinkers.

In our optimal tax framework consumers have heterogeneous demands for the differentiated products in the market. Each product contains a different quantity of ethanol (pure alcohol), as well as other characteristics. A consumer's total ethanol demand is derived from their product level demands. Consumption of ethanol is associated with negative externalities and the marginal externality that each consumer generates may be heterogeneous. A social planner chooses tax rates to maximise the sum of consumer surplus and tax revenue minus the external costs of consumption. If consumers' demands for different alcohol products are correlated with their marginal externality then it is optimal to levy different tax rates on the ethanol in different types of products. The planner can target the most socially harmful drinking by taxing more heavily the ethanol in products that are both disproportionally consumed by problem drinkers and for which an increase in price will lead to a relatively strong reduction in their total ethanol consumption. The welfare gain from moving from a single ethanol tax rate applied to all alcohol products to rates that vary across products depends crucially on the degree of heterogeneity in demands and externalities and the correlation between the two.

The optimal tax rates depend on the full set of own and cross price elasticities across products and how they vary with the marginal externalities that consumers create. In our empirical implementation we use a structural model of demand to solve for the optimal rates. A number of papers apply continuous choice demand methods to alcohol, either treating alcohol as a homogeneous composite commodity (see, inter alia, Baltagi and Griffin (1995), Manning et al. (1995)), or estimating demand over a set of broad alcohol types (e.g. Irvine and Sims (1993), Crawford et al. (1999)). In contrast to these papers, we are interested in capturing substitution patterns between differentiated alcohol products and the correlation in demands for different types of alcohol with how heavily people drink (which is a crucial driver of the size of externalities from an additional drink). This is central to our application because switching patterns and correlations with externalities are precisely what drive the potential for variation in tax rates across different forms of ethanol to improve on a single ethanol tax rate. We use a discrete choice framework that embeds the decision over whether to buy alcohol, what product to buy and in what size.

We estimate the model using longitudinal data on a panel of British households' alcohol purchases. These data contain repeated observations per household and accurate price and product information for disaggregate products. Consistently heavy drinkers (i.e. those with high total ethanol demands) systematically purchase a different mix of products than lighter drinkers; on average, they buy stronger and cheaper alcoholic beverages. We find they are much more willing to switch between different alcohol products in response to price changes, and less willing to switch away from alcohol altogether than lighter drinkers.

Alcohol markets are a natural setting in which to study optimal corrective taxes. The social costs of alcohol consumption are of concern across the developed world (World Health Organization (2014)). Negative consumption externalities associated with alcohol include: public healthcare costs, violent behaviour (e.g. Luca et al. (2015)), drink driving (e.g. Ruhm (1996), Jackson and Owens (2011), Hansen (2015)) and negative impacts on prenatally exposed children (Nilsson (2017)). There is considerable evidence that these externalities are non-linear in ethanol consumption. For example, in the US frequent binge drinkers represent 7% of the population, but drink 45% of the ethanol consumed by adults (US Department of Justice (2005)), and binge drinking accounts for roughly three quarters of the cost of excessive alcohol use (Centers for Disease Control and Prevention

(2016)). Despite the evidence that external costs are convex in alcohol consumption (and hence, at the margin, heavy drinkers tend to create much larger externalities than lighter drinkers), there is uncertainty about the degree of this convexity. We calibrate the mapping from alcohol demands into external costs as a weakly convex function of households' total ethanol demand and we show how our empirical optimal tax results vary across different degrees of convexity of this relationship.

We consider two alternative tax systems. The first is a single tax rate levied on ethanol (the consumption of which maps directly into externalities). The second is a multi rate system in which the planner can vary the tax rate levied on ethanol across a set of alcohol types (based on market segment – spirits, wine, beer or cider – and alcohol strength). If externalities are linear in ethanol and the same across people, then a single ethanol tax rate can achieve the first best; there are no welfare gains from moving to a multi rate system. However, the more convex is the externality function, the larger are welfare gains from being able to set different tax rates across different forms of ethanol. This is because the higher the degree of convexity, the larger the share of externalities are generated by the heaviest drinkers. This enables the planner to target the multi rate tax system more specifically on lowering the ethanol intake of this narrow set of households. By levying a relatively high tax rate on strong spirits the planner is able to target a larger share of the alcohol purchases of heavy than light drinkers, and is able to encourage them to switch to less strong alcohol products, hence lowering their level of ethanol consumption. The size of welfare gains from this additional flexibility depend on how concentrated externalities are among the heaviest drinkers – if, for instance, the 18% of households that purchase the most ethanol account for 95% of the external costs of drinking, the welfare gain from setting different rates is around £400 million.

Much of the recent literature in public economics has made use of sufficient statistics. Sufficient statistics can be informative about optimal tax policy and have the advantage that they can be estimated by reduced form methods, sidestepping the need to estimate a structural model (for discussion, see Chetty (2009)). We show that the first order conditions for the optimal tax rates can be expressed as functions of a small number of sufficient statistics, which capture how much consumers reduce their ethanol demand in response to tax changes and how the size of this reduction varies with the social costs of different consumers' alcohol consumption. If the policy maker is restricted to setting one tax rate on ethanol then these statistics are sufficient to determine whether the current rate is optimal. However, if the policy maker can set multiple tax rates (for example on wine, beer, spirits,...) then this approach requires us to place restrictions on patterns of consumer switching between products. We also derive sufficient statistic expressions for the optimal tax rates, under the additional restriction that these sufficient statistics are constant over non marginal tax changes.

We use our structural demand model to evaluate the sufficient statistics so that we can focus on the role played by the specific restrictions under which they are derived. We show that for both the single and multi rate tax systems, and for any degree of convexity of the externality function, the first order conditions evaluated using the sufficient statistics correctly tell us which direction current taxes should change to move toward the optimal. When the externality function is mildly convex in ethanol, the sufficient statistics expressions do a very good job of recovering the optimal single ethanol tax rate and do a reasonable job of approximating the optimal multi rate system. The poorer performance in the latter case is due to the patterns of cross product switching that are missed by the sufficient statistics approach. However, when the externality function is more convex, the expressions are less informative about the optimal rates.

Our work relates to several strands of the public finance literature. We focus on the ability of alcohol taxes to correct externalities. A number of papers have considered how public goods should be funded, or externalities corrected, when the government's revenue constraint must be satisfied using distortionary taxation (see Bovenberg and Goulder (1996) and Bovenberg and Goulder (2002) for a survey). Kaplow (2012) shows that, under weak separability of leisure from consumption, the income tax system can be adjusted to perfectly off-set any redistributional effects of corrective taxes. Akerlof (1978) notes the potential value of using observable characteristics to identify (or "tag") the potentially needy to improve the efficiency of the benefit system; others have applied this to labour income taxation (e.g. Mankiw and Weinzierl (2010), Weinzierl (2011)). We use a related idea based on the correlation in consumers' preferences for different alcohol products and their marginal externality to "tag" consumption that is likely to have high marginal external costs. Other papers consider the desirability of differential commodity taxation for revenue raising purposes, in the presence of non-linear income taxation (e.g. Atkinson and Stiglitz (1976), Cremer et al. (2001), Saez (2002), Laroque (2005), Kaplow (2006)). In contrast to these papers, we focus on the potential of differential taxation of products *within* a single market to improve welfare by correcting the externalities associated with consumption.

The rest of the paper is structured as follows. In the next section, we discuss the design of corrective taxes in markets with heterogeneous consumers and with many products that potentially generate externalities. We are primarily interested in alcohol taxes, but these results also apply in other settings, such as soda taxes. In Section 3 we describe our data on the UK alcohol market. In Section 4 we outline the empirical demand model and present our demand estimates. We use these along with our optimal tax framework to compute optimal tax rates, which we present, along with welfare results, in Section 5. A final section summarises and concludes. Additional details are provided in the Online Appendix.

2 Corrective tax design

2.1 Model set-up

Let $i \in \{1, ..., N\}$ index consumers; each consumer has income y_i. Let $j \in \{1, \ldots, J\}$ index alcohol products, available at post-tax prices $\mathbf{p} = (p_1, \ldots, p_J)'$. Each product contains a vector of characteristics, $\mathbf{x}_j$. One characteristic (i.e. an element of $\mathbf{x}_j$) is the amount of ethanol (pure alcohol) the product contains, denoted z_j. We denote the matrix of all product characteristics $\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_J)'$ and the vector of ethanol contents, $\mathbf{z} = (z_1, \ldots, z_J)'$.

We assume that consumer i's indirect utility is quasi-linear in the numeraire good and is given by

$$V_i(y_i, \mathbf{p}, \mathbf{x}) = \alpha_i y_i + v_i(\mathbf{p}, \mathbf{x}), \tag{2.1}$$

where α_i is the marginal utility of income and $v_i(\mathbf{p}, \mathbf{x})$ is the indirect utility that arises from the alcohol demands for consumer i. We denote the consumer's demand for product j by $q_{ij} = f_{ij}(\mathbf{p}, \mathbf{x})$ and the consumer's vector of demands by $\mathbf{q}_i = (q_{i1}, \ldots, q_{iJ})'$. Quasi-linear utility means alcohol demands do not depend directly on income; however heterogeneity in preferences (including the marginal utility of income) allows for demand functions to vary flexibly across consumers.

It is widely accepted that alcohol consumption can generate costs that are not considered by individuals when making alcohol consumption decisions, which justifies government intervention. We assume that the external cost associated with an individual's alcohol consumption is given by $\phi_i(Z_i)$, where $Z_i = \sum_j z_j q_{ij}$ denotes individual's i's total ethanol demand from all the products in the market. An implication of this form of external cost function is that, conditional on total ethanol demand, the marginal externality from drinking a unit of ethanol is the same across different types of alcohol. In this section we impose no further restriction on the shape of $\phi_i(\cdot)$, other than that it is continuous and differentiable. The total external cost from all consumers in the market is then:

$$\Phi = \sum_i \phi_i (Z_i) . \tag{2.2}$$

Consumers ignore the externality when making their choices, and the goal of the social planner is to use taxes to induce consumers to internalise the externality, while minimising the reduction in consumer surplus that arises due to the higher prices. We consider the social planner's problem of choosing alcohol taxes to maximise the sum of consumers' indirect utilities (given by equation 2.1) plus revenue raised from tax, R, minus the total external costs of consumption (given by equation 2.2).

We make two important assumptions about the planner's problem. First, we write the objective function in money metric form. This means we abstract from any questions of redistribution, focusing exclusively on the design of taxes to correct externalities. Kaplow (2012) shows that by accompanying externality correcting taxes with a distribution-neutral adjustment to the income tax system, we can offset the effects of the corrective taxes across the income distribution. However, it is harder to offset the impact of the taxes *within* income classes, if consumers with the same income have different preferences (Kaplow (1996)).

Second, we abstract from issues of market power by assuming that taxes are fully passed through to consumer prices and there is no producer surplus term in the planner's problem. This is because we want to focus on the role the tax system plays in correcting externalities and also because the UK grocery (including alcohol) market is very competitive by international standards. Recent papers by Miravete et al. (2016), Conlon and Rao (2015) and Seim and Waldfogel (2013) complement our work by studying the supply side of the US market for spirits. These papers consider how government regulations, such as post and hold regulations in Connecticut and the public monopoly in Pennsylvania, which are designed in part to limit alcohol consumption, interact with firm conduct. The UK alcohol market neither has public monopolies nor post and hold style regulations.

We consider specific (or unit) taxes levied on ethanol content. Let τ denote the tax rate applied to the ethanol content in product j; the post-tax price of product j is therefore $p_j = \tilde{p}_j + \tau z_j$, where $\tilde{p}_j$ is the pre-tax price of product j. We assume that non-price product characteristics do not change as a result of the tax.

Let $\boldsymbol{\tau}$ denote a vector of tax rates levied per unit of ethanol. We write indirect utility, tax revenue and the externality function directly as functions of $\boldsymbol{\tau}$. The social welfare function is:

$$W(\boldsymbol{\tau}) = \sum_i \left[y_i + \frac{v_i(\boldsymbol{\tau})}{\alpha_i} \right] + R(\boldsymbol{\tau}) - \Phi(\boldsymbol{\tau}). \tag{2.3}$$

2.2 Characterising tax policy

Optimal policy

First, it is clear that if the planner can set consumer specific tax rates, then the first best can be achieved by setting $\tau_i^* = \phi_i'(Z_i(\tau_i^*))$ for each consumer i. This is simply the Pigouvian result that the optimal consumer specific rate is set equal to the consumer's marginal consumption externality at that tax rate. However, in practice, setting consumer specific rates is infeasible for governments.

We consider optimal tax rates that are constrained to be the same across consumers. Let $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_K)'$ denote a set of tax rates. Tax rate τ_k applies to the set of products $\mathcal{K}_k$ (we refer to this as set k). $K \leq J$; $K = 1$ corresponds to a single rate ethanol tax, $K = J$ corresponds to tax rates that vary across all products and $K < J$ captures intermediate cases; most tax systems levy different tax rates on spirits, wine, beer etc.

Let $Z_{ik} = \sum_{j \in \mathcal{K}_k} q_{ij}(\boldsymbol{\tau}) z_j$ denote consumer i's ethanol demand from the products belonging to set k and $\frac{\partial Z_{ik}}{\partial \tau_l} = \sum_{j \in \mathcal{K}_k} \frac{\partial q_{ij}}{\partial \tau_l} z_j$ denote the derivative of ethanol demand from set k with respect to a change in the tax rate that applies to products in set l. Tax revenue in this case is given by $R(\boldsymbol{\tau}) = \sum_k \left(\tau_k \sum_i Z_{ik} \right)$.

Taking the derivative of the planner's problem (equation 2.3) with respect to tax rate τ_l and applying Roy's identity yields:

$$\frac{\partial W}{\partial \tau_l} = \sum_i \sum_k (\tau_k - \phi_i') \frac{\partial Z_{ik}}{\partial \tau_l}, \tag{2.4}$$

where $\phi_i' \equiv \phi_i'(Z_i)$. The optimal set of tax rates $\boldsymbol{\tau}^*$ are implicitly defined by setting the first order conditions to zero (equation 2.4 for $l = 1, ..., K$). In general, $\boldsymbol{\tau}^*$ depends on the full set of substitution patterns between the different sets of products and their correlation with the marginal externalities.

Role of consumer heterogeneity

In general, when externalities vary across consumers, setting rates that vary across sets of products improves welfare, relative to a single tax rate. Specifically, this is the case if demand for different types of alcohol are correlated with the marginal externalities that an individual's alcohol consumption creates (i.e. as long as it is not the case that $\text{cov}(\phi'_i, \frac{\partial Z_{ik}}{\partial \tau_l}) = 0 \; \forall \; (k,l))$. There are three obvious cases when these covariances are zero: (i) there is no heterogeneity in externalities, so $\phi'_i = \phi'$; (ii) there is no heterogeneity in demands, so $Z_{ik} = \bar{Z}_k \; \forall \; k$; or (iii) the heterogeneity in externalities and demands are uncorrelated. Under (i) all tax rates are set equal to the marginal externality, $\tau^*_k = \phi'$, and the first best is achieved; under (ii) and (iii) all tax rates are set equal to the average marginal externality, $\tau^*_k = \bar{\phi}' \equiv \frac{1}{N}\sum_i \phi'_i$, but the first best is not achieved.

When there is correlated heterogeneity in marginal externalities and demands, the optimal tax rate on a group of alcohol products is increasing in how popular the products are with individuals that generate large marginal externalities and it is increasing in how strongly those consumers reduce their ethanol demand in response to an increase in the tax rate. As we show in Section 5, the welfare gain due to moving from a single tax rate to tax rates that vary across different alcohol types depends crucially on the degree of heterogeneity in demand, externalities and their relationship.

A sufficient statistics approach

Recent work in the public economics literature has combined the advantages of reduced form strategies – transparent and credible identification – with the ability of structural models to make statements about welfare. This "sufficient statistics" approach has been widely applied to income taxation (see Feldstein (1999), Saez (2001), Gruber and Saez (2002), among many others), social insurance (e.g. Baily (1978), Gruber (1997), Chetty (2006)), and is becoming increasingly popular in other applications, see, for example, Jacobsen et al. (2016) for applications to externality correcting policies.

In our set up, the optimal tax rates are implicitly defined by a set of equations (equation 2.4 for $l = 1, ..., K$) and depend on the full matrix of own and cross tax effects. There are no closed form solutions for the tax rates, and so to recover τ^* we have to solve this system of non-linear equations, fully accounting for switching patterns and non-linearities in demands and externalities. To do this necessarily entails placing structure on the problem, and is relatively data intensive and computationally demanding. An alternative approach is to follow the sufficient statistics literature and express features of optimal taxes in terms of fewer objects that can be estimated using reduced form methods. This can obviate the need to estimate a structural model.

We can significantly simplify the expressions for optimal tax rates by placing restrictions on switching patterns across products. For example, if we were considering a separate wine and beer tax we might be willing to rule out switching between wine and beer products, i.e. assume that, for all consumers i, there are zero cross price effects across products in different sets, $\frac{\partial Z_{ik}}{\partial \tau_l} = 0$ for all $k \neq l$, while still allowing substitution within wine and within beer products. Jacobsen et al. (2016) make use of similar restrictions to derive sufficient statistics for evaluating the efficiency of environmental taxes. They assume that errors in tax rates – the difference between an existing tax rate and the Pigovian first best – are uncorrelated with own and cross demand slopes. This plays a role similar to assuming $\frac{\partial Z_{ik}}{\partial \tau_l} = 0$ for all $k \neq l$ in our set up.

In this case, the impact of a marginal change in τ_k on social welfare can be written:

$$\frac{\partial W}{\partial \tau_k} = N\bar{Z}'_k \left(\tau_k - \bar{\phi}' - \frac{\text{cov}(\phi'_i, Z'_{ik})}{\bar{Z}'_k} \right) \tag{2.5}$$

where $\bar{Z}'_k = \frac{1}{N}\sum_i \frac{\partial Z_{ik}}{\partial \tau_k}$ is the average own tax slope of ethanol demand for the product set k, $\bar{\phi}' = \frac{1}{N}\sum_i \phi'_i$ is the average marginal externality across consumers, and $\text{cov}(\phi'_i, Z'_{ik})$ denotes the covariance in the slope of ethanol demand and marginal externalities across consumers.

If the assumption of no cross price effects between sets of products holds, then evaluating equation (2.5) at the observed tax rates tells us whether the current level of tax applied to the product set k equals the optimal rate, and, if not, whether the rate should be lowered or raised to move towards the optimum. This condition depends on three "sufficient statistics": (i) the average marginal externality, $\bar{\phi}'$; (ii) the covariance between the marginal externality and own slope of demand for product set k, $\text{cov}(\phi'_i, Z'_{ik})$; and (iii) the average own slope of demand for set k, $\bar{Z}'_k$. If, say, there is evidence that marginal externalities are increasing in alcohol consumption, one could use variation in tax rates to estimate how sensitive demand for product set k is to tax changes and how this varies across light to heavy drinkers.

A special case of this formulation is that of a single tax rate applied to all products i.e. $K = 1$, with $\tau_k = \tau$. In this case, the assumption that there are no cross price effects between sets of products imposes no restrictions, since all products are contained in one single set. The sufficient statistics for this single tax rate are the same as above, but where the relevant own slope of demand is for the sum of ethanol from all products in the market, rather than a subset. In this case, optimal tax policy corresponds to that derived in Diamond (1973).

If the derivatives $\bar{\phi}'$ and Z'_{ik} are constant between the observed tax system and the optimal one (and under the assumption of zero cross price effects between sets of alcohol products), then the optimal tax rate for product set k can be expressed:

$$\tau_k^{**} = \bar{\phi}' + \frac{\text{cov}(\phi_i', Z_{ik}')}{\bar{Z}_k'} \tag{2.6}$$

The performance of this expression as an approximation to the optimal tax rates depends on the importance of cross price effects and the degree of non-linearity in demands and externalities. It is likely that the approximation will work best for a single rate ethanol tax (which entails no restrictions on cross price effects) and if the movement to the optimal rate entails only small changes in taxes from current rates.

Empirical implementation

In the following sections we develop a model of consumer demand in the alcohol market, allowing for heterogeneity in demand patterns. We couple this with a mapping between consumers' total ethanol demands and the externalities that their ethanol consumption creates, using evidence from government and medical sources. Our structural demand model enables us to compute optimal taxes according to equations 2.4. We show how varying the shape of the externality function, and hence the relationship between alcohol demands and marginal externalities, affects the welfare gain from moving from a single alcohol tax rates to multiple rates.

We are also interested in how much we can learn from the sufficient statistics about optimal alcohol taxes. We therefore use our empirical demand model to compare the optimal tax rates with the sufficient statistics expressions and present evidence on how well they perform. This provides a useful guide for policymakers who may not have a structural demand model at their disposal, but who may have some evidence on the sufficient statistics embedded in the approximate formula.

3 Data

We use data from the Kantar Worldpanel, which contain rich product information, repeated observations for each household, and accurate prices. Each participating household uses a hand held scanner to record all grocery products, at the UPC level, that are purchased and brought into the home. The data include details of transaction prices, product size, alcohol type and strength.[1] This type of data are becoming increasingly widely used in research (for example, see Aguiar and

[1]Strength is measured as percentage of alcohol-by-volume (ABV). This is defined as the number of millilitres of pure ethanol present in 100ml of solution at 20°C.

Hurst (2007) and Dubois et al. (2014)). For a detailed description of the data, see Griffith and O'Connell (2009) and Leicester and Oldfield (2009); Griffith et al. (2013) contains information on the alcohol segment of the data.

Our data have two substantial advantages over other data sources, such as cross sectional expenditure surveys (e.g. the Living Costs and Food Survey (LCFS) and the Consumer Expenditure Survey (CEX)) and intake diaries (e.g. the Health Survey for England (HSE) and National Health and Nutrition Survey (NHANES)). First, our data track households for a long period of time, meaning we can measure households' long run average alcohol purchases. Second, our data contain very detailed information on purchases of alcohol products, including transaction prices and alcohol contents. A drawback of our data is that they do not include purchases of on-trade alcohol (those made in restaurants and bars). Our data covers the 77% of purchases of alcohol that are made off-trade in supermarkets and liquor store (calculated using the LCFS). In Online Appendix A we show that the distribution of alcohol purchases from our data matches well with other data sources. We also show that the patterns of alcohol purchases are similar for both off and on-trade alcohol purchases in the UK.

3.1 Households

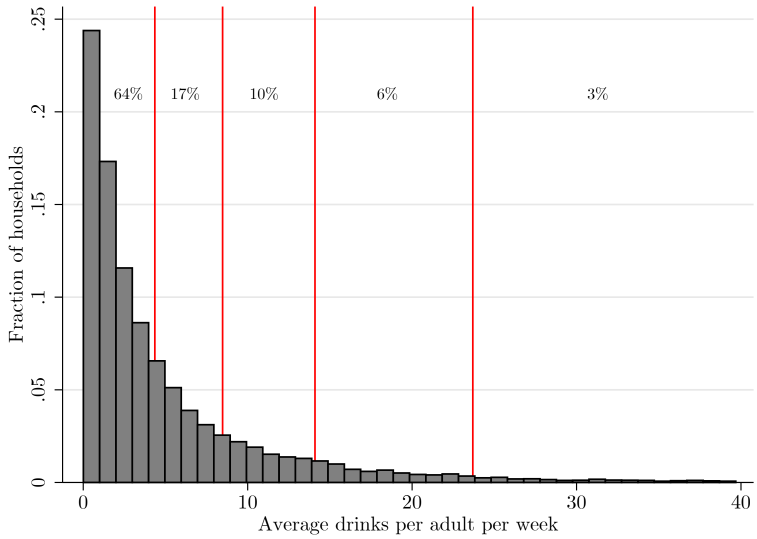
We use a sample (representative of the British population) of 11,634 households, which we observe buying alcohol in 2010 and 2011. We observe households for a minimum of 20 weeks in 2011, and for around 40 weeks per year, on average. We use the 2011 data to estimate demand for alcohol products and the 2010 data to group households based on how much alcohol they buy in this pre-sample period.

Conventions for measuring ethanol volume vary across countries. The US uses "standard drinks"; a standard drink contains 17.7 ml of ethanol. The UK, and many other European countries, use "units"; a unit contains 10 ml of ethanol. For each household we calculate the number of standard drinks that they purchase per adult household member in each week that we observe them in 2010. We take the average for each household across weeks to construct the household's average ethanol purchases in 2010. We observe each household for an average of 40 weeks in 2010, which means we measure whether households are consistently heavy drinkers.

Figure 3.1 plots the distribution of average drinks per adult per week across households. We refer to this as the distribution of ethanol purchases. We use this measure to group households into five quintiles, with each quintile accounting for 20% of all drinks purchased. 64% of households are in the first, or bottom, quintile: the lightest 64% of alcohol consumers account for 20% of all drinks bought. The fifth, or top, quintile accounts for 20% of drinks, but contains only 3% of households. We use these quintiles as conditioning variables in our demand estimation.

High average drinks per adult per week over a long period of time may be due to consumers drinking large amounts regularly or engaging in less regular very high consumption (binge drinking). Both types of drinking behaviour can lead to externalities, although the nature of these externalities may differ. In Online Appendix A we show that in both the UK and the US people who report consuming more ethanol also report drinking more days per week and are more likely to have reported binge drinking in the previous week.

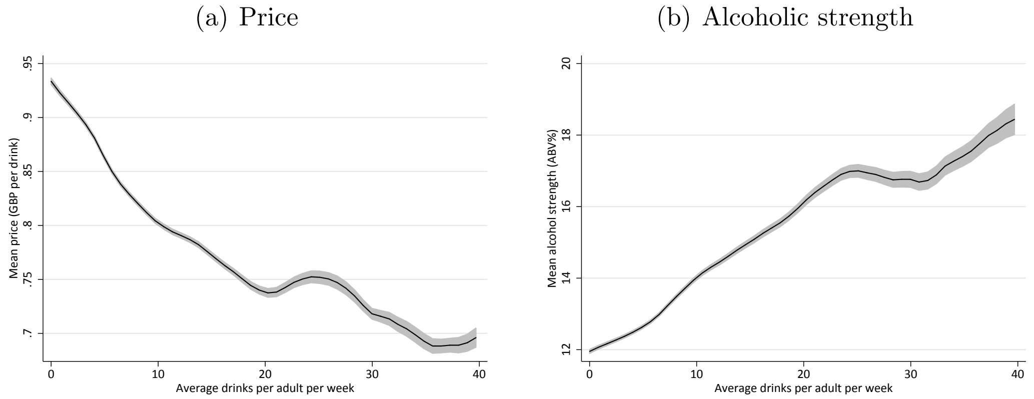
Figure 3.1: *Distribution of drinkers*



Notes: Distribution drawn across 11,634 households. The red lines mark the cutoffs between drinking quintiles, which each constitute 20% of total drinks purchased; these are located at: 4.4, 8.5, 14.1, 23.7 drinks per adult per week. The numbers show the percentage of households in each quintile. Numbers are based on the pre-sample period.

Drinking patterns vary substantially across the distribution of drinkers. If heavy drinkers generate larger marginal externalities than moderate drinkers, this is likely to lead to considerable welfare gains from having tax rates that vary across different types of alcohol. This variation in drinking patterns is apparent in Figure 3.2, which plots the relationship across households between mean average alcoholic strength (left hand panel) and price of products purchased (right hand panel) with the average number of drinks purchased per adult per week. Heavier drinkers tend to purchase stronger types of alcohol, on average. This is both because the heaviest drinking households buy proportionately more spirits, and less beer, than lighter drinkers, and they also buy stronger products within these broad categories. The heaviest drinkers also buy products that are cheaper in per-drink terms. This suggests that a tax system that increases the relative prices of strong and cheap products may successfully target the consumption of heavy drinkers. Whether this is indeed the case depends on: (i) how strongly different households (e.g. light versus heavy drinkers) switch away from the products in response to a tax rise; (ii) how strongly and to what alternative alcohols they switch; and (iii) what fraction of drinking externalities are accounted for by the set of heavy drinkers.

Figure 3.2: *Average alcoholic strength and price of products purchased, across distribution of drinkers*



Notes: For each household-week in 2011 we calculate the average price per drink and alcoholic strength. The lines plot the relationship between these variables and the average number of drinks purchased per adult per week, measured in the pre-sample period, for each household. The grey lines are 95% confidence intervals.

3.2 Products

In excess of 7000 alcohol UPCs (barcodes) and 3000 brands are recorded as being purchased in our data. Estimating a choice model with 7000 UPCs is likely to be neither feasible nor informative. We aggregate UPCs into 32 "products" to focus on the margins of substitution that are most relevant to our application, see Table 3.1. It is important that we capture heterogeneity in the shape of demand for sets of UPCs that are impacted similarly by alcohol tax changes and it is also important that we capture how changes in taxes and hence prices affect the total quantity of alcohol that households purchase. We are therefore careful not to aggregate over UPCs that have different alcohol strengths or UPCs likely to be subject to different tax treatment and, as much as possible, only aggregate across UPCs that are of a similar alcohol type, quality and price.

Table 3.1: *Product definition and characteristics*

(1) Product definition	(2) Top brand and within-product share (%)	(3) No. brands	(4) Mean ABV	(5) Market share (%)	(6) No. sizes
Beer					
Premium beer; ABV < 5%	Newcastle Brown Ale (6.1)	386	4.4	1.8	3
Premium beer; ABV ≥ 5%	Old Speckled Hen (16.5)	238	5.5	2.1	3
Mid-range bottled beer	Budweiser Lager (19.6)	94	4.7	4.6	3
Mid-range canned beer; ABV < 4.5%	Carlsberg Lager (28.8)	17	3.9	5.8	3
Mid-range canned beer; ABV ≥ 4.5%	Stella Artois Lager (72.0)	15	5.0	2.7	3
Budget beer	John Smiths Bitter (23.6)	72	4.2	3.2	3
Wine					
Red wine	Tesco Wine (6.2)	439	12.6	18.4	4
White wine	Tesco Red Wine (7.8)	327	12.1	17.1	4
Rose wine	Echo Falls Wine (8.6)	67	11.5	4.2	2
Sparkling wine	Lambrini Sparkling Wine (8.4)	125	9.2	3.1	2
Champagne	Lanson Champagne (12.7)	42	11.8	0.8	1
Port	Dows Port (22.0)	23	19.8	0.7	1
Sherry	Harveys Bristol Cream (18.7)	25	16.8	1.2	1
Vermouth	Martini Extra Dry (11.8)	33	15.0	0.6	1
Other fortified wines	Tesco Fortified Wine (21.8)	37	14.6	0.9	1
Spirits					
Premium gin	Gordons Gin (59.6)	21	38.3	1.6	2
Budget gin	Tesco Gin (22.3)	15	38.3	1.3	2
Premium vodka	Smirnoff Red Vodka (39.0)	54	37.6	3.1	2
Budget vodka	Tesco Vodka (31.4)	17	37.5	1.8	2
Premium whiskey	Jack Daniels Bourbon/Rye (19.6)	80	40.5	2.1	2
Budget whiskey	Bells Scotch Whiskey (18.7)	56	40.0	8.1	2
Liqueurs; ABV <30%	Baileys (25.9)	203	18.4	3.1	2
Liqueurs; ABV ≥30%	Southern Comfort (27.2)	41	37.0	0.8	2
Brandy	Tesco Brandy (22.1)	55	37.3	2.4	2
Rum	Bacardi White Rum (29.1)	58	37.1	2.0	2
Pre-mixed spirits	Gordons Gin+Tonic (14.7)	43	6.1	0.2	1
Alcopops	Smirnoff Ice Vodka Mix (17.3)	147	4.8	0.8	1
Cider					
Apple cider, <5% ABV	Magners Original Cider (26.9)	52	4.4	1.6	3
Apple cider, 5-6% ABV	Strongbow Cider (63.1)	49	5.3	2.0	3
Apple cider, >6% ABV	Scrumpy Jack Cider (18.7)	71	7.0	0.8	2
Pear cider	Bulmers Pear Cider (24.2)	33	4.9	0.7	2
Fruit cider	Jacques Fruit Cider (21.4)	48	4.4	0.5	2

Notes: Column (1) shows the product definition. Column (2) lists the brand that constitutes the largest share of spending within each product; its within-product expenditure share is shown in parentheses. Column (3) lists the number of brands within each product. Column (4) shows the mean alcoholic strength (ABV) of each product. Column (5) shows the share of the alcohol market accounted for by each product. Column (6) shows the number of bins used to divide the quantity distribution.

Table 3.1 shows that for many of our 32 products, one brand constitutes the majority of the spending on that product. However, other products consist of many smaller brands, for example, wine and premium bottled beer. We consider tax systems that vary tax rates across different types and strengths of alcohol. The most important consumer substitution resulting from changes in tax systems such as these is between different alcohol types and strengths. This switching is well captured by our 32 products. If we were interested in alcohol tax systems that set different rates for, say, Spitfire Kentish Ale and Badger Golden Glory – two different brands of premium beers, each with 4.5% ABV – then it would be important to model substitution within low strength premium beers.

To model the quantity of a product that households choose, we discretize the quantity distribution into a set of equally sized categories for each product. We allow the number of size categories to vary across products depending on how dispersed the quantity distribution is – in Table 3.1 we list the number of categories we use for each product. For example, for red wine, which constitutes around 18% of total alcohol spending, we define 4 categories – 1 bottle, 2 bottles, 3 bottles and more than 3 bottles. In total there are 69 product-sizes. In Online Appendix A we plot the distributions of quantity for each product, we also show that the distribution of drinks purchased per household-week computed using our discrete size categories very closely matches the distribution observed directly in the data.

3.3 Prices

For each product-size we compute a price index that we use in our model. The index captures price movements of the underlying UPCs that comprise the product. We compute a weighted average of the UPC prices using weights that are fixed over time.

Let b index UPC (or barcode), j index product, s index size, r index region, t index time, and f index retailer. The barcode b is sold at price ρ_{bft} in retailer f at time t. In the UK the main retailers set national prices. Let $\mathcal{B}_{js}$ denote the set of barcodes that belong to product j in size s. The region r, time t price index for product j in size s is:

$$p_{jsrt} = \sum_{b \in \mathcal{B}_{is}, f} w_{bfr} \rho_{bft}, \qquad w_{bfr} = \frac{N_{bfr}}{\sum_{b' \in \mathcal{B}_{is}, f'} N_{b'f'r}} \tag{3.1}$$

where N_{bfr} denotes the number of purchases of barcode b from retailer f in region r across the entire time period. The regional dimension to the weights captures geographical variation in retailer coverage.[2]

In practice, we compute 69 product-size price indices which vary over 12 months and 11 regions; this means we observe 9108 price points. In Online Appendix A we report average prices and plot the price series for each of the 69 product-size pairs. There is considerable differential time series variation in price across products. We discuss how this price variation allows us to identify the effect of price changes on demand in Section 4.4.

[2]We also allow the weights to vary across the five drinking quintiles, capturing the possibility that the popularity of UPCs within product-sizes varies across these quintiles. We omit a household quintile index for notational simplicity.

4 Demand model

We specify a model of consumer demand in the alcohol market. The model embeds the decision of whether or not buy alcohol, what product to choose and what quantity. It also incorporates heterogeneity in preferences, allowing for the possibility that the shape of product level demands are correlated with where households are in the distribution of ethanol purchases (and hence what level of externality their drinking is likely to create).[3]

4.1 Empirical demand specification

We model the alcohol purchase a household makes on a "purchase occasion". We define a purchase occasion as a week in which the household is recorded buying groceries. Alcohol is purchased on 54% of purchase occasions. On the remaining purchase occasions households choose the "outside option" of no alcohol. We model the decision over whether to buy alcohol and which option to choose as a discrete choice. A discrete choice demand framework rationalises zero purchases[4] and, due to the mapping of preferences into attribute space, does not suffer from the curse of dimensionality of continuous choice demand models. On 17% of household-week observations, a household purchases more than one (typically two or three) alcohol products. We treat this behaviour as the household undertaking multiple separate purchase occasions. In total we have data on 632,810 purchase occasions.

We index households by i and products by j. $j = 0$ denotes the option of purchasing no alcohol, $j = 1, ..., J$ indexes different alcohol products. Products are available to the consumer in discrete sizes, indexed by s. We model the decision over which product-size, (j, s), to select, with the option to purchase no alcohol denoted $(0, 0)$. We use t to index time (i.e. weeks).[5]

Household preferences are defined over characteristics of products, both observed (Gorman (1980), Lancaster (1971)) and unobserved (Berry (1994), Berry et al. (1995)). We assume that the utility that household i obtains from selecting option

[3]Although we condition the entire preference distribution on pre-sample ethanol consumption, we do not explicitly model state dependence. In Online Appendix B we provide some reduced form evidence that once preference heterogeneity is accounted for state dependence in demand appears not to be of first order importance.

[4]A household typically chooses one or a small number of alcohol products. This means we observe a multiplicity of zero demands. Most continuous choice models are derived under the assumption all demands are strictly positive. Estimation of such models using data with many zero purchases results in serious biases – see Wales and Woodland (1983).

[5]For households that purchase multiple (i.e. two or three) different alcohol products in a weeks, we have multiple observations per week.

(j, s) in period t is given by:

$$u_{ijst} = \nu(p_{jsrt}, \mathbf{x}_{jst}; \boldsymbol{\theta}_i) + \epsilon_{ijst}, \tag{4.1}$$

where p_{jsrt} is the price of option (j, s) in period t and region r, $\mathbf{x}_{jst}$ is a vector of option characteristics (including a time-varying unobserved attribute), and $\boldsymbol{\theta}_i$ is a vector of household level preference parameters. ϵ_{ijst} is an idiosyncratic shock distributed i.i.d. type I extreme value. We normalise the utility from purchasing no alcohol so that $u_{i00t} = \epsilon_{i00t}$.

Households select the option (j, s) that provides them with the highest utility. Integrating across the demand shocks, $\boldsymbol{\epsilon}_{it} = (\epsilon_{i00t}, ..., \epsilon_{iJSt})'$, yields conditional choice probabilities, which describe the probability that household i selects option (j, s) in week t, conditional on prices, product attributes and preferences. At the household level the conditional choice probability for option $j > 0, s > 0$ takes the closed form:

$$q_{ijst} = \frac{\exp(\nu(p_{jsrt}, \mathbf{x}_{jst}; \boldsymbol{\theta}_i))}{1 + \sum_{j'>0,s'>0} \exp(\nu(p_{j's'rt}, \mathbf{x}_{j's't}; \boldsymbol{\theta}_i))} \tag{4.2}$$

and expected utility is given by:

$$v_{it}(\mathbf{p}_{rt}, \mathbf{x}_t) = \ln \sum_{j>0,s>0} \exp\{\nu(p_{jsrt}, \mathbf{x}_{jst}; \boldsymbol{\theta}_i)\} + C \tag{4.3}$$

where C is a constant of integration that differences out when comparisons are made across two different tax regimes. Equations 4.2 and 4.3 give the empirical analogues for q_{ij} and v_i used in Section 2.

We assume the function ν takes the form:

$$\nu(p_{jsrt}, \mathbf{x}_{jst}; \boldsymbol{\theta}_i) = \alpha_i p_{jsrt} + \beta_i w_j + \sum_{m=1}^{4} 1[j \in \mathcal{M}_m] \cdot (\gamma_{im1} z_{js} + \gamma_{im2} z_{js}^2) + \xi_{ijt}. \tag{4.4}$$

We allow the size of the option, measured as the amount of pure alcohol (ethanol) in the option, z, to affect the utility from option (j, s) through a quadratic function with parameters that we allow to vary across the four segments of the alcohol market: beer, wine, spirits and cider – indexed $m = 1, ..., 4$. $\mathcal{M}_m$ denotes the set of options that belong to segment m. This allows for the possibility that households might value larger or smaller quantities of ethanol differently, depending on what type of alcohol they are buying. We also allow the product's alcoholic strength, w_j, and a household specific time varying unobserved product attribute, ξ_{ijt}, to affect the utility from option (j, s).[6]

4.2 Preference distribution

We model preference heterogeneity over observable attributes (price, strength and ethanol content) and over unobservable attributes (denoted by ξ_{ijt}). We use random coefficients to capture unobserved preference heterogeneity. To capture how preferences vary with where consumers are in the distribution of ethanol purchases we condition the random coefficients on the five quintiles defined in Section 3.1.

Observable product attributes

Let $d = 1, \ldots, D$ index the quintiles of households defined in Section 3.1 based on average drinks purchased per adult per week in the pre-sample period – these allocate households into five quintiles from lightest to heaviest drinkers with each quintile comprising 20% of drinks purchased. We model preferences over observed attributes as following a multivariate normal distribution, *conditional on household quintile d.* This means that we allow both the mean and the covariances of the preference parameters to vary across the distribution of drinkers. Specifically, denoting preferences on the first and second order ethanol content terms by $\boldsymbol{\gamma}_{i1} = (\gamma_{i11}, ..., \gamma_{i1M})'$ and $\boldsymbol{\gamma}_{i2} = (\gamma_{i21}, ..., \gamma_{i2M})'$, we assume:

$$\begin{pmatrix} \alpha_i \\ \beta_i \\ \boldsymbol{\gamma}_{i1} \\ \boldsymbol{\gamma}_{i2} \end{pmatrix} \Bigg| d \sim \mathcal{N} \left(\begin{pmatrix} \bar{\alpha}^d \\ \bar{\beta}^d \\ \bar{\boldsymbol{\gamma}}_{i1}^d \\ \bar{\boldsymbol{\gamma}}_{i2}^d \end{pmatrix}, \begin{pmatrix} \sigma_{\alpha\alpha}^d & \sigma_{\alpha\beta}^d & \sigma_{\alpha\gamma}^d & 0 \\ \sigma_{\alpha\beta}^d & \sigma_{\beta\beta}^d & \sigma_{\beta\gamma}^d & 0 \\ \sigma_{\alpha\gamma}^d & \sigma_{\beta\gamma}^d & \sigma_{\gamma\gamma}^d & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \right) \tag{4.5}$$

Note that mean within quintile preference for alcohol strength, $\bar{\beta}^d$, is not separately identified from the product effects (see below), so we normalise this to zero.

Unobservable product attributes

As we discuss in Section 4.4, the inclusion of unobservable product attributes, ξ_{ijt}, in our demand model is important for consistently estimating price effects. We assume that these unobserved effects can be decomposed into two components:

[6] The alcohol strength of a product is the amount of ethanol it contains per litre of product i.e. $w_j = z_{js}/L_{js}$, where L_{js} is the size in litres of product-size (j, s). It is only necessary to include two of these three variables to capture both preferences over product size and alcoholic strength.

$\xi_{ijt} = \eta_{ij} + \chi_{ikt}$. η_{ij} is a time-invariant product effect and χ_{ikt} is an alcohol type time effect.

The time-invariant product effects capture consumers' preferences over unobserved product attributes that are fixed over time. We allow the product effects to vary across the five household quintiles, capturing the possibility that preferences over unobserved product attributes are correlated with how heavily a household consumes alcohol. We also allow for a random component for the product effects that is common across products within each of the four segments of the market. This allows for the possibility that households' willingness to substitute between products in each of these segments differs from their willingness to switch between products in different segments. Formally, denoting the set of product effects in market segment m by the vector $\boldsymbol{\eta}_{im}$, we assume that $\bar{\boldsymbol{\eta}}_{im}|d \sim \mathcal{N}(\bar{\boldsymbol{\eta}}^d_m, \sigma^d_m)$.[7]

The alcohol type time effects capture variation in preferences for unobserved alcohol attributes over time (due, for instance, to the effects of advertising or seasonal demand patterns). We include these for a set of 8 alcohol types (listed in Table 5.1) and allow them to vary across quarters. We also allow the time effects to vary across households quintiles, so $\chi_{ikt} = \chi^d_{kt}$, allowing for the possibility that the extent of temporal variation in alcohol type demands is different across the total drinking distribution.

4.3 Estimation

For each of the five household quintiles there are a total of 9 parameters governing the means and 9 governing the covariance of the preference distribution for observable product attributes and there are a further 4 parameters governing the variance of the unobserved products effects. In addition there are the common product and alcohol type time effects. We condition the entire preference distribution on the households quintiles, so we can estimate the model separately quintile-by-quintile.

We estimate demand using maximum simulated likelihood. Conditional on the preference draws, the probability a household selects a given option on a given purchase occasion is given by the closed form of equation 4.2. This follows from our assumption that the idiosyncratic utility shocks, ϵ_{ijst}, are i.i.d. type I extreme value. To construct the likelihood function we integrate across the random coefficient distribution. Let $(1, ..., T_i)$ denote the stream of sampled purchase occasions on which we see decisions of household i and let (j^*_t, s^*_t) denote the option the household chooses on purchase occasion t. The contribution household i makes to

[7]In addition, we assume the segment random coefficients, condition on households quintile d are uncorrelated with each other and with the preferences on observable product attributes.

the likelihood function is then:

$$l_i = \ln \int \prod_{t=(1,\ldots,T_i)} q_{ij_t^* s_t^* t} dF(\boldsymbol{\theta})$$

No closed form for this integral exists, so we use simulation methods.

4.4 Identification of demand parameters

We use longitudinal micro data; for each household in our sample we observe many repeated choices. The vector of product prices that households face varies cross sectionally across regions and over time. How households adjust their behaviour in response to these changes aids identification of the preference parameters. A number of papers (e.g. Berry and Haile (2010), Berry et al. (2004)) have highlighted the powerful identifying role that micro data (compared with more commonly used market level data) plays in pinning down parameters in choice models.[8]

We exploit price variation that is driven by supply side factors, which include determinants of marginal cost such as input prices and alcohol tax rates. However, we may be concerned about variation in price driven by demand side factors, such as firms altering prices in response to fluctuations in demand. These sources of price variation are potentially problematic if it leads prices to be correlated with fluctuations in demand that are not controlled for and are therefore collected in the shock term ϵ_{ijst}.

To deal with this, we include a rich set of unobserved characteristics that control for a number of possible sources of price endogeneity arising from demand side price drivers. For instance, our vector of product effects controls for unobserved quality differences across products likely to be correlated with price and our time effects control for seasonality in demand and spikes in demand due to advertising campaigns. In addition, the practice of UK supermarkets of pricing products nationally limits the scope for geographical variation in prices driven by local demand shocks.[9] Nevertheless, we cannot rule out the possibility that there may by some residual omitted demand side variables correlated with prices.

[8] Berry and Haile (2010) and Fox and Gandhi (2016) establish conditions for nonparametric identification of random coefficients in random utility discrete choice models by placing restrictions on the covariate supports. Fox et al. (2012) show that the identification conditions are weaker in the case where ϵ_{ijst} shocks are distributed type I extreme value, and that even with cross sectional data the model is always identified if utilities are a function of linear indices with continuously distributed covariates.

[9] The large UK supermarkets, which make up over three quarters of the grocery market, agreed to implement a national pricing policy following the Competition Commission's investigation into supermarket behaviour (Competition Commission (2000)).

We therefore include a control function for price that isolates price variation driven by a set of instruments that we expect to shift firm costs, but not to directly impact on demand (see Blundell and Powell (2004) and, for multinomial discrete choice models, Petrin and Train (2010)). Our instrument set includes producer prices for beer and cider, which are likely to be drivers of the consumer price of beer and cider options. Producer prices are also likely to vary seasonally due to demand fluctuations, but we control for this through the time effect in our demand model. Also included are the sterling-euro and sterling-dollar exchange rates, which affect the price of imported alcohol, and alcohol duty rates. The main reason for regional variation in prices in the UK is differences in the geographical coverage of food retailers. To capture this we include as instruments the market share of the main retailers in each region. We also include the price of oil interacted with region to capture regional variation in transport costs. In Online Appendix B we describe variation in the instruments.

The F-stat for a test of the (ir)relevance of the instruments is 17.9, meaning we strongly reject the hypothesis of no relationship between price and the instruments. In demand estimation we control for the predicted residuals of the first stage regression. The residuals enter positively and statistically significantly and the price coefficients become more negative when the control function is included. This indicates that the omission of the control function would lead to a (modest) bias towards zero of the price coefficients.

4.5 Demand estimates and elasticities

In Online Appendix B we report the coefficient estimates for our demand model. The coefficients capturing the mean of the price preference distribution for each household quintile are all negative and statistically significant. The parameters capturing the variance of the preferences for price, strength and ethanol for each household quintile all indicate statistically significant within quintile preference heterogeneity. The covariance parameters show that, within each quintile, more price sensitive consumers typically have relatively strong preferences over quantity of ethanol and alcoholic strength. One exception is for the heaviest drinkers, for whom the less price sensitive consumers have stronger preferences for alcoholic strength. In Online Appendix B we also present estimates of the average of the mean product effects within each alcohol segment (relative to the utility from the outside option). The light drinking households in the bottom quintile of the the ethanol purchase distribution have the lowest mean product effects for each segment on average. However, the segment specific variance parameters indicate that, as with the observable product attributes, there is high degree of within quintile preference heterogeneity.

The demand model estimates generate a set of own and cross price elasticities that describe how households switch between all the options (product-sizes) in the market, as well as towards the no purchase outside option, in response to marginal price changes. After integrating across the unobserved preference heterogeneity we obtain a 70×69 matrix of elasticities for each household quintile. In Figure 4.1 we summarise this information.

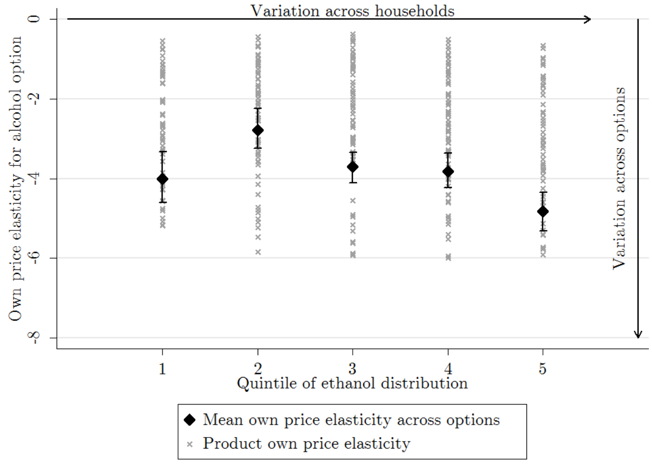
Panel (a) shows the own price elasticities and panel (b) shows the cross price elasticities. The vertical variation in the graphs is across alcohol options and the horizontal variation is across the five household quintiles. The grey dots represent the own price elasticities for each alcohol option (in panel (a)) and the cross price elasticities between pairs of alcohol options (in panel (b)). The black dots represent the mean elasticity for each quintile of households; the bars are 95% confidence intervals.[10] The graph highlights that variation in elasticities across product-sizes is substantial. It also shows some variation in the mean own price elasticity across quintile, with the top quintile, on average, having the most price elastic product level demand. However, the variation in the mean cross price elasticity across the household quintiles is much more striking. The mean cross price elasticity of households in the heaviest drinking top quintile is over 4.5 times as high as the mean for the lightest drinking bottom quintile. The heaviest drinkers are much more likely to respond to an increase in a product's price by switching to alternative products (rather than out of the market). A consequence of this is that when we simulate the overall price elasticity of demand for ethanol (i.e. what is the % change in demand that follows a 1% price increase in all alcohol) households in the top quintile are much less price sensitive; their own price elasticity is -0.95 compared with -2.07 for the bottom quintile.

The responsiveness of households' product demands to price changes is a crucial input into computing the optimal tax rates on different alcohol products. In the next section we show how variation in price responsiveness directly translates into the optimal tax rates.

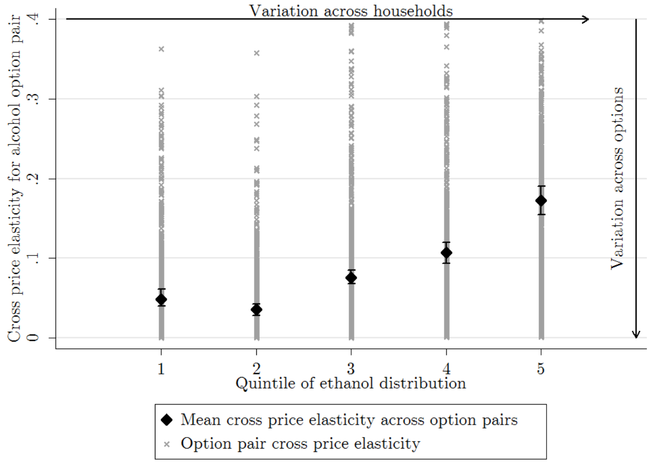
[10]We calculate confidence intervals by obtaining the variance-covariance matrix for the parameter vector estimates using standard asymptotic results. We then take 100 draws of the parameter vector from the joint normal asymptotic distribution of the parameters and, for each draw, compute the statistic of interest, using the resulting distribution across draws to compute Monte Carlo confidence intervals (which need not be symmetric around the statistic estimates).

Figure 4.1: *Summary of own and cross price elasticities*

(a) Own price elasticities



(b) Cross price elasticities



Notes: The grey markers represent alcohol option (product-size) level elasticities, computed separately for households in each of the five quintiles. The black markers are averages across these option level elasticities. Quintiles of the ethanol distribution are defined as follows: households are ranked according to how much ethanol they bought per adult per week in 2010, households that buy the first 20% of all ethanol purchased are in the first quintile, households that buy the next 20% are in the second quintile and so on. The bars are 95% confidence intervals.

5 Optimal alcohol taxes

In this section we combine our estimates of households' alcohol demands with the optimal tax framework from Section 2 to calculate optimal alcohol taxes.

5.1 Externality function

In Section 2 we specified the argument of the consumer level externality function, $\phi_i(.)$, to be total ethanol demand Z_i. Our data have details of alcohol purchases made by households, so we convert total ethanol demand into ethanol demand per adult (person aged 18 or over). We also place some additional structure on the externality function. We assume that ϕ_i is an increasing (weakly) convex function and that its shape does not vary across households i.e. $\phi_i(.) = \phi(.)$ for all i. This means that: (i) the marginal externality of a households' drinking is (weakly) increasing in its level of ethanol consumption, and (ii) differences in marginal externalities across people are driven by differences in their level of ethanol demand.[11]

We parametrise the externality function as quadratic with parameters, (ϕ_0, ϕ_1):

$$\phi(Z_{it}) = \phi_0 Z_{it} + \phi_1 Z_{it}^2 \tag{5.1}$$

(ϕ_0, ϕ_1) jointly determine the aggregate external cost and degree of convexity of the function.

We calibrate the externality function to match the aggregate external cost estimate based on a study by the UK Cabinet Office (2003). Using this study Cnossen (2007) categorises estimates of the various costs associated with alcohol misuse in the UK. The report estimates that the direct tangible social costs are £7.25 billion (in 2011 prices).[12]

There is a large body of evidence that suggests the external costs of drinking are highly concentrated among a small number of heavy drinkers (see Cnossen (2007) for a survey). Relatedly, there is a considerable amount of evidence that externalities from alcohol consumption are convexly increasing in ethanol consumed (and hence the marginal externality associated with an additional drink is increasing in number of drinks consumed). For example, there is evidence of a threshold effect with some diseases: at low levels of ethanol consumption, the risk of disease is not elevated, but this risk increases sharply above a certain point (see Lönnroth et al. (2008) for evidence on tuberculosis, and Rehm et al. (2010) for evidence on liver

[11] It is generally accepted that men generate more externalities from drinking than women. The World Health Organization argues this is principally a consequence of men drinking more rather than creating more externality for a given level of consumption; *"when the number of health and social consequences is considered for a given level of alcohol use or drinking pattern, sex differences for social outcomes reduce significantly"*. (World Health Organization (2014)).

[12] The estimate reported in the paper was £7.5 billion in 2001 prices; we uprate this to 2011 prices using the Retail Price Index (RPI) and scale to account for the fact that we are using data on alcohol purchases excluding those made in restaurants and pubs (off-trade purchases). We assume that the share of external costs generated by off-trade alcohol consumption is proportional to the number of units consumed off-trade (77%).

cirrhosis). Although there is considerable evidence that the external costs of alcohol consumption are convex, there is little evidence on the precise degree of convexity of the relationship. Therefore we remain agnostic about this and show how the optimal tax rates change as we vary the degree of convexity in the relationship from zero (i.e. constant marginal externality) to a high degree of convexity.

Specifically, we group together the 19% of heaviest drinking households that together purchase 60% of total ethanol; these are the households in the top three quintiles of the ethanol distribution. Henceforth, we describe these households as heavy drinkers, and the 81% households in the bottom two quintiles (or those that buy the remaining 40% of ethanol) as light drinkers. If the marginal external costs of drinking are constant, then the heavy drinkers would generate 60% of the external costs (as they buy 60% of the ethanol). As the convexity of the externality function increases, the share of costs generated by the heavy drinkers increases. We calibrate (ϕ_0, ϕ_1) to eight specifications, in which heavy drinkers generate: 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95% of the external costs of drinking. The assumption that the heavy drinkers generate 95% of the externality implies that the marginal externality of somebody who drinks 40 standard drinks a week is 8 times as large as the marginal externality of somebody who drinks 8 standard drinks a week. We plot the externality functions for the 60%, 65%, 80%, 95% in Online Appendix B.[13] The Centers for Disease Control and Prevention (2016) estimate that binge drinkers, who constitute 17% of the population, are responsible for 77% of the costs of excessive alcohol use.

5.2 Optimal tax rates

We consider two optimal tax systems. For the first we consider a situation in which the planner is constrained to set a single tax rate levied per unit of ethanol, common across all products. For the second, we consider the case in which the planner can set multiple tax rates, varying them across eight alcohol types; beer (>5% ABV), beer (≤5% ABV), wine (>14% ABV), wine (≤14% ABV), spirits (>20% ABV), spirits (≤20% ABV), cider (>5% ABV) and cider (≤5% ABV). For the multi rate system we choose to let rates vary across alcohol types in a way that is similar to existing tax systems in many countries. For instance, in the UK rates currently vary across spirits, beer, wine and cider with various bands in the latter three groups based on alcohol strength. However, the UK taxes on wine and cider are based on product volume rather than ethanol and, as we show below, the system is very far

[13]Note we place a lower bound of the externality function of zero. For the most convex specifications this binds at low levels of ethanol consumption.

from optimal. Allowing for tax rates to vary over more disaggregate alcohols (e.g. a different vodka and gin rate) simply magnifies our conclusions about the welfare gains of rate differentiation.

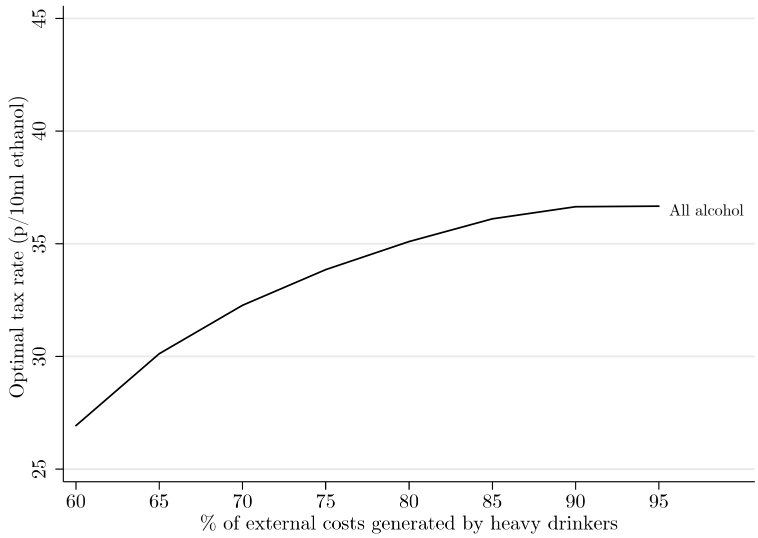
In Figure 5.1 we show how the optimal tax rates vary with how convex the externality function is. When the heavy drinking group account for 60% of the external costs, the marginal externality is constant. As we raise the fraction of externalities accounted for by the heavy drinking group the externality function becomes increasingly convex. Panel (a) shows the optimal single rate and panel (b) shows the optimal multi rate system.

The optimal single rate is 27p/10ml ethanol when marginal drinking externalities are constant, and increases as we vary the externality function to make it increasingly convex (rising to 36p/10ml ethanol when 95% of externalities arise from the set of heavy drinkers). The reason the optimal rate is higher the more convex is the function is that heavy drinkers reduce their ethanol more *in levels* (though not in percent terms) than light drinkers do as a consequence of an increase in alcohol prices. Therefore, the greater the share of aggregate externalities that are accounted for by the heavy drinkers, the more effective is tax at lowering the social costs of alcohol consumption and therefore the higher is the optimal rate.

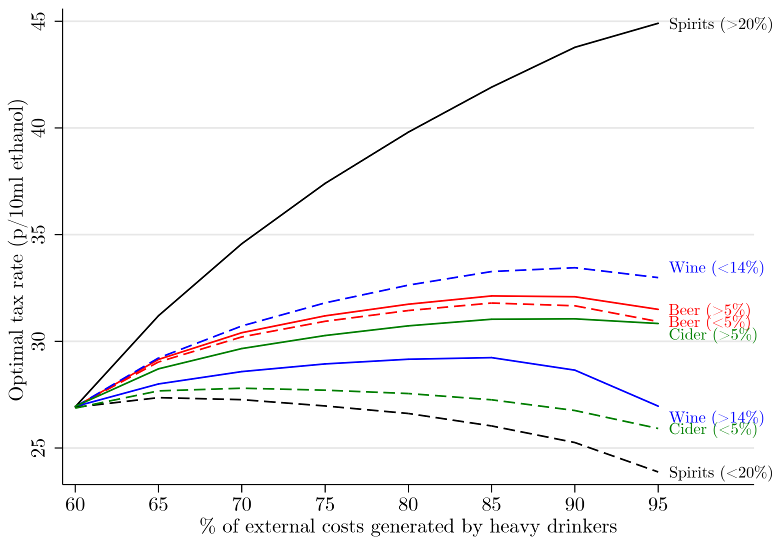
In the case of the multi rate system, all optimal rates are 27p/10ml ethanol when marginal externalities are constant. In this case there is no gain from rate differentiation and the optimal single and multi rate systems coincide. However, as the externality function becomes increasingly convex the optimal rates in the multi rate system diverge. High strength spirits (those with ABV>20%) attract the highest tax rate (over 40p/10ml ethanol when the heavy drinkers account for over 80% of externalities). Table wine (wine with ABV≤14%) attracts the next highest tax rate followed by beer and strong cider. The lowest rates apply to fortified wine (wine with ABV>14%), weak cider and weak spirits.

Figure 5.1: *Comparison of optimal tax rates under different externality function calibrations*

(a) Optimal single rate



(b) Optimal multi rate



Notes: The figures show the optimal tax rates under various calibrations of the convexity of the externality function, shown on the horizontal axis. Heavy drinkers are defined as the 19% of households that buy the most ethanol (and account for 60% of total ethanol purchased). The vertical axis show the optimal tax rate (p/10ml ethanol). The top panel shows the optimal single tax rate applied to all alcohol products and the bottom panel shows the optimal multi rate system applied to the 8 different alcohol types.

What drives the variation in tax rates?

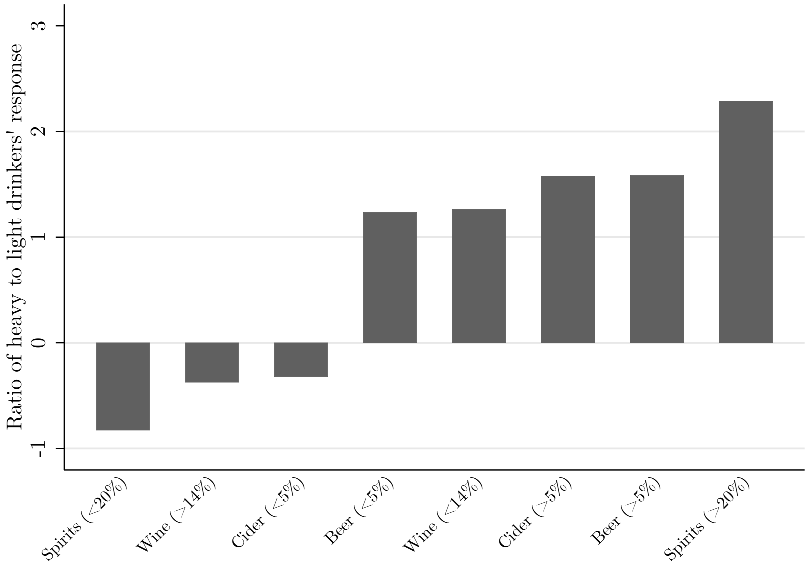
The variation in optimal rates is driven by the correlation between how strongly a tax rate induces households to switch away from ethanol and their marginal externalities. If, for instance, taxing an alcohol type induces heavy drinkers to switch strongly away from that alcohol type, without switching too strongly to alternative sources of ethanol, while leaving the decisions of light drinkers relatively unchanged, then that particular tax is effective at discouraging the most socially costly forms of drinking.

To illustrate empirically what is driving the optimal tax results we compute the change in the quantity of ethanol demanded resulting from a 1% increase in the price of each of the eight alcohol types, doing this separately for the set of heavy and light drinkers. In Figure 5.2 we show the ratio of the change for the heavy drinkers to the change for the light drinkers. An increase in the price of strong spirits reduces the ethanol demand of the heavy drinkers by 2.3 times as much as it reduces ethanol demand for the light drinkers. For beer, table wine and strong cider, a marginal increase in the price of products belonging to each of these alcohol types stimulates larger reductions in total ethanol from heavy drinkers compared with the light group by a factor of between 1.2 and 1.5. This helps explain why the optimal tax rate on strong spirits is greater than those on beer, table wine and strong cider. For weak spirits, fortified wine and weak cider, a marginal increase in the price of products in each of these groups actually increases the ethanol demand of the set of heavy drinkers (though it lowers demand among the lighter drinkers) – this explains why the bars for these alcohol types in Figure 5.2 are negative and why they are the alcohols with the lowest optimal rates. Raising the tax rate on these types of alcohols encourages the heavy drinkers to switch to alternative stronger alcohols, leading the optimal rates on these alcohol types to be relatively low.

Differences in the impact on ethanol demand of a change in tax rate for a given alcohol type across light and heavy drinkers may be due to: (i) differences in level of ethanol from that alcohol type; (ii) the strength of switching away from it (i.e. the alcohol type own tax effect); or (iii) differences in the propensity to switch to alternative alcohol types. In Table 5.1 we compare each of these between the set of light and heavy drinkers.

Columns (1) and (3) show the number of standard drinks per adult per week that the group of light and heavy drinkers get from each alcohol type; column (3) shows the ratio. Heavy drinkers tend to purchase more of each alcohol type. However, by the far the biggest discrepancy is for strong spirits – the 19% of households that comprise the group of heavy drinkers get, on average, over 4 times as much ethanol from this source as the 80% of lightest drinking households. In columns (4) and (5) we show the own price effects for the different alcohol types (i.e. the % change in ethanol demanded from the type following a 1% increase in the price of all products of that type) for the light and heavy drinkers and in column (6) we show the ratio between the two groups. Strong spirits is one of only two alcohol types that the heavy drinkers have a lower (in absolute terms) elasticity compared with lighter drinks. However, the much larger level of strong spirits demanded by the heavy drinkers means that they switch more strongly in level terms away from this alcohol type in response to a price rise. Finally in columns (7)–(9) we show the impact of changes in price for each alcohol type on overall ethanol demand, taking into account substitution towards other alcohols. Column (7), for light, and (8), for heavy drinkers, show the percentage change in total ethanol that results from a 1% increase in the price of each alcohol type; column (9) shows the ratio. For all alcohol types the heavy drinkers switch away from ethanol less in percent terms than the light drinkers. This is because heavy drinkers are much more inclined than lighter drinkers to respond by switching to alternative alcohol types.

Figure 5.2: *Response of heavy and light drinkers to increases in the price of different alcohol types*



Notes: For heavy and light drinkers we calculate the change in total ethanol demand as a result of a 1% increase in the price of the alcohol type shown on the horizontal axis. The bars show the ratio of the change for the heavy drinkers to the light drinkers. Heavy drinkers are defined as the 19% of households that buy the most ethanol (and account for 60% of total ethanol purchased); light drinkers buy the remaining 40%.

Hence, the relatively high tax rate on strong spirits is driven by three factors: (i) heavy drinkers get a large share of their ethanol from these products; (ii) they are reasonably price sensitive with respect to these products (though not as much as light drinkers); (iii) although they tend to switch to alternative alcohol types to a much greater extent than lighter drinkers, alternatives to strong spirits tend to contain much less alcohol. Therefore, overall, taxing strong spirits at a relatively high tax rate is an effective way to reduce the ethanol purchased by the heavy drinking group without imposing large costs on lighter drinkers.

Table 5.1: *Demand responses by alcohol types*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
				% reduction in ethanol demand following 1% increase in price of type, accounting for:					
	Number of standard drinks			Own price effect			Total effect		
	Light	Heavy	Ratio	Light	Heavy	Ratio	Light	Heavy	Ratio
Spirits (<20%)	0.16	0.20	1.2	-2.62	-2.89	1.1	-0.04	0.01	-0.3
Wine (>14%)	0.16	0.37	2.3	-2.34	-2.44	1.0	-0.03	0.01	-0.1
Cider (<5%)	0.19	0.28	1.5	-1.71	-2.20	1.3	-0.02	0.00	-0.1
Beer (<5%)	1.01	2.03	2.0	-3.14	-3.07	1.0	-0.47	-0.23	0.5
Wine (<14%)	1.70	3.36	2.0	-2.14	-2.59	1.2	-0.48	-0.24	0.5
Cider (>5%)	0.27	0.78	2.9	-2.53	-2.32	0.9	-0.09	-0.06	0.6
Beer (>5%)	0.24	0.63	2.6	-3.29	-3.59	1.1	-0.09	-0.06	0.6
Spirits (>20%)	0.90	3.91	4.4	-3.98	-2.73	0.7	-0.59	-0.54	0.9

Notes: Columns (1) and (2) show the number of standard drinks from each alcohol type for light and heavy drinkers respectively. Column (3) shows the ratio of column (2) to (1). We simulate a 1% increase in the price of each alcohol type and calculate the % change in ethanol demanded from that type (shown in columns (4) and (5)) and the % change in total ethanol demanded from all types (shown in columns (7) and (8)). Column (6) shows the ratio of column (5) to (4), and column (9) shows the ratio of column (8) to (7). Heavy drinkers are defined as the 19% of households that buy the most ethanol (and account for 60% of total ethanol purchased); light drinkers buy the remaining 40%.

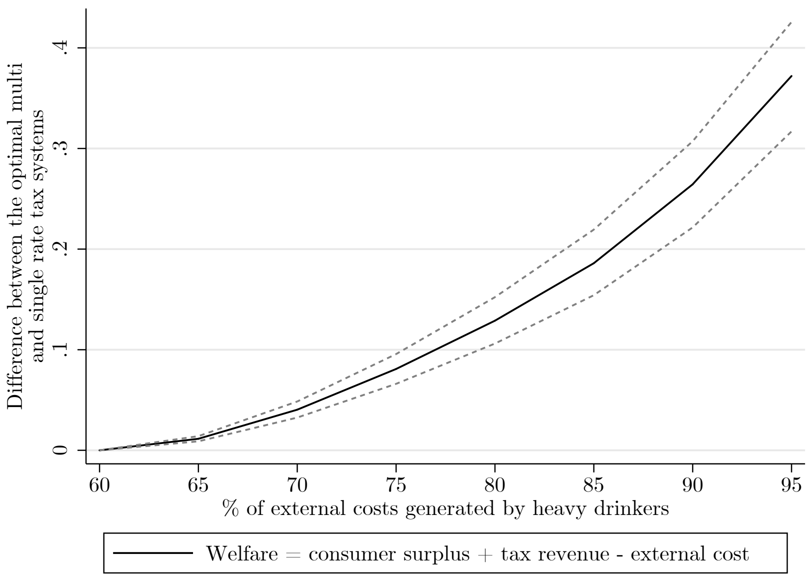
Welfare

In Figure 5.3 we show how social welfare under the optimal multi rate system differs from welfare under the optimal single rate system, and how this difference varies with the degree of convexity of the externality function. Panel (a) shows the impact on total social welfare (the sum of consumer surplus and tax revenue minus external costs) and panel (b) shows the impact on the constituent parts – consumer surplus, tax revenue and the external costs of drinking. When 60% of externalities are generated by the set of heavy drinkers (who, recall, consume 60% of ethanol) there is no difference in the optimal multi and single rate systems.

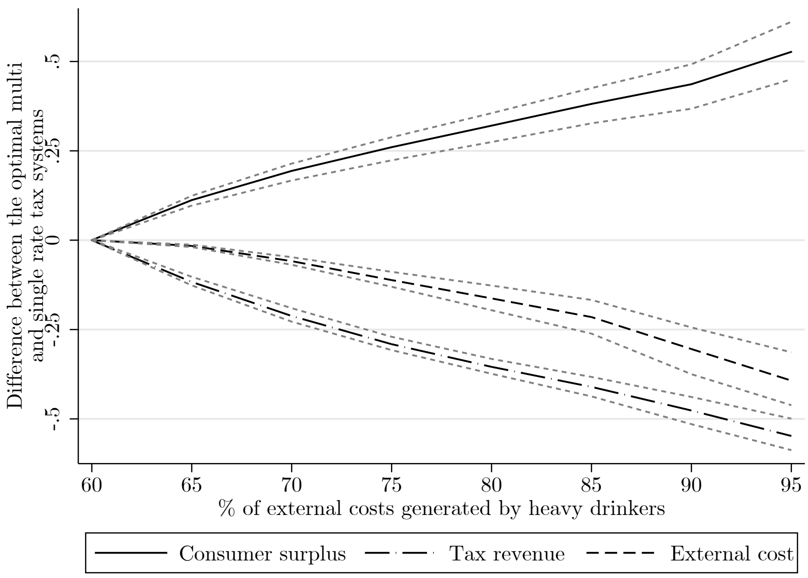
The welfare gain from moving from a single to multi rate system becomes increasingly large as we increase the degree of convexity of the externality function. If the heavy drinkers generate 95% of the external costs, the optimal multi rate system increases welfare by over £350 million, relative to an optimally set single rate system. To put this number into context, the UK Cabinet Office estimate that around £370 million is incurred by alcohol related accident and emergency visits (Cnossen (2007)).[14]

Figure 5.3: *Comparison of welfare under optimal single and multi rate tax systems*

(a) Total welfare



(b) Consumer surplus, tax revenue, externality



Notes: Differences are measured in £billion per year. Heavy drinkers are defined as the 19% of households that buy the most ethanol (and account for 60% of total ethanol purchased). 95% confidence intervals are shown in grey.

[14] The number reported in Cnossen (2007) is €447 million. We convert this to pounds and uprate to 2011 prices using the RPI.

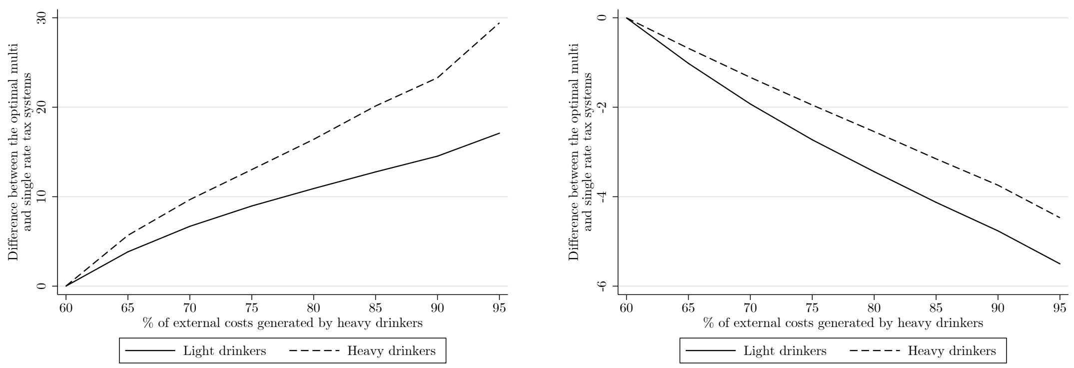
The welfare gain is driven both by lower external costs and higher consumer surplus. The flexibility afforded by the optimal multi rate system enables the planner to target the consumption of the heaviest drinkers by raising the tax rate on strong spirits and lowering the tax rates on other forms of alcohol below the the optimal single rate. This means the multi rate system is able more effectively to reduce ethanol consumption among the most heavy drinkers, while actually reducing the average tax rate on alcohol relative to the optimal single rate. As a consequence, on aggregate, consumer surplus rises even as the total external costs from drinking are brought down. However, as a result, the optimal multi rate system raises less tax revenue than the single rate system and this revenue loss offsets, to some extent, the welfare gains arising from higher consumer surplus and lower externalities.

The gains in consumer surplus associated with having an optimal multi rather than single rate system vary across households. In Figure 5.4 we show (panel (a)) the average consumer surplus change for both the group of light drinkers (these are the 81% of the lightest drinking households, who together account for 40% of all ethanol purchases) and for the heavy drinkers (the remaining 19% of households) and how this varies with the convexity of the externality function. For any strictly convex function both groups have, on average, larger consumer surplus under the multi rate system. Panel (b) shows that the reason for this is that for both light and heavy drinkers, the average tax rate they face for their alcohol is lower under the multi than single rate system. Although the light drinkers see the largest reduction in their average tax rate, their consumer welfare gain in £ terms is lower than for the heavy drinking group. This is because heavy drinkers purchases more ethanol than light drinkers – as a fraction of alcohol expenditure their consumer surplus gain is smaller than for the light drinkers.

The flexibility of the multi rate tax system (relative to a single ethanol tax rate) creates welfare gains through achieving higher average alcohol taxes for heavy relative to light drinkers, thus focusing on lowering the consumption of heavy drinkers while leaving relatively less affected the consumption of light drinkers. The optimal single tax rate on ethanol prescribes a relatively high tax on all alcohol, while the multi rate system can focus much more on reducing spirits consumption among the very heaviest drinkers. This lowers the social costs of drinking while also achieving consumer surplus gains for the majority of households.

Figure 5.4: *Welfare under optimal single and multi rate tax systems for heavy and light drinkers*

(a) Consumer surplus (b) Average tax rate faced



Notes: Difference in consumer surplus measured in £ per household per year. Difference in tax rate is p/10ml ethanol. We measure average tax rates for different households as the weighted average of the tax rates applied to different alcohol types, where the weights are the share of ethanol from each type for each household. Heavy drinkers are defined as the 19% of households that buy the most ethanol (and account for 60% of total ethanol purchased); light drinkers buy the remaining 40%.

Comparison to the UK system

Our primary focus is on how exploiting correlation between heterogeneous marginal externalities and demands can lead to significant welfare gains from varying tax rates across different forms of ethanol. However, we can also use our framework to assess how close the UK tax system gets to an optimal tax system. In the UK there are different tax rates levied on beer, spirits, wine and cider with some variation in rates across different ABV contents. However, both wine and cider are taxed per litre of product (rather than per amount of ethanol), and the system is not coherently targeted at consumers that generate large marginal externalities.

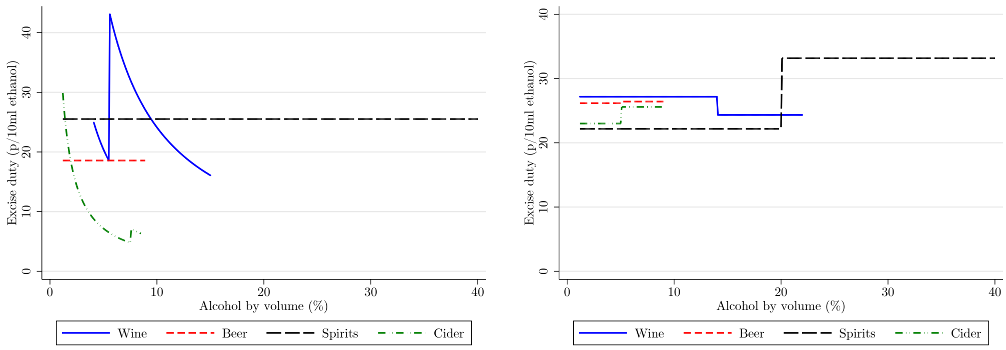
Figure 5.5 compares the optimal tax rates for the calibration in which 80% of the external costs of drinking are generated by heavy drinkers, and the UK tax system in 2011. In the UK, in addition to alcohol excise duty, there is a broad based Value Added Tax (VAT) levied at the rate of 20%. To make our optimal taxes comparable to the UK duty component, in Figure 5.5 we divide them by 1.2.[15] Panel (a) shows the current UK tax system, and panel (b) shows the optimal multi rate system. It is clear that the current system is far from optimal, significant welfare gains could be achieved from: (i) levying taxes on ethanol rather than on volume, (ii) increasing

[15] Our optimal tax estimates are for the total tax levied on alcohol. With a VAT tax of 20%, the optimal alcohol duty rates equal our estimates divided by 1.2.

the tax rate on cider, (iii) reducing the tax rate on spirits below 20% ABV, and increasing the rate on spirits above 20% ABV.

Figure 5.5: *Comparison of the current UK system with the optimal multi rate system*

(a) Current UK system (b) Optimal UK system



Notes: The optimal rates are shown for the calibration in which 80% of the external costs of drinking are generated by heavy drinkers. The UK tax rates are those in place in 2011.

Moving from the UK system to the optimal single rate ethanol tax would realise substantial welfare gains. This size of these gains depend both on the aggregate social costs of drinking and how concentrated they are among heavy drinkers. Maintaining the assumption that the aggregate social costs are £7.25 billion (from Cnossen (2007)), if the heaviest 19% of drinkers account for 80% of the externalities the welfare gain from moving from the UK to single rate system would be £1.23 billion. Moving instead to the optimal multi rate system would result in a further 10% improvement.[16] Although these precise numbers depend on the level and degree of convexity of the externality function, the fact that rationalising the UK tax system would achieve substantial welfare gains holds generally.

5.3 Sufficient statistics approach

In Section 2.2 we describe how we can use our optimal tax framework to provide "sufficient statistics" expressions for the optimal tax rates. The advantage of these is that they do not depend on switching patterns across different alcohol types (and their correlation with marginal externalities). Rather, they depend on (i) the average marginal externality, (ii) the average slope of ethanol demand (with respect to the tax rate in question), and (iii) the covariance between demand slopes and the marginal externality. In principle, these ethanol demand slopes (and how they vary across light and heavy drinkers, and hence with marginal externalities) could be

[16] See Online Appendix B for a breakdown of these numbers into consumer surplus, external costs and tax revenue, as well as a description of the UK tax system.

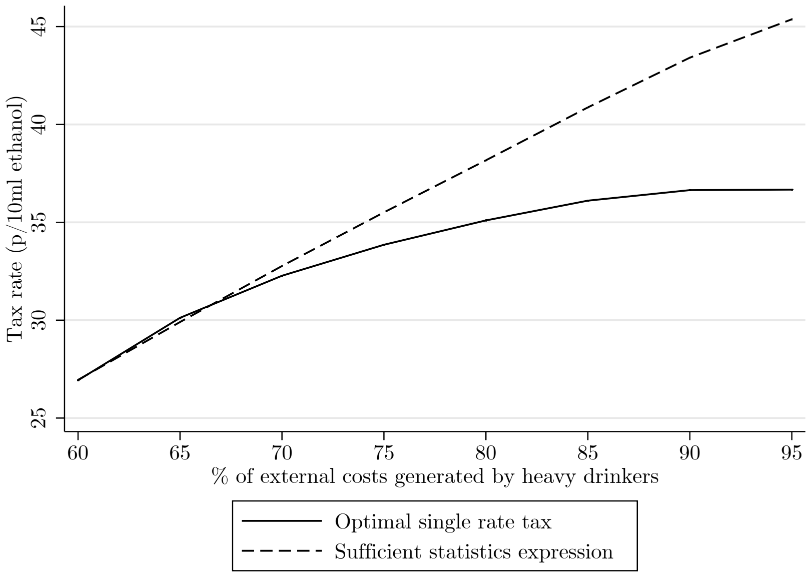
estimated using reduced form methods based on variation in tax rates observed in practice. Such an approach obviates the need to estimate a structure model, but the expressions only hold exactly under additional assumptions and their performance depends on how large deviations are from the these assumptions.

To assess the usefulness of the sufficient statistic approach in our context, we use ethanol demand slopes estimated using our demand model (evaluated at observed prices) to compute the sufficient statistics. This means that the comparisons between sufficient statistics expressions and the optimal rates implied by the structural demand model reflect how closely the additional sufficient statistics assumptions hold and do not conflate differences that arise due structural versus reduced form estimation.

We focus firstly on the optimal single rate ethanol tax. In this case the sufficient statistic expression is based on the assumption that both the slope of total ethanol demand with respect to a change in the price of all alcohols and marginal externalities are constant between the observed tax rates and the optimal rate. The more non-linear are demands and externalities the less likely this restriction will hold. In Figure 5.6 we show how both the optimal single rate tax (computed using the structural demand model) and the sufficient statistics expression vary with the degree of convexity of the externality function. At low levels of convexity the expression does very well at recovering the optimal tax rate. As the externality function becomes more convex, the sufficient statistics expression increasingly over estimates the optimal rate. This in part reflects the fact that as we increase the convexity of the externality function, the optimal rate becomes further from the UK average rate (which is roughly 27p/10ml ethanol), meaning the assumption of constant demand derivatives over the range of tax rates being considered becomes an increasingly less reasonable assumption.

The sufficient statistics expressions to the multi rate system are derived under a second assumption (additional to derivatives being constant between the observed and optimal tax rates). This is that there are zero cross price effects between the sets of products belonging to different alcohol types. This assumption allows us to analyse the first order conditions of the planner's problem on an alcohol type by type basis. However, the stronger cross alcohol type switching and the more correlated this is with marginal externalities, the less well the expressions will perform.

Figure 5.6: *Comparison of the optimal single rate and the sufficient statistics expression*



Notes: The optimal single rate is the same as shown in Figure 5.1(a). The sufficient statistics expression is computed using equation (2.6), where the statistics are evaluated at the current UK tax rate. Heavy drinkers are defined as the 19% of households that buy the most ethanol (and account for 60% of total ethanol purchased).

In Table 5.2 we provide evidence on how well the sufficient statistics expressions do. We consider two alternative degrees of convexity of the externality function; a mild degree (where the group of heavy drinkers account for 65% of the externalities) and a moderately large degree (where the group of heavy drinkers account for 80% of the externalities). In each case – columns (1) and (4) – we report the sign of the first order conditions for the planner's problem evaluated at observed prices with sufficient statistics (see equation 2.5). This does not require the assumption of constant demand derivatives between the observed rates and the optimal rates, but does impose the assumption of no cross price effects between alcohol types. A positive (negative) sign denotes that the sufficient statistics indicate the optimal tax is above (below) the UK level. For both the 65% and 80% convexity calibrations the sufficient statistics correctly indicate the direction tax rate should change to go towards the optimum.

In columns (2) and (3) of Table 5.2 we compare the difference in optimal tax rates with the UK rates and the sufficient statistics expressions for the 65% calibration.[17] Columns (5) and (6) do the same for the 80% calibration. For the low convexity case the sufficient statistics expressions get relatively close to the optimal rates (and are much closer than the UK system); they deviate by at most 3p/10ml. The

[17]The UK tax system imposes VAT on alcohol in addition to excise duty. In Table 5.2 we present optimal taxes prior to adjusting for VAT and compare to UK taxes inclusive of VAT.

sufficient statistics expressions typically overestimate the optimal rates. The reason is that heavy drinkers are more willing to switch between alcohol types than light drinkers. This means alcohol taxes are less effective at reducing the most socially costly alcohol consumption than when cross price effects between alcohol types are zero, which acts to lower optimal tax rates. For the more convex case the sufficient statistics expressions are much further from the optimum (and further than the UK rates are). This is due both to the fact that the sufficient statistics expressions ignore cross price effects between alcohol types, and assume that the statistics are constant between the observed and optimal tax rates.

Table 5.2: *Comparison of the optimal multi rates and the sufficient statistics expressions*

	(1)	(2)	(3)	(4)	(5)	(6)
	65% external costs generated by heavy drinkers			80% external costs generated generated by heavy drinkers		
	Sign of	Difference between optimal rate and:		Sign of	Difference between optimal rate and:	
Alcohol type	FOC	UK	SS approx	FOC	UK	SS approx
Spirits (<20%)	−	-14.35	-0.60	−	-15.09	-4.17
Cider (<5%)	+	10.05	-2.20	+	9.92	-10.50
Wine (>14%)	+	4.86	-1.96	+	6.01	-9.22
Cider (>5%)	+	17.89	-3.14	+	19.90	-14.82
Beer (<5%)	+	2.56	-1.61	+	4.97	-9.52
Beer (>5%)	+	2.17	-1.78	+	4.76	-10.31
Wine (<14%)	−	-3.08	-0.78	+	0.33	-5.91
Spirits (>20%)	−	-0.40	0.11	+	8.20	-2.87

Notes: Columns (1)–(3) show numbers for the calibration under which 65% of the external costs are generated by the heavy drinkers. Columns (4)–(6) show numbers for the calibration under which 80% of the external costs are generated by the heavy drinkers. Heavy drinkers are defined as the 19% of households that buy the most ethanol (and account for 60% of total ethanol purchased). Columns (1) and (4) show the sign of the planner's first order condition evaluated at the current UK taxes and using sufficient statistics. Columns (2) and (5) show the difference between the optimal tax rate and the UK tax rate, in pence/10ml ethanol. Columns (3) and (6) show the difference between the optimal tax rate and the sufficient statistics expression, in p/10ml ethanol.

6 Summary and conclusions

In this paper we consider corrective tax design in markets in which an externality generating commodity is available in many products and consumers potentially are heterogeneous in both the externalities that their consumption creates and in their demands. We focus on the alcohol market. There is much evidence that consumption of ethanol (which is available in many products bundled together with other attributes) is associated with externalities and these externalities are non-linear. We consider degrees of non-linearity in this relationship, ranging from marginal externalities that are constant in ethanol consumption to the majority of the social costs of alcohol being generated by a small group of heavy drinkers.

Our results show that varying tax rates across different forms of alcohol can lead to significant welfare gains relative to an optimally set single ethanol tax rate. Optimally varying rates exploits correlations in households' preferences for product attributes (and hence product level demands) with their total demand for ethanol, enabling the tax system to better target consumption that generates high externalities. Welfare gains are larger the more convex externalities are in ethanol consumption.

To implement our optimal tax framework empirically we estimate a structural demand model using detailed longitudinal micro data on households' alcohol purchases. We also investigate an alternative sufficient statistics approach. We show that the sufficient statistics are useful for indicating the direction of change in tax rates to move towards the optimal. When the externality function is not too convex, they also do a good job of approximating optimal tax rates, although this becomes less true for more convex specifications of externalities.

Our focus in this paper has been on the correction of externalities. We have considered a social planner that sets taxes to maximise the sum of consumer surplus and tax revenue minus external costs. The social planner does not take account of the existence of positive mark ups arising from imperfect competition. We have also assumed complete pass-through of tax to consumer prices. In the UK alcohol market we believe these are defensible abstractions; UK supermarkets, by international standards, are very competitive and policy is concerned with tackling excessive consumption. However, an important avenue for future research is to incorporate supply side considerations into the optimal tax framework.

The framework that we develop is well suited to other applications in which there are heterogeneous consumption externalities in differentiated product markets. For example, concern about obesity and the excess consumption of sugar has led to growing interest in sugar taxes. In this case specific groups may be more prone to generate externalities (including on their future self) – for instance, there are particular concerns surrounding children's sugar consumption. If there is correlation between the preferences for different soda products and the marginal externality of sugar consumption, then application of our model would shed light on the design of sugar taxes that reduce the externality while minimising the reduction in consumer surplus.

References

Aguiar, M. and E. Hurst (2007). Life-cycle prices and production. *American Economic Review 97*(5), 1533–1559.

Akerlof, G. A. (1978). The Economics of "Tagging" as Applied to the Optimal Income Tax, Welfare Programs, and Manpower Planning. *American Economic Review 68*(1), 8–19.

Atkinson, A. and J. Stiglitz (1976). The design of tax structure: Direct versus indirect taxation. *Journal of Public Economics 6*(1-2), 55–75.

Baily, M. N. (1978). Some aspects of optimal unemployment insurance. *Journal of Public Economics 10*(3), 379–402.

Baltagi, B. H. and J. M. Griffin (1995). A Dynamic Demand Model for Liquor: The Case for Pooling. *Review of Economics and Statistics 77*(3), 545–554.

Berry, S., J. Levinsohn, and A. Pakes (1995). Automobile prices in market equilibrium. *Econometrica 63*(4), 841–890.

Berry, S., J. Levinsohn, and A. Pakes (2004). Differentiated Products Demand Systems from a Combination of Micro and Macro Data: The New Car Market. *Journal of Political Economy 112*(1), 68–105.

Berry, S. T. (1994). Estimating discrete-choice models of product differentiation. *RAND Journal of Economics 25*(2), 242–262.

Berry, S. T. and P. A. Haile (2010). Nonparametric Identification of Multinomal Choice Demand Models with Heterogeneous Consumers. *Cowles Foundation Discussion Paper* (1718), 1–41.

Blundell, R. W. and J. L. Powell (2004). Endogeneity in semiparametric binary response models. *Review of Economic Studies 71*(3), 655–679.

Bovenberg, A. L. and L. H. Goulder (1996). Optimal Environmental Taxation in the Presence of Other Taxes: General- Equilibrium Analyses. *American Economic Review 86*(4), 985–1000.

Bovenberg, A. L. and L. H. Goulder (2002). Environmental taxation and regulation. *Handbook of Public Economics 3*, 1471–1545.

Centers for Disease Control and Prevention (2016). CDC - Data and Maps - Alcohol. http://www.cdc.gov/alcohol/data-stats.htm.

Chetty, R. (2006). A general formula for the optimal level of social insurance. *Journal of Public Economics 90*(10-11), 1879–1901.

Chetty, R. (2008). Moral Hazard versus Liquidity and Optimal Unemployment Insurance. *Journal of Political Economy 116*(2), 173–234.

Chetty, R. (2009). Sufficient Statistics for Welfare Analysis: A Bridge Between Structural and Reduced-Form Methods. *Annual Review of Economics 1*(1), 451–488.

Cnossen, S. (2007). Alcohol taxation and regulation in the European Union. *International Tax and Public Finance 14*(6), 699–732.

Conlon, C. T. and N. S. Rao (2015). The Price of Liquor is Too Damn High: Alcohol Taxation and Market Structure. *NYU Wagner Research Paper No. 2610118*.

Crawford, I., Z. Smith, and S. Tanner (1999). Alcohol Taxes, Tax Revenues, and the Single European Market. *Fiscal Studies 20*(3), 287–304.

Cremer, H., P. Pestieau, and J.-C. Rochet (2001). Direct Versus Indirect Taxation: The Design of the Tax Structure Revisited. *International Economic Review 42*(3), 781–800.

Diamond, P. A. (1973). Consumption externalities and corrective imperfect pricing. *The Bell Journal of Economics 4*(2), 526–538.

Dubois, P., R. Griffith, and A. Nevo (2014). Do Prices and Attributes Explain International Differences in Food Purchases? *American Economic Review 104*(3), 832–867.

Feldstein, M. (1999). Tax Avoidance and the Deadweight Loss of the Income Tax. *The Review of Economics and Statistics 81*(4), 674–680.

Fox, J. T. and A. Gandhi (2016). Nonparametric identification and estimation of random coefficients in multinomial choice models. *RAND Journal of Economics 47*(1), 118–139.

Fox, J. T., K. I. Kim, S. P. Ryan, and P. Bajari (2012). The random coefficients logit model is identified. *Journal of Econometrics 166*(2), 204–212.

Gorman, W. M. (1980). A Possible Procedure for Analysing Quality Differentials in the Egg Market. *Review of Economic Studies 47*(5), 843–856.

Griffith, R., A. Leicester, and M. O'Connell (2013). Price-based measures to reduce alcohol consumption. IFS Briefing Note 138.

Griffith, R. and M. O'Connell (2009). The use of scanner data for research into nutrition. *Fiscal Studies 30*(3-4), 339–365.

Gruber, J. (1997). The Consumption Smoothing Benefits of Unemployment Insurance. *American Economic Review 87*(1), 192–205.

Gruber, J. and E. Saez (2002, April). The elasticity of taxable income: Evidence and implications. *Journal of Public Economics 84*(1), 1–32.

Hansen, B. (2015). Punishment and Deterrence: Evidence from Drunk Driving. *American Economic Review 105*(4), 1581–1617.

Irvine, I. J. and W. A. Sims (1993). The Welfare Effects of Environmental Taxation. *Journal of Public Economics 52*(1), 101–119.

Jackson, C. K. and E. G. Owens (2011). One for the road: Public transportation, alcohol consumption, and intoxicated driving. *Journal of Public Economics 95*, 106–121.

Jacobsen, M. R., C. R. Knittel, J. M. Sallee, and A. A. van Benthem (2016). Sufficient statistics for imperfect externality-correcting policies. *NBER Working Paper 22063*.

Kaplow, L. (1996). The Optimal Supply of Public Goods and the Distortionary Cost of Taxation. *National Tax Journal 49*(4), 513–533.

Kaplow, L. (2006). On the undesirability of commodity taxation even when income taxation is not optimal. *Journal of Public Economics 90*(6–7), 1235–1250.

Kaplow, L. (2012). Optimal Control of Externalities in the Presence of Income Taxation. *International Economic Review 53*(2), 487–509.

Lancaster, K. (1971). *Consumer Demand: A New Approach*. Columbia University Press New York.

Laroque, G. R. (2005, April). Indirect taxation is superfluous under separability and taste homogeneity: A simple proof. *Economics Letters 87*(1), 141–144.

Leicester, A. and Z. Oldfield (2009). Using scanner technology to collect expenditure data. *Fiscal Studies 30*, 309–37.

Lönnroth, K., B. G. Williams, S. Stadlin, E. Jaramillo, and C. Dye (2008). Alcohol use as a risk factor for tuberculosis–a systematic review. *BMC Public Health 8*(1).

Luca, D. L., E. Owens, and G. Sharma (2015). Can Alcohol Prohibition Reduce Violence Against Women? *American Economic Review 105*(5), 625–629.

Mankiw, N. G. and M. Weinzierl (2010). The Optimal Taxation of Height: A Case Study of Utilitarian Income Redistribution. *American Economic Journal: Economic Policy 2*(1), 155–176.

Manning, W. G., L. Blumberg, and L. H. Moulton (1995). The demand for alcohol: The differential response to price. *Journal of Health Economics 14*(2), 123–148.

Miravete, E. J., K. Seim, and J. Thurk (2016). Naive Policy Design and Legislative Inertia in the Regulation of Alcohol. Technical report.

Nilsson, J. P. (2017). Alcohol Availability, Prenatal Conditions, and Long-Term Economic Outcomes. *Journal of Political Economy 125*(4), 1149–1207.

Petrin, A. and K. Train (2010). A Control Function Approach to Endogeneity in Consumer Choice Models. *Journal of Marketing Research 47*(1), 3–13.

Rehm, J., B. Taylor, S. Mohapatra, H. Irving, D. Baliunas, J. Patra, and M. Roerecke (2010). Alcohol as a risk factor for liver cirrhosis: A systematic review and meta-analysis. *Drug and alcohol review 29*(4), 437–445.

Ruhm, C. J. (1996). Alcohol policies and highway vehicle fatalities. *Journal of Health Economics 15*(4), 435–454.

Saez, E. (2001). Using Elasticities to Derive Optimal Income Tax Rates. *Review of Economic Studies 68*(1), 205–229.

Saez, E. (2002). The desirability of commodity taxation under non-linear income taxation and heterogeneous tastes. *Journal of Public Economics 83*(2), 217–230.

Seim, B. K. and J. Waldfogel (2013). Public Monopoly and Economic Efficiency: Evidence from the Pennsylvania Liquor Control Board's Entry Decisions. *American Economic Review 103*(2), 831–862.

UK Cabinet Office (2003). Alcohol misuse: How much does it cost? Technical report.

US Department of Justice (2005). *Drinking in America: Myths, Realities, and Prevention Policy.* Report by Office of Juvenile Justice and Delinquency Prevention Published: www.udetc.org/documents/Drinking_in_America.pdf, accessed 13 April 2014.

Wales, T. J. and A. D. Woodland (1983). Estimation of consumer demand systems with binding non-negativity constraints. *Journal of Econometrics 21*(3), 263–285.

Weinzierl, M. (2011). The Surprising Power of Age-Dependent Taxes. *Review of Economic Studies 78*(4), 1490–1518.

World Health Organization (2014). Global Status Report on Alcohol and Health. Technical report, Geneva: World Health Organization.

Appendix

For Online Publication

Tax design in the alcohol market

Rachel Griffith, Martin O'Connell and Kate Smith

A Data

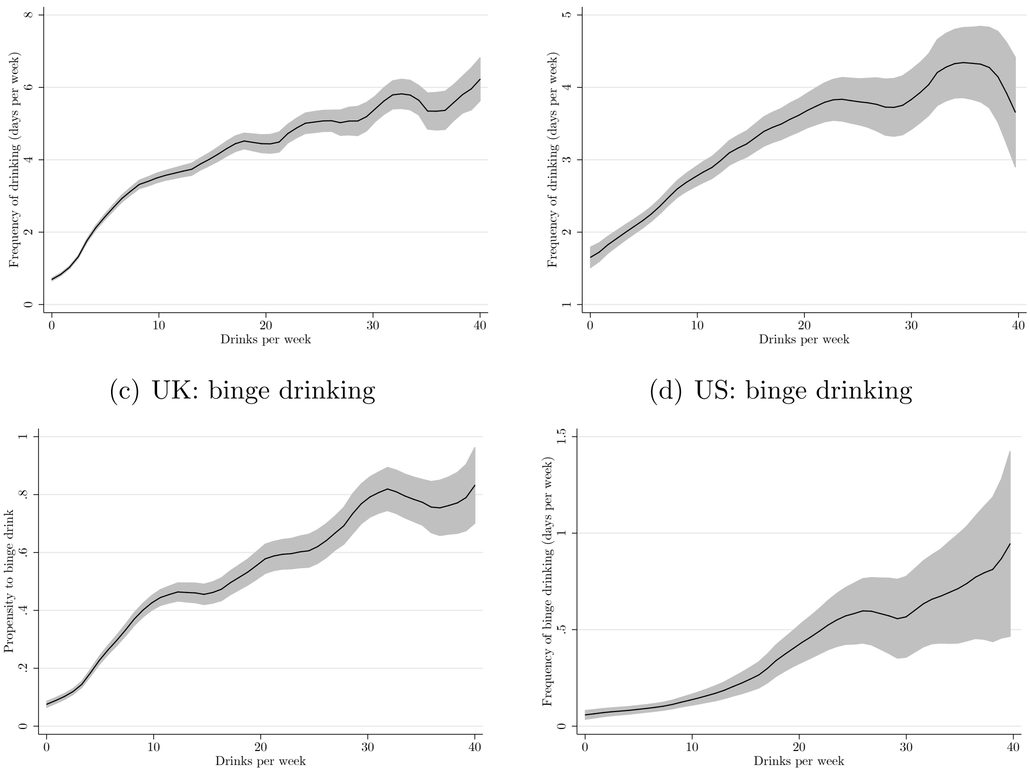
A.1 Alcohol purchase patterns

We use the total amount of ethanol (or standard drinks) per adult per week that a household purchases as the argument of the externality function. Figure A.1 uses the Health Survey for England (HSE) for the UK and the National Health and Examination Survey (NHANES) for US to show that drinks per week is strongly correlated with both the frequency of drinking and the propensity to binge drink. In particular, panels (a) and (b) show that in both the UK and US people that report consuming higher amounts of ethanol also report drinking more days per week. Panels (c) and (d) show that in both countries there is a positive relationship between consumers' total ethanol and whether they reported binge drinking in the previous week. In the rest of this subsection we describe how we use the HSE and NHANES data sets to create this figure.

Figure A.1: *Ethanol consumption, binge and high frequency drinking*

(a) UK: frequency of drinking (b) US: frequency of drinking

(c) UK: binge drinking (d) US: binge drinking



Notes: Panels (a) and (c) are drawn using data from the alcohol questionnaire component of the HSE. Panels (b) and (d) are drawn using data from the alcohol questionnaire and food and drink diary components of NHANES.

Health Survey for England (HSE)

HSE combines interviews and physical examinations to assess the health status of adults and children in the United Kingdom. We use data on 8281 individuals over the age of 18 in the 2011 survey. We use the alcohol questionnaire component of the survey. We use the derived variable *totalwu* – total units of alcohol per week, which is derived from questions about the individuals drinking habits to construct the units of ethanol per week variable used on the horizontal axis of Figures A.1(a) and A.1(c).

Figure A.1(a) uses the responses to question *d7many* ("How many days in the last 7 have you had a drink?") in a local polynomial regression to estimate the relationship between ethanol consumption and frequency of drinking. We use the responses to question *d7unitwg* ("Number of units drunk on the heaviest day in the past 7 days") to construct a variable indicating the propensity to binge drink. This is equal to 1 if the individual was male (female) and recorded drinking over 8

(6) units on their heaviest drinking day out of the past 7. Figure 1(c) uses a local polynomial regression to estimate the relationship between ethanol consumption and propensity to binge drink.

National Health and Examination Survey (NHANES)

NHANES combines interviews and physical examinations to assess the health and nutritional status of adults and children in the United States. We use data on 15,699 individuals over the age of 21 from the 2007 – 2011 surveys. We use two components from the survey.

The first is the diary component. Individuals record all foods and beverages consumed during the 24-hour period of the interview (midnight to midnight). Individuals are interviewed twice: the first dietary recall interview is collected in-person, and the second interview is collected by telephone 3 to 10 days later. To construct the variable measured on the horizontal axis of Figures 1(b) and 1(d) we average all ethanol consumed over the two separate diary days, and convert to standard drinks (1 standard drink = 14g ethanol).

The second component we use is the alcohol questionnaire, which focuses on lifetime and current use (past 12 months). We use the answers to two questions to draw Figures A.1(b) and A.1(d). Figure A.1(b) uses questions *ALQ120Q* ("How often did you drink alcohol over the past 12 months?") and *ALQ120U* (unit of measure for question *ALQ120Q*) to construct the average per week drinking frequency. Figure A.1(d) uses questions *ALQ141Q* ("On how many days over the past 12 months did you consume 4 or 5 alcoholic beverages?") and *ALQ141U* (unit of measure for question ALQ141Q) to construct the average number of days per week on which the individual engaged in binge drinking. Figures A.1(b) and A.1(d) fits local polynomial regressions between these variables and the ethanol consumption variable constructed from the diary data for the subset of individuals who record consuming non-zero quantities of ethanol in the diary (3234 individuals).

A.2 On versus off trade alcohol

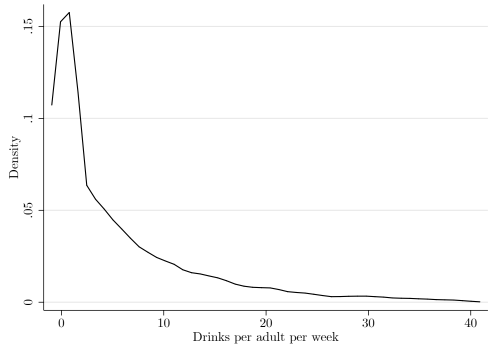
One of the advantages of the Kantar Worldpanel is that we can calculate how much alcohol people buy on average over a long period, as opposed to just making a one-off large purchase. Cross sectional expenditure surveys, (e.g. the Living Costs and Food Survey (LCFS) and the Consumer Expenditure Survey (CEX)) and intake diaries (e.g. the Health Survey for England (HSE) and National Health and Nutrition Survey (NHANES)), have much shorter reporting periods, which makes it harder to identify consistently heavy drinkers. Nonetheless, we use the questions in the HSE about average weekly alcohol consumption to verify whether our sample is representative of the distribution of drinkers. The HSE data cover alcohol consumption from purchases made off-trade and on-trade (in bars and restaurants); we scale up average standard drinks per adult per week in the Kantar data to account for the absence of on-trade alcohol purchases. In the HSE, 9% of individuals who report drinking in the last 12 months report a weekly consumption of more than 20 standard drinks; in our data this number is 7%. This suggests that we are doing a reasonable job at capturing the alcohol purchases of heavy drinkers (Health Survey for England (2016)).

Our data contain very detailed information on purchases of alcohol products off-trade, but they do not contain information on alcohol purchases on-trade (those made in restaurants and bars). The LCFS contains information on alcohol purchased both on- and off-trade. It is a two week diary survey with a sample of 3688 households in 2011 who record buying alcohol. Unlike the Kantar data, the data do not contain repeated observations for the same households over time, product level information, transaction prices nor any measure of alcohol strength. Nevertheless, we can use these data to get an idea of whether purchase patterns are similar between off-trade alone and on- and off-trade alcohol together. To do this we impute the strength of the alcohol categories collected in the LCFS. For instance, for the category beer we use 4% ABV – the average from the Kantar data. Based on this, in 2011, we compute that 77% of units of ethanol purchased was done so off-trade.

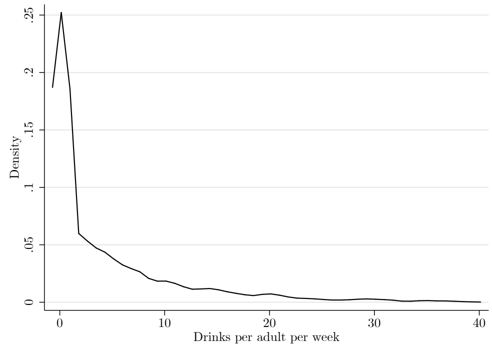
We also use this data to show that the patterns of alcohol purchases, and crucially how it varies with total ethanol demand, is similar for off- and on-trade alcohol. Figure A.2 plots the distributions of ethanol purchases for on- and off-trade. It shows that the shape of the distribution is similar for both on- and off-trade together, and off-trade only. Figure A.3 shows the how the share of ethanol from different alcohol segments varies across the total ethanol purchase distribution. The figure shows that the pattern of households with relatively large ethanol demands getting a relatively low share of their ethanol from beer and a relatively high share from spirits holds for both off-trade alone and on- and off-trade together. This suggests our focus on off-trade purchases is unlikely to result in a substantially different pattern of optimal taxes across products than would result if we estimated demand including the 23% of ethanol purchased on-trade.

Figure A.2: *Distribution of ethanol purchases: on- and off-trade*

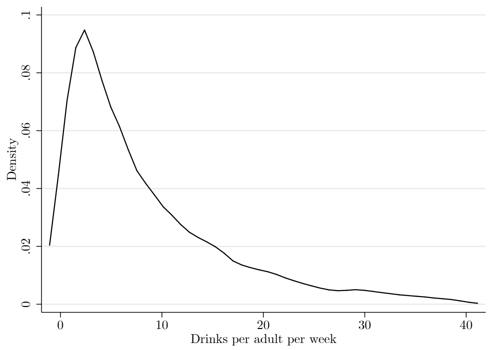
(a) On- and off-trade



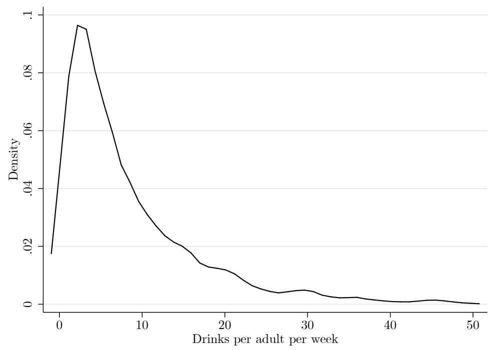
(b) Off-trade only



(c) On- and off-trade: conditional on buying



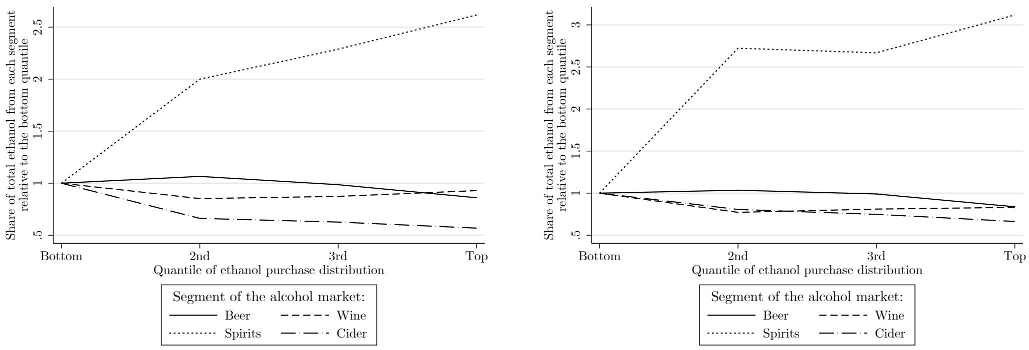
(d) Off-trade only: conditional on buying



Notes: Both panels use data from the Living Costs and Food survey 2011; the left hand panel shows alcohol purchases made on- and off-trade, the right hand panel shows alcohol purchased made off-trade only. We calculate the total amount of ethanol purchased per adult in each household over the two week period; the figures show the distribution of this variable. The top two figures show the unconditional distributions across all households; the bottom two figures show the distributions conditional on purchasing alcohol (at all (left), or off-trade only (right)) during the two week survey period.

Figure A.3: *Alcohol purchases: on- and off-trade*

(a) On- and off-trade (b) Off-trade only

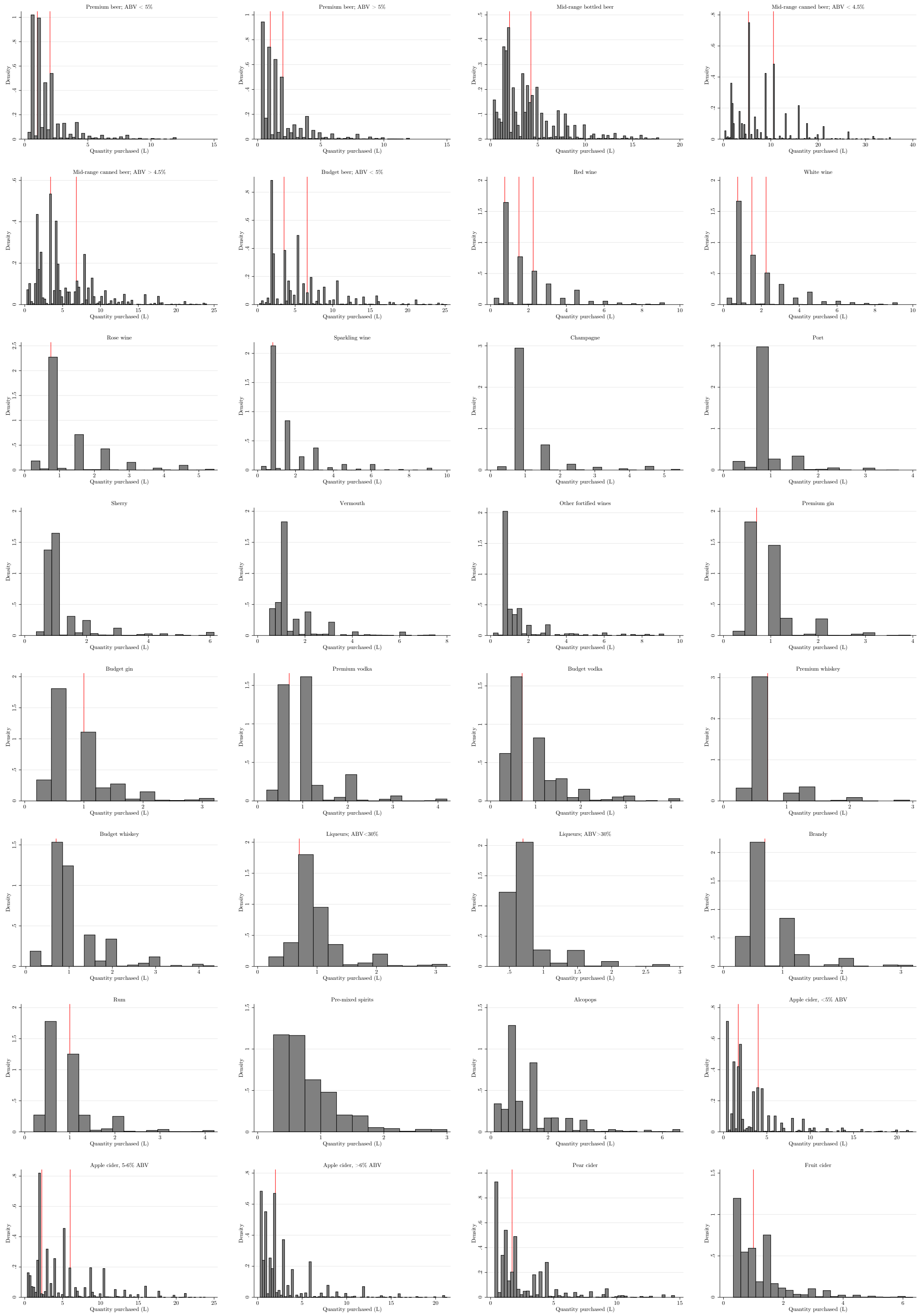


Notes: Both panels use data from the Living Costs and Food survey 2011; the left hand panel shows alcohol purchases made on- and off-trade, the right hand panel shows alcohol purchased made off-trade only. We calculate the total amount of ethanol purchased per adult in each household over the two week period, and divide households into quartiles based in this measure. The y-axis shows the proportion of ethanol that comes from beer, wine, spirits, cider relative to the bottom quantile.

A.3 Size definition

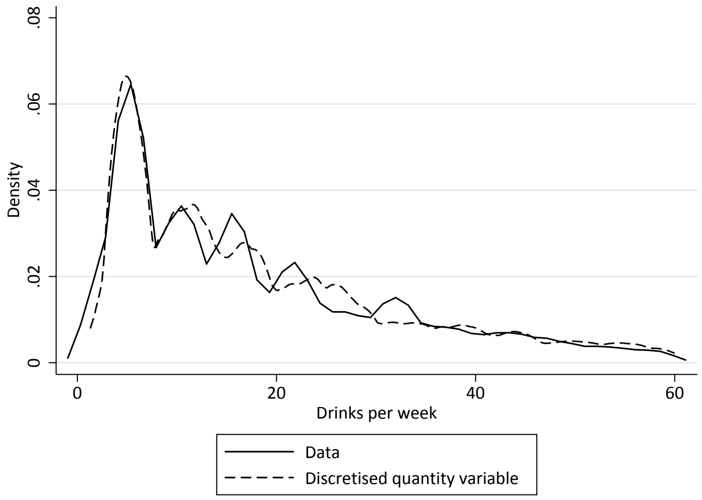
For each of the 32 alcohol products in our demand system we discretize the distribution of quantity purchased on individual purchase occasions by defining a set of equally sized categories. The number of size categories varies across products based on how dispersed the quantity distribution is. In Figure A.4 we plot the quantity distribution for each product and show the cutoff points that define the discrete size categories. In Figure A.5 we show the distribution of drinks per adult per week across household-weeks in the data and constructed based on our discretisation of the quantity distribution. The figure shows the distributions are very similar.

Figure A.4: *Distributions of quantity purchased within products*



Notes: Distributions of quantity purchased across weeks for each product. The red lines show the cutoffs used to discretize the quantity distribution.

Figure A.5: *Drinks distribution with discretized size variable*



Notes: The solid line plots the distribution of drinks purchased per household-week calculated using the raw data. The dashed line plots the distribution of drinks purchased per household-week calculated using the discretized quantity variable.

A.4 Prices

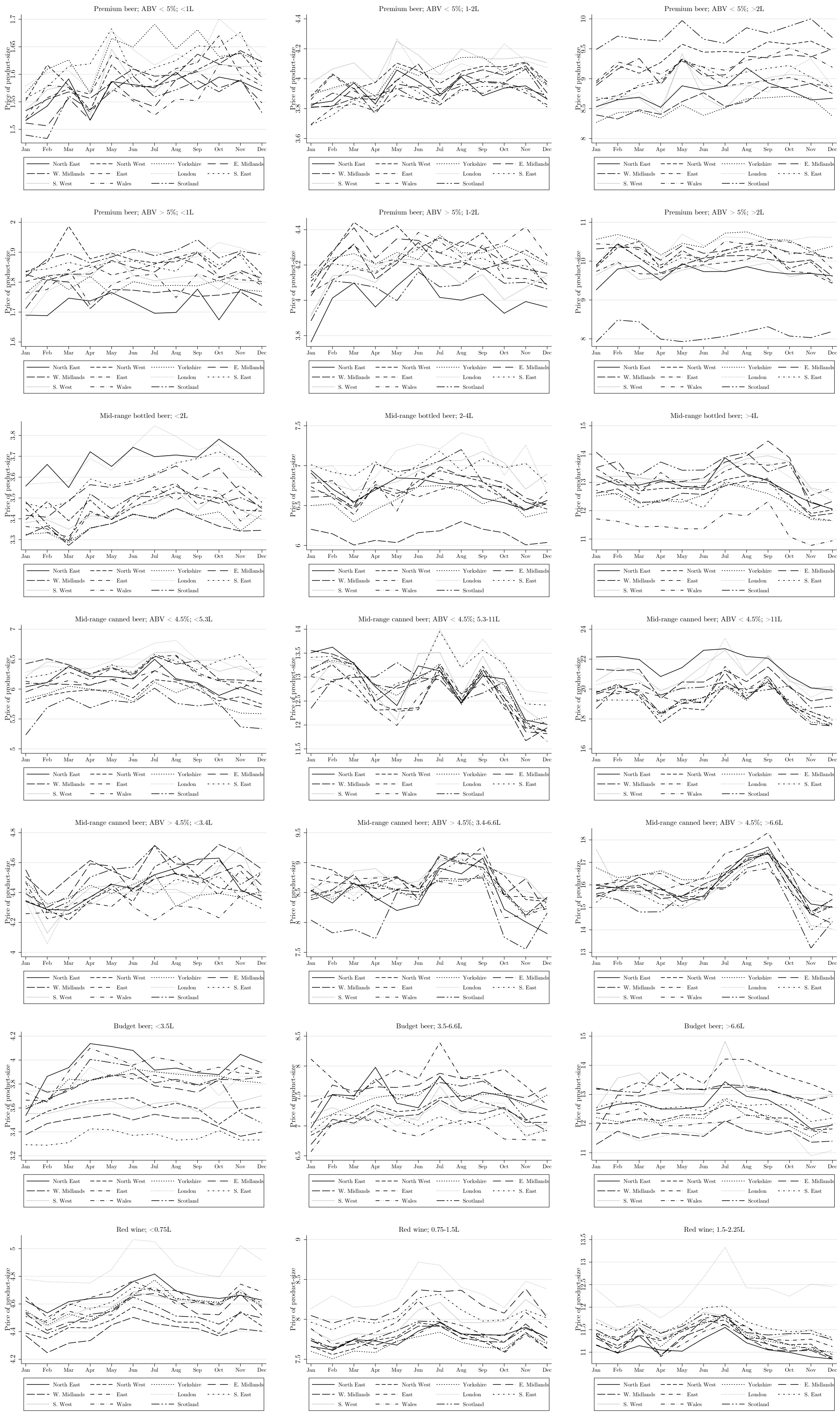
In Table A.1 we report the mean price for each product-size. This price is constructed as a fixed weight price index as described in the main paper. These prices vary geographically and through time. Figure A.6 depicts that variation.

Table A.1: *Product sizes and prices*

(1) Product definition	(2) Size	(3) Mean quantity (L)	(4) Mean price (£)
Beer			
Premium beer; ABV $<$ 5%	500ml	0.52	1.59
	1-2L	1.32	3.96
	2.5-8L	3.63	9.01
Premium beer; ABV $\geq$ 5%	500ml	0.52	1.82
	1-2L	1.35	4.19
	2.5-10L	3.59	9.92
Mid-range bottled beer	1-2L	1.43	3.50
	2.5-4L	3.01	6.71
	5-14L	6.58	12.79
Mid-range canned beer; ABV $<$ 4.5%	2-5L	3.47	6.15
	7-10L	8.13	12.80
	15-25L	14.72	19.93
Mid-range canned beer; ABV $\geq$ 4.5%	1-3L	1.94	4.43
	4-6L	4.16	8.58
	8-20L	8.94	15.96
Budget beer	2-4L	2.05	3.72
	4-6L	4.31	7.32
	8-20L	8.34	12.50
Wine			
Red wine	1x750ml	0.72	4.59
	2x750ml	1.22	7.90
	3x750ml	1.78	11.44
	4x750ml	3.08	18.35
White wine	1x750ml	0.72	4.44
	2x750ml	1.23	7.73
	3x750ml	1.76	10.99
	4x750ml	2.88	16.81
Rose wine	1x750ml	0.72	4.24
	2x750ml	1.79	10.14
Sparkling wine	1x750ml	0.74	5.24
	2x750ml	2.27	9.56
Champagne	1-2x750ml	1.16	26.54
Port	1-2x750ml	0.90	8.89
Sherry	1-2x750ml	1.20	7.85
Vermouth	1-2x750ml	1.32	7.12
Other fortified wines	1-2x750ml	1.33	6.63
Spirits			
Premium gin	1x700ml	0.69	11.74
	2x700ml	1.16	17.95
Budget gin	1x700ml	0.75	10.12
	2x700ml	1.27	14.89
Premium vodka	1x700ml	0.67	10.25
	2x700ml	1.16	16.17
Budget vodka	1x700ml	0.59	8.18
	2x700ml	1.14	14.71
Premium whiskey	1x700ml	0.67	19.60
	2x700ml	1.29	30.55
Budget whiskey	1x700ml	0.66	10.96
	2x700ml	1.21	16.17
Liqueurs; ABV $<$30%	1x700ml	0.64	7.82
	2x700ml	1.25	15.45
Liqueurs; ABV $\geq$30%	1x700ml	0.62	13.57
	2x700ml	1.14	21.38
Brandy	1x700ml	0.63	10.77
	2x700ml	1.11	17.43
Rum	1x700ml	0.79	12.26
	2x700ml	1.45	19.98
Pre-mixed spirits	700ml	0.70	4.37
Alcopops	1.3L	1.32	5.89
Cider			
Apple cider, $<$5% ABV	1L	0.91	2.66
	2-3L	2.45	3.94
	6-10L	7.10	10.15
Apple cider, 5-6% ABV	1-2L	1.62	2.77
	4L	3.79	5.49
	10-14L	8.41	10.39
Apple cider, $>$6% ABV	1-2L	1.21	3.28
	3-9L	4.40	7.06
Pear cider	1L	0.92	2.53
	3-6L	3.87	7.21
Fruit cider	750ml	0.68	2.48
	1-3L	1.98	6.75

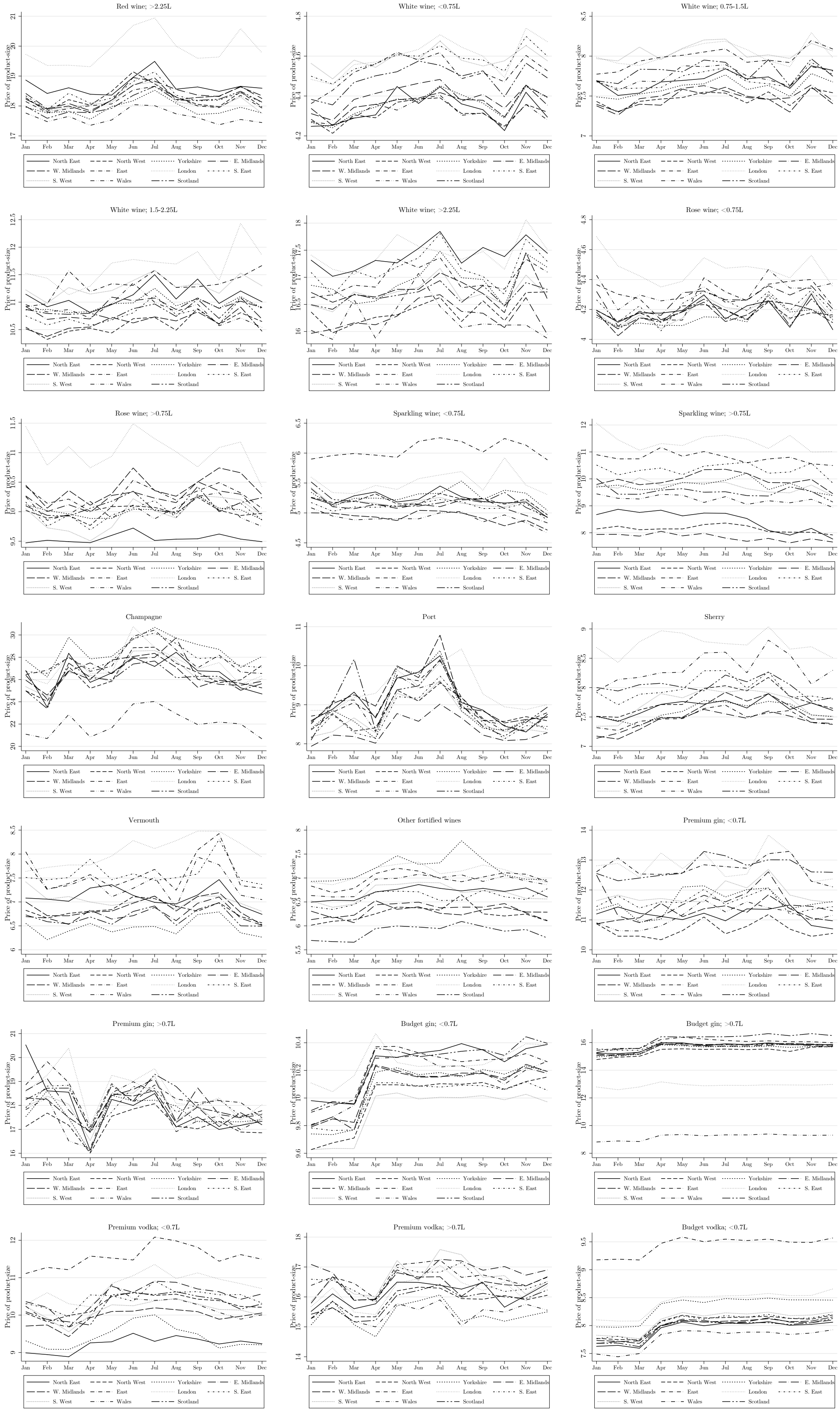
Notes: Mean quantity is the average quantity of each product purchased by households in a given week over the calendar year. Mean price is the average price (constructed as described in Section 3.3 of the paper) over regions and months in 2011.

Figure A.6: *Price indices for product-size pairs*



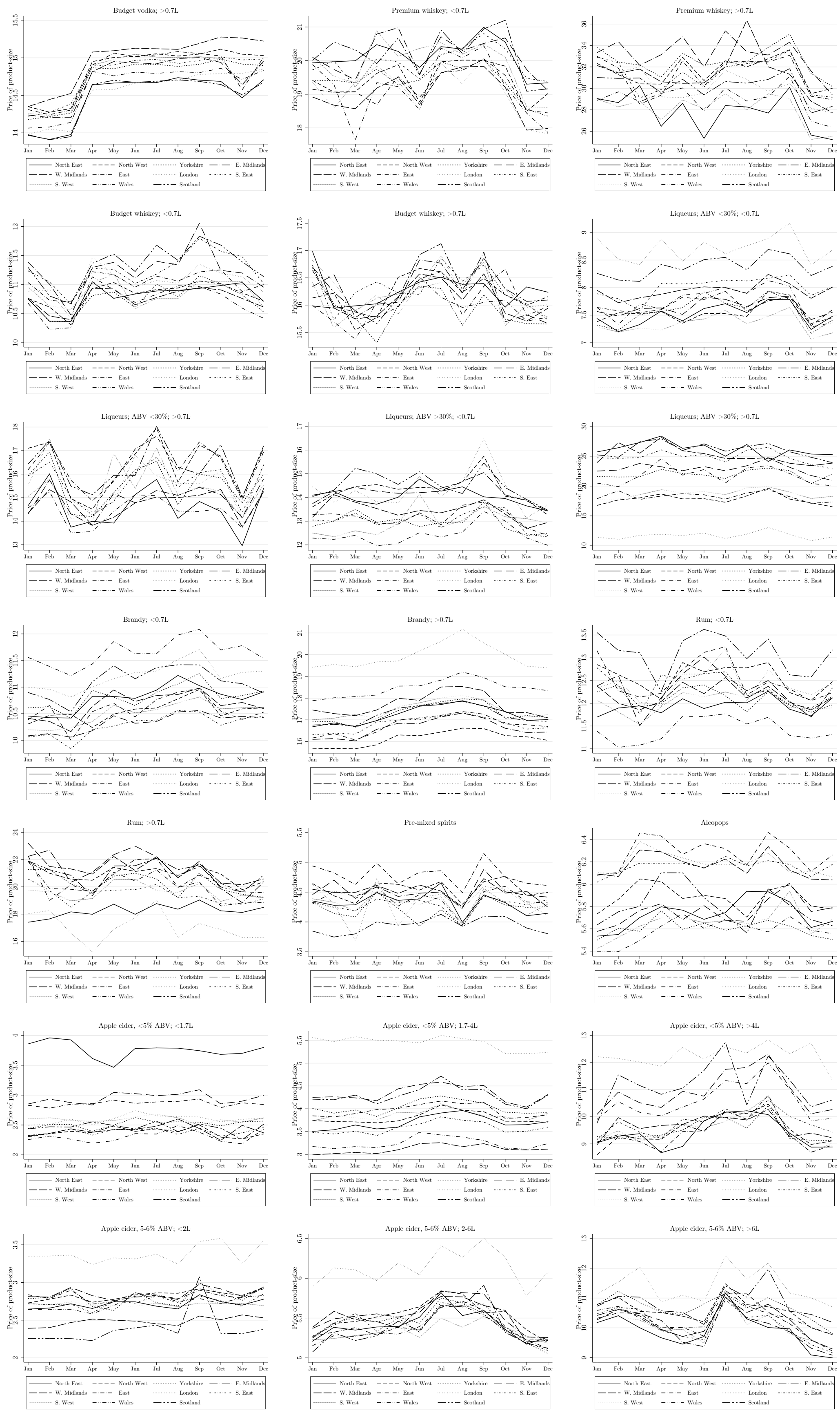
Notes: Prices are constructed as described in Section 3.3 of the paper.

Price indices for product-size pairs (cont.)



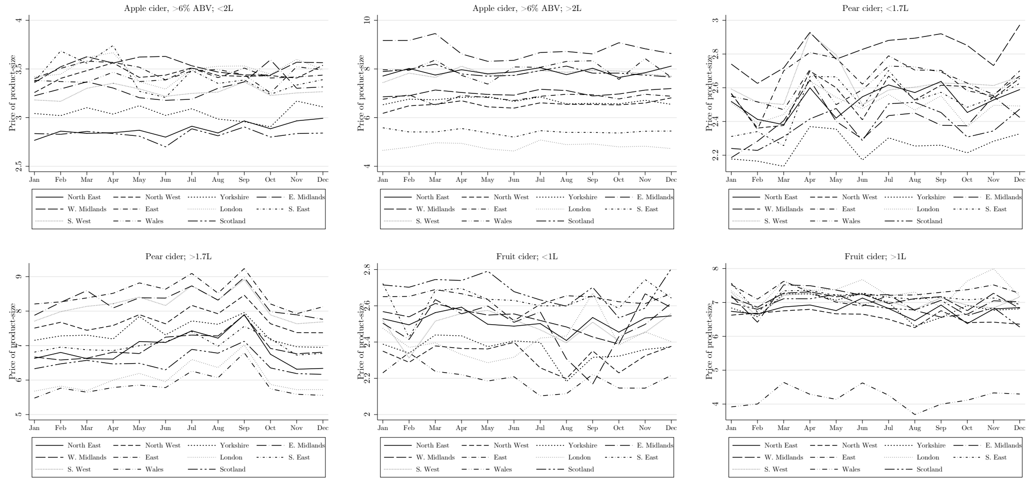
Notes: Prices are constructed as described in Section 3.3 of the paper.

Price indices for product-size pairs (cont.)



Notes: Prices are constructed as described in Section 3.3 of the paper.

Price indices for product-size pairs (cont.)



Notes: Prices are constructed as described in Section 3.3 of the paper.

B Empirical implementation

B.1 Short-run persistence and stockpiling

We allow for individual level heterogeneity in preferences and therefore statistical dependence in households' purchases, across time, through the random coefficients $\boldsymbol{\theta}_i$. However, we do not model state dependence arising, conditional on preference heterogeneity, from the effect of past purchases on current behaviour. Current choice may depend on past choices due to high frequency habit formation. It may also arise if households stockpile during sales periods (Hendel and Nevo (2006a)). We cannot categorically rule out these forms of state dependence but we can show some reduced form evidence that suggests that these forms of dynamics are not likely to be of first order importance once we take account of household level preference heterogeneity.

We test for evidence of habit formation by running two regressions. The dependent variable in the first regression is a dummy equal to one if a household purchases alcohol in a given week and the dependent variable in the second regression is, conditional on purchasing, how many standard drinks the household purchased. We regress these variables on the number of standard drinks the household purchased in each of the past eight weeks, plus week dummies. We estimate each regression both omitting household fixed effects and including them. When we omit the fixed effects, there is a moderate relationship between past behaviour and current behaviour – for instance, purchasing 1 more standard drink per adult two weeks previously is associated with an increase in the probability of purchasing alcohol of 0.36 percentage points and conditional on buying, is associated with purchasing 0.12 (or 0.9%) more standard drinks. However, once we include household fixed effects, these numbers fall to just 0.01 percentage points and 0.01 standard drinks, see Table B.1.

We also assess evidence for omitted state dependence arising from consumers stockpiling during sale periods; if short-run price reductions, such as a sale, lead to an increase in alcohol purchases, which are then stored rather than immediately consumed, this would lead us to over estimate the own price elasticities of demand (see e.g. Hendel and Nevo (2006a)). To test for such an effect we follow one of the suggestions in Hendel and Nevo (2006b). We assume that each household has a constant consumption rate (equal to their weekly average number of standard drinks purchased) and use this along with their purchases to compute an inventory for each household at the beginning of each week. We then regress: (i) the probability of purchase in a week and (ii) the number of standard drinks purchased (conditional on purchasing a positive amount) on this constructed inventory variable, week effects (which control for price changes, promotions, advertising etc.) and household fixed effects. Hendel and Nevo (2006b) argue that if stockpiling is present, then a high inventory is likely to lead to a lower probability of purchase or lower quantity purchased conditional on purchasing. In contrast, we find a very weak positive relationship between the inventory variable and both the probability and quantity of alcohol purchased (see Table B.2).

Table B.1: *Dependence of current purchase decisions on past alcohol purchases*

	(1) Purchase alcohol?	(2) Purchase alcohol?	(3) Quantity	(4) Quantity
– 1 week before	0.0025	-0.0009	0.0950	-0.0160
	(0.0001)	(0.0001)	(0.0027)	(0.0026)
– 2 weeks before	0.0036	0.0001	0.1225	0.0073
	(0.0001)	(0.0001)	(0.0027)	(0.0026)
– 3 weeks before	0.0034	-0.0000	0.1058	-0.0029
	(0.0001)	(0.0001)	(0.0027)	(0.0026)
– 4 weeks before	0.0035	0.0001	0.1121	0.0061
	(0.0001)	(0.0001)	(0.0027)	(0.0027)
– 5 weeks before	0.0032	-0.0002	0.1026	-0.0017
	(0.0001)	(0.0001)	(0.0028)	(0.0027)
– 6 weeks before	0.0031	-0.0003	0.0977	-0.0060
	(0.0001)	(0.0001)	(0.0028)	(0.0027)
– 7 weeks before	0.0031	-0.0004	0.1062	0.0015
	(0.0001)	(0.0001)	(0.0028)	(0.0027)
– 8 weeks before	0.0032	-0.0002	0.1052	0.0035
	(0.0001)	(0.0001)	(0.0028)	(0.0027)
Mean of dependent variable	0.3968	0.3968	12.3094	12.3094
Time effects?	Yes	Yes	Yes	Yes
Household fixed effects?	No	Yes	No	Yes

Notes: The dependent variable in columns (1) and (2) is a dummy equal to one if the household purchased alcohol in that week. The dependent variable in columns (3) and (4) is the number of standard drinks purchased per adult in that week, conditional on making a non-zero purchase. The table shows the estimated coefficients on the number of standard drinks purchased per adult in the preceding one, two, three, etc. weeks. Standard errors are shown in parentheses. Week effects are included, and household fixed effects are include in columns (2) and (4).

Table B.2: *Dependence of current purchase decisions on inventory*

	(1) Purchase alcohol?	(2) Purchase alcohol?
Inventory	0.0025	0.0955
	(0.0000)	(0.0010)
Mean of dependent variable	0.3968	12.3094
Time effects?	Yes	Yes
Household fixed effects?	Yes	Yes

Notes: The dependent variable in column (1) is a dummy equal to one if the household purchased alcohol in that week. The dependent variable in columns (2) is the number of standard drinks purchased per adult in that week, conditional on making a non-zero purchase. The table shows the estimated coefficients on a variable for the household's alcohol inventory. This is calculated by assuming that the household has a fixed level of consumption (equal to its mean purchases over the year). Standard errors are shown in parentheses. Week effects and household fixed effects are included in both regressions.

B.2 Control function and instruments

Prices vary over time for various reasons. In order to identify the causal impact of price on demand, we need to isolate variation in price that is driven by supply-side factors, for instance, due to changes in costs. For example, there were changes in tax rates applied to different alcohol products over our estimation period – see Figure B.3. There was also considerable variation in the EUR-GBP and USD-GBP exchange rates, see Figure B.1(a). Movements in the exchange rate are likely to affect the prices of products differentially, depending on whether they are imported directly, or use imported inputs. Figure B.1(b) shows that the factory gate prices for beer and cider changed differentially over 2011. On reason for geographical differences in prices is geographical differences in retailer coverage – see Table B.4.

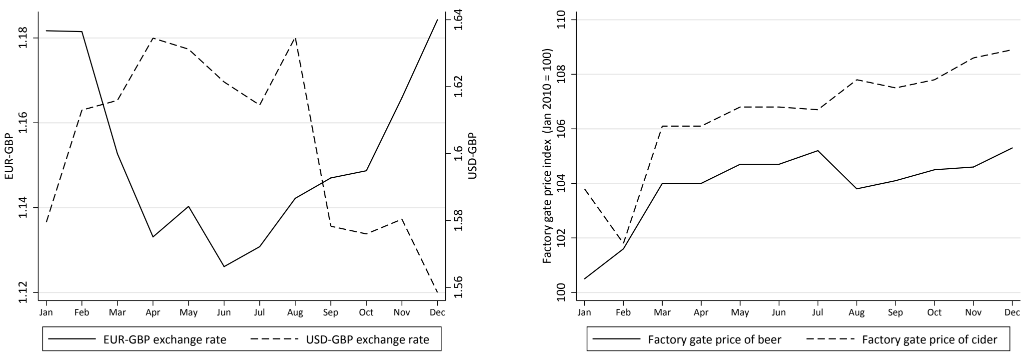
In the first stage regression our instruments are duty rates (interacted with options), exchange rates (interacted with options), beer and cider producer prices (interacted with appropriate options), regional retailer shares and oil prices (interacted with regions). The F-stat of the first stage is 17.9.

Table B.3: *Tax changes during 2011*

Segment	Applies to products:	Rate in Jan 2011:	Rate changes (month)
Beer	1.8-2.8% ABV	£17.32/litre ethanol	+1.25 (March); -9.28 (Oct)
	2.8-7.5% ABV	£17.32/litre ethanol	+1.25 (March)
	>7.5% ABV	£17.32/litre ethanol	+1.25 (March); +4.64 (Oct)
Wine	5.5-15% ABV (still)	£225.00/hectolitre product	+16.23 (March)
	15-22% ABV (still)	£299.97/hectolitre product	+21.64 (March)
	5.5-8.5% ABV (sparkling)	£217.83/hectolitre product	+15.72 (March)
	8.5-15% ABV (sparkling)	£288.20/hectolitre product	+20.79 (March)
Spirits	0-100% ABV	£23.80/litre ethanol	+1.72 (March)
Cider	1.2-7.5% ABV	£36.01/hectolitre product	-0.14 (March)
	7.5-8.5% ABV	£54.04/hectolitre product	-0.17 (March)

Figure B.1: *Exchange rates and factory gate prices, 2011*

(a) Exchange rates (b) Factory gate prices for beer and cider



Notes: Left hand panel plots the EUR-GBP and USD-GBP exchange rates over 2011 (data from Bank of England). Right hand panel plots the factory gate prices for beer and cider over 2011 (data from UK Office for National Statistics).

Table B.4: *Retailer coverage in different regions*

	Tesco	Sainsbury's	Asda	Morrisons	Discounter	Upmarket	Other
North East	19.3	10.0	27.6	18.7	5.9	2.6	16.0
North West	27.1	9.8	25.3	14.6	6.0	2.2	15.0
Yorkshire and Humber	23.3	9.9	21.3	23.3	4.8	2.5	14.9
East Midlands	29.9	13.2	17.8	15.8	5.4	2.6	15.3
West Midlands	26.0	15.6	19.6	14.8	6.2	2.8	14.9
East of England	39.3	16.5	13.6	9.6	4.2	4.4	12.4
London	31.7	26.2	11.9	7.2	3.3	6.4	13.2
South East	36.0	22.1	14.1	7.5	3.8	5.6	11.0
South West	34.5	17.4	14.9	11.4	6.0	4.2	11.5
Wales	34.0	6.8	22.1	12.1	7.7	2.3	14.9
Scotland	30.8	7.4	22.5	15.8	6.5	2.8	14.1

Notes: Numbers show the within-region grocery market share of the retailer chains listed in the first row. The discounters are Aldi and Lidl, the upmarket chains are Waitrose and Marks and Spencers, and Other are independent liquor stores.

B.3 Coefficient estimates

Table B.5 shows the coefficient estimates. The household quintiles are the five quintiles of the pre-sample ethanol per adult per week distribution. Panel A shows estimated parameters for the distribution of preferences over observable product characteristics, Panel B shows estimated parameters for the distribution of preferences over unobserved product characteristics. To estimate the model we randomly sample 500 households from each household group (with the exception of group 5, for which we use all households). For each drawn household we use 50 randomly sampled purchase occasion (or all of their purchase occasions if this is less than 50). The sampling lowers the computational burden of estimating the parameters and hence shortens estimation time (below 1 week). We conduct all post estimation analysis on the full sample.

Table B.5: *Estimated preference parameters*

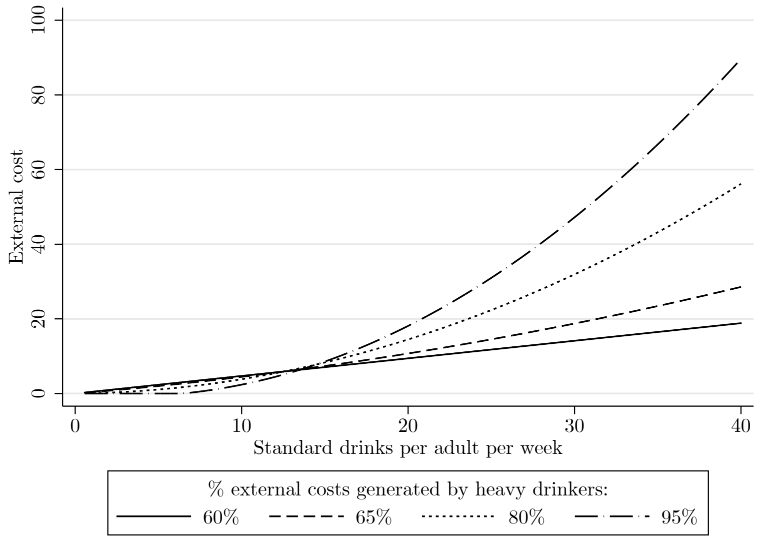
Household quintile:	1	2	3	4	5
Panel A: Preferences for observable product characteristics					
Means					
Price	-0.332	-0.256	-0.281	-0.319	-0.402
	(0.032)	(0.020)	(0.018)	(0.017)	(0.019)
Beer*Quantity of ethanol	0.140	0.163	0.163	0.193	0.240
	(0.018)	(0.012)	(0.010)	(0.010)	(0.011)
Wine*Quantity of ethanol	0.008	0.021	0.071	0.118	0.164
	(0.020)	(0.013)	(0.012)	(0.011)	(0.013)
Spirits*Quantity of ethanol	0.306	0.245	0.263	0.322	0.441
	(0.036)	(0.023)	(0.020)	(0.020)	(0.020)
Cider*Quantity of ethanol	0.069	0.088	0.123	0.179	0.193
	(0.019)	(0.012)	(0.011)	(0.011)	(0.011)
Beer*Quantity of ethanol^2	-0.002	-0.002	-0.001	-0.002	-0.002
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Wine*Quantity of ethanol^2	0.001	0.001	0.001	0.001	0.001
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Spirits*Quantity of ethanol^2	-0.004	-0.003	-0.003	-0.003	-0.004
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Cider*Quantity of ethanol^2	-0.001	-0.001	-0.002	-0.002	-0.002
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Variances ×100					
Price	1.825	1.694	2.161	0.001	1.210
	(0.211)	(0.148)	(0.171)	(0.003)	(0.114)
Quantity of ethanol	0.726	0.279	0.676	0.110	0.449
	(0.058)	(0.021)	(0.034)	(0.013)	(0.028)
Strength	0.185	0.268	0.388	0.504	0.471
	(0.023)	(0.015)	(0.019)	(0.024)	(0.023)
Covariances ×100					
Price*Quantity of ethanol	-0.921	-0.489	-1.084	0.008	-0.541
	(0.101)	(0.053)	(0.073)	(0.016)	(0.047)
Price*Alcohol strength	-0.054	-0.224	0.059	-0.006	0.167
	(0.033)	(0.029)	(0.021)	(0.012)	(0.022)
Quantity of ethanol*Alcohol strength	-0.137	-0.060	-0.239	-0.148	-0.195
	(0.026)	(0.012)	(0.016)	(0.010)	(0.018)
Panel B: Preferences for unobserved product characteristics					
Mean product effects for each segment					
Beer	-5.203	-4.524	-3.615	-3.592	-3.449
	(0.106)	(0.076)	(0.072)	(0.081)	(0.094)
Wine	-3.161	-2.786	-2.718	-1.863	-2.042
	(0.143)	(0.095)	(0.092)	(0.091)	(0.122)
Spirits	-8.221	-6.879	-6.065	-6.138	-6.730
	(0.333)	(0.240)	(0.217)	(0.215)	(0.233)
Cider	-4.749	-4.342	-4.175	-3.983	-3.265
	(0.144)	(0.106)	(0.096)	(0.106)	(0.110)
Variances					
Beer	1.994	2.273	2.126	2.593	2.151
	(0.121)	(0.121)	(0.107)	(0.137)	(0.150)
Wine	1.523	1.152	1.842	1.837	1.528
	(0.109)	(0.069)	(0.089)	(0.107)	(0.150)
Spirits	0.833	0.424	0.770	0.039	0.001
	(0.096)	(0.046)	(0.074)	(0.016)	(0.002)
Cider	2.938	5.624	2.987	4.245	2.844
	(0.244)	(0.356)	(0.196)	(0.260)	(0.186)
Product effects	Yes	Yes	Yes	Yes	Yes
Type-time effects	Yes	Yes	Yes	Yes	Yes
Control function	Yes	Yes	Yes	Yes	Yes
Number of households	500	500	500	500	351
Number of purchase occasions	21,638	22,820	23,616	23,958	16,959

Notes: Panel A shows estimated parameters for the distribution of preferences over observable product characteristics, Panel B shows estimated parameters for the distribution of preferences over unobserved product characteristics. Standard errors are reported below the coefficients.

B.4 Optimal tax results

For the optimal tax analysis we show results for when the set of heavy drinkers (the 19% of households that buy the most alcohol and together account for 60% of ethanol purchases) account for 60%, 65%, 70%, 75%, 80%, 85%, 90% and 95% of the total external costs of drinking. In Figure B.2 we plot the externality function for 60%, 65%, 80% and 95%. When the heavy drinkers account for 60% of externalities the function is linear. For any percentage above this the function is convex. For highly convex calibrations we impose a lower bound of zero on the function. This binds only at lower levels of ethanol demand.

Figure B.2: *Calibrated externality functions*



Notes: External cost is in £ per adult per week.

Table B.6 gives further details of the comparison of the UK tax system with the optimal single and multi rate systems. It breaks the welfare difference down into consumer surplus, external costs and tax revenue and for consumer surplus shows numbers for both the set of light and heavy drinkers. Moving to either optimal systems would involve reductions in consumer surplus and tax revenue. However, these would be more than made up for by reducing drinking externalities.

Table B.6: *Comparison of welfare under the optimal rates with UK system*

£bill per year unless stated	Difference with UK system for optimal: Single rate	Multi rate
Consumer surplus	-1.62	-1.30
of light drinkers (total)	-0.68	-0.47
of light drinkers (£ per household-year)	*-35.20*	*-24.28*
of heavy drinkers (total)	-0.94	-0.83
of heavy drinkers (£ per household-year)	*-140.61*	*-124.21*
External cost	-2.95	-3.11
Tax revenue	-0.10	-0.46
Social welfare	1.23	1.36

Notes: Differences are measured in £billion per year, unless stated. The numbers are shown for the calibration of the externality function under which the heavy drinkers generate 80% of the external costs of drinking. Heavy drinkers are defined as the 19% of households that buy the most ethanol (and account for 60% of total ethanol purchased)

References for Online Appendix

Health Survey for England (2016). Trend Tables: 2015. http://www.content.digital.nhs.uk/catalogue/PUB22616/HSE2015-Adult-trend-tbls.xlsx.

Hendel, I. and A. Nevo (2006a). Measuring the implications of sales and consumer inventory behavior. *Econometrica 74*(6), 1637–1673.

Hendel, I. and A. Nevo (2006b). Sales and Consumer Inventory. *RAND Journal of Economics 37*(3), 543–561.