Pre-school children's demand for sugar sweetened beverages: Evidence from stated-preference panel data

Anthony Scott, Peter Sivey, Ou Yang

Abstract

The aim of this paper is to examine the impact of price changes on children's consumption of sugar sweetened beverages. Using micro-level panel data obtained from a stated preference experiment, we specify a two-sided censoring semi-parametric demand system model with fixed effects. To overcome an estimation difficulty which is potentially a common issue to all applications studying micro-level consumption data, we propose a new consistent two-step estimation framework. The economic restrictions implied by consumption theory are imposed through a consistent and asymptotically efficient GMM estimator. We analyse the consumption behaviour of subjects through estimated expenditure and price elasticities. The partial elasticities of demand with respect to attributes of soft drinks are also examined. Our results show that the compensated own-price elasticities for Fizzy, Juice and Cordial are respectively -0.755, -0.100 and -0.811, and therefore, are all price-inelastic. All compensated cross-price elasticities are positive, suggesting net substitutes. While most average cross-drink effects of attributes are not statistically significant, we do observe on average, Fizzy being diet significantly increases its consumption by 27.8%. Being presented together with diet Cordial increases the consumption of Fizzy by 15.4% than otherwise. Healthier fizzy (no added colours or preservatives) significantly crowds out the consumption of juice. And, being present together with either diet Fizzy or Juice with no added sugar increases Cordial consumption than otherwise. To better inform policy, we also estimate our model respectively for rich and poor sub-samples, and our results highlight substantial discrepancies between their consumption behaviours

Key words: sugar sweetened beverages, consumption behaviour, panel data, demand system, censoring

1. Introduction

Consumption of sugar sweetened beverages (SSBs) exhibits strong associations with weight gain, obesity, and dental caries, especially in young children and for children of low socio-economic status (Malik, Schulze, and Hu 2006). These problems affect about one-third of children of pre-school age, with 13% of children aged 2-3 years old consuming SSBs every day (Wake et al. 2006; Dubois et al. 2007).

There are strong arguments, and numerous examples, of taxes on SSBs (Brownell and Frieden 2009). The use of taxes to improve population health is controversial. The evidence of a net welfare gain is mixed, and depends on the effects on the consumption of other foods and beverages (Sharma et al. 2014). Arguments as to whether such taxes are regressive depend on how the price elasticity of demand varies across sub-groups of the population (Sharma et al. 2014). Recent previous studies of the impact of taxation on consumption have either estimated average price elasticities (e.g. Finkelstein et al. 2013, Zhen et al. 2014, Briggs et al. 2013), or have examined heterogeneity amongst moderate and high consumers (Etilé and Sharma 2015) or different income groups (Sharma et al. 2014). Examining the impact of changes in price on high risk populations is therefore important in examining the overall effectiveness of taxation on population health.

The aim of this paper is to examine the impact of price changes on children's consumption of SSBs. We examine price and cross-price elasticities across SSBs. Usual datasets use household scanner data or aggregated data for small areas and so do not have information on the consumption of SSBs by children within households due to aggregation. Data disaggregated to below household level is generally not available. We use unique micro-data from a stated preference experiment administered to parents of children from a birth cohort study of 500 children (de Silva-Sanigorski et al. 2011). Stated preference experiments use hypothetical choices of goods to examine the impact of prices and other characteristics on choices.

Unlike stated preference discrete choice experiments which focus on choosing one good from several alternatives, our consumption experiment was designed to capture, first, the number of bottles of cordial, fruit juice and fizzy drink bought for the household, conditional on their price and other characteristics. Second, respondents were asked how many glasses of each were consumed by children in the household. This provides a continuous measure of consumption suited to analysis using a demand system approach that allows for i) the possibility that no soft drinks are consumed at all, ii) censoring (zero consumption of at least one SSB conditional on that the total consumption on all soft drinks is positive), iii) panel data (multiple scenarios per respondent). We also therefore contribute to the literature on the analysis of stated preference experiments. A particular advantage of such an experiment is that prices are presented to respondents exogenously. In addition, an experimental design is used to ensure that the variation in the attributes is orthogonal and that standard errors are minimised.

To study our unique experimental consumption data, a new semi-parametric fixed-effects censored demand system is proposed. To deal with the estimation difficulty arising from the possibility that no soft drinks are consumed at all, which is potentially a common issue to all applications studying micro-level consumption data, a two-step estimation strategy is developed. The application of such a semi-parametric fixed-effects censored demand system and two-step estimation framework is not exclusive to the experimental data of this study. It can actually be extensively applied to any micro-consumption panel data, such as household scanner data which have seen a growing number of applications in the recent consumption behaviour and marketing literature (such as Sharma et al. 2014, Andreyeva, Long, and Brownell 2010, Zhen et al. 2011, Kim, Allenby, and Rossi 2002, to just name a few), to build a demand system to examine the consumption for a bundle of very detailed goods, such as Fizzy, Juice and Cordial in this study. It is not clear in the existing literature how to deal with possible zero observations on total expenditure and this study is an attempt to fill this gap.

The plan of this paper is set out as follows. Section II introduces the consumption experiment and describes the data. Section III presents the model specification and estimation strategy. The estimation results and corresponding discussions are given in Section IV. The last section concludes this paper.

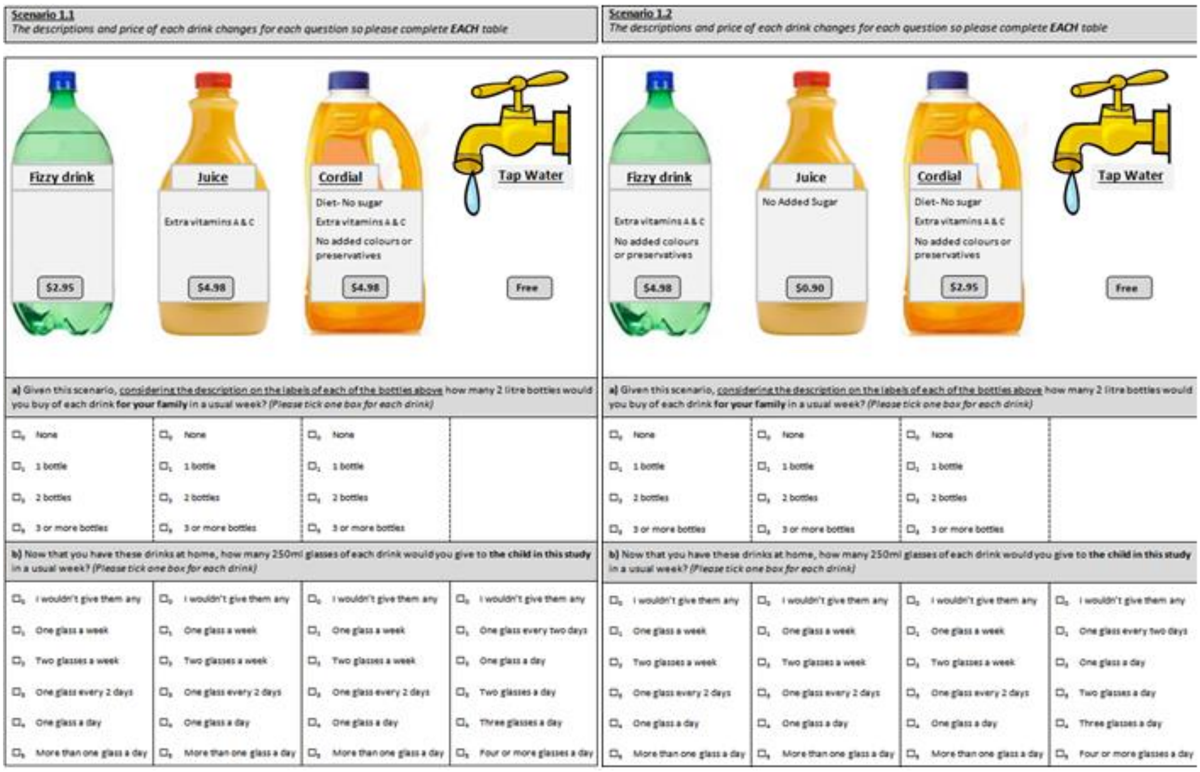
2. A consumption experiment and data

The consumption experiment (CE) consists of presenting survey respondents, the parents of 24-month-old children in the SPLASH study (de Silva-Sanigorski et al. 2011), with a series of hypothetical scenarios about the quantities of alternative drink types for their family's and children's consumption. The CE is a labelled design, where respondents choose consumption levels for four broad categories of drinks: Fizzy Drink, Juice, Cordial and Tap Water. The soft drink categories are characterised by four attributes: price, sugar content, added vitamins and no added colours or preservatives. The tap water category is not described by any attributes.

We undertook an extensive pre-piloting phase with in-depth interviews of 32 families to develop the four labelled drink categories, the attributes of the drinks, and the nature of the choice task. The pre-pilot was an iterative process, where initial designs were drafted, presented to potential respondents during interviews, and attributes and labels were refined before being presented again to potential respondents. This process broadly followed the recommendations of Coast et al. (2012) in that we avoided describing the latent construct (eg "the drink is tasty" or "the drink is healthy"), used in-depth interviews and broadly followed a constant-comparative approach to qualitative data collection and analysis. More details of the qualitative approaches used are detailed in de Silva-Sanigorski et al. (2011) and Hoare et al. (2014).

The choice context, attributes and levels were informed by three considerations. Firstly, some attributes were of particular policy interest, including price and sugar content of drinks. Secondly, we conducted an investigation of the websites of major Australian supermarket chains. This was a key step as it enabled us to ensure the hypothetical choices were as close as possible to real-world choices that parents would be making whilst shopping for drinks. Thirdly, all of our decisions were informed, verified and modified from the iterative process of the qualitative interviews.

Figure 1 Two examples of the shopping scenarios in the consumption experiment



To be a bit more specific, this consumption experiment features soft-drink consumption related questions across several hypothetical shopping scenarios. Figure 1 illustrates two examples of these hypothetical scenarios. As shown in the examples, the consumption experiment is set in the context of the main 'family shop' (e.g. Saturday shop in a supermarket). It was recognised in qualitative work that young children's drink consumption was particular to context and was particularly idiosyncratic out of the household (on trips or visiting friends and family) and on special occasions (Hoare et al. 2014), however it would be difficult to model consumption in all of these alternative contexts comprehensively. The regular family shop provides a well-understood context which accounts for a large proportion of a child's drink intake.

The design takes into account that the supermarket shop typically involves a choice of drinks for the family, not just for the child. So, for example, a large bottle of juice could be bought with the intention of providing drinks for adults and older children in the household as well as for young children. For this reason we ask responding parents to make two sequential consumption choices in each scenario: first they must decide how many bottles of each soft drink to buy for the week for the whole family; secondly, they must decide how many glasses of each drink they would give to their young child to drink for the week.

The four categories of drinks (fizzy drink, juice, cordial and tap water) were chosen as the most common broad categories of drinks given to young children. A decision was made early to exclude milk and milk-based drinks as they form a separate category of drinks which can be consumed for nutritional reasons. Tap water is included as a labelled drink category but is not described by the attributes. We assume tap water is regarded as free of charge and homogeneous to the families. The other three drink types can be described by all four attributes: price, sugar content, added vitamins, and no added colours or preservatives.

When choosing consumption levels for the family for each drink category, the respondent chooses the number of bottles of drink. We specify two-litre sized bottles as informed by the investigation of common Australian supermarket websites. For the choice of consumption for the child, we specify 250 millilitre glasses, for all four drink categories, including water.

Price is a key determinant of choice, displayed prominently in supermarkets, mentioned by interviewees as determining their choice and is of policy and academic interest. The three price levels chosen, $0.90, $2.95, and $4.98 per two litre bottle were designed to cover the full range of prices encountered in supermarkets.

The sugar content attribute is another key policy attribute in the study. The attribute has only two levels, ‘Diet-No Sugar’ or blank, implying ‘with sugar’. We chose this wording to match real-life labelling of drinks, ‘Diet’ or ‘No Sugar’ or very similar variants were used on the packaging of sugar free drinks, whereas highly sugar sweetened drinks were not labelled with regard to sugar content. One exception to this wording was for the 'Juice' drinks category, for which we used the wording 'No added sugar' instead of 'Diet-No Sugar', again matching the labelling most often used in supermarkets. The final two binary attributes represent common health claims made by soft drink labels: "Extra vitamins A and C" and "No added colours or preservatives". Each of these attributes is blank when there are no extra vitamins or when there may be added colours or preservatives. Also, as you can see in Figure 1, the attributes and prices of the three soft drinks vary across scenarios. If a drink boasts a certain attribute, for instance if in a scenario the fizzy drink features extra vitamins, then the extra vitamins label is shown on the bottle; otherwise, it is blank. This is exactly what happens in the real world, because most retailers tend to only advertise good aspects of their commodities.

In sum, the experiment consists of five attributes, three (sugar, vitamins and colour or preservative attributes) with two levels and two (soft drink type and price) with three levels giving $2^3 \times 3^2 = 72$ possible alternatives. Balancing statistical efficiency considerations, practical survey issues and minimising the demands on respondents, we choose to present each respondent with 9 shopping scenarios of three choice alternatives (juice, fizzy drink, cordial) and the tap water alternative. We produce four versions of the survey, allowing us to include 9x3x4=108 alternatives in total in four versions of the survey (each with 27 alternatives).

Our approach to producing the experimental design following the pre-piloting stage is in two stages: First we conduct a pilot study, estimating simple models with the data obtained. Secondly, we use the results from the pilot study as 'priors' to inform the design of the questionnaire in our main study in order to maximise statistical efficiency. This general approach follows the recent literature in stated preference discrete choice experiments (e.g.

Scott et al. 2013 and Sivey et al. 2012 are recent examples; Huber and Zwerina 1996 and Carlsson and Martinsson 2003 were seminal application in marketing and in health economics).

The experimental design is based on a multinomial logit model of drink choice. We specify a linear indirect utility function in which all attributes enter utility in an additive and separable manner. An initial orthogonal design was piloted among 35 responding families (giving 314 observations). The data from this pilot were analysed using a simple multinomial logit. We generated the final design by minimizing the D-error in the multinomial logit model, using the prior values from the pilot study results, to produce 36 choice scenarios of three choices (fizzy drink, juice, cordial) across the four versions. Each choice set in the survey also contained the 'tap water' choice alternative. Respondents were randomly allocated to one of the four blocks of choice sets in the questionnaire.

In the final panel data set, there are 282 parents whose consumption choices for their pre-school children are observed for 9 hypothetical shopping scenarios which are different in terms of attributes and prices of soft drinks. Hence, there are in total 2538 observations in our sample when laid out as one long cross-section, 78 of which contain missing values and are excluded from the sample.

3. Model Specification and Two-step Estimation Strategy

3.1 Model specification

This paper specifies a demand system to jointly study the soft-drink consumption data gathered from the SPLASH consumption experiment. In particular, we are interested in how changes in prices and attributes would affect parents' consumption decisions for their children. The demand system literature features modelling the demand of goods in budget share form on goodness of fit grounds, which also helps avoid heteroscedasticity (Leser 1963). The obvious difficulty with using budget share form is that for respondents who do not give their children any of the three soft drinks their total expenditure on soft drinks is zero. Budget shares are not defined or defined as missing values. As a matter of fact, it’s not just because a demand system is specified in the form of budget shares; rather, the neoclassical consumption theoretical framework is based on positive total budget constraint. Hence, a budget allocation analysis framework should not include observations with zero total expenditure.

Conditioning on positive total expenditure makes perfect technical sense in the macro aggregate consumption world, in that aggregate expenditure on any good in the budget is always positive and thereby, the total expenditure must also be positive. An enormous amount of previous exceptional studies were based on aggregate demand data (for example, Deaton and Muellbauer 1980a, Manser and McDonald 1988, Varian 1983, Christensen, Jorgenson, and Lau 1975, and Gallant 1981). Blundell, Pashardes, and Weber (1993) concluded that unless certain factors are controlled, aggregate data alone unlikely produce reliable estimates of structural price and income coefficients.

When it comes to micro individual-level consumption world, zero expenditures for certain goods, or even for the whole category of such goods as soft drinks, meat, etc., are certainly possible. Then, if a weakly separably preference is assumed as in, for example, Hoderlein and Mihaleva (2008), Chalfant (1987) and Lewbel (1989), the zero observations on total soft-drink expenditure could be considered as a result of the first-stage budget allocation problem in a multi-stage budgeting framework (Deaton and Muellbauer 1980a, Deaton and Muellbauer 1980b, Edgerton 1997). In particular, assuming weak separability between soft drinks and other alternative commercial drinks and foods, such as milk-based drinks, the subjects first make decisions on whether they would like to give some soft drinks to their children and then, how much to give them. If they have decided to give some soft drinks to their children, i.e. the total expenditure on total soft drinks for their children is positive, they proceed to the second stage to make decisions about how to allocate the total budget on soft drinks among the three drinks considered. It is noteworthy that at this second stage, it is also possible that respondents do not choose certain drinks for their children. As a result, although they are recorded as zeros in expenditure form when pooled together, some of these zeros might come from a different generating process than the others.

Table 1 and Table 2 summarise the number of total, positive and zero observations, apart from sample means and standard deviations for total expenditure on soft drinks and for shares of the three soft drinks considered. As shown in Table 1, for each of the four variables, i.e. total expenditure, fizzy share, juice share and cordial share, there are a substantial amount of zero observations, which requires serious consideration in the econometric analysis.

Table 1 Number of positive and zero observations for total soft-drink expenditure for children and shares of soft drinks

	Total obs.	Positive obs.	Zero obs.
Total expenditure	2460 (100%)	1232 (50.08%)	1228 (49.92%)
Fizzy share	1232 (100%)	214 (17.37%)	1018 (82.63%)
Juice share	1232 (100%)	1026 (83.28%)	206 (16.72%)
Cordial share	1232 (100%)	463 (37.58%)	769 (62.42%)

Note: observation is abbreviated to obs.

Table 2 Summary of total expenditure and shares of soft drinks for total, positive and zero observations

	Total obs.		Positive obs.	
	Mean	SD	Mean	SD
Total expenditure	0.692	(1.356)	1.382	(1.649)
Fizzy share	0.096	(0.255)	0.554	(0.350)
Juice share	0.745	(0.381)	0.895	(0.200)
Cordial share	0.159	(0.315)	0.422	(0.391)

Note: observation is abbreviated to obs. Standard deviation are given in parentheses.

In the literature, the two principal reasons for zero expenditures in microeconomic expenditure data are consumers at a corner solution for the commodity in question (Wales and Woodland 1983), and limited survey periods leading to infrequency of purchase (Deaton and Irish 1984). To our knowledge, most of the econometric techniques in the literature are developed to model economics non-consumption (for example Yen and Lin 2006, Meyerhoefer, Ranney, and Sahn 2005, Yen 2005, Perali and Chavas 2000, Heien and Wessells 1990). The only exception is Deaton and Irish (1984). Since our panel data come from a consumption experiment, it is admissible to assume that the zero expenditure observations, in our case, represent a genuine corner solution where the subjects deliberately choose not to consume particular soft drinks conditional on the attributes of the soft drinks in each scenario.

This study employs a fixed-effects censored demand system analysis framework, to account for the reported zero expenditure observations on certain soft drinks (i.e. choose to or not to give their children certain soft drinks). The fixed-effects censored demand system model is estimated using the micro-level panel data collected from the SPLASH consumption experiments. With the increasing availability of micro-data, the use of such individual-level data is preferable, since it avoids the problem of aggregation over individuals and often provides a large and statistically rich sample (Heien and Wessells 1990). As opposed to other comparable household-level consumption data, such as ACNielsen Australian Household Scanner Panel data and Australia Household Expenditure Survey data, another remarkable advantage of our data is that they give information on the consumption of a couple of popular sugar sweetened beverages by children within households.

Much of the recent empirical effort on censored demand system has been concerned with circumventing the “curse of dimensionality” associated with the theoretically consistent models proposed by Wales and Woodland (1983) and Lee and Pitt (1986, 1987). For example, Heien and Wessells (1990), Shonkwiler and Yen (1999) and Yen, Kan, and Su (2002) adopt a two-step procedure to reduce the computational burden from using a full information maximum likelihood estimator. Nonetheless, Arndt, Liu, and Preckel (1999) claimed that this procedure and its application to corner solutions are unable to account for the role of reservation prices. Instead, Arndt (1999) proposed to address this difficulty using maximum entropy (ME) techniques, and generate a simpler framework for the imposition of regularity conditions. However, the fact that the asymptotic properties of this estimator are not well understood in nonlinear applications limits its feasibility.

More recently, Perali and Chavas (2000) suggest a consistent multi-step approach to the problem, which involves single-equation tobit estimation of unrestricted demand parameters and minimum chi-square estimation to recover restricted demand parameters. Yen, Lin, and Smallwood (2003) use the simulation technique, as well as a quasi-maximum likelihood procedure, to facilitate the estimation of a censored demand system based on the Amemiya-Tobin general model structure. Yen and Lin (2006) adopt a sample selection approach to estimating a system involving a small number of commodities using a full information maximum likelihood estimator. While all of the above studies provide an approach to obtaining consistent estimates of disaggregated demand models, they are designed for cross-sectional data and thereby, they suffer from limited ability to control for heterogeneous preferences and limited variation in real price. To the best of our knowledge, Meyerhoefer, Ranney, and Sahn (2005) is the only work which extends this literature to the context of panel data. They proposed a consistent GMM estimation framework for censored demand system applications using panel data, and controlled for unobserved heterogeneity using a correlated random-effects specification.

Given the panel structure of our micro-level data, it seems natural for us to follow Meyerhoefer, Ranney, and Sahn (2005)'s estimation strategy. However, one technical difficulty is that built upon the neoclassical budget allocation framework, a general flexible demand system analysis model, such as AIDS and QUAIDS, requires positive expenditure to be observed for at least one of the three soft drinks; in other words, as discussed, subjects' total expenditure on all the three soft drinks has to be positive. Even though Meyerhoefer, Ranney, and Sahn (2005)'s censored demand system model is able to handle zero expenditure observations for certain goods, if a subject is observed to have purchased nothing, this observation has to be excluded from the estimation. This can also be easily seen from the use of logarithm of total expenditure on the right-hand side of the system specification as an explanatory variable. On the other hand, estimating average expenditure elasticity across the sample also involves evaluating the logarithm of total expenditure for each observation.

As shown in Table 1, 50.08% of the total 2460 non-missing observations have zero total expenditure on soft drinks. Employing Meyerhoefer, Ranney, and Sahn (2005)'s correlated random-effects censored demand system analysis framework will exclude these observations from estimation, which one might find similar to an incidental truncation problem. If a subject's decision about whether or not to give their children any soft drink is not systematically correlated to their decision about how much of each soft drink to give to their children, estimates conditional on the truncated sample (or equivalently, conditional on positive total expenditure on soft drinks) are still consistent; otherwise, a sample selection bias might result. Accordingly, a proper statistical test for this potential selection bias is needed.

Before proceeding to carrying out a statistical test for selection bias in the current context, we first introduce the share equations for a censored demand system model whereby price and expenditure elasticities can be estimated. Conditional on positive total expenditure on soft drinks, the subject makes decisions on how to allocate the total expenditure among individual soft drinks in scenarios given the price and attributes of each drink. In accordance with neoclassical consumption theory, assuming weakly separable preference, define the conditional direct utility function as $U(q_{jt}; d_{1jt}, \ldots, d_{Ljt}, \varphi_j)$, where t $(=1, \ldots, T)$ indexes scenarios, j $(=1, \ldots, J)$ denotes subjects or decision makers, $q_{jt} = (q_{1jt}, \ldots, q_{Kjt})'$ is a vector containing subject j's consumption levels for the kth soft drink in scenario t, d_{ljt} denotes the realisation of the l th $(=1, \ldots, L)$ attribute for subject j $(=1, \ldots, J)$ at scenario t $(=1, \ldots, T)$, and φ_j is a time invariant individual specific effect representing unobserved heterogeneity across subjects.

It is assumed that $U(.)$ represents a preference ordering of the PIGLOG form. Then, according to duality theory (Deaton and Muellbauer 1980b), the indirect utility function corresponding to Deaton and Muellbauer (1980a) can be specified as:

$$V_{jt}^{*} = \frac{\log c_{jt} - \alpha_0 - \sum_k \alpha_k \log p_{kjt} - \sum_k \sum_l \lambda_{kl} \log p_{kjt} d_{ljt} - \frac{1}{2}\sum_k \sum_i \tilde{\gamma}_{ki} \log p_{kjt} \log p_{ijt} - \sum_k \psi_k \log p_{kjt} \varphi_j}{\beta_0 \prod_k p_{kjt}^{\beta_k}} \tag{3.1}$$

where $\log c_{jt}$ represents the total expenditure on soft drinks at scenario t for household j and p_{kjt} denotes the price of soft drink k observed at scenario t by subject j.

The attributes of soft drinks and the individual specific effects are embedded into the demand model following a procedure named "demographic translating". This procedure is very general in the sense that the demographically extended demand system are still theoretically plausible, if the initial demand system is theoretically plausible (Pollak and Wales 1981, Pollak and Wales 1992).

Demand equations are conventionally represented in share form, to be more consistent with an assumption of homoscedasticity and to remove dependence on the numeraire (Fry, Fry, and McLaren 1996). Applying the logarithm version of Roy's Identity, the deterministic Marshallian uncompensated demand share equations of the demographically extended Almost Ideal Demand System (AIDS) can be obtained. To estimate the system of share equations, appending stochastic error terms gives rise to the econometric specification shown as follows:

$$w_{njt}^{*} = \alpha_n + \sum_l \lambda_{nl} d_{ljt} + \sum_k \gamma_{nk} \log p_{kjt} + \beta_n (\log c_{jt} - \log P_{jt}) + \rho_{nj} + u_{njt} \tag{3.2}$$

where w_{njt}^{*} is the expenditure share of soft drink $n\ (=1,\ldots K)$ at scenario t for household j,

$$\log P_{jt} = \alpha_0 + \sum_k \alpha_k \log p_{kjt} + \sum_k \sum_l \lambda_{kl} \log p_{kjt} d_{ljt} + \frac{1}{2}\sum_k \sum_i \gamma_{ki} \log p_{kjt} \log p_{ijt} + \sum_k \psi_k \log p_{kjt} \varphi_j,$$

and $\gamma_{ki} = \frac{1}{2}(\tilde{\gamma}_{ki} + \tilde{\gamma}_{ik})$. u_{njt} is an error term.

In order to linearize the above budget share equation and circumvent the problem that incorporating demand shifters in the intercepts renders the AIDS model invariant to units of measurement (Alston, Chalfant, and Piggott 2001). One way to solve the problem is to use a scale-invariant log-linear Laspeyres index, $\log P_{jt}^{S} = \sum_k w_k^o \log p_{kjt}$ where w_k^o is the mean share for soft drink k across all the subjects and all the scenarios, to replace $\log P_{jt}$ in the AIDS model, which has been shown by Moschini (1995) and Buse (1998) to have good approximation properties. This new price index can also reduce the potential for severe multicollinearity problem while reducing the burden of estimation. Homogeneity and symmetry restrictions implied from consumption theory can be imposed on the demand equations through restrictions on certain parameters as follows: $\sum_k \gamma_{ik} = 0$ and $\gamma_{ki} = \gamma_{ik}$.

The adding-up condition is not imposed *a priori*, because although the observed budget shares add up to one, the latent shares need not, which remains an issue yet to be resolved in this literature, and no attempt is made in this study to formally deal with this difficulty. Consequently, following the previous studies (for instance Meyerhoefer, Ranney, and Sahn 2005 and Perali and Chavas 2000), adding up is not imposed on the structural parameter estimates. It should be noted that this practice should have limited impact on the price coefficients since imposing both symmetry and homogeneity restrictions implies the γ's sum to zero across equations by default.

The share equations in (3.2) can be regarded as latent share equations (Wales and Woodland 1983). In reality, demand shares are bounded between zero and unity. Thus, observed shares w_{njt} relate to latent shares w_{njt}^* such that

$$w_{njt} = \begin{cases} 0 & \text{if} \quad w_{njt}^* < 0 \\ w_{njt}^* & \text{if} \quad 0 \le w_{njt}^* \le 1 \\ 1 & \text{if} \quad w_{njt}^* > 1 \end{cases},$$

From (3.2), it can be clearly seen that any observation with total expenditure, c_{jt}, being zero will be excluded from the estimation.

3.2 A variable addition test for selection bias

To test the significance of the potential sample selection bias, a variable addition test, similar in spirit to Wooldridge's (1995) variable addition tests for selection bias (also see Wooldridge 2010a), is proposed and applied in this study. In particular, we specify the selection mechanism as an equation of the Tobit form, as follows:

$$\begin{aligned} c_{jt}^* &= \alpha_0 + \sum_l \alpha_l d_{ljt} + \sum_k \beta_k \log p_{kjt} + \theta Income_j + \gamma q_{jt}^w + \eta_j + \varepsilon_{jt} \\ c_{jt} &= \max(0,\ c_{jt}^*) \end{aligned} \tag{3.3}$$

where, denoting observable explanatory variables as x_{jt}^{s*} and letting $\tilde{x}_j^{s*} \equiv \left(x_{j1}^{s*}, \ldots, x_{jT}^{s*} \right)'$, ε_{jt} is assumed to be independent of $\tilde{x}_j^{s*}$. $Income_j$ represents respondent j's total household income. Combining the latent equations in (3.2) and (3.3), for each soft drink $n\ (=1,\ldots K)$ introduces the following fixed-effects selection system:

$$w_{njt}^* = \alpha_n + \sum_l \lambda_{nl} d_{ljt} + \sum_k \gamma_{nk} \log p_{kjt} + \beta_n (\log c_{jt} - \log P_{jt}) + \rho_{nj} + u_{njt} \tag{3.4}$$

$$c_{jt}^* = \alpha_0 + \sum_l \alpha_l d_{ljt} + \sum_k \beta_k \log p_{kjt} + \theta Income_j + \gamma q_{jt}^w + \eta_j + \varepsilon_{jt} \tag{3.5}$$

Since the unobservable individual specific effect η_j in (3.5) is expected to be correlated with individual tap water consumption, using a Mundlak-type model (Mundlak 1978), this correlation can be modelled as a linear projection of η_j on the average tap water consumption across all the scenarios, denoted by $\overline{q_j^w}$:

$$\eta_j = \tau_1 \overline{q_j^w} + \nu_j \tag{3.6}$$

where ν_j is assumed to be independent of $\tilde{x}_j^{s*}$ with a zero mean normal distribution. Substituting in η_j , the selection equation (3.5) can be written as:

$$c_{jt}^* = \alpha_0 + \sum_l \alpha_l d_{ljt} + \gamma q_{jt}^w + \sum_k \beta_k \log p_{kjt} + \theta Income_j + \tau_1 \overline{q_j^w} + \xi_{jt} \tag{3.7}$$

where $\xi_{jt} \equiv \nu_j + \varepsilon_{jt}$, and $\xi_{jt} \sim N\left(0,\ \sigma_\xi^2\right)$. Assuming a weakly separable preference, the model (3.3) might be considered as a reduced form of a first/upper-stage budget allocation problem. It should also be mentioned that this test is under the assumption that the latent variable determining selection can be observed whenever it is nonnegative, but for the purpose of test, the selection mechanism does not have to be correctly specified in any sense, as it simply serves as a vehicle for obtaining a valid test (Wooldridge 1995)

If there is no selectivity bias, since w_{njt}^* in (3.4) is only partially observed, a normal linear fixed-effects estimation strategy for (3.4) still produces inconsistent estimates. Alan et al. (2014)'s semi-parametric estimator for two-sided censoring models with fixed effects is employed. Denote all the observable explanatory variables in (3.4) as x_{jt} and let $\tilde{x}_j \equiv \left(x_{j1}, \ldots, x_{jT_j}\right)'$ and $\tilde{\xi}_j \equiv \left(\xi_{j1}, \ldots \xi_{jT}\right)'$. Under the assumption that for any n , u_{njt} is identically distributed conditional on $\left(\rho_{nj}, \nu_j, \tilde{x}_j, \tilde{\xi}_j\right)$, the semi-parametric estimator conditional on $c_{jt}^* > 0$ is consistent and asymptotically normal (Alan et al. 2014). A necessary condition of this assumption is $E\left(u_{njt} \mid \rho_{nj}, v_j, \tilde{x}_j, \tilde{\xi}_j\right) = 0$. This also suggests a useful alternative that implies selectivity bias. The simplest such alternative is

$$E\left(u_{njt} \mid \rho_{nj}, v_j, \tilde{x}_j, \tilde{\xi}_j\right) = \theta_n \varepsilon_{jt} = \theta_n \left(\xi_{jt} - v_j\right), \qquad t = 1, 2, \ldots, T,$$

$$\omega_{njt} \equiv u_{njt} - \theta_n \left(\xi_{jt} - v_j\right) \text{ is identically distributed conditional on } \left(\rho_{nj}, v_j, \tilde{x}_j, \tilde{\xi}_j\right) \tag{3.8}$$

for some unknown scalar θ_n .

Under the alternative (3.8), we have

$$\begin{aligned} w^*_{njt} &= \alpha_n + \sum_l \lambda_{nl} d_{ljt} + \sum_k \gamma_{nk} \log p_{kjt} + \beta_n (\log c_{jt} - \log P_{jt}) + \rho_{nj} + \theta_n \left(\xi_{jt} - v_j\right) + \omega_{njt} \\ &= \alpha_n + \sum_l \lambda_{nl} d_{ljt} + \sum_k \gamma_{nk} \log p_{kjt} + \beta_n (\log c_{jt} - \log P_{jt}) + \theta_n \xi_{jt} + \sigma_{nj} + \omega_{njt} \end{aligned} \tag{3.9}$$

where $\sigma_{nj} \equiv \rho_{nj} - \theta_n v_j$. From (3.9), it follows that if we could observe ξ_{jt} , when $c^*_{jt} > 0$, then we could test the null hypothesis by including the ξ_{jt} as an additional regressor in the semi-parametric fixed-effects estimation and testing H_0 : $\theta_n = 0$ using standard methods. While ξ_{jt} is not observable, it can be estimated whenever $c^*_{jt} > 0$ because ξ_{jt} is simply the error in a Tobit model. Therefore, the following test for selection bias when $c_{jt} > 0$ is proposed:

Step 1: Estimate the equation (3.7) by pooled Tobit.

Step 2: When $c_{jt} > 0$, calculate the Tobit residuals:

$$\hat{\xi}_{jt} = c_{jt} - \left(\hat{\alpha}_0 + \sum_l \hat{\alpha}_l d_{ljt} + \hat{\gamma} q^w_{jt} + \sum_k \hat{\beta}_k \log p_{kjt} + \hat{\tau}_1 \overline{q^w_j} + \hat{\tau}_2 Income_j \right) \tag{3.10}$$

Step 3: Estimate the equation

$$
\begin{aligned}
w_{njt}^{*} = {} & \alpha_n + \sum_l \lambda_{nl} d_{ljt} + \sum_k \gamma_{nk} \log p_{kjt} + \beta_n (\log c_{jt} - \log P_{jt}) \\
& + \theta_n \hat{\xi}_{jt} + \sigma_{nj} + \omega_{njt},
\end{aligned}
\tag{3.11}
$$

using those observations for which $c_{jt} > 0$.

Step4: Test $\mathrm{H}_0:\ \theta_n = 0$ using the standard error of $\hat{\theta}_n$.

As mentioned above, w_{njt}^{*} is only partially observed, the normal linear fixed-effects estimation produces inconsistent estimates. This study employs Alan et al. (2014)'s consistent semi-parametric estimator for two-sided censoring models with fixed-effects to estimate parameters in equation (3.11). In particular, let δ_n denote coefficients in (3.11) to be estimated and x_{jt} denote the vector of all the observed explanatory variables in (3.11) exculding $\hat{\xi}_{jt}$ and let $\tilde{x}_j \equiv \left(x_{j1}, \ldots, x_{jT_j}\right)'$ and $\tilde{\xi}_j \equiv \left(\xi_{j1}, \ldots \xi_{jT}\right)'$. Under the hypothesis that ω_{njt} is identically distributed conditional on $\left(\rho_{nj}, v_j, \tilde{x}_j, \tilde{\xi}_j\right)$, δ_n can be consistently estimated Alan et al. (2014)'s semi-parametric estimator, and

$$
\hat{\delta}_n = \arg\min_{\delta} \sum_{j=1}^{J} \sum_{1<s<t<T_j} \frac{1}{T_j} U\left(w_{njt}, w_{njs}, \begin{pmatrix} x_{jt} - x_{js} \\ \hat{\xi}_{jt} - \hat{\xi}_{js} \end{pmatrix}' \delta \right)
\tag{3.12}
$$

where

$$
U\left(y_1, \mathrm{y}_2, d\right) \begin{cases}
1 + 2c_1 + c_1^2 - 2c_3 c_1 + 2c_3 c_2 + \left(y_1 - y_2 - c_2\right)^2 & \text{for } d < -1 \\
-2d - d^2 + 2c_1 + c_1^2 - 2c_3 c_1 + 2c_3 c_2 + \left(y_1 - y_2 - c_2\right)^2 & \text{for } -1 \le d < c_1 \\
-2c_3 d + 2c_3 c_2 + \left(y_1 - y_2 - c_2\right)^2 & \text{for } c_1 \le d < c_2 \\
\left(y_1 - y_2 - d\right)^2 & \text{for } c_2 \le d < c_3 \\
-2c_2 d + 2c_2 c_3 + \left(y_1 - y_2 - c_3\right)^2 & \text{for } c_3 \le d < c_4 \\
-d^2 + 2d + c_4^2 - 2c_4 - 2c_2 c_4 + 2c_2 c_3 + \left(y_1 - y_2 - c_3\right)^2 & \text{for } c_4 \le d < 1 \\
1 + c_4^2 - 2c_4 - 2c_2 c_4 + 2c_2 c_3 + \left(y_1 - y_2 - c_3\right)^2 & \text{for } d \ge 1
\end{cases}
$$

and

$c_1 = \min\{-y_2, y_1 - 1\}$, $c_2 = \max\{-y_2, y_1 - 1\}$, $c_3 = \min\{1 - y_2, y_1\}$ and $c_4 = \max\{1 - y_2, y_1\}$.

The rationale behind this estimator is that for example, if $E(\varepsilon x) = 0$, then one has the moment conditions $E\left[(y^* - x'\beta)x\right] = 0$, where y^* denotes the latent variable. However, with censoring, $y - x'\beta$ will not have the same properties as ε. The idea employed in Alan et al. (2014), and some others such as Powell (1986), Honoré (1992) and Honoré and Powell (1994), is to apply additional censoring to $y - x'\beta$ in such a manner that the resulting re-censored residual satisfies the conditions assumed on ε. The minimisation problem (3.12) has as first-order condition the sample analogue of moment conditions as follows:

$$E\left[\sum_{1<s<t<T_j} \frac{1}{T_j} u\left(w_{njt}, \mathrm{w}_{njs}, \Delta x_j{}'\delta_n\right)\Delta x_j\right] = 0 \tag{3.13}$$

where $\Delta x_j \equiv x_{jt} - x_{js}$

and

$$u(y_1, \mathrm{y}_2, d) = \begin{cases} 0 & \text{for } d < -1 \\ 1 + d & \text{for } -1 \le d < c_1 \\ \min\{1 - y_2,\ y_1\} & \text{for } c_1 \le d < c_2 \\ y_1 - y_2 - d & \text{for } c_2 \le d < c_3 \\ \max\{y_1 - 1,\ -y_2\} & \text{for } c_3 \le d < c_4 \\ d - 1 & \text{for } c_4 \le d < 1 \\ 0 & \text{for } d \ge 1 \end{cases}$$

Under H_0: $\theta_n = 0$,

$$\sqrt{J}\left(\hat{\delta}_n - \delta_n\right) \xrightarrow{d} N\left(0, \Gamma^{-1} S \Gamma^{-1}\right) \tag{3.14}$$

where Γ and V are consistently estimated as follows (Alan et al. 2014):

$$\hat{\Gamma} = \frac{1}{J}\sum_{j=1}^{J}\left[\sum_{s<t}\frac{1}{T_j}1\left\{-1<\left(x_{js}-x_{jt}\right)'\hat{\delta}_n<1\right\}\begin{pmatrix}1\left\{-1<\left(x_{js}-x_{jt}\right)'\hat{\delta}_n<w_{js}-1\right\}- \\ 1\left\{0<\left(x_{js}-x_{jt}\right)'\hat{\delta}_n<w_{js}\right\}- \\ 1\left\{-w_{jt}<\left(x_{js}-x_{jt}\right)'\hat{\delta}_n<0\right\}+ \\ 1\left\{1-w_{jt}<\left(x_{js}-x_{jt}\right)'\hat{\delta}_n<1\right\}\end{pmatrix}\left(x_{js}-x_{jt}\right)\left(x_{js}-x_{jt}\right)'\right] \tag{3.15}$$

and

$$\hat{S} = \frac{1}{J}\sum_{j=1}^{J}\hat{s}_j\hat{s}_j{}'$$

with

$$\hat{s}_j = \sum_{s<t}\frac{1}{T_j}u\left(w_{js}, w_{jt}, \left(x_{js}-x_{jt}\right)'\hat{\delta}_n\right)\left(x_{js}-x_{jt}\right).$$

3.3 Correcting for sample selection bias

In cases where the null hypothesis is rejected, the model has to be corrected for selection bias. To correct for selection bias, we need to formalise the selection mechanism and the assumption about the relationship among ρ_{nj}, u_{njt} and ξ_{jt}. We first formalise the selection mechanism.

Assumption 3.3.1:

Denote observable explanatory variables in (3.7) as x_{jt}^s and let $\tilde{x}_j^s \equiv \left(x_{j1}^s, \ldots, x_{jT}^s\right)'$. Define c_{jt}^* as in (3.7), where ξ_{jt} is independent of $\tilde{x}_j^s$ and $\xi_{jt} \sim N\left(0,\ \sigma_\xi^2\right)$.

In spirit to a conditional mean independence assumption in a linear fixed-effects estimation framework, we need the following assumption which allows us to correct for selection bias in the current nonlinear estimation framework.

Assumption 3.3.2:

$E\left(u_{njt} \mid \rho_{nj}, v_j, \tilde{x}_j, \tilde{\xi}_j\right) = \theta_n \varepsilon_{jt} = \theta_n\left(\xi_{jt} - v_j\right),\ t = 1, 2, \ldots, T,$ and $\left(\omega_{nj1}, \omega_{nj2}, \ldots, \omega_{njT_j}\right)$, where $\omega_{njt} \equiv u_{njt} - \theta_n\left(\xi_{jt} - v_j\right)$, are continuously distributed with a density that is continuous and positive everywhere and are identically distributed conditional on $\left(\rho_{nj}, v_j, \tilde{x}_j, \tilde{\xi}_j\right)$, while x_{jt} denote the vector of all the observed explanatory variables in (3.11) exculding $\hat{\xi}_{jt}$ and $\tilde{x}_j \equiv \left(x_{j1}, \ldots, x_{jT_j}\right)'$.

Under Assumptions 3.3.1 and 3.3.2, we have (3.11). Estimation for (3.11) proceeds exactly as in the test procedure in the previous section, except in cases that θ_n is different from zero, the asymptotic variance of the coefficient estimates in (3.11) $\hat{\delta}_n$ needs to be adjusted as in the following procedure, given the preliminary estimation of the coefficients, denoted by τ, in (3.7).

Step 1, *Step 2* and *Step 3* are carried out exactly as in the in the test procedure in the previous section.

Step4: to estimate the asymptotic variance of $\hat{\delta}_n$ using the results in the Appendix.

3.4 Generalized Method of Moments Estimation framework

Once the consistent equation-by-equation estimates are obtained for each soft drink, following Meyerhoefer, Ranney, and Sahn (2005), the cross-equation homogeneity and symmetry restrictions on γ_{nk}'s, implied from the consumption theory, are imposed through a minimum distance estimator using the sample analogue of moment conditions in (3.13), to derive consistent structural parameter estimates. Specifically, denote the drink-by-drink reduced-form parameter estimates for all share equations as $\delta = \left(\delta_1', \ \delta_2', \ \delta_3'\right)'$. The structural parameters, denoted by π, can be consistently estimated as:

$$\min_{\pi}\left(\hat{\delta} - m(\pi)\right)' \mathrm{W}\left(\hat{\delta} - m(\pi)\right)$$

where $\hat{\delta}$ are consistent estimates of the reduced-form parameters δ, which are obtained from drink-by-drink estimation, and W is the weighting matrix measuring the distance between the sample moments and the corresponding population moments. $m(.)$ is a function mapping π into δ, which is used to impose restrictions implied from demand theory on the reduced form parameters. π can be efficiently estimated if $\mathrm{W} = \Xi^{-1}$, where Ξ is the asymptotic covariance matrix of $\hat{\delta}$. It can be shown that $\Xi = \mathrm{H}^{-1} S \mathrm{H}^{-1}$ (Wooldridge 2010b).

Let $S_j = \left(S_{1j}{}', S_{2j}{}', S_{3j}{}'\right)'$ denote the set of the subject j's moment conditions in (3.13) for all the soft drinks and H_{nj} denote the univariate Hessian for soft drink n. Then, define $\mathrm{H}^{-1} = diag\left\{E\left(H_{1j}\right)^{-1}, \ldots, E\left(H_{Nj}\right)^{-1}\right\}$ and $S = E\left(S_j S_j'\right)$. Ξ can be consistently estimated by substituting in sample analogues.

3.5 Elasticity Formulae

The generic expressions for total expenditure and uncompensated price elasticities for any demand system are given by

$$E_n = \frac{\partial w_n}{\partial \log c} \frac{1}{w_n} + 1 \tag{3.16}$$

and

$$e_{ni} = \frac{\partial w_n}{\partial \log p_i} \frac{1}{w_n} - \delta^*_{ni}, \tag{3.17}$$

where δ^*_{ni} is the Kronecker delta, and the compensated price elasticities are derived using the Slutsky relationship: $\tilde{e}_{ni} = e_{ni} + s_i E_n$. Since, as explained in Honor (2008), the parameter estimates for the fixed-effects models can be converted to marginal effects by multiplying them by the fraction of observations that are not censored, for the demand system proposed in this study, E_n and e_{ni} can be expressed as follows:

$$E_n = \beta_n F_n \frac{1}{w_n} + 1 \tag{3.18}$$

and

$$e_{ni} = \gamma_{ni} F_n \frac{1}{w_n} - \delta^*_{ni}, \tag{3.19}$$

where F_n denotes the fraction of observations that are not censored for soft drink n and w_n is the share of drink n.

4. Results

All tests and estimations were carried out using the R programming language. The codes can be obtained from the authors upon request. As the null hypothesis that there is no selection bias is rejected for Fizzy and Cordial but not for Juice, the correction procedure is only implemented for Fizzy and Cordial. Table 3 presents the structural parameter estimates. Based on these estimates, the average expenditure and price elasticities are estimated and reported in Table 4. We have also estimated and examined the average partial elasticities of demand for drink n (=Fizzy, Juice and Cordial) with respect to the *j*th attribute of drink i, which can be generically expressed as:

$$P_{n,ij} = \frac{\partial \log x_n}{\partial d_{ij}} = \frac{\partial w_n}{\partial d_{ij}} \frac{1}{w_n} \tag{4.1}$$

where x_n denotes the demand for drink n, and is particularly expressed as

$$P_{n,ij} = \lambda_n F_n \frac{1}{w_n}, \tag{4.2}$$

where w_n is the share of drink n. The estimates and their standard errors are reported in Table 5. These results are important evidence for soft drink tax policy concerns and health-related, such obesity and dental health, campaigns.

As shown in Table 4, the average uncompensated own-price elasticities for all the three drinks are expectedly all negative. According to the sign of the average uncompensated cross-price elasticities, Fizzy and Juice and Juice and Cordial are on average treated as gross complements. After accounting for the income effects, the average compensated own-price elasticities of all three goods are also significantly negative, which is consistent with the law of demand, and all the average compensated cross-price elasticities are positive, although some are not statistically significant, indicating that on average all the three drinks are net substitutes to one another.

As has been noticed, the adding-up condition is not imposed *a priori*. It is interesting to see if this condition is still, at least approximately, satisfied. Since the adding-up condition is equivalent to the Engel aggregation constraint: $\sum w_n E_n = 1$ (Deaton and Muellbauer 1980b), given our estimated income elasticities $\hat{E} = (1.607,\ 1.129,\ 0.315)'$ and the mean shares of drinks $\bar{w} = (0.096,\ 0.745,\ 0.159)$, $\bar{w}'\hat{E} = 1.045$, which is very close to 1.

Table 5 presents the average estimated partial elasticities w.r.t attributes. In particular, these estimates can be interpreted as the percentage changes of drink consumption corresponding to changes of attributes. Taking Fizzy as an example, switching from normal Fizzy to diet Fizzy would increase the consumption of Fizzy on average by 27.8%. As shown in Table 5, most average own and cross partial elasticity estimates are not statistically significant, suggesting that on average, drink consumptions are not statistically significantly affected by most own and cross attribute variations. Nevertheless, we do observe that Fizzy being diet significantly increases its consumption (27.8%). Being presented together with diet Cordial increases the consumption of Fizzy by 15.4% than otherwise. Healthier fizzy (no added colours or preservatives) significantly crowd out the consumption of juice. And, being present together with either diet Fizzy or Juice with no added sugar increases its consumption than otherwise.

In order to illustrate what might have been achieved by taking into account the potential selection on total soft drink expenditure, we also estimate the proposed semi-parametric fixed-effects censored demand system model only using observations with positive total soft drink expenditure without undertaking the sample selection test and correction procedures. To put it another way, we make our estimates fully vulnerable to potential selection biases which might arise from the exclusion of observations with zero total soft drink expenditure and it would be very interesting to see how the main results would be different. The results are presented in Tables A1 through A3 in Appendix 2.

In comparison to Table 3, the structural parameter estimates in Table A1 tend to be smaller, and we observe fewer significant estimates in Table A1. In particular, in contrast to Table 3, the coefficient estimates of log real total expenditure for Fizzy and log cordial price are not statistically significant even at 10% significance level in Table A1. While contrasting Table A2 to Table 4, in terms of total expenditure and price elasticity estimates, it can be observed that without taking into potential selection bias tends to overestimate positive but underestimate negative estimates than otherwise. We also observe fewer estimates which are statistically significant (even at 10% significance level) in Table A2. For instance, while in Table 4 one percent increase in Fizzy price will significantly increase the consumption of Cordial by 0.254 percent, after accounting for the income effects, the counterpart in Table A2 is only 0.012 percent and not statistically significant at 10% level.

As for partial elasticity estimates, once again, we observe that the discrepancy between estimates of Table 5 and Table A3 is remarkable. For instance, while in Table 5 we respectively observe 10% significant and non-significant partial elasticities of the consumption of Fizzy with respect to Cordial Diet and Juice Diet, we observe exactly the opposite in Table A3. Also, in Table A3 for the Cordial consumption, while a statistically significant (at 1% level) negative partial elasticity w.r.t. Fizzy diet is observed, Table 5 shows a corresponding 10% statistically significant positive partial elasticity.

Accordingly, it's quite obvious that taking into account the potential selection biases on the total soft drink expenditure results in substantial differences in the estimates, which, in a perceivable manner, demonstrates significant selection bias.

To better inform policy, it is interesting to examine how the consumption behaviour of subjects from households (HHs) having higher income would be different from those from HHs with lower HH income. Hence, we explicitly split our sample into two parts at the point of median HH income in our sample, test and correct for potential selection bias on the total soft drink expenditure, and estimate the proposed semi-parametric fixed-effects demand system model respectively for the two parts.

In terms of estimation, the most salient discrepancy is that for the subjects with higher than median HH income, the null hypothesis of no selection bias with the total soft drink expenditure is rejected for all the three drinks considered. In contrast, the null is not rejected for any of the three drinks for the subjects with lower than median HH income. Consequently, the correction procedure is undertaken for all the drinks for the subjects with higher than median HH income, but is not carried out for any of the three drinks for the sub-sample with lower than median HH income. This discrepancy suggests that for those rich subjects, there might exist an underlying mechanism which drives subjects' overall perception towards soft drinks given the attributes and prices presented. Such a mechanism is systematically correlated with the rich subject's consumption decision for each of the three drinks. In contrast, such a correlation does not exist for any of the three drinks for the poor subjects.

The results for subjects with HH income lower than the median and subjects with higher than the median HH income are respectively given in Table A4 through Table A6 and Table A7 through Table A8. Contrasting Table A8 with Table A5, in terms of total expenditure elasticity, we observe that subjects with lower HH income have higher elasticity for Fizzy but lower elasticity for Juice and much lower elasticity for Cordial. In our case, the total expenditure on soft drinks might be better interpreted as an indicator of subjects' (or parents') overall openness or attitude towards soft drinks. Hence, conservative parents tend to give less or even no soft drinks to their children. The differences in total expenditure elasticities between subjects with lower and higher HH income suggest that when the subject becomes more open to soft drinks and decides to increase his/her total budget on soft drinks for their children by 1%, then the subject from HH with higher than median income (rich subject) only increases the Cordial consumption on average by 0.4%, while the subject from HH with lower than median income (poor subject) increases the Cordial consumption by 1.093%, and on average the subject with lower than median HH income tends to increase his/her Fizzy and Juice consumption more than the subject with higher than median HH income.

In terms of price elasticity, after accounting for the income effects, it seems that the rich subject's Fizzy and Cordial consumption is more elastic than the poor subject's by exhibiting lower negative compensated own price elasticities (-0.873 versus -0.535 and -0.768 versus -0.650), while the rich subject's Juice consumption is less elastic than the poor's. As for substitutability between different drinks, looking at cross-price compensated price elasticities, while there is very strong evidence in Table A8 for the rich subject evincing that Cordial is treated as a substitute to either Fizzy or Juice, for the poor subject in Table A5 such evidence does not exist. We even vaguely observe that Cordial might be treated as a complement to Fizzy or Juice, although the negative compensated cross-price elasticities are not statistically significant even at 10% significance level.

When it comes to partial elasticities w.r.t. attributes, we observe substantial discrepancies between the rich and the poor subjects. In particular, the rich subject seems to be remarkably more sensitive than the poor to the attributes of drinks. For instance, for the rich subject, in terms of partial elasticities w.r.t. own attributes, switching from normal Fizzy to Fizzy with extra vitamins A & C remarkably increases the weekly consumption of Fizzy by three times. Diet Fizzy also increases its consumption by 14.7%. Interestingly, while we expected parents to have a sense that Fizzy with no added colours or preservatives is healthier and therefore increases the consumption, we actually observe a significant decrease of 49.2%. The same phenomenon is observed for Cordial as well. In particular, we observe a decrease of 21.8% when the Cordial is labelled as no added colours or preservatives. As for Juice, only being labelled with extra vitamins A & C seems to statistically significantly increase its consumption by 5.9%. For the poor subject, we do not observe much statistically significant evidence. The only exception is that Juice with no added sugar interestingly decreases its consumption by 7.6%.

In regards to cross-attribute partial elasticities, it seems that for the rich subject, Juice with extra vitamins or Fizzy with extra vitamins tend to significantly increase one another's consumption. For instance, Juice with extra vitamins dramatically increases the weekly consumption of Fizzy by 2.5 times. We also observe that Fizzy with extra vitamins and Cordial with extra vitamins tend to crowd out each other's weekly consumption level respectively by 83.1% and 27.4%. Juice with no added sugar decreases Fizzy consumption by 29.3%, whereas Cordial drinks with no added colours or preservatives increase Fizzy consumption by 48.1%. Similarly, Fizzy and Juice with no added colours or preservatives respectively increase the Cordial consumption respectively by 28.5% and 14.6%. As a result, it is interesting to note that for the rich subject, when there is a statistically significant effect, a drink being labelled as no added colours or preservatives seems not to increase its own consumption yet increase consumption of other drinks. As for the poor subject, once again, we observe very little statistically significant evidence. However, it does show that Fizzy with extra vitamins significantly increases the consumption of Cordial by 51.2%.

Table 3 Structural Parameter Estimates

Drink	Variable	Coef.		S.E.
Fizzy				
	Fizzy Diet	0.196	**	(0.082)
	Fizzy Vitamins	0.101		(0.087)
	Fizzy Nocolours	-0.026		(0.055)
	Juice Diet	-0.049		(0.050)
	Juice Vitamins	0.109		(0.115)
	Juice Nocolours	-0.055		(0.100)
	Cordial Diet	0.109	*	(0.059)
	Cordial Vitamins	0.036		(0.075)
	Cordial Nocolours	0.128		(0.100)
	Log Fizzy Price	0.064	**	(0.025)
	Log Juice Price	-0.063	***	(0.022)
	Log Cordial Price	-0.001		(0.014)
	Log Real Total Expenditure	0.428	***	(0.104)
Juice				
	Fizzy Diet	-0.161		(0.111)
	Fizzy Vitamins	-0.018		(0.031)
	Fizzy Nocolours	-0.144	***	(0.045)
	Juice Diet	-0.054		(0.039)
	Juice Vitamins	-0.017		(0.072)
	Juice Nocolours	0.022		(0.039)
	Cordial Diet	-0.090		(0.104)
	Cordial Vitamins	0.004		(0.070)
	Cordial Nocolours	-0.015		(0.036)
	Log Fizzy Price	-0.063	***	(0.022)
	Log Juice Price	0.145	***	(0.029)
	Log Cordial Price	-0.082	***	(0.023)
	Log Real Total Expenditure	0.318	***	(0.037)
Cordial				
	Fizzy Diet	0.058	*	(0.033)
	Fizzy Vitamins	0.003		(0.042)
	Fizzy Nocolours	0.066		(0.044)
	Juice Diet	0.061	*	(0.031)
	Juice Vitamins	-0.015		(0.061)
	Juice Nocolours	0.008		(0.027)
	Cordial Diet	0.014		(0.057)
	Cordial Vitamins	-0.013		(0.045)
	Cordial Nocolours	0.000		(0.042)
	Log Fizzy Price	-0.001		(0.014)
	Log Juice Price	-0.082	***	(0.023)
	Log Cordial Price	0.082	***	(0.028)
	Log Real Total Expenditure	-0.404	***	(0.038)

Note: * Significant at 10%; ** Significant at 5%; *** Significant at 1%. Fizzy Diet: diet Fizzy; Fizzy Vitamins: Fizzy with extra vitamins; Fizzy Nocolours: Fizzy with no added colours or preservatives; Juice Diet: Juice with no added sugar; Juice Vitamins: Juice with extra vitamins; Juice Nocolours: Juice with no added colours or preservatives; Cordial Diet: diet Cordial; Cordial Vitamins: Cordial with extra vitamins; Cordial Nocolours: Cordial with no added colours or preservatives.

Table 4 Average Total Expenditure Elasticities and Uncompensated and Compensated Price Elasticities

		Fizzy	Juice	Cordial
Total expenditure		1.607*** (0.147)	1.129*** (0.015)	0.315*** (0.064)
Uncompensated	Fizzy	-0.909*** (0.035)	-0.026*** (0.009)	-0.001 (0.024)
	Juice	-0.090*** (0.031)	-0.941*** (0.012)	-0.138*** (0.038)
	Cordial	-0.001 (0.020)	-0.033*** (0.009)	-0.861*** (0.047)
Compensated	Fizzy	-0.755*** (0.032)	0.083*** (0.009)	0.029 (0.024)
	Juice	1.108*** (0.126)	-0.100*** (0.017)	0.097 (0.065)
	Cordial	0.254*** (0.032)	0.146*** (0.009)	-0.811*** (0.048)

Note: Standard errors are in parenthesis. * Significant at 10%; ** Significant at 5%; *** Significant at 1%.

Table 5 Average Partial Elasticities w.r.t. Attributes

Drink	Attributes	Coef.		S.E.
Fizzy				
	Fizzy Diet	0.278	**	(0.117)
	Fizzy Vitamins	0.143		(0.123)
	Fizzy Nocolours	-0.037		(0.077)
	Juice Diet	-0.070		(0.070)
	Juice Vitamins	0.154		(0.163)
	Juice Nocolours	-0.078		(0.142)
	Cordial Diet	0.154	*	(0.084)
	Cordial Vitamins	0.051		(0.106)
	Cordial Nocolours	0.182		(0.142)
Juice				
	Fizzy Diet	-0.065		(0.045)
	Fizzy Vitamins	-0.007		(0.013)
	Fizzy Nocolours	-0.058	***	(0.018)
	Juice Diet	-0.022		(0.016)
	Juice Vitamins	-0.007		(0.029)
	Juice Nocolours	0.009		(0.016)
	Cordial Diet	-0.036		(0.042)
	Cordial Vitamins	0.002		(0.029)
	Cordial Nocolours	-0.006		(0.014)
Cordial				
	Fizzy Diet	0.098	*	(0.056)
	Fizzy Vitamins	0.004		(0.072)
	Fizzy Nocolours	0.113		(0.074)
	Juice Diet	0.104	*	(0.053)
	Juice Vitamins	-0.025		(0.103)
	Juice Nocolours	0.013		(0.046)
	Cordial Diet	0.024		(0.097)
	Cordial Vitamins	-0.023		(0.077)
	Cordial Nocolours	0.000		(0.071)

Note: * Significant at 10%; ** Significant at 5%; *** Significant at 1%. Fizzy Diet: diet Fizzy; Fizzy Vitamins: Fizzy with extra vitamins; Fizzy Nocolours: Fizzy with no added colours or preservatives; Juice Diet: Juice with no added sugar; Juice Vitamins: Juice with extra vitamins; Juice Nocolours: Juice with no added colours or preservatives; Cordial Diet: diet Cordial; Cordial Vitamins: Cordial with extra vitamins; Cordial Nocolours: Cordial with no added colours or preservatives.

5. Summary and Conclusion

In this study, using panel data obtained from a soft drink consumption experiment, a new semi-parametric fixed-effects censored demand system is specified and estimated. In particular, to deal with the difficulty of a substantial proportion of zero observations for the total expenditure on soft drinks and for the expenditure on each drink, a new two-step estimation strategy is developed and a semi-parametric estimator for two-sided censoring models with fixed effects is employed. In addition, a consistent and asymptotically efficient GMM estimator is used to impose economic restrictions on the model and identify the underlying structural parameters.

This two-step estimation strategy also has the potential to recover a full multi-stage demand system estimation framework which has been extensively examined and applied to study aggregate consumption data, but it is not yet clear in the literature how to apply a multi-stage framework to study micro-level consumption data and worth pursuing in future research.

Based on our parameter estimates, the consumption behaviour of subjects is analysed through estimated income elasticities and uncompensated and compensated price elasticities. The partial elasticities of demand with respect to attributes of soft drinks are also estimated. These results provide valuable empirical evidence for soft-drink tax policy concerns and health-related, such obesity and dental health, campaigns.

Appendix

The estimation falls within the two-step M-estimation framework, and the semi-parametric estimator becomes a two-step estimator. To see this, it is helpful to substitute in (3.12) the expression of $\hat{\xi}_{jt}$ in (3.10) and rewrite the semi-parametric estimator on the selected sample as:

$$\begin{aligned}\hat{\delta}_n &= \arg\min_{\delta_n} \sum_{j=1}^{J} \sum_{1<s<t<T_j} \frac{1}{T_j} U\left(w_{njt}, w_{njs}, \left(\tilde{x}_{jt} - \tilde{x}_{js}\right)' \tilde{\delta}_n + \left(c_{jt} - c_{js}\right)\theta_n - \left(\widehat{x}_{jt} - \widehat{x}_{js}\right)' \hat{\tau}\theta_n \right) \\ &= \arg\min_{\delta_n} \sum_{j=1}^{J} \sum_{1<s<t<T_j} \frac{1}{T_j} U\left(w_{njt}, w_{njs}, \left(x_{jt}^c - x_{js}^c\right)' \left(\tilde{\delta}_n - \hat{\tau}^c\theta_n\right) + \left(c_{jt} - c_{js}\right)\theta_n - \left(x_{jt}^u - x_{js}^u\right)' \hat{\tau}^u\theta_n \right)\end{aligned} \tag{A.1}$$

where $\tilde{x}_{jt}$ denotes the vector of all the explanatory variables in (3.11) except $\hat{\xi}_{jt}$; $\tilde{\delta}_n$ denotes coefficient vector corresponding to $\tilde{x}_{jt}$; θ_n denotes the coefficient corresponding to $\hat{\xi}_{jt}$, so that $x_{jt} \equiv \left(\tilde{x}_{jt}', \hat{\xi}_{jt}\right)'$ and $\delta_n \equiv \left(\tilde{\delta}_n', \theta_n\right)'$; $\widehat{x}_{jt}$ denotes the vector of observable explanatory variables in (3.7) and $\hat{\tau}$ denotes corresponding coefficient estimates; x_{jt}^c denotes shared observable explanatory variables between (3.7) and (3.11) and x_{jt}^u denotes observable explanatory variables that only appear in the selection equation (3.7). The identification and consistency of this two-step estimator given τ can be clearly seen following the identification and consistency arguments in Alan et al. (2014).

Since the objective function $U(\cdot)$, as shown in (3.12), is not twice differentiable, the standard adjustment procedure (see for example Wooldridge 2010c) cannot be applied to show the asymptotic normality of $\hat{\delta}_n$. However, following Theorem 3.3 in Pakes and Pollard (1989), the normality of the two-step semi-parametric estimator can still be proved.

Let

$$h\left(x_j, \delta_n; \hat{\tau}\right) = \sum_{1<s<t<T_j} \frac{1}{T_j} u\left(w_{njt}, w_{njs}, \Delta x_j{}' \delta_n\right) \Delta x_j$$

(A.2)

and define functions:

$$G_J\left(a;b\right) = 1/J \sum_{j=1}^{J} h\left(x_j, a; b\right) \equiv 1/J \sum_{j=1}^{J} h_j\left(a;b\right)$$

(A.3)

and

$$G\left(a;b\right) = E\left(h_j\left(a;b\right)\right)$$

(A.4)

Since $\hat{\tau}$ is obtained from a Tobit estimation, a first-order representation for $\sqrt{J}\left(\hat{\tau}-\tau\right)$ can be obtained, which is written as (Wooldridge 2010c):

$$\sqrt{J}\left(\hat{\tau}-\tau\right) = J^{-1/2} \sum_{j=1}^{J} r_j\left(\tau\right) + o_p\left(1\right).$$

(A.5)

Throughout this appendix, the symbols $\|\cdot\|$ denotes not only the usual Euclidean norm but also a matrix norm: $\left\|\left(b_{ij}\right)\right\| = \left(\sum_{i,j} b_{ij}^2\right)^{1/2}$. LEMMA 1 states that the function $G\left(\delta_n; b\right)$ is differentiable at τ.

Lemma A1: Given finite moment conditions that $E\left(\left\|x_{jt}^c\right\|^2\right) < \infty$ and $E\left(\left\|x_{jt}^u\right\|^2\right) < \infty$, $G\left(\delta_n; b\right)$ is differentiable at τ with derivative matrix $\Gamma_\tau\left(\delta_n\right) = \left.\frac{dG\left(\delta_n; b\right)}{db}\right|_{b=\tau}$.

PROOF:

Let $x_{jt}^{a} = \left(x_{jt}^{c}{}', c_{jt}, x_{jt}^{u}{}' \right)'$ for any $t \le T_j$ and $\delta_n^a = \left(\left(\tilde{\delta}_n - b^c \theta_n \right)', \theta_n, -b^{u\prime} \theta_n \right)'$, so that $b = \left(b^{c\prime}, b^{u\prime} \right)'$. According to Theorem 1 in Alan et al. (2014), given

$$h_j\left(\delta_n; b\right) = \sum_{1<s<t<T_j} \frac{1}{T_j} u\left(w_{njt}, w_{njs}, \Delta x_j^{a\prime} \delta_n^a \right) \Delta x_j^{a\prime},$$

$$\begin{aligned}
\Gamma_\tau\left(\delta_n\right) &= \left. \frac{dE\left[h_j\left(\delta_n; b\right) \right]}{db} \right|_{b=\tau} \\
&= \left. E\left(\sum_{1<s<t<T_j} \left(\begin{array}{l} \frac{1}{T_j} 1\left\{ -1 < \left(\Delta x_j^{a\prime} \delta_n^a \right) < 1 \right\} \\ \left(\begin{array}{l} 1\left\{ -1 < \Delta x_j^{a\prime} \delta_n^a < w_{njs} - 1 \right\} - 1\left\{ 0 < \Delta x_j^{a\prime} \delta_n^a < w_{njs} \right\} \\ -1\left\{ -w_{njt} < \Delta x_j^{a\prime} \delta_n^a < 0 \right\} + 1\left\{ 1 - w_{njt} < \Delta x_j^{a\prime} \delta_n^a < 1 \right\} \end{array} \right) \Delta x_j^a \Delta x_j^{a\prime} \end{array} \right) \right) \frac{\partial \delta_n^a}{\partial b} \right|_{b=\tau},
\end{aligned}$$

where $\Delta x_j^a = x_{jt}^a - x_{js}^a$. Given the finite moment conditions, it is trivial to see that $\left\| \Gamma_\tau \right\| < \infty$.

Q.E.D.

In order to show that $G_n\left(\delta_n; \hat{\tau}\right)$ can be approximated by a well-behaved linear function, we also need the following lemma.

Lemma A2: For any sequence $\left\{ \chi_J \right\}$ of positive numbers such that $\chi_J \to 0$ as $J \to \infty$,

$$\sup_{\|b-\tau\|<\chi_J} \left\| G_J\left(\delta_n; b\right) - G\left(\delta_n; b\right) - G_J\left(\delta_n; \tau\right) \right\| = o_p\left(J^{-1/2} \right).$$

Let $\delta_n^a = \left(\left(\tilde{\delta}_n - b^c \theta_n \right)', \theta_n, -b^{u\prime} \theta_n \right)'$. Given $h_j\left(\delta_n; b\right) = \sum_{1<s<t<T_j} \frac{1}{T_j} u\left(w_{njt}, w_{njs}, \Delta x_j^{a\prime} \delta_n^a \right) \Delta x_j^{a\prime}$ and the differentiability from Lemma 1, Lemma 2 can be similarly proved as in the proof of Theorem 2 verifying condition (iii) for $l = 4$ in Honoré (1992).

Based on the previous two lemmas, the following lemma states that $G_J(\delta_n;\hat{\tau})$ converges in distribution to a normal distribution, given the asymptotic normality of $G_J(\delta_n;\tau)$.

Lemma A3: If $\sqrt{J}G_J(\delta_n;\tau)\xrightarrow{d}N(0,V)$, then

$$\sqrt{J}G_J(\delta_n;\hat{\tau})\xrightarrow{d}N(0,V^*)$$

where

$$V^*=E\left[\left(h_j(\delta_n;\tau)+\Gamma_\tau(\delta_n)r_j(\tau)\right)\left(h_j(\delta_n;\tau)+\Gamma_\tau(\delta_n)r_j(\tau)\right)'\right]$$

and

$$\Gamma_\tau(\delta_n)=\left.\frac{dG(\delta_n;b)}{db}\right|_{b=\tau}.$$

PROOF:

To establish asymptotic normality of $G_J(\delta_n;\hat{\tau})$, we first show that $G_J(\delta_n;\hat{\tau})$ is very well approximated by the linear function: $L_J^\tau(\delta_n;\hat{\tau})=\Gamma_\tau(\delta_n)(\hat{\tau}-\tau)+G_J(\delta_n;\tau)$. This follows directly from Lemma 1 and Lemma 2 together with the consistency of $\hat{\tau}$ and first-order representation in (A.5). Specifically, given the consistency of $\hat{\tau}$, we can always choose a positive sequence $\{\chi_J\}$ that converges to zero as J goes to infinity slowly enough to ensure that

$$P\left\{\|\hat{\tau}-\tau\|\le\chi_J\right\}\to 1.$$

With the probability tending to one, the supremum in the statement of Lemma A2 runs over a range that includes the random value $\hat{\tau}$. Hence,

$$\left\|G_J(\delta_n;\hat{\tau})-G(\delta_n;\hat{\tau})-G_J(\delta_n;\tau)\right\|\le o_p\left(J^{-1/2}\right)$$

Then, it follows that

$$\left\| G_J\left(\delta_n;\hat{\tau}\right) - L_J^{\tau}\left(\delta_n;\hat{\tau}\right) \right\| \le \left\| G_J\left(\delta_n;\hat{\tau}\right) - G\left(\delta_n;\hat{\tau}\right) - G_J\left(\delta_n;\tau\right) \right\| + \left\| G\left(\delta_n;\hat{\tau}\right) - \Gamma_{\tau}\left(\delta_n\right)\left(\hat{\tau}-\tau\right) \right\|$$
$$\le o_p\left(J^{-1/2}\right) + o_p\left(\left\|\hat{\tau}-\tau\right\|\right)$$
$$= o_p\left(J^{-1/2}\right).$$

Hence,

$$\begin{aligned}\sqrt{J}G_J\left(\delta_n;\hat{\tau}\right) &= \sqrt{J}G_J\left(\delta_n;\tau\right) + \Gamma_{\tau}\left(\delta_n\right)\sqrt{J}\left(\hat{\tau}-\tau\right) + o_p\left(1\right) \\ &= J^{-1/2}\sum\nolimits_{j=1}^{J} h\left(x_j,\delta_n;\tau\right) + \Gamma_{\tau}\left(\delta_n\right) J^{-1/2}\sum_{j=1}^{J} r_j\left(\tau\right) + o_p\left(1\right) \\ &= J^{-1/2}\sum\nolimits_{j=1}^{J}\left[h\left(x_j,\delta_n;\tau\right) + \Gamma_{\tau}\left(\delta_n\right) r_j\left(\tau\right)\right] + o_p\left(1\right),\end{aligned}$$

(A.6)

given the first-order representation of $\sqrt{J}\left(\hat{\tau}-\tau\right)$ in (A.5). Then, it follows that

$$\sqrt{J}G_J\left(\delta_n;\hat{\tau}\right) \xrightarrow{d} N\left(0,V^*\right),$$

where $V^* = E\left[\left(h_j\left(\delta_n;\tau\right) + \Gamma_{\tau}\left(\delta_n\right) r_j\left(\tau\right)\right)\left(h_j\left(\delta_n;\tau\right) + \Gamma_{\tau}\left(\delta_n\right) r_j\left(\tau\right)\right)'\right]$.

Q.E.D.

With these lemmas, we can derive the asymptotic distributions of the semi-parametric two-step estimator $\hat{\delta}_n$.

Theorem A1: Let x_{jt}^{c} denotes shared observable explanatory variables between (3.7) and (3.11) and x_{jt}^{u} denote observable explanatory variables that only appear in the selection equation (3.7). Define $x_{jt}^{a} \equiv \left(x_{jt}^{c\prime}, c_{jt}, x_{jt}^{u\prime}\right)'$. If

1. The parameter space, Δ, is compact, and the true value of the parameter is an interior point of Δ, $\delta_n \in \operatorname{int}\Delta$.
2. Assumption 3.3.1:

Denote observable explanatory variables in (3.7) as x_{jt}^{s} and let $\tilde{x}_{j}^{s} \equiv \left(x_{j1}^{s},\ldots,x_{jT}^{s}\right)'$. Define c_{jt}^{*} as in (3.7), where ξ_{jt} is independent of $\tilde{x}_{j}^{s}$ and $\xi_{jt} \sim N\left(0,\ \sigma_{\xi}^{2}\right)$.

3. Assumption 3.3.2:

Let $\tilde{x}_{j} \equiv \left(x_{j1},\ldots,x_{jT_j}\right)'$ and $\tilde{\xi}_{j} \equiv \left(\xi_{j1},\ldots\xi_{jT}\right)'$.

$E\left(u_{njt} \mid \rho_{nj}, v_{j}, \tilde{x}_{j}, \tilde{\xi}_{j}\right) = \theta_{n}\varepsilon_{jt} = \theta_{n}\left(\xi_{jt} - v_{j}\right),\ t = 1,2,\ldots,T_{j},$ and $\left(\omega_{nj1}, \omega_{nj2}, \ldots, \omega_{njT_j}\right)$, where $\omega_{njt} \equiv u_{njt} - \theta_{n}\left(\xi_{jt} - v_{j}\right)$, are continuously distributed with a density that is continuous and positive everywhere and are identically distributed conditional on $\left(\rho_{nj}, v_{j}, \tilde{x}_{j}, \tilde{\xi}_{j}\right)$, while x_{jt} denote the vector of all the observed explanatory variables in (3.11) excluding $\hat{\xi}_{jt}$.

4. Finite moment conditions: for any $t \le T$, $E\left(\left\|x_{jt}^{a}\right\|^{2}\right) < \infty$. For any $s,t \le T_{j}$, $s \ne t$, $E\left(u_{njt}^{2}\left\|x_{js}^{a} - x_{jt}^{a}\right\|^{2}\right) < \infty$, $E\left(u_{njs}^{2}\left\|x_{js}^{a} - x_{jt}^{a}\right\|^{2}\right) < \infty$, and $E\left(\eta_{j}^{2}\left\|x_{js}^{a} - x_{jt}^{a}\right\|^{2}\right) < \infty$.

5. While x_{jt} denotes the vector of all the observed explanatory variables in (3.11) excluding $\hat{\xi}_{jt}$, define $x_{jt}^{m} \equiv \left(x_{jt}, \xi_{jt}\right)'$. The matrix

$$E\left[\left(x_{jt}^{m} - x_{js}^{m}\right)\left(x_{jt}^{m} - x_{js}^{m}\right)' \mid -1 < \left(x_{jt}^{m} - x_{js}^{m}\right)' \delta_{n} < 1\right]$$

has full rank, for $s,t \le T_{j}$, $s \ne t$.

Then

$$\sqrt{J}\left(\hat{\delta}_{n} - \delta_{n}\right) \xrightarrow{d} N\left(0, \Gamma_{\delta_n}\left(\tau\right)^{-1} V^{*} \Gamma_{\delta_n}\left(\tau\right)^{-1}\right)$$

where $\hat{\delta}_n$ is defined as in (3.12) and $\Gamma_{\delta_n}(\tau) \equiv \left.\frac{\partial G(a;\tau)}{\partial a}\right|_{a=\delta_n}$, and $V^* \equiv E\left[g_j g_j{}'\right]$;

$g_j \equiv h_j(\delta_n;\tau) + \Gamma_\tau(\delta_n) r_j(\tau)$, where $r_j(\tau)$ is given as in (A.5) and $\Gamma_\tau(\delta_n) = \left.\frac{dG(\delta_n;b)}{db}\right|_{b=\tau}$.

PROOF:

First we prove $\sqrt{J}$-consistency. Given the consistency of the estimator $\hat{\delta}_n$ in (A.1), it allows us to choose a positive sequence $\{k_J\}$ that converges to zero as J goes to infinity slowly enough to ensure that

$$P\left\{\left\|\hat{\delta}_n - \delta_n\right\| \le k_J\right\} \to 1$$

(A.7)

The differentiability of $G(a;\tau)$ at δ_n can be justified by Theorem 1 in Alan et al. (2014). It can be similarly proved as in the proof of Theorem 2 verifying condition (iii) for $l = 4$ in Honoré (1992) that

$$\sup_{\|a-\delta_n\|\le k_J} \left\|G_J(a;\tau) - G(a;\tau) - G_J(\delta_n;\tau)\right\| = o_p\left(J^{-1/2}\right).$$

(A.8)

With the probability in (A.7) tending to one, the supremum in (A.8) runs over a range that includes the random value $\hat{\delta}_n$. Hence,

$$\left\|G_J\left(\hat{\delta}_n;\tau\right) - G\left(\hat{\delta}_n;\tau\right) - G_J(\delta_n;\tau)\right\| \le o_p\left(J^{-1/2}\right)$$

(A.9)

By the triangle inequality, the left-hand side of (A.9) is larger than

$$\left\|G\left(\hat{\delta}_n;\tau\right)\right\| - \left\|G_J\left(\hat{\delta}_n;\tau\right)\right\| - \left\|G_J(\delta_n;\tau)\right\|$$

(A.10)

Thus,

$$\left\| G\left(\hat{\delta}_n;\tau\right) \right\| \le o_p\left(J^{-1/2}\right) + \left\| G_J\left(\hat{\delta}_n;\tau\right) \right\| + \left\| G_J\left(\delta_n;\tau\right) \right\|$$

(A.11)

As $U(\bullet)$ is everywhere differentiable, $G_J\left(\hat{\delta}_n;\tau\right) = 0$. Also, as a direct consequence of the Central Limit Theorem, $\sqrt{J}G_J\left(\delta_n;\tau\right)$ converges in distribution to $N(0,V)$, where $V = E\left[h_j\left(\delta_n;\tau\right)h_j\left(\delta_n;\tau\right)'\right]$. Hence, it follows from (A.11) that

$$\left\| G\left(\hat{\delta}_n;\tau\right) \right\| \le o_p\left(J^{-1/2}\right) + \left\| G_J\left(\delta_n;\tau\right) \right\| = O_p\left(J^{-1/2}\right)$$

(A.12)

That is,

$$\left\| G\left(\hat{\delta}_n;\tau\right) \right\| = O_p\left(J^{-1/2}\right).$$

(A.13)

The differentiability of $G(a;\tau)$ at δ_n with a derivative matrix of full rank, according to Theorem 1 in Alan et al. (2014), implies that there exists a positive constant c for which,

$$\left\| G(a;\tau) \right\| \ge c\left\| a - \delta_n \right\| \text{ for } a \text{ near } \delta_n .$$

(A.14)

In particular, $\left\| \hat{\delta}_n - \delta_n \right\| = O_p\left(\left\| G\left(\hat{\delta}_n;\tau\right) \right\|\right) = O_p\left(J^{-1/2}\right)$.

To establish asymptotic normality of $\sqrt{J}\left(\hat{\delta}_n - \delta_n\right)$, we argue that $G_J(a;\tau)$ can be very well approximated by the linear function

$$L_J(a;\tau) = \Gamma_{\delta_n}(\tau)\left(a - \delta_n\right) + G_J\left(\delta_n;\tau\right),$$

(A.15)

with an approximation error of order $o_p\left(J^{-1/2}\right)$ at $\hat{\delta}_n$ and at the $\hat{\delta}_n^*$ that minimises $\left\| L_J(a;\tau) \right\|$ globally. For $\hat{\delta}_n$, this follows directly from the differentiability of $G(a;\tau)$ at δ_n with a derivative matrix of full rank and (A.9), together with the $\sqrt{n}$-consistency results already established. In particular,

$$\begin{aligned}\left\|G_J\left(\hat{\delta}_n;\tau\right)-L_J\left(\hat{\delta}_n;\tau\right)\right\| &\le \left\|G_J\left(\hat{\delta}_n;\tau\right)-G\left(\hat{\delta}_n;\tau\right)-G_J\left(\delta_n;\tau\right)\right\| \\ &\quad +\left\|G\left(\hat{\delta}_n;\tau\right)-\Gamma_{\delta_n}\left(\tau\right)\left(\hat{\delta}_n-\delta_n\right)\right\| \\ &\le o_p\left(J^{-1/2}\right)+o_p\left(\left\|\hat{\delta}_n-\delta_n\right\|\right) \\ &= o_p\left(J^{-1/2}\right).\end{aligned}$$

(A.16)

To correspond to a minimum of $\left\|L_J\left(a;\tau\right)\right\|$, the vector $\Gamma_{\delta_n}\left(\tau\right)\left(\hat{\delta}_n^*-\delta_n\right)$ must be equal to the projection of $-G_J\left(\delta_n;\tau\right)$ onto the column space of $\Gamma_{\delta_n}\left(\tau\right)$. Hence,

$$\Gamma_{\delta_n}\left(\tau\right)\left(\hat{\delta}_n^*-\delta_n\right)=-\Gamma_{\delta_n}\left(\tau\right)\left(\Gamma_{\delta_n}\left(\tau\right)'\Gamma_{\delta_n}\left(\tau\right)\right)^{-1}\Gamma_{\delta_n}\left(\tau\right)'G_J\left(\delta_n;\tau\right).$$

(A.17)

Then, since $\Gamma_{\delta_n}\left(\tau\right)$ is symmetric and full rank, it follows that

$$\sqrt{J}\left(\hat{\delta}_n^*-\delta_n\right)=-\sqrt{J}\Gamma_{\delta_n}\left(\tau\right)^{-1}G_J\left(\delta_n;\tau\right)$$

(A.18)

As the estimate $\hat{\tau}$ of the selection equation in (3.7) is used to consistently estimate τ, substituting $\hat{\tau}$ in (A.18) gives

$$\sqrt{J}\left(\hat{\delta}_n^*-\delta_n\right)=-\sqrt{J}\Gamma_{\delta_n}\left(\hat{\tau}\right)^{-1}G_J\left(\delta_n;\hat{\tau}\right)$$

According to continuous mapping theorem (Mann and Wald 1943), $\Gamma_{\delta_n}\left(\hat{\tau}\right)\xrightarrow{p}\Gamma_{\delta_n}\left(\tau\right)$. As a direct consequence of the Central Limit Theorem, $\sqrt{J}G_J\left(\delta_n;\tau\right)$ converges in distribution to $N\left(0,V\right)$. Hence, Lemma A3 induces $\sqrt{J}G_J\left(\delta_n;\hat{\tau}\right)\xrightarrow{d}N\left(0,V^*\right)$. Consequently, it follows that

$$\sqrt{J}\left(\hat{\delta}_n^*-\delta_n\right)\xrightarrow{d}N\left(0,\Gamma_{\delta_n}\left(\tau\right)^{-1}V^*\Gamma_{\delta_n}\left(\tau\right)^{-1}\right),$$

where $V^*=E\left[\left(h_j\left(\delta_n;\tau\right)+\Gamma_\tau\left(\delta_n\right)r_j\left(\tau\right)\right)\left(h_j\left(\delta_n;\tau\right)+\Gamma_\tau\left(\delta_n\right)r_j\left(\tau\right)\right)'\right]$.

Hence, $\hat{\delta}_n^*=\delta_n+O_p\left(J^{-1/2}\right)$ and the $\{k_J\}$ sequence can be assumed to satisfy

$$P\left\{\left\|\hat{\delta}_n^* - \delta_n\right\| \geq k_J\right\} \to 0$$

(A.19)

Since δ_n is an interior point of the parameter space of δ_n, Δ, (A.19) implies that $\hat{\delta}_n^*$ lies in Δ with probability tending to one. Hereafter, we shall act as if $\left\|\hat{\delta}_n^* - \delta_n\right\| < k_J$ and $\hat{\delta}_n^*$ always lies in Δ. Actually, it can be easily shown that the contributions from those values of $\hat{\delta}_n^*$ not satisfying these two requirements are eventually absorbed into an $o_p(1)$ error term.

Similarly as in (A.8) and (A.9), we can get

$$\left\|G_J\left(\hat{\delta}_n^*;\tau\right) - G\left(\hat{\delta}_n^*;\tau\right) - G_J\left(\delta_n;\tau\right)\right\| \leq o_p\left(J^{-1/2}\right)$$

(A.20)

Then, similarly as in (A.10) through (A.16), we have

$$\left\|G_J\left(\hat{\delta}_n^*;\tau\right) - L_J\left(\hat{\delta}_n^*;\tau\right)\right\| = o_p\left(J^{-1/2}\right)$$

(A.21)

Since $G_J(a;\tau)$ and $L_J(a;\tau)$ are close at both $\hat{\delta}_n$ and $\hat{\delta}_n^*$ and $\hat{\delta}_n^*$ minimises $\left\|L_J(a;\tau)\right\|$, $\hat{\delta}_n$ is close to minimising $\left\|L_J(a;\tau)\right\|$. So, it follows that

$$\begin{aligned}
\left\|L_J\left(\hat{\delta}_n;\tau\right)\right\| - o_P\left(J^{-1/2}\right) &\leq \left\|G_J\left(\hat{\delta}_n;\tau\right)\right\| \\
&\leq \left\|G_J\left(\hat{\delta}_n^*;\tau\right)\right\| + o_p\left(J^{-1/2}\right) \\
&\leq \left\|L_J\left(\hat{\delta}_n^*;\tau\right)\right\| + o_p\left(J^{-1/2}\right).
\end{aligned}$$

(A.22)

That is,

$$\left\|L_J\left(\hat{\delta}_n;\tau\right)\right\| = \left\|L_J\left(\hat{\delta}_n^*;\tau\right)\right\| + o_p\left(J^{-1/2}\right).$$

(A.23)

Squaring both sides gives that

$$\left\| L_J\left(\hat{\delta}_n;\tau\right)\right\|^2 = \left\| L_J\left(\hat{\delta}_n^*;\tau\right)\right\|^2 + o_p\left(J^{-1}\right).$$

(A.24)

The cross product term being absorbed into the $o_p\left(J^{-1}\right)$ is because, from the differentiability of $G\left(a;\tau\right)$ at δ_n, it follows

$$\left\| G\left(\hat{\delta}_n^*;\tau\right)\right\| \le \left\| \Gamma_{\delta_n}\left(\hat{\delta}_n^* - \delta_n\right)\right\| + o\left(\left\|\hat{\delta}_n^* - \delta_n\right\|\right) = O_p\left(J^{-1/2}\right).$$

(A.25)

Similarly as in (A.8) through (A.11), it follows

$$\left\| G_J\left(\hat{\delta}_n^*;\tau\right)\right\| \le o_P\left(J^{-1/2}\right) + \left\| G\left(\hat{\delta}_n^*;\tau\right)\right\| + \left\| G_J\left(\delta_n;\tau\right)\right\|,$$

(A.26)

which gives $\left\| G_J\left(\hat{\delta}_n^*;\tau\right)\right\| = O_p\left(J^{-1/2}\right)$. Hence, given (A.21), $\left\| L_J\left(\hat{\delta}_n^*;\tau\right)\right\|$ is of order $O_p\left(J^{-1/2}\right)$. The quadratic form $\left\| L_J\left(\delta;\tau\right)\right\|^2$ has a simple expansion

$$\left\| L_J\left(\delta;\tau\right)\right\|^2 = \left\| L_J\left(\hat{\delta}_n^*;\tau\right)\right\|^2 + \left\| \Gamma_{\delta_n}\left(\delta - \hat{\delta}_n^*\right)\right\|^2,$$

(A.27)

about its global minimum. The cross-product term vanishes, because

$$L_J\left(\delta;\tau\right) - L_J\left(\hat{\delta}_n^*;\tau\right) = \Gamma_{\delta_n}\left(\delta - \hat{\delta}_n^*\right),$$

(A.28)

which rearranges to

$$L_J\left(\delta;\tau\right) = \Gamma_{\delta_n}\left(\delta - \hat{\delta}_n^*\right) + L_J\left(\hat{\delta}_n^*;\tau\right).$$

(A.29)

Since $\hat{\delta}_n^*$ minimises $\left\| L_J\left(\delta;\tau\right)\right\|$, the residual vector $L_J\left(\hat{\delta}_n^*;\tau\right)$ must be orthogonal to the columns of Γ_{δ_n}.

Substituting $\hat{\delta}_n$ in (A.27) gives $\left\|L_J\left(\hat{\delta}_n;\tau\right)\right\|^2 = \left\|L_J\left(\hat{\delta}_n^*;\tau\right)\right\|^2 + \left\|\Gamma_{\delta_n}\left(\hat{\delta}_n - \hat{\delta}_n^*\right)\right\|^2$. Equating this to (A.24) gives $\left\|\Gamma_{\delta_n}\left(\hat{\delta}_n - \hat{\delta}_n^*\right)\right\| = o_p\left(J^{-1/2}\right)$. Since Γ_{δ_n} is full rank, this is equivalent to

$$\sqrt{J}\left(\hat{\delta}_n - \delta_n\right) = \sqrt{J}\left(\hat{\delta}_n^* - \delta_n\right) + o_p(1).$$

Then, it follows that

$$\sqrt{J}\left(\hat{\delta}_n - \delta_n\right) \xrightarrow{d} N\left(0, \Gamma_{\delta_n}{}^{-1} V^* \Gamma_{\delta_n}{}^{-1}\right).$$

Q.E.D.

We already know how to consistently estimate Γ_{δ_n}: use expression (3.15). $h_j\left(\delta_n;\tau\right)$ and $r_j\left(\tau\right)$ can be respectively by $h_j\left(\hat{\delta}_n;\hat{\tau}\right)$ and $r_j\left(\hat{\tau}\right)$. Γ_τ can be estimated as follows:

$$\hat{\Gamma}_\tau = \frac{1}{J}\sum_{j=1}^{J}\left[\sum_{s<t}\frac{1}{T_j}\left\{\begin{array}{l} 1\left\{-1 < \begin{pmatrix} x_{jt}-x_{js} \\ \hat{\xi}_{jt}-\hat{\xi}_{js}\end{pmatrix}' \hat{\delta}_n < 1\right\} \\ \left(\begin{array}{l} 1\left\{-1 < \begin{pmatrix} x_{jt}-x_{js} \\ \hat{\xi}_{jt}-\hat{\xi}_{js}\end{pmatrix}' \hat{\delta}_n < w_{js}-1\right\} - \\ 1\left\{0 < \begin{pmatrix} x_{jt}-x_{js} \\ \hat{\xi}_{jt}-\hat{\xi}_{js}\end{pmatrix}' \hat{\delta}_n < w_{js}\right\} - \\ 1\left\{-w_{jt} < \begin{pmatrix} x_{jt}-x_{js} \\ \hat{\xi}_{jt}-\hat{\xi}_{js}\end{pmatrix}' \hat{\delta}_n < 0\right\} + \\ 1\left\{1-w_{jt} < \begin{pmatrix} x_{jt}-x_{js} \\ \hat{\xi}_{jt}-\hat{\xi}_{js}\end{pmatrix}' \hat{\delta}_n < 1\right\} \end{array}\right) \begin{pmatrix} x_{jt}-x_{js} \\ \hat{\xi}_{jt}-\hat{\xi}_{js}\end{pmatrix}\left(-\left(\hat{x}_{jt}-\hat{x}_{js}\right)'\hat{\theta}_n\right) + \\ u\left(w_{js}, w_{jt}, \begin{pmatrix} x_{jt}-x_{js} \\ \hat{\xi}_{jt}-\hat{\xi}_{js}\end{pmatrix}' \delta_n\right)\begin{bmatrix} \mathbb{Z} \\ -\left(\hat{x}_{jt}-\hat{x}_{js}\right)' \end{bmatrix} \end{array}\right.\right]$$

where $\widehat{x}_{jt}$ denotes the vector of observed explanatory variables in (3.7), $\mathbb{Z}$ is a matrix of zeros, with the row dimension being equal to the length of x_{js} and column dimension being equal to the length of $\widehat{x}_{jt}$.

Appendix 2

Table A1 Structural Parameter Estimates for without correcting for selection

Drink	Variable	Coef.		S.E.
Fizzy				
	Fizzy Diet	0.172	**	(0.078)
	Fizzy Vitamins	0.139		(0.136)
	Fizzy Nocolours	0.053		(0.104)
	Juice Diet	-0.085	*	(0.051)
	Juice Vitamins	0.091		(0.153)
	Juice Nocolours	-0.151		(0.142)
	Cordial Diet	0.069		(0.068)
	Cordial Vitamins	-0.004		(0.087)
	Cordial Nocolours	-0.059		(0.183)
	Log Fizzy Price	0.125	**	(0.052)
	Log Juice Price	-0.100	***	(0.024)
	Log Cordial Price	-0.025		(0.052)
	Log Real Total Expenditure	0.232		(0.181)
Juice				
	Fizzy Diet	-0.122		(0.107)
	Fizzy Vitamins	-0.020		(0.031)
	Fizzy Nocolours	-0.119	***	(0.044)
	Juice Diet	-0.006		(0.040)
	Juice Vitamins	0.020		(0.070)
	Juice Nocolours	0.053		(0.039)
	Cordial Diet	-0.079		(0.102)
	Cordial Vitamins	0.036		(0.070)
	Cordial Nocolours	0.029		(0.037)
	Log Fizzy Price	-0.100	***	(0.024)
	Log Juice Price	0.152	***	(0.027)
	Log Cordial Price	-0.053	***	(0.020)
	Log Real Total Expenditure	0.309	***	(0.037)
Cordial				
	Fizzy Diet	-0.013		(0.153)
	Fizzy Vitamins	-0.014		(0.557)
	Fizzy Nocolours	0.087		(0.307)
	Juice Diet	-0.024		(0.251)
	Juice Vitamins	-0.120		(0.630)
	Juice Nocolours	-0.031		(0.424)
	Cordial Diet	0.077		(0.338)
	Cordial Vitamins	-0.109		(0.298)
	Cordial Nocolours	-0.036		(0.229)
	Log Fizzy Price	-0.025		(0.052)
	Log Juice Price	-0.053	***	(0.020)
	Log Cordial Price	0.078		(0.058)
	Log Real Total Expenditure	-0.396	***	(0.058)

Note: * Significant at 10%; ** Significant at 5%; *** Significant at 1%. Fizzy Diet: diet Fizzy; Fizzy Vitamins: Fizzy with extra vitamins; Fizzy Nocolours: Fizzy with no added colours or preservatives; Juice Diet: Juice with no added sugar; Juice Vitamins: Juice with extra vitamins; Juice Nocolours: Juice with no added colours or preservatives; Cordial Diet: diet Cordial; Cordial Vitamins: Cordial with extra vitamins; Cordial Nocolours: Cordial with no added colours or preservatives.

Table A2 Average Total Expenditure Elasticities and Uncompensated and Compensated Price Elasticities for without correcting for selection

		Fizzy	Juice	Cordial
Total expenditure		0.965*** (0.073)	1.062*** (0.011)	0.958*** (0.087)
Uncompensated	Fizzy	-1.083*** (0.260)	0.014 (0.028)	0.130 (0.573)
	Juice	0.177** (0.074)	-0.988*** (0.015)	-0.184 (0.504)
	Cordial	-0.141*** (0.034)	-0.040*** (0.010)	-1.061*** (0.388)
Compensated	Fizzy	-0.990*** (0.265)	0.117*** (0.028)	0.222 (0.572)
	Juice	0.895*** (0.035)	-0.197*** (0.018)	0.529 (0.498)
	Cordial	0.012 (0.034)	0.128*** (0.009)	-0.909** (0.391)

Note: Standard errors are in parenthesis. * Significant at 10%; ** Significant at 5%; *** Significant at 1%.

Table A3 Average Partial Elasticities w.r.t. Attributes for without correcting for selection

Drink	Attributes	Coef.		S.E.
Fizzy				
	Fizzy Diet	0.244	**	(0.111)
	Fizzy Vitamins	0.197		(0.193)
	Fizzy Nocolours	0.076		(0.148)
	Juice Diet	-0.12	*	(0.072)
	Juice Vitamins	0.129		(0.218)
	Juice Nocolours	-0.215		(0.201)
	Cordial Diet	0.098		(0.096)
	Cordial Vitamins	-0.006		(0.123)
	Cordial Nocolours	-0.083		(0.260)
Juice				
	Fizzy Diet	0.094		(0.073)
	Fizzy Vitamins	-0.049		(0.043)
	Fizzy Nocolours	-0.008		(0.012)
	Juice Diet	-0.048	***	(0.018)
	Juice Vitamins	-0.002		(0.016)
	Juice Nocolours	0.008		(0.028)
	Cordial Diet	0.021		(0.016)
	Cordial Vitamins	-0.032		(0.041)
	Cordial Nocolours	0.014		(0.028)
Cordial				
	Fizzy Diet	-0.090	***	(0.034)
	Fizzy Vitamins	0.524	***	(0.062)
	Fizzy Nocolours	-0.022		(0.260)
	Juice Diet	-0.023		(0.944)
	Juice Vitamins	0.148		(0.520)
	Juice Nocolours	-0.041		(0.426)
	Cordial Diet	-0.204		(1.068)
	Cordial Vitamins	-0.053		(0.719)
	Cordial Nocolours	0.130		(0.573)

Note: * Significant at 10%; ** Significant at 5%; *** Significant at 1%. Fizzy Diet: diet Fizzy; Fizzy Vitamins: Fizzy with extra vitamins; Fizzy Nocolours: Fizzy with no added colours or preservatives; Juice Diet: Juice with no added sugar; Juice Vitamins: Juice with extra vitamins; Juice Nocolours: Juice with no added colours or preservatives; Cordial Diet: diet Cordial; Cordial Vitamins: Cordial with extra vitamins; Cordial Nocolours: Cordial with no added colours or preservatives.

Table A4 Structural Parameter Estimates for observations with household weekly income lower than median

Drink	Variable	Coef.		S.E.
Fizzy				
	Fizzy Diet	0.066		(0.137)
	Fizzy Vitamins	-0.043		(0.172)
	Fizzy Nocolours	-0.032		(0.131)
	Juice Diet	-0.142	*	(0.084)
	Juice Vitamins	-0.131		(0.214)
	Juice Nocolours	0.036		(0.170)
	Cordial Diet	0.006		(0.129)
	Cordial Vitamins	-0.204		(0.166)
	Cordial Nocolours	0.204		(0.197)
	Log Fizzy Price	0.045		(0.047)
	Log Juice Price	-0.100	***	(0.032)
	Log Cordial Price	0.056		(0.038)
	Log Real Total Expenditure	0.213		(0.142)
Juice				
	Fizzy Diet	-0.109		(0.148)
	Fizzy Vitamins	-0.006		(0.042)
	Fizzy Nocolours	-0.157	**	(0.069)
	Juice Diet	0.039		(0.055)
	Juice Vitamins	0.129		(0.091)
	Juice Nocolours	0.023		(0.069)
	Cordial Diet	-0.112		(0.132)
	Cordial Vitamins	0.183	*	(0.097)
	Cordial Nocolours	-0.005		(0.051)
	Log Fizzy Price	-0.100	***	(0.032)
	Log Juice Price	0.111	***	(0.037)
	Log Cordial Price	-0.011		(0.038)
	Log Real Total Expenditure	0.308	***	(0.044)
Cordial				
	Fizzy Diet	0.109		(0.199)
	Fizzy Vitamins	-0.002		(0.184)
	Fizzy Nocolours	0.409		(0.284)
	Juice Diet	-0.241		(0.462)
	Juice Vitamins	-0.611		(0.564)
	Juice Nocolours	-0.027		(0.353)
	Cordial Diet	-0.079		(0.319)
	Cordial Vitamins	-0.791		(0.802)
	Cordial Nocolours	0.088		(0.178)
	Log Fizzy Price	0.056		(0.038)
	Log Juice Price	-0.011		(0.038)
	Log Cordial Price	-0.045		(0.047)
	Log Real Total Expenditure	-0.318	***	(0.080)

Note: * Significant at 10%; ** Significant at 5%; *** Significant at 1%. Fizzy Diet: diet Fizzy; Fizzy Vitamins: Fizzy with extra vitamins; Fizzy Nocolours: Fizzy with no added colours or preservatives; Juice Diet: Juice with no added sugar; Juice Vitamins: Juice with extra vitamins; Juice Nocolours: Juice with no added colours or preservatives; Cordial Diet: diet Cordial; Cordial Vitamins: Cordial with extra vitamins; Cordial Nocolours: Cordial with no added colours or preservatives.

Table A5 Average Total Expenditure Elasticities and Uncompensated and Compensated Price Elasticities for observations with household weekly income lower than median

		Fizzy	Juice	Cordial
Total expenditure		1.094*** (0.063)	1.054*** (0.018)	1.093*** (0.062)
Uncompensated	Fizzy	-0.656** (0.333)	0.089* (0.047)	-0.132 (0.532)
	Juice	0.076 (0.079)	-1.003*** (0.024)	-1.317 (1.335)
	Cordial	-0.170*** (0.055)	-0.048*** (0.016)	-0.854*** (0.297)
Compensated	Fizzy	-0.535 (0.336)	0.205*** (0.046)	-0.012 (0.532)
	Juice	0.845*** (0.056)	-0.262*** (0.032)	-0.548 (1.342)
	Cordial	0.034 (0.055)	0.148*** (0.015)	-0.650** (0.296)

Note: Standard errors are in parenthesis. * Significant at 10%; ** Significant at 5%; *** Significant at 1%.

Table A6 Average Partial Elasticities w.r.t. Attributes for observations with household weekly income lower than median

Drink	Attributes	Coef.		S.E.
Fizzy				
	Fizzy Diet	0.112		(0.232)
	Fizzy Vitamins	-0.073		(0.291)
	Fizzy Nocolours	-0.054		(0.221)
	Juice Diet	-0.239	*	(0.142)
	Juice Vitamins	-0.222		(0.362)
	Juice Nocolours	0.060		(0.287)
	Cordial Diet	0.009		(0.219)
	Cordial Vitamins	-0.346		(0.282)
	Cordial Nocolours	0.344		(0.333)
Juice				
	Fizzy Diet	0.103		(0.069)
	Fizzy Vitamins	-0.052		(0.071)
	Fizzy Nocolours	-0.003		(0.020)
	Juice Diet	-0.076	**	(0.033)
	Juice Vitamins	0.019		(0.026)
	Juice Nocolours	0.062		(0.044)
	Cordial Diet	0.011		(0.033)
	Cordial Vitamins	-0.054		(0.064)
	Cordial Nocolours	0.089	*	(0.047)
Cordial				
	Fizzy Diet	-0.018		(0.063)
	Fizzy Vitamins	0.512	***	(0.074)
	Fizzy Nocolours	0.181		(0.331)
	Juice Diet	-0.003		(0.306)
	Juice Vitamins	0.680		(0.473)
	Juice Nocolours	-0.401		(0.770)
	Cordial Diet	-1.017		(0.938)
	Cordial Vitamins	-0.045		(0.588)
	Cordial Nocolours	-0.132		(0.532)

Note: * Significant at 10%; ** Significant at 5%; *** Significant at 1%. Fizzy Diet: diet Fizzy; Fizzy Vitamins: Fizzy with extra vitamins; Fizzy Nocolours: Fizzy with no added colours or preservatives; Juice Diet: Juice with no added sugar; Juice Vitamins: Juice with extra vitamins; Juice Nocolours: Juice with no added colours or preservatives; Cordial Diet: diet Cordial; Cordial Vitamins: Cordial with extra vitamins; Cordial Nocolours: Cordial with no added colours or preservatives.

Table A7 Structural Parameter Estimates for observations with household weekly income higher than median

Drink	Variable	Coef.		S.E.
Fizzy				
	Fizzy Diet	0.132	*	(0.079)
	Fizzy Vitamins	3.597	***	(0.441)
	Fizzy Nocolours	-0.440	***	(0.098)
	Juice Diet	-0.263	***	(0.055)
	Juice Vitamins	3.146	***	(0.349)
	Juice Nocolours	-0.028		(0.095)
	Cordial Diet	-0.070		(0.074)
	Cordial Vitamins	-0.245	**	(0.101)
	Cordial Nocolours	0.431	***	(0.079)
	Log Fizzy Price	-0.020		(0.029)
	Log Juice Price	0.018		(0.030)
	Log Cordial Price	0.002		(0.011)
	Log Real Total Expenditure	0.699	***	(0.066)
Juice				
	Fizzy Diet	0.118		(0.139)
	Fizzy Vitamins	0.112	**	(0.051)
	Fizzy Nocolours	-0.060		(0.046)
	Juice Diet	0.009		(0.038)
	Juice Vitamins	0.170	*	(0.090)
	Juice Nocolours	-0.040		(0.048)
	Cordial Diet	0.187		(0.136)
	Cordial Vitamins	0.022		(0.066)
	Cordial Nocolours	0.018		(0.030)
	Log Fizzy Price	-0.099	**	(0.040)
	Log Juice Price	0.087	**	(0.035)
	Log Cordial Price	-0.105	***	(0.024)
	Log Real Total Expenditure	0.113	*	(0.064)
Cordial				
	Fizzy Diet	-0.033		(0.056)
	Fizzy Vitamins	-0.481	***	(0.137)
	Fizzy Nocolours	0.165	***	(0.056)
	Juice Diet	0.044		(0.028)
	Juice Vitamins	-0.428	***	(0.152)
	Juice Nocolours	0.084	**	(0.040)
	Cordial Diet	-0.031		(0.092)
	Cordial Vitamins	-0.042		(0.056)
	Cordial Nocolours	-0.126	***	(0.027)
	Log Fizzy Price	0.002		(0.011)
	Log Juice Price	-0.105	***	(0.024)
	Log Cordial Price	0.103	***	(0.027)
	Log Real Total Expenditure	-0.347	***	(0.039)

Note: * Significant at 10%; ** Significant at 5%; *** Significant at 1%. Fizzy Diet: diet Fizzy; Fizzy Vitamins: Fizzy with extra vitamins; Fizzy Nocolours: Fizzy with no added colours or preservatives; Juice Diet: Juice with no added sugar; Juice Vitamins: Juice with extra vitamins; Juice Nocolours: Juice with no added colours or preservatives; Cordial Diet: diet Cordial; Cordial Vitamins: Cordial with extra vitamins; Cordial Nocolours: Cordial with no added colours or preservatives.

Table A8 Average Total Expenditure Elasticities and Uncompensated and Compensated Price Elasticities for observations with household weekly income higher than median

		Fizzy	Juice	Cordial
Total expenditure		1.780*** (0.074)	1.039*** (0.022)	0.400*** (0.068)
Uncompensated	Fizzy	-1.023*** (0.032)	-0.034** (0.014)	0.004 (0.020)
	Juice	0.020 (0.033)	-0.970*** (0.012)	-0.182*** (0.041)
	Cordial	0.002 (0.013)	-0.036*** (0.008)	-0.822*** (0.046)
Compensated	Fizzy	-0.873*** (0.027)	0.053*** (0.014)	0.037* (0.021)
	Juice	1.410*** (0.086)	-0.159*** (0.022)	0.131* (0.077)
	Cordial	0.243*** (0.016)	0.104*** (0.009)	-0.768*** (0.044)

Note: Standard errors are in parenthesis. * Significant at 10%; ** Significant at 5%; *** Significant at 1%.

Table A9 Average Partial Elasticities w.r.t. Attributes for observations with household weekly income higher than median

Drink	Attributes	Coef.		S.E.
Fizzy				
	Fizzy Diet	0.147	*	(0.088)
	Fizzy Vitamins	4.015	***	(0.493)
	Fizzy Nocolours	-0.492	***	(0.109)
	Juice Diet	-0.293	***	(0.061)
	Juice Vitamins	3.512	***	(0.389)
	Juice Nocolours	-0.031		(0.106)
	Cordial Diet	-0.078		(0.082)
	Cordial Vitamins	-0.274	**	(0.113)
	Cordial Nocolours	0.481	***	(0.088)
Juice				
	Fizzy Diet	0.041		(0.048)
	Fizzy Vitamins	0.039	**	(0.018)
	Fizzy Nocolours	-0.021		(0.016)
	Juice Diet	0.003		(0.013)
	Juice Vitamins	0.059	*	(0.031)
	Juice Nocolours	-0.014		(0.017)
	Cordial Diet	0.065		(0.047)
	Cordial Vitamins	0.008		(0.023)
	Cordial Nocolours	0.006		(0.010)
Cordial				
	Fizzy Diet	-0.056		(0.097)
	Fizzy Vitamins	-0.831	***	(0.237)
	Fizzy Nocolours	0.285	***	(0.096)
	Juice Diet	0.076		(0.049)
	Juice Vitamins	-0.740	***	(0.263)
	Juice Nocolours	0.146	**	(0.069)
	Cordial Diet	-0.053		(0.159)
	Cordial Vitamins	-0.072		(0.096)
	Cordial Nocolours	-0.218	***	(0.047)

Note: * Significant at 10%; ** Significant at 5%; *** Significant at 1%. Fizzy Diet: diet Fizzy; Fizzy Vitamins: Fizzy with extra vitamins; Fizzy Nocolours: Fizzy with no added colours or preservatives; Juice Diet: Juice with no added sugar; Juice Vitamins: Juice with extra vitamins; Juice Nocolours: Juice with no added colours or preservatives; Cordial Diet: diet Cordial; Cordial Vitamins: Cordial with extra vitamins; Cordial Nocolours: Cordial with no added colours or preservatives.

References

lan, Sule, Bo E. Honoré, Luojia Hu, and Søren Leth-Petersen. 2014. “Estimation of Panel Data Regression Models with Two-Sided Censoring or Truncation.” *Journal of Econometric Methods* 3: 1–20.

Alston, J. M., J. A. Chalfant, and N. E. Piggott. 2001. “Incorporating Demand Shifters in the Almost Ideal Demand System.” *Economics Letters* 70 (1): 73–78.

Andreyeva, Tatiana, Michael W Long, and Kelly D Brownell. 2010. “The Impact of Food Prices on Consumption: A Systematic Review of Research on the Price Elasticity of Demand for Food.” *American Journal of Public Health* 100 (2): 216–22. doi:10.2105/ajph.2008.151415.

Arndt, C. 1999. “Demand For Herbicide In Corn: An Entropy Approach Using Micro-Level Dataa.” *Journal of Agricultural and Resource Economics* 24 (01): 204–21.

Arndt, C., S. Liu, and P.l Preckel. 1999. “On Dual Approaches to Demand Systems Estimation in the Presence of Binding Quantity Constraints.” *Applied Economics* 31 (8): 999–1008.

Blundell, Richard, Panos Pashardes, and G. Weber. 1993. “What Do We Learn about Consumer Demand Patterns from Micro Data?” *The American Economic Review* 83 (3). JSTOR: 570–97. http://www.jstor.org/stable/10.2307/2117534.

Briggs, A. D. M., O. T. Mytton, R. Tiffin, M. Rayner, P. Scarborough, and A. Kehlbacher. 2013. “Overall and Income Specific Effect on Prevalence of Overweight and Obesity of 20% Sugar Sweetened Drink Tax in UK: Econometric and Comparative Risk Assessment Modelling Study.” *BMJ* 347: f6189.

Brownell, K. D., and T. R. Frieden. 2009. “Ounces of Prevention — The Public Policy Case for Taxes on Sugared Beverages.” *New England Journal of Medicine* 360 (18): 1805–8.

Buse, A. 1998. “Testing Homogeneity in the Linearized Almost Ideal Demand System.” *American Journal of Agricultural Economics* 80 (1): 208–20.

Carlsson, Fredrik, and Peter Martinsson. 2003. “Design Techniques for Stated Preference Methods in Health Economics.” *Health Economics* 12 (May 2001): 281–94. doi:10.1002/hec.729.

Chalfant, JA. 1987. “A Globally Flexible, Almost Ideal Demand System.” *Journal of Business & Economic Statistics* 5 (2): 233–42. http://www.tandfonline.com/doi/abs/10.1080/07350015.1987.10509581.

Christensen, L. R., D. W. Jorgenson, and L. J. Lau. 1975. “Transcendental Logarithmic Utility Functions.” *The American Economic Review* 65 (3): 367–83.

Coast, Joanna, Hareth Al-Janabi, Eileen J Sutton, Susan A Horrocks, A Jane Vosper, Dawn R Swancutt, and Terry N Flynn. 2012. “Using Qualitative Methods for Attribute Development for Discrete Choice Experiments: Issues and Recommendations.” *Health*

Economics 21 (6): 730–41.

de Silva-Sanigorski, Andrea, Elizabeth Waters, Hanny Calache, Michael Smith, Lisa Gold, Mark Gussy, Anthony Scott, Kathleen Lacy, and Monica Virgo-Milton. 2011. “Splash!: A Prospective Birth Cohort Study of the Impact of Environmental, Social and Family-Level Influences on Child Oral Health and Obesity Related Risk Factors and Outcomes.” *BMC Public Health* 11 (505): 1–9.

Deaton, A., and M. Irish. 1984. “Statistical Models for Zero Expenditures in Household Budgets.” *Journal of Public Economics* 23: 59–80.

Deaton, A., and J. Muellbauer. 1980a. “An Almost Ideal Demand System.” *American Economic Review* 70 (3): 312–26.

———. 1980b. *Economics and Consumer Behavior*. Cambridge; New York: Cambridge University Press.

Dubois, L., A. Farmer, M. Girard, and K. Peterson. 2007. “Regular Sugar-Sweetened Beverage Consumption between Meals Increases Risk of Overweight among Preschool-Aged Children{A Figure Is Presented}.” *Journal of the American Dietetic Association* 107 (6): 924–34.

Edgerton, David L. 1997. “Weak Separability and the Estimation of Elasticities in Multistage Demand Systems.” *American Journal of Agricultural Economics* 79 (1): 62–79.

Etilé, F., and A. Sharma. 2015. “Do High Consumers of Sugar-Sweetened Beverages Respond Differently to Price Changes? A Finite Mixture IV-Tobit Approach.” *Health Economics*.

Finkelstein, E. A., C. Zhen, M. Bilger, J. Nonnemaker, A. M. Farooqui, and J. E. Todd. 2013. “Implications of a Sugar-Sweetened Beverage (SSB) Tax When Substitutions to Non-Beverage Items Are Considered.” *Journal of Health Economics* 32 (1): 219–39.

Fry, J. M., T. R. Fry, and K. R. McLaren. 1996. “The Stochastic Specification of Demand Share Equations: Restricting Budget Shares to the Unit Simplex.” *Journal of Econometrics* 73 (2): 377–85.

Gallant, A. R. 1981. “On the Bias in Flexible Functional Forms and an Essentially Unbiased Form.” *Journal of Econometrics* 15: 211–45.

Heien, D., and C. R. Wessells. 1990. “Demand Systems Estimation with Microdata: A Censored Regression Approach.” *Journal of Business & Economic Statistics* 8 (3). JSTOR: 365–71.

Hoare, Alexandria, Monica Virgo-Milton, Rachel Boak, Lisa Gold, Elizabeth Waters, Mark Gussy, Hanny Calache, Michael Smith, and Andrea M de Silva. 2014. “A Qualitative Study of the Factors That Influence Mothers When Choosing Drinks for Their Young Children.” *BMC Research Notes* 7: 430.

Hoderlein, S., and S. Mihaleva. 2008. “Increasing the Price Variation in a Repeated Cross Section.” *Journal of Econometrics* 147 (2): 316–25.

Honoré, B. E. 1992. “Trimmed LAD and Least Squares Estimation of Truncated and Censored Regression Models with Fixed Effects.” *Econometrica* 60 (3): 533–65.

———. 2008. *On Marginal Effects in Semiparametric Censored Regression Models*.

Honoré, B. E., and J. L. Powell. 1994. “Pairwise Difference Estimators of Censored and Truncated Regression Models.” *Journal of Econometrics* 64 (1-2): 241–78.

Huber, Joel, and Klaus Zwerina. 1996. “The Importance of Utility Balance in Efficient Choice Designs.” *Journal of Marketing Research* 33 (3): 307–17. doi:10.2307/3152127.

Kim, Jaehwan, Greg M. Allenby, and Peter E. Rossi. 2002. “Modeling Consumer Demand for Variety.” *Marketing Science* 21 (3): 229–50. doi:10.1287/mksc.21.3.229.143.

Lee, L. F., and M. M. Pitt. 1986. “Microeconometric Demand Systems with Binding Nonnegativity Constraints: The Dual Approach.” *Econometrica* 54 (5): 1237–42.

———. 1987. “Microeconometric Models of Rationing, Imperfect Markets, and Non-Negativity Constraints.” *Journal of Econometrics* 36 (1-2): 89–110.

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

Lewbel, A. 1989. “Identification and Estimation of Equivalence Scales under Weak Separability.” *The Review of Economic Studies* 56 (2): 311–16.

Malik, V. S., M. B. Schulze, and F. B. Hu. 2006. “Intake of Sugar-Sweetened Beverages and Weight Gain: A Systematic Review.” *American Journal of Clinical Nutrition* 84 (2): 274–88.

Mann, H. B., and a. Wald. 1943. “On Stochastic Limit and Order Relationships.” *The Annals of Mathematical Statistics* 14 (3): 217–26. doi:10.1214/aoms/1177731415.

Manser, M. E., and R. J. McDonald. 1988. “An Analysis of Substitution Bias in Measuring Inflation, 1959-85.” *Econometrica* 56 (4): 909–30.

Meyerhoefer, C. D., C. K. Ranney, and D. E. Sahn. 2005. “Consistent Estimation of Censored Demand Systems Using Panel Data.” *American Journal of Agricultural Economics* 87 (3): 660–72.

Moschini, G. 1995. “Units of Measurement and the Stone Index in Demand System Estimation.” *American Journal of Agricultural Economics* 77 (1): 63–68.

Mundlak, Y. 1978. “On the Pooling of Time Series and Cross Section Data.” *Econometrica* 46: 69–85.

Pakes, Ariel, and David Pollard. 1989. “Simulation and the Asymptotics of Optimization Estimators.” *Econometrica* 57 (5): 1027–57. doi:10.2307/1913622.

Perali, Federico, and Jean-Paul Chavas. 2000. “Estimation of Censored Demand Equations from Large Cross-Section Data." *American Journal of Agricultural Economics* 82 (4): 1022–37.

Pollak, R. A., and T. J. Wales. 1992. *Demand System Specification and Estimation*. Oxford University Press.

Pollak, R.A., and T.J. Wales. 1981. "Demographic Variables in Demand Analysis." *Econometrica* 49 (6): 1533–51.

Powell, J. L. 1986. "Symmetrically Trimmed Least Squares Estimation for Tobit Models." *Econometrica* 54 (6): 1435–60. doi:10.2307/1914308.

Scott, Anthony, Julia Witt, John Humphreys, Catherine Joyce, Guyonne Kalb, Sung Hee Jeon, and Matthew McGrail. 2013. "Getting Doctors into the Bush: General Practitioners' Preferences for Rural Location." *Social Science and Medicine* 96: 33–44.

Sharma, A., K. Hauck, B. Hollingsworth, and L. Siciliani. 2014. "The Effects of Taxing Sugar-Sweetened Beverages across Different Income Groups." *Health Economics* 23 (9): 1159–84.

Shonkwiler, J.S., and S.T. Yen. 1999. "Two-Step Estimation of a Censored System of Equations." *American Journal of Agricultural Economics* 81 (4): 972–82.

Sivey, Peter, Anthony Scott, Julia Witt, Catherine Joyce, and John Humphreys. 2012. "Junior Doctors' Preferences for Specialty Choice." *Journal of Health Economics* 31: 813–23.

Varian, H. R. 1983. "Non-Parametric Tests of Consumer." *The Review of Economic Studies* 50 (1): 99–110.

Wake, M, P Hardy, L Canterford, M Sawyer, and J B Carlin. 2006. "Overweight, Obesity and Girth of Australian Preschoolers: Prevalence and Socio-Economic Correlates." *International Journal of Obesity* 31 (7): 1044–51.

Wales, T.J., and A.D. Woodland. 1983. "Estimation of Consumer Demand Systems with Binding Non-Negativity Constraints." *Journal of Econometrics* 21: 263–85.

Wooldridge, J. M. 1995. "Selection Corrections for Panel Data Models under Conditional Mean Independence Assumptions." *Journal of Econometrics* 68: 115–32.

———. 2010a. "Chapter 19 Censored Data, Sample Selection, and Atrition." In *Econometric Analysis of Cross Section and Panel Data*, Second Ed, 775–851.

———. 2010b. "Chapter 14 Generalized Method of Moments and Minimum Distance Estimation." In *Econometric Analysis of Cross Section and Panel Data*, Second Ed, 525–58.

———. 2010c. "Chapter 12 M-Estimation, Nonlinear Regression, and Quantile Regression." In *Econometric Analysis of Cross Section and Panel Data*, Second Ed, 397–462.

Yen, S. T., B. H. Lin, and D. M. Smallwood. 2003. "Quasi-and Simulated-Likelihood

Approaches to Censored Demand Systems: Food Consumption by Food Stamp Recipients in the United States." *American Journal of Agricultural Economics* 85 (2): 458–78.

Yen, S.T. 2005. "A Multivariate Sample-Selection Model: Estimating Cigarette and Alcohol Demands with Zero Observations." *American Journal of Agricultural Economics* 87 (May). JSTOR: 453–66.

Yen, S.T., K. Kan, and S.J. Su. 2002. "Household Demand for Fats and Oils: Two-Step Estimation of a Censored Demand System." *Applied Economics* 34 (14): 1799–1806.

Yen, S.T., and B.H. Lin. 2006. "A Sample Selection Approach to Censored Demand Systems." *American Journal of Agricultural Economics* 88 (August): 742–49.

Zhen, C, M K Wohlgenant, S Karns, and P Kaufman. 2011. "Habit Formation and Demand for Sugar-Sweetened Beverages." *American Journal of Agricultural Economics* 93 (1): 175–93.

Zhen, C., E. A. Finkelstein, J. M. Nonnemaker, S. A. Karns, and J. E. Todd. 2014. "Predicting the Effects of Sugar-Sweetened Beverage Taxes on Food and Beverage Demand in a Large Demand System." *American Journal of Agricultural Economics* 96 (1): 1–25.