THE UNIVERSITY OF MELBOURNE

FACULTY OF BUSINESS & ECONOMICS

Melbourne Institute Working Paper Series

Working Paper No. 25/16

Preschool Children's Demand for Sugar Sweetened Beverages: Evidence from Stated-Preference Panel Data

Ou Yang, Peter Sivey, Andrea M. de Silva and Anthony Scott

MELBOURNE INSTITUTE®
of Applied Economic and Social Research

Preschool Children's Demand for Sugar Sweetened Beverages: Evidence from Stated-Preference Panel Data*

Ou Yang[†], Peter Sivey[‡], Andrea M. de Silva[§] and Anthony Scott[‖]

[†] **Department of Econometrics and Business Statistics, Monash University**

[‡] **Department of Economics and Finance, La Trobe University**

[§] **Melbourne Dental School, The University of Melbourne**

[‖] **Melbourne Institute of Applied Economic and Social Research, The University of Melbourne**

Melbourne Institute Working Paper No. 25/16

ISSN 1447-5863 (Online)

ISBN 978-0-73-405222-3

August 2016

* This paper was funded by Australian Research Council Linkage Grant (LP0989576) and by Dental Health Services Victoria. We would like to acknowledge all the families and health services involved in the study; the late Professor Elizabeth Waters without whom this research would not have been possible; and other investigators and researchers: Hanny Calache, Lisa Gold, Mark Gussy, Jacqueline Martin-Kerry, Michael Smith, Elyse O'Callaghan and Monica Virgo-Milton. Correspondence to: Anthony Scott, email <a.scott@unimelb.edu.au>.

Melbourne Institute of Applied Economic and Social Research
The University of Melbourne
Victoria 3010 Australia
***Telephone* (03) 8344 2100**
***Fax* (03) 8344 2111**
***Email* melb-inst@unimelb.edu.au**
***WWW Address* http://www.melbourneinstitute.com**

Abstract

Consumption of sugar sweetened beverages exhibits strong associations with weight gain, obesity, and dental caries, especially in young children. 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 that 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. Our results show that the uncompensated own-price elasticities for sugar-sweetened beverages range from -0.83 to -0.94, demonstrating inelastic but substantial price effects. The marginal effects of demand with respect to nutritional attributes of sugar-sweetened beverages are negligible overall, but are strongest for those in low-income households. High-income households are less responsive to price and not responsive at all to non-price attributes.

Keywords: 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 *et al.* 2006). These problems affect about one-third of children of preschool age, with over 13% of children aged 2-3 years old consuming SSBs every day (Wake *et al.* 2006; Dubois *et al.* 2007).

The public policy debate around taxing SSBs has been gaining momentum in recent years (Brownell and Frieden 2009). Concurrently, a growing body of research evidence addresses questions such as whether taxes on SSBs are effective in reducing consumption and whether large changes in the relative prices of SSBs would eventually reduce obesity (Andreyeva *et al.* 2010, Sturm *et al.* 2010, Powell *et al.* 2009, Fletcher *et al.* 2010a, Fletcher *et al.* 2010b). 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. Recent 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 such as young children 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. In particular, we examine price and cross-price elasticities across SSBs. We make three contributions to the literature. First, 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.

A particular advantage of this approach is that prices and other product attributes are presented to respondents exogenously. This is not the case with revealed preference data where a number of econometric methods need to be used for identification. A further advantage is that an experimental design is used to ensure that the researcher controls the variation in the attributes to maximise the efficiency of the subsequent model estimation. A disadvantage is the potential for hypothetical bias, in that results may not match those in 'real' markets and stated preferences may not be incentive compatible (Beck *et al.* 2016). This is an issue about external validity. The evidence that examines external validity for stated preference experiments in transport, marketing, environment and health is mixed with some studies showing evidence of bias and others not. However, it is also not clear *a priori* the direction of the bias and its nature is likely to be context-specific, influenced by the design of the experiment, and influenced by the choice task (Lancsar and Swait 2014). Indeed in many cases in health, as is the case for this paper, revealed preference data do not exist or could not be collected (Lancsar and Swait 2014).

Unlike stated preference discrete choice experiments, which focus on choosing one good from several alternatives, our stated preference experiment was designed to elicit consumption of SSBs rather than discrete choices. This provides a continuous measure of consumption suited to analysis using a demand system approach that needs to allow for i) the possibility that no SSBs are consumed at all, ii) censoring (zero consumption of at least one SSB conditional on that the total consumption on all SSBs is positive), iii) panel data (multiple scenarios per respondent). This approach contributes to the literature on the analysis of stated preference experiments.

Secondly, we introduce a new semi-parametric fixed-effects censored demand system to overcome a shortcoming in the existing literature. Although there have been a growing number of studies using household expenditure panel data (Zhen *et al.* 2014; Finkelstein *et al.* 2013; Zhen *et al.* 2011; Kim *et al.* 2002; Dong *et al.* 2015; Kyureghian *et al.* 2013; Nayga 1995), these studies generally aggregate household expenditure over the time dimension to generate quarterly or yearly expenditure data and study them using models designed in a cross sectional setting. This has limited the ability of previous studies to control for preference heterogeneity (Meyerhoefer *et al.* 2005, Zhen *et al.* 2014). By using a panel data estimator, our model is better able to control for such heterogeneity.

Finally, one general difficulty that might be encountered using individual expenditure panel data for a specific small group of goods, such as SSBs, is that it might be observed that some households do not purchase any good within this specific group.

In such cases, building a conditional demand system using existing demand system models will exclude observations with zero total expenditure of the group of interest.

Even if one might argue that a composite *numeraire* outside good representing expenditure for all other goods and services can be included to build a complete demand system, having an extremely large *numeraire* good in comparison with the small group of goods of interest might produce numerical difficulty in estimation and impair the accuracy of numerical procedures. In cases where researchers only observe or collect expenditure data for a specific small group of goods for individuals of particular interest within the household, building a conditional demand system is necessary. We investigate this issue by studying how the exclusion of observations with zero expenditure on the group of goods of interest would affect the estimates of a demand system and if consistent estimates can still be obtained by introducing a two-step estimation strategy.

Our results show that the uncompensated own-price elasticities for our three drink categories, Fizzy Drink (carbonated drinks), Juice and Cordial, are respectively -0.943, -0.949 and -0.832 demonstrating inelastic but substantial price effects, which are similar across drink categories. While we use a unique stated-preference approach, these elasticities are broadly comparable with the findings in the previous literature for households using household survey or scanner data, for example a review by Andreyeva *et al.* (2010) found an elasticity range of -0.8 to -1.0 for soft drinks. Another more recent review by Powell *et al.* (2013) found a range of -0.71 to -2.26 for regular carbonated soda, a range of -0.69 to -1.91 for fruit drinks and a range of -0.87 to -1.26 for aggregated SSB category. Recent studies such as Sharma *et al.* (2014) and Zhen *et* *al.* (2014) report similar results. Our results seem to support of the claim that efforts to raise the prices of SSBs would influence consumption.

All compensated cross-price elasticities are positive, suggesting the drink categories are net substitutes. While most average effects of non-price attributes are not statistically significant, we do observe on average, the “No added colours or preservatives” attribute for Juice significantly increasing its consumption by 2.7%. In order to better inform policy, we also estimate our model for high and low-household income sub-samples respectively, and our results highlight substantial discrepancies between their consumption behaviours.

The outline of this paper is set out as follows. Section II introduces the stated preference 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 final section concludes.

2. Data

Data were collected as part of longitudinal birth cohort of 500 children in south-western Victoria, Australia (de Silva-Sanigorski *et al.*., 2011). The stated preference experiment (SPE) consisted of presenting survey respondents, the parents of 24-month-old children, with a series of hypothetical scenarios about the quantities of alternative drink types purchased for their family’s and children’s consumption. The SPE is a labelled design, where respondents choose consumption levels for four broad categories of drinks: Fizzy Drink, Juice, Cordial and Tap Water. The SSB categories were characterised by four attributes: price, sugar content, added vitamins and no added colours or preservatives, discussed in more detail below. The Tap Water category is not described by any attributes.

We undertook extensive pre-piloting 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") 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). For an example consider the drink category labels "Fizzy Drink" (similar to "Soda" in the United States) and "Cordial" (fruit-based concentrated sugar syrup which is consumed diluted with tap water) which are common colloquial terms used in Australia for carbonated SSBs and diluted SSBs. The use of these terms was developed, modified and verified through the pre-pilot process.

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 to ensure the hypothetical choices were as close as possible to real-world choices that parents would be making whilst shopping for drinks, though we did not focus on specific brands. Thirdly, all of our decisions were informed, verified and modified from the iterative process of the qualitative interviews.

Our stated preference experiment features SSB consumption related questions across several hypothetical shopping scenarios. **Error! Reference source not found.** illustrates one example of these hypothetical scenarios. As shown in the examples, the stated preference 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 asked responding parents to make two sequential consumption choices in each scenario: first they must decide how many bottles of each SSB 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 section of the survey containing the experiment included simple instructions for respondents and a practice question: “Imagine you are doing your weekly shop for your FAMILY at the supermarket and you are choosing which drinks to buy.

- “Once you have decided what drinks to buy you must decide what drink to give to your child (in this study).
- You may purchase the items shown below in different forms (e.g. cans of soft drink, Juice boxes, 1.25L soft drinks etc.); HOWEVER for the purpose of this question IMAGINE that these 2 litre bottles and tap water are the ONLY options available to you.

- For each drink, the price, sugar content, and whether each drink has extra vitamins or added colours and preservatives changes. Therefore you MUST fill in all 9 tables.
- There are no right or wrong answers
- Below is an example to show you how to answer each question."

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 (i.e. in place of food). 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 are 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 in 2012.

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 in supermarkets, ‘Diet’ or ‘No Sugar’ or very similar variants were used on the packaging of sugar free drinks, whereas highly sugar sweetened drinks were not explicitly labelled on the front of the bottle 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 SSB 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. The experiment consists of five attributes, three (sugar, vitamins and colour or preservative attributes) with two levels and two (SSB type and price) with three levels giving $2^3 \times 3^2 = 72$ possible alternatives.

As illustrated in Figure 1, the attributes and prices of the three SSBs vary across scenarios. The variation in attribute levels was generated using an experimental design based on a multinomial logit model of drink choice. This 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). We specified a linear indirect utility function in which all attributes enter in an additive and separable manner.

Our approach to producing the experimental design following the pre-piloting stage is in two stages. First we conducted a pilot study among 35 responding families (giving 314 observations) that used an orthogonal design. The data from this pilot were analysed using a simple multinomial logit model. Secondly, we use the coefficient estimates from the pilot study as ‘priors’ in the full experimental design in order to maximise statistical efficiency.

Both pilot and main experimental designs were created using Ngene 1.1 (ChoiceMetrics 2012).

We generated the final design through simulation of a multinomial choice model using the prior values from the pilot study results. Results from these simulations were used to choose a set of attribute levels that minimised the D-error= $(\det\Omega^{-1})^{\frac{1}{K}}$, where Ω^{-1} is the covariance matrix of the multinomial logit model and K is the number of attributes (Huber and Zwerina 1996) in the multinomial logit model. The chosen design produced 36 choice scenarios of three choices (Fizzy Drink, Juice, Cordial). Balancing statistical efficiency considerations, practical survey issues and minimising the demands on respondents, we choose to present each respondent with nine shopping scenarios of three types of SSBs (Juice, Fizzy Drink, Cordial) and the tap water alternative, and so produced four versions of the survey. There were 9x3x4=108 alternatives across the four surveys. Respondents were randomly allocated to one of the four blocks of choice sets in the questionnaire. The survey was mailed to a birth cohort of families of 500 children in south-west Victoria, Australia (de Silva-Sanigorski *et al.*., 2011).

3. Model Specification and Two-step Estimation Strategy

3.1 Model specification

This paper specifies a demand system to jointly study the SSB consumption data gathered from the SPLASH stated preference 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 SSBs their total expenditure on SSBs is zero. Budget shares are not defined or defined as missing values. The neoclassical consumption theoretical framework is based on a positive total budget constraint. Hence, a budg*et al.*location 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 *et al.* 1975, and Gallant 1981). Blundell *et al.* (1993) concluded that unless certain factors are controlled for, aggregate data alone are unlikely to produce reliable estimates of structural price and income coefficients.

When it comes to micro individual-level consumption, zero expenditures for certain goods, or even for the whole category of such goods as SSBs, meat, etc., are common and need to be accounted for in the analysis. If a weakly separable preference is assumed as in, for example, Hoderlein and Mihaleva (2008), Chalfant (1987) and Lewbel (1989), the zero observations on total SSB expenditure could be considered a result of the first-stage budg*et al.*location problem in a multi-stage budgeting framework (Deaton and Muellbauer 1980a, Deaton and Muellbauer 1980b, Edgerton 1997). In particular, assuming weak separability between SSBs and other alternative commercial drinks, such as milk-based drinks, or water, the subjects first make decisions on whether they would like to give some SSBs to their children and then, how much to give them. If they have decided to give some SSBs to their children, i.e. the total expenditure on total SSBs for their children is positive, they proceed to the second stage to make decisions about how to allocate the total budget on SSBs 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.

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 Nielsen Homescan data, another advantage of our data is that they give information on the consumption of popular sugar sweetened beverages specifically by children within households, a key sub-group of interest to policy makers.

We drop observations with missing values for at least one variable producing a dataset including 2,381 valid observations of each alternative, and forming an unbalanced panel of 276 individuals observed over 9 scenarios. Table 1 and Table 2 summarise the number of total, positive and zero observations, and sample means and standard deviations for total SSB expenditure and the respective expenditures of the three drinks. 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. In our sample, there are 94 (34.06%) out of 276 subjects who report giving no SSBs (i.e. zero total SSB expenditure) at all to their children for all the nine scenarios, which account for 70.9% of zero total SSB expenditure observations. The remaining 29.1% of observations with zero total expenditure will therefore be influenced by the attributes in the particular hypothetical scenario.

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 recent literature are developed to model non-consumption (for example Yen and Lin 2006, Meyerhoefer *et al.* 2005, Yen 2005, Perali and Chavas 2000, Heien and Wessells 1990). The only exception is Deaton and Irish (1984), where on top of non-consumption, additional zeros that may arise because of durability so that no purchase is recorded for some households over the limited period of the survey, are taken into account. Of note is that nondurable goods such as non-alcoholic beverages may still exhibit a certain degree of durability if consumers stockpile. Given the design and context of the stated preference experiment, it is admissible to believe that our data are not too vulnerable to the stockpiling issue. Thus, the zero expenditure observations, in our case, represent a genuine corner solution where the subjects deliberately choose not to consume particular SSBs conditional on the attributes of the SSBs in each scenario.

Much of the recent empirical efforts on censored demand system have focused on cross-sectional data. Hence, they suffer from limited ability to control for heterogeneous preferences and limited variation in real price (for example, Yen and Lin 2006, Yen *et al.* 2003, Yen 2005, Yen *et al.* 2002, Perali and Chavas 2000, Arndt 1999, and Heien and Wessells 1990). To the best of our knowledge, Meyerhoefer *et al.* (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 *et al.* (2005)'s estimation strategy. However, one technical difficulty is that built upon the neoclassical budg*et al.*location 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 SSBs; in other words, as discussed, subjects' total expenditure on all the three SSBs has to be positive. Even though Meyerhoefer *et al.* (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.

As shown in Table 1, 50.1% of the total 2381 observations record zero total expenditure on SSBs. Employing Meyerhoefer *et al.* (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 SSB is not systematically correlated to their decision about how much of each SSB to give to their children, estimates conditional on the truncated sample (or equivalently, conditional on positive total expenditure on SSBs) 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.

This study proposes a fixed-effects censored demand system analysis framework to account for the reported zero expenditure observations on specific SSBs (i.e. choose to or not to give their children certain drinks). Conditional on positive total expenditure on SSBs, the subject makes decisions on how to allocate the total expenditure among individual SSBs in scenarios given the price and attributes of each drink. In accordance with neoclassical consumption theory, assuming weakly separable preferences, the conditional direct utility function is defined 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 SSB in scenario t, d_{ljt} denotes the realisation of the lth $(=1,\ldots,L)$ 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:

$$V_{jt}^{*} = \frac{\log c_{jt} - \alpha_0 - \sum_k \alpha_k \log p_{kjt} - \sum_k \sum_l \lambda_{kl} \log p_{kjt} d_{ljt} - \frac{1}{2} \sum_k \sum_i \tilde{\gamma}_{ki} \log p_{kjt} \log p_{ijt} - \sum_k \psi_k \log p_{kjt} \varphi_j}{\beta_0 \prod_k p_{kjt}^{\beta_k}} \tag{0.1}$$

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

The attributes of SSBs 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 is 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 *et al.* 1996). Applying the logarithm version of Roy's Identity, the Marshallian uncompensated demand share equations of the demographically extended Almost Ideal Demand System (AIDS) can be derived. To estimate the system of share equations, appending stochastic error terms gives rise to the econometric specification shown as follows:

$$w_{njt}^{*} = \alpha_n + \sum_l \lambda_{nl} d_{ljt} + \sum_k \gamma_{nk} \log p_{kjt} + \beta_n (\log c_{jt} - \log P_{jt}) + \rho_{nj} + u_{njt} \tag{0.2}$$

where w_{njt}^{*} is the latent expenditure share of SSB $n\ (=1,\ldots K)$ 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.

Following Deaton and Muellbauer (1980a) and others (for example, Edgerton 1997, Capps Jr *et al.* 2003 and Meyerhoefer *et al.* 2005), the above budget share equation is linearised by replacing the highly non-linear P_{jt} by a price index. The Stone price index suggested by Deaton and Muellbauer (1980a) renders the linearised AIDS model inconsistent with the "Closed Under Unit Scaling" property (Moschini 1995). One way to solve this 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 SSB 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 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 shown 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 *et al.* 2005 and Perali and Chavas 2000), adding up is not imposed on the structural parameter estimates. It is however noteworthy 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 (0.2) represent latent shares (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

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

From (0.2), it can be seen that any observation with total expenditure, c_{jt}, of 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 particular, we specify the selection mechanism as an equation of the Tobit form, as follows:

$$c_{jt}^{*} = \alpha_0 + \sum_l \alpha_l d_{ljt} + \sum_k \beta_k \log p_{kjt} + \theta c_{jt}^{a} + \gamma q_{jt}^{w} + \eta_j + \varepsilon_{jt}$$
$$c_{jt} = \max(0,\ c_{jt}^{*}) \tag{0.3}$$

where p_{kjt} denotes the price of SSB k observed at scenario t by subject j, d_{ljt} denotes the realisation of the l th attribute for subject j at scenario t, c_{jt}^{a} denotes the total family SSB expenditure by subject j in the first question in the experiment, q_{jt}^{w} is subject j 's tap water consumption at scenario t, η_j is subject j 's unobservable specific effect, and the random effect is denoted by ε_{jt}. Denote the vector of observable explanatory variables at scenario t as x_{jt}^{s*} and let $\tilde{x}_j^{s*} \equiv \left(x_{j1}^{s*},\ldots,x_{jT}^{s*}\right)'$. ε_{jt} is assumed to be independent of $\tilde{x}_j^{s*}$.

Combining the latent equations in (0.2) and (0.3), for each SSB $n\ (=1,\ldots K)$ introduces the following fixed-effects selection system:

$$w_{njt}^{*} = \alpha_n + \sum_l \lambda_{nl} d_{ljt} + \sum_k \gamma_{nk} \log p_{kjt} + \beta_n (\log c_{jt} - \log P_{jt}) + \rho_{nj} + u_{njt} \tag{0.4}$$

$$c_{jt}^{*} = \alpha_0 + \sum_l \alpha_l d_{ljt} + \sum_k \beta_k \log p_{kjt} + \theta c_{jt}^{a} + \gamma q_{jt}^{w} + \eta_j + \varepsilon_{jt} \tag{0.5}$$

Although all the SSB attributes and prices are orthogonal to the individual specific effect ρ_{nj} in (3.4), the total SSB expenditure for the children $\log c_{jt}$ is not exogenous to ρ_{nj}.

To allow for η_j in the selection equation (3.5) to be correlated with q_{jt}^{w} and c_{jt}^{a}, we can specify a Mundlak-type model (Mundlak 1978; Vella 1992). In particular, assuming η_j depends only on the scenario average of q_{jt}^{w} and c_{jt}^{a}, this correlation can be modelled as a linear projection of η_j on the scenario average tap water consumption and total family SSB expenditure, denoted by $\overline{q_j^{w}}$ and $\overline{c_{jt}^{a}}$:

$$\eta_j = \tau_1 \overline{q_j^{w}} + \tau_2 \overline{c_{jt}^{a}} + \nu_j \tag{0.6}$$

where v_j is assumed to be independent of $\tilde{x}_j^{s*}$ with a zero mean normal distribution. Substituting in η_j, the selection equation (0.5) can be written as:

$$c_{jt}^* = \alpha_0 + \sum_l \alpha_l d_{ljt} + \theta c_{jt}^a + \gamma q_{jt}^w + \sum_k \beta_k \log p_{kjt} + \tau_1 \overline{q_j^w} + \tau_2 \overline{c_{jt}^a} + \xi_{jt} \qquad (0.7)$$

where $\xi_{jt} \equiv v_j + \varepsilon_{jt}$, and $\xi_{jt} \sim N\left(0,\ \sigma_\xi^2\right)$. It should be noted that this test is valid under the condition that the latent variable c_{jt}^* 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 (0.4) is only partially observed, a normal linear fixed-effects estimation strategy for (0.4) still produces inconsistent estimates. Alan *et al.* (2014)'s semi-parametric estimator for two-sided censoring models with fixed effects is employed. Denote the vector of all the observable explanatory variables at scenario t in (0.4) 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 $\left(\rho_{nj}, v_j, \tilde{x}_j, \tilde{\xi}_j\right)$, 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

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

$$\omega_{njt} \equiv u_{njt} - \theta_n \left(\xi_{jt} - v_j\right) \text{ is identically distributed conditional on } \left(\rho_{nj}, v_j, \tilde{x}_j, \tilde{\xi}_j\right) \qquad (0.8)$$

for some unknown scalar θ_n.

Under the alternative (0.8), we have

$$\begin{aligned} w_{njt}^{*} &= \alpha_n + \sum_l \lambda_{nl} d_{ljt} + \sum_k \gamma_{nk} \log p_{kjt} + \beta_n (\log c_{jt} - \log P_{jt}) + \rho_{nj} + \theta_n \left(\xi_{jt} - v_j \right) + \omega_{njt} \\ &= \alpha_n + \sum_l \lambda_{nl} d_{ljt} + \sum_k \gamma_{nk} \log p_{kjt} + \beta_n (\log c_{jt} - \log P_{jt}) + \theta_n \xi_{jt} + \sigma_{nj} + \omega_{njt} \end{aligned} \tag{0.9}$$

where $\sigma_{nj} \equiv \rho_{nj} - \theta_n v_j$. From (0.9), 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 $\mathrm{H}_0: \theta_n = 0$. 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 (0.7) by pooled Tobit.

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

$$\hat{\xi}_{jt} = c_{jt} - \left(\hat{\alpha}_0 + \sum_l \hat{\alpha}_l d_{ljt} + \hat{\theta} c_{jt}^{a} + \hat{\gamma} q_{jt}^{w} + \sum_k \hat{\beta}_k \log p_{kjt} + \hat{\tau}_1 \overline{q_j^{w}} + \hat{\tau}_2 \overline{c_{jt}^{a}} \right) \tag{0.10}$$

Step 3: Estimate the equation

$$\begin{aligned} w_{njt}^{*} &= \alpha_n + \sum_l \lambda_{nl} d_{ljt} + \sum_k \gamma_{nk} \log p_{kjt} + \beta_n (\log c_{jt} - \log P_{jt}) \\ &\quad + \theta_n \hat{\xi}_{jt} + \sigma_{nj} + \omega_{njt}, \end{aligned} \tag{0.11}$$

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

Step4: Test $\mathrm{H}_0: \theta_n = 0$.

As mentioned above, since w^*_{njt} is only partially observed, the linear fixed-effects estimator produces inconsistent estimates. Let δ_n denote coefficients in (0.11) to be estimated and x_{jt} denote the vector of all the observed explanatory variables in (0.11) exculding $\hat{\xi}_{jt}$. 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 $\left(\rho_{nj}, v_j, \tilde{x}_j, \tilde{\xi}_j\right)$, Alan *et al.* (2014)'s semi-parametric estimator produces consistent estimates. The details of this estimator can be found in Appendix 1. Of note is that an explicit exclusion restriction is need, that is, the set of the observable explanatory variables of the share equation (3.5) is a strict subset of that of the selection equation (3.4). Otherwise, including the Tobit residual in the share equation, this estimation suffers from perfect multicollinearity.

Two variables only appear in the selection equation of the total SSB expenditure for the children: the total SSB expenditure for the whole family answered by subjects in the first question of the questionnaire in the experiment, given the two-sequential-question design of the experiment, and the consumption level of tap water of the children. For the former, it is admissible to assume if the total SSB expenditure for the family is changed, this might or might not affect the total SSB expenditure for the children; however, given the attributes and prices of the three SSBs are unchanged, the relative amounts consumed of the three SSBs by the children are not affected. As for the latter, the consumption level of tap water, a weakly separable preference for the SSBs and the tap water is assumed. In particular, it is assumed that the consumption level of tap water affects subjects' decision on total SSB expenditure for their children; however, it doesn't affect subjects' budg*et al.*location decision or the relative amounts consumed (substitutability) of the three SSBs by the children. Such an assumption of a weak separable preference between the SSBs and the tap water is a fairly weak assumption.

As is common in the estimation of consumer demand, this assumption is invoked for estimation, but is not tested (see for example, Haag *et al.* 2009, Dong *et al.* 2007, Yen *et al.* 2003, and Carpentier and Guyomard 2001).

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 (0.7) 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 (0.7), where ξ_{jt} is independent of $\tilde{x}_{j}^{s}$ and $\xi_{jt} \sim N\left(0,\ \sigma_{\xi}^{2}\right)$.

In the spirit of 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, ω_{njt}, which is equal to $u_{njt} - \theta_{n}\left(\xi_{jt} - v_{j}\right)$, is continuously distributed with a density that is continuous and positive everywhere and is identically distributed conditional on $\left(\rho_{nj}, v_{j}, \tilde{x}_{j}, \tilde{\xi}_{j}\right)$ across t.

Under Assumptions 3.3.1 and 3.3.2, we have (0.11). Estimation for (0.11) proceeds exactly as in the test procedure in the previous section, except in cases that θ_{n} is significantly different from zero, the asymptotic variance of the coefficient estimates in (0.11) $\hat{\delta}_n$ needs to be adjusted as in the following procedure, given the preliminary estimation of the coefficients, denoted by τ, in (0.7).

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

Step4: to adjust for and estimate the asymptotic variance of $\hat{\delta}_n$ using the results in the Appendix 2.

3.3 Generalized Method of Moments Estimation framework

Once the consistent equation-by-equation estimates are obtained for each SSB, following Meyerhoefer *et al.* (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 (A1.2), 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 by:

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

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 $W = \Xi^{-1}$, where Ξ is the asymptotic covariance matrix of $\hat{\delta}$. Let $S_j = \left(S_{1j}', S_{2j}', S_{3j}'\right)'$ denote the set of the subject j's moment conditions in (A1.2) for all the SSBs and H_{nj} denote the univariate Hessian for SSB n. Then, define $\mathrm{H}^{-1} = diag\left\{E\left(H_{1j}\right)^{-1}, \ldots, E\left(H_{Nj}\right)^{-1}\right\}$ and $S = E\left(S_j S_j'\right)$. $\Xi = H^{-1} S H^{-1}$ and can be consistently estimated by substituting in sample analogues (Wooldridge 2010b).

3.4 Elasticity Formulae

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

$$E_n = \frac{\partial w_n}{\partial \log c} \frac{1}{w_n} + 1 \tag{0.12}$$

and

$$e_{ni} = \frac{\partial w_n}{\partial \log p_i} \frac{1}{w_n} - \delta_{ni}^{*}, \tag{0.13}$$

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 Honoré (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:

$$E_n = \beta_n F_n \frac{1}{w_n} + 1 \tag{0.14}$$

and

$$e_{ni} = \gamma_{ni} F_n \frac{1}{w_n} - \delta^*_{ni}, \tag{0.15}$$

where F_n denotes the fraction of observations that are not censored for SSB n. We have also estimated and examined the average marginal effect of the *j*th attribute of the drink n (=Fizzy, Juice and Cordial) on the demand of the drink, which can be generically expressed as:

$$P_{n,j} = \frac{\partial \log x_n}{\partial d_{nj}} = \frac{\partial w_n}{\partial d_{nj}} \frac{1}{w_n} \tag{0.16}$$

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

$$P_{n,j} = \lambda_n F_n \frac{1}{w_n}, \tag{0.17}$$

where w_n is the share of drink n.

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 significantly rejected at the 5% significance level for Fizzy Drink but not for Juice and Cordial, the correction procedure is only implemented for the Fizzy Drink equation. Table A1 in Appendix 3 presents estimates of the selection equation.

4. Results

The average uncompensated and compensated own-price elasticities for all the three drinks are all negative and statistically significant (Table 3), and range from -0.83 to -0.94 (uncompensated), and from -0.11 to -0.79 (compensated). In addition, we find the average compensated cross-price elasticities are all positive, although some are not statistically significant, indicating that on average all the three drinks are net substitutes to one another.

Table 4 presents the average estimated marginal effects with respect to non-price attributes for the full sample. These estimates can be interpreted as the percentage change in drink consumption corresponding to a change in the level of each attribute. Taking Juice as an example, switching from regular Juice to Juice with no added colours or preservatives would increase consumption (the mean consumption level is 1.594) on average by 2.7%. However, except for the effect of the attribute "no added colours or preservatives" on the consumption of Juice, none of the attributes are statistically significant in influencing the consumption of the three drinks.

In order to illustrate the effect of accounting for selection bias, we also estimate the proposed semi-parametric fixed-effects censored demand system model only using observations with positive total SSB 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 that might arise from exclusion of observations with zero total SSB expenditure. The resulting elasticities are shown in Table A1 and A2 in Appendix 3. Although there is no clear pattern of over- or under-estimation for uncompensated price elasticities, if selection bias is unaccounted for, six out of the eight significant compensated price elasticities become lower in absolute value. This provides evidence that in the presence of a large proportion of zeros, elasticities may be downwards biased.

As for the marginal effect estimates for the non-price attributes, most of the estimates without accounting for selection bias are still insignificant. However, besides the marginal effect of “no added colours or preservatives” on Juice consumption, the attribute “Diet” is also statistically significant in influencing the consumption of Fizzy Drink. Recall that while estimating for the full sample, a selection bias is only significant and corrected for Fizzy Drink. The contrasting results clearly show that not correcting for the selection bias would have potentially resulted in misleading results.

To better understand the distributional implications of our results, we examine how the consumption behaviour of children from high-income households would be different from those from low-income households. Previous research has shown slightly larger price effects for low-income households (Zhen *et al.* 2014, Sharma *et al.* 2014). Hence, we split our sample into two parts at the point of median household income, test and correct for potential selection bias on the total SSB expenditure, and estimate the proposed semi-parametric fixed-effects demand system model respectively for the two samples.

For the high-income respondents, the null hypothesis of no selection bias is not rejected at the 5% significance level for any of the three drinks. In contrast, the null is rejected at the 5% level for both Fizzy Drink and Juice for the low-income respondents. The total expenditure elasticities for the three drinks are broadly similar across the samples (Table 3). In particular, compared to high-income respondents, the consumption of Fizzy Drink and Cordial for those on low incomes is slightly less elastic while the consumption of Juice is slightly more elastic. In the case of this study, the total household expenditure on SSBs might be better interpreted as an indicator of respondents’ (or parents’) overall openness or attitude towards SSBs. Hence, conservative parents tend to give less or even no SSBs to their children.

In terms of price elasticities, after accounting for the effects on total expenditure, the consumption of all the three drinks for those on high incomes are less price sensitive than those on low incomes (Fizzy Drink: -0.798 vs -0.845; Juice: -0.100 vs -0.134; Cordial: -0.740 vs -0.803). Cross-price compensated price elasticities are mostly higher in absolute value for those on low incomes. For instance, for those on low incomes the consumption of Cordial is more elastic to a change in the price of Fizzy Drinks than that for those on high incomes (0.290 vs 0.172), suggesting a stronger net substitutability between Fizzy Drinks and Cordial for those on low incomes.

When it comes to marginal effects of attributes on consumption shown side by side in Table 4, we observe some large differences between the high and low-income subjects. None of the attributes of interest are statistically significant in influencing the consumption of the three drinks for those on high incomes. The attributes matter more for those on low incomes, though the pattern of results are mixed. For those on low incomes the strongest results are for Fizzy, where Fizzy Diet drink and Fizzy Vitamins would lead to lower consumption than Fizzy (-15.4% and -20.6% respectively), but Fizzy no colours and preservatives leads to higher consumption (28.4%) by children of parents in low income households. The negative effects on consumption for Fizzy Diet and Fizzy vitamins is counter-intuitive, if these are considered to be healthy drinks. However, it is possible that parents think that Fizzy drinks with vitamins and diet have a worse taste or have a stronger preference against ‘additives’ of any kind, even if they are marketed to be healthy, such as artifcial sweetners in Diet and added vitamins. Juice no colours and preservatives also leads to 13.6% higher consumption compared to Juice. This also shows that the effect of no colours and preservatives for Juice we found in the full sample is driven by the preferences of those on low incomes.

5. Discussion

In this paper, we specify and estimate a new semi-parametric fixed-effects censored demand system to analyse demand for sugar-sweetened beverages for small children. Our data originates from a stated preference experiment where prices are varied exogenously. Hence, we circumvent the endogeneity of using unit values as prices in revealed preference survey data, which is an important issue in many previous studies. In order to deal with the difficulty of a substantial proportion of zero observations for the total expenditure on SSBs and for the expenditure on each drink, we develop a new two-step estimation strategy and employ a semi-parametric estimator for two-sided censoring models with fixed effects. 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 framework which has been applied extensively to aggregate consumption data, but has not yet been applied to micro-level consumption data. This would be a worthy area for future research. By contrasting estimates with and without taking into account the endogenous selection of respondents who report zero expenditure, we identify significant differences, suggesting selection bias in the unadjusted models.

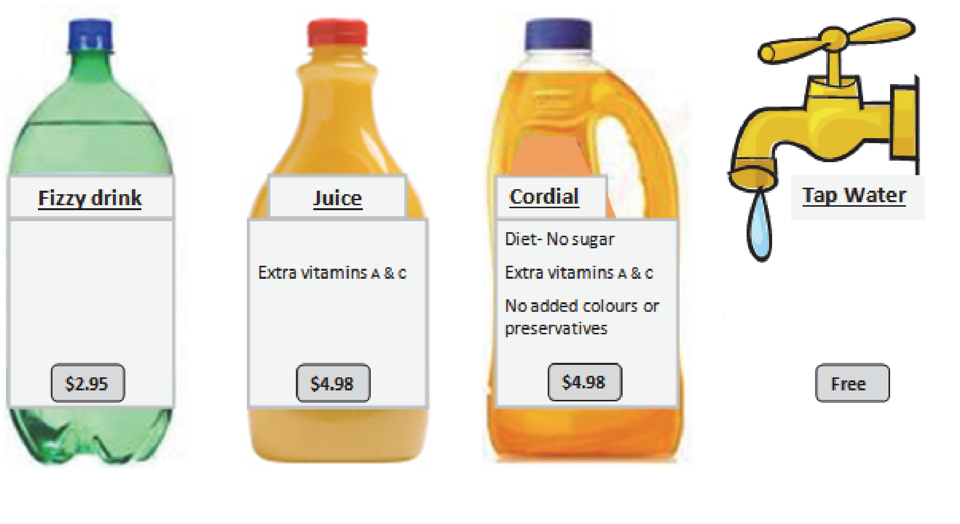
Overall our results suggests that parents of young children will reduce their children's consumption of SSBs if prices were to increase, say through a rise in tax. The size of the elasticities is similar in magnitude to other studies that focus on the purchase of SSBs for the whole family. The literature was summarised in a review by Andreyeva *et al.* (2010) which found an uncompensated elasticity range of -0.8 to -1.0 for soft drinks. Over eighty percent of these earlier studies used time-series data or cross-sectional household survey data and many do not account for price endogeneity. A more recent review by Powell *et al.* (2013) further extended Andreyeva *et al.* (2010)'s review and found a range of -0.71 to -2.26 for regular carbonated soda (equivalent to our Fizzy Drink), a range of -0.69 to -1.91 for fruit drinks and a range of -0.87 to -1.26 for aggregated SSB category. More recent studies have attempted to account for the endogeneity of prices using area-level price indices (Sharma *et al.*., 2014) or adjacent-area price indices (Zhen *et al.*., 2014) as instrumental variables within a demand-system framework. Zhen *et al.* (2014) finds elasticities of -1.0 and -0.96 for regular and diet soda (equivalent to our Fizzy Drink), and -1.6 for Juice. Sharma *et al.* (2014) finds a slightly lower elasticity of -0.6 for soda (Fizzy Drink) and slightly larger elasticities for Cordial (-0.98) and Juice (-1.2). In general, our results are more similar to those from Sharma *et al.* (2014), which could be explained by the similar market context in Australia.

The similarity of our results with the literature suggest that parents have similar price elasticities for their children as for other members of the household. This is encouraging given that the stated preference methodology is based on hypothetical consumption rather than on observed consumption, but has the advantage of exogenous prices and an experimental design. The external validity of our results could be questioned given the sample is a birth cohort of young children in a specific geographic area, and so further research is required to confirm our findings in other settings.

Increasing the price of one type of SSB will lead parents to substitute other types of SSB, and so any price increase through taxation would need to be across all SSBs to encourage substitution to other non-SSBs drinks. Overall, those in low income households are more responsive to price changes, more responsive to cross-price changes, and more responsive to changes in non-price attributes. Those in high income households are less sensitive to price and do not seem to be influenced at all by non-price attributes such as 'diet' 'added vitamins', and 'no added colours or preservatives'. This suggests that labelling of SSBs seem to have little impact overall, but a stronger impact in low income households, especially for Fizzy drinks. Our study answers the call for more evidence on the effect of prices on consumption of SSBs by particularly at-risk groups (Andreyeva *et al.* 2010).

Figure 1 One example of the shopping scenarios
in the stated preference experiment

<u>**Scenario 1.1**</u>
*The descriptions and price of each drink changes for each question so please complete **EACH** table*



a) Given this scenario, <u>considering the description on the labels of each of the bottles above</u> how many 2 litre bottles would you buy of each drink **for your family** in a usual week? *(Please tick one box for each drink)*

Fizzy drink	Juice	Cordial	Tap Water
☐ None	☐ None	☐ None	
☐ 1 bottle	☐ 1 bottle	☐ 1 bottle	
☐ 2 bottles	☐ 2 bottles	☐ 2 bottles	
☐ 3 or more bottles	☐ 3 or more bottles	☐ 3 or more bottles	

b) Now that you have these drinks at home, how many 250ml glasses of each drink would you give to **the child in this study** in a usual week? *(Please tick one box for each drink)*

Fizzy drink	Juice	Cordial	Tap Water
☐ I wouldn't give them any	☐ I wouldn't give them any	☐ I wouldn't give them any	☐ I wouldn't give them any
☐ One glass a week	☐ One glass a week	☐ One glass a week	☐ One glass every two days
☐ Two glasses a week	☐ Two glasses a week	☐ Two glasses a week	☐ One glass a day
☐ One glass every 2 days	☐ One glass every 2 days	☐ One glass every 2 days	☐ Two glasses a day
☐ One glass a day	☐ One glass a day	☐ One glass a day	☐ Three glasses a day
☐ More than one glass a day	☐ More than one glass a day	☐ More than one glass a day	☐ Four or more glasses a day

Table 1 Number and percentage of positive and zero observations for total SSB expenditure and shares of SSBs respectively for the whole sample and the high and low-income subsamples[1].

	Total observations.			Positive observations.			Zero observations.		
	Full sample	High-income	Low-income	Full sample	High-income	Low-income	Full sample	High-income	Low-income
Total	2381	1380	1001	1188	643	545	1193	737	456
	100%	100%	100%	49.9%	46.6%	54.5%	50.1%	53.4%	45.6%
Fizzy	1188	643	545	203	89	114	985	554	431
	100%	100%	100%	17.1%	13.8%	20.9%	82.9%	86.2%	79.1%
Juice	1188	643	545	991	555	436	197	88	109
	100%	100%	100%	83.4%	86.3%	80.0%	16.6%	13.7%	20.0%
Cordial	1188	643	545	456	207	249	732	436	296
	100%	100%	100%	38.4%	32.2%	45.7%	61.6%	67.8%	54.3%

Notes: 1. High-income is defined as the subsample with household weekly income higher than median; Low-income is defined as the subsample with household weekly income lower than median.

Table 2 SSB expenditure and consumption levels for the three drinks[1, 2, 3]

	Total observations.			Positive observations.		
	Full sample	High-income	Low-income	Full sample	High-income	Low-income
	Mean (sd)	Mean (sd)	Mean (sd)	Mean (sd)	Mean (sd)	Mean (sd)
Total expenditure	0.700	0.602	0.835	1.403	1.292	1.534
	(1.368)	(1.247)	(1.510)	(1.663)	(1.565)	(1.765)
Fizzy expenditure	0.084	0.067	0.107	0.986	1.046	0.940
	(0.445)	(0.448)	(0.440)	(1.199)	(1.452)	(0.960)
Juice expenditure	0.564	0.493	0.661	1.355	1.227	1.518
	(1.162)	(1.040)	(1.306)	(1.474)	(1.338)	(1.618)
Cordial expenditure	0.052	0.041	0.067	0.271	0.273	0.269
	(0.184)	(0.168)	(0.202)	(0.343)	(0.355)	(0.333)
Fizzy consumption	0.230	0.179	0.300	2.692	2.770	2.632
	(1.032)	(0.968)	(1.111)	(2.427)	(2.725)	(2.176)
Juice consumption	1.594	1.463	1.775	3.830	3.638	4.076
	(2.872)	(2.772)	(2.997)	(3.355)	(3.346)	(3.355)
Cordial consumption	0.696	0.500	0.965	3.632	3.333	3.880
	(2.095)	(1.756)	(2.464)	(3.504)	(3.339)	(3.623)

Note: 1. High-income is defined as the subsample with household weekly income higher than median; Low-income is defined as the subsample with household weekly income lower than median. 2. The unit of measurement of expenditure is one Australia dollar. 3. The unit of measurement of consumption is one 250ml glass of drink.

Table 3 Average Total Expenditure Elasticities and Uncompensated and Compensated Price Elasticities

		Full sample			High-income			Low-income		
		Fizzy	Juice	Cordial	Fizzy	Juice	Cordial	Fizzy	Juice	Cordial
Total expenditure		1.603 ***	1.124 ***	0.374 ***	1.496 ***	1.088 ***	0.400 ***	1.255 ***	1.134 ***	0.318 ***
		(0.115)	(0.017)	(0.064)	(0.098)	(0.018)	(0.064)	(0.219)	(0.026)	(0.090)
Uncompensated	Fizzy	-0.943 ***	-0.046 *	-0.011	-0.916 ***	-0.051 **	-0.033 **	-0.983 ***	-0.067 **	0.050
		(0.029)	(0.026)	(0.018)	(0.025)	(0.024)	(0.016)	(0.033)	(0.030)	(0.032)
	Juice	-0.013 *	-0.949 ***	-0.038 ***	-0.015 **	-0.953 ***	-0.032 ***	-0.019 **	-0.926 ***	-0.054 ***
		(0.007)	(0.008)	(0.009)	(0.007)	(0.009)	(0.007)	(0.009)	(0.016)	(0.012)
	Cordial	-0.012	-0.156 ***	-0.832 ***	-0.047 **	-0.158 ***	-0.795 ***	0.048	-0.184 ***	-0.864 ***
		(0.021)	(0.037)	(0.043)	(0.022)	(0.037)	(0.045)	(0.031)	(0.040)	(0.034)
Compensated	Fizzy	-0.794 ***	1.148 ***	0.249 ***	-0.798 ***	1.122 ***	0.172 ***	-0.845 ***	0.810 ***	0.290 ***
		(0.027)	(0.099)	(0.026)	(0.021)	(0.095)	(0.021)	(0.044)	(0.153)	(0.051)
	Juice	0.092 ***	-0.112 ***	0.144 ***	0.071 ***	-0.100 ***	0.117 ***	0.105 ***	-0.134 ***	0.163 ***
		(0.007)	(0.015)	(0.010)	(0.007)	(0.016)	(0.008)	(0.008)	(0.027)	(0.012)
	Cordial	0.022	0.123 **	-0.771 ***	-0.015	0.155 **	-0.740 ***	0.083 ***	0.039	-0.803 ***
		(0.021)	(0.062)	(0.044)	(0.022)	(0.068)	(0.045)	(0.030)	(0.082)	(0.039)

Note: Standard errors are in parenthesis. * Significant at 10%; ** Significant at 5%; *** Significant at 1%. High-income: subsample with household weekly income higher than median. Low-income: subsample with household weekly income lower than median.

Table 4 Average Partial Elasticities w.r.t. Attributes

Drink	Attributes	Full sample	High-income	Low-income
Fizzy	Fizzy Diet	-0.050 (0.062)	-0.025 (0.109)	-0.154 * (0.084)
	Fizzy Vitamins	-0.030 (0.042)	0.028 (0.046)	-0.206 *** (0.050)
	Fizzy Nocolours	0.061 (0.053)	-0.069 (0.056)	0.284 *** (0.109)
Juice	Juice Diet	-0.011 (0.012)	-0.017 (0.013)	0.006 (0.023)
	Juice Vitamins	0.003 (0.015)	0.005 (0.018)	0.058 (0.037)
	Juice Nocolours	0.027 * (0.015)	0.001 (0.013)	0.136 *** (0.026)
Cordial	Cordial Diet	-0.050 (0.078)	-0.049 (0.098)	0.022 (0.115)
	Cordial Vitamins	-0.047 (0.051)	-0.062 (0.054)	0.009 (0.064)
	Cordial Nocolours	0.031 (0.058)	-0.020 (0.052)	0.056 (0.102)

Note: * Significant at 10%; ** Significant at 5%; *** Significant at 1%. High-income: the subsample with household weekly income higher than median. Low-income: the subsample with household weekly income lower than median. 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.

Appendix 1

Let x_{jt} denote the vector of all the observed explanatory variables in (0.11) excluding $\hat{\xi}_{jt}$, and $\tilde{x}_j \equiv \left(x_{j1}',\ldots,x_{jT_j}'\right)'$ 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 by solving the following minimisation problem:

$$\hat{\delta}_n = \arg\min_{\delta} \sum_{j=1}^{J} \sum_{1<s<t<T_j} \frac{1}{T_j} U\left(w_{njt}, w_{njs}, \begin{pmatrix} x_{jt} - x_{js} \\ \hat{\xi}_{jt} - \hat{\xi}_{js} \end{pmatrix}' \delta \right) \qquad \text{(A1.1)}$$

where

$$U\left(y_1, \mathrm{y}_2, d\right) \begin{cases} 1 + 2c_1 + c_1^2 - 2c_3c_1 + 2c_3c_2 + \left(y_1 - y_2 - c_2\right)^2 & \text{for } d < -1 \\ -2d - d^2 + 2c_1 + c_1^2 - 2c_3c_1 + 2c_3c_2 + \left(y_1 - y_2 - c_2\right)^2 & \text{for } -1 \le d < c_1 \\ -2c_3 d + 2c_3 c_2 + \left(y_1 - y_2 - c_2\right)^2 & \text{for } c_1 \le d < c_2 \\ \left(y_1 - y_2 - d\right)^2 & \text{for } c_2 \le d < c_3 \\ -2c_2 d + 2c_2 c_3 + \left(y_1 - y_2 - c_3\right)^2 & \text{for } c_3 \le d < c_4 \\ -d^2 + 2d + c_4^2 - 2c_4 - 2c_2c_4 + 2c_2c_3 + \left(y_1 - y_2 - c_3\right)^2 & \text{for } c_4 \le d < 1 \\ 1 + c_4^2 - 2c_4 - 2c_2c_4 + 2c_2c_3 + \left(y_1 - y_2 - c_3\right)^2 & \text{for } d \ge 1 \end{cases}$$

and

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

The rationale behind this estimator is that for example, if $E\left(\varepsilon x\right) = 0$, then one has the moment conditions $E\left[\left(y^* - x'\beta\right)x\right] = 0$, 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 (A1.1) has as first-order condition the sample analogue of moment conditions as follows:

$$E\left[\sum_{1<s<t<T_j} \frac{1}{T_j} u\left(w_{njt}, \mathrm{w}_{njs}, \Delta x_j{}' \delta_n\right) \Delta x_j\right] = 0 \tag{A1.2}$$

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

and

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

Under $\mathrm{H}_0:\ \theta_n = 0$,

$$\sqrt{J}\left(\hat{\delta}_n - \delta_n\right) \xrightarrow{d} N\left(0, \Gamma^{-1} S \Gamma^{-1}\right) \tag{A1.3}$$

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

$$\hat{\Gamma} = \frac{1}{J}\sum_{j=1}^{J}\left[\sum_{s<t}\frac{1}{T_j}1\left\{-1<\left(x_{js}-x_{jt}\right)'\hat{\delta}_n<1\right\}\begin{pmatrix}1\left\{-1<\left(x_{js}-x_{jt}\right)'\hat{\delta}_n<w_{js}-1\right\}- \\ 1\left\{0<\left(x_{js}-x_{jt}\right)'\hat{\delta}_n<w_{js}\right\}- \\ 1\left\{-w_{jt}<\left(x_{js}-x_{jt}\right)'\hat{\delta}_n<0\right\}+ \\ 1\left\{1-w_{jt}<\left(x_{js}-x_{jt}\right)'\hat{\delta}_n<1\right\}\end{pmatrix}\left(x_{js}-x_{jt}\right)\left(x_{js}-x_{jt}\right)'\right] \tag{A1.4}$$

and

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

with

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

Appendix 2

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 (A1.1) the expression of $\hat{\xi}_{jt}$ in (3.10) and rewrite the semi-parametric estimator on the selected sample as:

$$\begin{aligned}\hat{\delta}_n &= \arg\min_{\delta_n} \sum_{j=1}^{J} \sum_{1<s<t<T_j} \frac{1}{T_j} U\left(w_{njt}, w_{njs}, \left(\tilde{x}_{jt} - \tilde{x}_{js}\right)' \tilde{\delta}_n + \left(c_{jt} - c_{js}\right)\theta_n - \left(\hat{x}_{jt} - \hat{x}_{js}\right)' \hat{\tau}\theta_n \right) \\ &= \arg\min_{\delta_n} \sum_{j=1}^{J} \sum_{1<s<t<T_j} \frac{1}{T_j} U\left(w_{njt}, w_{njs}, \left(x_{jt}^c - x_{js}^c\right)' \left(\tilde{\delta}_n - \hat{\tau}^c\theta_n\right) + \left(c_{jt} - c_{js}\right)\theta_n - \left(x_{jt}^u - x_{js}^u\right)' \hat{\tau}^u\theta_n \right)\end{aligned} \quad \text{(A2.1)}$$

where $\tilde{x}_{jt}$ denotes the vector of all the explanatory variables in (0.11) 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 $\delta_n \equiv \left(\tilde{\delta}_n', \theta_n\right)'$; $\hat{x}_{jt}$ denotes the vector of observable explanatory variables in (0.7) and $\hat{\tau}$ denotes corresponding coefficient estimates; x_{jt}^c denotes shared observable explanatory variables between (0.7) and (0.11) and x_{jt}^u denotes observable explanatory variables that only appear in the selection equation (0.7), i.e. household income and consumption level of tap water. The consistency of the estimator in the share equation estimation given τ, can be clearly seen following the consistency arguments in Alan *et al.* (2014).

Since the objective function $U(\cdot)$, as shown in (A1.1), 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

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

(A2.2)

and define functions:

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

(A2.3)

and

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

(A2.4)

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):

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

(A2.5)

Throughout this appendix, the symbols $\|\cdot\|$ denotes not only the usual Euclidean norm but also a matrix norm: $\left\|\left(b_{ij}\right)\right\| = \left(\sum_{i,j} b_{ij}^2\right)^{1/2}$.

Theorem A1: Let x_{jt}^c denotes shared observable explanatory variables between (0.7) and (0.11) and x_{jt}^u denote observable explanatory variables that only appear in the selection equation (0.7). Define $x_{jt}^a \equiv \left(x_{jt}^{c}{}', c_{jt}, x_{jt}^{u}{}'\right)'$. If

1. Assumption 3.3.1:

   Denote observable explanatory variables in (0.7) 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 (0.7), where ξ_{jt} is independent of $\tilde{x}_j^s$ and $\xi_{jt} \sim N\left(0,\ \sigma_{\xi}^2\right)$.

2. Assumption 3.3.2:

   $E\left(u_{njt} \mid \rho_{nj}, v_j, \tilde{x}_j, \tilde{\xi}_j\right) = \theta_n \varepsilon_{jt} = \theta_n\left(\xi_{jt} - v_j\right)$, $t = 1, 2, \ldots, T_j$, and, ω_{njt}, which is equal to $u_{njt} - \theta_n\left(\xi_{jt} - v_j\right)$, is continuously distributed with a density that is continuous and positive everywhere and is identically distributed conditional on $\left(\rho_{nj}, v_j, \tilde{x}_j, \tilde{\xi}_j\right)$ across t.

3. Finite moment conditions: for any $t \le T$, $E\left(\left\|x_{jt}^a\right\|^2\right) < \infty$ and $E\left(\left\|x_{jt}^a\right\|^4\right) < \infty$. 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 $E\left(\alpha_n^2 \left\|x_{js}^a - x_{jt}^a\right\|^2\right) < \infty$.

4. While x_{jt} denotes the vector of all the observed explanatory variables in (0.11) 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 $s,t \le T_j$, $s \ne t$.

5. We make the following definitions: $\Gamma_{\delta_n}(\tau) \equiv \left.\frac{\partial G(a;\tau)}{\partial a}\right|_{a=\delta_n}$ $\Gamma_{\tau}(\delta_n) = \left.\frac{dG(\delta_n;b)}{db}\right|_{b=\tau}$. The expectations in $\Gamma_{\delta_n}(\tau)$ and $\Gamma_{\tau}(\delta_n)$ are finite and $\Gamma_{\delta_n}(\tau)$ and $\Gamma_{\tau}(\delta_n)$ are of full rank.

Then

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

where $\hat{\delta}_n$ is defined as in (A1.1) 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 $\Gamma_\tau(\delta_n) = \left.\frac{dG(\delta_n;b)}{db}\right|_{b=\tau}$.

Before we proceed to the proof of Theorem A1, we first introduce and prove three useful lemmas. In particular, Lemma 1 states that the function $G(\delta_n;b)$ is differentiable at τ.

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

PROOF:

Let $x_{jt}^a = \left(x_{jt}^{c\,\prime}, c_{jt}, x_{jt}^{u\,\prime}\right)'$ for any $t \le T_j$ and $\delta_n^a = \left(\left(\tilde{\delta}_n - b^c\theta_n\right)', \theta_n, -b^{u\prime}\theta_n\right)'$, and then b is a vector of b^c and b^u, i.e. $b = \left(b^{c\prime}, b^{u\prime}\right)'$. According to Theorem 1 in Alan *et al.* (2014), given $h_j(\delta_n;b) = \sum_{1<s<t<T_j} \frac{1}{T_j} u\left(w_{njt}, w_{njs}, \Delta x_j^{a\prime}\delta_n^a\right)\Delta x_j^{a\prime}$,

$$\Gamma_\tau(\delta_n) = \left.\frac{dE\left[h_j(\delta_n;b)\right]}{db}\right|_{b=\tau}$$

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

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

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

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

PROOF:

Using the linear operator notation of Pakes and Pollard (1989), with v_n denoting the standardised empirical process $\sqrt{n}(P_n - P)$, it suffices to prove that

$$\begin{aligned}
&\sup_{\|b-\tau\|<\chi_J} \left\| G_J(\delta_n;b) - G(\delta_n;b) - G_J(\delta_n;\tau) \right\| \\
&= \sup_{\|b-\tau\|<\chi_J} \left\| \left[G_J(\delta_n;b) - G(\delta_n;b)\right] - \left[G_J(\delta_n;\tau) - G(\delta_n;\tau)\right] \right\| \\
&= J^{-1/2} \sup_{\|b-\tau\|<\chi_J} \left\| v_J h(x_j,\delta_n;b) - v_J h(x_j,\delta_n;\tau) \right\| \\
&= o_p\left(J^{-1/2}\right).
\end{aligned}$$

It follows from Lemma 2.17 of Pakes and Pollard (1989), that it suffices to demonstrate that (a) $\mathcal{H} \overset{\text{def}}{=} \{h_j(\delta_n;b):\ b \in B\}$ (B is the parameter space of τ) is a Euclidian class with an envelope H for which $E(H^2) < \infty$, and (b) $E\left[h_j(\delta_n;b)^2\right]$ is continuous at $b = \tau$. It is clear that $\mathcal{H}$ is bounded in norm by $\left(\|w^*_{njs}\| + \|w^*_{njt}\|\right)\|x^a_{js} - x^a_{jt}\|$. These are squared integrable by the finite moment conditions. It is also straightforward to see that $E\left[h_j(\delta_n;b)^2\right]$ is continuous at $b = \tau$. As a result, what remains to be shown is that $\mathcal{H}$ is Euclidian. $\mathcal{H}$ is Euclidian by Lemma 2.13 of Pakes and Pollard (1989) (see the expression for u in Appendix 1).

Q.E.D.

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

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

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

where

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

and

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

PROOF:

To establish asymptotic normality of $G_J\left(\delta_n;\hat{\tau}\right)$, we first show that $G_J\left(\delta_n;\hat{\tau}\right)$ is very well approximated by the linear function: $L_J^\tau\left(\delta_n;\hat{\tau}\right)=\Gamma_\tau\left(\delta_n\right)\left(\hat{\tau}-\tau\right)+G_J\left(\delta_n;\tau\right)$. This follows directly from Lemma 1 and Lemma 2 together with the consistency of $\hat{\tau}$ and first-order representation in (A2.5). Specifically, given the consistency of $\hat{\tau}$, we can always choose a positive sequence $\left\{\chi_J\right\}$ that converges to zero as J goes to infinity slowly enough to ensure that

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

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,

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

Then, it follows that

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

Hence,

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

(A2.6)

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 $\hat{\delta}_n$.

PROOF OF THEOREM A1:

First we prove $\sqrt{J}$-consistency. Given the consistency of the estimator $\hat{\delta}_n$ in (A2.1) and $\hat{\tau}$, it allows us to choose a positive sequence $\{k_J\}$ that converges to zero as J goes to infinity slowly enough to ensure that

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

(A2.7)

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 Lemma A2 to show that

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

By the triangle inequality, it is straightforward to see that

$$\begin{aligned} &\left\| G_J(a;\hat{\tau}) - G(a;\hat{\tau}) - G_J(\delta_n;\hat{\tau}) \right\| \le \\ &\left\| G_J(a;\hat{\tau}) - G(a;\hat{\tau}) - G_J(\delta_n;\hat{\tau}) + G(\delta_n;\hat{\tau}) \right\| + \left\| G(\delta_n;\tau) - G(\delta_n;\hat{\tau}) \right\|. \end{aligned}$$

Note that $G(\delta_n;\tau)=0.$ Then, together with the fact that since $G(\delta_n;b)$ is continuously differentiable in *b* (see Lemma A1), $\left\| G(\delta_n;\tau) - G(\delta_n;\hat{\tau}) \right\| = O_p\left(J^{-1/2}\right),$ it follows that

$$\sup_{\|a-\delta_n\|\le k_J} \left\| G_J(a;\hat{\tau}) - G(a;\hat{\tau}) - G_J(\delta_n;\hat{\tau}) \right\| = O_p\left(J^{-1/2}\right). \tag{A2.8}$$

With the probability in (A2.7) tending to one, the supremum in (A2.8) runs over a range that includes the random value $\hat{\delta}_n$. Hence,

$$\left\| G_J\left(\hat{\delta}_n;\hat{\tau}\right) - G\left(\hat{\delta}_n;\hat{\tau}\right) - G_J(\delta_n;\hat{\tau}) \right\| \le O_p\left(J^{-1/2}\right) \tag{A2.9}$$

By the triangle inequality, the left-hand side of (A2.9) is larger than

$$\left\| G\left(\hat{\delta}_n;\hat{\tau}\right) \right\| - \left\| G_J\left(\hat{\delta}_n;\hat{\tau}\right) \right\| - \left\| G_J(\delta_n;\hat{\tau}) \right\| \tag{A2.10}$$

Thus,

$$\left\| G\left(\hat{\delta}_n;\hat{\tau}\right) \right\| \le O_p\left(J^{-1/2}\right) + \left\| G_J\left(\hat{\delta}_n;\hat{\tau}\right) \right\| + \left\| G_J(\delta_n;\hat{\tau}) \right\| \tag{A2.11}$$

As $U(\cdot)$ is everywhere differentiable, the sample counterparts of the moment conditions imply that $G_J\left(\hat{\delta}_n;\hat{\tau}\right)=0$. Also, as a direct consequence of the Central Limit Theorem, $\sqrt{J}G_J(\delta_n;\tau)$ converges in distribution to $N(0,V)$, where $V = E\left[h_j(\delta_n;\tau) h_j(\delta_n;\tau)' \right]$.

Hence, according to Lemma A3, it follows that $\left\| G_J\left(\delta_n;\hat{\tau}\right)\right\| = O_p\left(J^{-1/2}\right)$. Then, it follows from (A2.11) that

$$\left\| G\left(\hat{\delta}_n;\hat{\tau}\right)\right\| \le O_p\left(J^{-1/2}\right) + \left\| G_J\left(\delta_n;\hat{\tau}\right)\right\| = O_p\left(J^{-1/2}\right) \tag{A2.12}$$

That is,

$$\left\| G\left(\hat{\delta}_n;\hat{\tau}\right)\right\| \le O_p\left(J^{-1/2}\right). \tag{A2.13}$$

The differentiability of $G\left(a;\hat{\tau}\right)$ 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,

$$\left\| G\left(a;\hat{\tau}\right)\right\| \ge c\left\| a-\delta_n\right\| \text{ for } a \text{ near } \delta_n. \tag{A2.14}$$

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

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

$$L_J\left(a;\hat{\tau}\right) = \Gamma_{\delta_n}\left(\hat{\tau}\right)\left(a-\delta_n\right) + G_J\left(\delta_n;\hat{\tau}\right), \tag{A2.15}$$

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 $\left\| L_J\left(a;\hat{\tau}\right)\right\|$ globally. For $\hat{\delta}_n$, this follows directly from the differentiability of $G\left(a;\hat{\tau}\right)$ at δ_n with a derivative matrix of full rank and (A2.9), together with the $\sqrt{n}$-consistency results already established. In particular,

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

(A2.16)

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

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

(A2.17)

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

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

(A2.18)

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 $N(0,V)$. 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

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

(A2.19)

Since δ_n is an interior point of the parameter space of δ_n, Δ, (A2.19) 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 (A2.8) and (A2.9), we can get

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

(A2.20)

Then, similarly as in (A2.10) through (A2.16), we have

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

(A2.21)

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

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

(A2.22)

That is,

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

(A2.23)

Squaring both sides gives that

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

(A2.24)

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

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

(A2.25)

Likewise, as in (A2.8) through (A2.11), it follows

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

(A2.26)

which gives $\left\|G_J\left(\hat{\delta}_n^*;\hat{\tau}\right)\right\| \leq O_p\left(J^{-1/2}\right)$. Hence, given (A2.21), $\left\|L_J\left(\hat{\delta}_n^*;\hat{\tau}\right)\right\|$ is of order $O_p\left(J^{-1/2}\right)$. The quadratic form $\left\|L_J\left(a;\hat{\tau}\right)\right\|^2$ has a simple expansion

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

(A2.27)

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

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

(A2.28)

which rearranges to

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

(A2.29)

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

Substituting $\hat{\delta}_n$ in (A2.27) gives $\left\|L_J\left(\hat{\delta}_n;\hat{\tau}\right)\right\|^2 = \left\|L_J\left(\hat{\delta}_n^*;\hat{\tau}\right)\right\|^2 + \left\|\Gamma_{\delta_n}\left(\hat{\tau}\right)\left(\hat{\delta}_n - \hat{\delta}_n^*\right)\right\|^2$. Equating this to (A2.24) gives $\left\|\Gamma_{\delta_n}\left(\hat{\tau}\right)\left(\hat{\delta}_n - \hat{\delta}_n^*\right)\right\| = o_p\left(J^{-1/2}\right)$. Since $\Gamma_{\delta_n}\left(\hat{\tau}\right)$ is of full rank (at least in the asymptotics), 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

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

Q.E.D.

We already know how to consistently estimate $\Gamma_{\delta_n}(\tau)$: use expression (A1.4). $h_j(\delta_n;\tau)$ and $r_j(\tau)$ can be respectively by $h_j\left(\hat{\delta}_n;\hat{\tau}\right)$ and $r_j(\hat{\tau})$. $\Gamma_\tau(\delta_n)$ can be estimated as follows:

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

where $\hat{x}_{jt}$ denotes the vector of observed explanatory variables in (0.7), $\mathbf{Z}$ 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 $\hat{x}_{jt}$.

Appendix 3

Table A1 Average Total Expenditure Elasticities and Uncompensated and Compensated Price Elasticities for without correcting for selection

		Fizzy	Juice	Cordial
Total expenditure		1.508 *** (0.138)	1.135 *** (0.016)	0.284 *** (0.077)
Uncompensated	Fizzy	-0.808 *** (0.045)	-0.136 *** (0.029)	-0.057 (0.054)
	Juice	-0.038 *** (0.008)	-0.940 *** (0.008)	-0.022 *** (0.008)
	Cordial	-0.064 (0.062)	-0.090 *** (0.031)	-0.846 *** (0.074)
Compensated	Fizzy	-0.668 *** (0.039)	0.988 *** (0.105)	0.188 *** (0.069)
	Juice	0.068 *** (0.008)	-0.095 *** (0.014)	0.162 *** (0.008)
	Cordial	-0.038 (0.062)	0.122 * (0.070)	-0.800 *** (0.074)

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

Table A2 Average Partial Elasticities w.r.t. Attributes for without correcting for selection

Drink	Attributes	Coef.		S.E.
Fizzy				
	Fizzy Diet	0.196	***	(0.065)
	Fizzy Vitamins	0.065		(0.063)
	Fizzy Nocolours	0.081		(0.064)
Juice				
	Juice Diet	-0.015		(0.011)
	Juice Vitamins	0.009		(0.015)
	Juice Nocolours	0.042	***	(0.014)
Cordial				
	Cordial Diet	0.038		(0.329)
	Cordial Vitamins	-0.077		(0.248)
	Cordial Nocolours	-0.152		(0.299)

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

Alan, 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.
Andreyeva, T., M. W. Long, and K. 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): 216–222. 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): 204–221.
Beck, M. J., S. Fifer, and J. M. Rose. 2016. "Can You Ever Be Certain? Reducing Hypothetical Bias in Stated Choice Experiments via Respondent Reported Choice Certainty." *Transportation Research Part B: Methodological* 89: 149–167.
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): 570–597.
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): 1805–1808.
Buse, A. 1998. "Testing Homogeneity in the Linearized Almost Ideal Demand System." *American Journal of Agricultural Economics* 80 (1): 208–220.
Capps Jr, O., J. Church, and A. Love. 2003. "Specification Issues and Confidence Intervals in Unilateral Price Effects Analysis." *Journal of Econometrics* 113 (1): 3–31.
Carlsson, F., and P. Martinsson. 2003. "Design Techniques for Stated Preference Methods in Health Economics." *Health Economics* 12 (4): 281–294.
Carpentier, A, and H Guyomard. 2001. "Unconditional Elasticities in Two-Stage Demand Systems: An Approximate Solution." *American Journal of Agricultural Economics* 83 (1): 222–229.
Chalfant, JA. 1987. "A Globally Flexible, Almost Ideal Demand System." *Journal of Business & Economic Statistics* 5 (2): 233–242.
ChoiceMetrics. 2012. "Ngene 1.1.1 User Manual & Reference Guide." Sydney: Australia.
Christensen, L. R., D. W. Jorgenson, and L. J. Lau. 1975. "Transcendental Logarithmic Utility Functions." *The American Economic Review* 65 (3): 367–383.
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): 730–741.
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): 312–326.
———. 1980b. *Economics and Consumer Behavior*. Cambridge; New York: Cambridge University Press.
Dong, D., C. G. Davis, and H. Stewart. 2015. “The Quantity and Variety of Households’ Meat Purchases: A Censored Demand System Approach.” *Agricultural Economics* 46: 99–112. doi:10.1111/agec.12143.
Dong, D., H. M. Kaiser, and O. Myrland. 2007. “Quantity and Quality Effects of Advertising: A Demand System Approach.” *Agricultural Economics* 36 (3): 313–324. doi:10.1111/j.1574-0862.2007.00209.x.
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): 924–934.
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): 219–239.
Fletcher, J. M., D. E. Frisvold, and N. Tefft. 2010a. “The Effects of Soft Drink Taxes on Child and Adolescent Consumption and Weight Outcomes.” *Journal of Public Economics* 94: 967–974. doi:10.1016/j.jpubeco.2010.09.005.
Fletcher, J. M., D. Frisvold, and N. Tefft. 2010b. “Can Soft Drink Taxes Reduce Population Weight?” *Contemporary Economic Policy* 28 (1): 23–35. doi:10.1111/j.1465-7287.2009.00182.x.
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): 377–385.
Gallant, A. R. 1981. “On the Bias in Flexible Functional Forms and an Essentially Unbiased Form.” *Journal of Econometrics* 15: 211–245.
Haag, B R, S Hoderlein, and K Pendakur. 2009. “Testing and Imposing Slutsky Symmetry in Nonparametric Demand Systems.” *Journal of Econometrics* 153 (1): 33–50. doi:10.1016/j.jeconom.2009.04.003.
Heien, D., and C. R. Wessells. 1990. “Demand Systems Estimation with Microdata: A Censored Regression Approach.” *Journal of Business & Economic Statistics* 8 (3): 365–371.
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): 316–325.
Honoré, B. E. 1992. “Trimmed LAD and Least Squares Estimation of Truncated and Censored Regression Models with Fixed Effects.” *Econometrica* 60 (3): 533–565.
———. 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): 241–278.

Huber, Joel, and Klaus Zwerina. 1996. “The Importance of Utility Balance in Efficient Choice Designs.” *Journal of Marketing Research* 33 (3): 307–317. doi:10.2307/3152127.
Kim, Jaehwan, Greg M. Allenby, and Peter E. Rossi. 2002. “Modeling Consumer Demand for Variety.” *Marketing Science* 21 (3): 229–250. doi:10.1287/mksc.21.3.229.143.
Kyureghian, G., R. M. Nayga, and S. Bhattacharya. 2013. “The Effect of Food Store Access and Income on Household Purchases of Fruits and Vegetables: A Mixed Effects Analysis.” *Applied Economic Perspectives and Policy* 35 (1): 69–88. doi:10.1093/aepp/pps043.
Lancsar, E., and J. Swait. 2014. “Reconceptualising the External Validity of Discrete Choice Experiments.” *PharmacoEconomics* 32 (10): 951–965.
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): 311–316.
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): 274–288.
Mann, H. B., and a. Wald. 1943. “On Stochastic Limit and Order Relationships.” *The Annals of Mathematical Statistics* 14 (3): 217–226. 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): 909–930.
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): 660–672.
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.
Nayga, R. M. 1995. “Determinants of U . S . Household Expenditures on Fruit and Vegetables : A Note and Update.” *Journal of Agricultural and Applied Economics* 27 (2): 588–594.
Pakes, Ariel, and David Pollard. 1989. “Simulation and the Asymptotics of Optimization Estimators.” *Econometrica* 57 (5): 1027–1057. 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) (November): 1022–1037.
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): 1533–1551.
Powell, J. L. 1986. “Symmetrically Trimmed Least Squares Estimation for Tobit Models.” *Econometrica* 54 (6): 1435–1460. doi:10.2307/1914308.
Powell, L. M., J. F. Chriqui, and F. J. Chaloupka. 2009. “Associations between State-Level Soda Taxes and Adolescent Body Mass Index.” *Journal of Adolescent Health* 45 (3 SUPPL.): S57–S63. doi:10.1016/j.jadohealth.2009.03.003.
Powell, L. M., J. F. Chriqui, T. Khan, R. Wada, and F. J. Chaloupka. 2013. “Assessing the Potential Effectiveness of Food and Beverage Taxes and Subsidies for Improving Public Health: A Systematic Review of Prices, Demand and Body Weight Outcomes.” *Obesity Reviews* 14 (2): 110–128. doi:10.1111/obr.12002.

Scott, A., J. Witt, J. Humphreys, C. Joyce, G. Kalb, S. Jeon, and M. 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): 1159–1184.

Sivey, P., A. Scott, J. Witt, C. Joyce, and J. Humphreys. 2012. “Junior Doctors’ Preferences for Specialty Choice.” *Journal of Health Economics* 31 (6): 813–823.

Sturm, R., L. M. Powell, J. F. Chriqui, and F. J. Chaloupka. 2010. “Soda Taxes, Soft Drink Consumption, and Children’s Body Mass Index.” *Health Affairs* 29 (5): 1052–1058. doi:10.1377/hlthaff.2009.0061.

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

Vella, F. 1992. “Simple Tests for Sample Selection Bias in Censored and Discrete Choice Models.” *Journal of Applied Econometrics* 7 (4): 413–421. doi:10.1002/jae.3950070407.

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): 1044–1051.

Wales, T.J., and A.D. Woodland. 1983. “Estimation of Consumer Demand Systems with Binding Non-Negativity Constraints.” *Journal of Econometrics* 21: 263–285.

Wooldridge, J. M. 1995. “Selection Corrections for Panel Data Models under Conditional Mean Independence Assumptions.” *Journal of Econometrics* 68: 115–132.

———. 2010a. “Chapter 19 Censored Data, Sample Selection, and Atrition.” In *Econometric Analysis of Cross Section and Panel Data*, Second Ed, 775–851. MIT: Cambridge, MA.

———. 2010b. “Chapter 14 Generalized Method of Moments and Minimum Distance Estimation.” In *Econometric Analysis of Cross Section and Panel Data*, Second Ed, 525–558. MIT: Cambridge, MA.

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

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): 458–478.

Yen, S.T. 2005. “A Multivariate Sample-Selection Model: Estimating Cigarette and Alcohol Demands with Zero Observations.” *American Journal of Agricultural Economics* 87 (May): 453–466.

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): 742–749.

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): 175–193.

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. doi:10.1093/ajae/aat049.