Essays in Industrial Organization and Applied Econometrics

Linqi Zhang

A dissertation

submitted to the Faculty of

the department of Economics

in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

Boston College
Morrissey College of Arts and Sciences
Graduate School

March 2024

Essays in Industrial Organization and Applied Econometrics

Linqi Zhang

Dissertation Committee:

Arthur Lewbel (Chair)

Charles Murry

Richard L. Sweeney

This dissertation comprises three essays on empirical industrial organization (IO) and applied econometrics. The first and third chapters focus on identification approaches in structural models, with the first chapter dedicated to addressing limitations in demand modeling, while the third chapter studies identification in a triangular two-equation system. The second chapter applies modern econometric tools to understand policy-related topics in IO.

The first chapter deals with identification in structural demand modeling, and generalizes the current framework in the literature to achieve a more accurate estimation of differentiated products demand. Within the framework of Berry (1994) and Berry, Levinsohn, and Pakes (1995), the existing empirical industrial organization literature often assumes that market size is observed. However, the presence of an unobservable outside option is a common source of mismeasurement. Measurement errors in market size lead to inconsistent estimates of elasticities, diversion ratios, and counterfactual simulations. I explicitly model the market size, and prove point identification of the market size model along with all demand parameters in a random coefficients logit (BLP) model. No additional data beyond what is needed to estimate standard BLP models is required. Identification comes from the exogenous variation in product characteristics across markets and the nonlinearity of the demand system. I apply the method to a merger simulation in the carbonated soft drinks (CSD) market in the US, and find that assuming a market size larger than the true estimated size would underestimate merger price increases.

Understanding consumer demand is not only central to studying market structure and

competition but also relevant to the study of public policy such as taxation. In the second chapter, we examine household demand for sugar-sweetened beverages (SSB) in the U.S. Our goal is to understand the distributional effect of soda taxes across demographic groups and market segments (at-home versus away-from-home). Using a novel dataset that includes at-home and away-from-home food purchases, we study who is affected by soda taxes. We nonparametrically estimate a random coefficient nested logit model to exploit the rich heterogeneity in preferences and price elasticities across households, including SNAP participants and non-SNAP-participant poor. By simulating its impacts, we find that soda taxes are less effective away-from-home while more effective at-home, especially by targeting the total sugar intake of the poor, those with high total dietary sugar, and households without children. Our results suggest that ignoring either segment can lead to biased policy implications.

In the final chapter, we show that a standard linear triangular two equation system can be point identified, without the use of instruments or any other side information. We find that the only case where the model is not point identified is when a latent variable that causes endogeneity is normally distributed. In this non-identified case, we derive the sharp identified set. We apply our results to Acemoglu and Johnson's (2007) model of life expectancy and GDP, obtaining point identification and comparable estimates to theirs, without using their (or any other) instrument.

Contents

Acknowledgments

I am profoundly grateful to my advisor, Arthur Lewbel, for his invaluable support throughout my doctoral journey. His guidance has shaped my academic trajectory and fostered my growth as a researcher. Without his expertise and encouragement, this dissertation would not have been possible.

I extend my thanks to Charles Murry and Richard Sweeney for their insightful feedback on my research and for generously investing their time in preparing me for the job market.

Additionally, I would like to thank Matthew Masten for his mentorship during my early professional development. His presence epitomizes the finest aspects of the academic world.

Finally, I thank my partner, Zhanhan, who has shown me the possibility of setting up a new life and kept me sane in a world *"(filled) with doubts, an eternal dissatisfaction with yourself, vain attempts to improve, and failures, and an eternal expectation of the happiness that has eluded you and is not possible for you"*[1].

[1] *Anna Karenina*, by Leo Tolstoy, translated by Richard Pevear and Larissa Volokhonsky.

List of Tables

List of Figures

Chapter 1

Identification and Estimation of Market Size in Discrete Choice Demand Models

LINQI ZHANG

1.1 Introduction

Aggregate demand models of differentiated products are crucial for analyzing market power and firm competition in a wide range of industries. The most widely adopted estimation approach developed by Berry (1994) and Berry, Levinsohn, and Pakes (1995) (hereafter referred to as BLP) involves using observed aggregate market shares. Constructing market shares requires researchers to observe the size of the market. Market size consists of all observed sales (the inside goods) plus all potential purchases (the outside goods and no-purchase). Potential purchases are generally unobservable and are therefore a source of possible mismeasurement of market size.[1]

Many empirical results are sensitive to market size (see section 1.2 for details and examples). Yet how to choose market size in demand models has received limited formal

[1] For instance, when estimating airline demand, a market is typically defined as an origin-destination pair of cities. This raises questions about how to determine the number of potential flyers – whether it comprises only those currently traveling by other means, individuals who might opt for travel with lower prices, or the entire population of end-point cities, some of whom may never travel to the destination.

attention in the literature. A few researchers have commented on this problem,[2] but provide little guidance on what to do about unobserved or mismeasured market size.

A common empirical choice is to assume the market size equals the population of the market times a constant.[3] For example, in the demand for soft drinks, this constant represents the maximum amount an individual can potentially consume, which is not observed or estimated in general but chosen ad hoc based on institutional background or consumer behavior. It is important to note that this constant is not a free normalization as it affects the estimates of preferences and counterfactual simulations.

This paper shows how to correct for the unknown market size in random coefficients BLP and other related demand models. For example, in the case where market size is a constant times the observed population, I provide sufficient conditions to point identify and estimate this constant along with all the other parameters of the BLP model. More generally, market size can be point identified and estimated when it is a general function of observed variables and unknown parameters. So, for example, in an airline demand model, market size can be a function of the population in the origin city, population in the destination city, city characteristics like being a hub or not, and a vector of unknown parameters that are identified and estimated along with the rest of the BLP model.

Identification exploits two important features: exogenous variation that shifts quantities across markets and the nonlinearity of the demand model. It does not rely on other information such as micro-moments or additional data beyond those typically used in standard BLP. A key insight is that any exogenous changes in product characteristics affect the total sales of inside goods, and the responsiveness of total sales to this variation depends on the true size of the market. Why does this variation have extra identifying power for parameters beyond ordinary demand coefficients? In section 1.3, I show that the log of product

[2] For example, Berry (1994) states that "issues that might be examined include questions of how to estimate market size when this is not directly observed".

[3] Well-known examples include Nevo (2001), Petrin (2002), Rysman (2004), Berto Villas-Boas (2007), Berry and Jia (2010), Ho, Ho, and Mortimer (2012), Ghose, Ipeirotis, and Li (2012), and Eizenberg and Salvo (2015a), among others.

share in the plain multinomial logit model is linear in product characteristics but nonlinear in market size parameters, making identification possible. More formally, identification is based on conditional moment restrictions and full rank conditions. By explicitly computing the associated Jacobian matrix, I provide low-level assumptions on instruments that serve to identify the market size.

In addition to proving this identification results, I also (a) derive the bias caused by mismeasured market size; (b) establish a test to detect the relevance of instruments for the market size parameter; (c) show identification in models where market size is an unknown function of observed variables; (d) provide stronger conditions that permit point identification and estimation of market size, even when the demand model is not known or nonparametric (e.g., in Berry and Haile (2014)'s nonparametric BLP framework), which allows for testing market size specifications without estimating the demand model; (e) offer simpler identification results for the plain multinomial logit model, e.g., employing market fixed effects.

I demonstrate how the proposed method is related to but different from commonly used approaches – implement a nested logit model or market fixed effects – that aim to reduce biases from unknown market size. I formally demonstrate that these existing approaches have some theoretical basis and intuition. Nevertheless, they are not equivalent to my approach and cannot eliminate all biases. Furthermore, I highlight that a special case of nonparametric estimation of random coefficients is equivalent to estimating the market size, but it requires imposing particular assumptions on the distribution of random coefficients.

Based on these identification results, I apply the proposed method to a merger simulation of carbonated soft drink companies. I specifically select the soda market for several reasons: First, it is a market frequently studied using structural methods that involve estimating random coefficients logit models. Second, the existing literature lacks a consensus on how to define market size. Third, this market satisfies the conditions for strong identification, which I will state in the Monte Carlo simulation section. In the merger analysis, I use both the proposed method and the standard BLP to estimate demand, while assuming a Bertrand competition among firms. Using the estimated market size of 12 servings per week, I predict a price effect that is 31% higher compared to the literature's assumption of 17 servings per week. This market size estimate also suggests that defining market size based on per capita consumption of all non-alcoholic beverages (a common practice in the literature) may be too large. Additionally, in Appendix 1.K, I present a second merger analysis using the constructed cereal data from Nevo (2000). These counterfactual simulations demonstrate substantial gains from the proposed correction.

Furthermore, in the Monte Carlo simulations (Appendix 1.H), I examine what parameters are most sensitive to errors in market size measurements and assess whether adding random coefficients helps mitigate bias. I also show that the proposed approach performs well, particularly when the true share of the outside option is not extremely large, and so my method will generally be useful in applications.

One argument for not correcting the market size issue is the belief that random coefficients or a nesting parameter can partially account for the bias. For example, Miller and Weinberg (2017) state that the nesting parameter ensures that estimates are not too sensitive to the market size measure. Some calculations, such as own- and cross-price elasticities, may exhibit less sensitivity when the model includes random coefficients, as demonstrated in Rysman (2004), Iizuka (2007), and Duch-Brown et al. (2017). However, my simulations and empirical study reveal that more flexible demand models do not fully eliminate biases. Biases are more pronounced in certain calculations, such as outside good elasticities, outside good diversion ratios, choice probabilities, and aggregate price elasticities.[4] This can lead to substantially different results for empirical questions, particularly those related to the outside option share, such as the willingness-to-pay for a new good (see discussion in Conlon and Mortimer 2021), tax or subsidy policies (dependent on aggregate elasticities), and merger analysis (see section 1.2 for examples).

[4] See also Conlon and Mortimer (2021) Table 4, which shows that outside diversion ratios and aggregate elasticities are sensitive to market size in both the BLP automobile application and Nevo's cereal application.

Furthermore, in the Department of Justice (DOJ) documents, the word "market size" appears at a high frequency, implying that the size of a market by itself is a piece of critical and useful information for firms and regulators.[5] This suggests that obtaining a consistent estimate of the true market size is important in itself, in addition to its use in removing model estimate biases.

The proposed method in this paper is transparent and simple to implement. It requires estimating only a few extra nonlinear parameters, along with the standard BLP estimation. Researchers may have tried to estimate market size, but the lack of identification theorem and the unsatisfactory empirical performance or numerical issues with the estimator have hindered the widespread adoption of market size estimation in applied work. I provide conditions under which the market size is identified, discuss the data variation that facilitates identification, and propose tests to assess the relevance of these instruments. I hope this paper can alleviate researchers' uncertainty about the market size. Moreover, whenever the market size itself is important to practitioners or regulators, this method can serve as a means to infer the size of the market. Note, that although the solution is simple, it goes beyond merely adding a regressor or market fixed effects.

The next section is a literature review. In section 1.3, I start with a multinomial logit demand model to illustrate the problem of mismeasured market size and provide identification results. In section 1.4, the results are generalized to the random coefficients logit model. Section 1.5 provides extensions. Section 1.6 presents an empirical application. Section 1.7 summarizes additional results provided in the supplemental appendix, and section 1.8 concludes.

[5] At the DOJ/FTC merger workshop, Newmark (2004) emphasizes the significance of market size/population in price-concentration studies for merger cases. Additionally, firms predict product quantities on the basis of potential market size. The Comments of DOJ on Joint Application Of American Airlines Et Al. state that "To model the benefits of an alliance ... Given a fixed market size, passengers are assigned based on relative attractiveness of different airline offerings."

1.2 Literature Review

In the empirical industrial organization literature, market size is often assumed rather than observed or estimated. Notable examples include previous works that use the number of households in the US as the market size in analyzing the automobile industry (such as BLP and Petrin 2002). Some researchers realize this problem and conduct sensitivity analysis. For instance, various merger analysis papers, including Ivaldi and Verboven (2005), Weinberg and Hosken (2013), Bokhari and Mariuzzo (2018) and Wollmann (2018), perform robustness checks on market size assumptions, using different logit-based models and demand specifications, and find that market size impacts simulated price changes and consumer welfare.

Several other papers also recognize the issue and explicitly incorporate market size estimation into demand models. Bresnahan and Reiss (1987) and Greenstein (1996) both specify market size as a linear function of market characteristics, though theirs is a vertical model rather than BLP. Berry, Carnall, and Spiller (2006) estimate a scaling factor similar to this paper, however, they do not discuss identification as I do, and they do not allow for market size being a more general function of multiple measures. Chu, Leslie, and Sorensen (2011) utilize supply side pricing conditions as additional moments to estimate market size. While their approach does not impose functional form assumptions, it requires one to observe the marginal costs of firms. Sweeting, Roberts, and Gedge (2020) and Li et al. (2022) estimate a generalized gravity equation and define market size as proportional to the expected total passengers predicted from the gravity equation but leave the choice of the proportionality factor to the researcher. Hortaçsu, Oery, and Williams (2022) estimate a Poisson arrival process and use the arrival rate as a proxy measure of market size. Their method applies to settings with individual choice data, whereas I focus on aggregate data.

The closest study to ours is Huang and Rojas (2014), which provides theoretically-founded methods to deal with the market size problem in a random coefficients logit setting, by approximating the unobserved market size as a linear function of market characteristics

(Chamberlain's device). They employ the control function method to handle price endogeneity as in Petrin and Train (2010). By doing so, the unobserved market size becomes an additive term outside of the nonlinear part of the demand function. In contrast, ours is built on the standard BLP framework, where market size enters the moment restrictions in a nonlinear manner. Huang and Rojas (2014)'s method largely relies on the linear additivity and thus can not extend directly to the BLP framework.[6] Their primary focus is on removing bias, while this paper also aims to identify and estimate the market size.

Two other papers have looked at issues arising in constructing market shares. Gandhi, Lu, and Shi (2020) handle the problem of zeros in market share data. Berry, Linton, and Pakes (2004) take into account sampling errors in estimating shares from a sample of consumers. While both papers deal with errors in aggregate market shares, the present paper tackles a different problem, inherent to the model itself rather than features of the data sample. The goal of this paper is to address the more fundamental problem of the unobserved share of the outside option and that all shares will be inconsistent in the limit. Unlike sampling errors that diminish as the sample size increases, the errors I address persist and do not vanish.

More recently, theoretical literature on the identification and estimation of random coefficients aggregate demand model has been growing. Berry and Haile (2014) and Gandhi and Houde (2019) highlight that identification of BLP demand models requires instruments for not only endogenous prices but also endogenous market shares. Other studies that discuss the role of instruments in BLP models include Reynaert and Verboven (2014), and Conlon and Gortmaker (2020). I contribute to this literature by providing low-level conditions on instruments for identification of random coefficients in the standard BLP model, both with and without identifying market size.

Recent work generalizes the parametric demand models to more flexible nonparametric,

[6] Petrin and Train (2010)'s control function approach is an alternative to the BLP approach in dealing with the price endogeneity; which method to use will be application-specific. This discussion is outside the scope of the present paper.

nonseparable demand systems. Nonparametric identification of aggregate demand models is studied by Berry and Haile (2014), Gandhi and Houde (2019), Lu, Shi, and Tao (2021), and Dunker, Hoderlein, and Kaido (2022), among others. This paper also provides conditions for identification of market size in nonparametric specified demand models.

1.3 The Multinomial Logit Demand Model

I first briefly review the setup of a plain multinomial logit model without random coefficients or individual-level covariates. Throughout this section, I assume exogenous prices to simplify the exposition and focus more on market size identification. I then propose a simple model of market size. The model contains assumptions on the unobserved outside shares. Combining both models, I provide, in Theorem 1, assumptions under which demand parameters and market size can be identified. In Appendix 1.E, Theorem 1 is extended to a nested logit model, highlighting the discrepancies and connections between a nest structure and the market size model.

1.3.1 Demand Model

Suppose that we observe T independent markets. A market can refer to a single region in a single time period. Let $\mathscr{J}_t = (1, \cdots, J_t)$ be the set of differentiated products in market t, referred to as inside goods. Let $j = 0$ denote the outside option. As in Berry (1994), I assume the indirect utility of consumer i for product j in market t is characterized by a linear index structure

$$U_{ijt} = X'_{jt}\beta + \xi_{jt} + \varepsilon_{ijt},$$

which depends on a vector of observed market-specific product characteristics $X_{jt} \in \mathbb{R}^L$, unobserved characteristics ξ_{jt}, and idiosyncratic tastes of consumers ε_{ijt}. Consumer tastes are assumed to be independently and identically distributed across consumers and products, with extreme value type I distribution.

Let the average utility index of product j at market t be denoted as $\delta_{jt} = X'_{jt}\beta + \xi_{jt}$, with the mean utility for the outside option being normalized as $\delta_{0t} = 0$.

Let π_{jt} denote the true conditional probability of choosing product j in market t. Each consumer chooses the product that gives rise to the highest utility. This defines the set of unobserved consumer tastes that corresponds to the purchase of good j. The probability of choosing good j is obtained by integrating out over the distribution of consumer tastes ε_{ijt}. Given the functional form and parametric assumptions, the true choice probability takes an analytic form:

$$\pi_{jt} = \frac{exp(\delta_{jt})}{1+\sum_{k=1}^{J_t} exp(\delta_{kt})} \quad \forall j \in \mathscr{J}_t, \quad \text{and } \pi_{0t} = \frac{1}{1+\sum_{k=1}^{J_t} exp(\delta_{kt})}.$$

In a plain logit context, the nonlinear demand system can be inverted to solve for δ_{jt} as a function of choice probabilities, yielding

$$\ln(\pi_{jt}/\pi_{0t}) = X'_{jt}\beta + \xi_{jt} \quad \forall j \in \mathscr{J}_t. \tag{1.1}$$

If the value of π_{jt} and π_{0t} were observed, parameters β can be consistently estimated by regressing $\ln(\pi_{jt}/\pi_{0t})$ on X_{jt}. Generalized Method of Moments (GMM) estimators can be constructed based on the mean independence condition $E(\xi_{jt} \mid X_{jt}) = 0$. The conditions I have imposed so far are standard assumptions made in Berry (1994) and the empirical IO literature, which are sufficient to identify the demand parameters β when the market size is correctly measured and therefore π_{jt} and π_{0t} are observed without errors.

1.3.2 Market Size Model

In this subsection I provide modeling assumptions for the unobserved π_{jt} and π_{0t}. These assumptions allow us to characterize the connection between unobserved probabilities and measures of market size. I then combine these assumptions with the demand system to obtain a new model which I will later prove identification.

Define r^*_{jt} by

$$r^*_{jt} = \frac{\pi_{jt}}{\sum_{k=1}^{J_t} \pi_{kt}} \quad \forall j \in \mathscr{J}_t, \tag{1.2}$$

which is the true conditional choice probability of choosing product j, conditional on purchasing any inside goods. Using equations (1.1) and (1.2), we have

$$\ln\left(r_{jt}^{*}\right) = \ln\left(\frac{\pi_{0t}}{1-\pi_{0t}}\right) + X_{jt}'\beta + \xi_{jt} \quad \forall j \in \mathscr{J}_t. \tag{1.3}$$

Let N_{jt} be the observed sales of good j in market t, and let $N_t^{total} = \sum_{j=1}^{J_t} N_{jt}$ denote the total observed sales of all goods. We observe r_{jt}, where $r_{jt} = N_{jt}/N_t^{total}$ represents the fraction of total purchases spent on good j in market t, and therefore does not depend on the outside option or the size of the total market. I call these r_{jt} *relative shares*, and assume $r_{jt} = r_{jt}^{*}$. In Appendix 1.C, I relax this assumption and allow the true r_{jt}^{*} to be unobservable, introducing sampling errors or measurement errors in r_{jt}.

In general, r_{jt} would be observable along with N_t^{total}. In most empirical contexts, we might directly observe N_{jt}. For example, the number of passengers on flights by airline j in city pair t, or servings of cereals of brand j sold in city t. From these observed N_{jt} we can calculate r_{jt} and N_t^{total}. In other applications, r_{jt} and N_t^{total} might come from separate sources. For instance, r_{jt} could be the fraction of a set of sampled consumers who buy product j in time period t, and N_t^{total} could be separate estimates of total sales in time t.

The issue with not observing market size is not observing π_{0t}. If the total market size were directly observed, we could calculate π_{0t} from the observed N_t^{total} and the market size. However, observing only the relative shares r_{jt} for all J_t goods does not provide sufficient information to determine π_{0t}. Therefore, we need to specify a model for the unobserved outside share. Compared to equation (1.1), the model of equation (1.3) offers the advantage that only the first term on the right side depends on the outside share, and thus it is easier and more natural to impose assumptions on this additively separable term.

Let M_t be some observed population or quantity measure of market t that we believe is related to the true market size. For instance, if a market is defined to be a city, M_t could be the population size (e.g. Nevo 2001; Berto Villas-Boas 2007; Rysman 2004; Ho, Ho, and Mortimer 2012; and Ghose, Ipeirotis, and Li 2012). Alternatively, M_t could be a prediction of total product sales or the number of passengers on a flight (e.g. Sweeting, Roberts, and

Gedge 2020; Li et al. 2022; and Backus, Conlon, and Sinkinson 2021). Let $W_t = M_t / N_t^{total}$ denote *observed market to sales*. As discussed earlier, it is both natural and necessary to place assumptions on π_{0t}. For now, I assume that

$$\ln\left(\frac{\pi_{0t}}{1-\pi_{0t}}\right) = \ln\left(\gamma W_t - 1\right) \tag{1.4}$$

for some constant γ. In Appendix 1.C, I relax equation (1.4) by introducing a random noise term v_t, so that this relationship is approximate rather than exact. In section 1.4, I further generalize the model by allowing π_{0t} to depend on multiple γ's[7].

The model of equation (1.4) is sensible for the following reasons. In the conventional approach, market size is assumed to be a known constant γ multiplied by an observed population measure M_t. In this case, $1 - \pi_{0t} = N_t^{total} / \gamma M_t$ would equal $1/\left(\gamma W_t\right)$, and thus $\ln\left(\pi_{0t}/\left(1-\pi_{0t}\right)\right)$ would equal $\ln\left(\gamma W_t - 1\right)$. Equation (1.4) treats the usual constant γ as unknown rather than known. Furthermore, equation (1.4) is consistent with a deeper economic model, which I elaborate on in section 1.5.1.

Putting the above equations and assumptions together we get the estimating equation

$$\ln\left(r_{jt}\right) = \ln\left(\gamma W_t - 1\right) + X_{jt}'\beta + \xi_{jt} \quad \forall j \in \mathscr{J}_t \tag{1.5}$$

In Appendix 1.B, I demonstrate the bias introduced in estimating β when employing the conventional approach of equation (1.1) with a mismeasured market size. For example, if the market size used in estimation is larger than the true size, the model exhibits a positive correlation between the price of good j and the measurement error, and a negative correlation between the price and its own market share. As a result, the estimated price coefficient will be biased downward (in absolute value), indicating an underestimation of price sensitivity.

[7] An alternative approach to relax this modeling assumption, which I do not explore in the present paper, is to consider γ as a function of observed market-level covariates that affect preferences. I leave this possibility for future research.

1.3.3 Identification

Here I provide identification of model (1.5). Unknown parameters in this model include the market size parameter γ and demand coefficients β. My approach allows for the identification of both the true market size and demand parameters, relying on variation across multiple markets. To achieve this, one would need to observe data from many markets. In Appendix 1.D, I present an alternative approach utilizing market fixed effects. However, while the market fixed effects approach identifies demand parameters, it does not provide identification of the market size.

Assumption 1. *$E\left(\xi_{jt} \mid Q_t, X_{1t}, \ldots, X_{J_t t}\right) = 0$, where Q_t represents instruments for W_t. W_t and Q_t are continuously distributed. The number of markets $T \to \infty$.*

Assumption 1 assumes that the additive error ξ_{jt} is mean independent of product characteristics and some instrument Q_t, and that the regressors have a continuous distribution. Note that the nonlinear variable W_t in equation (1.5) is endogenous since it is a function of quantities. The instrument Q_t can take the form of a vector or a scalar. For the sake of convenience, Theorem 1 employs a scalar Q_t. The large T assumption is necessary as the theorem is based on a conditional expectation conditioning on Q_t, and the derivatives of the conditional expectation. These derivatives would be estimated using nonparametric regression techniques such as kernel regression or local polynomials (Li and Racine 2007). Assuming Q_t is continuous, it asymptotically requires observing all values of Q_t on its support, hence needs T to approach infinity. Moreover, this assumption implies that the instrument Q_t can not be a binary variable.

Theorem 1. *Given Assumption 1 and equation (1.5), let Γ be the set of all possible values of γ, if*

1. *function $f(c,q,x)$ is twice differentiable in (c,q) for every $x \in supp(X_{jt})$, where*

$$f(c,q,x) = E\left(\ln\left(r_{jt}\right) - \ln\left(cW_t - 1\right) \mid Q_t = q, X_{jt} = x\right),$$

2. *and* $\partial E\left(-\frac{W_t}{cW_t-1} \mid Q_t = q, X_{jt} = x\right)/\partial q > 0$ *or* < 0 *for all* $c \in \Gamma$,

then γ *and* β *are identified.*

The proof of Theorem 1, provided in Appendix 1.A, works by showing that there exists q and x such that $g(c,q,x) = 0$ has a unique solution c, where $g(c,q,x) = \partial f(c,q,x)/\partial q$. To provide an idea of what the restrictions in the theorem entail, consider the simplest (although not possible in theory) case where W_t is exogenous. Then W_t serves as an instrument for itself, i.e. $Q_t = W_t$, and so the sufficient condition $\partial E(-\frac{W_t}{cW_t-1} \mid Q_t = q, X_{jt} = x)/\partial q = \partial(-\frac{w}{cw-1})/\partial w = 1/(cw-1)^2 > 0$ is satisfied. On the other hand, if W_t is endogenous but the instrument Q_t is independent of W_t (conditional on X_{jt}), then $E(-\frac{W_t}{cW_t-1} \mid Q_t = q, X_{jt} = x) = E(-\frac{W_t}{cW_t-1} \mid X_{jt} = x)$, which does not depend on q. Therefore the derivative with respect to q would be zero, violating the condition. Generally, the second condition in Theorem 1 is a nonlinear analog of the traditional relevance restriction required in the classical linear IV model, requiring W_t to vary with Q_t in a certain way.

Identification requires an instrument Q_t, which varies with the market total sales and is uncorrelated with the error term ξ_{jt}. A simple candidate satisfying these conditions is the sum of exogenous characteristics of all products in market t. Since product characteristics affect the utilities consumers get and lead to variations in quantities across regions or time periods, the relevance condition is in general satisfied. The exogeneity condition is also satisfied because the error term ξ_{jt} is not only mean independent of characteristics of product j, but also of all other products in market t, making it mean independent of the sum of all products. This resembles the standard BLP instrument, and a detailed discussion of this type of instrument, known as "functions of inside regressors", is deferred to the next section.

An exogenous price change, perhaps driven by tax or subsidy policies, can also serve as an external instrument to identify the market size. For example, to study the demand for alcohol or soda, sin taxes on these products can be utilized to construct the required instruments. Intuitively, after a tax implementation, one can observe the decrease in market quantity, which represents the proportion of consumers who switch to the outside option. In a logit model, substitutions are proportional to the true market shares. The degree of substitution to the outside option depends on the true market size. This suggests that if we observe any variations in the outside diversion (due to changes in Q_t), we can infer the true market size.

Estimation of the model of equation (1.5) based on Theorem 1 is straightforward. It could be done by a standard GMM estimation or nonlinear two-stage least squares estimation using Q_t as instruments.

Visual Intuition

After presenting the formal identification results, I offer visual intuition. As discussed above, exogenous variations in characteristics across markets allow us to observe consumers entering or exiting the outside option, enabling the identification of γ. This exogenous variation is typically already present in the data. For example, in a market with two goods like Coke and Pepsi, when the characteristics of Pepsi get worse, total quantities decrease, leading to an increase in W_t. The identification of γ follows from the relative increase in Coke's shares. If γ is large, a significant number of Pepsi consumers might divert to the outside option, resulting in minimal diversion to Coke. Conversely, if γ is small, it implies that fewer consumers are on the margin. Thus, when Pepsi worsens, more Pepsi users would divert to Coke rather than the outside option.

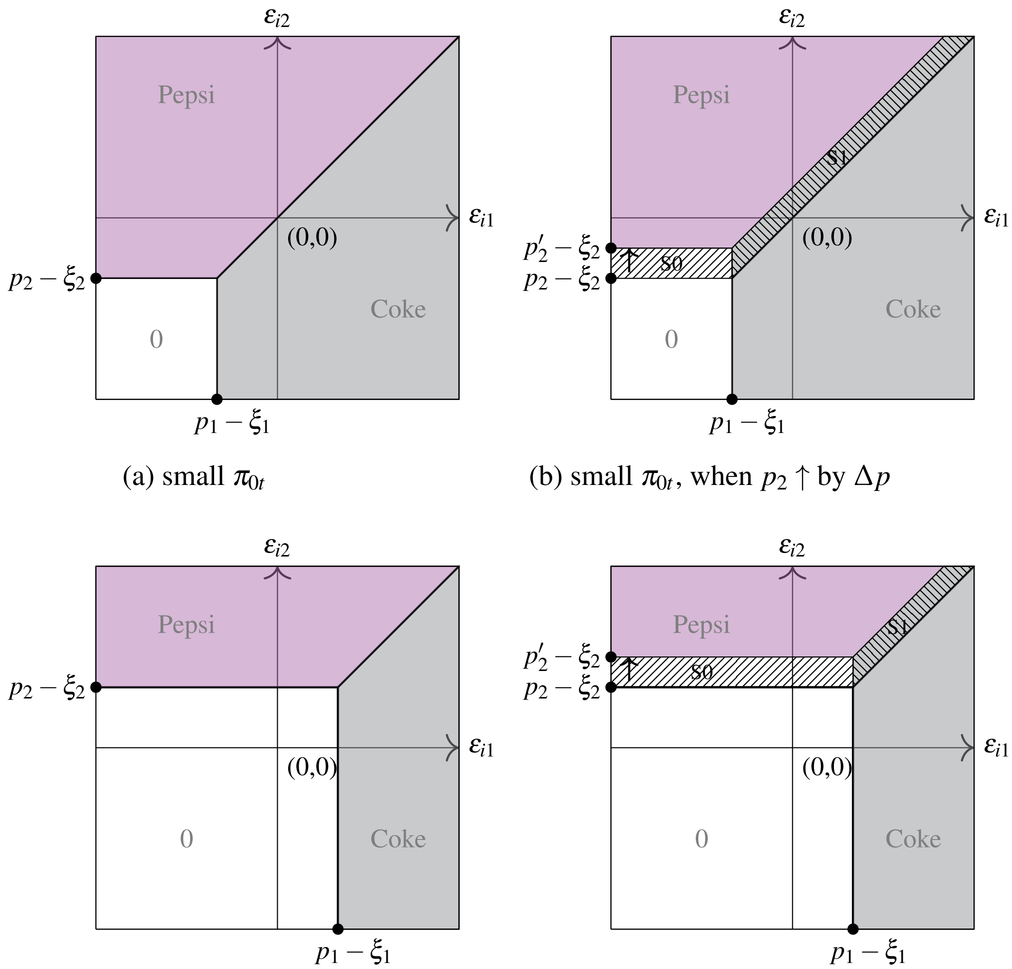
Figure 1.1 illustrates the aforementioned intuition. In a simplified model where $\delta_{jt} = -p_{jt} + \xi_{jt}$, with two goods ($j = 1$ Coke and $j = 2$ Pepsi), the space of ε_{ij} is partitioned into three regions, each corresponding to the choice of $j = 0, 1, 2$ (Berry and Haile 2014 and Thompson 1989). The measure of consumers in each region, i.e. integral of ε over the region, reflects choice probabilities. For example, $Pr(j = 1 \mid p, \xi) = Pr(\varepsilon_{i1} > p_1 - \xi_1; \varepsilon_{i1} > \varepsilon_{i2} + (p_1 - \xi_1) - (p_2 - \xi_2))$. In Figure 1.1, panel (c) illustrates a larger probability of choosing the outside option compared to panel (a), given a fixed known density function of

ε_{ij}. Since the true choice probability π_0 is unknown but only the relative inside good shares r_j are known, then the question we ask is whether the true data generating process (dgp) corresponds to panel (a) or panel (c).

Panels (a) and (b) of Figure 1.1 depict a dgp where the true π_{0t} is small. Panels (c) and (d) show similar graphs but with large true π_{0t}. When the price of good 2 increases, the changes in choice probabilities π_{0t} and π_{1t} are captured by shaded boundaries S_0 and S_1. In panel (b), the price increase prompts more consumers to switch to good 1, while in panel (d), the same price change leads to more consumers switching to the outside option. The relative diversion to the outside option compared to good 1, which is known, relies on the original sizes of each region, which is unknown, and this relationship provides identification of the underlying market size.

In the logit model, we have $(\partial \pi_{1t}/\partial p_{2t})/(\partial \pi_{0t}/\partial p_{2t}) = \pi_{1t}/\pi_{0t}$. As both sides of the equation are ratios, the unobserved choice probabilities can be transformed into observed sales, $(\partial N_{1t}/\partial p_{2t})/(\partial N_{0t}/\partial p_{2t}) = N_{1t}/N_{0t}$. Note that $\partial N_{0t}/\partial p_{2t}$ is observable since the total sales decrease is just the increase in N_{0t}, and vice versa. Thus, the ratio of derivatives on the left side of the equation and N_{1t} are all observed from data, which can help identify the unobserved outside market size N_{0t}.

With the inclusion of random coefficients in section 1.4, the simplified example depicted in Figure 1.1 may not hold anymore. This is because an observed increase in substitution to good 1 could be attributed to good 1 and good 2 being closer substitutes. However, the underlying intuition remains valid: even without the independence of irrelevant alternatives (IIA) property, cross-product substitutions are still functions of the true choice probabilities. Thus, the level of substitution to the outside good will depend on the true market shares. Relative changes in quantities of inside versus outside goods can be exploited to recover the true market size.



(a) small π_{0t}

(b) small π_{0t}, when $p_2 \uparrow$ by Δp

(c) big π_{0t}

(d) big π_{0t}, when $p_2 \uparrow$ by Δp

Figure 1.1: Intuition for Identification in Multinomial Logit Demand Model

1.3.4 The Nested Logit Demand Model

In Appendix 1.E, I establish formal identification of market size in a nested logit demand model. Here I briefly summarize the intuition. Consider the case where all goods are divided up into two nests, one with the outside good as the only choice and the other containing all inside goods. Using our notation, the estimating equation is a nonlinear function of the market size parameter γ and the nesting parameter ρ,

$$\ln(r_{jt}) = \frac{1}{1-\rho}\ln(\gamma W_t - 1) + X_{jt}'\frac{\beta}{1-\rho} + \frac{\xi_{jt}}{1-\rho},$$

and the total derivative with respect to these two parameters has independent variation. I leverage instruments that shift W_t to separately identify γ and ρ.

1.4 The Random Coefficients Logit Demand Model

I generalize our previous results to the random coefficients demand model. I begin by introducing the notation and model assumptions, and then present sufficient conditions for model identification and suggest valid instruments. I then discuss the testing of instrument relevance and provide intuition for identification. Additionally, I derive results for market fixed effects and demonstrate that fixed effects would not be a viable solution for an unknown market size.

1.4.1 Demand Model and Market Size

The utility of consumer i for product j in market t is now given by

$$U_{ijt} = X_{jt}'\beta_i + \xi_{jt} + \varepsilon_{ijt}, \tag{1.6}$$

where $\beta_i = (\beta_{i1}, \cdots, \beta_{iL})$. The individual-specific taste parameter for the l-th characteristics can be decomposed into a mean level term β_l and a deviation from the mean $\sigma_l \nu_{il}$:

$$\beta_{il} = \beta_l + \sigma_l \nu_{il}, \quad \text{with } \nu_i \sim f_\nu(\nu)$$

where ν_{il} captures consumer characteristics. The consumer characteristics could be either observed individual characteristics or unobserved characteristics. When estimating demand models, what econometricians usually have are aggregate data, where no observed individual characteristics are available. Therefore, in the current analysis, I assume ν_{il} are some unobserved characteristics with a known distribution f_ν. The extension to include observed consumer characteristics will be straightforward if there are individual-level data.

As in section 1.3, let δ_{jt} denote the mean utility $X_{jt}'\beta + \xi_{jt}$. Combining equations we have

$$U_{ijt} = \delta_{jt} + \sum_l \sigma_l x_{jtl}^{(2)} \nu_{il} + \varepsilon_{ijt},$$

where $X_{jt}^{(2)} = (x_{jt1}^{(2)}, \cdots, x_{jtL'}^{(2)})$ is a $L' \times 1$ subvector of X_{jt} that has random coefficients and is the nonlinear components of the indirect utility function.

After integrating out over the logit error ε_{ijt}, the true aggregate choice probability is

$$\pi_{jt}\left(\delta_t, X_t^{(2)}; \sigma\right) = \int \frac{exp(\delta_{jt} + \sum_l \sigma_l x_{jtl}^{(2)} \nu_{il})}{1 + \sum_{k=1}^{J_t} exp(\delta_{kt} + \sum_l \sigma_l x_{ktl}^{(2)} \nu_{il})} f_\nu(\nu) d\nu, \tag{1.7}$$

where the arguments in the choice probability function are mean utilities $\delta_t = (\delta_{1t}, \cdots, \delta_{J_t t})$, nonlinear attributes $X_t^{(2)} = (X_{1t}^{(2)}, \cdots, X_{J_t t}^{(2)})$ and taste parameters $\sigma = (\sigma_1, \cdots, \sigma_{L'})$. The choice probability is written as a function of δ_t, $X_t^{(2)}$ and σ in order to highlight its dependence on the mean utilities, nonlinear attributes, and parameters of the model. I suppress the dependence of the choice probability function on ν_i for brevity. The mean utility of outside good is normalized to $\delta_{0t} = 0$.

I next consider a general model of market size. Let $M_t = (M_{1t}, \cdots, M_{Kt})$ be a vector of measures of the market size, and $\gamma = (\gamma_1, \gamma_2)$, $\gamma_1 = (\gamma_{11}, \cdots, \gamma_{K1})$ and $\gamma_2 = (\gamma_{12}, \cdots, \gamma_{K2})$ are two vectors of market size parameters. To ease the exposition, I again assume $r_{jt}^* = r_{jt}$. Observational errors in r_{jt} and other disturbances in the mismeasurement are therefore assumed away. Recall that N_{jt} is the observed sales of each good and N_t^{total} is the total sales of all inside goods. Assumption 2 formalizes the modeling assumption.

Assumption 2. *(a) The observed N_t^{total} and M_t are linked to the unobserved true choice probability $\pi_{0t}\left(\delta_t, X_t^{(2)}; \sigma\right)$ by*

$$1 - \pi_{0t}\left(\delta_t, X_t^{(2)}; \sigma\right) = \frac{N_t^{total}}{\sum_{k=1}^K \gamma_{k1} M_{kt}^{\gamma_{k2}}}.$$

(b) The unobservable true conditional choice probability r_{jt}^* is equal to the observed r_{jt}, i.e.

$$\frac{\pi_{jt}\left(\delta_t, X_t^{(2)}; \sigma\right)}{\sum_{k=1}^{J_t} \pi_{kt}\left(\delta_t, X_t^{(2)}; \sigma\right)} = \frac{N_{jt}}{N_t^{total}}.$$

The market size formula $\sum \gamma_{k1} M_{kt}^{\gamma_{k2}}$ has several appealing features. Taking the airline market as an example, suppose M_{1t} is the population of city A (a small market) and M_{2t} is the population of city B (a big market). The true size of a market defined by these two end-point cities could be $M_{1t}^2 + 3M_{2t}^2$. First, this formula allows for different coefficients for each term. For instance, city B might have a larger coefficient due to being a major transportation hub. Second, it accommodates nonlinearity in M_t. In the airline example, larger metropolitan areas are more likely to have alternative transportation options, such as high-speed rail or highways in multiple directions. Under Assumption 2, the implicit system of demand equations in a given market t is given by

$$\frac{N_t}{\sum_{k=1}^{K} \gamma_{k1} M_{kt}^{\gamma_{k2}}} = \pi_t\left(\delta_t, X_t^{(2)}; \sigma\right), \tag{1.8}$$

where $N_t = (N_{1t}, \cdots, N_{J_t t})$ and $\pi_t(\cdot) = (\pi_{1t}(\cdot), \cdots, \pi_{J_t t}(\cdot))$ represent vectors of observed quantities and *choice probability functions*.

1.4.2 Identification

In a standard BLP model, the link between the choice probability $\pi_{jt}(\delta_t, X_t^{(2)}; \sigma)$ predicted by the model and the observed market shares is crucial. The key to identification and estimation in a standard BLP model is to recover the mean utility δ_t as a function of the observed variables and parameters, by the *inversion* of the demand equation system. This paper builds on the same form of demand inversion while replacing observed market shares with the unobserved ones.

The identification argument can be summarized into two parts: First, I show that for any given parameters (γ, σ) and data (N_t, M_t, X_{jt}), the implicit system of equations (1.8) has a unique solution δ_t for each market[8]. This is supported by Proposition 1, which establishes the existence and uniqueness of demand inversion as shown in Berry (1994) and Berry, Levinsohn, and Pakes (1995), adapted to our framework (see also Berry and Haile

[8] As equation (11) in Berry (1994) shows, the system of market shares used to solve for δ consists of only the inside goods $j = 1, \cdots, J$, not including s_{0t}. However, the existence of good 0 is important both because it has economic meaning, and also because it serves as a technical device, see Berry, Gandhi, and Haile (2013) for a discussion.

(2014) for demand inversion in nonparametric models). Second, once we have a unique sequence of *inverse demand function* $\delta_{jt}(N_t, M_t, X_t^{(2)}; \gamma, \sigma)$, we can construct a corresponding sequence of *residual function* $\xi_{jt}(N_t, M_t, X_t; \gamma, \sigma, \beta)$, which will be defined later. Identification is then based on conditional moment restrictions, and we will require unique solutions to the associated unconditional moment conditions at the true parameter values.

Proposition 1. *Let equations (1.7) and (1.8) hold. Define the function* $g_t : \mathbb{R}^{J_t} \to \mathbb{R}^{J_t}$, *as* $g_t(\delta_t) = \delta_t + \ln(N_t) - \ln(\sum_{k=1}^{K} \gamma_{k1} M_{kt}^{\gamma_{k2}}) - \ln(\pi_t(\delta_t, X_t^{(2)}; \sigma))$. *Given any choice of the model parameters* (γ, σ) *and any given* $(N_t, M_t, X_t^{(2)})$, *there is a unique fixed point* $\delta_t(N_t, M_t, X_t^{(2)}; \gamma, \sigma)$ *to the function* g_t *in* $\mathbb{R}^{J_t}$.

The proof of Proposition 1 closely follows the contraction mapping argument in Berry, Levinsohn, and Pakes (1995). I show that all the conditions in the contraction mapping theorem are satisfied in our setting with the extra vector of γ. Therefore, the function $g(\delta)$ is a contraction mapping.

Proposition 1 shows that there is a unique fixed point δ_t to the function $g_t(\delta_t)$. Now, let $\theta = (\gamma, \sigma, \beta) \in \Theta$ be the full vector of model parameters of dimension $dim(\theta)$. Given the inverse demand function $\delta_{jt}(N_t, M_t, X_t^{(2)}; \gamma, \sigma)$, I define the residual function as

$$\xi_{jt}(N_t, M_t, X_t; \theta) = \delta_{jt}\left(N_t, M_t, X_t^{(2)}; \gamma, \sigma\right) - X'_{jt}\beta. \tag{1.9}$$

The uniqueness of $\delta_{jt}(N_t, M_t, X_t^{(2)}; \gamma, \sigma)$ implies a unique sequence of $\xi_{jt}(N_t, M_t, X_t; \theta)$. Following Berry, Levinsohn, and Pakes (1995), Berry and Haile (2014), and Gandhi and Houde (2019), I will assume that the unobserved structural error term is mean independent of a set of exogenous instruments Z_t, based off which I will later construct unconditional moment conditions. Specifically, I replace the exogenous restriction in section 1.3 with the following conditional moment restriction.

Assumption 3. *Let* $Z_t = (Z_{1t}, \cdots, Z_{Jt})$. *The unobserved product-specific quality is mean independent of a vector of instruments* Z_t*:*

$$E\left(\xi_{jt}(N_t, M_t, X_t; \theta_0) \mid Z_t\right) = 0.$$

Define $h_{jt}(\theta) = \xi_{jt}(N_t, M_t, X_t; \theta)\phi_j(Z_t)$*, where* $\phi_j(Z_t)$ *is a* $m \times 1$ *vector function of the instruments with* $m \geq dim(\theta)$*. Then the conditional moment restriction implies*

$$E\left(h_{jt}(\theta_0)\right) = 0.$$

The instrument vector Z_t typically includes a subvector of X_t that contains exogenous characteristics and excluded price instruments such as cost shifters. The assumption posits that the structural error is mean independent not only of the exogenous covariates of product j but also of all other products. Similar to standard BLP models, two types of instruments are generally required: (i) price instruments and (ii) instruments that identify nonlinear parameters (σ and γ). I will discuss these instruments in detail in the next subsection.

Showing function $g_t(\delta_t)$ has a unique fixed point δ_t is only a necessary condition for identification. To complete the proof of point identification, we need conditions that are sufficient for the existence of a unique solution to the moments.

Definition 1. θ_0 *is globally identified if and only if the equations* $E\left(h_{jt}(\theta)\right) = 0$ *have a unique solution at* $\theta = \theta_0$*. In other words,*

$$E\left(h_{jt}(\tilde{\theta})\right) = 0 \iff \tilde{\theta} = \theta_0, \text{ for all } \tilde{\theta} \in \Theta \tag{1.10}$$

θ_0 *is locally identified if (1.10) holds only for* $\tilde{\theta}$ *in an open neighborhood of* θ_0*.*

I formally define *local identification* in Definition 1. Assumption 4 in Berry and Haile (2014) and equation (5) in Gandhi and Houde (2019) both impose a similar high-level identification assumption to (1.10). Theorem 5.1.1 in Hsiao (1983) (in line with Fisher 1966 and Rothenberg 1971) provides sufficient rank conditions for the identification assumption stated above to hold locally, which I summarize in Proposition 2.

Proposition 2 (Theorem 5.1.1 in Hsiao 1983)**.** *If* θ_0 *is a regular point, a necessary and sufficient condition that* θ_0 *be a locally isolated solution is that the* $m \times dim(\theta)$ *Jacobian matrix formed by taking partial derivatives of* $E\left(h_{jt}(\theta)\right)$ *with respect to* θ*,* $\nabla_\theta E\left(h_{jt}(\theta)\right)$ *has rank* $dim(\theta)$ *at* θ_0*.*

The idea of using full rank conditions to establish identification in nonlinear simultaneous equations models dates back to Fisher (1966) and Rothenberg (1971). See Hsiao (1983) for a comprehensive review. The application of full rank conditions for achieving local identification is seen in various studies, including McConnell and Phipps (1987), Iskrev (2010), Qu and Tkachenko (2012), Milunovich and Yang (2013), and Gospodinov and Ng (2015). Using Proposition 2, I can now establish an identification theorem for the random coefficients demand model with an unobserved market size.

Theorem 2. *Under Assumptions 2 and 3, if the rank of*

$$E\left[\phi_j(Z_t)\frac{\partial\delta_{jt}(N_t,M_t,X_t^{(2)};\gamma,\sigma)}{\partial\gamma'} \quad \phi_j(Z_t)\frac{\partial\delta_{jt}(N_t,M_t,X_t^{(2)};\gamma,\sigma)}{\partial\sigma'} \quad \phi_j(Z_t)X'_{jt}\right]$$

is $dim(\theta)$ at θ_0, then θ is locally identified.

Standard BLP models require a rank condition similar to the one stated in Theorem 2, but not the same because it does not have the extra γ rows and columns in the Jacobian matrix. These moments depend on the inverse demand function, which lacks a closed-form expression, making it challenging to directly verify full column rank. However, I show that the full rank condition is generally satisfied due to the high nonlinearity of the demand system. The rank condition is testable using the test of the null of underidentification proposed by Wright (2003).

Sufficient Conditions for Identification

I replace the high-level rank condition with some low-level conditions on instruments. The identification theorem imposes an assumption regarding the rank of the Jacobian matrix. This rank condition will generally hold because the total derivative of the demand system (1.8) with respect to parameters exhibits independent variation. To verify the rank of the Jacobian matrix, I calculate the derivatives of $h_{jt}(\theta)$. The Jacobian matrix encompasses four sets of derivatives: derivatives with respect to γ_1, γ_2, σ and β, respectively. By utilizing the implicit function theorem for a system of equations (Sydsæter et al. 2008) and applying the Cramer's rule, the first two sets of derivatives can be explicitly computed as

$$
\begin{aligned}
J_1 &= \frac{\partial h_{jt}(\theta)}{\partial \gamma_{k1}} = \begin{vmatrix} \frac{\partial \pi_{1t}}{\partial \delta_{1t}} & \cdots & \frac{\partial \pi_{1t}}{\partial \delta_{Jt}} \\ \vdots & \ddots & \vdots \\ \frac{\partial \pi_{Jt}}{\partial \delta_{1t}} & \cdots & \frac{\partial \pi_{Jt}}{\partial \delta_{Jt}} \end{vmatrix}^{-1} \underbrace{\begin{vmatrix} \frac{\partial \pi_{1t}}{\partial \delta_{1t}} & \cdots & \pi_{1t} & \cdots & \frac{\partial \pi_{1t}}{\partial \delta_{Jt}} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \frac{\partial \pi_{Jt}}{\partial \delta_{1t}} & \cdots & \pi_{Jt} & \cdots & \frac{\partial \pi_{Jt}}{\partial \delta_{Jt}} \end{vmatrix}}_{(\pi_{1t},\cdots,\pi_{Jt})' \text{ is in the } j\text{-th column}} \frac{M_{kt}^{\gamma_{k2}}}{\sum_k \gamma_{k1} M_{kt}^{\gamma_{k2}}} \phi_j(Z_t), \\
&= \Psi_{jt}\left(\delta_t, X_t^{(2)}; \sigma\right) \frac{M_{kt}^{\gamma_{k2}}}{\sum_k \gamma_{k1} M_{kt}^{\gamma_{k2}}} \phi_j(Z_t),
\end{aligned} \tag{1.11}
$$

and

$$
J_2 = \frac{\partial h_{jt}(\theta)}{\partial \gamma_{k2}} = \Psi_{jt}\left(\delta_t, X_t^{(2)}; \sigma\right) \frac{\gamma_{k1} \ln(M_{kt}) M_{kt}^{\gamma_{k2}}}{\sum_k \gamma_{k1} M_{kt}^{\gamma_{k2}}} \phi_j(Z_t) \tag{1.12}
$$

where J_1 and J_2 are $m \times 1$ vectors, and $\Psi_{jt}(\delta_t, X_t^{(2)}; \sigma)$ denotes the product of the first two matrix determinants in equation (1.11). I emphasize its dependence on δ_t and $X_t^{(2)}$ because the partial derivatives of π_{jt} with respect to δ_{jt} and δ_{kt} are functions of mean utilities and characteristics of all products. I provide the calculation of these partial derivatives in Appendix 1.L. The Jacobian determinant of $(\pi_{1t}, \cdots, \pi_{Jt})'$ with respect to $(\delta_{1t}, \cdots, \delta_{Jt})$ is different from zero, so the condition of implicit function theorem is satisfied.

Identification fails when two or more parameters enter the demand system in a manner that makes it impossible to distinguish them. In such cases, the associated columns of the Jacobian matrix become linearly dependent. For example: if M_t were independent of $\phi_j(Z_t)$ and all other components in the demand model, we would essentially have $E(\partial h_{jt}/\partial \gamma_{k1}) = cE(\partial h_{jt}/\partial \gamma_{k2})$, for some non-zero constant c. This would make it impossible to separately identify γ_{k1} and γ_{k2}, neither could we distinguish γ_{k1} and γ_{j1} for $j \neq k$.To disentangle the γ vector, we require some instruments that change M_t exogenously. For example, if M_t is population, then instruments could be expansions of highways in a city.

The third group of derivatives is

$$
\begin{aligned}
J_3 &= \frac{\partial h_{jt}(\theta)}{\partial \sigma_l} = \begin{vmatrix} \frac{\partial \pi_{1t}}{\partial \delta_{1t}} & \cdots & \frac{\partial \pi_{1t}}{\partial \delta_{Jt}} \\ \vdots & \ddots & \vdots \\ \frac{\partial \pi_{Jt}}{\partial \delta_{1t}} & \cdots & \frac{\partial \pi_{Jt}}{\partial \delta_{Jt}} \end{vmatrix}^{-1} \underbrace{\begin{vmatrix} \frac{\partial \pi_{1t}}{\partial \delta_{1t}} & \cdots & -\frac{\partial \pi_{1t}}{\partial \sigma_l} & \cdots & \frac{\partial \pi_{1t}}{\partial \delta_{Jt}} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \frac{\partial \pi_{Jt}}{\partial \delta_{1t}} & \cdots & -\frac{\partial \pi_{Jt}}{\partial \sigma_l} & \cdots & \frac{\partial \pi_{Jt}}{\partial \delta_{Jt}} \end{vmatrix}}_{(-\partial \pi_{1t}/\partial \sigma_l, \cdots, -\partial \pi_{Jt}/\partial \sigma_l)' \text{ is in the } j\text{-th column}} \phi_j(Z_t) \\
&= \Phi_{jt}(\delta_t, X_t^{(2)}; \sigma)\phi_j(Z_t),
\end{aligned}
$$

where I let the product of the two determinants of J_3 be denoted as $\Phi_{jt}(\delta_t, X_t^{(2)}; \sigma)$. Comparing J_3 with J_1 (or J_2), the first determinant term of $\Phi_{jt}(\delta_t, X_t^{(2)}; \sigma)$ and $\Psi_{jt}(\delta_t, X_t^{(2)}; \sigma)$ are identical. The difference lies in the j-th column of the second determinant term, which is $(-\partial \pi_{1t}/\partial \sigma_l, \cdots, -\partial \pi_{Jt}/\partial \sigma_l)'$ for J_3, and $(\pi_{1t}, \cdots, \pi_{Jt})'$ for J_1 and J_2. Observe that the derivative $\partial \pi_{jt}(\delta_t, X_t^{(2)}; \sigma)/\partial \sigma_l$ and $\pi_{jt}(\delta_t, X_t^{(2)}; \sigma)$ are not perfectly collinear[9], implying that $\Psi_{jt}(\delta_t, X_t^{(2)}; \sigma)$ is not perfect multicollinear with $\Phi_{jt}(\delta_t, X_t^{(2)}; \sigma)$ in general. The column vectors of the Jacobian matrix are therefore linearly independent as long as we have a sufficient number of instruments that are correlated with $\Psi_{jt}(\delta_t, X_t^{(2)}; \sigma)$ and $\Phi_{jt}(\delta_t, X_t^{(2)}; \sigma)$, respectively.

Lemma 1. *Suppose γ is a scalar. Let $\phi_j^{(1)}(Z_t)$, $\phi_j^{(2)}(Z_t)$ and $\phi_j^{(3)}(Z_t)$ be subvectors of $\phi_j(Z_t)$. The rank condition for identification given in Theorem 4 is satisfied if $E(\phi_j^{(1)}(Z_t)X_t')$ is non-singular, the support of $\phi_j(Z_t)$ does not lie in a proper linear subspace of $\mathbb{R}^{dim(\theta)}$, and there are instruments that satisfy*

$$Cov\left(\Psi_{jt}\left(\delta_t, X_t^{(2)}; \sigma\right), \phi_j^{(2)}(Z_t)\right) \neq 0, \tag{1.13}$$

and

$$Cov\left(\Phi_{jt}\left(\delta_t, X_t^{(2)}; \sigma\right), \phi_j^{(3)}(Z_t)\right) \neq 0, \tag{1.14}$$

[9]Specifically, for the j-th column of the above matrices, we have

$$\pi_{jt}\left(\delta_t, X_t^{(2)}; \sigma\right) = \int \pi_{jti}\left(\delta_t, X_t^{(2)}; \sigma\right) f_\nu(\nu) d\nu \quad \text{for } J_1 \text{ (or } J_2\text{), and}$$

$$\frac{\partial \pi_{jt}\left(\delta_t, X_t^{(2)}; \sigma\right)}{\partial \sigma_l} = \int \pi_{jti}\left(\delta_t, X_t^{(2)}; \sigma\right)\left(x_{jtl}^{(2)} - \sum_{k=1}^{J} x_{ktl}^{(2)} \pi_{kti}\left(\delta_t, X_t^{(2)}; \sigma\right)\right) \nu_{il} f_\nu(\nu) d\nu \quad \text{for } J_3.$$

where $\phi_j^{(2)}(Z_t)$ is of dimension one, and $\phi_j^{(3)}(Z_t)$ has the same dimension as σ.

Collectively, to identify market size parameters, two sets of instruments are required: (1) shifters of market size measures M_t, and (2) variables that provide exogenous variations in quantities. In the simple case where $\gamma_2 = 0$ and γ_1 is a scalar, we only need the second set of instruments, which is the same as those needed for identifying random coefficients. Note that for standard BLP, assumptions similar to those in Lemma 1 are necessary, but we only need instruments that satisfy condition (1.14). However, when γ is a vector of dimension greater than one, we need an additional source of variation to identify elements of the γ vector, specifically through variation in measures of market size.

Valid potential instruments that satisfy (1.13) and (1.14) are functions of exogenous product characteristics. This means that the proposed method can be implemented without requiring a new class of outside instruments over and above those commonly used in BLP models, or any additional independent variations in data. Examples of commonly used instruments of this type include: (i) BLP instruments, which are sums of product characteristics of other products produced by the same firm, and the sums of product characteristics offered by rival firms, and (ii) differentiation instruments, which are sums of differences of products in characteristics space (Gandhi and Houde 2019). The rationale behind Gandhi and Houde's differentiation instruments is that demand for a product is mostly influenced by other products that are very similar in the characteristics space. However, the validity of differentiation instruments depends on the symmetry property of the demand function, which has not been shown in my model. Since the introduction of γ breaks the symmetry property that was used to derive these instruments, one can no longer treat the outside option the same as inside goods. Therefore, in the empirical section, I use BLP-type instruments to obtain the main results and employ differentiation instruments as a robustness check. Another set of valid instruments is Chamberlain's (1987) optimal instrument, as implemented in BLP by Reynaert and Verboven (2014). The optimal instrument is the expected value of the Jacobian of inverse demand function, which, in the context of this paper, is equivalent to using $E(\Psi_{jt}(\delta_t, X_t^{(2)}; \sigma) \mid Z_t)$ and $E(\Phi_{jt}(\delta_t, X_t^{(2)}; \sigma) \mid Z_t)$ as instruments.

1.4.3 Relevance of Instruments

Gandhi and Houde (2019) show that the relevance of instruments in BLP models can be tested by estimating a plain logit regression on product characteristics and instruments, with the coefficients determining the strength of these instruments. I re-define the parameters and show that the same test of instrument relevance can be applied in the setting of this paper, for both the random coefficients and the market size parameter.

Gandhi and Houde (2019) use λ to denote the vector of parameters that determine the joint distribution of the random coefficients. Here I follow this notation and extend it to include the market size parameters. Specifically, let $\lambda_\sigma = \sigma$, $\lambda_{\gamma_1} = \gamma_1 - 1$ and $\lambda_{\gamma_2} = \gamma_2$, and $\lambda = (\lambda_\sigma, \lambda_{\gamma_1}, \lambda_{\gamma_2})$ be the full vector of nonlinear parameters in the model. By absorbing λ_γ into the conditioning parameter vector, we rewrite equation (1.9) as

$$\xi_{jt}(N_t, M_t, X_t; \theta) = \delta_{jt}\left(N_t, M_t, X_t^{(2)}; \lambda\right) - X_{jt}'\beta. \tag{1.15}$$

Equation (1.15) encompasses equation (1.9) and is similar to equation (4) in Gandhi and Houde (2019). Here I have (N_t, M_t) instead of the observed market shares s_t in their function.

The endogenous problem arises for λ_σ and λ_γ because the inverse demand function depends on quantities N_t (or market shares) of all products, and these endogenous quantities interact nonlinearly with λ_σ and λ_γ in the inverse demand function. Therefore, we need instrumental variables for quantities (or market shares) of products to identify λ_σ and λ_γ. This is the nonlinear simultaneous equations model that has been previously studied by Jorgenson and Laffont (1974) and Amemiya (1974). Unlike in linear models, where the strength of instruments can be assessed by linear regression of endogenous variables on excluded instruments, for nonlinear models, how to detect weak instruments is not obvious.

I use the method as in Gandhi and Houde (2019) to test the relevance of instruments for identifying λ_σ and λ_γ, which I summarize here. By equation (7) in Gandhi and Houde (2019), the reduced form of the inverse demand function $E\left(\delta_{jt}(N_t, M_t, X_t^{(2)}; \lambda) \mid Z_t\right)$ can be approximated by a linear projection onto functions of instruments:

$$E\left(\delta_{jt}\left(N_t, M_t, X_t^{(2)}; \lambda\right) \mid Z_t\right) \approx \phi_j(Z_t)'\alpha.$$

Definition 1 in Gandhi and Houde (2019) provides a practical method referred to as "IIA-test" to detect the strength of the instruments by evaluating the inverse demand function at $\lambda = 0$ (suppose the true parameters are $\lambda_0 \neq 0$). Evaluating the inverse demand function at $\lambda_\sigma = \lambda_{\gamma_1} = \lambda_{\gamma_2} = 0$, we have

$$\begin{aligned} E\left(\delta_{jt}\left(N_t, M_t, X_t^{(2)}; \lambda = 0\right) \mid Z_t\right) &= E\left(\ln\left(\frac{N_{jt}}{M_t - \sum_{j=1}^{J_t} N_{jt}}\right) \mid Z_t\right) \\ &\approx X_{jt}'\alpha_1 + \alpha_p \hat{P}_{jt} + \phi_j^{-X}(Z_t)'\alpha_2, \end{aligned}$$

where $\hat{P}_{jt}$ is the projection of prices on X_t and price instruments, and $\phi_j^{-X}(Z_t)$ is a subvector of instruments excluding X_t. Note that $\hat{P}_{jt}$ is constructed based on exogenous variables and thus satisfied the mean independence restriction of Assumption 3. The regression relates the observed product quantities to product characteristics and functions of instruments. The null hypothesis of the test is that the model exhibits IIA preference and market shares calculated by N_{jt}/M_t are not mismeasured. One can reject the null hypothesis when the parameter vector α_2 in the reduced form regression is different from zero. On the other hand, when α_2 is close to zero, it indicates that the instruments are weak.

1.4.4 Intuition for Identification

I provide additional intuition for separately identifying γ and σ. First, I show how the intuition for identification in a plain logit model can be applied here. Second, I provide a brief numerical example to visually illustrate the identification.

In section 1.3, I show that γ is identified in a plain logit model by the exogenous variation in W_t. Recall that $W_t = M_t / N_t^{total}$. Rewriting equation (1.5) gives us an alternative way of understanding γ identification in the plain logit model:

$$\ln \frac{r_{jt}}{W_t} = \ln \gamma + \ln \left(\frac{\exp(X_{jt}'\beta + \xi_{jt})}{1 + \sum_{k=1}^{J_t} \exp(X_{kt}'\beta + \xi_{kt})} \right).$$

The left side of the regression is observed, and it is linear in γ, but nonlinear in β. As shown earlier, γ is identified in this regression. Same logic carries over to the random coefficients case. For a scalar γ, we can rewrite equation (1.8) as

$$\ln \frac{r_{jt}}{W_t} = \ln \gamma + \ln \left(\int \frac{\exp(X_{jt}'\beta + \xi_{jt} + \sum_l \sigma_l x_{jtl} \nu_{il})}{1 + \sum_{k=1}^{J_t} \exp(X_{kt}'\beta + \xi_{kt} + \sum_l \sigma_l x_{ktl} \nu_{il})} f_\nu(\nu) d\nu \right),$$

which is again linear in γ, but nonlinear in β and σ. I can exploit the same nonlinearity as in the simple logit case to distinguish γ and (β, σ).

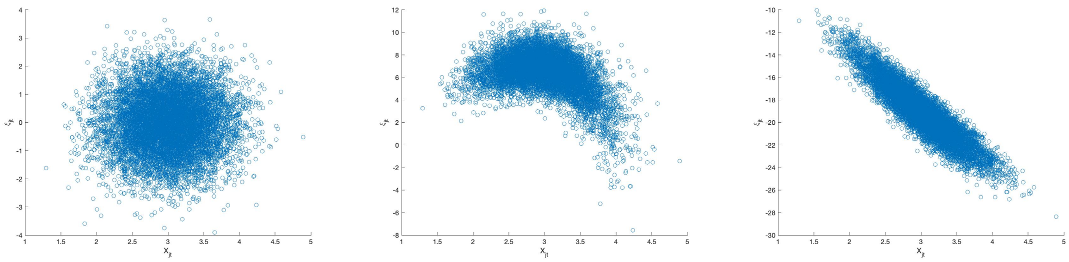
A Numerical Illustration

For the numerical illustration, I consider a model that has only one nonlinear parameter σ. The utility to consumer i for product j in market t is $U_{ijt} = \sigma \nu_i X_{jt} + \xi_{jt} + \varepsilon_{ijt}$, and the market size is parameterized by a single scalar γ. Equation (1.8) can be written as $\frac{N_{jt}}{\gamma M_t} = \int \frac{\exp(\xi_{jt} + \sigma \nu_i X_{jt})}{1 + \sum_{k=1}^{J} (\xi_{kt} + \sigma \nu_i X_{kt})} f_\nu(\nu) d\nu$.

If we do not have any additional conditional moment restrictions, γ is not point identified. To see this, recognize that for a given wrong value $\tilde{\gamma}$, one can construct a corresponding wrong $\tilde{\xi}_{jt}$ that fits equally well by letting $\tilde{\xi}_{jt}$ be given by $\frac{N_{jt}}{\tilde{\gamma} M_t} = \int \frac{\exp(\tilde{\xi}_{jt} + \sigma \nu_i X_{jt})}{1 + \sum_{k=1}^{J} (\tilde{\xi}_{kt} + \sigma \nu_i X_{kt})} f_\nu(\nu) d\nu$.

Put differently, for any value of $\tilde{\gamma}$, the implied $\tilde{\xi}_{jt}$ will adjust to set the predicted choice probabilities equal to the observed shares $N_{jt} / \tilde{\gamma} M_t$. That is why we need Assumption 3 $E(\xi_{jt}(\theta_0) \mid Z_t) = 0$ to normalize the location of ξ_{jt}. Following a similar idea in Gandhi and Nevo (2021), in Figure 1.2, I visually illustrate the intuition for identification and why one can distinguish γ and σ.

Figure 1.2 plots X_{jt} against the implied residual function $\xi_{jt}(\sigma, \gamma)$ for different values of (σ, γ). As depicted in Figure 1.2(a), there is no correlation between ξ and the X at the true parameter values. Figure 1.2(b) shows that when σ is different from the truth, it exhibits a hump-shaped correlation and Figure 1.2(c) shows that when γ is different from the truth, there is a linear correlation. For the wrong σ or γ to fit the data, ξ would have to be correlated with the instruments. Therefore once we assume that ξ is mean independent of X, we are shutting down this channel (as in Gandhi and Nevo 2021). Only at the true parameter values can we match the market shares. Furthermore, the graphs with wrong σ or wrong γ have different shapes, which provide information to distinguish these two parameters.



(a) $\sigma = \sigma^0, \gamma = \gamma^0$ (b) $\sigma > \sigma^0, \gamma = \gamma^0$ (c) $\sigma = \sigma^0, \gamma > \gamma^0$

Figure 1.2: Intuition for Identification in Random Coefficients Logit

Notes: The figure shows a scatter plot of ξ_{jt} and the characteristics X_{jt} under three scenarios. (a) $\sigma = \sigma^0 = 5, \gamma = \gamma^0 = 1$, (b) $\sigma = 15, \gamma = \gamma^0 = 1$, and (c) $\sigma = \sigma^0 = 5, \gamma = 4$.

1.4.5 Market Fixed Effects

In Appendix 1.D I show that in a plain logit model, by including market fixed effects in the regression, one could obtain consistent estimators of β without observing or estimating the true market size. Here, I briefly discuss why the same approach cannot be taken in the random coefficients case. The more detailed derivation is provided in Appendix 1.G.

For plain multinomial logit, when the choice probabilities of all products are rescaled by the same factor, it implies that the quality (mean utility δ) of inside goods has changed by the same amount. These quality gaps can be captured using market dummies. In contrast, for random coefficients logit, the difference in choice probabilities is also driven by consumer taste heterogeneity. Mean utilities δ alone do not pin down choice probabilities. Consequently, when rescaling shares, the implied quality gap varies across alternatives, depending on individual heterogeneous preferences. Market fixed effects cannot fully capture this additional preference variation.

1.5 Extensions

1.5.1 Nonparametric Random Coefficients

In this section I show that identifying and estimating market size in the form of γM_t can be equivalent to nonparametric identification and estimation of a peculiar form of random coefficients. It is in this unusual sense that one may rationalize the belief that random coefficients can compensate for failure to correctly observe market size. More specifically, consider a model with indirect utility given by equation (1.6) and $\beta_i \sim F(\beta)$ follows an unknown distribution. Identifying and estimating $F(\beta)$ can be done nonparametrically. Following the approach of Fox, Kim, and Yang (2016a) (Example 1 in their paper), using a sieve space approximation to the distribution of random coefficients, we can write

$$\pi_{jt}(\delta_{jt};\sigma) = \sum_{r=1}^{R} \sigma_r \frac{\exp(\delta_{jt} + \sum_l \eta_l^r x_{jtl})}{1 + \sum_{k=1}^{J_t} \exp(\delta_{kt} + \sum_l \eta_l^r x_{ktl})} \tag{1.16}$$

with restrictions

$$\sum_{r=1}^{R} \sigma_r = 1 \text{ and } 0 \leq \sigma_r \leq 1,$$

where $\eta_l = (\eta_l^1 \cdots \eta_l^R)$ is a fixed grid of values chosen by researchers. Parameters to be estimated are the weights $\sigma = (\sigma_1 \cdots \sigma_R)$. The associated maximum likelihood estimator was originally proposed for estimation with individual choice data. Here instead I apply this approach in a BLP setting where only aggregate level data is available.

Consider a special case where there are only two types of consumers ($R = 2$), and we aim to identify the probability mass of each type of consumer. Suppose, without loss of generality, that only the constant term has a random coefficient. Let $\eta_1 = -\infty$ and $\eta_2 = 0$ (any values other than 0 would be absorbed into the constant term of δ). The model reduces to

$$\pi_{jt}(\delta_{jt};\sigma) = \sigma_2 \frac{\exp(\delta_{jt})}{1 + \sum_{k=1}^{J_t} \exp(\delta_{kt})},$$

where the equality follows from $\eta_1 = -\infty$ and $\eta_2 = 0$. Note that σ_2 plays the same role as the scalar γ discussed in section 1.3 for the simple logit model. This result can be extended to $R > 2$. If an element of η is $-\infty$, it implies that certain consumers will never purchase any inside goods under any circumstances. These consumers should not be considered *potential* consumers and should be excluded from the measure of market size. In general, the most flexible model of this kind can be approximated by a distribution with a probability mass at negative infinity. Estimating random coefficients in this way allows for flexible consumer tastes and accounts for the unobserved market size of the form γM_t.

Nonparametric random coefficients can address the unknown market size issue if the distribution follows the specified form. This might be where the intuition that random coefficients can partially resolve the problem originates. In a standard BLP model (1.7) with the common distributional assumption $\nu \sim N(0,1)$, since the normal distribution has unbounded support, if the estimated value of $\hat{\sigma}$ in model (1.7) is large, a random draw ν_i with a large negative value from the normal distribution can result in $\hat{\sigma}\nu_i$ approaching negative infinity, similar to $\eta_1 = -\infty$.

Identification of random coefficients distribution of this particular type (one that has a probability mass point at negative infinity) would require strong assumptions. In the literature on nonparametric identification of random coefficients for aggregate demand, Berry and Haile (2014) and Dunker, Hoderlein, and Kaido (2022) prove identification of random coefficients without any restriction on the distribution (i.e., allow for infinite absolute moments). However, both require full/large support of product characteristics or prices (e.g., Assumption 3.3(i) in Dunker, Hoderlein, and Kaido 2022).

Moreover, estimating the random coefficients distribution using a sieve space approximation might not be feasible in the BLP setting. While Wang (2022) proposes a sieve BLP estimation for aggregate demand, the implementation differs significantly from the approach in Fox, Kim, and Yang (2016a), and the choice probability cannot be expressed in the form of equation (1.16). Furthermore, sieve BLP requires the number of instruments to be at least the number of parameters, which corresponds to the dimension of the sieve space (unless we have a moment condition for each j instead of pooling across products, like in Wang 2022). This suggests that an unfeasibly large amount of instruments would be required.

1.5.2 Nonparametric Identification of Market Size

Under stronger conditions, the parametric model of market size considered in prior sections can be extended to a more general specification where the true market size is an unknown function of the vector of measures $M_t \in \mathbb{R}^K$. For the moment, I consider the plain logit setting. I replace the model of true market size with $s(M_t)$, where $s(\cdot)$ is an unknown function. Under this assumption, the estimating equation becomes

$$\ln(r_{jt}) = \ln\left(\frac{s(M_t)}{N_t^{total}} - 1\right) + X_{jt}'\beta + \xi_{jt},$$

which is a partially linear regression with an endogenous nonparametric part studied by Ai and Chen (2003) (see also Newey and Powell 2003 and Chen and Pouzo 2009; see Robinson 1988 for an exogenous nonparametric part). Implicitly, I allow market size measures to be endogenous in the sense that $E(M_t\xi_{jt}) \neq 0$. Identification of β and $s(\cdot)$ can be achieved by imposing assumptions similar to those in Ai and Chen (2003). I summarize it in the following theorem.

Theorem 3. *Let $\Lambda_c^b(\cdot) = \{g \in \Lambda^b(\cdot) : \|g\|_{\Lambda^b} \leq c < \infty\}$ be a Hölder ball with radius c, where $\|g\|_{\Lambda^b}$ is the Hölder norm of order b. Let $Y_t = (N_t^{total}, M_t)$, $Z_{jt} = (X_{jt}, Q_t)$, and $dim(Q_t) = dim(Y_t) = K+1$. Suppose the following hold: (i) $E(\xi_{jt} \mid Z_{jt}) = 0$; (ii) The conditional distribution of Y_t given Z_{jt} is complete; (iii) $s(\cdot) \in \Lambda_c^b(\mathbb{R}^K)$; (iv) $E\left(\ln\left(\frac{s(M_t)}{N_t^{total}} - 1\right) \mid Z_{jt}\right) \notin$ linear span(X_{jt}), and $E\left(X_{jt}X_{jt}'\right)$ is non-singular. Then β and $s(\cdot)$ are identified.*

The proof follows from Newey and Powell (2003) and Proposition 3.1 in Ai and Chen (2003), relying on the completeness of the conditional distribution[10]. Ai and Chen (2003) propose a sieve minimum distance estimator to estimate β and $s(\cdot)$. By restricting the

[10]See Lehmann and Romano (2005) for the concept of statistical completeness. Andrews (2017) provides examples of distributions that are complete.

unknown function to a Hölder space, the function is smooth and one can approximate it using a wide range of sieve basis.

1.5.3 Identification With a Nonparametric Demand Model

The identification and estimation in sections 3 and 4 are based on parametric demand models with logit error terms and known distribution of the random variable ν. However, in some applications, these distribution assumptions on individual tastes may appear to be arbitrary and relatively strong. Thus, I generalize the results to a fully nonparametric model of BLP in the spirit of Berry and Haile (2014) to accommodate less restrictive consumer preferences. The demand system is as equation (1.8), but with an unknown function $\pi_t(\cdot)$ replacing the regular logit formula and an unknown function $s(\cdot)$ being the true market size, yielding

$$\frac{N_{jt}}{s(M_t)} = \pi_j\left(\delta_t, X_t^{(2)}\right), \quad j = 1, \cdots, J. \tag{1.17}$$

The following results show that under a stronger exogeneity condition, (1) the market size function $s(\cdot)$ can be identified up to scale, without even knowing the whole demand model, and (2) the rest of the demand model can be identified nonparametrically.

Theorem 4. *Assume that M_t is continuously distributed, and is independent of (ξ_t, X_t). Assume that $s(m)$ is differentiable in m. Then $s(m) = e^{\int g(m)}\tilde{c}$ is identified up to a constant $\tilde{c}$, where $g(m) = \partial E(\ln(N_{jt}) \mid m)/\partial m$.*

To illustrate how we gain identification of γ from outside of the demand model, I first consider a market size model of the form $M_t^{\gamma_2}$. Taking log on both sides of the demand equations, we have $\ln(N_{jt}) = \gamma_2 \ln(M_t) + \ln(\pi_j(\delta_t, X_t^{(2)}))$. Given that M_t is independent of ξ_t, X_t, and thus independent of δ_t and $X_t^{(2)}$, we have the following conditional expectation

$$E\left(\ln(N_{jt}) \mid M_t\right) = \gamma_2 \ln(M_t) + E\left(\ln\left(\pi_j(\delta_t, X_t^{(2)})\right)\right),$$

from which we can identify γ_2 by construction: that is, $\gamma_2 = \partial E\left(\ln(N_{jt}) \mid M_t = m\right)/\partial \ln(m)$. When taking derivative with respect to $\ln(M_t)$, the demand function term π_j drops out because of the assumption that the market size measure M_t is exogenous. It means that if the observed N_{jt} increased more than double as we double M_t, the true market size must be growing at an increasing rate in M_t. Moreover, we can use these estimates to test the specification of the market size model, e.g., testing if a linear model of market size holds, without estimating the whole BLP model.

A second example is when the true market size takes the form of $M_{1t} + \gamma_1 M_{2t}$. Under the same assumption that M_t is independent of ξ_t and X_t, we can identify γ_1 by $\gamma_1 = \left(\partial E(\ln(N_{jt}) \mid M_{1t} = m_1, M_{2t} = m_2)/\partial m_2\right) / \left(\partial E(\ln(N_{jt}) \mid M_{1t} = m_1, M_{2t} = m_2)/\partial m_1\right)$.

After establishing point identification of market size, the empirical shares on the left hand side of equation (1.17) are identified. It would suffice to impose assumptions made in Berry and Haile (2014) to obtain nonparametric identification of the demand model.

1.6 Empirical Application: A Merger Analysis

Market size plays a crucial role in merger analysis. The analysis of unilateral effects hinges on whether an increase in the price of one product will lead consumers to choose an alternative in the market; also important is whether the consumer will divert to an outside option. Throughout this section, I assume that firms are under a static Nash-Bertrand pricing game. As I show in Appendix 1.I, market shares (or market sizes) used in estimation not only affect estimates of marginal effects ($\beta's$) but also enter firms' first-order conditions for pricing. Thus, assumptions about market size can influence firms' markup and consumer surplus. The formal pricing conditions of the firm's problem are provided in Appendix 1.I.

Suppose there are two firms each producing a single product. According to Pakes (2017), the upward pricing pressure (UPP) of good 1 depends on the substitution between good 1 and good 2, as well as the markup of good 2. The size of the outside market matters for a firm's optimization problem and, therefore, has a substantial effect on the estimated markup. More generally, in mergers involving multiple firms and products, the strategic complements result in all market participants increasing their prices, which in turn generates substitution to the outside option.

Intuitively, if the market size used in estimation is larger than the true size, the diversion to the outside option tends to be overstated. In the case of a merger, overestimating the outside option diversion suggests that the merged firm would maintain a relatively low price to prevent consumers from switching to the outside alternative. For example, Weinberg and Hosken (2013) study the breakfast syrup and motor oil industries using a plain logit model and demonstrates that simulated price changes decrease as the potential market size increases.

In this section, I apply the proposed method to analyze the price effects of a hypothetical merger in the Carbonated Soft Drink (CSD) market. In Appendix 1.K, I have a second merger analysis in the Ready-to-Eat Cereal market showing that our method works in different empirical contexts.

1.6.1 Carbonated Soft Drink (CSD) Market

The soft drink market has received significant attention in the literature, primarily driven by health and regulatory concerns. The conventional discrete choice model remains a widely used approach in modeling consumer purchasing behavior in this field of research.

The soft drink market is suitable for this study due to three key factors. First, the existing literature lacks a consensus on how to define market size. It is measured either by multiplying the population by the potential maximum soft drinks consumption (Eizenberg and Salvo 2015a assumed the constant to be six liters per week), or by multiplying the population by the per capita consumption of non-alcoholic beverages (as in Lopez and Fantuzzi 2012a, Liu, Lopez, and Zhu 2014; Lopez, Liu, and Zhu 2015; Liu and Lopez 2016; and Zheng, Huang, and Ross 2019). In the former case, the maximum weekly consumption can only be justified by considering consumer behavior, while for the latter case, it is not obvious which non-alcoholic beverages should be regarded as outside alternatives[11].

Second, this industry is one where we generally believe the outside option is not too large. Our simulation findings suggest that the proposed method achieves stronger identification in cases where the true choice probability of the outside option is not excessively large. While one do not observe the true outside share ex-ante, goods with frequent purchases tend to have a relatively small outside market. To see why, consider an extreme scenario where the prices of all soft drink products drop to zero. Consumers who never consume soda will not suddenly enter the market, even if the products are free, whereas soda drinkers are already regular purchasers. Therefore, we would not anticipate a significant increase in total sales, indicating that the potential consumption in the market is not exceptionally large[12].

The third reason for applying our method to the soda market is the occurrence of several horizontal mergers in the soft drink industry in recent years. For example, in 2018, the Coca-Cola Company acquired Costa Coffee, and PepsiCo acquired SodaStream in the same year.

1.6.2 Data

I use a panel of weekly scanner data from NielsenIQ for our analysis. The NielsenIQ scanner data provides comprehensive information on prices, sales, and product attributes, including package size, flavor, and nutritional contents. The dataset covers 202 designated market areas (DMAs) in the US and spans 52 weeks, encompassing the period from January 2019 to December 2019. I aggregate the dataset from the retailer level to the market level. Consistent with the literature, I define a market as a combination of a specific DMA and

[11] When one uses individual purchase data, the analogous definitions of market size could be slightly different. For example, Marshall (2015a) assumes the choice of outside options occurs when a trip is completed without the purchase. Bonnet and Réquillart (2013a) assume a narrower outside option, which is observed choices of alternative beverages.

[12] In contrast, the airline market is an example where the outside market can be substantial, reaching as high as 99%. For instance, if all airline tickets become free, there would likely be a surge in demand for airline flights.

week, resulting in a total of 10504 DMA-week markets[13].

In addition to the NielsenIQ data, I augment the dataset with input price information, which serves as excluded price instruments. This includes raw sugar prices from the US Department of Agriculture, Economic Research Service; local wage from the U.S. Bureau of Labor Statistics; as well as electricity and fuel prices from the US Department of Energy, Energy Information Administration.

Following Eizenberg and Salvo (2015a), I aggregate flavors and products in different sized packages into 15 brand-groups, denoted as $j = 1, \cdots, 15$ (e.g., Coca-Cola Cherry 12-oz and Coca-Cola Original 16.9-oz are treated as the same brand). Following Dubé (2005a), I consider diet and regular drinks as separate brands due to their distinct target demographics and separate advertising and promotion strategies within the industry. These brand categories include 11 brands owned by the three leading companies. The 12th and 13th brand categories represent aggregate private label (PL) brands for regular and diet drinks, respectively. To account for numerous niche brands (each with a volume share below 1 percent), I aggregate them into the 14th and 15th brand categories for regular and diet drinks, respectively. By doing so, I implicitly assume that product differentiation among these small brands is not of importance in the context of our study. I limit the sample to soft drinks sold in package types that have substantial sales, specifically including the 12-pack of 12-oz cans, 67.6-oz bottle, 6-pack of 16.9-oz bottles, 20-oz bottle, and 8-pack of 12-oz cans. These five package sizes dominate in terms of volume sales compared to other package types.

Table 1.1 shows volume shares of the carbonated soft drink category for each firm averaged across DMAs. These shares represent the volume sold of brands produced by a specific manufacturer divided by the total volume sold in the entire carbonated soft drink category. The brands from the largest manufacturer hold a share of 35.07 percent.

[13] I drop markets with extremely large or small sales relative to their respective populations, leaving us with $9,658$ markets.

1.6.3 Demand Model

As in section 1.4, the indirect utility of consumer i in market t from consuming brand j is given by

$$U_{ijt} = \delta_{jt} + \sigma \nu_i P_{jt} + \varepsilon_{ijt}.$$

The term δ_{jt} denotes a market-specific, individual-invariant mean utility from brand j: $\delta_{jt} = X'_{jt}\beta + \alpha P_{jt} + \xi_{jt}$. The vector X_{jt} includes in-store presence, brand fixed effects, seasonal effects and region fixed effects. In-store presence is measured by the proportion of stores within a market that carry a particular brand. Brand fixed effects capture the time invariant unobserved product characteristics, while seasonal effects capture temporal demand fluctuations. P_{jt} represents the price of brand j, and ξ_{jt} denotes demand shocks specific to a brand-market combination, observable to consumers but unobservable to the econometrician. The second term $\sigma \nu_i P_{jt}$ introduces consumer heterogeneity. ν_i follows a standard normal distribution. Finally, the utility function includes the term ε_{ijt}, representing consumer and brand-specific shocks that follow the Extreme Value Type I distribution and are iid across consumers, brands, and markets[14].

One issue is that in-store presence could be endogenous due to correlation with the unobservables ξ_{jt}. I address this potential endogeneity concern by flexibly controlling for brand-, quarter- and region-specific fixed effects. With a rich set of fixed effects included, the unobservables that remain are brand-region specific demand shocks that vary by time. I assume retailers or firms do not observe these demand shocks when making product assortment decisions. It is worth noting that in-store presence has been used as an exogenous covariate in previous studies such as Eizenberg and Salvo (2015a). Similarly, in the airline industry, *carrier presence* is often considered as an exogenous attribute. The economic interpretation of in-store presence in the present context aligns closely with carrier presence

[14]One thing worth noting is that because each consumer i can appear more than once in a week, the assumption that ε_{ijt} is independent across i might be violated. However, assuming independence is standard in the literature, and we think random coefficients partly account for correlation for a consumer. Therefore, in this analysis, I will not deal with correlation in ε.

in the airline market. Just as carrier presence may raise concerns of endogeneity, it has typically been addressed through via fixed effects.

Table 1.2 provides summary statistics for prices and in-store presence in the dataset. The prices and in-store presence are averaged across all UPCs within each brand, weighted by the volume sales of UPCs. The last three columns of Table 1.2 show the percentage of variance explained by brand, DMA, and month dummy variables. The results indicate that a majority of the variation in prices and in-store presence is attributed to differences between brands. After accounting for this brand-level variation, the remaining variation is primarily driven by disparities across geographic areas.

1.6.4 Market Size Definition

I define one serving of soft drink as 12 ounces. In calculating the market share of the outside good, Eizenberg and Salvo (2015a) assume a potential weekly consumption of 6 liters (approximately 17 servings) per household. Similarly, Zheng, Huang, and Ross (2019) use as γ the documented average per capita consumption of non-alcoholic beverages, including CSDs, water, juice, tea and sports drinks. The average consumption is around 30 ounces per person per day, equivalent to 17.5 servings per week. Other studies, such as Lopez and Fantuzzi (2012a), Liu, Lopez, and Zhu (2014), Lopez, Liu, and Zhu (2015), and Liu and Lopez (2016), also utilize per capita consumption of non-alcoholic beverages as a proxy for market size. The specific proportional factor varies depending on the inclusion of different beverages as outside options. For example, Liu, Lopez, and Zhu (2014) include milk consumption, while Zheng, Huang, and Ross (2019) do not. The per capita weekly consumption of non-alcoholic beverages in Liu, Lopez, and Zhu (2014) reaches as high as 32 servings, nearly double the amount used in Zheng, Huang, and Ross (2019).

These choices of market size are somewhat subjective. Eizenberg and Salvo (2015a) have shown that their results are not qualitatively sensitive to the market size assumption. However, in alternative counterfactual exercises like merger simulations, the market size assumption could play a more substantial role. It is important to note that Eizenberg and Salvo (2015a) use the scanner data from Brazil, while the study in this paper employs US data. Therefore, the assumed value of $\gamma = 17$ may be more appropriate for their dataset[15].

The market size assumptions can be expressed in our notation as γM_t, where M_t represents the total population in a DMA area. Throughout this section, all comparisons will be made with regard to assuming $\gamma = 17$ servings[16]. Specifically, I estimate γ along with other demand parameters and calculate elasticities and diversion ratios. I then simulate the merger using two potential market sizes: one assumes a market size of 17 servings per week, and the other assumes a market size of $\hat{\gamma}$ servings per week.

1.6.5 Instruments

To address the likely correlation of the demand errors ξ_{jt} with prices, as well as identify the random coefficients and market size parameters, I employ three sets of instruments. The first two sets are standard excluded instruments suggested by Berry and Haile (2014) and have been widely used in empirical studies (e.g. Eizenberg and Salvo 2015a; Petrin and Train 2010; and Nevo 2001).

The first set of price instruments belongs to the *Hausman-type* instrument, proposed by Hausman, Leonard, and Zona (1994). Specifically, the instrument for the price of brand j in a given DMA is the average price of this brand in other DMAs belonging to the same Census Region. These instruments provide variation across brands and DMAs, and are valid due to the correlation of prices across geographic regions through a common cost structure. However, the Hausman-type instruments could be problematic if demand unobservables are correlated across markets (e.g., launching a national campaign). To lessen this concern, I control for DMA-specific, brand-level in-store presence, which partially absorbs

[15] Another distinction between Eizenberg and Salvo (2015a) and other papers that use the US data is that the market size in Eizenberg and Salvo (2015a) is calculated based on the number of households, whereas others use the population. Here, I adopt the population measure. A potential market size of 17 servings per household is smaller than 17 servings per capita.

[16] I use 17 servings per week only as a baseline level to be compared to. It could be any other numbers.

common demand shocks.

The second class of price instruments consists of cost shifters. Specifically, I use input prices such as electricity prices, fuel prices and local wages. These cost shifters are excluded from the demand equation but affect prices through the supply side.

The third set of instruments serves to identify random coefficients and market size parameters. Here I use the traditional BLP type instruments. Specifically, they involve sums over exogenous characteristics of brands produced by the same company and sums over rival brands. I construct this class of instruments based on in-store presence and fitted values of prices. The fitted values of prices are obtained by regressing prices on X_{jt} and excluded price instrument. The projection of prices on exogenous variables would be mean independent of the unobservables ξ_{jt}. This exogenous variation in price facilitates the identification of the parameters associated with heterogeneity in price sensitivity. As a robustness check, I also use the differentiation instruments proposed by Gandhi and Houde (2019).

To see why the constructed instruments (based on in-store presence[17]) identify market size, consider a scenario where the in-store presence of 7Up increases. This change, possibly due to supply side factors like reduced transportation costs or the establishment of a new distribution hub, results in consumers encountering 7Up more frequently on store shelves. With this change in the physical environment of retail stores, one would observe consumers switching from alternative drinks and outside option to 7Up. Assuming all other factors remain constant, if we observe a substantial decrease in Sprite sales without an increase in overall soda consumption, it suggests a small potential market size, because little changes are from the extensive margin.

[17] If stores make assortment decisions after the realization of all demand shocks (as assumed in Ciliberto, Murry, and Tamer 2021), fixed effects may not fully address the endogeneity of in-store presence. As an alternative, though not explored in this paper, one can use exogenous changes in soda taxes as instruments.

1.6.6 Results

Table 1.3 reports five sets of demand model estimates. The first two columns correspond to plain logit and random coefficients logit models, where γ is estimated along with other demand parameters. Columns 3 to 5 are standard BLP estimates assuming $\gamma = 17$. Column 3 replicates the specification of column 2, while column 4 introduces an additional random coefficient on the constant term to capture unobserved preferences for the outside option. In column 5, DMA-week specific fixed effects are included. The strength of instruments, measured by the F-statistic of an IIA-test (as discussed in section 1.4.3), is 2819 with a p-value of 0.00, rejecting the null hypothesis of weak instruments.

The estimated values of γ are 12.478 and 11.767 for the plain logit and random coefficients logit models, respectively[18]. These estimates are lower than the range assumed in the literature (between 17.5 and 32), suggesting that a market size defined based on per capita consumption of all non-alcoholic beverages may be too large. It implies that not all beverage categories should be considered as outside alternatives to soda[19].

In columns 1 and 2 of Table 1.3, the estimated price sensitivities are -8.748 and -9.86. The estimate of random coefficient parameter σ in column 2 is 1.952 and is statistically significant, indicating a rejection of the plain logit model. Column 3, assuming $\gamma = 17$, exhibits higher price sensitivity (-13.033) and a larger standard deviation (4.395) in the preference for price. This aligns with what one would expect when assuming a larger potential market size. Column 4, which includes a second random coefficient on the constant term, produces estimates comparable to column 3. The estimate of σ for the constant term is small in magnitude -0.09 and statistically insignificant. In the last column, with mar-

[18] To verify that the estimated γ achieves global minimum for the random coefficients logit model, in Appendix 1.J I plot the GMM objective function over a grid of values for γ. The figure suggests that there are no multiple minima within the specified interval. However, the function is not steep around the minimum, which could pose challenges for numerical optimization.

[19] In 2019, the soft drink consumption per person per week in the US is approximately 107 ounces, or 8.9 servings. See: `https://www.ibisworld.com/us/bed/per-capita-soft-drink-consumption/1786/`. This reassures that our estimated value of potential consumption, which amounts to 12 servings, is reasonable.

ket fixed effects, the estimate of price sensitivity is much lower. Precisely estimating σ becomes challenging, with extremely large standard errors, which is expected due to the inclusion of near $10,000$ dummy variables in the GMM estimation. Therefore, there is limited exogenous variation to identify the random coefficient.

Table 1.4 provides estimated own-price elasticities and outside-good diversion ratios. Column 1 reports the elasticities based on our estimate of $\hat{\gamma} = 12$. The own-price elasticities range from -3.651 to -1.887, which is consistent with previous literature[20]. Note that PLs have lower own-price elasticities compared to other brands. This can be attributed to PLs being composite brands consisting of numerous niche products. The demand for an entire category are expected be less elastic than for each individual product. Furthermore, Steiner (2004) and Hirsch, Tiboldo, and Lopez (2018), find that PLs face relative inelastic demand due to limited interbrand substitution within a store. The outside-good diversion ratios exceed 60% for all brands, with PLs exhibiting the highest diversion ratio. This indicates that when faced with a price increase, iconsumers are more likely to cease purchasing rather than switch to branded alternatives, which is what one would expect to see if there exists a high degree of store loyalty.

The remaining columns in Table 1.4 are based on estimates from columns 3 to 5 of Table 1.3. Assuming $\gamma = 17$ when the true value is $\gamma = 12$, the biases in own-price elasticities are small. However, the biases in outside diversion ratios are more substantial, with a difference of 9 percentage points for PLs and approximately 3 to 4 percentage points for other brands, indicating even less substitutions across brands. Including a second random coefficient on the constant term yields results similar to those in column 2. This is mainly due to the fact that the estimated σ for the constant term is not significantly different from zero. The inclusion of market fixed effects leads to slightly lower own-price elasticities and higher outside diversion ratios. Although the results with market fixed effects are comparable to

[20]For example, the estimated own-price elasticities in Dubé (2005a) are in the range of -3 to -6. Lopez, Liu, and Zhu (2015) report elasticities between -1 and -2. The magnitude of elasticities varies with the aggregation level of product.

our estimates, the standard error of the random coefficient estimate is so large that we can not conclude any statistically significant results. The key takeaway from Table 1.4 is that none of the commonly employed solutions produce elasticities and diversion ratios close to those obtained using our estimated market size. Additionally, I provide estimates of aggregate elasticities in Appendix 1.J, which allow one to assess the impact of hypothetical soda taxes.

Finally, I simulate a merger between the largest manufacturer and private label manufacturers. The merger simulation abstracts away from cost reduction, or changes in the model of competition (e.g. coordination between other firms). Table 1.5 shows the percentage change in prices for the merging products. In column 1, the estimates (approximately 2.22% to 8.41% price increases) are reasonably comparable to those of Dubé (2005a), who estimated the price effect after a simulated merger between two leading manufacturers. The merger simulations predict larger price increases for the PLs than products of the leading manufacturer. This results from the relatively lower own-price elasticities of PLs, and is consistent with previous findings on higher pricing margins for PLs.

In columns 2 and 3, which assume $\gamma = 17$, the price effects of the merger for brands owned by the merging parties tend to be underestimated. The bias is the most pronounced for PLs. Simulated price increases are approximately 8 percent when the market size parameter is estimated to be 12, while assuming $\gamma = 17$ yields a price increase of 5.5 percent, biased by 31%. For brands from the leading manufacturer, the simulated price effects are relatively lower with the assumed $\gamma = 12$, although I acknowledge that the differences are not economically significant. In the last column, the estimate is relatively closer to our estimates but is imprecisely estimated with large standard errors.

In summary, both the diversion ratios and merger simulations generated by different market sizes vary and may lead to different policy evaluations. As the potential market size increases, the simulated price changes display a monotonic decrease.

1.7 Additional Results

The online supplemental appendix to this paper contains additional theoretical results, another empirical application, proofs of Theorems, and an extensive set of Monte Carlo experiments.

Some additional technical results include deriving the direction of bias, adding errors to the market size specification, identifying market size in a nested logit model, analyses of model identification with market fixed effects, and identification with a Bernoulli distributed random coefficient. There are also extra results for the CSD application, including price elasticities of market demand, which is useful in evaluating a simulated soda tax. The appendix also presents a second empirical analysis in the ready-to-eat cereal market to verify the method's applicability to different empirical contexts.

Three Monte Carlo experiments are conducted. The first evaluates whether random coefficients remove bias induced by incorrect market size assumptions. The second explores how sensitive parameter estimates and elasticities are to market size assumptions in a random coefficients logit model. The third experiment assesses the performance of our proposed method. Simulation results suggest that our estimator works well, particularly when the true outside good share is not too large.

1.8 Conclusions

This paper shows that market size is point identified in aggregate discrete choice demand models. Point identification relies on observed substitution patterns induced by exogenous variation in product characteristics and the nonlinearity of the demand model. The required data are conventional market-level data used in standard BLP estimation. I illustrate the results using Monte Carlo simulations and provide an empirical application to merger analysis in the soft drink industry. Our application shows that correctly measuring market size is economically important. For instance, I find that assuming a market size larger than the true size leads to a non-negligible downward bias in the estimated merger price increase, which could affect the conclusions of the merger evaluation. Apart from the merger application, my results would also have important implications for social welfare, markup calculations, tax and subsidy policies, and the entry of new firms.

Potential areas for future theoretical research include deriving conditions for strong identification and instrument selection, extending the model to micro-BLP which uses individual choice data, and allowing for dependence among logit errors to make the results applicable to panel data settings as in Khan, Ouyang, and Tamer (2021).

In the application, I consider a scalar γ. A possible extension would be to allow γ to vary based on market characteristics, such as demographic composition and the number of retail stores. It would also be useful to test my model in an industry where the true market size is known, such as the pharmaceutical market, where researchers generally know the number of patients, which can be considered as the potential market size. Another possibility for further work is generalizing the model to empirical contexts where inside good quantity rather than outside option is mismeasured or unknown, such as the consumption of informal goods or services (Pissarides and Weber 1989).

Acknowledgments

Researcher(s)' own analyses calculated (or derived) based in part on data from Nielsen Consumer LLC and marketing databases provided through the NielsenIQ Datasets at the Kilts Center for Marketing Data Center at The University of Chicago Booth School of Business. The conclusions drawn from the NielsenIQ data are those of the researcher(s) and do not reflect the views of NielsenIQ. NielsenIQ is not responsible for, had no role in, and was not involved in analyzing and preparing the results reported herein. This work received support from the Institute for Humane Studies under grant no. IHS017122.

Tables

Table 1.1: Manufacturer-Level Volume Shares of Carbonated Soft Drink

	Regular (%)	Diet (%)	Total (%)
Manufacturer A	22.19	12.88	35.07
Manufacturer B	12.25	6.87	19.12
Manufacturer C	7.17	2.7	9.87
Private Label	5.09	5.44	10.53
Others	13.04	12.36	25.4

Notes: Volume shares are the volume sold of a specific manufacturer divided by the total volume sold of the carbonated soft drink category.

Table 1.2: Prices and In-store Presence of Brands in Sample

	Mean	Median	Std	Min	Max	Brand Variation	DMA Variation	Month Variation
Prices ($ per 12 oz.)	0.40	0.39	0.12	0.11	2.75	39.73%	39.50%	0.50%
In-store Presence	0.50	0.51	0.22	0.01	1.00	75.12%	13.44%	0.06%

Notes: Variance contribution of brands, DMAs and months is the R-squared value added by each variable when it is added to the regression of price (or in-store presence) on the other two independent variables. In-store presence: the proportion of stores with the given brand in stock.

Table 1.3: Baseline Demand Estimation Results

	Estimate γ		Assume $\gamma = 17$ servings		
	Plain Logit	RC Logit	RC Logit	RC Logit with two RC's	RC Logit with Market FE
Means β					
Price	-8.748	-9.860	-13.033	-12.793	-5.245
	(0.084)	(0.222)	(0.289)	(0.434)	(0.311)
In-store Presence	3.281	3.311	3.309	3.314	5.061
	(0.022)	(0.022)	(0.023)	(0.024)	(0.019)
Standard Deviations σ					
Price		1.952	4.395	4.257	0.007
		(0.211)	(0.155)	(0.247)	(53.834)
Constant				-0.090	
				(1.189)	
Market Size Parameter					
γ	12.478	11.767			
	(0.263)	(0.210)			
Product Fixed Effects	Yes	Yes	Yes	Yes	Yes
Seasonal Effects	Yes	Yes	Yes	Yes	No
Region Fixed Effects	Yes	Yes	Yes	Yes	No
DMA-Week (Market) Fixed Effects	No	No	No	No	Yes

Notes: This table reports demand model estimates. Columns 1 and 2 correspond to plain logit and random coefficients logit models, and γ is to be estimated. Columns 3 to 5 are standard BLP estimates assuming $\gamma = 17$. Column 3 replicates the specification of column 2. Column 4 introduces an additional random coefficient on the constant term and column 5 includes market fixed effects. Standard errors in parentheses. Constant terms are omitted due to collinearity with product fixed effects.

Table 1.4: Demand Elasticities and Diversion Ratios

	RC Logit with $\hat{\gamma} = 12$	RC Logit Assuming $\gamma = 17$	RC Logit with two RC's Assuming $\gamma = 17$	RC Logit with Market FE Assuming $\gamma = 17$
Own-Price Elasticities				
Product 1	-3.398	-3.362	-3.351	-2.097
Product 2	-3.597	-3.493	-3.482	-2.224
Product 3	-3.651	-3.528	-3.518	-2.262
Private Label R	-1.887	-2.181	-2.151	-1.000
Outside-Good Diversion Ratios				
Product 1	62.8%	66.0%	66.5%	78.5%
Product 2	60.3%	63.0%	63.5%	77.2%
Product 3	59.8%	62.4%	62.9%	77.0%
Private Label R	68.4%	77.7%	77.7%	76.9%

Notes: This table reports estimates of elasticities and diversion ratio. Columns 1 is based on a random coefficients logit model with estimated γ. Columns 2 to 4 assume $\gamma = 17$. Column 2 replicates the specification of column 1. Column 3 introduces an additional random coefficient on the constant term and column 4 includes market fixed effects. To save space, only top-3 regular drink products are reported in the table. R represents regular.

Table 1.5: Simulated Percentage Price Effects for Merging Firms' Brands

	RC Logit with $\hat{\gamma} = 12$	RC Logit Assuming $\gamma = 17$	RC Logit with two RC's Assuming $\gamma = 17$	RC Logit with Market FE Assuming $\gamma = 17$
Manufacturer A Products	2.33	1.65	1.65	2.80
	2.37	1.66	1.67	2.85
	2.22	1.58	1.58	2.66
	2.49	1.73	1.73	3.01
Private Label R	8.41	5.64	5.66	10.14
Private Label DT	8.21	5.56	5.57	9.83

Notes: This table reports the percentage price change after a simulated merger between Manufacturer A and private label manufacturers. Columns 1 is based on a random coefficients logit model with estimated γ. Columns 2 to 4 assume $\gamma = 17$. Column 2 replicates the specification of column 1. Column 3 introduces an additional random coefficient on the constant term and column 4 includes market fixed effects. To save space, only merging firms' brands are reported in the table. R represents regular. DT stands for diet.

Appendix

1.A Proofs

Proof of Theorem 1. By the mean independence condition given in Assumption 1, we have

$$E\left(\ln\left(r_{jt}\right) \mid Q_t = q, X_{jt} = x\right) = E\left(\ln\left(\gamma W_t - 1\right) \mid Q_t = q, X_{jt} = x\right) - x'\beta.$$

Taking derivative with respect to q yields

$$0 = \frac{\partial E\left(\ln\left(r_{jt}\right) - \ln\left(\gamma W_t - 1\right) \mid Q_t = q, X_{jt} = x\right)}{\partial q}.$$

Let Γ be the set of all possible values of γ. For any given constant $c \in \Gamma$, define the function

$$g(c,q,x) = \frac{\partial E\left(\ln\left(r_{jt}\right) - \ln\left(cW_t - 1\right) \mid Q_t = q, X_{jt} = x\right)}{\partial q}$$

We observe r_{jt}, W_t, Q_t and X_{jt}. For any constant c, observed q and x, we can therefore nonparametrically identify $g(c,q,x)$. In order to show point identification, we need to verify that there exists at most one value of $c \in \Gamma$ such that $g(c,q,x) = 0$ for all observed $q \in \text{Supp}(Q_t)$ and $x \in \text{Supp}(X_{jt})$. Taking the derivative of $g(c,q,x)$ with respect to c, we have

$$\frac{\partial^2 E\left(\ln\left(r_{jt}\right) - \ln\left(cW_t - 1\right) \mid Q_t = q, X_{jt} = x\right)}{\partial c \partial q} = \frac{\partial E\left(-\frac{W_t}{cW_t - 1} \mid Q_t = q, X_{jt} = x\right)}{\partial q}.$$

The identification then follows from the assumption that there exists (q,x) on the support of (Q_t, X_{jt}) such that $\partial E\left(-\frac{W_t}{cW_t-1} \mid Q_t = q, X_{jt} = x\right) / \partial q$ is strictly positive or strictly negative for all $c \in \Gamma$.

Given γ, the model becomes equivalent to a standard multinomial choice model, and therefore β is identified the same way. □

Lemma 2 is the contraction mapping theorem in the appendix from Berry, Levinsohn, and Pakes (1995).

Lemma 2. *Consider the metric space* $(\mathbb{R}^J, d)$ *with* $d(x,y) = \|x-y\|$. *Let* $g : \mathbb{R}^J \to \mathbb{R}^J$ *have the properties:*

(1) $\forall \delta \in \mathbb{R}^J$, $f(\delta)$ *is continuously differentiable, with,* $\forall k$ *and* j,
$$\frac{\partial g_k(\delta)}{\partial \delta_j} \geq 0,$$
and
$$\sum_{j=1}^{J} \frac{\partial g_k(\delta)}{\partial \delta_j} < 1.$$

(2) $\min_j \inf_\delta g_j(\delta) = \underline{\delta} > -\infty$.*(There is a lower bound to* $g_j(\delta)$, *denoted* $\underline{\delta}$*)*

(3) There is a value $\overline{\delta}$, *with the property that if for any* j, $\delta_j \geq \overline{\delta}$, *then for some* k, $g_k(\delta) < \delta_k$.

Then, there is a unique fixed point δ^* *to* g *in* $\mathbb{R}^J$.

Proof of Proposition 1. The implicit system of equations is solved for each market, therefore we drop the t subscript in the proof to simplify the notation. We show the proposition for a scalar γ. Let $s_j = N_j/M$ and $s_0 = 1 - \sum_j N_j/M$. We obtain the generalized proposition by replacing $\ln(s_j/\gamma)$ with $\ln(N_j/\sum \gamma_1 M^{\gamma_2})$ Now we show that the function $g(\delta) = \delta + \ln(s) - \ln(\gamma) - \ln(\pi(\delta;\sigma))$ satisfies the three conditions in Lemma 2.

(1) The function $g(\delta)$ is continuously differentiable by the differentiability of the predicted choice probability function $\pi(\delta;\sigma)$.

First we want to show that
$$\frac{\partial g_j(\delta)}{\partial \delta_j} = 1 - \frac{1}{\pi_j(\delta;\sigma)} \frac{\partial \pi_j(\delta;\sigma)}{\partial \delta_j} \geq 0$$

Take the derivative of $\pi_j(\delta;\sigma)$ with respect to δ_j, we have

$$
\begin{aligned}
& \frac{\partial \pi_j(\delta;\sigma)}{\partial \delta_j} \\
= \; & \int \frac{exp(\delta_j + \sum_l \sigma_l x_{jl} \nu_{il}) \left(1 + \sum_{k=1}^{J_t} exp(\delta_k + \sum_l \sigma_l x_{kl} \nu_{il})\right)}{\left(1 + \sum_{k=1}^{J_t} exp(\delta_k + \sum_l \sigma_l x_{kl} \nu_{il})\right)^2} \\
& - \frac{(exp(\delta_j + \sum_l \sigma_l x_{jl} \nu_{il}))^2}{\left(1 + \sum_{k=1}^{J_t} exp(\delta_k + \sum_l \sigma_l x_{kl} \nu_{il})\right)^2} f_\nu(\nu) d\nu \\
= \; & \int \frac{exp(\delta_j + \sum_l \sigma_l x_{jl} \nu_{il})}{1 + \sum_{k=1}^{J_t} exp(\delta_k + \sum_l \sigma_l x_{kl} \nu_{il})} - \left(\frac{exp(\delta_j + \sum_l \sigma_l x_{jl} \nu_{il})}{1 + \sum_{k=1}^{J_t} exp(\delta_k + \sum_l \sigma_l x_{kl} \nu_{il})}\right)^2 f_\nu(\nu) d\nu \\
= \; & \pi_j(\delta;\sigma) - \int \left(\frac{exp(\delta_j + \sum_l \sigma_l x_{jl} \nu_{il})}{1 + \sum_{k=1}^{J_t} exp(\delta_k + \sum_l \sigma_l x_{kl} \nu_{il})}\right)^2 f_\nu(\nu) d\nu
\end{aligned}
$$

Then we can rewrite the derivative of function $g_j(\delta)$ as

$$
\begin{aligned}
\frac{\partial g_j(\delta)}{\partial \delta_j} &= 1 - \frac{1}{\pi_j(\delta;\sigma)} \frac{\partial \pi_j(\delta;\sigma)}{\partial \delta_j} \\
&= \frac{1}{\pi_j(\delta;\sigma)} \int \left(\frac{exp(\delta_j + \sum_l \sigma_l x_{jl} \nu_{il})}{1 + \sum_{k=1}^{J_t} exp(\delta_k + \sum_l \sigma_l x_{kl} \nu_{il})}\right)^2 f_\nu(\nu) d\nu,
\end{aligned}
$$

which is non-negative because $\pi_j(\delta;\sigma)$ is strictly positive, and the integrand of the second term is continuous and strictly positive, hence the integral over any closed integral is strictly positive, so the same must hold over the entire real line.

Take the derivative of $\pi(\delta;\sigma)$ with respect to δ_j, we have

$$
\frac{\partial \pi_k(\delta;\sigma)}{\partial \delta_j} = -\int \frac{exp(\delta_k + \sum_l \sigma_l x_{kl} \nu_{il}) exp(\delta_j + \sum_l \sigma_l x_{jl} \nu_{il})}{\left(1 + \sum_{k=1}^{J_t} exp(\delta_k + \sum_l \sigma_l x_{kl} \nu_{il})\right)^2} f_\nu(\nu) d\nu.
$$

Therefore the derivative of $g_k(\delta)$ with respect to δ_j is

$$
\begin{aligned}
\frac{\partial g_k(\delta)}{\partial \delta_j} &= -\frac{1}{\pi_k(\delta;\sigma)} \frac{\partial \pi_k(\delta;\sigma)}{\partial \delta_j} \\
&= \frac{1}{\pi_k(\delta;\sigma)} \int \frac{exp(\delta_k + \sum_l \sigma_l x_{jl} \nu_{il}) exp(\delta_j + \sum_l \sigma_l x_{jl} \nu_{il})}{\left(1 + \sum_{k=1}^{J_t} exp(\delta_k + \sum_l \sigma_l x_{kl} \nu_{il})\right)^2} f_\nu(\nu) d\nu,
\end{aligned}
$$

which is non-negative because $\pi_k(\delta;\sigma)$ and the integrand of the second term are strictly positive.

To show the condition $\sum_{j=1}^{J} \partial g_k(\delta)/\partial\delta_j < 1$, note that increasing all the δ_j including δ_0 simultaneously will not change the market shares, implying that $\sum_{j=0}^{J} \partial\pi_k(\delta;\sigma)/\partial\delta_j = 0$. Then

$$\sum_{j=1}^{J} \frac{\partial\pi_k(\delta;\sigma)}{\partial\delta_j} = -\frac{\partial\pi_k(\delta;\sigma)}{\partial\delta_0} > 0$$

We can therefore establish the condition that the derivatives of g_k sums to less than one

$$\sum_{j=1}^{J} \frac{\partial g_k(\delta)}{\partial\delta_j} = 1 - \frac{1}{\pi_k(\delta;\sigma)} \sum_{j=1}^{J} \frac{\partial\pi_k(\delta;\sigma)}{\partial\delta_j} < 1.$$

(2) Rewrite $g_j(\delta)$ as

$$\begin{aligned} g_j(\delta) &= \ln(s_j) - \ln(\gamma) - \ln\left(D_j(\delta)\right), \\ \text{where } D_j(\delta) &= \int \frac{exp(\sum_l \sigma_l x_{jl} \nu_{il})}{1 + \sum_{k=1}^{J_t} exp(\delta_k + \sum_l \sigma_l x_{kl} \nu_{il})} f_\nu(\nu) d\nu. \end{aligned}$$

A lower bound of g_j can be obtained by letting all of δ_k go to $-\infty$, then $D_j(\delta) \to \int exp(\sum_l \sigma_l x_{jl} \nu_{il}) f_\nu(\nu) d\nu$. So the lower bound on $g_j(\delta)$ is

$$\underline{\delta} \equiv \ln(s_j) - \ln(\gamma) - \ln\left(\int exp(\sum_l \sigma_l x_{jl} \nu_{il}) f_\nu(\nu) d\nu\right)$$

(3) The proof of this part follows Berry (1994). He shows condition (3) of Lemma 2 is satisfied by first showing that if for any product j, $\delta_j \geq \overline{\delta}$, then there is at least one element k with $\pi_k(\delta;\sigma) > s_k/\gamma$.

To construct a $\overline{\delta}$ that satisfies the above requirement, first set all of δ_k (other than good j and outside good) to $-\infty$. Define $\overline{\delta}_j$ to be the value of δ_j that makes $\pi_0(\delta;\sigma) = 1 - (1 - s_0)/\gamma$. Then define $\overline{\delta} = \max_j \overline{\delta}_j$.

Now if there is any element of δ with $\delta_j > \overline{\delta}$, then $\pi_0(\delta;\sigma) < 1 - (1 - s_0)/\gamma$. It then follows from $\sum_{j=0}^{J} \pi_j(\delta;\sigma) = 1$ that $\sum_{j=1}^{J} \pi_j(\delta;\sigma) > \sum_{j=1}^{J} s_j/\gamma$. Thus there is at least one good k with $\pi_k(\delta;\sigma) > s_k/\gamma$, which implies $g_k(\delta) < \delta_k$:

$$\begin{aligned}
& \pi_k(\delta;\sigma) > \frac{s_k}{\gamma} \\
\iff \quad & \ln\left(\pi_k(\delta;\sigma)\right) > \ln(s_k) - \ln(\gamma) \\
\iff \quad & \ln(s_k) - \ln(\gamma) - \ln\left(\pi_k(\delta;\sigma)\right) < 0 \\
\iff \quad & g_k(\delta) = \delta_k + \ln(s_k) - \ln(\gamma) - \ln\left(\pi_k(\delta;\sigma)\right) < \delta_k
\end{aligned}$$

□

Proof of Theorem 2. Assuming enough regularity to take the derivative inside the expectation and applying the dominated convergence theorem, we have $\nabla_\theta E\left(h_{jt}(\theta)\right) = E\left(\nabla_\theta h_{jt}(\theta)\right)$. The Jacobian matrix is

$$\begin{aligned}
E\left(\nabla_\theta h_{jt}(\theta)\right) &= E\left[\frac{\partial h_{jt}(\theta)}{\partial \gamma'} \quad \frac{\partial h_{jt}(\theta)}{\partial \sigma'} \quad \frac{\partial h_{jt}(\theta)}{\partial \beta'}\right] \\
&= E\left[\phi_j(Z_t)\frac{\partial \delta_{jt}(N_t, M_t, X_t^{(2)};\gamma,\sigma)}{\partial \gamma'} \quad \phi_j(Z_t)\frac{\partial \delta_{jt}(N_t, M_t, X_t^{(2)};\gamma,\sigma)}{\partial \sigma'} \quad \phi_j(Z_t)X_{jt}'\right]
\end{aligned}$$

Recall that $h_{jt}(\theta) = (\delta_{jt}(N_t, M_t, X_t^{(2)};\gamma,\sigma) - X_{jt}'\beta)\phi_j(Z_t)$. The first derivative of the above matrix is an $m \times 2K$ vector. $\partial \pi_{jt}(\delta_t;\sigma)/\partial \sigma'$ is a $1 \times L$ row vector, so the second derivative of the above matrix is an $m \times L$ matrix. Similarly, the dimension of the last derivative is $m \times L$. The identification proof follows directly from Lemma 2 and the rank condition that the Jacobian matrix has rank K. □

Proof of Lemma 1. To ease notation in the proof, we drop the subscript j and t and suppress the dependence of Φ and Ψ on $(\delta_t, X_t^{(2)};\sigma)$, and the dependence of ϕ on Z. We make a simplifying assumption w.l.o.g.: Suppose X are exogenous and thus can serve as its own instruments, i.e. $\phi^{(1)} = X$. When γ is a scalar, the Jacobian matrix reduces to

$$\begin{pmatrix}
E\left(\begin{pmatrix}\phi^{(2)} \\ \phi^{(3)}\end{pmatrix}\begin{pmatrix}\frac{1}{\gamma}\Psi \\ \Phi\end{pmatrix}'\right) & E\left(\begin{pmatrix}\phi^{(2)} \\ \phi^{(3)}\end{pmatrix}X'\right) \\
E\left(X\begin{pmatrix}\frac{1}{\gamma}\Psi \\ \Phi\end{pmatrix}'\right) & E\left(XX'\right)
\end{pmatrix},$$

and recall that

$$A = E\left(\begin{pmatrix}\phi^{(2)}\\ \phi^{(3)}\end{pmatrix}\begin{pmatrix}\frac{1}{\gamma}\Psi\\ \Phi\end{pmatrix}'\right) \quad B = E\left(\begin{pmatrix}\phi^{(2)}\\ \phi^{(3)}\end{pmatrix}X'\right)$$

$$C = E\left(X\begin{pmatrix}\frac{1}{\gamma}\Psi\\ \Phi\end{pmatrix}'\right) \quad D = E\left(XX'\right),$$

Let $X = (1, \tilde{X}')'$. Denote $\Omega = (E(\tilde{X}\tilde{X}') - E(\tilde{X})E(\tilde{X}'))^{-1}$, then we have

$$D^{-1} = \begin{pmatrix} 1 + E(\tilde{X}')\Omega E(\tilde{X}) & -E(\tilde{X}')\Omega \\ -\Omega E(\tilde{X}) & \Omega \end{pmatrix},$$

and

$$A - BD^{-1}C = \frac{1}{\gamma}\left(Cov\left(\begin{pmatrix}\phi^{(2)}\\ \phi^{(3)}\end{pmatrix}, (\Psi,\Phi)\right) - Cov\left(\begin{pmatrix}\phi^{(2)}\\ \phi^{(3)}\end{pmatrix}, \tilde{X}'\right)\Omega Cov(\tilde{X}, (\Psi,\Phi))\right)$$

For the Jacobian matrix to have full rank, we make a technical assumption that $det(A - BD^{-1}C) \neq 0$. This assumption is generically satisfied when

$$Cov\left(\begin{pmatrix}\phi^{(2)}\\ \phi^{(3)}\end{pmatrix}, (\Psi,\Phi)\right)$$

has full rank. Note that given the regularity assumptions in the Lemma, when the above matrix has full rank, $det(A - BD^{-1}C)$ equals zero only at a set of measure zero. $\square$

Proof of Theorem 4. Assuming $M_t \perp (\xi_t, X_t)$, we take log and conditional expectation on both sides

$$E\left(\ln(N_{jt}) \mid M_t\right) = \ln(s(M_t)) + E\left(\ln\left(\pi_j(\delta_t, X_t^{(2)})\right)\right).$$

Take derivative w.r.t. m

$$\frac{\partial E\left(\ln(N_{jt}) \mid M_t = m\right)}{\partial m} = \frac{\partial \ln(s(M_t))}{\partial m} \equiv g(m),$$

from which $g(m)$ is identified. Then $\ln(s(M_t)) = \int g(m) + c$ is identified up to location. Thus,

$$s(m) = e^{\int g(m)}\tilde{c}$$

is identified up to scale.

$\square$

Proof of Theorem 5. By Assumption 4, the conditional mean function is

$$E\left(\ln\left(r_{jt}\right) \mid X_{jt}=x\right)=\kappa_t+x'\beta \quad \forall t\in(1,\cdots,T).$$

If X_{jt} is continuous, then $\partial E\left(\ln\left(r_{jt}\right) \mid X_{jt}=x\right)/\partial x=\beta$. If X_{jt} is discrete, then $E\left(\ln\left(r_{jt}\right) \mid X_{jt}=x_1\right)-E\left(\ln\left(r_{jt}\right) \mid X_{jt}=x_2\right)=(x_1-x_2)'\beta$. β is therefore identified given that the support of X_{jt} does not lie in a proper linear subspace of $\mathbb{R}^{\dim(X)}$ for $t=1,\cdots,T$ and X_{it} does not contain a constant.

Now that we have shown β is identified, the conditional mean function becomes

$$E\left(\ln\left(r_{jt}\right) \mid X_{jt}=x\right)-x'\beta=\kappa_t \quad \forall t\in(1,\cdots,T).$$

The left hand side is identified, and each of the T equations pins down a unique κ_t. Therefore $\left(\kappa_1,\cdots,\kappa_T\right)$ are identified. $\square$

Proof of Theorem 6. By the mean independence condition given in Assumption 1, we have

$$E\left(\ln\left(r_{jt}\right) \mid Q_t=q,X_{jt}=x\right)=\frac{1}{1-\sigma}E\left(\ln\left(\gamma W_t-1\right) \mid Q_t=q,X_{jt}=x\right)-x'\frac{\beta}{1-\sigma}.$$

Taking first-order derivative with respect to q yields

$$\frac{\partial E\left(\ln\left(r_{jt}\right) \mid Q_t=q,X_{jt}=x\right)}{\partial q}=\frac{1}{1-\sigma}\frac{\partial E\left(\ln\left(\gamma W_t-1\right) \mid Q_t=q,X_{jt}=x\right)}{\partial q}. \tag{1.18}$$

Taking second-order derivative with respect to q yields

$$\frac{\partial^2 E\left(\ln\left(r_{jt}\right) \mid Q_t=q,X_{jt}=x\right)}{\partial q^2}=\frac{1}{1-\sigma}\frac{\partial^2 E\left(\ln\left(\gamma W_t-1\right) \mid Q_t=q,X_{jt}=x\right)}{\partial q^2}. \tag{1.19}$$

Define functions

$$g(q,x)=\frac{\partial E\left(\ln\left(r_{jt}\right) \mid Q_t=q,X_{jt}=x\right)}{\partial q},$$

and

$$h(\gamma,q,x)=\frac{\partial E\left(\ln\left(\gamma W_t-1\right) \mid Q_t=q,X_{jt}=x\right)}{\partial q}.$$

Dividing equation (1.19) by (1.18) yields

$$\frac{\partial g(q,x)}{\partial q}\frac{1}{g(q,x)}=\frac{\partial h(\gamma,q,x)}{\partial q}\frac{1}{h(\gamma,q,x)}$$

Let Γ be the set of all possible values of γ. For any given $c \in \Gamma$, define function

$$f(c,q,x) = \frac{\partial h(c,q,x)}{\partial q} \frac{1}{h(c,q,x)} - \frac{\partial g(q,x)}{\partial q} \frac{1}{g(q,x)}.$$

We observe r_{jt}, W_t, Q_t and X_{jt}. For any constant c and observed q and x, we can therefore nonparametrically identify $f(c,q,x)$. In order to show point identification of γ, we need to verify that there exists at most one value of $c \in \Gamma$ such that $f(c_q,q,x) = 0$ for all observed $q \in \text{Supp}(Q_t)$ and $x \in \text{Supp}(X_{jt})$. Taking the derivative of $f(c,q,x)$ with respect to c, we have

$$\begin{aligned}
\frac{\partial f(c,q,x)}{\partial c} &= \frac{\partial^2 (h(c,q,x))}{\partial q \partial c} \frac{1}{h(c,q,x)} - \frac{\partial h(c,q,x)}{\partial q} \frac{h(c,q,x)}{\partial c} \frac{1}{h(c,q,x)^2} \\
&= \frac{1}{h(c,q,x)} \frac{\partial^2 E\left(\frac{W_t}{cW_t - 1} \mid Q_t = q, X_{jt} = x\right)}{\partial q^2} - \\
&\quad \frac{1}{h(c,q,x)^2} \frac{\partial^2 E\left(\ln(cW_t - 1) \mid Q_t = q, X_{jt} = x\right)}{\partial q^2} \frac{\partial E\left(\frac{W_t}{cW_t - 1} \mid Q_t = q, X_{jt} = x\right)}{\partial q}.
\end{aligned}$$

The identification of γ then follows from the assumption that there exists (q,x) on the support of (Q_t, X_{jt}) such that $\frac{\partial f(c,q,x)}{\partial c}$ is strictly positive or strictly negative for all $c \in \Gamma$.

Given a unique γ, and the assumption that $\frac{h(\gamma,q,x)}{g(q,x)} \neq 0$, we can solve for σ explicitly as

$$\sigma = 1 - \frac{h(\gamma,q,x)}{g(q,x)}.$$

Given γ and σ, the model reduces to a standard multinomial logit model, and $\beta/(1-\sigma)$ is identified in a linear regression model. Given $\beta/(1-\sigma)$ and σ, we can solve for β. $\square$

1.B Bias Caused by Mismeasured Market Size

I show that the usual approach that estimates demand based on equation (1.1) with a mismeasured market size will lead to biased estimates of β. To see this, suppose the true model is given by equation (1.5) with true value of $\gamma \neq 1$. Without loss of generality, let $s_{jt} = N_{jt}/M_t$ and $s_{0t} = (M_t - N_t^{total})/M_t$ denote the mismeasured market shares calculated based on the incorrect assumption that market size is $\tilde{\gamma} M_t$, with $\tilde{\gamma} = 1$. Define μ_{jt} to be the difference between the true choice probabilities $\ln(\pi_{jt}/\pi_{0t})$ and the mismeasured market shares $\ln(s_{jt}/s_{0t})$, so it gives the model that relates observed market shares to covariates and errors

$$\ln\left(\frac{s_{jt}}{s_{0t}}\right) = X'_{jt}\beta + \xi_{jt} + \mu_{jt},$$

with

$$\begin{aligned}\mu_{jt} &= \ln\left(\frac{s_{jt}}{s_{0t}}\right) - \ln\left(\frac{\pi_{jt}}{\pi_{0t}}\right) \\ &= \ln\left(\frac{\gamma W_t - 1}{W_t - 1}\right) \\ &= \ln\left(1 \Big/ \left(\frac{1}{\gamma} + \left(\frac{1}{\gamma} - 1\right)\frac{1-\pi_{0t}}{\pi_{0t}}\right)\right)\end{aligned}$$

by construction. The first equality is by the definition of μ_{jt}. The second equality follows from the definition of mismeasured market shares and equations (1.1) and (1.5). The third equality follows from equation (1.4). It is not reasonable to believe that π_{0t} would be independent of X_{jt} because by the model, π_{0t} depends on the characteristics of all goods. One possible technique to fix the problem is using a standard 2SLS regression or GMM with appropriate instruments. In this case, a valid instrument should be correlated with the demand covariates X_{jt}, and in the meanwhile, uncorrelated with π_{0t}, which again is a function of X_{jt}. In general, it is unlikely to construct an instrument that satisfies both restrictions.

Using the relationship provided above, we can predict the direction of the bias: Suppose that the observed market size is larger than the true size (i.e. $\gamma < 1$), the model predicts that the price of good j will be positively correlated with μ_{jt}, and negatively correlated with its own market share. Therefore, the estimate of the price coefficient will be biased downward (in absolute value), implying an underestimated price sensitivity.

1.C Extension of the Simple Logit Case

r_{jt} and r^*_{jt} are defined as in section 1.3. Now we assume

$$\ln(r_{jt}) = \ln(r^*_{jt}) + e_{jt}.$$

Here, e_{jt} is the error in $\ln(r_{jt})$ that we will later assume to have mean zero. It can include sampling errors, measurement errors, or aggregate unobserved heterogeneity in individual utility.

Then we assume that the mismeasurement in W_t relative to π_{0t} takes the form

$$\ln\left(\frac{\pi_{0t}}{1-\pi_{0t}}\right) = \ln(\gamma W_t - 1) + v_t$$

for some constant γ and some random mean zero noise v_t. I add the error term v_t to account for this relationship being approximate rather than exact. With the additional v_t, $1-\pi_{0t}$ would approximately equal $1/(\gamma W_t)$, and therefore $\ln(\pi_{0t}/(1-\pi_{0t}))$ would approximately equal $\ln(\gamma W_t - 1)$.

Putting the above equations and assumptions together we get the estimating equation

$$\ln(r_{jt}) = \ln(\gamma W_t - 1) + X_{jt}'\beta + u_{jt} \quad \forall j \in \mathscr{J}_t$$

where

$$u_{jt} = \xi_{jt} + e_{jt} + v_t.$$

To achieve identification as in section 1.3, we only need to modify the mean independence assumption such that $E(u_{jt} \mid Q_t, X_{1t}, \ldots, X_{J_t t}) = 0$, where everything else is defined as in section 1.3.

1.D Market Fixed Effects Approach for Simple Logit

Returning to equation (1.5), observe that the term with the unknown π_{0t} shows up additively, and it varies by market, not by product. I could allow for separate intercepts for each market to capture the unknown π_{0t}. The inclusion of the market level intercepts allows for unobserved aggregate market effects of the kind introduced by the presence of outside goods. Let $(\kappa_1, \cdots, \kappa_T)$ denote the aggregate market-varying and product-invariant parameters, then we can rewrite the model of equation (1.5) as

$$\ln(r_{jt}) \quad = \quad \kappa_t + X_{jt}'\beta + u_{jt} \quad \text{for each } t = 1, \cdots, T.$$

Assumption 4. *$E(u_{jt} \mid X_{jt}) = 0$ for all $t \in (1, \cdots, T)$. The support of X_{jt} does not lie in a proper linear subspace of $\mathbb{R}^L$.*

The conditional mean in Assumption 4 takes expectation across all products j for a fixed market t. Assumption 4 first assumes all X_{jt} are exogenous characteristics. Prices are taken to be exogenous throughout the context of the plain logit model for expositional purposes. I will relax this assumption in the next section. Assumption 4 also imposes no multicollinearity requirements on X_{jt}.

Theorem 5. *Let Assumption 4 hold. Let β^0 be the coefficient on the constant. Normalize $\beta^0 = 0$. Then $(\kappa_1, \cdots, \kappa_T, \beta)$ are identified.*

The proofs are in the appendix. Theorem 5 indicates that all parameters are identified except for the constant. This result has straightforward and important implications for how one can deal with the unobserved market size. In particular, when we observe data from a single market ($T = 1$), estimating κ_t resembles estimating the constant term. The desirable thing is that it would only bias the estimate of the constant in the consumer's indirect utility function and does not affect estimates of elasticities. For $T \geq 2$, when there are repeated measures of the same market/region over multiple time periods, or when we have cross-sectional data from more than one market/region, including market or time dummies in the model ensures consistent estimation of all parameters but the constant.

However, this method comes with some costs. First, it incurs efficiency loss because the data variation across markets is not exploited. In addition, the choice probabilities will not be identified because the true market size is not identified, which puts limitations on the study of, for example, diversions, mergers, and product entry or exit as these questions depend heavily on choice probabilities. Moreover, coefficients of market-level regressors will not be identified, so we cannot estimate marginal effects of any market characteristics. The biggest limitation is that this method relies on the functional form of the model specification. It works only in the plain logit model as a special case and cannot be generalized to the random coefficients demand model (see section 1.4.5).

1.E Identification of Market Size in Nested Logit Model

Following the nested logit framework in McFadden (1977) and Cardell (1997), we assume the utility of consumer i for product j belonging to group g is

$$U_{ijt} = \delta_{jt} + \zeta_{igt} + (1-\rho)\varepsilon_{ijt},$$

where $\delta_{jt} = X_{jt}'\beta + \xi_{jt}$ and ε_{ijt} is independently and identically distributed with extreme value type I distribution as before. The unobserved group specific taste ζ_{igt} follows a distribution such that $\zeta_{igt} + (1-\rho)\varepsilon_{ijt}$ is also distributed extreme value. ρ measures the correlation of unobserved utility among products in group g. A larger value of ρ indicates greater correlation within nest. When $\rho = 0$, the within group correlation of unobserved utility is zero, and the nested logit model degenerates to the plain multinomial logit model.

Berry (1994) shows that demand parameters β and ρ can be consistently estimated from a linear regression similar to the logit equation (1.1), with an additional regressor $\ln(\pi_{j|gt})$,

$$\ln(\pi_{jt}/\pi_{0t}) = X_{jt}'\beta + \rho\ln(\pi_{j|gt}) + \xi_{jt}, \tag{1.20}$$

where $\pi_{j|gt}$ is the conditional choice probability of product j given that a product in group g is chosen.

Consider the case where all goods are divided up into two nests, with the outside good as the only choice in group $g = 0$ and all inside goods belonging to group $g = 1$. In this case, $\pi_{j|gt} = r_{jt}^*$ for $j \neq 0$, where r_{jt}^* is defined in section 1.3.2. Then we can rewrite (1.20) as

$$\ln\left(r_{jt}^*\right) = \frac{1}{1-\rho}\ln\left(\frac{\pi_{0t}}{1-\pi_{0t}}\right) + X_{jt}'\frac{\beta}{1-\rho} + \frac{\xi_{jt}}{1-\rho}.$$

Following the same exposition of the market size model as in section 1.3.2, we assume equation (1.4) hold. Combining above equations and assumptions we get the estimating equation for the nested logit model

$$\ln\left(r_{jt}\right) = \frac{1}{1-\rho}\ln\left(\gamma W_t - 1\right) + X_{jt}'\frac{\beta}{1-\rho} + \frac{\xi_{jt}}{1-\rho}. \tag{1.21}$$

Theorem 6. *Given Assumption 1 and equation (1.21), let Γ be the set of all possible values of γ, if*

1. *all relevant first and second order derivatives exist,*

2. $\partial f(c,q,x)/\partial c > 0$ *or* < 0 *for all* $c \in \Gamma$, *where*

$$f(c,q,x) = \frac{\partial h(c,q,x)}{\partial q}\frac{1}{h(c,q,x)} - \frac{\partial g(q,x)}{\partial q}\frac{1}{g(q,x)},$$
$$g(q,x) = \frac{\partial E\left(\ln\left(r_{jt}\right) \mid Q_t = q, X_{jt} = x\right)}{\partial q},$$
$$h(c,q,x) = \frac{\partial E\left(\ln\left(cW_t - 1\right) \mid Q_t = q, X_{jt} = x\right)}{\partial q},$$

3. *and* $h(c,q,x) \neq 0$ *for all* $c \in \Gamma$.

Then γ, β and ρ are identified.

The proof of theorem 6 works by showing that there exists q and x such that $f(c,q,x) = 0$ has a unique solution of c. In practice, if Q_t is a scalar random variable, we can use Q_t and any nonlinear function of Q_t as instruments to estimate γ and ρ. Nonlinear functions of Q_t (e.g. $\sqrt{Q_t}$ or Q_t^2) will have additional explanatory power to separately identify γ and ρ.

I exploit the variation in W_t and Q_t, and the nonlinearity of the estimating equation to identify the model. Though theoretically we can distinguish γ and ρ, it can be seen from equation (1.21) that separately identifying the two parameters is hard without strong instruments. If $\gamma W_t - 1$ were close to zero or if the logarithm were not in the equation, ρ tends to be not identified. I can also see this from a first order Taylor expansion around $W_t = \overline{W}$ (White 1980), where $\overline{W}$ is the mean of W_t. The coefficient of the Taylor series depends on both γ and ρ. This result partly confirms the commonly held intuition that a nest structure can mitigate biases caused by unknown market size. A Monte Carlo simulation for the nested logit model is available upon request.

One might be concerned that the identification result of theorem 6 relies on the functional form assumption we made in equation (1.4). There might exist some different functional form assumption of market size which would make γ and ρ unidentified. For example, the model would be unidentified by letting the true market size be $(\exp(\gamma\tilde{W}_t) + 1)N_t^{total}$, for some variable $\tilde{W}_t$. In this case, equation (1.21) reduces to $\ln(r_{jt}) = 1/(1-\rho)\gamma\tilde{W}_t + X_{jt}'\beta/(1-\rho) + \xi_{jt}$. However, a market size model of this form is odd and lack of economic meaning.

1.F RCL with Bernoulli Distribution

Suppose $J = 1$. Consumers choose either purchasing or not purchasing (i.e., the outside good). Consumer i's purchasing decision is given by

$$Y_{it} = \mathbb{1}[\beta_{0i} + X_t\beta_{1i} + \xi_t + \varepsilon_{it} \geq 0],$$

where X_t is a scalar random variable, ε_{it} is standard logistically distributed, ξ_t are unobserved random errors, and (β_{0i}, β_{1i}) are two random coefficients with $\beta_{0i} = \beta_0 + \sigma_0\nu_i$ and $\beta_{1i} = \beta_1 + \sigma_1\nu_i$.

To get an analytic formula for the predicted market share, we assume that ν_i follows a Bernoulli distribution

$$\nu_i = \begin{cases} 0, & \text{with probability } \frac{1}{2} \\ 1, & \text{with probability } \frac{1}{2}. \end{cases}$$

Let $\delta_t = \beta_0 + X_t\beta_1 + \xi_t$. The overall true market share in market t is

$$\begin{aligned} \pi_t(\delta_t;\sigma) &= E\left[\frac{\exp(\beta_{0i} + X_t\beta_{1i} + \xi_t)}{1 + \exp(\beta_{0i} + X_t\beta_{1i} + \xi_t)} \mid X_t, \xi_t\right] \\ &= \frac{1}{2}\cdot\frac{\exp(\delta_t)}{1+\exp(\delta_t)} + \frac{1}{2}\cdot\frac{\exp(\delta_t + \sigma_0 + X_t\sigma_1)}{1+\exp(\delta_t + \sigma_0 + X_t\sigma_1)}, \end{aligned}$$

Now suppose that the true market size is γM_t, and the observed market share is $s_t = N_t^{total}/M_t$. Then the observed and true market share would be linked by $s_t = \gamma\pi_t$. Following BLP, we can implicitly solve for δ_t by equating $\frac{s_t}{\gamma} = \pi_t(\delta_t;\sigma)$.

Identification would be based on a set of conditional moment restrictions $E(\xi_t \mid Z_t) = 0$, where Z_t is a vector of instruments.

To simplify things and focus only on the constant term, suppose there were no X's, so

$$\pi_t(\delta_t;\sigma_0) = \frac{1}{2}\cdot\frac{\exp(\delta_t)}{1+\exp(\delta_t)} + \frac{1}{2}\cdot\frac{\exp(\delta_t+\sigma_0)}{1+\exp(\delta_t+\sigma_0)},$$

and $\delta_t = \beta_0 + \xi_t$. Assume that we have two instruments Z_{1t} and Z_{2t} satisfying

$$E\begin{bmatrix} \xi_t \\ \xi_t Z_{1t} \\ \xi_t Z_{2t} \end{bmatrix} = 0.$$

Since $\xi_t = \delta_t - \beta_0$, we can rewrite the above moment conditions as

$$E\begin{bmatrix} \delta_t - \beta_0 \\ (\delta_t-\beta_0)Z_{1t} \\ (\delta_t-\beta_0)Z_{2t} \end{bmatrix} = 0. \tag{1.22}$$

Note that δ_t is solved from the demand system, so it is a function of (σ_0,γ). For the unknown parameters $(\beta_0,\sigma_0,\gamma)$ to be (locally) point identified, we would need there to be a unique solution to the moment conditions (1.22). A sufficient condition is that the Jacobian matrix with respect to $(\beta_0,\sigma_0,\gamma)$ is non-singular.

Let $\pi_t^0 \equiv \exp(\delta_t)/(1+\exp(\delta_t))$ and $\pi_t^1 \equiv \exp(\delta_t+\sigma_0)/(1+\exp(\delta_t+\sigma_0))$. Let

$$g(\beta_0,\sigma_0,\gamma) = \begin{pmatrix} \delta_t - \beta_0 \\ (\delta_t-\beta_0)Z_{1t} \\ (\delta_t-\beta_0)Z_{2t} \end{pmatrix}$$

denote the 3×1 function. The Jacobian matrix would be

$$E\begin{bmatrix} \dfrac{-\pi_t^1(1-\pi_t^1)}{\pi_t^0(1-\pi_t^0)+\pi_t^1(1-\pi_t^1)} & \dfrac{1}{\gamma}\dfrac{-(\pi_t^0+\pi_t^1)}{\pi_t^0(1-\pi_t^0)+\pi_t^1(1-\pi_t^1)} & -1 \\ \dfrac{-\pi_t^1(1-\pi_t^1)}{\pi_t^0(1-\pi_t^0)+\pi_t^1(1-\pi_t^1)}Z_{1t} & \dfrac{1}{\gamma}\dfrac{-(\pi_t^0+\pi_t^1)}{\pi_t^0(1-\pi_t^0)+\pi_t^1(1-\pi_t^1)}Z_{1t} & -Z_{1t} \\ \dfrac{-\pi_t^1(1-\pi_t^1)}{\pi_t^0(1-\pi_t^0)+\pi_t^1(1-\pi_t^1)}Z_{2t} & \dfrac{1}{\gamma}\dfrac{-(\pi_t^0+\pi_t^1)}{\pi_t^0(1-\pi_t^0)+\pi_t^1(1-\pi_t^1)}Z_{2t} & -Z_{2t} \end{bmatrix},$$

where the first column is the derivative of $E[g(\beta_0,\sigma_0,\gamma)]$ with respect to σ_0, the second column is the derivative with respect to γ and the third column is the derivative with respect to β_0. For the above Jacobian matrix to be non-singular, we would require some relevance assumptions:

$$Cov\left(\frac{-\pi_t^1(1-\pi_t^1)}{\pi_t^0(1-\pi_t^0)+\pi_t^1(1-\pi_t^1)}, Z_t\right) \neq 0, \quad Cov\left(\frac{-(\pi_t^0+\pi_t^1)}{\pi_t^0(1-\pi_t^0)+\pi_t^1(1-\pi_t^1)}, Z_t\right) \neq 0.$$

When the relevance assumptions are satisfied, the Jacobian matrix is non-singular and therefore the moment conditions (1.22) have a unique solution. In practice, we need enough instruments that satisfy the mean independence assumption and also correlate with the market shares. When there are X's in the model and when there are more than one product, potential extra instruments can be exogenous X's of competing products in the same market or the competitiveness of the market. This is because exogenous characteristics of competing products $k \neq j$ enter the market share function of product j so would in general satisfy the relevance assumption.

1.G RCL with Market Fixed Effects

By Assumption 3, we have $E\left[\left(\delta_{jt}(N_{jt}, M_t, X_t^{(2)}; \gamma_0, \sigma_0) - X_{jt}'\beta_0\right)\phi_j(Z_t)\right] = 0$. I can rewrite the moment condition as

$$E\left[\left(\delta_{jt}\left(N_{jt}, M_t, X_t^{(2)}; \tilde{\gamma}, \sigma_0\right) - X_{jt}'\beta_0 + \delta_{jt}\left(N_{jt}, M_t, X_t^{(2)}; \gamma_0, \sigma_0\right)\right.\right.$$
$$\left.\left. - \delta_{jt}\left(N_{jt}, M_t, X_t^{(2)}; \tilde{\gamma}, \sigma_0\right)\right)\phi_j(Z_t)\right] = 0, \quad (1.23)$$

where $\tilde{\gamma} \in \Gamma$ can be any value in the parameter space of γ. Suppose one assumes the market size coefficient is $\tilde{\gamma}$ and implements the estimation following the standard BLP procedure, then the probability limit of the empirical moment used in estimation would be $E\left[\left(\delta_{jt}(N_{jt}, M_t, X_t^{(2)}; \tilde{\gamma}, \sigma_0) - X_{jt}'\beta_0\right)\phi_j(Z_t)\right]$. Now we explore the possibility of consistently estimating the parameters β and σ by adding market-level fixed effects like what we did in the plain logit case. The question then arises as to whether the term showing up in equation (1.23), $\delta_{jt}(N_{jt}, M_t, X_t^{(2)}; \gamma_0, \sigma_0) - \delta_{jt}(N_{jt}, M_t, X_t^{(2)}; \tilde{\gamma}, \sigma_0)$, is invariant across products in a given market. If yes, then this gap can be captured by a product-invariant parameter κ_t, and the true moment condition (1.23) would be $E\left[\left(\delta_{jt}\left(N_{jt}, M_t, X_t^{(2)}; \tilde{\gamma}, \sigma_0\right) - X_{jt}'\beta_0 - \kappa_t\right)\phi_j(Z_t)\right] =$

0, from which we can consistently estimate σ and β by including market-level dummies, and the choice of $\tilde{\gamma}$ would be a free normalization.

I verify this by looking at the changes in δ_{jt} resulting from changes in γ. First consider the plain logit model, where δ_{jt} has an analytic form. For a scalar γ, the derivative with respect to γ is

$$\frac{\partial \delta_{jt}\left(N_{jt}, M_t, X_t^{(2)}; \gamma\right)}{\partial \gamma} = -\frac{1}{\gamma} - \frac{\sum_k (N_{kt}/M_t)}{\gamma^2 - \gamma \sum_k (N_{kt}/M_t)},$$

which depends only on t, implying that the variation in δ_{jt} as γ changes is not product specific and thus $\delta_{jt}(N_{jt}, M_t, X_t^{(2)}; \gamma_0) - \delta_{jt}(N_{jt}, M_t, X_t^{(2)}; \tilde{\gamma})$ can be captured by κ_t. This is the reason why we can use market fixed effects to capture the unobserved outside option in the logit model.

Now consider random coefficients logit. Suppose $J = 2$, we have

$$\frac{\partial \delta_{1t}\left(N_{jt}, M_t, X_t^{(2)}; \gamma, \sigma\right)}{\partial \gamma} = \begin{vmatrix} \frac{\partial \pi_{1t}}{\partial \delta_{1t}} & \frac{\partial \pi_{1t}}{\partial \delta_{2t}} \\ \frac{\partial \pi_{2t}}{\partial \delta_{1t}} & \frac{\partial \pi_{2t}}{\partial \delta_{2t}} \end{vmatrix}^{-1} \begin{vmatrix} \frac{\pi_{1t}}{\gamma} & \frac{\partial \pi_{1t}}{\partial \delta_{2t}} \\ \frac{\pi_{2t}}{\gamma} & \frac{\partial \pi_{2t}}{\partial \delta_{2t}} \end{vmatrix},$$

$$\text{and } \frac{\partial \delta_{2t}\left(N_{jt}, M_t, X_t^{(2)}; \gamma, \sigma\right)}{\partial \gamma} = \begin{vmatrix} \frac{\partial \pi_{1t}}{\partial \delta_{1t}} & \frac{\partial \pi_{1t}}{\partial \delta_{2t}} \\ \frac{\partial \pi_{2t}}{\partial \delta_{1t}} & \frac{\partial \pi_{2t}}{\partial \delta_{2t}} \end{vmatrix}^{-1} \begin{vmatrix} \frac{\partial \pi_{1t}}{\partial \delta_{1t}} & \frac{\pi_{1t}}{\gamma} \\ \frac{\partial \pi_{2t}}{\partial \delta_{1t}} & \frac{\pi_{2t}}{\gamma} \end{vmatrix},$$

respectively. The denominators are identical for $j = 1, 2$. When $j = 1$, the determinant in the numerator is $\frac{1}{\gamma}\left(\int \pi_{1ti} f_\nu(\nu) d\nu\right)\left(\int \pi_{2ti}(1-\pi_{2ti}) f_\nu(\nu) d\nu\right) + \frac{1}{\gamma}\left(\int \pi_{2ti} f_\nu(\nu) d\nu\right)\left(\int \pi_{1ti}\pi_{2ti} f_\nu(\nu) d\nu\right)$. Similarly, when $j = 2$, the determinant in the numerator is $\frac{1}{\gamma}\left(\int \pi_{2ti} f_\nu(\nu) d\nu\right)\left(\int \pi_{1ti}(1-\pi_{1ti}) f_\nu(\nu) d\nu\right) + \frac{1}{\gamma}\left(\int \pi_{1ti} f_\nu(\nu) d\nu\right)\left(\int \pi_{1ti}\pi_{2ti} f_\nu(\nu) d\nu\right)$. The two are equivalent only when ν is not random and the individual choice probabilities are identical. I can see that it is the individual heterogeneity which enters through the random coefficients that makes $\partial \delta_{jt}(N_{jt}, M_t, X_t^{(2)}; \gamma, \sigma)/\partial \gamma$ depend on j. Overall, $\delta_{jt}(N_{jt}, M_t, X_t^{(2)}; \gamma_0, \sigma_0) - \delta_{jt}(N_{jt}, M_t, X_t^{(2)}; \tilde{\gamma}, \sigma_0)$ would have a j subscript and cannot be captured market fixed effects.

1.H Monte Carlo Simulations

The data generating process for the simulation datasets follows closely that in Armstrong (2016), but we only consider small J environments to avoid the weak instruments problem Armstrong raised. Prices are endogenously generated from a demand and supply model, where firms compete a la Bertrand in the market. In the baseline design of the Monte Carlo study, the number of products varies across markets. $2/3$ of markets have 20 products per market, and the remaining $1/3$ of markets have 60 products in the market. Each firm has 2 products. Other choices of number of products per firm do not significantly alter the results. I consider a relatively small sample size of $T = 100$. I use $R = 1000$ replications of each design.

Consumer utility is given by the random coefficients model described in Section 1.3

$$U_{ijt} = \beta_0 + (\beta_p + \sigma \nu_i) P_{jt} + \beta_1 X_{1,jt} + \xi_{jt} + \varepsilon_{ijt}, \tag{1.24}$$

where ν_i is generated from a standard normal distribution. Firm marginal cost is $MC_{jt} = \alpha_0 + \alpha_1 X_{1,jt} + \alpha_2 X_{S,jt} + \eta_{jt}$. ξ_{jt} and η_{jt} are generated from a mean-zero bivariate normal distribution with standard deviations $\sigma_\xi = \sigma_\eta = 0.8$ and covariance $\sigma_{\xi\eta} = 0.2$. $X_{1,jt}$ and the excluded cost shifter $X_{S,jt}$ are drawn from a uniform $(0,1)$ distribution and independent of each other. All random variables are independent across products j and markets t.

The true values of cost parameters are $(\alpha_0, \alpha_1, \alpha_2) = (2,1,1)$. Demand coefficients and the random coefficient take different values depending on designs.

I compute the true choice probabilities π_{jt} in accordance with equation (1.7). By equations (1.4), we can compute $N_{jt}/M_t = \gamma \pi_{jt}$, where the true value is $\gamma = 1$ throughout the Monte Carlo exercise. In the estimation, one assumes a possibly wrong $\tilde{\gamma}$ and uses the mismeasured $s_{jt} \equiv N_{jt}/\tilde{\gamma} M_t$ as the observed market shares.

The instruments we use in the GMM estimation in all experiments are

$$Z_{jt} = (1, X_{1,jt}, \sum_{k=1}^{J_t} X_{1,kt}, \sum_{k \in \mathscr{J}_f} X_{1,kt}, X_{S,jt}, X_{S,jt}^2),$$

where product j is produced by firm f and $\mathscr{J}_f$ is the set of all products produced by firm f.

I include BLP-type instruments or Gandhi and Houde differentiation instruments as well as functions of excluded cost instruments. The optimization algorithm we use for the GMM estimation is the gradient-based quasi-Newton algorithm (fminunc in MATLAB).

1.H.1 Random Coefficients on Constant Term and Price

The first simulation is designed to assess whether and to what extent random coefficients removes the biases induced from the wrong market size. I generate data from a plain logit model ($\sigma = 0$ in the model of equation (1.24)). It is widely believed that random coefficients partly take over the role of γ and can fix issues caused by unobserved market size. To see if this is true, for each of the $1,000$ simulated datasets, we consider three values of $\tilde{\gamma}$ ($\tilde{\gamma} = 1,2,4$) and estimate both the correctly specified plain logit model and the random coefficients model with a random coefficient on the constant term and price, respectively. I assume that the true demand coefficients are $\beta = (2,-1,2)$.

Tables 1.6 to 1.8 report results from estimating the plain logit model and the more flexible random coefficients models. Each table shows results for three different assumed market size $\tilde{\gamma}$. I report estimates of β, σ, and nonlinear functions of demand parameters, including the own- and cross-price elasticities, and diversion ratios averaged across products for the first market. Reported summary statistics of each parameter estimate across simulations are the mean (MEAN), the standard deviation (SD), and the median (MED).

In Table 1.6, comparing to estimates for the specification with correctly measured market size ($\tilde{\gamma} = 1$) in the first three columns, the means of β's change monotonically as we increase the assumed market size, and their standard deviations change as well. The implied elasticities and diversion ratios are all sensitive to the assumed market size. When we quadruple the assumed market size, the mean of the own-price elasticity increases from -5.99 to -4.17, the cross-price elasticity decreases from 0.077 to 0.028, the individual diversion ratio falls by half and the diversion to the outside good rises from around 17% to 79%.

Table 1.7 shows the results for estimating the random coefficients model with a random coefficient on the constant term. Although the incorrectly assumed market size results in biased estimates of β's, the own-price elasticities and individual diversion ratios of $\tilde{\gamma} = 2,4$ are comparable to the ones of $\tilde{\gamma} = 1$. The cross-price elasticities of the model with incorrectly assumed market size are also closer to those of $\tilde{\gamma} = 1$, relative to the plain logit model in Table 1.6 (decreases from 0.078 to 0.069 versus from 0.077 to 0.028). In contrast, the biases in the outside good elasticity and outside good diversion ratio remain large. When we quadruple the assumed market size, the mean of outside good diversion ratio rises from roughly 17% to 27% and the outside-good price elasticity decreases from 0.077 to 0.007.

In Table 1.8, we estimate the model with a random coefficient on price. Including the random coefficient improves especially the estimates of own- and cross-price elasticities as well as individual diversion ratios, similar to those in Table 1.7.

Although not shown in the table, we also experimented with different numbers of products per market. The design where the number of products varies across markets generally yields larger biases than the design where the number of products is fixed.

Finally, in Table 1.9, we report the estimates from our proposed method of equation (1.5). Results are based on the IV-GMM estimation that uses cost shifters and sum of characteristics as instruments for both price and the observed market to sales variable W_t defined in Section 1.3. Estimates of β and γ are very close to the true values, with small standard deviations. The implied elasticities and diversion ratios are quite comparable to the estimates of the logit model with correctly assumed market size shown in the first three columns of Table 1.6.

To summarize, we find that including a random coefficient on either term accounts for the incorrectly assumed $\tilde{\gamma}$, so that the biases in certain calculations are relatively small. This finding is consistent with the intuition that σ partly corrects for the mismeasured market size. However, biases in other substitution patterns such as cross-price elasticities, outside-good elasticities and diversion ratios are not fully removed.

Table 1.6: Monte Carlo Results: Plain Logit, True $\gamma = 1$

		$\tilde{\gamma} = 1$			$\tilde{\gamma} = 2$			$\tilde{\gamma} = 4$		
	TRUE	MEAN	SD	MED	MEAN	SD	MED	MEAN	SD	MED
β_0	2	1.99	0.318	2.006	-1.205	0.534	-1.192	-2.401	0.594	-2.379
β_p	-1	-0.998	0.056	-1.002	-0.731	0.094	-0.732	-0.688	0.105	-0.691
β_1	2	1.998	0.076	2	1.725	0.105	1.724	1.681	0.114	1.681
Own-Elasticity		-5.994	0.354	-6.006	-4.415	0.584	-4.418	-4.17	0.649	-4.181
Cross-Elasticity		0.077	0.005	0.077	0.028	0.004	0.028	0.013	0.002	0.013
Outside-Good Elasticity		0.077	0.005	0.077	0.028	0.004	0.028	0.013	0.002	0.013
Diversion Ratio		0.014	0	0.014	0.007	0	0.007	0.003	0	0.003
Outside-Good Diversion		0.167	0.027	0.166	0.587	0.013	0.586	0.794	0.007	0.794

Notes: The table reports the empirical mean (MEAN), the standard deviation (SD), and the median (MED) of the demand parameters, the implied price elasticities and diversion ratios for the first market. The GMM estimates are based on $1,000$ generated data sets of sample size $T = 100$ and varied J. The true model is a plain logit model, with $\gamma = 1$. Parameters are estimated from the plain logit model assuming $\tilde{\gamma} = 1, 2, 4$.

Table 1.7: Monte Carlo Results: Random Coefficient on Constant Term, True $\gamma = 1$

		$\tilde{\gamma} = 1$			$\tilde{\gamma} = 2$			$\tilde{\gamma} = 4$		
	TRUE	MEAN	SD	MED	MEAN	SD	MED	MEAN	SD	MED
σ	0	0.037	0.273	0	3.998	0.168	3.992	5.116	0.172	5.11
β_0	2	2.039	0.343	2.05	0.86	0.333	0.862	-1.806	0.321	-1.79
β_p	-1	-1.003	0.057	-1.005	-1.001	0.058	-1.003	-1.001	0.058	-1.003
β_1	2	2.003	0.076	2.005	2.004	0.078	2.005	2.004	0.078	2.005
Own-Elasticity		-6.022	0.357	-6.031	-6.018	0.364	-6.029	-6.02	0.365	-6.03
Cross-Elasticity		0.078	0.005	0.078	0.069	0.005	0.069	0.068	0.005	0.068
Outside-Good Elasticity		0.077	0.005	0.077	0.017	0.001	0.017	0.007	0	0.007
Diversion Ratio		0.014	0	0.014	0.013	0	0.013	0.012	0	0.012
Outside-Good Diversion		0.166	0.027	0.165	0.255	0.01	0.255	0.271	0.009	0.271

Notes: The table reports the empirical mean (MEAN), the standard deviation (SD), and the median (MED) of the demand parameters, the implied price elasticities and diversion ratios for the first market. The GMM estimates are based on $1,000$ generated data sets of sample size $T = 100$ and varied J. The true model is a plain logit model, with $\gamma = 1$. Parameters are estimated from a random coefficients model with the random coefficient on the constant term, assuming $\tilde{\gamma} = 1, 2, 4$.

Table 1.8: Monte Carlo Results: Random Coefficient on Price, True $\gamma = 1$

	TRUE	$\tilde{\gamma} = 1$ MEAN	SD	MED	$\tilde{\gamma} = 2$ MEAN	SD	MED	$\tilde{\gamma} = 4$ MEAN	SD	MED
σ	0	0.013	0.064	0	0.712	0.057	0.712	0.92	0.044	0.919
β_0	2	2.063	0.534	2.057	2.946	0.417	2.951	2.879	0.408	2.88
β_p	-1	-1.005	0.074	-1.006	-1.39	0.071	-1.389	-1.86	0.084	-1.858
β_1	2	2.006	0.09	2.006	2.013	0.08	2.013	2.013	0.08	2.014
Own-Elasticity		-6.034	0.434	-6.031	-6.005	0.402	-6.013	-6.026	0.403	-6.032
Cross-Elasticity		0.078	0.007	0.078	0.065	0.006	0.065	0.063	0.005	0.063
Outside-Good Elasticity		0.078	0.005	0.078	0.025	0.002	0.025	0.01	0.001	0.01
Diversion Ratio		0.014	0	0.014	0.012	0	0.012	0.011	0	0.011
Outside-Good Diversion		0.167	0.027	0.165	0.308	0.019	0.308	0.329	0.02	0.329

Notes: The table reports the empirical mean (MEAN), the standard deviation (SD), and the median (MED) of the demand parameters, the implied price elasticities and diversion ratios for the first market. The GMM estimates are based on $1,000$ generated data sets of sample size $T = 100$ and varied J. The true model is a plain logit model, with $\gamma = 1$. Parameters are estimated from a random coefficients model with the random coefficient on price, assuming $\tilde{\gamma} = 1,2,4$.

Table 1.9: Monte Carlo Results: Estimating γ in the Plain Logit Model

	TRUE	MEAN	SD	MED
γ	1	1.001	0.011	1.001
β_0	2	1.99	0.341	1.993
β_p	-1	-0.999	0.058	-1
β_1	2	1.999	0.077	2
Own-Elasticity		-5.996	0.362	-6.004
Cross-Elasticity		0.077	0.005	0.077
Outside-Good Elasticity		0.077	0.005	0.077
Diversion Ratio		0.014	0	0.014
Outside-Good Diversion		0.168	0.028	0.167

Notes: The table reports the empirical mean (MEAN), the standard deviation (SD), and the median (MED) of the demand parameters, the implied price elasticities and diversion ratios for the first market. The GMM estimates are based on $1,000$ generated data sets of sample size $T = 100$ and varied J. The true model is a plain logit model. Parameters β and γ are estimated from IV-GMM estimations using excluded cost shifters and BLP instruments.

1.H.2 Sensitivity to Market Size Assumption

The second experiment complements the first experiment. I now generate data from a random coefficients model, with a random coefficient for the price. More specifically, we assume that $\beta = (2, -2, 2)$, and $\sigma = 1$. For each of the $1,000$ simulated datasets, we estimate the random coefficients model and consider four values of $\tilde{\gamma}$ ($\tilde{\gamma} = 1, 2, 4, 8$).

This experiment is designed to assess how parameter estimates and the implied substitution patterns vary with market size assumptions in a random coefficients logit model.

Table 1.10 shows results of demand estimates and the implied statistics. Some general tendencies stand out. First, consumer heterogeneity (σ) and disutility for price (β_p) tend to be overestimated as $\tilde{\gamma}$ increases. The direction of biases in β_0 is ambiguous. Second, the implied elasticities and diversion ratios give similar results as those in Table 1.8. The outside-good elasticities and the outside-good diversion ratios are most sensitive to the choice of $\tilde{\gamma}$. The cross-price elasticities are also affected, but not as sensitive as the former two calculations. However, biases in elasticities and diversion ratios tend not to be monotonic in $\tilde{\gamma}$. For instance, $\tilde{\gamma} = 2$ leads to an upward bias of the diversion to outside good (from around 17% to 20%), but $\tilde{\gamma} = 4$ gives a modest downward bias of the outside-good diversion (from 17% to 16%). The extreme case, which imposes $\tilde{\gamma} = 8$, results in a much larger bias (from 17% to 25%). Hence, imposing different assumptions of the market size is not a simple rescaling of the calculations. This again confirms that random coefficients logit models do not correct for all biases induced by wrong market size assumptions.

Table 1.10: Sensitivity to Market Size Assumptions in Random Coefficients Logit, True $\gamma = 1$

	TRUE	$\tilde{\gamma} = 1$ MEAN	$\tilde{\gamma} = 1$ SD	$\tilde{\gamma} = 1$ MED	$\tilde{\gamma} = 2$ MEAN	$\tilde{\gamma} = 2$ SD	$\tilde{\gamma} = 2$ MED	$\tilde{\gamma} = 4$ MEAN	$\tilde{\gamma} = 4$ SD	$\tilde{\gamma} = 4$ MED
σ	1	1	0.034	0.999	1.413	0.036	1.413	2.646	0.173	2.629
β_0	2	2.012	0.447	1.999	1.431	0.396	1.418	2.164	0.604	2.143
β_p	-2	-2.001	0.068	-2	-2.68	0.069	-2.681	-4.604	0.273	-4.577
β_1	2	1.998	0.054	2.001	1.984	0.055	1.987	2	0.055	2.001
Own-Elasticity		-7.095	0.328	-7.079	-6.922	0.334	-6.913	-7.025	0.392	-6.986
Cross-Elasticity		0.077	0.005	0.076	0.071	0.004	0.071	0.075	0.005	0.074
Outside-Good Elasticity		0.029	0.003	0.029	0.011	0.001	0.011	0.004	0	0.004
Diversion Ratio		0.014	0	0.014	0.014	0	0.014	0.014	0.001	0.014
Outside-Good Diversion		0.175	0.025	0.176	0.201	0.022	0.201	0.167	0.033	0.168
		$\tilde{\gamma} = 8$								
σ	1	2.427	0.048	2.426						
β_0	2	-1.252	0.416	-1.247						
β_p	-2	-4.307	0.091	-4.306						
β_1	2	1.91	0.066	1.909						
Own-Elasticity		-5.84	0.445	-5.826						
Cross-Elasticity		0.052	0.004	0.052						
Outside-Good Elasticity		0.002	0	0.002						
Diversion Ratio		0.013	0	0.013						
Outside-Good Diversion		0.247	0.023	0.246						

Notes: The table reports the empirical mean (MEAN), the standard deviation (SD), and the median (MED) of the demand parameters, the implied price elasticities and diversion ratios for the first market. The GMM estimates are based on 1,000 generated data sets of sample size $T = 100$ and varied J. The true model is a random coefficients logit model with a random coefficient for price, with $\gamma = 1$. Parameters are estimated from the random coefficients model, assuming $\tilde{\gamma} = 1, 2, 4, 8$.

1.H.3 Market Size Estimation in Random Coefficients Logit

The third experiment enables us to assess the performance of our proposed method. As we discussed in Section 1.3, it suffices to use the same set of BLP-type instruments to estimate the market size parameter γ in addition to the random coefficient parameter σ.

The baseline design (design 1) is the same as before: $2/3$ of markets have 20 products per market and the rest of markets have 60 products in the market. The true values of demand parameters are $\beta = (2, -2, 2)$. I consider two alternative designs, changing either the market structure or demand parameters. In design 2, we use the same set of parameters $\beta = (2, -2, 2)$ as design 1, but assume all markets have 20 products. This leads to less variation in the true outside share π_{0t} across markets. In design 3, we use the same market structure as design 1, but assume $\beta = (2, -3, 2)$. This particular choice of parameters leads to larger true outside share π_{0t}, and less variation of π_{0t} in design 3 than in design 1. The average π_{0t} across $1,000$ simulated samples is 0.55 for design 1, while 0.9 for design 3.

Tables 1.11 and 1.12 report results from each design. In addition to the mean, the standard deviation, and the median, we also report the 25% quantile (LQ), the 75% quantile (UQ), the root mean squared error (RMSE), the mean absolute error (MAE), and the median absolute error (MDAE).

Table 1.11 shows results for the baseline design. The primary parameter of interest, γ, tends to be estimated precisely, with the RMSE being 0.2. Estimates of β and σ are mostly close to the true parameter values, and the RMSEs are small. Only the estimate of the constant term coefficient β_0 is somewhat variable, having a larger RMSE of 0.9. Although not reported in the main tables, we have estimated the same specification replacing BLP-type instruments with Gandhi and Houde differentiation instruments. The resulting estimates are qualitatively similar overall but somewhat more precise with smaller RMSEs.

In Panel A of Table 1.12, estimates from design 2 are generally noisier than those in design 1, with most RMSEs in the range of 0.7 to 1.3. The median of estimates remains close to the true values. Although γ and demand parameters are less precisely estimated in design 2, our proposed estimation is still more preferable to making wrong assumptions of the market size. As shown in the table, the mean of γ estimates is 1.447, which is closer to the true value than any $\tilde{\gamma} > 1.5$. Panel B provides results for design 3. γ, σ and β_p appear to be difficult to be precisely estimated, with large standard deviations. Intuitively, when the shares of the outside option are too large, the variation of market shares of inside goods is squeezed. The limited variation in data leads to the poor performance of the estimator.

This confirms that our proposed estimator works well particularly in cases where the true outside good share is not too large and has enough variation across markets.

Table 1.11: Estimating γ in the Random Coefficients Logit Model, Design 1

	TRUE	MEAN	SD	LQ	MED	UQ	RMSE	MAE	MDAE
γ	1	1.032	0.211	0.861	1.004	1.195	0.213	0.178	0.173
σ	1	0.969	0.226	0.805	1.019	1.16	0.228	0.19	0.169
β_0	2	1.655	0.924	1.146	1.842	2.296	0.985	0.704	0.517
β_p	-2	-1.956	0.358	-2.26	-2.036	-1.686	0.361	0.303	0.273
β_2	2	1.989	0.059	1.95	1.994	2.026	0.06	0.047	0.038

Notes: The table report summary statistics of the demand parameters. The GMM estimates are based on $1,000$ generated data sets of sample size $T = 100$ and varied J. The true model is a random coefficients logit model with a random coefficient for price. Parameters β, σ and γ are estimated from IV-GMM estimations using excluded cost shifters and BLP instruments. Design 1: $\beta = (2, -2, 2)$, varied number of products per market.

Table 1.12: Estimating γ in the Random Coefficients Logit Model, Alternative Designs

	TRUE	MEAN	SD	LQ	MED	UQ	RMSE	MAE	MDAE
					Panel A: Design 2				
γ	1	1.447	1.188	0.887	1.006	1.711	1.269	0.607	0.222
σ	1	1.169	0.712	0.913	1.034	1.291	0.732	0.312	0.156
β_0	2	1.744	0.835	1.285	1.771	2.287	0.873	0.663	0.511
β_p	-2	-2.273	1.109	-2.483	-2.052	-1.863	1.142	0.502	0.255
β_2	2	1.991	0.077	1.936	1.994	2.044	0.078	0.062	0.052
					Panel B: Design 3				
γ	1	2.234	2.143	0.67	1.011	3.452	2.472	1.574	0.457
σ	1	2.518	5.15	0.795	0.994	2.223	5.367	1.743	0.287
β_0	2	1.844	1.511	1.309	1.835	2.305	1.518	0.659	0.511
β_p	-3	-5.351	7.901	-4.938	-2.988	-2.665	8.24	2.731	0.537
β_2	2	1.989	0.119	1.958	1.994	2.028	0.12	0.046	0.034

Notes: The table report summary statistics of the demand parameters. The GMM estimates are based on $1,000$ generated data sets of sample size $T = 100$ and varied J. The true model is a random coefficients logit model with a random coefficient for price. Parameters β, σ and γ are estimated from IV-GMM estimations using excluded cost shifters and BLP instruments. Design 2: $\beta = (2, -2, 2)$, fixed number of products per market. Design 3: $\beta = (2, -3, 2)$, varied number of products per market.

1.I Pricing Conditions in Merger Analysis

Assume that firms are under a static Nash-Bertrand pricing game. Following the steps and notation in Weinberg and Hosken (2013), let $\mathscr{J}_f$ denote the set of all products produced by firm f. The first-order condition for product j produced by firm f can be written as

$$\sum_{k \in \mathscr{J}_f} \left(\frac{p_k - mc_k}{p_k} \right) \eta_{k,j} \pi_k + \pi_j = 0, \tag{1.25}$$

where mc is the marginal costs, and $\eta_{k,j}$ is the elasticity of product k with respect to the price of j. This yields a system of J equations in each market. Using observed prices, market shares, and the price elasticities computed from the estimated demand, one can solve for the marginal costs.

After a merger, firms' profit functions change and the equilibrium prices firms optimally choose would also change. If firm f merged with firm g, holding the characteristics and marginal costs of all their products constant, the merged firm's first-order conditions become:

$$\sum_{k\in\mathscr{J}_f}\left(\frac{p_k-mc_k}{p_k}\right)\eta_{k,j}\pi_k+\sum_{h\in\mathscr{J}_g}\left(\frac{p_h-mc_h}{p_h}\right)\eta_{h,j}\pi_h+\pi_j=0,$$

based off which one can use the recovered marginal costs and estimated demand to solve for the post-merger equilibrium prices.

To demonstrate how a wrong market size can undermine the conclusion of a merger analysis, we substitute the formula of price elasticities into equation (1.25), giving

$$-\sum_{k\in\mathscr{J}_f}(p_k-mc_k)\int\beta_{pi}\pi_{ji}\pi_{ki}dF(\beta_{pi})+\pi_j=0.$$

The market size affects three things: the estimated random coefficient on price β_{pi}, the estimated individual choice probabilities π_{ji} and π_{ki}, and the share π_j itself.

1.J Additional Results for the CSD Application

1.J.1 Aggregate Price Elasticity

I provide additional results for the soft drink application. First, we calculate the price elasticity of aggregate demand, which is the percentage change in total sales for soft drinks when the prices of all soft drinks increase. Note that we can link aggregate demand directly to the outside share, by recognizing that without an outside option defined in the model, the aggregate market demand is perfectly inelastic. More formally, in a simple logit model, the price elasticity of aggregate demand can be calculated by $\alpha\pi_0\hat{p}$, where α is the price coefficient and $\hat{p}$ the average price.

This aggregate elasticity can be thought of as the market-level response to a proportional tax imposed on all products. It is economically important, for example, when policymakers aim to evaluate the effectiveness and targeting of soda taxes.

Figure 1.3 illustrates the estimated aggregate elasticities of demand in each market when $\gamma = 17$ and 12, respectively. With a larger market size, the aggregate elasticity falls (in absolute value). The direction of bias is same as those found in Conlon and Mortimer (2021). Moreover, it not only changes the mean level but also the overall distribution across markets. This finding confirms that market size definition is relevant for questions that affect all products in a market.

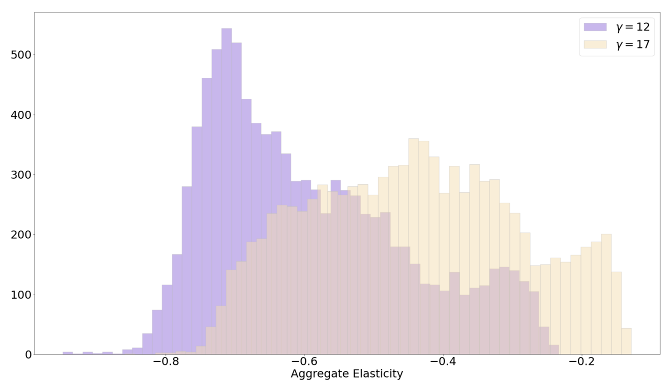


Figure 1.3: Distribution of Aggregate Elasticities across Markets
Notes: The figure shows the aggregate elasticities of demand across markets for $\gamma = 12$ and 17.

1.J.2 Profiled GMM Objective Function

I plot the GMM objective function while keeping γ fixed over a grid of values and re-optimizing the remaining parameters with the weighting matrix fixed. There are no multiple minima within the specified interval. However, the function is not steep around the minimum, which could pose challenges for numerical optimization. Stronger instruments may help improve parameter identification and numerical optimization.

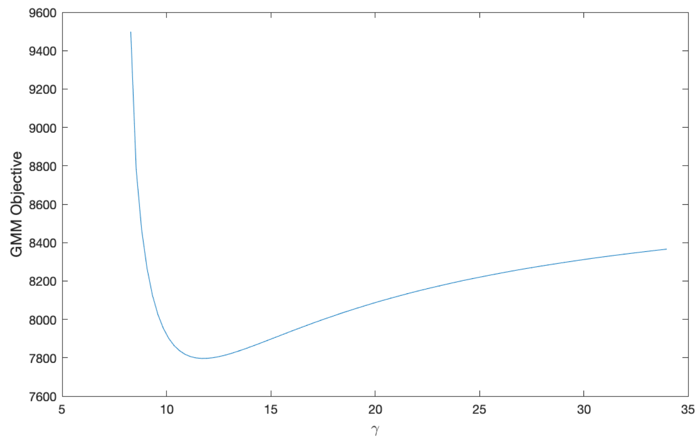


Figure 1.4: Profiled GMM Objective

Notes: The figure shows the profiled GMM objective. γ is fixed while the remaining parameters are re-optimized.

1.K Merger Analysis: Ready-to-Eat Cereal Market

The data in Nevo (2000) is simulated from a model of demand and supply, and consists of 24 brands of the ready-to-eat cereal products for 94 markets. Nevo's specification contains a price variable and brand fixed effects. The variables that enter the non-linear part of the model are the constant, price, sugar content and a mushy dummy. For each market 20 iid simulation draws are provided for each of the non-linear variables. In addition to the unobserved tastes, v_i, demographics are drawn from the current population survey (CPS) for 20 individuals in each market. It allows for interactions between demographics such as income and the child dummy with price, sugar content and the mushy dummy, capturing heterogeneity on the tastes for product characteristics across demographic groups. To instrument for the endogenous variables (prices and market shares), Nevo (2000) uses as instruments the prices of the brand in other cities, variables that serve as proxies for the marginal costs , distribution costs and so on.

A market is defined as a city-quarter pair and thus the market size is the total potential number of servings. Nevo assumes the potential consumption is one serving of cereal per day. Using notations in this paper, the assumed market potential is therefore $1 \cdot M_t$, where M_t is the population in city t in this case.

The baseline specification replicates that in Nevo (2000). I calculate the estimated own- and cross-price elasticities and diversion ratios, which are the mean of all entries of the elasticity/diversion ratio matrix over the 94 markets. The results demonstrate the average substitution patterns between products. On the basis of the baseline estimation, we consider a hypothetical merger analysis between two multi-products firms. Post-merger equilibrium prices are solved from the Bertrand first order condition. Consumer surplus claculations are provided to show the impacts of the hypothetical merger. Next, we consider an alternative choice of potential market size. I rescale the market shares for all inside goods by a factor of $1/2$, which is equivalent to taking the potential market size to be double as large as in the baseline case. I resimulate the merger using the rescaled market shares. Finally, we assume the true market size is γ servings per person per day, estimate γ and repeat the merger simulation.

Table 1.13 reports the demand coefficients and the implied mean elasticities and diversion ratios. The baseline estimation replicates the results in Nevo (2000). Interestingly, doubling the market size has little impact on the estimates of demand coefficients β and σ. The baseline estimation has a price coefficient of -32 and the rescaled of -28.9. However, translating it to elasticities and diversion ratios, we see a substantial increment in the diversion to outside option. In particular, the average outside-good diversion increase from 37.5% to 60.2%. These estimates imply that, if one assumed a larger market size, more consumers would switch to outside good rather than alternative substitutes upon an increase in price of inside goods. The third column presents the estimated γ and the associated demand estimates. $\hat{\gamma} = 0.78$ means that the true market size is a potential daily consumption of 0.78 servings per person. The implied market size is smaller than the baseline case, leading to a lower true diversion ratio. My estimate of γ makes economic sense and has a small standard error. Given γ estimate being 0.78, we can calculate the outside share is about 40%. It is a relatively small outside share so the identification is strong in the current context.

In order to quantify the overall effect of uncertainty in market size on merger analysis, we look at the impact on both the simulated prices and consumer surplus. Figure 1.5 plots the distribution of percentage price changes pre- and post-merger, where the three curves plot the baseline case, rescaled case and the case for our estimate of γ. Predicted price increase is the smallest when we assume $\gamma = 2$. When the potential market size is two times the baseline case, prices of the merging brands respond relatively less to the merger, with a median increase of 5.4%. While in the baseline case, the median price increase is 10.7% for the merging brands. Under the true estimated market size $\hat{\gamma} = 0.78$, the predicted price increase is larger than assuming $\gamma = 1$. This is consistent with our intuition: when there are less people substitute to outside good, the merging firms will have a greater increase in market power.

Next we consider the implications of our estimates for the consumer surplus change after the merger.[21] As expected, we predict a larger decrease in consumer surplus when the price increase is high. Overall, different market sizes affect how much we predict a merger harms consumer welfare.

[21] The consumer surplus is the expected value of the highest utility one can get measured in dollar values. It is calculated by $CS = \sum_{i=1}^{NS} w_{it} CS_{it}$, where the consumer surplus for individual i is

$$CS = \ln\left(1 + \sum_{j \in J_t} \exp V_{ijt}\right) \bigg/ \left(-\frac{\partial V_{i1t}}{\partial p_{1t}}\right), \text{ and } V_{ijt} \equiv U_{ijt} - \varepsilon_{ijt}.$$

Table 1.13: Parameter Estimates for the Cereal Demand

	Baseline (M_t)	Rescaled ($2M_t$)	Estimate γ
β_{price}	-32	-28.9	-35.817
	(2.304)	(3.294)	(7.055)
σ_{cons}	0.375	0.245	0.684
	(0.120)	(0.156)	(0.329)
σ_{price}	1.803	3.312	2.134
	(0.920)	(0.972)	(1.737)
σ_{sugar}	-0.004	0.016	-0.029
	(0.012)	(0.014)	(0.029)
σ_{mushy}	0.086	0.025	0.173
	(0.193)	(0.192)	(0.269)
$\sigma_{cons\times inc}$	3.101	3.223	4.119
	(1.054)	(0.875)	(1.799)
$\sigma_{cons\times age}$	1.198	0.7	2.118
	(1.048)	(0.682)	(1.755)
$\sigma_{price\times inc}$	4.187	-2.936	8.979
	(4.638)	(5.155)	(152.358)
$\sigma_{price\times child}$	11.75	10.87	14.495
	(5.197)	(4.747)	(7.515)
$\sigma_{sugar\times inc}$	-0.19	-0.143	-0.295
	(0.035)	(0.032)	(0.081)
$\sigma_{sugar\times age}$	0.028	0.027	0.024
	(0.032)	(0.033)	(0.038)
$\sigma_{mushy\times inc}$	1.495	1.396	1.526
	(0.648)	(0.470)	(0.898)
$\sigma_{mushy\times age}$	-1.539	-1.251	-1.919
	(1.107)	(0.677)	(1.675)
γ			0.779
			(0.062)
Mean own-elasticity	-3.702	-3.682	-3.804
Mean cross-elasticity	0.095	0.061	0.121
Mean outside-good diversion	0.375	0.602	0.226

Notes: The first column is the baseline estimation where market potential is 1 serving per person per day. The second column is the rescaled estimation where the market potential is 2 servings per person per day. In the third column we estimate the market size parameter γ.

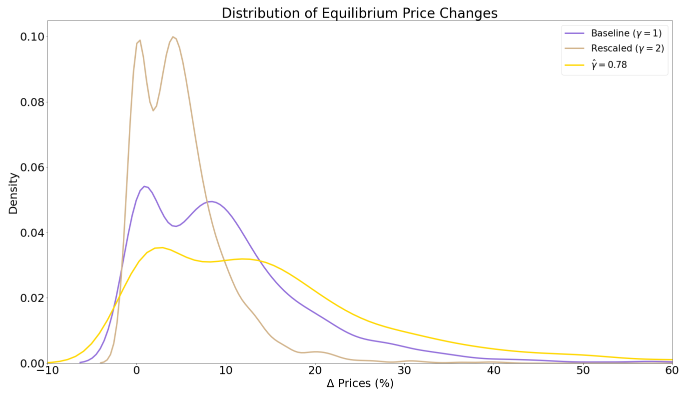


Figure 1.5: Equilibrium Price Changes

Notes: The figure shows changes in equilibrium prices after a merger between firms 1 and 2.

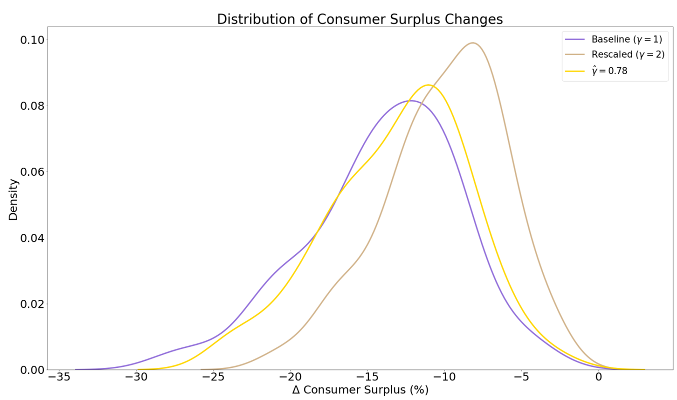


Figure 1.6: Consumer Surplus Changes

Notes: The figure shows changes in equilibrium prices after a merger between firms 1 and 2.

1.L Additional Derivations

Partial Derivatives of π_{jt}

The partial derivatives of π_{jt} with respect to δ_{jt} and δ_{kt} are functions of mean utilities and characteristics of all products:

$$\frac{\partial \pi_{jt}}{\partial \delta_{jt}} = \int \pi_{jti}\left(\delta_t, X_t^{(2)}; \sigma\right)\left(1 - \pi_{jti}\left(\delta_t, X_t^{(2)}; \sigma\right)\right) f_\nu(\nu) d\nu,$$

$$\frac{\partial \pi_{jt}}{\partial \delta_{kt}} = -\int \pi_{jti}\left(\delta_t, X_t^{(2)}; \sigma\right) \pi_{kti}\left(\delta_t, X_t^{(2)}; \sigma\right) f_\nu(\nu) d\nu,$$

where

$$\pi_{jti}\left(\delta_t, X_t^{(2)}; \sigma\right) = \frac{exp\left(\delta_{jt} + \sum_l \sigma_l x_{jtl}^{(2)} \nu_{il}\right)}{1 + \sum_{k=1}^{J_t} exp\left(\delta_{kt} + \sum_l \sigma_l x_{ktl}^{(2)} \nu_{il}\right)}.$$

The partial derivatives of π_{jt} with respect to σ_l is

$$\frac{\partial \pi_{jt}\left(\delta_t, X_t^{(2)}; \sigma\right)}{\partial \sigma_l} = \int \pi_{jti}\left(\delta_t, X_t^{(2)}; \sigma\right)\left(x_{jtl}^{(2)} - \sum_{k=1}^{J} x_{ktl}^{(2)} \pi_{kti}\left(\delta_t, X_t^{(2)}; \sigma\right)\right) \nu_{il} f_\nu(\nu) d\nu$$

Relevance of Instruments

The legitimacy of treating λ_γ and λ_σ alike in section 1.4.3 is shown below. I first recognize that for any given (N_t, M_t, X_t) and model parameters, the residual function in equation (1.9) can be rewritten as

$$\xi_{jt}\left(\frac{N_t}{\sum_k(\lambda_{\gamma_{k1}} + 1) M_t^{\lambda_{\gamma_{k2}}}}, X_t; \lambda_\sigma, \beta\right) = \delta_{jt}\left(\frac{N_t}{\sum_k(\lambda_{\gamma_{k1}} + 1) M_t^{\lambda_{\gamma_{k2}}}}, X_t^{(2)}; \lambda_\sigma\right) - X_{jt}'\beta. \quad (1.26)$$

When $\lambda_{\gamma_1} = \lambda_{\gamma_2} = 0$, and let s_t denote the usual observed shares N_t/M_t, the residual function reduces to

$$\xi_{jt}(s_t; \lambda_\sigma, \beta) = \delta_{jt}(s_t; \lambda_\sigma) - X_{jt}'\beta,$$

which is equivalent to equation (4) in Gandhi and Houde (2019). When λ_γ is different from zero, the residual function would depend nonlinearly on λ_γ as well. The residual function is not linear in λ_γ because $\partial \delta_{jt}/\partial \lambda_\gamma$ is a function that depends on λ_γ.

The linear approximation in section 1.4.3 can also be obtained from linearizing the inverse demand function around the true λ_0

$$\begin{aligned}\delta_{jt}\left(N_t,M_t,X_t^{(2)};\lambda\right) &\approx \delta_{jt}\left(N_t,M_t,X_t^{(2)};\lambda_0\right)+\sum_l(\lambda_{\sigma_l}-\lambda_{\sigma_l 0})f^{\sigma}_{l,jt}+\sum_k(\lambda_{\gamma_k}-\lambda_{\gamma_k 0})f^{\gamma}_{k,jt}\\ &= X'_{jt}\beta_0+\xi_{jt}+\sum_l(\lambda_{\sigma_l}-\lambda_{\sigma_l 0})f^{\sigma}_{l,jt}+\sum_k(\lambda_{\gamma_k}-\lambda_{\gamma_k 0})f^{\gamma}_{k,jt},\end{aligned}$$

with $f^{\sigma}_{l,jt}=\partial\delta_{jt}(N_t,M_t,X_t^{(2)};\lambda_0)/\partial\sigma_l$, $f^{\gamma}_{k,jt}=\partial\delta_{jt}(N_t,M_t,X_t^{(2)};\lambda_0)/\partial\gamma_k$. Note that $f^{\sigma}_{l,jt}$ and $f^{\gamma}_{k,jt}$ depend on the vector of δ_t and $X_t^{(2)}$.

Chapter 2

Who is Most Affected by Soda Taxes? Evidence from Purchases At-Home and Away-From-Home

XIRONG LIN AND LINQI ZHANG

2.1 Introduction

It is well known that sugary drinks can have negative effects on health, including a correlation with diabetes, heart disease, and childhood obesity (Currie 2009, Currie et al. 2010, Gortmaker, Long, and Wang 2009, Griffith et al. 2020, Cutler, Glaeser, and Shapiro 2003). Experts suggest that in the developed world sugar consumption is far above the recommended level. Therefore, the resulting individual and social costs of the internalities (that is, the ignored future costs of current consumption) and externalities (costs that are borne by others) of sugar consumption have attracted policymakers' concern. Many countries have implemented taxes on sugar-sweetened beverages (SSB) in order to discourage soft drink consumption.[1] Soft drinks are also a main contributor to the sugar consumption of

[1] See Allcott, Lockwood, and Taubinsky 2019b for a summary of countries that implement sugar-sweetened beverage taxes.

vulnerable individuals: the young, high sugar consumers, and the poor. Whether soda taxes can be effective in reducing sugar consumption and improving welfare depends crucially on how demand responses vary across different demographic groups and may also vary across consumption locations (at-home versus away-from-home).

In this paper, we assess whether soda taxes are effective at lowering the sugar consumption of households by taking into account all channels (at-home, on-the-go, and restaurants) of households' SSB consumption. We account for heterogeneity in household taste for SSB and price sensitivity in these different channels in order to comprehensively and precisely estimate the demand responses of households who are targeted by the policy (i.e., households with children, high sugar consumers, SNAP (Supplemental Nutrition Assistance Program) participants, and non-SNAP-participant poor). We reveal new evidence on the important but understudied away-from-home segment (on-the-go and restaurants) of the market. With novel data, we can estimate a model of consumer choice in both at-home and away-from-home segments. We uncover household preferences in each segment, and we then simulate the impacts of a soda tax, allowing for different pass-through rates in each segment.

Extending the existing literature, we make two main advances in this paper. First, we document descriptive patterns of household SSB purchases at-home, on-the-go, and in restaurants. We use a unique dataset, USDA's National Household Food Acquisition and Purchase Survey (FoodAPS). The main advantage of FoodAPS, as compared to other popular scanner datasets like Nielsen Homescan Data, is its complete coverage of all food purchased from all sources and of both packaged and random-weight items.[2] FoodAPS was filed between April 2012 and January 2013, which was before the enactment of any SSB taxes in the U.S.. Hence, we cannot exploit any variation in tax policy by time or region. Instead, we utilize the rich demographic information to build a demand model and

[2]Nielsen Homescan Data are known to under-report at-home purchases, not record random-weight purchases with sufficient product detail to understand diet behavior, and not provide any away-from-home information (Zhen et al. 2023).

simulate counterfactual tax experiments.

Most papers on SSB taxes look at only one of the segments, at-home (Allcott, Lockwood, and Taubinsky 2019a, Bollinger and Sexton 2018), on-the-go (Dubois, Griffith, and O'Connell 2020), or in restaurants (Moran et al. 2019). In fact, the latter two channels have rarely been studied although they actually constitute a large fraction of household expenditures. For example, Americans drink 52% of SSB calories at-home and 48% of SSB calories away-from-home.[3] Different from in other countries, U.S. households purchase a considerable amount of SSB from the nation's largest chain restaurants, particularly when combination meals or kids' menu items are ordered (Moran et al. 2019).[4] These facts imply that any SSB tax analysis missing one of the segments will be an incomplete documentation of the impact of soda taxes on household sugar intake.

Second, we estimate a discrete choice demand model by adapting the nonparametric framework of Fox et al. 2011 and Fox, Kim, and Yang 2016b. Prior work that uses a similar approach includes Nevo, Turner, and Williams 2016 and Blundell, Gowrisankaran, and Langer 2020. The estimation technique is nonparametric in the sense that it estimates the distribution of random coefficients over a fixed grid of potential values, rather than assuming that the random coefficients are drawn from a known distribution. This is important given the recent empirical finding in Dubois, Griffith, and O'Connell 2020 that preferences vary with demographics in ways that would be difficult to capture by specifying a priori the distribution of random coefficients.[5] Our model allows us to use the rich demographic information in FoodAPS, including SNAP participation and eligibility and household income and composition, in order to reveal the diverse preferences and elasticities across household types.

[3]See Kit et al. 2013.

[4]For example, in the U.K. data from Dubois, Griffith, and O'Connell 2020, on-the-go purchases are three times as large as that of restaurant purchases, while the U.S. sample shows the opposite pattern.

[5]Dubois, Griffith, and O'Connell 2020 overcome this problem by estimating an individual-level demand model using U.K. households SSB purchases on-the-go. We cannot follow the same strategy because it is not possible to have individual-level consumption information at-home and in restaurants. Hence, we choose the framework of Fox et al. 2011 and Fox, Kim, and Yang 2016b as a middle ground. Throughout the paper, we compare our model, findings, and implications to Dubois, Griffith, and O'Connell 2020.

The only other paper we know that investigates SSB taxes with FoodAPS is Zhen et al. 2023. They estimate the Exact Affine Stone Index model (Lewbel and Pendakur 2009) of food and study the correlation between obesity and food consumption. We instead estimate a characteristics type of demand model of SSB consumption. We emphasize on the heterogeneity of SSB demand in terms of SNAP participation and income inequality. The similarity between us is that we both point out the importance of away-from-home consumption in household demand and policy implications.

We find that preferences and elasticity vary with demographics in terms of SNAP participation, income, the existence of children, and household dietary sugar. Consistent with previous literature, low-income households have the strongest preferences for SSB. But among the poor, SNAP eligible nonparticipants have weaker preferences than SNAP poor and are the most elastic to price among all groups. Elder households and those without children have weaker preferences for SSB and are more sensitive to price. Lastly, SSB preferences exhibit an increasing relationship with dietary sugar while price elasticity exhibits a decreasing relationship with dietary sugar.

In terms of heterogeneity among segments, we find that consumers have diverse preferences and elasticity at-home versus away-from-home. For example, low-income households have strong preferences for SSB at-home while high-income households prefer more SSB in restaurants. Those with higher dietary sugar from SSB obtain much more SSB at-home than away-from-home and vice versa. In terms of price responsiveness, overall the average elasticity across all groups at-home is larger than that away-from-home. For heterogeneity along the line of SNAP participation and income, elasticities vary widely across groups both at-home and away-from-home. But for heterogeneity in the dimension of total dietary sugar from SSB, there is minimal variation in elasticities away-from-home but large differences in elasticities at-home.

Our findings suggest that on average, the current taxes of the form and size implemented in the U.S. lead to reductions of around 18.12 percent, 5.75 percent, and 14.53

percent in the total sugar intakes from SSBs at-home, away-from-home, and in total. We find that soda taxes are less effective away-from-home and there is little variation in responses across households. One reason is that households (mainly the high-income households) who have strong preferences for SSB away-from-home are also less price sensitive and hence have small reductions in sugar intake from SSB. In contrast, we find substantial variation in demand responses at-home across households. Soda taxes are relatively effective at targeting the total sugar intake of the poor, those with high sugar consumption, and households without children.[6] Lastly, our results suggest that ignoring any segment will lead to biased policy implications. For example, Dubois, Griffith, and O'Connell 2020 find that the soda tax is less successful at targeting those with high total dietary sugar for the on-the-go segment. However, we find that the total (at-home and away-from-home) reduction in sugar intake from SSB is largest for those households if we account for all segments (the reduction almost doubles that of the high-income households).

One major debate about SSB taxes is the concern that they are regressive, i.e., the poor spend a disproportionately large fraction of expenditures on SSB and they end up bearing the largest share of the tax burden. We use compensating variation (the amount of money that an individual needs to reach her pre-tax utility level after the imposition of an excise tax) as our welfare measure for soda taxes and compare it across household groups by income and SNAP participation. Unlike the previous literature which often finds a larger compensating variation for low-income households, we find that even though low-income households obtain more added sugar from SSB, their compensating variation is not much higher than that of high-income households. That is because soda taxes are based on volume rather than the amount of sugar in each drink. Given the fact that, high-income households obtain more soft drinks away-from-home, this was not accounted for by the

[6]Following previous literature like Dubois, Griffith, and O'Connell 2020, we measure the effectiveness of SSB tax in terms of the level reduction, rather than percentage reduction in sugar from SSB. For example, those with high total dietary sugar have the strongest preferences for SSB while the lowest price elasticity. However, the soda tax is effective at targeting these households because their level-reduction in sugar from SSB is still the largest across groups even though their percentage reduction in sugar from SSB is the lowest.

previous literature. Meanwhile, both groups can obtain similar amounts of soft drinks while the sugar amount in each drink is much higher in low-income households. These findings suggest that household preference heterogeneity in each purchase segment (at-home or away-from-home) and their preference for the specific drink types (in terms of the amount of sugar) together determine the welfare cost of a soda tax. Ignoring either segment can lead to biased policy implications of soda taxes.

Literature Review This paper belongs to a burgeoning literature on the effect of soda taxes.[7] One strand of this literature exploits a specific SSB tax implementation or reform, used as quasi experiments, in order to estimate the effects of those reforms on household SSB expenditures (Seiler, Tuchman, and Yao 2021, Rojas and Wang 2017, Bollinger and Sexton 2018). Most of these studies use retailer-side scanner data in order to study the overall impact of an SSB tax on prices and consumption. The findings speak to the effect of the specific reform and there are different results for different places. For example, Seiler, Tuchman, and Yao 2021 find that a soda tax in Philadelphia is passed through at an average rate of 97% and demand decreases by 46%. However, accounting for cross-border shopping reduces the demand response by 20%. Rojas and Wang 2017 compare SSB tax pass-through rates and volume sales in Washington DC and Berkeley CA. They find that both retail price and volume sales in Washington react sharply while in Berkeley retail price reacts only marginally with no effect on the volume of sales. Yet, these empirical studies do not provide any mechanisms that explain the conflicting findings or allow further re-valuation of alternative reforms in order to derive the most effective policy. Moreover, Allcott, Lockwood, and Taubinsky 2019b and Rozema and Rees-Jones 2018 point out that interest groups' advertising campaigns and public debates can also confound the effects of SSB taxes.

In contrast, our paper belongs to the other strand of the literature that include papers with structural models of household demand for SSB and that simulates counterfactual ex-

[7]See Allcott, Lockwood, and Taubinsky 2019a and Cawley et al. 2019 for overviews of the theory and evidence of an SSB tax.

ercises of alternative tax policies. These include Dubois, Griffith, and O'Connell 2020, Allcott, Lockwood, and Taubinsky 2019a, Wang 2015, Bonnet, Requillart, et al. 2012, Harding and Lovenheim 2017, and Chernozhukov, Hausman, and Newey 2019. We complement these papers by exploiting a novel dataset with consumer SSB consumption in all segments as well as rich demographic information on the policy targeted groups. Our nonparametric demand model estimates and the counterfactual findings have meaningful implications for the effectiveness of soda taxes. Similar to Ramsey 1927, Diamond and Mirrlees 1971a, 1971b, and Miravete, Seim, and Thurk 2020, we find that preference heterogeneity among consumers and variation in demand elasticities by segments are important aspects in the design of optimal tax schemes.

Our paper also complements the literature on food and beverage consumption with consumer side scanner data at-home (Aguiar and Hurst 2007, Lin 2023, and Dubois, Griffith, and Nevo 2014), away-from-home (Dubois, Griffith, and O'Connell 2020, O'Connell and Smith 2020, and Griffith, Jin, and Lechene 2022, Saksena et al. 2018, Moran et al. 2019), or both (Okrent and Alston 2012).

The rest of this paper is structured as follows. In Section 2.2 we describe the data and the soft drinks market. We present a detailed descriptive analysis of the consumption patterns across diverse demographic groups and retail segments. In Section 2.3 we present the demand model, the identification, the nonparametric estimation, and the empirical estimates of the model parameters. In section 2.4 we simulate the counterfactual tax exercise and discuss its implications on effectiveness, targeting, and regressivity of an SSB tax. Section 2.5 concludes.

2.2 The Nonalcoholic Drinks Market

We model household behavior in the nonalcoholic drinks market. Nonalcoholic drinks include soft drinks (e.g., carbonated drinks, commonly referred to as soda, fruit drinks, sport and energy drinks, and sweetened coffee and tea), alternative sugary drinks (fruit and vegetable juice, unsweetened coffee and tea, flavored milk), and bottled water. soda taxes are typically imposed on soft drinks that contain added sugar. Diet drinks and drinks with natural sugar like fruit juice are normally exempted from soda taxes .

We focus on household purchases at-home (grocery store, supermarket, pharmacy, club store, dollar store, gas station/market), on-the-go (convenience store, vending machine, and retail store), and in the restaurants (coffee shop and care, fast-food outlet, restaurants, drinking places, other store and farmers market). Drinks market at-home have been extensively studied in the literature, while there are not many studies on drinking away-from-home. But it is very important to study this segment because near half of the nonalcoholic drinks are purchased away-from-home in the U.S. (Table 2.2). We document the purchase behavior in the three segments separately in this section. However, due to the small fraction of the on-the-go segment out of total SSB expenditures (10%), in the demand estimation, we aggregate drinks on-the-go and restaurants into one segment, that is, drinks away-from-home.

We start by documenting household nonalcoholic drinks purchases in the FoodAPS dataset, e.g., the prices paid, expenditures and expenditure shares, products bought, and places types. We then relate the purchase behavior to household demographic characteristics such as household income, age, the existence of children, and overall dietary sugar intake from SSB.

2.2.1 FoodAPS

The data that we use is the Food Acquisition and Purchase Survey (FoodAPS), which was fielded from April 2012 through January 2013. The FoodAPS data collect information on foods that all household members acquired from all sources over a 7-day period, for a sample of 4,826 US households.[8]

[8] For more information regarding the collection of the data, please refer to the Appendix Section 2.A.

There are three main advantages of FoodAPS. First, to our knowledge, the FoodAPS away-from-home survey is unique. Second, FoodAPS includes packaged and random-weight food (like fruit and vegetables), which are not available in other scanner data.[9] Third, Clay et al. 2016 compare FoodAPS to other scanner datasets and find that FoodAPS suffers less from underreporting due to missing items scanned or unrecorded shopping trips.[10]

Another nice feature of FoodAPS is its classification of households into four groups: (1) SNAP participants, (2) Households with income below the poverty line but do not participate in SNAP, (3) Households with income at or above 100 percent and less than 185 percent of the poverty guideline, and (4) Households with income equal to or greater than 185 percent of the poverty guideline. Other scanner data like Nielsen Homescan Data do not have information on SNAP eligibility or participation.

One caveat of FoodAPS is that it only collects information over a 7-day period and households might not purchase all items in need in that given period. However, households in FoodAPS are surveyed in different time from April 2012 through January 2013. We expect the spread of household purchases across different weekdays, weekends, months, and seasons will mitigate the concern.[11] The average SSB volume consumption at-home in FoodAPS is 43 and 70 liters per year (imputed from weekly volume) for high and low income households (Figure 2.2a). The numbers are 50 and 85 liters per year in Allcott, Lockwood, and Taubinsky 2019a Figure 1. The lower volume found in FoodAPS can be due to the short survey period and infrequency of purchase.

Table 2.1 describes the distribution of place types where households purchase SSB. In total, the sample contains 3,556 households who make any SSB purchases during the data collection week. Among them, 514 households buy SSB on-the-go, 1,837 households

[9]For example, Nielsen Homescan Data only collects information of random-weight items for a subset of households in certain years.

[10]For example, they find that the mean at-home spending in the Consumer Network scanner panel is 26% lower than that of FoodAPS.

[11]We utilize this feature to incorporate time controls in the demand model.

buy SSB at-home, and 2,115 households make SSB purchases in restaurants. We observe more than 50 percent of households purchasing SSB only in one of the three segments. For households who make SSB purchases in only two of the three segments, most of them make it at-home and in restaurants. Only 11 percent of the households purchase SSB in all three segments during the week. We describe some key features of the data in the next section.[12]

Table 2.1: Place Types

	At-home	Restaurant	On-the-go	At-home + Restaurant	At-home + On-the-go	Restaurant + On-the-go	At-home + Restaurant + On-the-go
Number of households	745	989	124	914	178	212	394
Percent of sample	20.95	27.81	3.49	25.7	5.01	5.96	11.08

Notes: The table shows different place type combinations at which households make any SSB purchases during the data collection week. Columns 1 to 3 show the number of households who purchase SSB in only one of the three place types. Columns 4 to 6 show the number of households who purchase SSB in two out of the three place types. The last column is the number of households who purchase SSB in all three place types.

2.2.2 Places, Prices, and Products

Places

Consumers visit different retailers when they shop at-home, on-the-go, or in restaurants. This implies that the prices and choice sets that they are faced with in a choice occasion will also differ. In Table 2.2, we describe the types of retailers and the expenditure share of drinks at each type. In total, the away-from-home segment accounts for 46% of the total SSB spending, while that number is 54% for the at-home segment. Previous literature that ignore the away-from-home drinks spending miss a large fraction of household total spending on drinks. Under the away-from-home segment, convenience store, vending machine, and retail store together can be classified as the on-the-go segment. They account for only 18% of spending in the away-from-home segment. This suggests that previous literature like Dubois, Griffith, and O'Connell 2020 and O'Connell and Smith 2020 who only study

[12]For more details, please refer to Appendix section **??**

SSB at-home or on-the-go overlook 38% of household total SSB spending that happens in restaurants and cafes.

Table 2.2: Expenditure share (%)

Away-from-home	46%	At-home	54%
Convenience store	10%	Combination grocery/other	3%
Retail store	6%	Dollar store	4%
Coffee shop and cafe	15%	Convenience store	4%
Fast-food outlet	34%	Gas station/market	2%
Restaurants	31%	Grocery store, large	1%
Drinking place	1%	Grocery store, medium	1%
Vending machine	2%	Pharmacy	2%
Other store and farmers market	0%	Super store	44%
		Supermarket	35%
		Club stores	4%

Notes: Numbers are % of drinks spending, in at-home and away-from-home segments, by retailer.

We do not model household choices over which place to shop in. We assume that the decision is driven by factors such as the proximity of nearby super stores or restaurants and overall preferences for different segments and place types (for which we control in demand). We assume away from the possibility that consumers choose places in order to search for a temporarily low price for a specific drink. The assumption is reasonable because we find that consumers tend to choose nearby places to shop in FoodAPS. Previous evidence also shows that fixed shopping costs lead consumers to undertake their grocery shopping in one or a small number of stores. Similar assumption is also made in O'Connell and Smith 2020.

Products

In Table 2.6, we describe the products of soft drinks available in the U.S. market, their percentage of transactions, and mean prices.[13] The current SSB tax is placed on the first four categories: soft drinks, fruit drinks, sport and energy drinks, and sweetened coffee and

[13] For more details about the construction of products, please refer to the Appendix section 2.B.

tea.[14] We classify whether a product is purchased in one of the two segments: at-home or away-from-home (on-the-go, or in restaurants). We also allow products to differ by packaging formats (regular, large, or multi-pack).[15] As a result, a product in the demand estimation is defined as either a category-package-segment, or a category-segment combination, for fruit juice, unsweetened coffee and tea, flavored milk and water, we aggregate across different sizes. In total, these non-SSB beverage categories account for less than 16 percent of the market.

From Table 2.6, we find that households in our sample on average purchase an SSB product in 34 percent of choice occasions. Among them, at-home purchases of SSB products account for around 16 percent of choice occasions, while away-from-home purchases of SSB products account for around 18 percent of choice occasions. The most frequently purchased product is the regular soft drinks away-from-home, which account for 9.4 percent of choice occasions. When consumers do not purchase any of the drinks during a choice occasion, we assume them to choose the "outside options": either the household purchases other food (e.g. meat or snacks) if it was a trip to a food store, or the household purchases a meal without ordering drinks if it was a trip to an eating place.

Unlike most of the previous literature where a product is defined as a brand-size combination (e.g., Dubé 2005b, Marshall 2015b, Dubois, Griffith, and O'Connell 2020), we abstract away from the substitution across brands because brand information is not available in the public file of FoodAPS. Also in the away-from-home segment, most items have missing UPCs. Instead, we aim to measure how consumers substitute between regular coke and diet coke rather than between regular Cola and regular Pepsi.

[14]For details regarding the construction of products, please see Section 2.B in the Appendix.

[15]Regular size is defined as a container size smaller than 32 ounces. Large size is defined as a container size larger than 32 ounces.

Prices

For each transaction, we observe the type of store the consumer shops and the total expenditures on each UPC. We calculate the transaction price as the expenditure on a UPC per unit purchased. In Figure 2.6, we show the average prices of all SSB purchased across all choice occasions in each segment by household income, age of the primary respondent, and total added sugar from SSB. First, panel (a) shows that overall, SSB prices are the highest in restaurants, second highest on-the-go, and the lowest at-home. There is a positive relationship between prices and income in all three segments. The slope is the steepest in the restaurants segment. In other words, richer households buy much more expensive SSB in restaurants. Second, panel (b) shows that the age of the primary respondent is positively correlated with SSB prices in restaurants while negatively correlated with SSB prices on-the-go. Together with panel (a), this implies that richer and older households pay much higher price for SSB products in restaurants. Lastly, SSB price is negatively correlated with weekly added sugar from SSB in all three segments. In other words, those with higher dietary sugar buy cheaper SSB products, no matter at-home, on-the-go, or in restaurants.

Price varies across products, time period and retailer type. The price of beverage products sold in a meal bundle is imputed by the price of other items of the same product within the same place category. Specifically we apply a linear interpolation of transaction price on package size for missing values of price, and perform this calculation separately for each product-place category combination. Figure **??** in the Appendix shows the variation in price across demographic groups.

For details about the composition of household SSBs purchases, please refer to Appendix section 2.B.

2.2.3 Demographics

Income

In Figure 2.1, we show that income is negatively correlated with household shares of SSB expenditures on food at-home while positively correlated with household shares of SSB expenditures in restaurants. This pattern mainly reflects two commonly found evidence in the empirical literature. First, poor people are more likely to have a less healthy diet, like high dietary sugar. Second, income and restaurant expenditures are positively correlated. Figure 2.1 also shows that household SSB expenditure shares on-the-go do not vary much across the deciles of household income.

The above findings have significant implications on how current literature on SSB consumption could go wrong by using only data on food at-home. For papers that only use food at-home SSB expenditures, they will miss the important asymmetry of income and SSB expenditures. First, they will mistakenly conclude that lower-income households purchase more SSB than higher-income households. Second, their elasticities estimates will also miss the heterogeneity in household demand responses by household income and segments. For papers that only use on-the-go SSB expenditures, they will not find any income effect in SSB demand, which is certainly not true from the evidence of Figure 2.1. Moreover, in the US, SSB consumption on-the-go constitutes very small shares, only 10%, of households overall SSB expenditures.

Another important implication from Figure 2.1 is related to regressivity of an SSB tax. Empirical evidence suggests that poor households spend more on SSB than richer households.[16] This implies that an SSB tax will fall disproportionately more on the poor. However, we see that high-income households actually spend more on SSB in restaurants than poor households. So the overall regressivity of an SSB taxes is unknown when we account for both food at-home and food away-from-home.

[16] For example, Allcott, Lockwood, and Taubinsky 2019b find that in the Nielsen Consumer Panel dataset, households with annual income below $10,000 purchase about 101 liters of SSB per adult each year, while households with income above $100,000 purchase only half that amount.

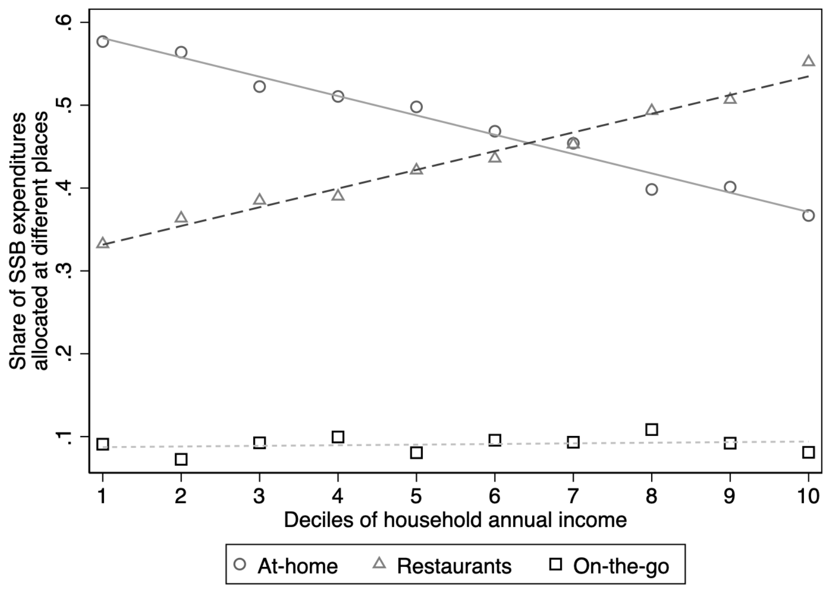


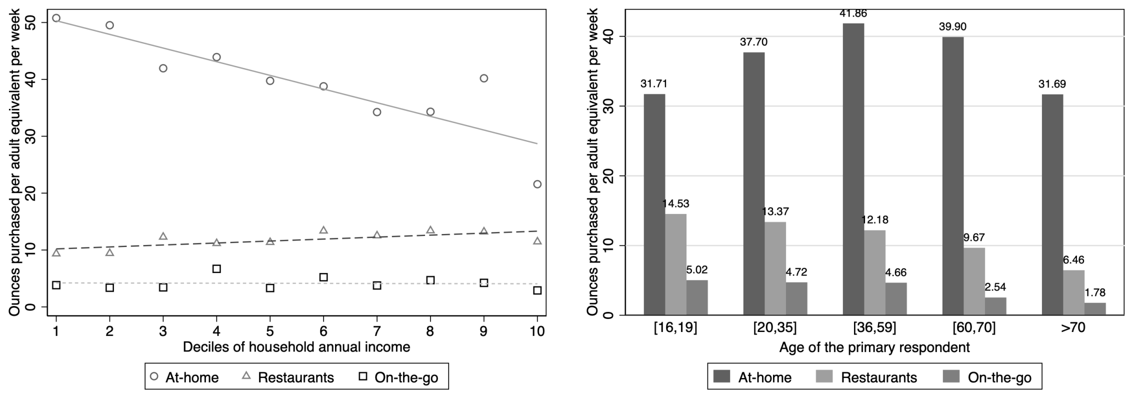
Figure 2.1: Average share of SSB expenditures with respect to income

SSB Consumption in Volume

Figure 2.2 plots the average purchased volume of SSB per adult equivalent per week with respect to household income and the age of the primary respondent.

Similar to Figure 2.1, income and SSB volume consumption are still negatively correlated. However, different from Figure 2.1, there is much smaller in magnitude the increasing trend between income and SSB volume consumption in restaurants. It implies that the steeper positive trend between income and SSB budget shares in restaurants in Figure 2.1 is mainly due to the price effect. That is, richer households buy slightly more SSB products in restaurants but much more expensive products there.

Panel (b) in Figure 2.2 turns out to be very different from Figure 2.7. First, across all age groups, the volume consumption of SSB in restaurants is much lower than that at-home. Combined with the similar budget shares of SSB at-home and in restaurants in Figure 2.2, this finding implies that prices should be much higher for SSB products in restaurants than at-home. Second, there is an inverted U-shape relationship between the age of the primary respondent and SSB volume purchased at-home. In other words, middle age families buy the most SSB at-home. There is a decreasing trend between age and SSB volume purchased in restaurants. This is consistent with the evidence found in Martin et al. 2020.



(a) By deciles of household annual income

(b) By age of the primary respondent

Figure 2.2: SSB volume consumption at-home, on-the-go, and in restaurants

Notes: The figure shows weekly SSB purchases (in oz.) per adult equivalent at different places varies by distribution of household equivalized income and age groups. In plot (a), the household annual income is equivalized by the OECD-modified equivalence scale. In plot (b), age groups are classified according to the same cutoffs as in the FoodAPS dataset.

SNAP Poor and Non-SNAP Poor

FoodAPS dataset has direct information on SNAP participation and income. They classify households into four types: SNAP participants, SNAP non-SNAP-participant poor (income $<$ 100% of the Federal Poverty Guideline), medium (income $>=$ 100% and income $<$ 185% of FPG) and higher (income $>=$ 185% of FPG) income households.[17] Table 2.3 reports the descriptive statistics for each group. Compared to SNAP participants, the non-SNAP poor have much lower household income, slightly older primary respondent, and fewer children.

First, in terms of the SSB consumption, we find that SNAP households have the largest volume consumption, SSB expenditures, and total grams of added sugar from SSB. In con-

[17] As reported by the Census, in 2019, 17% of those who were eligible for SNAP benefits did not participate in the program (see https://www.census.gov/library/stories/2021/02/demographic-snapshot-not-everyone-eligible-for-food-assistance-program-receives-benefits.html). The finding that non-SNAP poor households have lower income than SNAP participants is consistent with a previous study that documents SNAP patterns using CPS data (Gundersen 2021).

trast, non-SNAP-participant poor have the lowest value for all the previous three variables. In other words, SNAP participants eat the least healthy while SNAP non-SNAP-participant poor eat the most healthy, even compared to higher income households. This finding contradicts with previous evidence that low-income households in general eat less healthy than higher income households. However, this evidence is possibly driven by the fact that a large fraction of low income households are SNAP participants and SNAP benefits cover soft drink purchases. Previous literature find that SNAP households' shopping cart consist of lots of soda.[18]

Second, non-SNAP rich households turn out to have the second highest SSB consumption in volume and expenditures, as well as total added sugar from SSB. This finding is also very different from previous analyses that only look at consumption at-home. They normally find rich households to spend less on SSB. Rich households turn out to eat healthy at-home but unhealthy away-from-home. This finding highlights the importance of accounting for away-from-home consumption in order to evaluate households' overall diet quality and sugar intake.

Figure 2.3 provides a breakdown of the SSB consumption by places. It shows that SNAP households consume the highest ounces of SSB per adult equivalent per week at-home. Non-SNAP poor have similar SSB consumption at-home as higher income households. In terms of SSB consumed in restaurants, higher income households have higher ounces purchased. These findings suggest that the main difference in SSB consumption between SNAP and non-SNAP poor is driven by at-home purchases. Again, this can be driven by the fact that SNAP benefits can be spent on soft drinks.

[18] See O'Connor (2017) https://www.nytimes.com/2017/01/13/well/eat/food-stamp-snap-soda.html

Table 2.3: Descriptive Statistics by SNAP Status

	SNAP		Non-SNAP Poor		Non-SNAP Medium		Non-SNAP Rich	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Equivalized household annual income	12725.10	11461.10	6863.13	3208.24	15261.88	2634.37	43886.84	29479.16
Age of the primary respondent	42.92	15.51	44.95	16.96	48.10	19.25	47.33	16.59
Number of children	1.36	1.47	0.91	1.36	0.89	1.32	0.67	1.03
Household size (adult equivalents)	3.23	1.77	2.78	1.72	2.72	1.70	2.67	1.37
Total grocery expenditure	114.29	109.26	85.36	81.43	89.57	78.74	124.07	109.95
Total SNAP EBT amount reported	60.85	89.36	0.00	0.00	0.00	0.00	0.00	0.00
SSB volume consumption (ounce)	65.39	100.17	45.94	76.33	49.95	75.91	51.38	88.30
SSB expenditure (dollar)	2.68	3.64	1.98	2.51	2.18	2.77	2.64	3.41
Added sugar from SSB (gram)	184.54	282.57	129.65	214.08	140.07	212.77	141.35	254.04
Observations	1536		318		820		2009	

Notes: Grocery expenditure is the total dollar amount a household spent during a one week period in stores for at-home consumption. SNAP EBT amount reported is the total SNAP EBT amount reported by respondent or corrected by value observed on receipt when SNAP EBT payment is used for acquisition. SSB volume consumption, SSB expenditure and added sugar from SSB are measured at a per adult equivalent level.

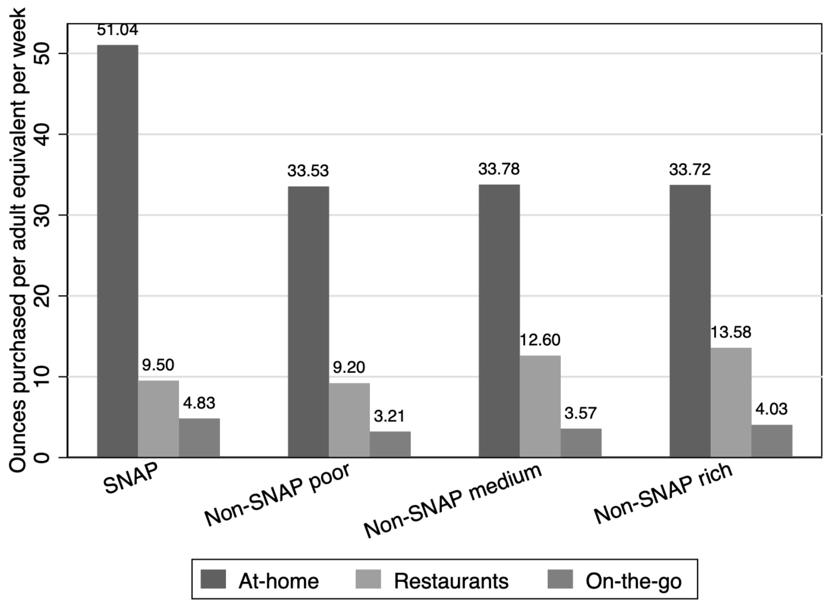


Figure 2.3: SSB volume consumption with respect to targeted group

2.2.4 Summary

To summarize, prices of SSB are the highest in restaurants, middle on-the-go, and the lowest at-home. Higher income households purchase more expensive soda drink in all three segments and buy drinks in restaurants more frequently. SNAP participants buy more SSB compared to nonparticipant poor, whose spending on SSB is even lower than higher-income households. Heavy sugar consumers tend to be lower income households and purchase less expensive soda drinks in all three segments. Once we account for drinks away-from-home, the regressivity concern of soda taxes becomes less serious because higher income households will be taxed more in restaurants while lower income households will be taxed more at-home. Who bares the most tax burden is ambiguous unless we have a demand model that accounts for all three segments. Section 2.C in the Appendix also shows how SSB purchases vary by age and existence of children.

Previous literature normally focus on the poor, children, and heavy sugar consumers as the main targets of soda taxes . For these three types of households, restaurants purchases seem to be less important for them compared to at-home purchases. However, these households still buy certain amount of soda drinks away-from-home. We may underestimate the effect of taxes on their total SSB demand if we ignore away-from-home purchases. On the other hand, they can have different price elasticities in the at-home and away-from-home segment. For example, if they are more price sensitive in the at-home segment, then we may overestimate the effectiveness of taxes because they may still buy a certain part of SSB from the away-from-home segment.

The other main critical benefit of including restaurants data in the analysis is the finding that higher income households will also be largely affected by an SSB taxes because they purchase more SSB, and potentially more expensive SSB, in restaurants. This finding suggests that the regressivity concern of an SSB taxes may be less serious than we expect. We will calculate the regressivity in the next chapters by predicting households' counterfactual SSB expenditure shares in terms of total expenditures on food and drinks in the three segments when we simulate tax incidences.

2.3 Model and Estimated Coefficients

In this section we estimate a structural model of non-alcoholic beverage demand. We employ a random coefficients nested logit model to create more flexible substitution patterns.[19] We estimate the random coefficient flexibly following Fox et al. 2011 and Fox, Kim, and Yang 2016b, for two reasons: First, Dubois, Griffith, and O'Connell 2020 finds that unobserved preferences might not be fully captured by specifying a priori the distribution of random coefficients. Thus, we relax the often-made assumption of normal distribution in BLP applications. Second, the method of Fox et al. 2011 and Fox, Kim, and Yang 2016b has the advantage of achieving a flexible distribution while maintaining computational tractability. We then use the model to evaluate counterfactual tax policies that could

[19] See Grigolon and Verboven 2014 for a comprehensive comparison between the random coefficients nested logit, random coefficients logit and nested logit models.

reduce SSB consumption.

2.3.1 A Model of Non-Alcoholic Beverage Demand

We index consumers by $i \in \{1, \cdots, N\}$. Each consumer visits a retailer $r \in \{1, \cdots, R\}$ at time t and makes a transaction or incurs choice occasion $\tau \in \{1, \cdots, \mathscr{T}\}$. Let $r(\tau)$ and $t(\tau)$ denote the specific retailer and time that the consumer visits the retailer. Notice that the retailer here can be grocery stores, convenience stores, vending machines, or restaurants. We index the non-alcoholic beverage products by $j \in \{1, \cdots, J\}$, as those defined in Table 2.6.

When making a decision, the choice set facing consumers contains purchasing options that are available to the consumers on each specific trip. This means that when a consumer visits a grocery store, she only considers drinks available at the store. Similarly, on a trip to the food at-home place, the choice sets facing the consumer only include food at-home drink products. We denote the choice set by $\Omega_{r(\tau)}$.

We allow for the possibility that a consumer instead chooses either other non-beverage products like meat or snacks in a store, or purchases a meal without ordering any drinks in a restaurant. We refer to these as "outside options". We indicate outside options by $j = 0$, and the choice set $\Omega_{r(\tau)}$ includes the outside option. [20]

We partition the choice set $\Omega_{r(\tau)}$ to two disjoint subsets denoted by C_0 and C_1. They are also called nests. C_0 is the nest of outside options. C_1 is the nest of all available products in the choice set. The indirect utility of a consumer i on choice occasion τ from product j in the nest C_g where $g \in \{0, 1\}$ is given by

$$U_{ij\tau} = V_{ij\tau} + \varepsilon_{ij\tau} \tag{2.1}$$

where

$$V_{ij\tau} = \alpha_i p_{jr(\tau)t(\tau)} + \eta_i s_j + \mathbf{x}'_{ij}\beta, \tag{2.2}$$

[20] Our definition of outside options is most close to that of Marshall 2015b, assuming the outside option is chosen when a shopping trip is observed with no purchase of any inside goods.

and the utility obtained from choosing the outside options is

$$U_{i0\tau} = \varepsilon_{i0\tau}. \tag{2.3}$$

The term $p_{jr(\tau)t(\tau)}$ denotes the price of product j, which varies over time t and across retailer types r.[21] The variable s_j is an indicator variable of an SSB product. $\mathbf{x}_{ij}$ is a vector of observed product characteristics (including the constant term) and their interactions with household demographics. Specifically, $\mathbf{x}_{ij}$ include package size (measured in ounces),[22] an indicator variable of products in pack, indicator variables of drink categories, season fixed effects, and retailer-drink category fixed effects.[23] $\varepsilon_{ij\tau}$ is an error term following the generalized extreme value distribution, with cumulative distribution of the following form

$$\exp\left(-\sum_{g=0}^{1}\left(\sum_{j\in C_g} e^{-\varepsilon_{ij\tau}/\lambda}\right)^{\lambda}\right),$$

which gives rise to the nested logit structure. For this distribution of $\varepsilon_{ij\tau}$, the idiosyncratic error terms are correlated within a given nest. For any two products belonging to different nests, the error terms are uncorrelated. The nest parameter λ measures dissimilarity among products within a nest. A value of $\lambda = 1$ indicates that $\varepsilon_{ij\tau}$ are uncorrelated within nests and the model degenerates to the standard logit model. As λ decreases, the correlation within nests rises. The nested logit assumption implies that products in the same nest are closer substitutes than products in different nests. In our context, the two nests are the outside options and all products in the choice set, respectively. This modeling assumption implies that all available products are considered closer substitutes than the outside option.

Allowing for preference heterogeneity is essential in capturing realistic demand features. The demand model here is flexible in that it incorporates preference heterogeneity

[21] Even though the price paid varies across households (Figure 2.6), we do not construct household-level prices for products because of the endogeneity concern. That is, the individual-level prices can be correlated with the error term $\varepsilon_{ij\tau}$. The practice is common in demand estimation with scanner dataset (e.g., Griffith et al. 2020, Lin 2023). For details regarding the construction of $p_{jr(\tau)t(\tau)}$, please see Section 2.D in the Appendix.

[22] For details about the construction of package size, please refer to Appendix section 2.B.

[23] Retailers are defined in Table 2.2. Drink categories are defined in Table 2.7.

through two aspects. The first is through the idiosyncratic error component $\varepsilon_{ij\tau}$, as we previously discussed. The second is through the taste heterogeneity for product attributes. Specifically, we let taste parameters like α_i and η_i vary by household observed and unobserved characteristics. We define the marginal (dis)utility of price and taste for SSB as the following[24]

$$\alpha_i = \alpha_0 + \mathbf{v}_i'\alpha_1 + \mu_i \quad \eta_i = \mathbf{v}_i'\eta_1,$$
$$\text{where } \mu_i \sim F(\mu).$$

$\mathbf{v}_i$ denotes observed household demographics, and μ_i is a random coefficient and captures unobserved preference related to prices. By allowing α_i and η_i to depend on household characteristics, we allow different consumers to have different price sensitivity when making purchases for at-home and away-from-home consumption and different tastes for SSB products.

The observed household demographics $\mathbf{v}_i$ include a joint variable of household income and SNAP participation, the age of the primary respondent, whether the household has children, and the household's overall sugar intake.[25] The unobserved household characteristics, or random coefficient μ_i can include individual household information that affects the purchasing decision, yet unobservable to econometricians.[26] Prior empirical work on random coefficients logit model usually make parametric assumptions on the distribution of the random coefficient $F(\mu)$, e.g., a normal distribution. We will relax this assumption and estimate $F(\mu)$ using a fixed grid approach, following Fox et al. 2011 and Fox, Kim,

[24]We do not put a constant term η_0 in the η_i function because $\mathbf{x}_{ij}$ contains indicator variables of drink categories, which leads to the mean level of the taste coefficient for SSB η_0 unidentified due to perfect collinearity problem.

[25]The correlation between "SNAP households" indicator and "households with children" indicator is only -0.2 and hence we include both indicators in the model. We also find that the average age of children is similar across sugar intake groups and hence we do not include it in the model.

[26]Following Bonnet and Dubois 2010 and Eizenberg and Salvo 2015b, among others, we only have a random coefficient on price.

and Yang 2016b, to allow for more flexible unobserved taste heterogeneity.[27]

Conditional on the unobservable μ_i, the joint probability of a consumer choosing a product j is

$$P_{ij\tau}(\mu_i) = P_{ij\tau|C_g}(\mu_i) P_{iC_g\tau}(\mu_i),$$

where $P_{ij\tau|C_g}(\mu_i)$ is the probability of choosing a product j conditional on a product in the nest C_g being chosen. $P_{iC_g\tau}(\mu_i)$ is the marginal probability of choosing a product in the nest C_g, conditional on the unobservable μ_i. As shown in McFadden 1978, the joint probability of choosing product $j \in C_g$ (conditional on μ_i) takes the nested logit formula

$$P_{ij\tau}(\mu_i) = \frac{e^{V_{ij\tau}(\mu_i)/\lambda} \left(\sum_{k \in C_g} e^{V_{ik\tau}(\mu_i)/\lambda} \right)^{\lambda - 1}}{\sum_{l=0}^{1} \left(\sum_{k \in C_l} e^{V_{ik\tau}(\mu_i)/\lambda} \right)^{\lambda}}. \tag{2.4}$$

The unconditional probability of consumer i choosing product $j \in C_g$ in a choice occasion τ is

$$P_{ij\tau} = \int \frac{e^{V_{ij\tau}(\mu)/\lambda} \left(\sum_{k \in C_g} e^{V_{ik\tau}(\mu)/\lambda} \right)^{\lambda - 1}}{\sum_{l=0}^{1} \left(\sum_{k \in C_l} e^{V_{ik\tau}(\mu)/\lambda} \right)^{\lambda}} dF(\mu). \tag{2.5}$$

2.3.2 Identification

Our identification is similar to Dubois, Griffith, and O'Connell 2020 except that we only know retailer types as those in Table 2.2 rather than specific retailers.[28]

The main identification challenge is to isolate the causal effect of price on demand for at-home and away-from-home products. That is, the parameter vector α. We rely on two sources of variation to identify the price effects. First, conditional on time and retailer-drink type effects, we exploit the variation in prices of the same product in different retailer types across time. The identification assumption is that consumers do not choose retailers when they make consumption choices for a specific product. Instead, retail choices are

[27] Birchall and Verboven 2022 and Miravete, Seim, and Thurk 2022 highlight that the functional form of the indirect utility function implicitly imposes restrictions on demand curvature. Miravete, Seim, and Thurk 2022 specifically suggest that incorporating a flexibly distributed price random coefficient not only offers greater flexibility in substitution patterns, but also allowing for a wider range of estimable demand curvature.

[28] FoodAPS does not have brand information and we are not able to exploit consumers' taste differentials across Pepsi versus Coca Cola.

more driven by convenience factors like distance to school and workplace.[29] Second, we utilize price variation for the same product in the same retailer at the same time but across different containers and sizes.

To address potential endogeneity bias arising from the correlation between the unobserved error term $\varepsilon_{ij\tau}$ and prices, we control for time and retailer-drink category fixed effects. These fixed effects capture demand shocks that vary at these levels, which could have driven price fluctuations. By including these fixed effects, we capture most of the unobservable factors, leaving only time-varying shocks specific to the retailer-drink type combination; for example, consumers may exhibit a growing preference for soda relative to fruit juice at convenience stores. The assumptions we adopt are that stores do not observe such demand shocks in the current period when making price decisions, and that any residual retailer-drink type-specific price variation across time reflects supply-side changes, such as differences in transportation costs due to adapting a new distribution network. Therefore, the inclusion of a rich set of fixed effects in our demand model should mitigate any bias resulting from unobservable factors that might be correlated with prices.

2.3.3 Estimation

Following recent literature, we estimate the at-home and away-from-home segment separately and obtain distinct preferences parameters for these two segments.[30] We refer to the method of Fox et al. 2011 and Fox, Kim, and Yang 2016b to estimate the random coefficients. The method has the advantage of achieving a flexible distribution, while maintaining

[29] A survey conducted by the National Retail Federation or NRF showed that 93 percent of shoppers said they are more likely to choose to shop at a specific retailer based on convenience (https://www.nasdaq.com/articles/convenience-is-priority-for-consumers:-survey-2020-01-15)

[30] The other papers that estimate different food segments separately include O'Connell and Smith 2020 and Dubois, Griffith, and O'Connell 2020, both studying the at-home and on-the-go segments. They follow Browning and Meghir 1991 to test for non-separabilities between segments and find no evidence of demand dependence between the two segments. Following them, we also conduct a separability test (see Section 2.E in the Appendix). The results support separability between at-home and away-from-home soft drinks consumption. Meanwhile, we have also tried to estimate the two segments jointly. The computation is burdensome due to the large choice set included.

computational tractability.

Consider a fixed grid $\mathscr{M}_R = (\mu^1, \cdots, \mu^R)$, where R represents the number of grid points. One can interpret R as the number of discrete household types.[31] We assume each μ_i is a draw from the set of values $(\mu^1, \cdots, \mu^R)$ and each grid point μ^r occurs with probability γ^r, for $r = 1, \cdots, R$. Given the choice of $\mathscr{M}_R$, we estimate the weights $\gamma = (\gamma^1, \cdots, \gamma^R)$ on the grid points. We impose the constraints $0 \leq \gamma^r \leq 1, \forall r$, and $\sum_{r=1}^{R} \gamma^r = 1$, such that μ has a well-defined distribution.

For each household type r, we can rewrite choice probability (2.4) by replacing μ_i with μ^r:

$$P_{ij\tau}(\mu^r) = \frac{e^{V_{ij\tau}(\mu^r)/\lambda} \left(\sum_{k \in C_g} e^{V_{ik\tau}(\mu^r)/\lambda} \right)^{\lambda - 1}}{\sum_{l=0}^{1} \left(\sum_{k \in C_l} e^{V_{ik\tau}(\mu^r)/\lambda} \right)^{\lambda}}. \tag{2.6}$$

Let $\alpha = (\alpha_0, \alpha_1)$ and $\eta = \eta_1$. Denote by $\theta = (\alpha, \beta, \eta, \gamma)$ the vector of preference parameters. Using choice probabilities defined above, we calculate the likelihood function defined by

$$\mathscr{L}(\theta) = \sum_i \sum_\tau \sum_{j \in \Omega_{r(\tau)}} d_{ij\tau} \log \left(\sum_{r=1}^{R} \gamma^r P_{ij\tau}(\mu^r) \right), \tag{2.7}$$

where θ is the vector of parameters to be estimated and $d_{ij\tau}$ is an indicator variable equal to one if consumer i chose product j on choice occasion τ and zero otherwise. The parameters are estimated using maximum likelihood estimator.

2.3.4 Estimated Coefficients and Elasticities

In Table 2.4 and 2.5, we summarize the estimated at-home and away-from-home preference parameters obtained by maximizing the likelihood function (equation (2.7)). The baseline group in the coefficient of α and η is SNAP participants without children with zero sugar

[31] In our estimation, we varied the choice of R and the results did not change qualitatively. The range of $\mathscr{M}_R$ is chosen to be of a similar magnitude to the price coefficients estimated from an initial regression of plain logit model without random coefficients.

intake.[32] The estimates of consumers' disutility of price and taste for SSB parameters reveal a large degree of heterogeneity across households, who differ by income, SNAP participation, the existence of children, and total sugar intake. More importantly, the patterns are very different at-home versus away-from-home. We describe how the parameters differ across households and between the two segments in details below.

Price Disutility For the at-home segment (Table 2.4), Non-SNAP low income and medium income households are even more price sensitive than the baseline households, as suggested by the interaction coefficients -0.074 and -0.038. In contrast, the positive interaction coefficient 0.072 of high income households implies that they are less price sensitive compared to the baseline group.

The picture is different for the away-from-home segment (Table 2.5). Nonparticipants low income households are slightly even more price sensitive than participants (-0.031). Medium and high income households are both less price sensitive than the baseline group (0.079 and 0.13). Households with children are less price sensitive both at-home (0.056) and away-from-home (0.026) than the baseline group.

There is a decreasing relationship between price sensitivity and household total sugar intake from SSB at-home. The difference in the marginal disutility of price between high and low sugar households is 0.21 (0.257 minus 0.044). In contrast, the difference becomes very small across low and high sugar taking households away-from-home (ranging from 0.13 to 0.147 for low and higher sugar households).

Taste for SSBs. First, we find that there is a decreasing relationship between the taste for SSB and household income, both at-home and away-from-home.[33] Second, comparing SNAP participants and nonparticipant poor, we find that the latter has weaker taste for SSB,

[32] We interact the coefficients of price and taste for SSB with four demographic variables. They are income and SNAP participation joint variable (SNAP participants, nonparticipants low income households, medium, and high income households), the age of the primary respondent, an indicator variable for households with children, and household total added sugar intake (from SSB) groups (zero sugar households, low, medium, and high sugar consuming households).

[33] This is suggested by the negative coefficients of the interaction between SSB and household income and SNAP participation status (-0.017 to -0.481 at-home and -0.093 to -0.164 away-from-home).

both at-home (-0.017) and away-from-home (-0.093). Lastly, SNAP participants turn out to have the strongest taste for SSB no matter at-home or away-from-home. This is consistent with the empirical evidence that SNAP participants buy a lot of SSB because SNAP benefits are allowed to be spent on SSB.[34]

There exists a positive linear relationship between the taste for SSB and household total sugar intake from SSB. The difference in the marginal utility from SSB between high and low sugar households is as large as 1.57. For the away-from-home segment, we find that across all sugar intake groups of households, they all have very large positive marginal utility from SSB (coefficients ranging from 4.889 to 5.548) compared to the at-home segment. High sugar consuming households still have slightly larger marginal utility of SSB compared to low sugar consumers.

Other Parameters of Interest We then look at the right panel of Table 2.4 and 2.5. Multi-pack has a negative impact on households marginal utility (-0.341) at-home. Package size has a positive effect (0.006) on households marginal utility at-home but a large negative effect away-from-home (-0.037). This finding is intuitive as we expect that consumers prefer small convenient package size on-the-go or in restaurants compared to in the grocery stores.

The nesting parameter λ is 1 in the at-home segment (Table 2.4). This implies that the nests are not significant and the demand model degenerates to a random coefficients logit model without nest structure. The interpretation is that when a specific drink product in the at-home segment (e.g., grocery store) becomes unavailable, the probabilities of choosing another drink product and choosing other grocery item such as meat and dairy would increase by the same proportion. The nesting parameter λ is 0.832 in the away-from-home segment (Table 2.5). The value implies that when consumers visit the away-from-home segment (e.g., restaurants), they are more likely to switch between drink products than to

[34]The article “In the Shopping Cart of a Food Stamp Household: Lots of Soda“ can be found on https://www.nytimes.com/2017/01/13/well/eat/food-stamp-snap-soda.html

switch from drink products to a dish.[35]

Table 2.4: Random Coefficients Nested Logit Demand Estimates, At-Home

	Estimate	SE		Estimate	SE
α			β		
Price	-0.285	0.011	Constant	-2.791	0.048
Price × non-SNAP poor	-0.074	0.003	Multi-pack	-0.341	0.024
Price × non-SNAP med	-0.038	0.002	Package size	0.006	0.000
Price × non-SNAP rich	0.072	0.04			
Price × age	-0.009	0.002			
Price × child	0.056	0.003	λ		
Price × sugar low	0.044	0.008	Nesting parameter	1.000	0.008
Price × sugar med	0.137	0.008			
Price × sugar high	0.257	0.008	μ	Weight	
			-0.1	0.74	
η			-0.078	0.24	
SSB × non-SNAP poor	-0.017	0.029	-0.056	0.01	
SSB × non-SNAP med	-0.226	0.023			
SSB × non-SNAP rich	-0.481	0.027	Drink category FE	Yes	
SSB × age	-0.032	0.008	Retailer-drink category FE	Yes	
SSB × child	0.284	0.027	Time FE	Yes	
SSB × sugar low	1.014	0.059			
SSB × sugar med	1.778	0.052			
SSB × sugar high	2.584	0.051	Number of choice occasions	23384	

Notes: We estimate demand on a sample of 4,412 households on 23,384 At-home choice occasions. Consumers choose from the products in At-home segments including the outside options. The reference group is SNAP households that consumed zero added sugar from SSB within a week. The coefficients of interaction between price and other demographic groups represent the change relative to the baseline level. Non-SNAP income group indicators are constructed by ERS, using household income measures and adjusted by poverty guidelines. The level of sugary diet is constructed based on weekly total added sugar from SSB a household has. We include a random coefficient for price. We report three μ^r with the highest estimated weights.

Elasticities We report aggregate price elasticities for the at-home and away-from-home segment by demographic groups in Table 2.13.[36]

We find substantial differences in elasticities across demographic groups and between the two segments. In Panel A of Table 2.13, SNAP participants are more elastic away-

[35] For more discussion on the demand estimates, please see Section 2.F in the Appendix.

[36] Specifically, we simulate a one percent increase in the price of all SSBs in any segment to predict SSB (own demand effect) and non-SSB purchases (cross demand effect). The elasticities are then calculated as the change in quantity demand divided by the price change. We also show the product-level price elasticity for the at-home and away-from-home segment in Tables 2.14 to 2.17 in Appendix Section 2.G.

Table 2.5: Random Coefficients Nested Logit Demand Estimates, Away-From-Home

	Estimate	SE		Estimate	SE
α			β		
Price	-0.515	0.031	Constant	-0.903	0.044
Price × non-SNAP poor	-0.031	0.007	Package size	-0.037	0.001
Price × non-SNAP med	0.079	0.006			
Price × non-SNAP rich	0.130	0.007			
Price × age	0.041	0.004			
Price × child	0.026	0.008	λ		
Price × sugar low	0.130	0.008	Nesting parameter	0.832	0.019
Price × sugar med	0.134	0.006			
Price × sugar high	0.147	0.008	μ	Weight	
			-0.1	0.79	
η			-0.011	0.10	
SSB × non-SNAP poor	-0.093	0.047	-0.078	0.05	
SSB × non-SNAP med	-0.020	0.035			
SSB × non-SNAP rich	-0.164	0.026	Drink category FE	Yes	
SSB × age	-0.099	0.008	Retailer-drink category FE	Yes	
SSB × child	0.089	0.024	Time FE	Yes	
SSB × sugar low	4.889	0.196			
SSB × sugar med	5.336	0.201			
SSB × sugar high	5.548	0.204	Number of choice occasions	23539	

Notes: We estimate demand on a sample of 3,977 households on 23,539 away-from-home choice occasions. Consumers choose from the products in away-from-home segments including the outside options. The reference group is SNAP households that consumed zero added sugar from SSB within a week. The coefficients of interaction between price and other demographic groups represent the change relative to the baseline level. Non-SNAP income group indicators are constructed by ERS, using household income measures and adjusted by poverty guidelines. The level of sugary diet is constructed based on weekly total added sugar from SSB a household has. We include a random coefficient for price. We report three μ^r with the highest estimated weights.

from-home while all other three non-SNAP poor, medium, and higher income households are more elastic at-home. In Panel B, households without children are more elastic at-home and are more elastic than household with children. In Panel C, households across all levels of sugar intake from SSB are more elastic at-home. The low sugar intake households are the most elastic while the high sugar intake households are the least elastic. This pattern is consistent with findings from Dubois, Griffith, and O'Connell 2020, who find that the soda taxes is less effective in targeting households with high total dietary sugar.

Discussion On one hand, our estimated elasticities for SSB at-home are of a similar magnitude to the existing literature on SSB demand with highly aggregated level of products. For example, Lopez and Fantuzzi 2012b estimate an own price elasticity of -0.58 for all caloric carbonated soft drinks (these are SSBs in our paper but excluding juice, energy drink, and sweetened tea and coffee). Andreyeva, Long, and Brownell 2010 collect own price elasticities for soft drink categories from 14 studies. The mean own price elasticities across the 14 studies is -0.79, with a 95% confidence interval of $[-0.33, -1.24]$.[37]

On the other hand, the demand for an aggregate-level category defined as in Table 2.13 here would in general be less elastic than the demand estimated from more disaggregated brand-level individual products. For example, Bonnet and Réquillart 2013b report a brand specific own price elasticities to be between -2.13 and -3.95. Dubé 2005b estimates a brand-level own price elasticities ranging between -2 and -4. The reason is that compared to a broadly defined category, brand-level products have more competition and substitution across each other.

2.4 Effects of a soda taxes on Sugar Intake

We use our demand estimates to simulate the introduction of a tax levied on SSB. We consider a tax rate of 1 cent per ounce. This is similar to the level of tax under the U.S. Soft Drinks Industry Levy.[38]

Let Ω_{SSB} denote the set of SSB products, r a soda taxes rate and q_j the volume in ounce. We assume the post-tax prices, p_j^{post} are given by

$$p_j^{post} = \begin{cases} p_j^{pre} + rq_j & \forall j \in \Omega_{SSB} \\ p_j^{pre} & \forall j \notin \Omega_{SSB} \end{cases}$$

We study the impact of the tax on household at-home and away-from-home sugar con-

[37] Their definition of soft drink categories are slightly different from ours. Their narrowest definition is carbonated soft drinks, and the broadest definition is non-alcoholic beverages.

[38] As of 2022, seven cities and counties in the United States have introduced an SSB taxes. The level of current excise taxes on SSB ranges from 1 to 2 cents per ounce, with five out of the seven cities being 1 cent per ounce.

sumption. Our main results assume 100 percent pass-through of soda taxes for both segments, given the evidence of almost 100 percent pass-through.[39] We also try setting the pass-through rate of soda taxes for the away-from-home segment to be 70 percent given the empirical finding in Cawley et al. 2021 using restaurants data in Boulder Colorado. They are reported in the Appendix. The counterfactual results are quite similar in both settings.

2.4.1 The Effectiveness of an SSB Tax

Our tax simulations suggest that consumers who purchase SSB will, on average, lower the amount of sugar they purchase from SSB at-home by 14.39g per week, away-from-home by 2.56 per week, and in total by 15.72g.[40] The average percentage reduction is 18.12 percent at-home and 5.75 percent away-from-home. The distribution of reductions in sugar in total is right skewed with the seventy-fifth, ninetieth, and ninety-fifth percentiles being 20.48g, 34.80g, and 47.61g.

An important aspect about the effectiveness of an SSB tax is whether it successfully achieves the reductions in sugar amongest the targeted groups of consumers: low-income households, in particular, SNAP participants and SNAP-eligible nonparticipants, households with children, and those with high total weekly dietary sugar. In Figure 2.4, we show how the effect of tax vary across these demographic characteristics. Panel (a) to (f) show how the mean reduction in sugar and the percentage reduction in sugar varies across SNAP status and income level, households with or without children, and total weekly dietary

[39]The literature that estimate SSB tax pass-through rate includes Cawley and Frisvold 2017, Grogger 2017, Berardi et al. 2016, Bergman and Hansen 2019, and Falbe et al. 2015. They tend to find that taxes are fully shifted to consumers, or even overshifted. The most recent papers like Cawley et al. 2021 find a pass-through rates of 71.1% on taxed drinks in Boulder, Colorado using hand-collected retail store data and 74.2% using restaurant data. Marinello et al. 2021 find similar price increases (82%) of bottled regular soda and diet soda in fast-food restaurants in Oakland, CA.

[40]Note that the total reduction in level (15.72g) is lower than the sum of the reduction at-home (14.39g) and the reduction away-from-home (2.56g). This is because the at-home, away-from-home, and the total segment is calculated based on a sample of 4412 households, 3977 households, and 4683 households respectively. There are some households with no purchase in one of the two segments, and hence for them the total reduction in SSB only comes from the other segment where they have made purchases.

sugar, separately for the at-home and away-from-home segment.

Panel (a), (c), and (e) show that the tax on sugary soft drinks achieves relatively large reductions in total sugar (at-home and away-from-home) among low-income households, households without children, and households with high weekly added sugar intake from SSB.

However, if we look at the at-home and away-from-home segments separately, we see very diverse patterns between the two segments. The reduction in sugar away-from-home is much smaller than that at-home, and it is small for all demographic groups. In other words, there is not much variation in the reduction in sugar away-from-home for all groups because the reduction is small: only around 3g per adult equivalent per week or 156g per adult equivalent per year. The percentage reduction in sugar is around 6 percent for all demographic groups (panel (a) and (b)).

The small variation in the effect of SSB tax on sugar intake away-from-home is supported by our previous descriptive findings in section 2.2.3. Households across income groups have similar share of SSB expenditures on-the-go. However, high-income households have much higher share in restaurants compared to low-income households. Meanwhile, the elasticity estimates in Table 2.13 show that high-income households are less price sensitive and hence their reduction in sugar in restaurants given the soda taxes is low. Both results lead to the finding here that the reduction in sugar of the high-income households is similar to that of the low-income households away-from-home.

Given the small impact of the tax on sugary soft drinks on the away-from-home segment, all variations in the reduction in total sugar (at-home and away-from-home) across demographic groups is driven by the variation of that in the at-home segment.

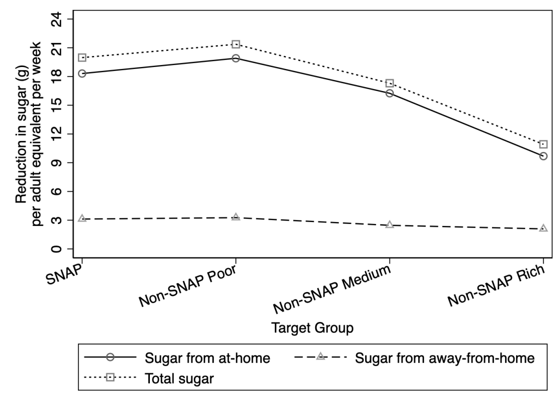
Low-income households are both more likely to be impacted by the policy and, conditional on this, have higher reductions in total SSB consumption than high-income households (panel (a)). The total reduction in sugar of SNAP and non-SNAP poor is around 21g per adult equivalent per week, which doubles the reduction in sugar of the high-income households. Panel (b) shows that while the average percent reduction in total sugar is slightly lower for the SNAP households (17 percent versus 23 percent across non-SNAP poor and medium-income households), this group obtains a relatively large amount of sugar from products targeted by the tax. This means their level reductions is large.

Panel (c) and (d) show that households without children exhibit higher reduction in both the average level, as well as the average percent in total sugar from SSBs. Our finding implies that the effectiveness of soda taxes for the young might not be as large as being found in Dubois, Griffith, and O'Connell 2020.

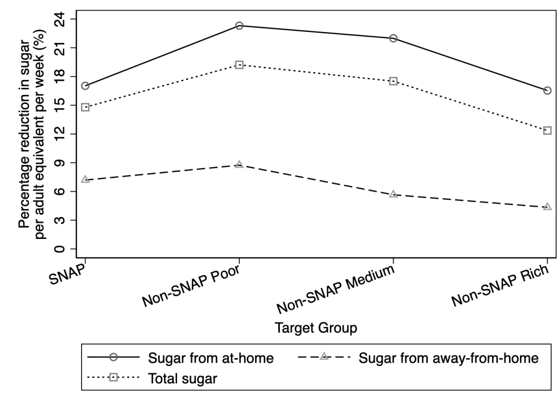
Panel (e) shows that the level reduction in total sugar is positively associated with household total sugar intake from SSBs. In particular, the difference in the reduction in total sugar between high and low total dietary sugar households is as large as 20g per adult equivalent per week.

We find that their response to soda taxes is smaller in percentage terms (panel (f)): for instance, the reduction for households with high decile of added sugar intake from SSB is over 14 percentage points below that for the low decile. They find that the reduction for the top decile of the dietary sugar distribution is over 4 percentage points below that for the bottom decile).

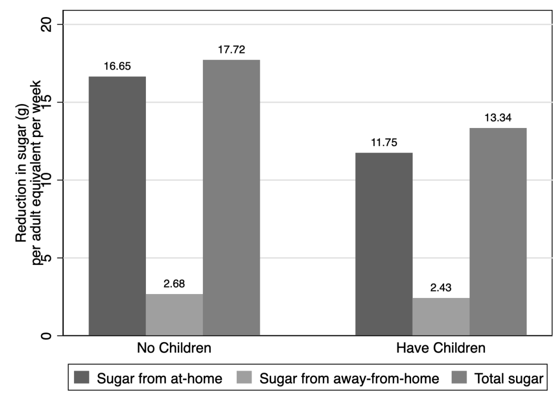
The difference in response across demographic groups can be supported by the pattern of preference variation. Even though the low-income households, those with children, and with high levels of sugar from SSB all have relatively strong SSB preferences, unlike the other groups those with children are less sensitive to prices. For more counterfactual results, please refer to Appendix section 2.H.



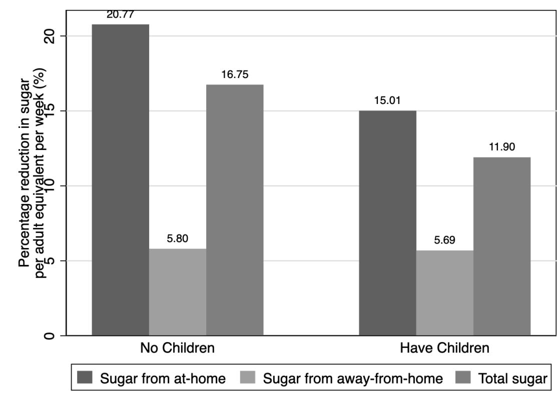
(a) By SNAP status and income level



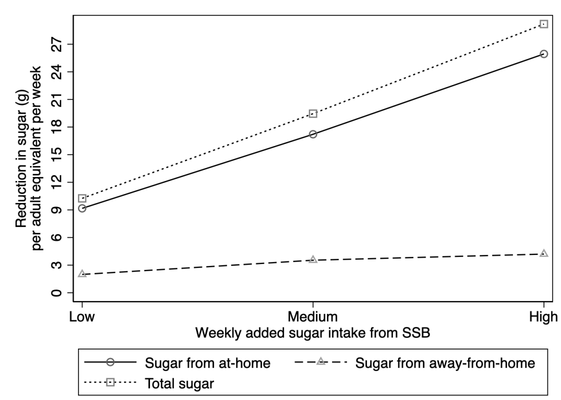
(b) By SNAP status and income level



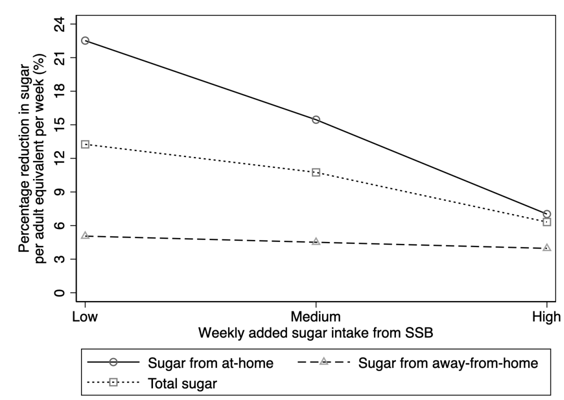
(c) By households with or without children



(d) By households with or without children



(e) By weekly added sugar intake



(f) By weekly added sugar intake

Figure 2.4: Reductions in Sugar From Drinks

Notes: The at-home segment is calculated based on 4,412 households in the sample, the away-from-home segment is based on 3,977 households, and the total is based on 4,683 households. Figures show how average reductions in SSB consumption varies across SNAP status and income groups, households with or without children and level of sugary diet. Figures (a), (c) and (e) show declines in level; figures (b), (d) and (f) show declines in percentage. In all figures, the pass-through rates is 100 percent in the at-home and away-from-home segments.

2.4.2 Consumer Welfare and Redistribution

The final question that we ask in this paper is: given the impact of SSB tax on household SSB demand, how will the tax affect consumer welfare? In particular, the tax will create an economic burden on consumers since it raises the price consumers pay. Moreover, with a higher price consumers can obtain less SSBs under the same total expenditure compared to under no tax regime. It is likely that for consumers who purchase SSBs, they will incur a welfare loss through this channel.

To answer this question, we use the standard Small and Rosen 1981 formula to calculate the compensating variation: the additional amount of money an individual would need to reach their initial utility following a change in prices. The compensating variation for a consumer i on choice occasion τ is given by

$$CV_{i\tau} = \frac{1}{\alpha_i}\left[\ln\left(\sum_{l=0}^{1}\left(\sum_{k\in C_l} e^{V_{ik\tau}^{post}(\mu_i)/\lambda}\right)^{\lambda}\right) - \ln\left(\sum_{l=0}^{1}\left(\sum_{k\in C_l} e^{V_{ik\tau}^{pre}(\mu_i)/\lambda}\right)^{\lambda}\right)\right], \quad (2.8)$$

where $V_{ik\tau}$ is defined in Section 2.3.1. We then integrated $CV_{i\tau}$ over the distribution of random coefficients and choice occasions to obtain the weekly (per adult equivalent) average compensating variation.

In Figure 2.5 panels (a)-(c) we describe how average compensating variation varies across SNAP status and income groups, households with or without children, and weekly added sugar intake. Compensating variation is determined by how exposed is the consumer to the tax (that is, whether the consumer buy a lot of the taxed goods) and how willing the consumer is to substitute towards other goods. In the at-home segment, low-income households and those with high weekly added sugar from SSB obtain more sugar from soft drinks and therefore are more exposed to the tax. Even without accounting for any behavioral effect, they would have higher compensating variation. After accounting for the behavioral effect, Figure 2.5 shows that the compensating variation remains high for low-income households and those with high weekly added sugar from SSB.

For the away-from-home segment, the picture is slightly reversed if we look at panel (a). The compensating variation is higher for higher-income households even though the difference is not as large as that in the at-home segment. Notice that even though Figure 2.4 panel (a) shows that the reductions in sugar from SSBs is similar across income groups in the away-from-home segment, the price of SSB is higher in this segment. In other words, higher-income households purchase more SSBs and more expensive SSBs away-from-home than poor households. That is why their compensating variation in the away-from-home segment is slightly higher than that of the poor households. Overall, the total compensating variation (accounting for both segments) is higher for low-income especially SNAP households. Panel (c) shows that high sugar diet households would have higher compensating variation in both the at-home and way-from-home segments. The difference in compensating variation between high and low sugar diet households is much larger in the at-home than the way-from-home segment. This finding is also consistent with previous result in Figure 2.4 panel (e) that the reduction in sugar is the highest for household with higher weekly added sugar intake from SSB, especially in the at-home segment.

The other thing to notice is that although the level of sugar reduction in Figure 2.4 is relatively low for the away-from-home segment, the associated compensating variation is somewhat comparable to that of the at-home segment. This is also due to the higher price of the away-from-home sugary drinks. In other words, for a similar level of economic burden, the tax in the away-from-home segment reduces sugar intake by a much smaller amount. For example, Figure 2.5 panel (c) shows that, in order to retrieve the utility before the tax, a medium sugar diet consumer would require a compensating variation of \$0.26 and \$0.23 for a 6 ounces and 1 ounce reduction in sugar from SSB for the at-home and away-from-home segment respectively. Furthermore, given that the SSB tax is based on volume, the benefits from tax revenue would be similar in the two segments. These imply that the welfare cost will be larger in the away-from-home segment than the at-home segment for a similar level of sugar reduction from SSB.

Even though the findings here suggest that the compensating variation is the highest among low-income households and those with high total dietary sugar, it does not imply that the total negative effect of the tax is the largest for these groups or the tax harms these groups the most. This is because consumers might purchase too much sugary soft drinks without considering the associated future costs (internality). Compensating variation only reflects part of the total consumer welfare effect of a sugary tax.

Policymakers are particularly concerned with low-income households. To provide an intuitive sense of the quantitative measure here, we find that in response to the tax consumers who have low income (regardless of SNAP participation), on average, reduce at-home sugar consumption from SSB by 21g per adult equivalent per week (annual: $21 \times 52 = 1092g$) and have average compensating variation of \$0.33 (annual: $\$0.33 \times 52 = \$17.16$). They reduce away-from-home sugar consumption from SSB by 3g per adult equivalent per week (annual: $3 \times 52 = 156g$) and have average compensating variation of \$0.15 (annual: $\$0.15 \times 52 = \$7.8$). The total reduction in sugar consumption from SSB is 24g per adult equivalent per week (annual: $24 \times 52 = 1248g$) and have average total compensating variation of \$0.43 (annual: $\$0.43 \times 52 = \$22.36$).

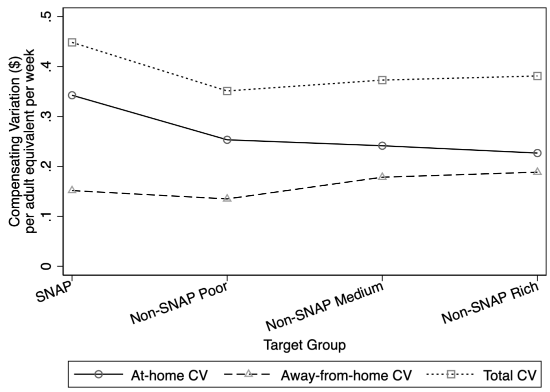
Following Dubois, Griffith, and O'Connell 2020, we use a typical sugary soft drink (a can of Coca-Cola), as our standard unit of comparison; a can of Coca-Cola in the United States is 12 oz (355 ml) and contains 35g of sugar. If we assume that consumers receive no benefits from the tax revenue raised, then this implies that the internality from a can of Coca-Cola would need to be at least \$0.55 (that is $17.16 * (35/1092)$) at-home and \$0.22 (that is $7.8 * (35/1248)$) away-from-home for this group on average to benefit from the tax. The value \$0.55 at-home is over 5 times larger than the average internality from sugar sweetened soft drinks estimated in Allcott, Lockwood, and Taubinsky 2019a.[41]

However, we have not accounted for any benefits from the tax revenue raised or any

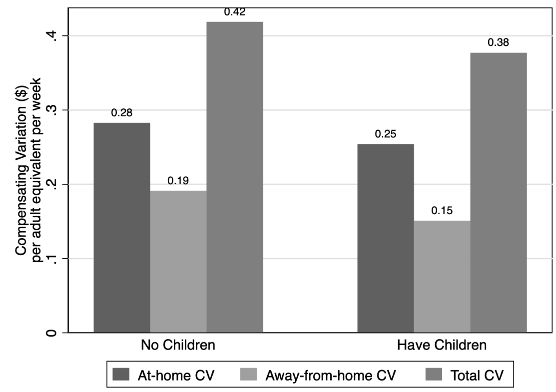
[41] The value found in Dubois, Griffith, and O'Connell 2020 for the on-the-go segment only is over 7 times larger than the average internality from sugar sweetened soft drinks estimated in Allcott, Lockwood, and Taubinsky 2019a.

savings from the averted externalities (for example, the health care costs). We could further compare the CV to the calculated tax revenue per consumer for each group. For example, on average, the current tax raises \$15.79 per adult equivalent over a year at-home.[42] The annual CV of SNAP consumers, \$17.68, is only slightly higher than the average tax revenue \$15.79. If the tax revenue can be distributed lump-sum back to the consumers, there only need to be a small positive amount of internality from the reduced soda consumption so that the SNAP participants can benefit from the tax. A similar calculation shows that the average tax revenue per consumer is \$7.32 away-from-home, which is very closed to the CV of SNAP consumers (\$7.8), implying a higher probability of the tax to be beneficial to them.

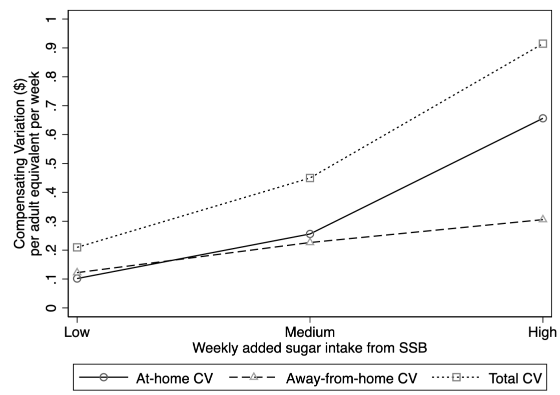
[42] Average tax revenue is calculated based on the post-tax volume consumption of SSB predicted by the model estimates and then multiplied by a tax rate of \$0.01 per ounce. The calculation method is similar to Dubois, Griffith, and O'Connell 2020.



(a) By SNAP status and income level



(b) By households with or without children



(c) By weekly added sugar intake

Figure 2.5: Revealed Consumer Welfare Effect

Notes: The at-home segment is calculated based on 4,412 households in the sample, the away-from-home segment is based on 3,977 households, and the total is based on 4,683 households. Figures show how average compensating variation varies across SNAP status and income groups, households with or without children and level of sugary diet. In all figures, the pass-through rate is 100 percent in the at-home and away-from-home segments.

Another concern about excise taxes is that they are regressive: lower-income households consume more of the taxed goods and hence bare more of the tax burden compared to higher-income households. Figure 2.1 confirms that, in the case of sugary soft drinks, low-income households are more likely to be soft drink purchasers and to get more sugar from these products; those in the bottom half of the distribution and who purchase soft drinks on average obtain 25 percent more sugar from these products. Notice that previous literature that only look at at-home or on-the-go segment along might overestimate this regressivity concern because low-income households have higher SSB shares at-home compared to high-income households. However, Figure 2.1 shows that high-income households obtain more SSB shares in restaurants and this finding mitigates the regressivity concern.

In sum, we find that the total CV is not largely different across income and SNAP participation group. Only SNAP households have 10 percent higher CV than other groups. In other words, we do not find the soda taxes to be regressive when we account for both at-home and away-from-home segments. Why is it the case that, low-income households obtain more total sugar (at-home plus away-from-home) from SSB, but their CV is not much larger than higher-income households? The reason is that higher-income households obtain more soft drinks and more expensive drinks away-from-home. Even though low-income households obtain more sugar from SSB especially at-home, but the tax is based on volume rather than the amount of sugar contained in each drink. In other words, low-income and high-income households might buy similar amount (ounces) of soft drinks across segments but the sugar amount in each drink is much higher for low-income households.

2.5 Conclusion

Beyond the focus on alcohol, tobacco, and gambling, sin taxes recently have been focused on the promotion of healthy eating. That is, the government has extended taxes to food and drinks. There is one main question related to assessing an effective an SSB tax. Who are most affected by a soda tax, and who bears the most of a soda taxes burden. This question is critical in assessing the effectiveness of an SSB tax in deterring excess levels of sugar intake and the welfare change of consumers due to a soda tax.

In this paper, we study the above questions by exploiting a novel dataset that cover household SSB demand from all channels (at-home, on-the-go, and in restaurants) for a representative sample of U.S. households. We utilize the rich demographic information on SNAP participation and eligibility and household income and composition, and non-parametrically estimate a flexible random coefficient nested logit model to document the heterogeneity in preferences and elasticity across household groups.

We find that preferences and elasticity vary with demographics in terms of SNAP participation, income, the existence of children, and the household total sugar from SSB. Such variation pattern is also different for the at-home and away-from-home segment. We find that soda taxes are less effective away-from-home and there is little variation in responses across households. In contrast, we find substantial variation in demand responses at-home across households. soda taxes are relatively effective at targeting the total sugar intake of the poor, those with high sugar consumption, and households without children for the at-home segment. Lastly, our results suggest that ignoring any segment will lead to biased policy implications on the targeting and effectiveness of soda tax.

We also find that, contrary to previous literature, the SSB tax is not highly regressive. The difference in compensating variation is much smaller than the difference in total sugar reduction across household income groups, especially when accounting for the away-from-home segment.

Firms will respond by adjusting their product types, pricing, advertising, the invention of new products, etc. Our results therefore only speak to the short to medium term effect of an SSB tax. Future research on incorporating the firm side responses into the picture will be very interesting and worth exploring.

Appendix

2.A FoodAPS Data Collection Process

Households in the FoodAPS data use scanners to scan all grocery purchases brought into the home. In both the food at-home and food away-from-home surveys we know what products (at the bar code, UPC level) were purchased, the product attributes, and the transaction price. We also observe information on the household and individual attributes, such as household size and composition, demographic characteristics, income, and participation in food assistance programs.

A screening interview determined whether the household at a sampled residence was eligible to participate in FoodAPS based on household income and SNAP participation. If eligible, the FoodAPS screener identified the main food shopper or meal planner in the household and invited him or her to participate in the week-long data collection.

The PR was asked to complete two in-person interviews and to call the study's telephone center for three brief telephone interviews regarding food acquisition events over the course of one week. The PR food book included both Blue pages to report details for "food at home" and Red pages to report "food away from home" acquisitions. The PR was responsible for recording food acquisitions by members under 11 years old.

Households were asked to scan barcodes on foods, save their receipts from stores and restaurants, and write information in their food books. For food-at-home acquisitions, the scanned barcodes were intended to be the primary source of item-level descriptions, while the receipts were intended to provide the price or expenditure information for each item. The Food Book (Blue) pages would provide the rest of the information and saved receipts would be used to verify this information and/or fill in missing information from the Blue page. For food-away-from-home acquisitions, the phone calls were intended to be the main source of item descriptions, details about the event, and price/expenditure information. The Red pages were reviewed to identify and capture any information that had not been reported during a phone call.

2.B Details of Data Construction

We measure the SSB consumption as the sum of all purchases of SSB in ounces during the data collection week. Similarly, we measure the SSB expenditure as the total spending on SSB during a seven-day period. We further construct the fraction of SSB expenditures spent at-home, in the restaurants, and on-the-go respectively, given by the SSB expenditures in each segment divided by the total SSB expenditures. The deciles of income are constructed using the household average monthly income.[43]

In our estimation, we separately estimate the at-home and away-from-home demand for soft drinks. We use information on the food at-home purchases of 4,412 households and food away-from-home purchase of 3,977 households. For either segment, we define a choice occasion as a trip in which a household makes a purchase of any good (including SSB drinks, non-SSB drinks or foods). When purchasing drinks for consumption at home, households choose a single item 51 percent of the time, whereas for consumption outside the home, 83 percent of the time the households choose a single item. On the remaining trips, the household chooses more than one type of drink product. In these cases, we treat each purchase in the multi-purchase transaction as a separate choice occasion. We observe households on an average of 9.7 choice occasions in our estimation sample. In total, the

[43] We did the same exercise using household *equivalized* income to account for household size and composition. It does not lead to significant difference.

sample contains 46,921 choice occasions. For over 95 percent of households we observe more than five choice occasions.

Product

The way we construct products is the following. We aggregate millions of UPCs (bar code) into ten product categories defined by the FoodAPS: soft drinks, fruit drinks, sport and energy drinks, sweetened coffee and tea, diet drinks, fruit and vegetable juice, unsweetened coffee and tea, flavored milk, flavored water, and water. By comparing the definition of drinks in the FoodAPS and the definition of taxable non-alcoholic drinks in the city and county websites in the U.S., we conclue that the current SSB tax is placed on the first four categories: soft drinks, fruit drinks, sport and energy drinks, and sweetened coffee and tea.[44]

We further classify whether a product is purchased in one of the two segments: at-home or away-from-home (on-the-go, or in restaurants). We also allow products to differ by packaging formats (regular, large, or multi-pack).[45] As a result, a product in the demand estimation is defined as either a category-package-segment, or a category-segment combination[46].

[44] Information regarding the taxable non-alcoholic beverages can be found in the city and county websites. See, e.g., `https://www.seattle.gov/license-and-tax-administration/business-license-tax/other-seattle-taxes/sweetened-beverage-tax.`.

[45] Regular size is defined as a container size smaller than 32 ounces. Large size is defined as a container size larger than 32 ounces.

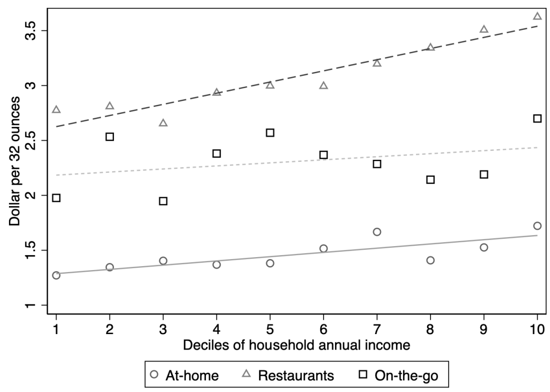
[46] As in Dubois, Griffith, and O'Connell 2020, for fruit juice, unsweetened coffee and tea, flavored milk and water, we aggregate across different sizes. In total, these non-SSB beverage categories account for less than 16 percent of the market.

Table 2.6: Products

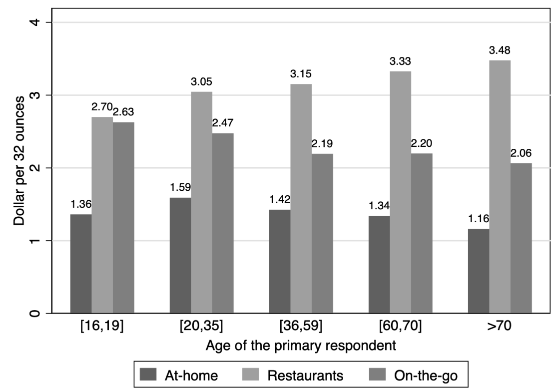
Product	Percentage	Price (dollar)	Product	Percentage	Price (dollar)
AH Regular Soft Drinks	1.387	1.457	AFH Regular Soft Drinks	9.409	1.505
AH Large-Bottle Soft Drinks	3.335	1.337	AFH Large-Bottle Soft Drinks	1.176	1.684
AH Pack Soft Drinks	3.097	3.507	AFH Fruit Drinks	1.703	2.055
AH Regular Fruit Drinks	1.473	1.294	AFH Large-Bottle Fruit Drinks	0.102	2.201
AH Large-Bottle Fruit Drinks	2.630	2.141	AFH Regular Sport and Energy Drinks	0.663	1.974
AH Pack Fruit Drinks	1.174	2.611	AFH Large-Bottle Sport and Energy Drinks	0.051	3.069
AH Regular Sport and Energy Drinks	1.552	1.171	AFH Regular Sweetened Coffee and Tea	4.288	2.097
AH Pack Sport and Energy Drinks	0.460	6.406	AFH Large-Bottle Sweetened Coffee and Tea	0.258	1.682
AH Regular Sweetened Coffee and Tea	0.431	1.624	AFH Regular Diet Drinks	2.421	1.506
AH Large-Bottle Sweetened Coffee and Tea	0.810	2.485	AFH Large-Bottle Diet Drinks	0.318	1.528
AH Pack Sweetened Coffee and Tea	0.104	5.000	AFH Fruit and Vegetable Juice	0.825	1.703
AH Regular Diet Drinks	0.603	1.359	AFH Unsweetened Coffee and Tea	4.245	1.666
AH Large-Bottle Diet Drinks	1.147	1.486	AFH Flavored Milk	1.023	2.095
AH Pack Diet Drinks	0.961	3.970	AFH Flavored and Enhanced Water	0.151	1.728
AH Fruit and Vegetable Juice	3.331	3.064	AFH Water	2.231	1.245
AH Unsweetened Coffee and Tea	0.311	1.997			
AH Flavored Milk	0.654	2.471	AH Outside Options	22.292	0.000
AH Flavored and Enhanced Water	1.485	1.439	AFH Outside Options	21.300	0.000
AH Water	2.596	2.506	Total Number of Choice Occasions	46921	46921

Notes: At-home segment is abbreviated as AH. Away-from-home segment is abbreviated as AFH. Regular size is defined as a container size smaller than 32 ounces. Large size is defined as a container size larger than 32 ounces. Multi-pack is defined as a pack with more than one unit of bottles/cans. AH (AFH) outside options refer to any foods or drinks except for nonalcoholic bevarages that are obtained from AH (AFH) segment. The second column shows the percentage of choice occasions where the indicated product is purchased, based on transactions made by the 4,683 households in the estimation sample. Prices are averaged across choice occasions.

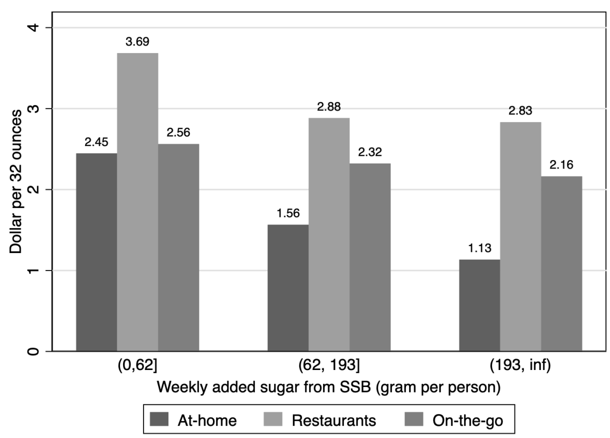
Prices



(a) By deciles of household annual income



(b) By age of the primary respondent



(c) By total added sugar from SSB

Figure 2.6: Average price of SSB

Notes: The figure shows how purchase price varies across households with or without children, age groups, and total added sugar from SSB. In plot (a), the household annual income is equivalized by the OECD-modified equivalence scale. In plot (b), age groups are classified according to the same cutoffs as in the FoodAPS dataset. Plot (c) is restricted to households who have positive amount of added sugar from SSB. The cutoff levels are the terciles of the total sugar from SSB.

Package Size

For a given UPC, the multi-pack information is not provided in the FoodAPS data. That is, we do not observe how many of those goods appear in a given pack (e.g. a 6 pack of soda). The *pkgsize* variable is the multiplication of the size of individual packaging and the number of individual packaging. Thus it measures the size of an item defined by a given UPC. There is another variable *quantity* in the item level data, which indicates how many of that item is purchased.

First, we restrict the sample to households who purchased a single item. Sometimes the *pkgsize* and *quantity* information is inconsistent with the UPC information. For example, an item is a 20 oz bottle-8 counts according to the UPC, so the correct item-level size should be $20*8$ oz. However, in the FoodAPS data, some of them are incorrectly recorded as $pkgsize = 160$ and $quantity = 8$, yielding a size of $160*8$ oz. It is hard to identify items with wrong size information like this. Therefore we eliminate all transactions with $quantity > 1$. The fraction of households buying multiple units of a given item is small (around 5%). By doing so, we abstract away multi-units purchasing behavior. Note that we did not eliminate multi-pack items. For example, we allow for the households to buy a 6 pack of coke, but we don't allow for 6 bottles of single-bottle coke.

Second, we look up the UPC code to recover the multi-pack information using the *pkgsize* variable. Almost always, a given *pkgsize* corresponds to a unique combination of number of goods in a pack and the size of the individual packaging good. Thus we are able to identify whether the item is a multi-pack product and the size of each single bottle using the *pkgsize* variable.

Third, near 15% of SSB transactions have missing package size information. Only less than 10% of the missing package size information might be imputed using UPC code. There are two things we need to impute: (a) package size, i.e., size of the item in ounces; (b) package type, i.e., regular single-bottle, large single-bottle, or multi-pack. We impute the missing package size by linear interpolation of package size on transaction price. This calculation is performed separately for each drink category within a given month. Products with imputed weight less than or equal to 32 ounces are classified as regular-size single-bottle. Products with imputed weight less than or equal to 32 ounces are either large single-bottle or multi-pack. In order to impute the missing information of package type, we take the subsample of observations that are either large single-bottle or multi-pack with non-missing information on the package type and fit a random forest classification model of package type (either large-bottled or multi-pack) on package size in ounce and transaction price. Based on the estimated model, we predict the package type for the set of observations with missing values.

Composition

Table 2.7 shows the average share of non-alcoholic beverage expenditures allocated at each type of products. For SSB purchased at-home, the top three purchased products, measured by shares of total beverage expenditures, are soft drinks (26.7%), fruit and vegetable juice (18.1%), and fruit drinks (17.7%). For SSB purchased away-from-home, the top three purchased products are soft drinks (42.6%), sweetened coffee and tea (18.9%), and unsweetened coffee and tea (12.7%). In total, SSB (the first four types) accounts for 54.4% of beverage expenditures in at-home market segment and it accounts for 72.4% in away-from-home purchases.

Tables 2.8-2.11 show the share of beverage expenditures in different products at-home or away-from-home by demographic groups. We find that different demographic groups have very different basket of goods under at-home and away-from-home categories. For example, Table 2.8 shows that households with no children purchase more diet drinks at-home and unsweetened coffee and tea away-from-home than households with children. This suggests that households with children have stronger taste for sugary drinks. Table 2.9 shows that demand differ by age, and different patterns for at-home and away-from-home products. For at-home segment, younger households purchase more soft drinks and fruit drinks at-home while older households purchase more fruit and vegetable juice. For away-from-home segment, all households like soft drinks but older households don't purchase much fruit drinks as younger households and tend to drink more unsweetened coffee and tea. Table 2.10 shows that households with different level of sugary diet also have different baskets. The higher the total added sugar from SSB a household has, the more soft drinks and fruit drinks this household purchases. Instead, low sugary diet households purchase more diet drinks and water at-home. They also purchase more sweetened and unsweetened coffee and tea away-from-home than high sugary diet households.

Table 2.7: Average Share of Non-alcoholic Beverage Expenditures Allocated at Different Products

	At-home	Away-from-home
Soft drinks	0.267	0.426
Fruit drinks	0.177	0.083
Sport and energy drinks	0.054	0.026
Sweetened coffee and tea	0.046	0.189
Diet drinks	0.098	0.065
Fruit and vegetable juice	0.181	0.022
Unsweetened coffee and tea	0.010	0.127
Flavored milk	0.029	0.036
Flavored and enhanced water	0.034	0.004
Water	0.105	0.022

Notes: The table shows how share of non-alcoholic beverage expenditures are allocated at ten product types for at-home and away-from-home, respectively. The expenditure shares are averaged across all households.

Table 2.8: Average Share of Non-alcoholic Beverage Expenditures Allocated at Different Products, by Households with or without Children

	At-home		Away-from-home	
	No Children	Have Children	No Children	Have Children
Soft drinks	0.241	0.290	0.386	0.465
Fruit drinks	0.149	0.203	0.072	0.093
Sport and energy drinks	0.044	0.064	0.022	0.029
Sweetened coffee and tea	0.052	0.040	0.197	0.181
Diet drinks	0.119	0.078	0.082	0.048
Fruit and vegetable juice	0.208	0.155	0.020	0.025
Unsweetened coffee and tea	0.013	0.007	0.166	0.091
Flavored milk	0.026	0.032	0.033	0.039
Flavored and enhanced water	0.042	0.026	0.003	0.006
Water	0.105	0.105	0.020	0.024

Notes: The table shows how share of non-alcoholic beverage expenditures are allocated at ten product types for at-home and away-from-home, respectively. The expenditure shares are averaged across households with or without children.

Table 2.10: Average Share of Non-alcoholic Beverage Expenditures Allocated at Different Products, by Total Added Sugar from SSB

	At-home			Away-from-home		
Total added sugar level	Low	Medium	High	Low	Medium	High
Soft drinks	0.168	0.268	0.409	0.436	0.472	0.492
Fruit drinks	0.155	0.208	0.214	0.089	0.094	0.088
Sport and energy drinks	0.052	0.075	0.057	0.020	0.032	0.032
Sweetened coffee and tea	0.041	0.044	0.064	0.233	0.195	0.189
Diet drinks	0.142	0.074	0.053	0.052	0.047	0.048
Fruit and vegetable juice	0.210	0.172	0.087	0.020	0.017	0.017
Unsweetened coffee and tea	0.011	0.011	0.008	0.101	0.089	0.073
Flavored milk	0.046	0.022	0.018	0.030	0.030	0.036
Flavored and enhanced water	0.048	0.025	0.022	0.003	0.003	0.006
Water	0.128	0.101	0.067	0.015	0.019	0.018

Notes: The table shows how share of non-alcoholic beverage expenditures are allocated at ten product types for at-home and away-from-home, respectively. The expenditure shares are averaged across households of different levels of total added sugar from SSB.

Table 2.11: Average Share of Non-alcoholic Beverage Expenditures Allocated at Different Products, by Household Annual Income

	At-home			Away-from-home		
Income level	Low	Medium	High	Low	Medium	High
Soft drinks	0.339	0.276	0.192	0.492	0.439	0.369
Fruit drinks	0.167	0.182	0.179	0.081	0.080	0.087
Sport and energy drinks	0.048	0.054	0.060	0.025	0.029	0.023
Sweetened coffee and tea	0.043	0.047	0.046	0.152	0.189	0.212
Diet drinks	0.074	0.095	0.122	0.055	0.064	0.072
Fruit and vegetable juice	0.161	0.182	0.196	0.033	0.019	0.019
Unsweetened coffee and tea	0.010	0.009	0.010	0.108	0.113	0.155
Flavored milk	0.027	0.023	0.040	0.029	0.040	0.036
Flavored and enhanced water	0.020	0.031	0.048	0.005	0.005	0.003
Water	0.109	0.101	0.107	0.020	0.022	0.023

Notes: The table shows how share of non-alcoholic beverage expenditures are allocated at ten product types for at-home and away-from-home, respectively. The expenditure shares are averaged across households of different income groups.

Table 2.9: Average Share of Non-alcoholic Beverage Expenditures Allocated at Different Products, by Age of the Primary Respondent

	At-home					Away-from-home				
Age group	[16,19]	[20,35]	[36,59]	[60,70]	>70	[16,19]	[20,35]	[36,59]	[60,70]	>70
Soft drinks	0.333	0.274	0.282	0.228	0.190	0.400	0.477	0.419	0.360	0.313
Fruit drinks	0.307	0.197	0.178	0.121	0.168	0.145	0.086	0.089	0.057	0.053
Sport and energy drinks	0.024	0.064	0.056	0.037	0.030	0.045	0.038	0.024	0.007	0.001
Sweetened coffee and tea	0.008	0.048	0.043	0.060	0.035	0.220	0.180	0.193	0.194	0.184
Diet drinks	0.029	0.077	0.099	0.135	0.129	0.035	0.042	0.070	0.094	0.102
Fruit and vegetable juice	0.171	0.172	0.158	0.242	0.271	0.044	0.020	0.024	0.020	0.021
Unsweetened coffee and tea	0.003	0.012	0.009	0.013	0.002	0.085	0.095	0.112	0.220	0.270
Flavored milk	0.000	0.030	0.029	0.025	0.037	0.013	0.032	0.041	0.028	0.039
Flavored and enhanced water	0.042	0.025	0.038	0.040	0.028	0.000	0.004	0.006	0.002	0.001
Water	0.083	0.102	0.109	0.099	0.110	0.014	0.025	0.022	0.018	0.016

Notes: The table shows how share of non-alcoholic beverage expenditures are allocated at ten product types for at-home and away-from-home, respectively. The expenditure shares are averaged across households of different age groups.

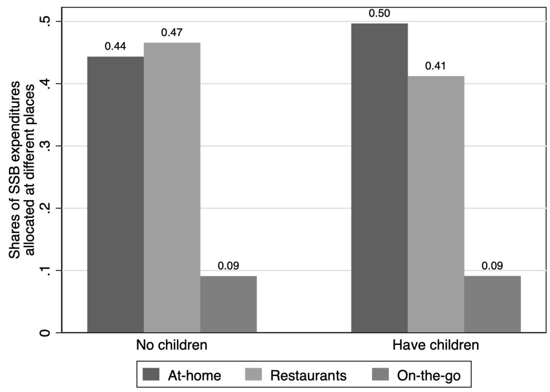
2.C Age, Existence of Children, and Overall Added Sugar

Figure 2.7 shows the average share of SSB expenditures at-home, on-the-go, and in restaurants by whether households have children, the age of the primary respondent, and overall added sugar from SSB.

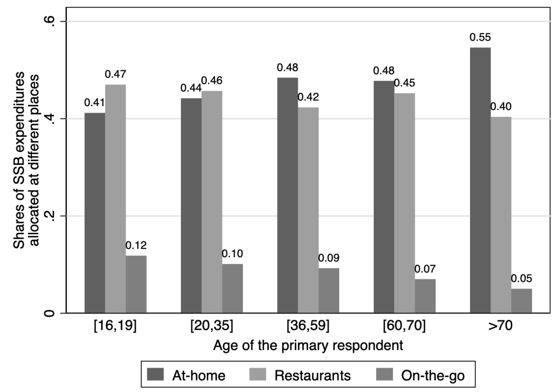
First, compared with households without children, households with children spend slightly larger shares (50%) of SSB at-home and slightly smaller shares (41%) of SSB in restaurants. Both groups spend 9% of SSB expenditure on-the-go. Second, there is an increasing trend of the age of the primary respondent with respect to SSB shares at-home, while a decreasing trend of that with respect to SSB shares on-the-go. There is barely any significant relationship between SSB shares in restaurants and primary respondent's age. Lastly, on average, households with a higher sugar intake are more likely to spend SSB budget shares at home while less likely to spend SSB shares in restaurants. Combined with previous evidence that lower-income households consume more at home than in restaurants, the finding here simply reflects that poor households spend larger expenditure shares at home, eat less healthy diet, tend to have higher sugar intake, and are more likely to purchase SSB at home. There is little variation in SSB purchases on-the-go across groups of overall added sugar intake.

2.D Price Construction in Demand Estimation

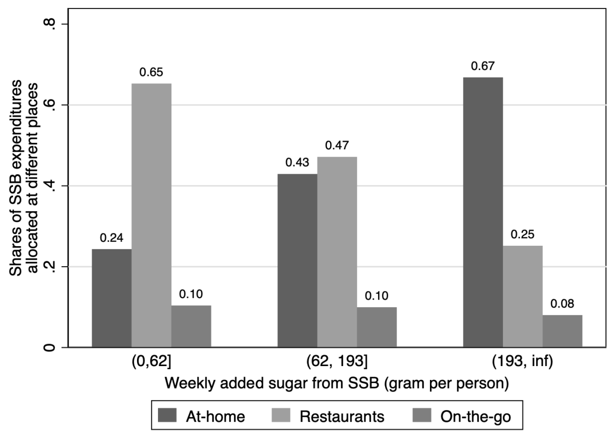
We construct the average price p_j for each product type j defined in Table 2.6. We average the transaction prices of all UPCs belonging to a particular product type in a given month in a specific retailer across all consumers. More specifically, each product type contains $\mathscr{K}_j$ UPCs indexed by $k = 1, ..., \mathscr{K}_j$. We denote the price paid by consumer i for each UPC at time t in a retailer r as $p_{kj}(i, r(\tau), t(\tau))$ and the price index for each product type j as $p_{jr(\tau)t(\tau)}$. We will discuss the notation in more detail in Section 2.3. The resulting price for each product type is a product-month-retailer tuple. For example, assume the total number



(a) By households with or without children



(b) By age of the primary respondent



(c) By total added sugar from SSB

Figure 2.7: Average share of SSB expenditures allocated at-home, on-the-go, and in restaurants

Notes: The figure shows how average share of SSB expenditures allocated at different places varies across households with or without children, age groups, and total added sugar from SSB. In plot (b), age groups are classified according to the same cutoffs as in the FoodAPS dataset. Plot (c) is restricted to households who have positive amount of added sugar from SSB. The cutoff levels are the terciles of the total sugar from SSB.

of transactions under product type j happened at time t in a retailer r is $\mathscr{T}_{jr(\tau)t(\tau)}$, we compute the product price index as

$$p_{jr(\tau)t(\tau)} = \frac{1}{\mathscr{T}_{jrt}} \sum_{\tau=1}^{\mathscr{T}_{jrt}} \sum_{k=1}^{\mathscr{K}_j} p_{kj}(i, r(\tau), t(\tau))$$

If there are no transactions happened for a product-month-retailer tuple, we impute the price using the average price of the same product in the same retailer in last month. If a product-retailer type involves no transactions in all time periods, we treat the price as a missing value. For example, purchases of large-bottle fruit drinks are never observed in drinking places. In the later section where we introduce the demand model, we will assume that products with missing prices are not included in consumers' choice set. In the previous example, this assumption implies that if a consumer visits a drinking place, large-bottle fruit drinks are not on the menu. We present details of data construction and how missing values are dealt with in Appendix Section 2.B.

A typical problem faced by researchers in discrete choice demand estimation is that the prices of products not chosen by the consumers are not observed. Using the mean prices to proxy for these unobserved prices may induce measurement error problem. These errors are "Berkson" errors (Berkson 1950). Schennach 2013 proposes a solution to fix the measurement error in prices in continuous demand models. Methods that can be applied to the discrete choice framework is an interesting research question that is worth future research.

2.E Separability Test

We consider two sets of separability tests. First, we include the average price of away-from-home soda in the at-home demand equations and vice versa. Second, we include a dummy for whether there were any away-from-home SSB purchases, and vice versa. More specifically, we estimate

$$q_i^{\text{AH}} = x_i'\beta + \gamma p_i^{\text{AFH}} + \varepsilon_i,$$

and

$$q_i^{\text{AH}} = x_i'\beta + \alpha d_i^{\text{AFH}} + \varepsilon_i,$$

and similar for the away-from-home demand equations. q_i^{AH} is the volume consumption of SSB at-home for household i, p_i^{AFH} is the average prices of away-from-home soda, and d_i^{AFH} is a dummy variable that equals one if there were any away-from-home purchases of SSB. Endogeneity might arise because households that demand more soda at-home might have unobserved characteristics that also cause them to purchase soda away from home, or visit places that offer low (or high) price soda. Due to the short time period of the data, it is infeasible to include a full set of household fixed effects to control for the unobservables. Instead, We include a rich set of household characteristics, x_i, to deal with potential endogeneity. x_i include the constant term, household size, whether the household has children, income, age of primary respondent, average BMI, diet status, and knowledge of nutrition.

Table 2.12: Separability Tests

	AH Demand	AFH Demand	AH Demand	AFH Demand
At-home price	-98.90***	0.892	-87.79***	
	(8.410)	(2.295)	(6.581)	
Away-from-home price	2.135	-12.52***		-13.05***
	(5.911)	(1.613)		(1.188)
At-home soda purchased				2.884
				(3.823)
Away-from-home soda purchased			19.99	
			(16.44)	
Child	59.60*	11.58	51.43**	8.334
	(23.46)	(6.401)	(19.55)	(4.788)
Household size	26.37***	8.649***	31.51***	9.769***
	(7.053)	(1.925)	(5.877)	(1.507)
Age	-0.535	0.0298	-0.371	-0.0623
	(0.633)	(0.173)	(0.499)	(0.123)
BMI	5.118***	0.535	4.540***	0.616*
	(1.296)	(0.354)	(1.069)	(0.272)
Constant	240.0**	51.26*	161.3**	51.43***
	(74.91)	(20.45)	(57.40)	(14.85)
Income group	Yes	Yes	Yes	Yes
Diet status	Yes	Yes	Yes	Yes
Nutrition fact	Yes	Yes	Yes	Yes

Notes: Standard errors in parentheses. Estimates of categorical variables are omitted from the table. * $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

In Table 2.12, we report our results. For example, column 1 shows that the estimated value of γ is 2.135, with a large standard error 5.911, suggesting that the price of away-from-home soda has insignificant effect on the demand at home. Similarly, column 3 shows that whether there were any away-from-home soda purchases has no significant impact on the amount of soda consumed at home ($\hat{\alpha} = 19.99$ and standard error is 16.44). The results supporting separability here are consistent with other results in the literature.

2.F Discussion on Demand Estimates

The price coefficients we obtained in both segments are quite small in magnitude compared to prior work.[47] However, this can be explained by two facts. The first is that the nesting parameter rescales all demand parameters. Grigolon and Verboven 2014 find that to make the price coefficient in a nested logit model comparable to the ones from standard logit model, the price parameter should be rescaled by α/λ. Thus, with a nesting parameter less than 1, the rescaled price coefficient should be larger in magnitude than those reported in Table 2.4 and 2.5. The second main reason is due to our definition of "products". Instead of choosing among specific brands or narrowly defined products, we define "products" in our analysis as drink categories such as regular soft drinks, large-bottle diet soft drinks, a pack of juice drink, etc. Furthermore, the magnitude of estimated coefficients varies depending on the unit of the variables. Therefore, it would be more meaningful to consider price elasticities rather than parameter values.

2.G Additional Results of Price Elasticity

In this section we show product level elasticities. Tables 2.14 and 2.15 report the demand change for alternative drink options resulting from a 1% price increase of each product. Tables 2.16 and 2.17 provide the full matrix of own- and cross-price elasticities for all products. For example, Tables 2.14 shows that a 1% increases in price of the regular sized soft drink category would result in a reduction of demand of 0.25% while demand for non-SSB drinks would rise by 0.012%.

[47] Using supermarket data in a developing country, Marshall 2015b reports an average marginal (dis)utility of price of -6.15. Dubois, Griffith, and O'Connell 2020 uses data in UK covering on-the-go purchases to study soda demand. Their estimate of mean level of price preference is -3.15,

Table 2.13: Aggregate-Level Price Elasticity

	Effect of 1 percent increase in the price of SSB on			
	Own demand effect	Cross demand for non-SSB	Own demand effect	Cross demand for non-SSB
	At-Home		Away-From-Home	
Panel A: SNAP status and income level				
SNAP	-0.267	0.130	-0.335	0.197
Non-SNAP poor	-0.473	0.161	-0.396	0.202
Non-SNAP medium	-0.374	0.126	-0.233	0.136
Non-SNAP rich	-0.233	0.063	-0.194	0.089
Panel B: Households with or without children				
No children	-0.348	0.093	-0.248	0.098
Have children	-0.244	0.111	-0.243	0.150
Panel C: High vs low sugar consumers				
Low	-0.620	0.091	-0.278	0.116
Medium	-0.393	0.137	-0.249	0.171
High	-0.163	0.119	-0.206	0.178

Notes: We simulate the effect of a 1 percent price increase for all SSB belonging to the at-home market segment and all SSB belonging to the away-from-home segment, respectively. In column 1, we report the change in demand for those products. In column 2 we report the effect on demand for non-SSB products. Elasticities are computed separately by demographic groups and are averaged across markets.

Table 2.14: At-Home Product Level Price Elasticities

	Own	FAH SSB	FAH non-SSB
Soft Drinks	-0.247	0.014	0.012
Large-Bottle Soft Drinks	-0.235	0.018	0.016
Pack Soft Drinks	-0.467	0.031	0.025
Fruit Drinks	-0.245	0.008	0.007
Large-Bottle Fruit Drinks	-0.341	0.016	0.013
Pack Fruit Drinks	-0.488	0.009	0.008
Sport and Energy Drinks	-0.239	0.005	0.005
Pack Sport and Energy Drinks	-0.875	0.011	0.008
Sweetened Coffee and Tea	-0.252	0.002	0.002
Large-Bottle Sweetened Coffee and Tea	-0.427	0.005	0.005
Pack Sweetened Coffee and Tea	-0.619	0.005	0.004
Diet Drinks	-0.368	0.003	0.005
Large-Bottle Diet Drinks	-0.390	0.004	0.007
Pack Diet Drinks	-0.854	0.007	0.010
Fruit and Vegetable Juice	-0.702	0.027	0.040
Unsweetened Coffee and Tea	-0.447	0.001	0.003
Flavored Milk	-0.545	0.004	0.006
Flavored and Enhanced Water	-0.377	0.005	0.008
Water	-0.559	0.022	0.032

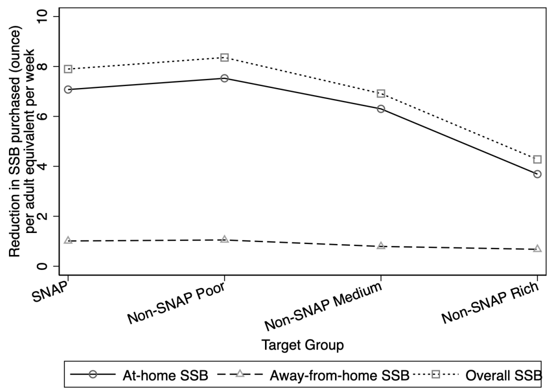
Notes: For each of the products we compute the change in demand for that product, for other SSB alternatives and for non-SSB alternatives resulting from a 1% price increase. Numbers are averaged across time and place types.

Table 2.15: Away-From-Home Product Level Price Elasticities

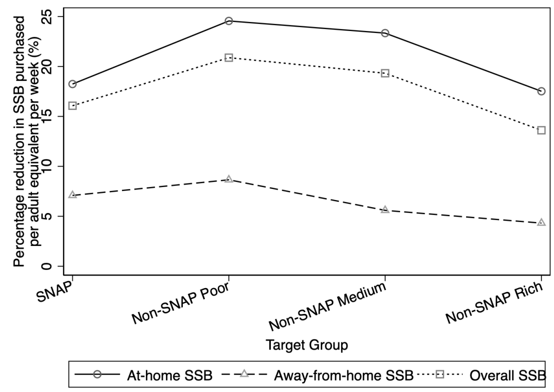
	Own	FAFH SSB	FAFH non-SSB
FAFH Soft Drinks	-0.277	0.065	0.049
FAFH Large-Bottle Soft Drinks	-0.374	0.016	0.011
FAFH Fruit Drinks	-0.442	0.016	0.011
FAFH Large-Bottle Fruit Drinks	-0.557	0.004	0.003
FAFH Sport and Energy Drinks	-0.423	0.011	0.008
FAFH Large-Bottle Sport and Energy Drinks	-0.738	0.003	0.002
FAFH Sweetened Coffee and Tea	-0.410	0.048	0.034
FAFH Large-Bottle Sweetened Coffee and Tea	-0.382	0.005	0.004
FAFH Diet Drinks	-0.364	0.020	0.027
FAFH Large-Bottle Diet Drinks	-0.396	0.004	0.005
FAFH Fruit and Vegetable Juice	-0.443	0.006	0.008
FAFH Unsweetened Coffee and Tea	-0.348	0.033	0.044
FAFH Flavored Milk	-0.431	0.006	0.008
FAFH Flavored and Enhanced Water	-0.426	0.002	0.003
FAFH Water	-0.343	0.012	0.017

Notes: For each of the products we compute the change in demand for that product, for other SSB alternatives and for non-SSB alternatives resulting from a 1% price increase. Numbers are averaged across time and place types.

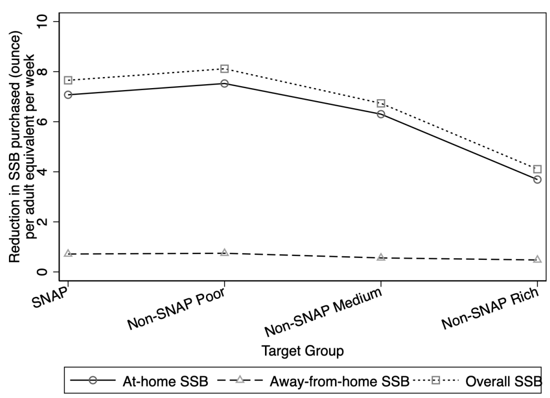
2.H Additional Counterfactual Results



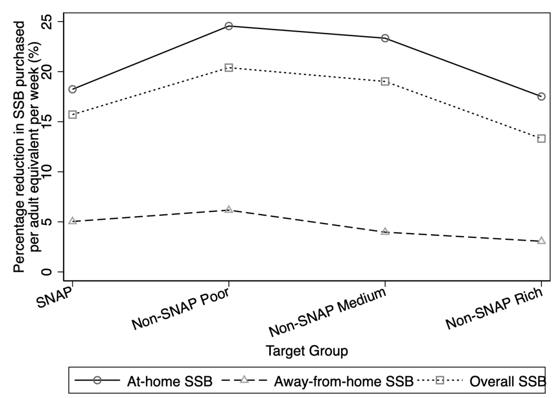
(a) Level reductions, at-home 100% pass-through, away-from-home 100% pass-through



(b) Percentage reductions, at-home 100% pass-through, away-from-home 100% pass-through



(c) Level reductions, at-home 100% pass-through, away-from-home 70% pass-through



(d) Percentage reductions, at-home 100% pass-through, away-from-home 70% pass-through

Figure 2.8: Reductions in SSB Purchases by SNAP Status and Income Level

Notes: Figures show how average reductions in SSB consumption varies across SNAP status and income groups. In all figures, the pass-through rates is 100 percent in the at-home segment. Figures (a) and (b) show the results with 100 percent pass-through rates in away-from-home, and (c) and (d) are 70 percent pass-through rates in away-from-home.

Table 2.16: At-Home Product Level Own- and Cross-Price Elasticities

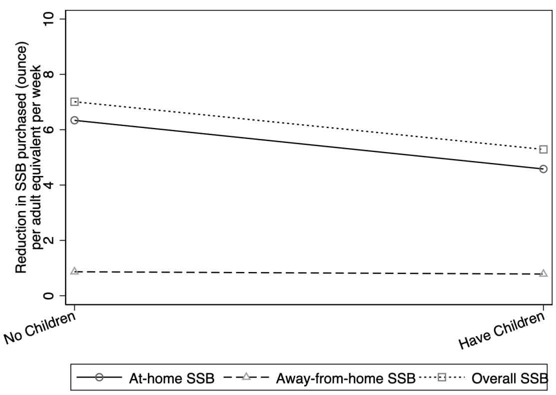
	Soft	L Soft	P Soft	Fruit	L Fruit	P Fruit	Sport	P Sport	Sw Coffee	L Sw Coffee	P Sw Coffee	Diet	L Diet	P Diet	Juice	Unsw Coffee	Milk	Flav Water
Soft	-0.247	0.014	0.014	0.014	0.014	0.015	0.014	0.013	0.014	0.014	0.011	0.012	0.012	0.013	0.012	0.012	0.012	0.012
L Soft	0.018	-0.235	0.018	0.018	0.018	0.018	0.018	0.017	0.018	0.018	0.014	0.016	0.016	0.016	0.016	0.016	0.016	0.015
P Soft	0.031	0.031	-0.467	0.031	0.031	0.033	0.031	0.029	0.031	0.031	0.028	0.024	0.024	0.028	0.025	0.024	0.024	0.024
Fruit	0.008	0.008	0.008	-0.245	0.008	0.008	0.009	0.008	0.008	0.008	0.009	0.007	0.007	0.007	0.007	0.007	0.007	0.007
L Fruit	0.016	0.016	0.016	0.016	-0.341	0.015	0.016	0.015	0.016	0.016	0.017	0.013	0.013	0.013	0.013	0.013	0.013	0.013
P Fruit	0.010	0.009	0.009	0.009	0.009	-0.488	0.010	0.009	0.009	0.010	0.010	0.007	0.007	0.010	0.008	0.007	0.007	0.007
Sport	0.006	0.006	0.005	0.006	0.006	0.006	-0.239	0.005	0.006	0.006	0.004	0.005	0.005	0.006	0.005	0.005	0.005	0.005
P Sport	0.011	0.011	0.011	0.011	0.011	0.012	0.011	-0.875	0.011	0.011	0.012	0.007	0.007	0.009	0.008	0.008	0.008	0.007
Sw Coffee	0.002	0.002	0.002	0.002	0.002	0.003	0.002	0.002	-0.252	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002
L Sw Coffee	0.005	0.006	0.005	0.006	0.006	0.006	0.006	0.005	0.006	-0.427	0.004	0.004	0.005	0.005	0.005	0.005	0.005	0.004
P Sw Coffee	0.005	0.005	0.005	0.005	0.005	0.004	0.005	0.005	0.005	0.005	-0.619	0.003	0.003	0.003	0.004	0.004	0.004	0.003
Diet	0.004	0.003	0.003	0.004	0.003	0.004	0.004	0.003	0.004	0.003	0.002	-0.368	0.006	0.006	0.005	0.006	0.006	0.006
L Diet	0.005	0.005	0.004	0.005	0.005	0.005	0.005	0.004	0.005	0.005	0.004	0.008	-0.390	0.008	0.007	0.008	0.008	0.008
P Diet	0.008	0.008	0.007	0.008	0.007	0.009	0.008	0.006	0.008	0.007	0.006	0.011	0.011	-0.854	0.010	0.011	0.010	0.011
Juice	0.028	0.028	0.026	0.028	0.027	0.027	0.029	0.022	0.028	0.027	0.028	0.042	0.042	0.039	-0.702	0.041	0.040	0.042
Unsw Coffee	0.002	0.002	0.001	0.002	0.002	0.002	0.002	0.001	0.002	0.002	0.001	0.003	0.003	0.002	0.002	-0.447	0.003	0.003
Milk	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.003	0.004	0.004	0.004	0.007	0.007	0.007	0.006	0.007	-0.545	0.007
Flavored Water	0.005	0.005	0.004	0.005	0.005	0.005	0.005	0.004	0.005	0.005	0.005	0.008	0.008	0.009	0.007	0.008	0.008	-0.377
Water	0.024	0.024	0.022	0.024	0.023	0.022	0.024	0.018	0.024	0.022	0.021	0.035	0.035	0.031	0.032	0.034	0.033	0.035

Notes: For each of the at-home products we compute the own- and cross-price elasticities. Numbers are averaged across time and place types.

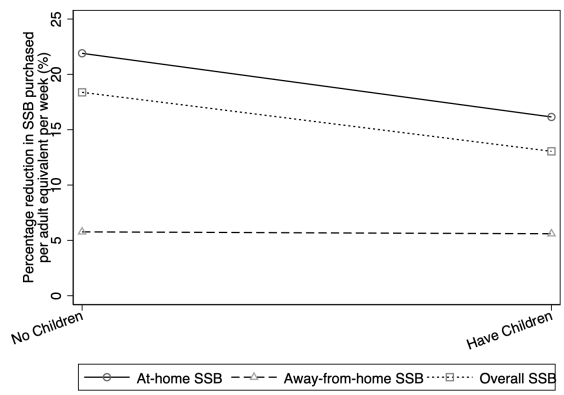
Table 2.17: Away-From-Home Product Level Own- and Cross-Price Elasticities

	Soft	L Soft	Fruit	L Fruit	Sport	L Sport	Sw Coffee	L Sw Coffee	Diet	L Diet	Juice	Unsw Coffee	Milk	Flav Water	Water
Soft	-0.277	0.067	0.066	0.066	0.066	0.065	0.066	0.075	0.050	0.055	0.049	0.048	0.051	0.050	0.050
L Soft	0.016	-0.374	0.015	0.018	0.016	0.013	0.016	0.020	0.012	0.014	0.013	0.011	0.013	0.012	0.013
Fruit	0.016	0.016	-0.442	0.015	0.016	0.012	0.016	0.015	0.011	0.010	0.011	0.011	0.011	0.011	0.011
L Fruit	0.004	0.005	0.004	-0.557	0.004	0.002	0.004	0.004	0.003	0.003	0.003	0.003	0.004	0.003	0.003
Sport	0.011	0.012	0.011	0.012	-0.423	0.020	0.011	0.013	0.008	0.010	0.009	0.008	0.009	0.008	0.009
L Sport	0.003	0.003	0.003	0.004	0.003	-0.738	0.003	0.004	0.002	0.003	0.002	0.002	0.003	0.002	0.002
Sweetened Coffee	0.048	0.051	0.046	0.055	0.047	0.018	-0.410	0.024	0.038	0.018	0.036	0.034	0.040	0.037	0.036
L Sweetened Coffee	0.005	0.006	0.005	0.006	0.005	0.005	0.005	-0.382	0.004	0.005	0.004	0.004	0.004	0.004	0.004
Diet	0.020	0.014	0.020	0.016	0.020	0.019	0.020	0.016	-0.364	0.021	0.019	0.027	0.022	0.030	0.019
L Diet	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.005	0.005	-0.396	0.005	0.005	0.005	0.005	0.005
Juice	0.006	0.007	0.006	0.006	0.006	0.006	0.006	0.007	0.007	0.010	-0.443	0.009	0.008	0.007	0.010
Unsweetened Coffee	0.034	0.036	0.032	0.037	0.033	0.025	0.033	0.024	0.046	0.032	0.045	-0.348	0.048	0.044	0.047
Milk	0.006	0.007	0.006	0.008	0.006	0.004	0.006	0.005	0.009	0.007	0.009	0.008	-0.431	0.009	0.009
Flavored Water	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.001	0.003	0.002	0.002	0.003	0.002	-0.426	0.002
Water	0.013	0.014	0.012	0.010	0.012	0.005	0.012	0.014	0.011	0.019	0.019	0.017	0.013	0.011	-0.343

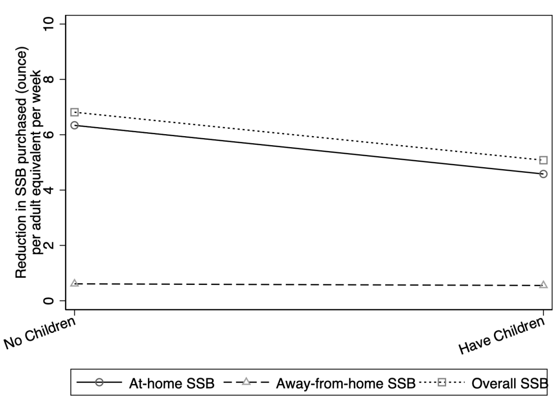
Notes: For each of the away-from-home products we compute the own- and cross-price elasticities. Numbers are averaged across time and place types.



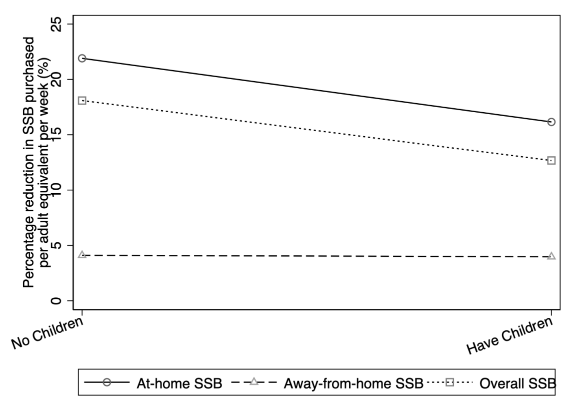
(a) Level reductions, at-home 100% pass-through, away-from-home 100% pass-through



(b) Percentage reductions, at-home 100% pass-through, away-from-home 100% pass-through



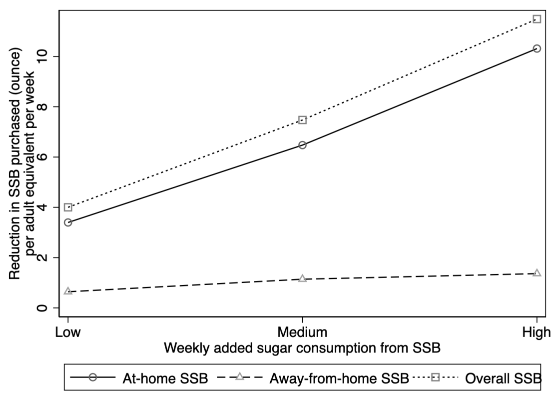
(c) Level reductions, at-home 100% pass-through, away-from-home 70% pass-through



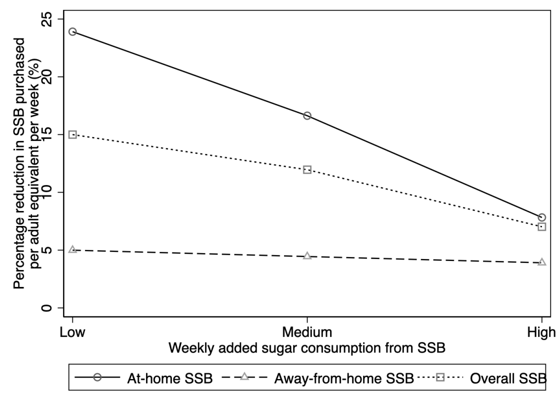
(d) Percentage reductions, at-home 100% pass-through, away-from-home 70% pass-through

Figure 2.9: Reductions in SSB Purchases by Households with or without Children

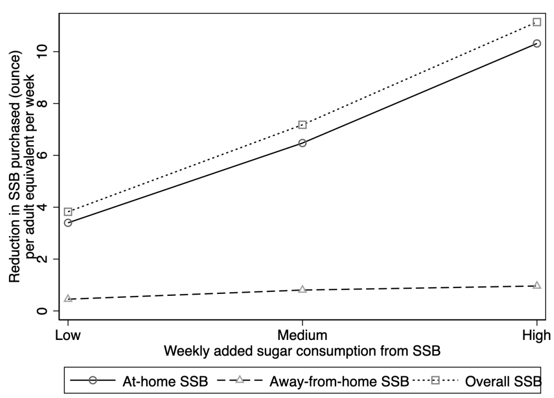
Notes: Figures show how average reductions in SSB consumption varies across households with or without children. In all figures, the pass-through rates is 100 percent in the at-home segment. Figures (a) and (b) show the results with 100 percent pass-through rates in away-from-home, and (c) and (d) are 70 percent pass-through rates in away-from-home.



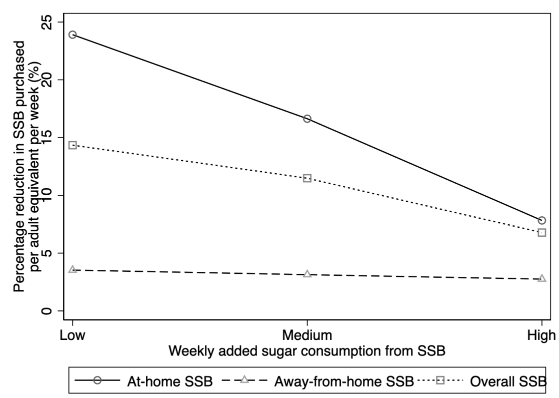
(a) Level reductions, at-home 100% pass-through, away-from-home 100% pass-through



(b) Percentage reductions, at-home 100% pass-through, away-from-home 100% pass-through



(c) Level reductions, at-home 100% pass-through, away-from-home 70% pass-through



(d) Percentage reductions, at-home 100% pass-through, away-from-home 70% pass-through

Figure 2.10: Reductions in SSB Purchases by Weekly Added Sugar Intake

Notes: Figures show how average reductions in SSB consumption varies across level of sugary diet. In all figures, the pass-through rates is 100 percent in the at-home segment. Figures (a) and (b) show the results with 100 percent pass-through rates in away-from-home, and (c) and (d) are 70 percent pass-through rates in away-from-home.

Chapter 3

Identification of a Triangular Two Equation System Without Instruments

ARTHUR LEWBEL, SUSANNE M. SCHENNACH, AND LINQI ZHANG

3.1 Introduction

Consider a standard linear triangular structural model

$$Y = X'b_1 + \varepsilon_1 \tag{3.1}$$

$$W = \gamma Y + X'b_2 + \varepsilon_2 \tag{3.2}$$

for some endogenous variables Y and W, exogenous covariates X, and unobserved errors ε_1 and ε_2. For example, W could be a worker's wages or earnings and Y could be her level of schooling. Or, as in our later empirical application, W could be a country's GDP growth and Y a health measure like growth in life expectancy. The primary goal is identification of γ, the direct causal effect of Y on W, though we will also obtain identification of b_1, b_2, and the joint distribution of the errors.[1]

[1] Throughout this paper we focus on the traditional homogeneous effects model where γ is a constant, rather than a heterogeneous treatment effects model.

The main obstacle to identification and estimation of γ is that ε_1 and ε_2 may be correlated, because both depend on a common unobserved U (ability in the case of schooling and wages, technology in the case of GDP and health). That is, in its simplest form,

$$\varepsilon_1 = U + V \quad \text{and} \quad \varepsilon_2 = \beta U + R \tag{3.3}$$

where U, V, and R are unobserved, mutually independent (conditional on X) random variables and β is a constant. After projecting off covariates X, the V and R errors represent idiosyncratic shocks to Y and W, while U is what makes Y an endogenous regressor in the W equation.

Similar triangular structural models arise whenever we have one variable Y affecting another variable W, and a common unobservable that affects them both. For example, consider a two period dynamic model with autocorrelated errors. In this case W equals Y in a subsequent time period, and U represents the autocorrelation in the errors. Another example is production, where W could be a firm's value-added output per unit of capital, Y is the firm's labor per unit of capital, and U is unobserved entrepreneurship, which affects both productivity and the chosen level of inputs.

Such models are traditionally identified in econometrics by finding an instrument, i.e., a variable that correlates with Y but not ε_2, or equivalently, a variable that correlates with V but not U or R. However, such instruments can be difficult to find. For example, Card (1995), Card (2001) and others propose using measures of access to schooling, such as distance to or cost of colleges in one's area, as wage equation instruments, while others raise objections to the validity of these instruments, e.g. Carneiro and Heckman (2002). Other wage equation instruments may raise fewer questions of validity but can be weak, like Angrist and Krueger (1991) and Angrist and Krueger (2001) quarter of birth instruments.

Similarly, Acemoglu and Johnson (2007) propose using changes in predicted mortality, constructed based on innovations in health care, as an instrument for life expectancy growth Y in their regression of GDP growth W on Y. However, such health innovations could be correlated with other technological advances that increase GDP, leading to instrument in-validity. Comparable questions can be raised regarding the instruments or identifying side information in other similar studies, such as Aghion, Howitt, and Murtin (2010), who find a positive γ, in contrast to Acemoglu and Johnson (2007) negative γ. Ecevit (2013) summarizes results from eleven similar studies, finding estimates of γ that range from strongly negative to insignificant to strongly positive. This range of estimates raises serious questions regarding the validity of instruments or other side information that different authors use to identify γ.[2]

Rather than propose any new instrument, we address the more fundamental question of whether and when this model can be point identified and estimated *without* side information such as instruments whose validity can be hard to ascertain (noting that the alternative of a randomized experiment is not feasible for a macro question like this). If so, then we can estimate the model without relying on side information, and/or test the validity of side information like instruments via overidentification tests.

We provide conditions for point identification of the model

$$Y = U + V \tag{3.4}$$

$$W = \gamma Y + \beta U + R \tag{3.5}$$

with U, V, and R being unobserved, mutually independent random variables with unknown distributions. The same identification theorem can then be applied conditioning on covariates X, to show point identification of more general models, where the entire distributions of U, V, and R could depend nonparametrically on X. A special case of this general identification result is then identification of equations (3.1), (3.2) and (3.3). In this special case, variables V and R that depend nonparametrically on X in equations (3.4) and (3.5) are instead replaced with $X'\beta_1 + V$ and $X'\beta_2 + R$, where these new V and R do not depend on

[2] Of course, differences are also due to variation in data sets and in how Y and W are defined and constructed. As another way to explain these differing results, Cervellati and Sunde (2011) suggest that the true effect might be non-monotonic.

X.[3]

Our main result is surprising: under minimal regularity assumptions, the coefficients γ and β, and the distributions of U, V, and R (and b_1 and b_2 in that model) are all point identified without instruments or other side information, unless either U or V is normally distributed (after appropriately conditioning on or projecting off covariates X). So, for example, Y having bounded support would be a sufficient condition for point identification, since that would rule out normality of U or V.

In addition to proving this general identification result, we also: 1. Provide a few low order moments yielding simple GMM estimators of the model, 2. Show how infinitely many additional moments conditions can be systematically constructed to provide identification under weaker conditions, 3. Provide the sharp identified set for the coefficients γ and β in the case where either U or V is normal and hence point identification fails, 4. Investigate the behavior of these GMM estimators in some Monte Carlo exercises, and 5. Provide an empirical application where we establish that our identification and estimation strategy is viable even with a very small sample size. Specifically, we estimate the Acemoglu and Johnson (2007) model without using any instruments, and obtain estimates that are very similar to what they found with their instrument.

Instrumental variables estimation of the model has the advantage that it only requires assumptions regarding first and second moments of the covariates, errors, and instruments. In contrast, our assumptions regarding U, V, and R are, implicitly, restrictions on all moments. However, there are a number of mitigating factors. First, some of our results, such as Lemma 3 below, only rely on lower order moments. Second, our main theorem works via convolutions, and so our independence assumptions can be relaxed to subindependence, as defined and described in Schennach (2019), who points out that subindependence is arguably as weak as a conditional mean assumption in terms of the dimensionality

[3]More generally, U, V and R could be heteroskedastic, or otherwise have higher moments that depend in unknown ways on X, but estimation would then become more complicated. One possibility would be replacing the GMM estimators we provide with conditional moment GMM, conditioning on X. More simply, heteroskedasticity could be parameterized, with parameters estimated as part of the GMM.

of the restrictions imposed. Third, our independence assumption is actually conditional on other covariates, so, e.g., the identification can handle arbitrary heteroskedasticity and dependence of higher moments on regressors. Similarly, if, e.g., U is ability, then identification only requires ability to be conditionally (sub)independent from other unobserved factors, conditional on covariates. Nevertheless, given our required assumptions, these results should be most useful when instruments either don't exist, or might be invalid.

The identification of equations (3.4) and (3.5) without instruments has been previously considered by Rigobon (2003), Klein and Vella (2010), and Lewbel (2012), but these results neither nest nor are nested by ours because they *require* that the errors be heteroskedastic, and identification is obtained by imposing varying restrictions on the structure of that heteroskedasticity.[4]

A number of special cases of our results do appear in the literature, but all of them assume $\gamma = 0$, and so they omit the most important feature of the model in applications like ours. Kotlarski (1967) is the special case of our model where it is known that $\gamma = 0$ and $\beta = 1$, and in that case Kotlarski's Lemma shows that point identification of the distribution of all the latent variables holds even under normality. Similarly, Reiersøl (1950) uses a special case of our model where it is known that $\gamma = 0$ and Y plays the role of a measurement of U contaminated by an error V and establishes conditions under which β would be identified. As noted in Lewbel (2022), with $\gamma = 0$ and Reiersøl's identification of β, one could rewrite Reiersøl's model as $Y = U + V$ and $W/\beta = U + R/\beta$, and then apply Kotlarski's lemma

[4]Rigobon (2003) and Klein and Vella (2010) impose different parametric restrictions on the error variances, while Lewbel (2012) imposes a nonparametric restriction. For simplicity we assume homoskedastic errors, but by conditioning our identification theorems on X, we could allow for general heteroskedastity as well, at the expense of likely weaker identification and more complicated estimators.

to the joint distribution of Y and W/β to identify the distributions of U, V, and R.[5]

Our results, showing necessary and sufficient conditions to identify the more general model of equations (3.4) and (3.5) with unknown nonzero γ, turns out to be a difficult extension. In particular, the methods of proof used by Reiersøl (1950) and Kotlarski (1967) do not extend to our problem. The proof of our main result instead relies on similar tools as Khatri and Rao (1972) or Rao (1966) and Rao (1971) (see also Comon (1994) reference to Darmois (1953)).

Some limitations of our results should be acknowledged upfront. We assume that the coefficients γ and β are constants. So, e.g., our results do not immediately extend to random coefficients, such as treatment effects with unobserved heterogeneity, or to nonlinearity in the dependence of W on Y. However, this limitation may be mitigated to some extent by allowing the distributions of the unobservables to be unknown functions of covariates. Another important restriction on our results is that we require U to be a scalar. While this is a common assumption (as in the examples cited earlier), there are other situations where one might expect a vector of unobservable shocks like U to affect both Y and W, and our identification results would then not apply. We provide examples in Supplement D. Finally, a limitation for empirical work is that our estimators depend on higher than second moments of the data, and such moments can lead to very imprecise estimates when sample sizes are small.

In section 3.2, we provide a few simple moments that will often suffice to point identify our model, and can be used to construct a correspondingly simple GMM estimator. In section 3.3, we present our general identification results, including constructing more

[5] A special case of non-normality is when the components U and V are asymmetric. Lewbel (1997) and Erickson and Whited (2002) exploit asymmetry to construct simple estimators for the Reiersøl (1950) model. See also Bierens (2012). Other papers propose estimators for models like equations (3.4) and (3.5) with $\gamma = 0$, by assuming that coefficients like β are point identified using higher moments, but without explicitly characterizing when that is possible. Examples include Bonhomme and Robin (2010), Fruehwirth, Navarro, and Takahashi (2016), and Navarro and Zhou (2017). A related result, showing identification of direction of causality in models under nonnormality, is Peters, Janzing, and Schölkopf (2017). Generalizations of Kotlarski's lemma to models with more components (but again still assuming $\gamma = 0$) include Székely and Rao (2000) and Li and Zheng (2020). A nonlinear extension of Reiersøl (1950) is Schennach and Hu (2013).

moments like those in section 3.2, and showing that, with minimal regularity, the model is point identified as long as both U and V are not normal. In sections 3.4 and 3.5 we derive the sharp identified set when either U or V is normal, and derive some inequalities regarding our model relative to ordinary least squares. Section 3.6 provides a Monte Carlo analysis of our simple GMM estimators. In section 3.7 we provide an empirical application based on Acemoglu and Johnson (2007), in which we obtain estimates comparable to theirs, without using their (or any other) instrument. Section 3.8 concludes with some suggestions for further work.

3.2 Simple Identification and Estimation

We begin with a simple special case of our general results, by providing some moments that can easily be used to identify and estimate (by standard GMM) the models described in the introduction. These results are not as general as our main identification theorem, but are likely to suffice for many empirical applications.

We first consider identification and estimation of equations (3.4) and (3.5) without covariates X, and then we extend the results to equations (3.1) and (3.2).

Assumption 5. *We observe the joint distribution of two real valued, nondegenerate random variables Y and W.*

With data, we could assume independent, identically distributed observations of Y and W, and then identify their joint distribution to satisfy Assumption 5 using the Glivenko Cantelli theorem.

Assumption 6. *The unobserved real valued random variables U, V, and R are mean zero and mutually independent,*[6] *with unknown distributions.*

Assumption 7. *R has finite variance, and U and V each have finite fourth moments.*

[6]Independence can be weakened to subindependence (Schennach (2019)).

Assumption 8. *The unknown constants γ and β are real valued, finite, and $\beta > 0$.*

We can assume our data Y and W have been demeaned, rationalizing the assumption that the unobservables have mean zero. To see why we need a sign restriction on β, observe that we can rearrange equations (3.4) and (3.5) to get $W = (\gamma + \beta)Y - \beta V + R$, which, except for the sign of β, is observationally equivalent to the original model, switching the roles of V and U. Usually, the sign of β should be clear from the economics of the application, e.g., in a returns to schooling model, $\beta > 0$ is a natural assumption, since it says that unobserved ability that increases (decreases) education outcomes will increase (decrease) wages. If we instead believed β was negative, we could just replace Y with $-Y$ everywhere to make β positive (redefining γ, U, and V accordingly).

We also rule out $\beta = 0$, because if $\beta = 0$ then it would be pointless to separately identify V and U. Moreover, having $\beta = 0$ is nonsensical in the types of applications we consider, since it would mean that Y is exogenous, making identification and estimation of γ trivial.

Substituting equation (3.4) into equation (3.5) gives the reduced form expression for W

$$W = \gamma V + \alpha U + R \quad \text{with} \quad \alpha = \gamma + \beta \tag{3.6}$$

The following Lemma provides two moments that can often suffice to point identify γ and α, which then trivially also point identifies β.

Lemma 3. *Let Assumptions 5-8 and equations (3.4) and (3.5) (and therefore also equation 3.6) hold. Then*

$$E\left[(W - \gamma Y)(W - \alpha Y)Y\right] = 0 \tag{3.7}$$

$$cov\left[(W - \gamma Y)(W - \alpha Y), Y^2\right] - 2E\left(WY - \gamma Y^2\right)E\left(WY - \alpha Y^2\right) = 0 \tag{3.8}$$

Proofs are all in Supplement A. The proof of Lemma 3 works by substituting $W - \gamma Y = \beta U + R$ and $W - \alpha Y = -\beta V + R$ into equations (3.7) and (3.8), and then uses the mutual independence of U, V, and R to verify that these equations hold.

Lemma 3 provides two equations in the two unknowns α and γ. If we solve the first equation for α and substitute that into the second, we obtain a quadratic in γ. The sign restriction that $\beta > 0$ then determines which root is the correct one for γ.

We later provide the formal conditions under which these two equations suffice to point identify α and γ. The main condition, derived in Theorem 7 below, is equation (3.21). Equation (3.21) shows that the main cases in which equations (3.7) and (3.8) by themselves fail to provide point identification are when U and V have the exact same distribution, or when both are symmetrically distributed, or if either U or V is normally distributed. We later show that infinitely many additional equations in α, γ, Y and W can be constructed, based on higher moments of Y and W than those used in Lemma 3. These higher moments can help identify α and γ in applications where Lemma 3 does not suffice.

A simple estimator for α and β can be constructed by rewriting equations (3.7) and (3.8) as moment conditions, and applying standard method of moments or GMM. One can immediately check that these equations take the form

$$E\left(YW-\mu_{yw}\right)=0, \qquad E(Y^2-\mu_{yy})=0 \tag{3.9}$$

$$E\left[\left(W-\gamma Y\right)\left(W-\left(\gamma+\beta\right)Y\right)Y\right]=0 \tag{3.10}$$

$$E\left[\left(W-\gamma Y\right)\left(W-\left(\gamma+\beta\right)Y\right)\left(Y^2-\mu_{yy}\right)-2\left(\mu_{yw}-\gamma\mu_{yy}\right)\left(W-\left(\gamma+\beta\right)Y\right)Y\right]=0 \tag{3.11}$$

where $\mu_{yw}=E\left(YW\right)$ and $\mu_{yy}=E\left(Y^2\right)$. The parameters μ_{yw} and μ_{yy} are estimated along with γ and β by putting equations (3.9), (3.10), and (3.11) into any standard GMM estimation routine. One could replace β with e^b in these equations to impose the sign restriction that $\beta > 0$.

Lemma 3 uses up to fourth moments of the data. Based on results derived in the next section, in Supplement B we provide additional equations (using up to fifth moments) that can provide overidentification of γ and β, or point identification in some cases where Lemma 3 does not suffice.

Let σ_U^2, σ_V^2, and σ_R^2 denote the variances of the error components U, V, and R. It may be of economic interest to estimate these variances, to identify how much of the variance of the model errors is due to unobserved ability U versus the idiosyncratic components V and R. From the model we have $E\left((W-\gamma Y)Y\right) = \beta\sigma_U^2$, $E\left(Y^2\right) = \sigma_U^2 + \sigma_V^2$, and $E\left((W-\gamma Y)^2\right) = \beta^2\sigma_U^2 + \sigma_R^2$, which implies

$$\sigma_U^2 = E\left((W-\gamma Y)Y\right)/\beta, \quad \sigma_V^2 = E\left(Y^2\right) - \sigma_U^2, \quad \sigma_R^2 = E\left((W-\gamma Y)^2\right) - \beta^2\sigma_U^2 \quad (3.12)$$

Given estimates of β and γ, we can replace the expectations in equation (3.12) with sample averages to estimate these variances.

Alternatively, we can estimate these variances jointly with the model parameters by observing that

$$\mu_{yy} = \sigma_U^2 + \sigma_V^2, \quad \mu_{yw} = \beta\sigma_U^2 + \gamma\left(\sigma_U^2 + \sigma_V^2\right). \quad (3.13)$$

So, in equations (3.9), (3.10), and (3.11) we can replace μ_{yy} and μ_{yw} with their expressions in equation (3.13), and apply GMM using those equations along with the additional equation

$$E\left((W-\gamma Y)^2 - \beta^2\sigma_U^2 - \sigma_R^2\right) = 0 \quad (3.14)$$

to simultaneously estimate β, γ, σ_U^2, σ_V^2, and σ_R^2. We can further replace σ_U^2 with $\sigma_U^2 = e^{\tau_U}$ and similarly for σ_V^2 and σ_R^2, to impose the constraint that variances are positive. See Supplement B for details on these moments.

Higher moments of U, V, and R can be estimated analogously. Alternatively, as discussed later, once we have have identified and estimated β and γ, we can apply Kotlarski's Lemma to recover the entire distributions of U, V, and R.

We can also easily extend this identification and associated estimation to allow for covariates. Suppose we have the model

$$Y = b_1'X + U + V \quad (3.15)$$

$$W = \gamma Y + b_2' X + \beta U + R \tag{3.16}$$

where X is exogenous and is therefore uncorrelated with U, V, and R. The reduced form for W is now

$$W = (\gamma b_1 + b_2)' X + (\gamma + \beta) U + \gamma V + R$$

So we can estimate the coefficient vectors b_1 and b_2 along with γ and β by replacing Y and W in equations (3.9), (3.10), and (3.11) with $Y - b_1' X$ and $W - (\gamma b_1 + b_2)' X$, respectively and estimate those moments along with the moments

$$E\left(\left(W - (\gamma b_1 + b_2)' X\right) X\right) = 0, \quad E\left((Y - b_1' X) X\right) = 0 \tag{3.17}$$

The complete set of moments for estimating this model via GMM, which we use in our empirical application, is provided in Supplement B.

Although we did not find this to be the case in our application, when GMM models are substantially overidentified (many more moments then parameters) it is sometimes preferable to only use a subset of available moments for estimation. Since our estimator takes the form of standard GMM, in these cases the existing literature on empirical choice of moments in standard GMM estimation might be applied. See, e.g., Andrews and Lu (2001), Caner (2009), and Liao (2013).

For simplicity, these estimators assumed the errors U, V, and R are homoskedastic, and similarly have higher moments that do not depend on X. This could be relaxed to allow higher moments of these errors to depend in unknown ways on X, by letting the assumptions of Lemma 3 hold conditional on X, thereby replacing the unconditional moments of equations (3.7) and (3.8) with conditional moments. Corresponding estimators would then, however, be much more complicated, and parameters like the error variances would need to be replaced by nonparametric functions of X.

3.3 General Point Identification

We now provide a more general and systematic analysis of the identification of our model, using more information than the low order moments of Lemma 3. We provide four main results. First, we show that it is possible to construct infinitely many moments like those of Lemma 3, which can be used to construct simple GMM estimators, and we give the conditions under which these moments point identify the coefficients α and γ (equivalently, β and γ). Second, we apply Kotlarski's lemma to point identify the distributions of U, V, and R given point identification of α and γ. Third, we demonstrate that, using the entire joint distribution of Y and W (instead of just some moments) the only case where point identification is not possible is when U or V (or both) are normal. Finally, in the not point identified case, we fully characterize the sharp identified set.

We make extensive use of the characteristic function and its logarithm. Knowing the (log) characteristic function of a vector of random variables is equivalent to knowing the joint distribution of those variables (Theorem 3.1.1 in Lukacs (1970)).

Definition 2. *Given two random variables Y and W, let $\phi_{Y,W}(\zeta,\xi) \equiv E\left[e^{i\zeta Y + i\xi W}\right]$ denote their joint characteristic function. Similarly for a single random variable, let $\phi_Y(\zeta) \equiv E\left[e^{i\zeta Y}\right]$. Moreover, let $\Phi_{Y,W}(\zeta,\xi) \equiv \ln \phi_{Y,W}(\zeta,\xi)$ and $\Phi_Y(\zeta) \equiv \ln \phi_Y(\zeta)$ denote log characteristic functions (which are also called cumulant generating functions).*

Definition 3. *Given two random variables Y and W, define the cumulant of order k,ℓ (Lukacs (1970), p. 27) as*

$$\Phi_{Y,W}^{k,\ell} \equiv \left[\frac{\partial^{k+\ell}\Phi_{Y,W}(\zeta,\xi)}{i^{k+\ell}\partial\zeta^k\partial\xi^\ell}\right]_{\zeta=0,\xi=0}.$$

Similarly for a single random variable, define the cumulant of order k as

$$\Phi_Y^k \equiv \left[\frac{\partial^k\Phi_Y(\zeta)}{i^k\partial\zeta^k}\right]_{\zeta=0}.$$

All cumulants can be expressed in terms of standard moments, as obtained by an explicit differentiation of the log characteristic function and by exploiting the characteristic function moment theorem (e.g. $E\left[Y^k\right]=\left[\frac{\partial^k\phi(\xi)}{i^k\partial\xi^k}\right]_{\xi=0}$)[7]. Also note that the joint and marginal characteristic functions as well as the corresponding cumulants are directly related, e.g., $\phi_Y(\zeta)=\phi_{Y,W}(\zeta,0)$, $\Phi_Y(\zeta)=\Phi_{Y,W}(\zeta,0)$ and $\Phi_Y^k=\Phi_{Y,W}^{k,0}$.

With these tools in hand, we are ready to state a general identification result based on moment constraints. As in Lemma 3, we start by rewriting the model of equations (3.4) and (3.5) in the reduced form of equations (3.4) and (3.6), and focus on the parameters α and γ.

Theorem 7. *Let Assumptions 5, 6, and Equations (3.4) and (3.6) hold. Assume $-\infty<\gamma<\alpha<\infty$ and let*

$$M_p(\alpha,\gamma)\equiv\Phi_{Y,W}^{1+p,2}-\alpha^2\Phi_Y^{3+p}-(\gamma+\alpha)\left(\Phi_{Y,W}^{2+p,1}-\alpha\Phi_Y^{3+p}\right). \tag{3.18}$$

Let $q,\tilde{q}\in\mathbb{N}\equiv\{0,1,\ldots\}$ with $q<\tilde{q}$. If $E\left[|U|^{\tilde{q}}\right]$, $E\left[|V|^{\tilde{q}}\right]$ and $E\left[|R|^{\tilde{q}}\right]$ exist and $\Phi_Y^{3+\tilde{q}}\Phi_{Y,W}^{2+q,1}\neq\Phi_Y^{3+q}\Phi_{Y,W}^{2+\tilde{q},1}$ (or, equivalently, if $\Phi_U^{3+\tilde{q}}\Phi_V^{3+q}\neq\Phi_V^{3+\tilde{q}}\Phi_U^{3+q}$), then the moment constraints

$$M_q(\alpha,\gamma)=0 \tag{3.19}$$

$$M_{\tilde{q}}(\alpha,\gamma)=0 \tag{3.20}$$

point identify the parameters of the model as $(\alpha,\gamma)=(\alpha_+,\alpha_-)$, where

$$\alpha_\pm=\frac{F^{3012}}{2F^{3021}}\pm\sqrt{\left(\frac{F^{3012}}{2F^{3021}}\right)^2+\frac{F^{1221}}{F^{3021}}}$$

and where $F^{abcd}\equiv\Phi_{Y,W}^{a+\tilde{q},b}\Phi_{Y,W}^{c+q,d}-\Phi_{Y,W}^{a+q,b}\Phi_{Y,W}^{c+\tilde{q},d}$.

The proof, provided in Supplement A, proceeds by a judicious choice of cumulants of (Y,W) that do not depend on cumulants of R, and by exploiting the fact that cumulants of (Y,W) of order k,ℓ that share the same value of $k+\ell$ involve the same cumulants of U and V with prefactors that only differ in how they depend on α and γ. These observations then lead to specific functions of cumulants that can be analytically solved for α and γ.

Note that Theorem 7 also relies on Assumption 4, here rephrased as $-\infty<\gamma<\alpha<\infty$. Had we assumed $-\infty<\alpha<\gamma<\infty$ instead, then essentially the same Theorem would hold

[7] For high-order cumulants, these otherwise tedious algebraic manipulations could be handled with symbolic algebra packages.

except that now α and γ would be point identified by $(\alpha,\gamma)=(\alpha_-,\alpha_+)$. We next formally show that Theorem 7 contains Lemma 3 as a special case.

Corollary 1. *The assumptions of Theorem 7 with $q=0$ and $\tilde{q}=1$ imply that the assumptions of Lemma 3 hold. Equations (3.19) and (3.20) in Theorem 7 with $q=0$ and $\tilde{q}=1$ are equivalent to equations (3.7) and (3.8) in Lemma 3.*

Equations (3.9), (3.10), and (3.11), used for GMM estimation of α and γ, were obtained by converting equations (3.7) and (3.8) into moments suitable for GMM. Equivalently, equations (3.9), (3.10), and (3.11) could have been directly derived from $M_0(\alpha,\gamma)=0$ and $M_1(\alpha,\gamma)=0$. This is done explicitly in the proof of Corollary 1.

As noted above, all cumulants can be expressed in terms of standard moments, specifically, cumulants equal sums of products of moments. To fit within a GMM framework, the cumulants in the expressions $M_p(\alpha,\gamma)=0$, after being converted to functions of moments, must be linearized. This is done by introducing nuisance parameters. To illustrate, the cumulant Φ_Y^4 appears in the equation $M_1(\alpha,\gamma)=0$. Now Φ_Y^4 equals $E\left[Y^4\right]-3\left[E\left(Y^2\right)\right]^2$, so, e.g., to convert the expression $\Phi_Y^4=c$ into a form suitable for GMM, we rewrite this expression as $E\left[Y^4-3Y^2\mu_{YY}-c\right]=0$ and $E\left[Y^2-\mu_{YY}\right]=0$, using the nuisance parameter μ_{YY} that was introduced in the previous section.

Theorem 7 shows that one can obtain any number of additional, potentially overidentifying, moments to use for GMM estimation, based on the fact $M_p(\alpha,\gamma)=0$ holds for any nonnegative integer p (as long as the associated moments of U, V, and R exist). We illustrate this in Supplement B, where, in addition to the moments based on Lemma 3, we provide the additional moments suitable for GMM estimation that are obtained from $p=2$. In our later Monte Carlo simulations and empirical application, we provide results using the exactly identifying set of GMM moments based on $p=0$ and 1, and also using the generally over identifying set of GMM moments based on $p=0,1$ and 2.

Theorem 7 provides explicit conditions under which any pair of cumulant functions $M_q(\alpha,\gamma)=0$ and $M_{\tilde{q}}(\alpha,\gamma)=0$ suffice to identify the parameters α and γ. In particular, point identification based on the moments in Lemma 3, corresponding to $M_0(\alpha,\gamma) = 0$ and $M_1(\alpha,\gamma) = 0$, requires that $\Phi_U^4\Phi_V^3 \neq \Phi_V^4\Phi_U^3$, or equivalently

$$\left(E\left(U^4\right) - 3\left[E\left(U^2\right)\right]^2\right)E\left(V^3\right) - \left(E\left(V^4\right) - 3\left[E\left(V^2\right)\right]^2\right)E\left(U^3\right) \neq 0. \tag{3.21}$$

The left-hand side of (3.21) turns out to be proportional to the determinant of the Jacobian of the moment conditions (3.7) and (3.8) evaluated at the true value of the parameters:

$$\beta\begin{bmatrix} E\left[V^3\right] & -E\left[U^3\right] \\ E\left[V^4\right] - 3\left(E\left[V^2\right]\right)^2 & -E\left[U^4\right] + 3\left(E\left[U^2\right]\right)^2 \end{bmatrix}. \tag{3.22}$$

This connection is expected, since having a nonsingular Jacobian at the true parameter values is a necessary condition for point identification.

Condition (3.21) is violated, for instance, if either U or V is normal, or if both U and V are symmetric, or if both U and V have the exact same distribution. If we add the additional moments corresponding to $M_2(\alpha,\gamma) = 0$, then point identification only requires that at least one of the inequalities $\Phi_U^4\Phi_V^3 \neq \Phi_V^4\Phi_U^3$, $\Phi_U^5\Phi_V^3 \neq \Phi_V^5\Phi_U^3$, or $\Phi_U^5\Phi_V^4 \neq \Phi_V^5\Phi_U^4$, hold. For example, if the second of these holds then Theorem 7 applies with $q = 0$ and $\widetilde{q} = 2$. If more than one of these inequalities holds, then we are generally overidentified.

Once the parameters α and γ have been identified, the full distribution of all unobservables can be determined under the following Assumption.[8]

Assumption 9. *The characteristic functions of U,V and R are nonvanishing on the real line.*

Corollary 2. *If Assumptions 5, 6, 9 and Equations (3.4) and (3.6) hold, $E\left[|Y|\right] < \infty$ and if α,γ are point identified, then the distributions of U, V and R are point identified from the*

[8] This can be relaxed to nonvanishing everywhere, except at isolated points, under slightly stronger moment existence conditions; see Schennach (2004) and Evdokimov and White (2012).

joint distribution of Y and W through

$$\Phi_V(\xi) = \int_0^\xi \frac{E\left[iYe^{i\zeta\frac{W-\alpha Y}{\gamma-\alpha}}\right]}{E\left[e^{i\zeta\frac{W-\alpha Y}{\gamma-\alpha}}\right]} d\zeta \tag{3.23}$$
$$\Phi_U(\zeta) = \Phi_Y(\zeta) - \Phi_V(\zeta)$$
$$\Phi_R(\xi) = \Phi_W(\xi) - \Phi_U(\alpha\xi) - \Phi_V(\gamma\xi).$$

A more explicit expression for the distributions of these unobserved variables can be obtained by an inverse Fourier transform. For instance, if V admits a density, it is given by

$$f_V(v) = (2\pi)^{-1}\int_{-\infty}^{\infty} \exp(\Phi_V(\xi))\, e^{-i\xi v} d\xi \tag{3.24}$$

and similarly for the other densities. More general distributions (e.g. discrete and/or singular) can be recovered as well, if equation (3.24) is interpreted in the appropriate measure theoretic sense.

Although Theorem 7 is quite general, it does require the condition $\Phi_U^{3+\tilde{q}}\Phi_V^{3+q} \neq \Phi_V^{3+\tilde{q}}\Phi_U^{3+q}$ to deliver identification, so it is natural to ask whether this is fundamentally necessary. It is in fact possible to formulate an estimation strategy that relaxes this condition. For instance, as discussed above, one could stack the moment conditions of the form (3.19) and (3.20) obtained with different values of $(q, \tilde{q})$. The resulting moment conditions would only fail to identify (α, γ) if the condition $\Phi_U^{3+\tilde{q}}\Phi_V^{3+q} \neq \Phi_V^{3+\tilde{q}}\Phi_U^{3+q}$ fails simultaneously for all the choices of q and $\tilde{q}$ considered.

An even more general strategy could be to start from the fundamental relationships between the log characteristic functions of the observables and unobservables ($\Phi_{Y,W}(\zeta,\xi) = \Phi_U(\zeta + \alpha\xi) + \Phi_V(\zeta + \gamma\xi) + \Phi_R(\xi)$) and cast identification as an optimization problem that minimizes deviations between the observed quantities (i.e. $\Phi_{Y,W}(\zeta,\xi)$) and predicted quantities:

$$(\alpha, \gamma, \Phi_U, \Phi_V, \Phi_R) \tag{3.25}$$
$$= \arg\min_{(\alpha,\gamma,\Phi_U,\Phi_V,\Phi_R)} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left|\Phi_U(\zeta + \alpha\xi) + \Phi_V(\zeta + \gamma\xi) + \Phi_R(\xi) - \Phi_{Y,W}(\zeta,\xi)\right|^2 d\xi d\zeta,$$

subject to $\alpha \geq \gamma$, zero mean constraints ($\Phi_U'(0) = 0, \Phi_V'(0) = 0, \Phi_R'(0) = 0$) and that

(Φ_U, Φ_V, Φ_R) be valid log characteristic functions. This approach circumvents requiring existence of the moments $E\left[|U|^{\tilde{q}}\right]$, $E\left[|V|^{\tilde{q}}\right]$ and $E\left[|R|^{\tilde{q}}\right]$. However, the introduction of nuisance functions (Φ_U, Φ_V, Φ_R) would complicate estimation, as these would have to be parameterized by series or other expansions to construct a corresponding sieve estimator. An estimator based on Equation (3.25) would be obtained replacing $\Phi_{Y,W}(\zeta, \xi)$ by its sample analogue and trimming or downweighting the high-frequency tails in the integral.

The question remains, do there exist situations where neither this nor any other estimator can consistently estimate the model, due to lack of point identification? The following theorem fully addresses this question, by showing that there exist cases that are not point identified. However, all such cases are when U or V (or both) are normal.

This differs from, and is simpler than, Reiersøl (1950) well-known result in linear univariate errors-in-variables models, where the nonidentified cases arise when the model contains normal factors (see below). However, the required methods of proof differ significantly. For instance, the presence of two slope parameters α and γ (instead of one), and the presence of both latent variables U and V in both equations of the model, prevents us from using Reiersøl's proof method, which is based on the fact that two functions of different variables that are equal to each other must be constant. In our case, we have sums of many different functions of different variables on each side of an equality, and possible cancellation between terms that complicates the argument significantly.

Assumption 10. $E\left[|U|^3\right], E\left[|V|^3\right], E\left[|R|^3\right]$ *are finite.*

Theorem 8. *Let Assumptions 5, 6, 9, 10 and Equations (3.4) and (3.6) hold and assume that $-\infty < \gamma < \alpha < \infty$. If neither U nor V are normally distributed, then α, γ are uniquely determined by the joint distribution of Y and W by Equation (3.25).*

Note that U or V normal implies Y has full real line support, so having the support of Y be bounded is a simple sufficient condition for point identification. In the next section, we address what happens when either U or V (or both) are normally distributed.

3.4 Set Identification

In the case where Theorem 8 does not apply, so that the parameters are not point identified, the objective function of Equation (3.25) is maximized over a set rather than at a single point. In order to precisely characterize this *identified set*, we first need to introduce the notion of *factor*, which is used by Reiersøl (1950) and by Schennach and Hu (2013).

Definition 4. *If a random variable Z can be decomposed as $Z = Z_1 + Z_2$ where Z_1 and Z_2 are independent, then Z_1 and Z_2 are called* factors *of Z. (The term* factor *can also be used to refer to the distributions of these variables.)*

While for given characteristic functions $\phi_{Z_1}(\xi)$ and $\phi_{Z_2}(\xi)$, we automatically have that $\phi_Z(\xi) = \phi_{Z_1}(\xi)\,\phi_{Z_2}(\xi)$ by the convolution theorem, the notion of factor embodies the fact that, if one is instead given the two characteristic functions $\phi_Z(\xi)$ and $\phi_{Z_1}(\xi)$, it is not automatic that there exists a random variable Z_2 with characteristic function $\phi_{Z_2}(\xi) = \phi_Z(\xi)/\phi_{Z_1}(\xi)$. The inverse Fourier transform of $\phi_{Z_2}(\xi)$, may not actually yield a proper probability measure (it could assign negative weights to some sets, for instance).

Next we consider what it means for a random variable to have a normal factor.

Lemma 4. *Let Z be an observed zero mean random vector. Then Z admits a unique decomposition into two unobserved zero mean independent factors*

$$Z = Z_g + Z_n, \tag{3.26}$$

where Z_g is Gaussian with variance $\bar{\Lambda}$ and Z_n has no Gaussian factors. Furthermore, the variance of Z_g is determined (from the observed distribution of Z) from the unique $\bar{\Lambda}$ such that

$$\bar{\Lambda} - \Lambda \text{ is positive semidefinite} \iff \phi_Z(\xi)\exp(\xi'\Lambda\xi/2) \text{ is a characteristic function.}$$

(Note that either Z_g or Z_n or both could be zero.)

Intuitively, Lemma 4 indicates that the decomposition into a Gaussian and a non-Gaussian factor can, in principle, be found by attempting to deconvolve Z by a Gaussian of variance Λ and seeking the "largest" (in a positive definite sense) possible Λ that will still yield a proper distribution. In Fourier representation, this amounts to dividing $\phi_Z(\xi)$ by $\exp(-\xi'\Lambda\xi/2)$ and checking if the result is a valid characteristic function (e.g., by verifying if the inverse Fourier transform is a nonnegative measure). An alternative check for the validity of a given function $\phi(\xi)$ to be a valid characteristic function can be based on Bochner's Theorem (Theorem 4.2.2 in Lukacs (1970): ϕ is a characteristic function iff $\sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j^* \phi(\xi_i - \xi_j) \geq 0$ for all $c_1, \ldots, c_n \in \mathbb{C}$ for all $\xi_1, \ldots, \xi_n \in \mathbb{R}$ for all integer $n \geq 1$ (Bochner's Theorem also includes the conditions that $\phi(\xi)$ be continuous and $\phi(0) = 1$ but these are automatically satisfied in our context.)

Using Lemma 4, we can decompose the observed $Z = (Y, W)$ into Gaussian (g) and non-Gaussian (n) factors

$$(Y, W) = (Y_g, W_g) + (Y_n, W_n) \tag{3.27}$$

This decomposition can be accomplished without the knowledge of α or γ. The non-Gaussian or Gaussian nature of the two factors is important in our context, because it is associated with the features that can or cannot be point-identified. This type of decomposition is not a purely theoretical construct; it can be empirically implemented. Independent Component Analysis techniques, which are widely used in signal processing, (see Hyvärinen and Oja (2000) for a review) specifically rely on such decompositions into Gaussian and non-Gaussian components.

Define

$$B_s = \frac{E[W_s Y_s]}{E[Y_s^2]} \tag{3.28}$$

$$D_s = \frac{E[W_s^2] E[Y_s^2] - (E[W_s Y_s])^2}{(E[Y_s^2])^2} \geq 0 \tag{3.29}$$

where the subscript s is either set to "g" , or to "n" , or is removed. We can now state our set-identification theorem:

Theorem 9. *Let Assumptions 5, 6 and Equations (3.4) and (3.6) hold and assume that* $E[Y^2]$, $E[W^2]$, $E[R^2] < \infty$ *and that* $-\infty < \gamma < \alpha < \infty$. *Then, the following bounds (illus-*

trated in Figure 3.1) are sharp:

1. *If both U and V are Gaussian (and $E\left[Y^2\right] > 0$), then*

$$\alpha \geq B_g \tag{3.30}$$

$$B_g - \frac{D_g}{\alpha - B_g} \leq \gamma \leq B_g. \tag{3.31}$$

2. *If V is Gaussian but U is not (and $E\left[Y_n^2\right], E\left[Y_g^2\right] > 0$), then*

$$\alpha = B_n \tag{3.32}$$

$$B_g - \frac{D_g}{\alpha - B_g} \leq \gamma \leq B_g. \tag{3.33}$$

3. *If U is Gaussian but V is not (and $E\left[Y_n^2\right], E\left[Y_g^2\right] > 0$), then*

$$\gamma = B_n \tag{3.34}$$

$$B_g \leq \alpha \leq B_g + \frac{D_g}{B_g - \gamma}. \tag{3.35}$$

For each of the possible values of (α, γ) in the set given by Theorem 9, there corresponds a unique implied distribution for U, for V, and for R, given by Corollary 2. To distinguish between the three cases in Theorem 9, we have that case 1 holds only if Y is normal, in case 2 $B_n > B$, and in case 3 $B_n < B$.

Although the quantities B_n, B_g, D_n, D_g are, in principle, observable quantities, they may be difficult to estimate. For this reason, we also provide below a coarser bound that is only based on the covariances matrix of the observed Y and W:

Corollary 3. *The following bounds on α, γ always hold:*

$$\alpha \geq B$$

$$B - \frac{D}{\alpha - B} \leq \gamma \leq B.$$

It is no accident that these bounds have the same form as Case 1 of Theorem 9: Both are solely based on covariance information, but in the Gaussian case, covariances exhaust all available information and yield sharp inequalities while, in general, that is not the case. This looser bound is also related to the measurement error bounds in Frisch (1934). If one is willing to rely on this relaxed bound, then a simple GMM estimator for the resulting identified set could be obtained based on the moment conditions

$$E\left[\alpha^2\sigma_U^2+\gamma^2\left(Y^2-\sigma_U^2\right)+\sigma_R^2-W^2\right] \quad = \quad 0 \tag{3.36}$$

$$E\left[\alpha\sigma_U^2+\gamma\left(Y^2-\sigma_U^2\right)-YW\right] \quad = \quad 0 \tag{3.37}$$

while optimizing over $\alpha,\gamma,\sigma_U^2,\sigma_R^2$, subject to the constraints $\gamma<\alpha$ (equivalent to $\beta>0$), $\sigma_U^2\geq 0$ and $\sigma_R^2\geq 0$. These moment conditions are obtained from Equations (66) and (67) in the proof of Theorem 9, without extracting the Gaussian parts. The bounds of Corollary 3 are also obeyed in the case of point identified models, since they are obtained solely from positive variance considerations that must always be satisfied. This implies that, if one is unsure whether Y is normal or not, the moment conditions (3.36) and (3.37) could be stacked with the ones of Theorem 7 to yield an estimator that is robust to loss of point identification.[9]

3.5 Ordinary Least Squares

It is instructive to analyze in more detail how the parameters of our model relate to the slope coefficient of a naive OLS regression (in the population limit). The coefficient B given by Equation (3.28) is the slope coefficient of the least-square regression of W on Y (in the population limit). Regardless of whether the model is point identified or not, an implication of the model (i.e., of equations (3.4) and (3.5)) is that B always lies between γ and α. This can be immediately verified by observing that

$$B=\frac{E\left[YW\right]}{E\left[Y^2\right]}=\frac{E\left[\left(U+V\right)\left(\alpha U+\gamma V\right)\right]}{E\left[\left(U+V\right)^2\right]}=\frac{\alpha E\left[U^2\right]+\gamma E\left[V^2\right]}{E\left[U^2\right]+E\left[V^2\right]}=\alpha\lambda+\gamma\left(1-\lambda\right) \tag{3.38}$$

where $\lambda=E\left[U^2\right]/\left(E\left[U^2\right]+E\left[V^2\right]\right)$ and so lies between zero and one. So in particular, if $\beta>0$ we get $\gamma\leq B\leq\alpha$.

This type of inequality has been noted before in the context of estimating returns to

[9] In this case the maximizing estimands could be sets rather than points, requiring nonstandard inference.

education (e.g. by Card (2001), in a more detailed model that allows for some individual heterogeneity). In particular, in the returns to schooling context, we would expect both β and γ to be positive (because unobserved ability U should affect schooling Y and wages W in the same direction, and increased schooling should increase wages). By the above analysis, this in turn means that we would expect $0 < \gamma \leq B$.

However, as noted by Card (2001), most returns to schooling empirical applications yield estimates of γ, using instrumental variables methods, that are greater than B, which contradicts this inequality and hence also contradicts the model. One possible explanation for this contradiction is that, in the returns to schooling context, Y may also contain significant measurement error. Standard attenuation bias under classical measurement error implies that the ordinary least squares coefficient B is biased towards zero relative to γ, which if $0 < \gamma$ would imply $B < \gamma$. If the model is correct for returns to education, but in addition Y is mismeasured, then B could be either larger or smaller than γ, depending on the relative magnitude of the measurement error.

3.6 Monte Carlo

To assess the finite sample performance of our simple GMM estimators, we generate data from the model of equations (3.4) and (3.5) without covariates. All of our designs are chosen to satisfy equation (3.21), so the model is point identified just from the moments in Lemma 1.[10] The true values of the coefficients are $\gamma = \beta = 1$. It is widely recognized that estimators based on higher moments can behave poorly with small sample sizes, so to see if our estimators suffer from these issues, we work with relatively small sample sizes of $n = 100$ and $n = 400$.

We generate $5,000$ replications of four different designs. In design 1, U is log normal while V and R are each standard Gumbel. In design 2, U is log normal while V and R are

[10]In particular in all of our designs, U and V have different, non-normal distributions, and at least one is asymmetically distributed. U, V, and R are also mutually independent and centered at mean zero.

uniform. We then reverse these, making U Gumbel and V and R log normal in design 3, and making U uniform with V and R log normal in design 4. For each design, we report results using two different estimators. The exactly identified estimator is GMM using moments corresponding to Lemma 1, given by equations (77), (78), and (79) (without covariates, so $\widetilde{Y} = Y$ and $\widetilde{W} = W$), as given in Supplement B. The over-identified estimator is GMM using these same equations, plus equations (81) and (82) of Supplement B.

Tables C1 to C4 of the Supplement report results from designs 1 to 4, respectively. Each Table has four panels, corresponding to the two different GMM estimators, each with the two different sample sizes. We report estimates of γ, β, the error component variances σ_U^2, σ_V^2, and σ_R^2, and, when over-identified, μ_{WW}. Reported summary statistics of each parameter estimate across the simulations are the mean (MEAN), the standard deviation (SD), the 25% quantile (LQ), the median (MED), the 75% quantile (UQ), the root mean squared error (RMSE), the mean absolute error (MAE), and the median absolute error (MDAE).

Some general tendencies stand out in these simulations. First, consider the trade off between the exactly identified vs over identified estimators. The latter uses more information, but that information takes the form of up to fifth order moments, which can be noisy and more sensitive to outliers. In general we find that the overidentified estimator performs better than the exactly identified estimator, particularly at the larger sample size.

The primary parameter of interest, γ, tends to be estimated reasonably precisely in all of the designs, with most RMSEs in the range of .3 to .7. In contrast, β is generally much less precisely estimated, often having much larger RMSEs (except in design 2). Estimates of the variances σ_U^2, σ_V^2, and σ_R^2, are mostly similar to each other, usually being less precise than γ but more than β. The estimate of μ_{WW} is noisier, since it only appears in the highest order moment equations of the over identified model. The designs where U was log normal (designs 1 and 2) generally had more accurate estimates than the other designs. We conclude that our estimator performs reasonably well even with rather small sample sizes.

3.7 GDP and Life Expectancy

There is a long literature studying the causal effect of health on economic growth. Examples include Acemoglu and Johnson (2007) (which we will hereafter refer to as AJ), Well (2007), Lorentzen, McMillan, and Wacziarg (2008), Aghion, Howitt, and Murtin (2010), Cervellati and Sunde (2011), Ecevit (2013), Bloom, Canning, and Fink (2014), and Bloom et al. (2019).

Based on a neo-classical growth model, AJ estimate a model in the form of equations (3.1) and (3.2), where Y is the change in the log of a country's life expectancy at birth between 1940 and 1980, W is the change in that country's log GDP in the same time span, and X is either just a constant, or a constant and a measure of the country's quality of institutions, or a constant and GDP per capita in 1930. The main goal is estimation of γ, the coefficient of Y in the W equation.

AJ observe that ordinary least squares estimation of the W equation is inconsistent, because the health measure Y is endogenous, with improvements and investments in a country's productive technology over time positively impacting both health outcomes and GDP. This technology change corresponds to our unobserved factor U (with $\beta > 0$) in equations (3.15) and (3.16), while V and R are the idiosyncratic shocks to health and economic outcomes, respectively.

To deal with the endogeneity caused by U, AJ construct an instrument, called predicted mortality, that combines each country's 1940 mortality rates from specific diseases with a set of global interventions that addressed those diseases. As noted in the introduction, one may question the validity of such constructed instruments.

In Table 3.1, columns labeled 2SLS1, 2SLS2, and 2SLS3 in Panel A are replications of selected results appearing in Table 9 of AJ.[11] These are AJ's estimates using two stage least squares (2SLS) with the above listed combinations of covariates X, and using their

[11] Our data are provided by AJ. Life expectancy is from UN data sources and the League of Nation reports. Pre-war GDP data are from Maddison (2003), and post-war data are from the UN. See AJ for details.

predicted mortality instrument. AJ's ordinary least squares (OLS) estimate of γ (corresponding to B in the previous section) is -0.81, while their 2SLS estimates of γ are considerably larger in magnitude, ranging from -1.316 to -1.643. As we noted earlier, having $\gamma < B$, as AJ find, is an implication of our model when $\beta > 0$. Note that the sample size is quite small in this application, with only 47 countries. Nevertheless, AJ's estimates of γ are statistically significant.[12]

Now suppose we had not observed predicted mortality, or we are uncertain of its validity as an instrument. We can instead consider applying our GMM estimators. First, consider the distribution of Y. Assuming (measured) life expectancy is bounded away from zero, log life expectancy is bounded, which suffices for point identification since it rules out U or V being normal.[13] We therefore attempt to apply our GMM estimators.

In Table 3.1, we report two sets of GMM estimates along with AJ's 2SLS results. Columns labeled GMM1, GMM2, and GMM3 are GMM estimates of equations (3.15) and (3.16), which do not make use of the predicted mortality instrument in any way. Specifically, these are estimates based on the over-identifying set of moments given by equations (77) to (82) in Supplement B. The last three columns of Table 3.1 then give GMM estimates that use both our over-identifying set of moments and the additional moment given by AJ's instrument (as discussed at the end of Supplement B).[14]

Panel A in Table 3.1 reports the main parameter of interest γ, and also reports b_2, the other covariate coefficients in equation (3.16). The variables in columns (4) and (7)

[12] Our standard errors in columns (1)-(3) of Table 3.1 differ from those reported by AJ. AJ's estimates are from *ivreg* in Stata 9. We use *ivregress 2sls*, which replaced *ivreg* as of Stata 10. *ivreg* and *ivregress* can give different robust standard error estimates, because *ivreg* uses HC1 (MacKinnon and White 1985) robust standard errors while *ivregress 2sls* uses HC0 (Huber-White). Also, to reduce the number of coefficients in GMM estimation, we differenced the data while AJ used level data with fixed effects. Since T=2, these are asymptotically equivalent estimators.

[13] More heuristically, if Y is close to normal, then it may be that U or V is close to normal. Y has a skewness of 0.170 and a kurtosis of 1.791, which is reasonably far from normal in terms of the low order moments our GMM estimator is based on. The p-value of a Shapiro-Wilk test of normality of Y is .02, rejecting normality, and even lower if one tests the residuals after regressing Y on either of the covariates in X.

[14] These GMM models are estimated in Stata, using the vce(robust) option to compute standard errors.

have been demeaned so there is no constant.[15] Our main takeaway from Panel A of Table 3.1 is that our estimates of γ are quite comparable to AJ's. In GMM1 and GMM2, the estimates of γ are -1.984 and -1.241, virtually the same range as AJ's 2SLS estimates, and are statisically significant. GMM3 gives an estimate of a lower magnitude -0.383, but this estimate is statistically insignificant with a very large standard error, suggesting that our higher moment based estimator is imprecise for this particular combination of covariates and small sample size. The last three columns of Table 3.1, which combine both our moments and the AJ instrument, give estimates very close to those of AJ, with somewhat smaller standard errors, which is exactly what one would expect to see if both sets of moments are valid and if AJ's instrument is strong. In the bottom row of Table 3.1 we report Hansen's J-test; we do not reject validity of the joint set of overidentifying restrictions in any of the GMM estimates.

Panels B and C of Table 3.1 provide the other estimated parameters of the model. Panel C gives the estimated b_1 coefficients from equation (3.15), while Panel B gives the estimates of β and the estimated variances of our error components. β appears to be difficult to precisely estimate, with large standard errors[16]. In the specifications where γ is statistically significant, the variance of U (the source of endogeneity in the model) is much smaller than the variances of the idiosyncratic components V and R, but very precisely estimated with small standard errors.

Later tables have the same format as Table 3.1, providing additional results. In Table 3.2, we re-estimate the model using the exactly identified set of moments from Lemma 3. As expected with fewer moments, these estimates are less efficient, and turn out to be quite a bit noisier than those of Table 3.1. GMM5, with the quality of institutions as the covariate, is still reasonably comparable to AJ with γ of -1.401, while now both GMM4

[15]In Supplement B: Moments for GMM Estimation, it is noted that "For the model without covariates, one can replace b_1 and b_2 with zero in the above expressions, and drop equation (80). Note that in this case Y and W should be demeaned." . In columns (4) and (7), we don't have covariates so Y and W are demeaned.

[16]In contrast α is, like γ, much more precisely estimated, but apparently the difference $\beta \equiv \alpha - \gamma$ is harder to pin down.

and GMM6 are insignificant and more variable. The estimates combining these moments with AJ's instrument behave as before.

We also perform a number of robustness checks in Supplement D, using alternative outcome variables that AJ considered in their Tables 8-9. These additional outcomes are log population, log births, percentage of population under age 20, log GDP, and log GDP per working age population. Some of the alternative outcomes suffer from the issue that U might also contain measurement error, and in those cases, our identification results would not apply. The results of our GMM estimators with other outcomes are generally more erratic than with log per capita GDP. The estimates that combine our moments and the AJ instrument remain comparable to AJ's 2SLS estimates.

We conclude that, in all specifications where the standard errors were small enough to yield statistically significant results, our estimates based on higher moments, without side information, are very close to those obtained by AJ that required an instrument.

3.8 Conclusions

We have shown that a standard linear triangular structural model is generally point identified, without an instrument or other side information that is generally used to identify such models. We illustrate the result with Monte Carlo simulations and in an empirical application. Our application shows that, without using an instrument, GMM estimation of moments based on the model yields estimates close to those that were obtained by previous authors using an instrument. Even when instruments are available, our estimator could be usefully combined with instrument based moments to either increase estimation precision by adding more moments to the model, or to provide overidentifying moments that might be used for specification testing.

What makes point identification possible is the assumed error structure, which takes the standard form of a scalar common component U in each equation, plus additional scalar idiosyncratic components V and R. One goal for future work could include deriving alternative estimators for the model. These could include estimators that allow U, V, and R to depend nonparametrically on covariates X (e.g., allowing heteroskedasticity of unknown form), and estimators that make direct use of all the information in Theorem 8, perhaps based directly on characteristic functions rather than moments. Other possibilities for further work include extending the model to more equations, allowing the common component U to affect outcomes nonlinearly, and extending the model to also allow for measurement error in Y. Based on Card (2001), this last extension would likely be needed for returns to education applications.

Acknowledgments

Susanne Schennach acknowledges support from NSF grant SES-1950969. Vincent Starck is gratefully acknowledged for valuable comments.

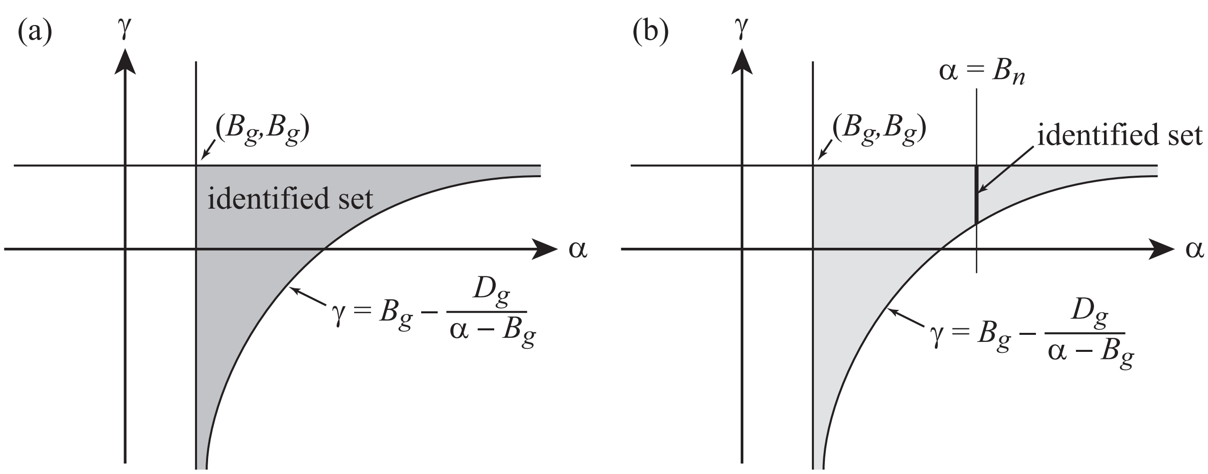


Figure 3.1: Identified set of Theorem 9 for (a) Case 1 and (b) Case 2 (Case 3, analogous to Case 2, is not shown).

Tables

Table 3.1: Over identified moments: Base sample 1940 and 1980

	(1) 2SLS1	(2) 2SLS2	(3) 2SLS3	(4) GMM1	(5) GMM2	(6) GMM3	(7) GMM1+AJ	(8) GMM2+AJ	(9) GMM3+AJ
Panel A. Dependent Variable: Growth in GDP per Capita 1940-1980									
Life expectancy	-1.316***	-1.643***	-1.589*	-1.984**	-1.241**	-0.383	-1.341***	-1.642***	-1.573*
	(0.382)	(0.521)	(0.876)	(0.911)	(0.626)	(0.383)	(0.334)	(0.520)	(0.866)
Institutions		-0.0490			-0.0291			-0.0489	
		(0.0418)			(0.0472)			(0.0417)	
Initial (1930) value of			-0.0730			0.149			-0.0638
dependent variable			(0.198)			(0.112)			(0.193)
Constant	1.336***	1.681***	1.990		1.448***	-0.127		1.680***	1.910
	(0.124)	(0.367)	(1.807)		(0.445)	(0.983)		(0.366)	(1.760)
Panel B. β and variances									
β				2.319***	4.527	54.28*	3.966**	9.523***	1.215
				(0.743)	(11.71)	(115.1)	(2.277)	(5.420)	(1.601)
σ_U^2				0.0147***	0.00171**	6.68e-07	0.00467***	0.00152***	0.0136***
				(0.0136)	(0.00486)	(0)	(0.00436)	(0.00192)	(0.0155)
σ_V^2				0.0150***	0.0177***	0.0139***	0.0260***	0.0179***	4.72e-05
				(0.0143)	(0.00552)	(0.00444)	(0.00516)	(0.00412)	(0.0155)
σ_R^2				0.0547***	0.0943**	0.120***	0.0586***	1.75e-09	0.123***
				(0.0383)	(0.0973)	(0.0263)	(0.0250)	(0)	(0.0319)
μ_{ww}				0.147***	0.143***	0.124***	0.137***	0.143***	0.125***
				(0.0261)	(0.0271)	(0.0206)	(0.0235)	(0.0270)	(0.0207)
Panel C. Dependent Variable: Growth in Life Expectancy 1940-1980									
Institutions		-0.0310***			-0.0496***			-0.0496***	
		(0.00755)			(0.00997)			(0.00996)	
Initial (1930) value of			-0.117***			-0.184***			-0.185***
dependent variable			(0.0310)			(0.0223)			(0.0212)
Constant		0.324***	1.122***		0.579***	1.757***		0.580***	1.760***
		(0.0595)	(0.267)		(0.0523)	(0.173)		(0.0522)	(0.165)
Observations		47	47	47	47	47	47	47	47
Hansen J				0.122	0	0.000493	0.850	0.000109	0.0598
p-val				0.727	1	0.982			

Notes: In all models, the endogenous regressor is the changes in log life expectancy between 1940 and 1980. 2SLS1 is the two-stage least squares regression of growth in GDP per capita on growth in life expectency, using predicted mortality as the instrument. 2SLS2 includes a measure of quality of institutions as exogenous covariate. 2SLS3 adds the initial (1930) value of log GDP per capita. GMM1-GMM3 are the same models as 2SLS1-2SLS3 estimated by GMM estimators based on over identified moments. GMM1-GMM3+AJ combine our over identified moments and the AJ moment, i.e. $E(IV\varepsilon) = 0$. The last row reports the p value of the J statistics under the null hypothesis that the overidentifying restrictions are valid.

Table 3.2: Exactly identified moments: Base sample 1940 and 1980

	(1) 2SLS1	(2) 2SLS2	(3) 2SLS3	(4) GMM4	(5) GMM5	(6) GMM6	(7) GMM4+AJ	(8) GMM5+AJ	(9) GMM6+AJ
Panel A. Dependent Variable: Growth in GDP per Capita 1940-1980									
Life expectancy	-1.316***	-1.643***	-1.589*	-3.636	-1.401	-0.340	-1.344***	-1.634***	-1.599*
	(0.382)	(0.521)	(0.876)	(4.184)	(9.818)	(1.022)	(0.348)	(0.520)	(0.871)
Institutions		-0.0490			-0.0370			-0.0478	
		(0.0418)			(0.483)			(0.0416)	
Initial (1930) value of dependent variable			-0.0730			0.156			-0.0706
			(0.198)			(0.195)			(0.194)
Constant	1.336***	1.681***	1.990		1.541	-0.195		1.670***	1.969
	(0.124)	(0.367)	(1.807)		(5.663)	(1.836)		(0.365)	(1.775)
Panel B. β and variances									
β				3.091	13.61	1.589	2.228	0.805	1.235
				(3.588)	(2,713)	(5.015)	(1.235)	(6.184)	(1.190)
σ_U^2				0.0278***	0.000712	8.72e-06	0.00804***	0.0177	0.0138***
				(0.0107)	(0.128)	(0.00959)	(0.00714)	(0.136)	(0.0123)
σ_V^2				0.00253	0.0187	0.0139***	0.0220***	0.00170	0.000140
				(0.00960)	(0.128)	(0.0133)	(0.00718)	(0.136)	(0.0100)
σ_R^2				0.101***	9.05e-07	0.123***	0.0863***	0.126***	0.123***
				(0.0307)	(28.70)	(0.0306)	(0.0198)	(0.0843)	(0.0317)
Panel C. Dependent Variable: Growth in Life Expectancy 1940-1980									
Institutions		-0.0310***			-0.0496***			-0.0494***	
		(0.00755)			(0.00997)			(0.00996)	
Initial (1930) value of dependent variable			-0.117***			-0.184***			-0.184***
			(0.0310)			(0.0223)			(0.0219)
Constant		0.324***	1.122***		0.579***	1.761***		0.578***	1.758***
		(0.0595)	(0.267)		(0.0522)	(0.173)		(0.0522)	(0.170)
Observations		47	47	47	47	47	47	47	47
Hansen J							2.039	0.00564	0.0289
p-val							0.153	0.940	0.865

Notes: In all models, the endogenous regressor is the changes in log life expectancy between 1940 and 1980. 2SLS1 is the two-stage least squares regression of growth in GDP per capita on growth in life expectency, using predicted mortality as the instrument. 2SLS2 includes a measure of quality of institutions as exogenous covariate. 2SLS3 adds the initial (1930) value of log GDP per capita. GMM4-GMM6 are the same models as 2SLS1-2SLS3 estimated by GMM estimators based on exactly identified moments. GMM4-GMM6+AJ combine our exactly identified moments and the AJ moment, i.e. $E(IV\varepsilon)=0$. The last row reports the p value of the J statistics under the null hypothesis that the overidentifying restrictions are valid.

Bibliography

Acemoglu, Daron, and Simon Johnson. 2007. “Disease and development: the effect of life expectancy on economic growth”. *Journal of political Economy* 115:925–985.

Aghion, Philippe, Peter Howitt, and Fabrice Murtin. 2010. *The relationship between health and growth: when Lucas meets Nelson-Phelps*. Tech. rep. National Bureau of Economic Research.

Aguiar, Mark, and Erik Hurst. 2007. “Life-cycle prices and production”. *American Economic Review* 97:1533–1559.

Ai, Chunrong, and Xiaohong Chen. 2003. “Efficient estimation of models with conditional moment restrictions containing unknown functions”. *Econometrica* 71:1795–1843.

Allcott, Hunt, Benjamin B Lockwood, and Dmitry Taubinsky. 2019a. “Regressive sin taxes, with an application to the optimal soda tax”. *The Quarterly Journal of Economics* 134:1557–1626.

— . 2019b. “Should we tax sugar-sweetened beverages? An overview of theory and evidence”. *Journal of Economic Perspectives* 33:202–27.

Amemiya, Takeshi. 1974. “The nonlinear two-stage least-squares estimator”. *Journal of econometrics* 2:105–110.

Andrews, Donald WK. 2017. “Examples of L2-complete and boundedly-complete distributions”. *Journal of econometrics* 199:213–220.

Andrews, Donald WK, and Biao Lu. 2001. "Consistent model and moment selection procedures for GMM estimation with application to dynamic panel data models". *Journal of econometrics* 101:123–164.

Andreyeva, Tatiana, Michael W Long, and Kelly D Brownell. 2010. "The impact of food prices on consumption: a systematic review of research on the price elasticity of demand for food". *American journal of public health* 100:216–222.

Angrist, Joshua D, and Alan B Krueger. 1991. "Does compulsory school attendance affect schooling and earnings?" *The Quarterly Journal of Economics* 106:979–1014.

— . 2001. "Instrumental variables and the search for identification: From supply and demand to natural experiments". *Journal of Economic perspectives* 15:69–85.

Armstrong, Timothy B. 2016. "Large market asymptotics for differentiated product demand estimators with economic models of supply". *Econometrica* 84:1961–1980.

Backus, Matthew, Christopher Conlon, and Michael Sinkinson. 2021. *Common ownership and competition in the ready-to-eat cereal industry*. Tech. rep. National Bureau of Economic Research.

Berardi, Nicoletta, et al. 2016. "The impact of a 'soda tax'on prices: evidence from French micro data". *Applied Economics* 48:3976–3994.

Bergman, U Michael, and Niels Lynggård Hansen. 2019. "Are Excise Taxes on Beverages Fully Passed through to Prices?: The Danish Evidence". *FinanzArchiv* 75:323–356.

Berkson, Joseph. 1950. "Are there two regressions?" *Journal of the american statistical association* 45:164–180.

Berry, Steve, Oliver B Linton, and Ariel Pakes. 2004. "Limit theorems for estimating the parameters of differentiated product demand systems". *The Review of Economic Studies* 71:613–654.

Berry, Steven. 1994. "Estimating discrete-choice models of product differentiation". *The RAND Journal of Economics*: 242–262.

Berry, Steven, Michael Carnall, and Pablo T Spiller. 2006. "Airline hubs: costs, markups and the implications of customer heterogeneity". *Competition policy and antitrust*.

Berry, Steven, Amit Gandhi, and Philip Haile. 2013. "Connected substitutes and invertibility of demand". *Econometrica* 81:2087–2111.

Berry, Steven, and Philip A Haile. 2014. "Identification in differentiated products markets using market level data". *Econometrica* 82:1749–1797.

Berry, Steven, and Panle Jia. 2010. "Tracing the woes: An empirical analysis of the airline industry". *American Economic Journal: Microeconomics* 2:1–43.

Berry, Steven, James Levinsohn, and Ariel Pakes. 1995. "Automobile prices in market equilibrium". *Econometrica: Journal of the Econometric Society*: 841–890.

Berto Villas-Boas, Sofia. 2007. "Vertical relationships between manufacturers and retailers: Inference with limited data". *The Review of Economic Studies* 74:625–652.

Bierens, Herman J. 2012. *Robust methods and asymptotic theory in nonlinear econometrics*. Vol. 192. Springer Science & Business Media.

Birchall, Cameron, and Frank Verboven. 2022. "Estimating Substitution Patterns and Demand Curvature in Discrete-Choice Models of Product Differentiation".

Bloom, David E, David Canning, and Günther Fink. 2014. "Disease and development revisited". *Journal of Political Economy* 122:1355–1366.

Bloom, David E, et al. 2019. *Health and economic growth: reconciling the micro and macro evidence*. Tech. rep. National Bureau of Economic Research.

Blundell, Wesley, Gautam Gowrisankaran, and Ashley Langer. 2020. "Escalation of scrutiny: The gains from dynamic enforcement of environmental regulations". *American Economic Review* 110:2558–2585.

Bokhari, Farasat AS, and Franco Mariuzzo. 2018. "Demand estimation and merger simulations for drugs: Logits v. AIDS". *International Journal of Industrial Organization* 61:653–685.

Bollinger, Bryan, and Steven Sexton. 2018. "Local excise taxes, sticky prices, and spillovers: evidence from Berkeley's soda tax". *Sticky Prices, and Spillovers: Evidence from Berkeley's Soda Tax (January 12, 2018)*.

Bonhomme, Stéphane, and Jean-Marc Robin. 2010. "Generalized non-parametric deconvolution with an application to earnings dynamics". *The Review of Economic Studies* 77:491–533.

Bonnet, Céline, and Pierre Dubois. 2010. "Inference on vertical contracts between manufacturers and retailers allowing for nonlinear pricing and resale price maintenance". *The RAND Journal of Economics* 41:139–164.

Bonnet, Céline, Vincent Requillart, et al. 2012. "Sugar policy reform, tax policy and price transmission in the soft drink industry". *Toulouse School of Economics*.

Bonnet, Céline, and Vincent Réquillart. 2013a. "Impact of Cost Shocks on Consumer Prices in Vertically-Related Markets: The Case of The French Soft Drink Market". *American Journal of Agricultural Economics* 95:1088–1108.

— . 2013b. "Tax incidence with strategic firms in the soft drink market". *Journal of Public Economics* 106:77–88.

Browning, Martin, and Costas Meghir. 1991. "The effects of male and female labor supply on commodity demands". *Econometrica: Journal of the Econometric Society*: 925–951.

Caner, Mehmet. 2009. "Lasso-type GMM estimator". *Econometric Theory* 25:270–290.

Card, David. 2001. "Estimating the return to schooling: Progress on some persistent econometric problems". *Econometrica* 69:1127–1160.

— . 1995. "Using geographic variation in college proximity to estimate the return to schooling in Louis Christofides, Kenneth E. Grant and Robert Swidinsky". *Aspects of Labour Market Behaviour, essays in honour of John Vanderkamp*.

Cardell, N Scott. 1997. “Variance components structures for the extreme-value and logistic distributions with application to models of heterogeneity”. *Econometric Theory* 13:185–213.

Carneiro, Pedro, and James J Heckman. 2002. “The evidence on credit constraints in post-secondary schooling”. *The Economic Journal* 112:705–734.

Cawley, John, and David E Frisvold. 2017. “The pass-through of taxes on sugar-sweetened beverages to retail prices: the case of Berkeley, California”. *Journal of Policy Analysis and Management* 36:303–326.

Cawley, John, et al. 2019. “The economics of taxes on sugar-sweetened beverages: a review of the effects on prices, sales, cross-border shopping, and consumption”. *Annual review of nutrition* 39:317–338.

Cawley, John, et al. 2021. “The pass-through of a tax on sugar-sweetened beverages in Boulder, Colorado”. *American Journal of Agricultural Economics* 103:987–1005.

Cervellati, Matteo, and Uwe Sunde. 2011. “Life expectancy and economic growth: the role of the demographic transition”. *Journal of economic growth* 16:99–133.

Chen, Xiaohong, and Demian Pouzo. 2009. “Efficient estimation of semiparametric conditional moment models with possibly nonsmooth residuals”. *Journal of Econometrics* 152:46–60.

Chernozhukov, Victor, Jerry A Hausman, and Whitney K Newey. 2019. *Demand analysis with many prices*. Tech. rep. National Bureau of Economic Research.

Chu, Chenghuan Sean, Phillip Leslie, and Alan Sorensen. 2011. “Bundle-size pricing as an approximation to mixed bundling”. *The American Economic Review*: 263–303.

Ciliberto, Federico, Charles Murry, and Elie Tamer. 2021. “Market structure and competition in airline markets”. *Journal of Political Economy* 129:2995–3038.

Clay, Marie, et al. 2016. *Comparing National Household Food Acquisition and Purchase Survey (FoodAPS) Data With Other National Food Surveys’ Data*. Tech. rep.

Comon, Pierre. 1994. "Independent component analysis, a new concept?" *Signal processing* 36:287–314.

Conlon, Christopher, and Jeff Gortmaker. 2020. "Best practices for differentiated products demand estimation with pyblp". *The RAND Journal of Economics* 51:1108–1161.

Conlon, Christopher, and Julie Holland Mortimer. 2021. "Empirical properties of diversion ratios". *The RAND Journal of Economics* 52:693–726.

Currie, Janet. 2009. "Healthy, wealthy, and wise: Socioeconomic status, poor health in childhood, and human capital development". *JoUrnal of economIc lIteratUre* 47:87–122.

Currie, Janet, et al. 2010. "The effect of fast food restaurants on obesity and weight gain". *American Economic Journal: Economic Policy* 2:32–63.

Cutler, David M, Edward L Glaeser, and Jesse M Shapiro. 2003. "Why have Americans become more obese?" *Journal of Economic perspectives* 17:93–118.

Darmois, George. 1953. "Analyse générale des liaisons stochastiques: etude particulière de l'analyse factorielle linéaire". *Revue de l'Institut international de statistique*: 2–8.

Diamond, Peter A, and James A Mirrlees. 1971a. "Optimal taxation and public production I: Production efficiency". *The American economic review* 61:8–27.

— . 1971b. "Optimal taxation and public production II: Tax rules". *The American Economic Review* 61:261–278.

Dubé, Jean-Pierre. 2005a. "Product differentiation and mergers in the carbonated soft drink industry". *Journal of Economics & Management Strategy* 14:879–904.

— . 2005b. "Product differentiation and mergers in the carbonated soft drink industry". *Journal of Economics & Management Strategy* 14:879–904.

Dubois, Pierre, Rachel Griffith, and Aviv Nevo. 2014. "Do prices and attributes explain international differences in food purchases?" *American Economic Review* 104:832–867.

Dubois, Pierre, Rachel Griffith, and Martin O'Connell. 2020. "How well targeted are soda taxes?" *American Economic Review* 110:3661–3704.

Duch-Brown, Néstor, et al. 2017. "The impact of online sales on consumers and firms. Evidence from consumer electronics". *International Journal of Industrial Organization* 52:30–62.

Dunker, Fabian, Stefan Hoderlein, and Hiroaki Kaido. 2022. "Nonparametric identification of random coefficients in endogenous and heterogeneous aggregate demand models". *arXiv preprint arXiv:2201.06140*.

Eizenberg, Alon, and Alberto Salvo. 2015a. "The rise of fringe competitors in the wake of an emerging middle class: An empirical analysis". *American Economic Journal: Applied Economics* 7:85–122.

— . 2015b. "The rise of fringe competitors in the wake of an emerging middle class: An empirical analysis". *American Economic Journal: Applied Economics* 7:85–122.

Erickson, Timothy, and Toni M Whited. 2002. "Two-step GMM estimation of the errors-in-variables model using high-order moments". *Econometric Theory* 18:776–799.

Evdokimov, Kirill, and Halbert White. 2012. "Some extensions of a lemma of Kotlarski". *Econometric Theory* 28:925–932.

Falbe, Jennifer, et al. 2015. "Higher retail prices of sugar-sweetened beverages 3 months after implementation of an excise tax in Berkeley, California". *American journal of public health* 105:2194–2201.

Fisher, Franklin M. 1966. *The identification problem in econometrics*. McGraw-Hill.

Fox, Jeremy T, Kyoo il Kim, and Chenyu Yang. 2016a. "A simple nonparametric approach to estimating the distribution of random coefficients in structural models". *Journal of Econometrics* 195:236–254.

— . 2016b. "A simple nonparametric approach to estimating the distribution of random coefficients in structural models". *Journal of Econometrics* 195:236–254.

Fox, Jeremy T, et al. 2011. “A simple estimator for the distribution of random coefficients”. *Quantitative Economics* 2:381–418.

Frisch, Ragnar. 1934. “Statistical confluence analysis by means of complete regression systems”. *(No Title)*.

Fruehwirth, Jane Cooley, Salvador Navarro, and Yuya Takahashi. 2016. “How the timing of grade retention affects outcomes: Identification and estimation of time-varying treatment effects”. *Journal of Labor Economics* 34:979–1021.

Gandhi, Amit, and Jean-François Houde. 2019. *Measuring substitution patterns in differentiated products industries*. Tech. rep. National Bureau of Economic Research.

Gandhi, Amit, Zhentong Lu, and Xiaoxia Shi. 2020. “Estimating demand for differentiated products with zeroes in market share data”. *Available at SSRN 3503565*.

Gandhi, Amit, and Aviv Nevo. 2021. *Empirical Models of Demand and Supply in Differentiated Products Industries*. Tech. rep. National Bureau of Economic Research.

Ghose, Anindya, Panagiotis G Ipeirotis, and Beibei Li. 2012. “Designing ranking systems for hotels on travel search engines by mining user-generated and crowdsourced content”. *Marketing Science* 31:493–520.

Gortmaker, S, M Long, and YC Wang. 2009. *The negative impact of sugar-sweetened beverages on children’s health: A research synthesis, Robert Wood Johnson Foundation. Retrieved August 10, 2010*.

Gospodinov, Nikolay, and Serena Ng. 2015. “Minimum distance estimation of possibly noninvertible moving average models”. *Journal of Business & Economic Statistics* 33:403–417.

Greenstein, Shane M. 1996. “From superminis to supercomputers: Estimating surplus in the computing market”. In *The economics of new goods*, 329–372. University of Chicago Press.

Griffith, Rachel, Wenchao Jin, and Valérie Lechene. 2022. "The decline of home-cooked food". *Fiscal Studies* 43:105–120.

Griffith, Rachel, et al. 2020. "What's on the menu? Policies to reduce young people's sugar consumption". *Fiscal studies* 41:165–197.

Grigolon, Laura, and Frank Verboven. 2014. "Nested logit or random coefficients logit? A comparison of alternative discrete choice models of product differentiation". *Review of Economics and Statistics* 96:916–935.

Grogger, Jeffrey. 2017. *Soda taxes and the prices of sodas and other drinks: evidence from Mexico*.

Gundersen, Craig. 2021. "A proposal to reconstruct the Supplemental Nutrition Assistance Program (SNAP) into a universal basic income program for food". *Food Policy* 101:102096.

Harding, Matthew, and Michael Lovenheim. 2017. "The effect of prices on nutrition: comparing the impact of product-and nutrient-specific taxes". *Journal of Health Economics* 53:53–71.

Hausman, Jerry, Gregory Leonard, and J Douglas Zona. 1994. "Competitive analysis with differenciated products". *Annales d'Economie et de Statistique*: 159–180.

Hirsch, Stefan, Giulia Tiboldo, and Rigoberto A Lopez. 2018. "A tale of two Italian cities: brand-level milk demand and price competition". *Applied Economics* 50:5239–5252.

Ho, Katherine, Justin Ho, and Julie Holland Mortimer. 2012. "The use of full-line forcing contracts in the video rental industry". *American Economic Review* 102:686–719.

Hortaçsu, Ali, Aniko Oery, and Kevin R Williams. 2022. *Dynamic price competition: Theory and evidence from airline markets*. Tech. rep. National Bureau of Economic Research.

Hsiao, Cheng. 1983. "Identification". *Handbook of econometrics* 1:223–283.

Huang, Dongling, and Christian Rojas. 2014. "Eliminating the outside good bias in logit models of demand with aggregate data". *Review of Marketing Science* 12:1–36.

Hyvärinen, Aapo, and Erkki Oja. 2000. "Independent component analysis: algorithms and applications". *Neural networks* 13:411–430.

Iizuka, Toshiaki. 2007. "Experts' agency problems: evidence from the prescription drug market in Japan". *The Rand journal of economics* 38:844–862.

Iskrev, Nikolay. 2010. "Local identification in DSGE models". *Journal of Monetary Economics* 57:189–202.

Ivaldi, Marc, and Frank Verboven. 2005. "Quantifying the effects from horizontal mergers in European competition policy". *International Journal of Industrial Organization* 23:669–691.

Jorgenson, Dale W, and Jean-Jacques Laffont. 1974. "Efficient estimation of nonlinear simultaneous equations with additive disturbances". In *Annals of Economic and Social Measurement, Volume 3, number 4*, 615–640. NBER.

Khan, Shakeeb, Fu Ouyang, and Elie Tamer. 2021. "Inference on semiparametric multinomial response models". *Quantitative Economics* 12:743–777.

Khatri, CG, and C Radhakrishna Rao. 1972. "Functional equations and characterization of probability laws through linear functions of random variables". *Journal of Multivariate Analysis* 2:162–173.

Kit, Brian K, et al. 2013. "Trends in sugar-sweetened beverage consumption among youth and adults in the United States: 1999–2010". *The American journal of clinical nutrition* 98:180–188.

Klein, Roger, and Francis Vella. 2010. "Estimating a class of triangular simultaneous equations models without exclusion restrictions". *Journal of Econometrics* 154:154–164.

Kotlarski, Ignacy. 1967. "On characterizing the gamma and the normal distribution". *Pacific Journal of Mathematics* 20:69–76.

Lehmann, Erich Leo, and Joseph P Romano. 2005. *Testing statistical hypotheses*. Vol. 3. Springer.

Lewbel, Arthur. 1997. “Constructing instruments for regressions with measurement error when no additional data are available, with an application to patents and R&D”. *Econometrica: journal of the econometric society*: 1201–1213.

— . 2022. “Kotlarski with a factor loading”. *Journal of Econometrics* 229:176–179.

— . 2012. “Using heteroscedasticity to identify and estimate mismeasured and endogenous regressor models”. *Journal of Business & Economic Statistics* 30:67–80.

Lewbel, Arthur, and Krishna Pendakur. 2009. “Tricks with Hicks: The EASI demand system”. *American Economic Review* 99:827–863.

Li, Qi, and Jeffrey Scott Racine. 2007. *Nonparametric econometrics: theory and practice*. Princeton University Press.

Li, Siran, and Xunjie Zheng. 2020. “A generalization of Lemma 1 in Kotlarski (1967)”. *Statistics & Probability Letters* 165:108814.

Li, Sophia, et al. 2022. “Repositioning and market power after airline mergers”. *The RAND Journal of Economics*.

Liao, Zhipeng. 2013. “Adaptive GMM shrinkage estimation with consistent moment selection”. *Econometric Theory* 29:857–904.

Lin, Xirong. 2023. “Food Demand and Cash Transfers: A Collective Household Approach with Homescan Data”. *Journal of Economic Behavior and Organization* 212:233–259.

Liu, Yizao, and Rigoberto A Lopez. 2016. “The impact of social media conversations on consumer brand choices”. *Marketing Letters* 27:1–13.

Liu, Yizao, Rigoberto A Lopez, and Chen Zhu. 2014. “The impact of four alternative policies to decrease soda consumption”. *Agricultural and Resource Economics Review* 43:53–68.

Lopez, Rigoberto A, and Kristen L Fantuzzi. 2012a. “Demand for carbonated soft drinks: implications for obesity policy”. *Applied Economics* 44:2859–2865.

— . 2012b. “Demand for carbonated soft drinks: implications for obesity policy”. *Applied Economics* 44:2859–2865.

Lopez, Rigoberto A, Yizao Liu, and Chen Zhu. 2015. “TV advertising spillovers and demand for private labels: the case of carbonated soft drinks”. *Applied Economics* 47:2563–2576.

Lorentzen, Peter, John McMillan, and Romain Wacziarg. 2008. “Death and development”. *Journal of economic growth* 13:81–124.

Lu, Zhentong, Xiaoxia Shi, and Jing Tao. 2021. “Semi-nonparametric estimation of random coefficient logit model for aggregate demand”. *Available at SSRN 3503560*.

Lukacs, Eugene. 1970. “Characteristic functions”. Second edition, Griffin, London.

Maddison, Angus. 2003. *Development centre studies the world economy historical statistics: historical statistics*. OECD publishing.

Marinello, Samantha, et al. 2021. “Longer-term impacts of sugar-sweetened beverage taxes on fast-food beverage prices: evidence from Oakland, California, 2-year post-tax”. *Public Health Nutrition* 24:3571–3575.

Marshall, Guillermo. 2015a. “Hassle costs and price discrimination: An empirical welfare analysis”. *American Economic Journal: Applied Economics* 7:123–46.

— . 2015b. “Hassle costs and price discrimination: An empirical welfare analysis”. *American Economic Journal: Applied Economics* 7:123–146.

Martin, Crescent B, et al. 2020. “Nonalcoholic beverage consumption among adults: United States, 2015–2018”.

McConnell, Kenneth E, and Tim T Phipps. 1987. “Identification of preference parameters in hedonic models: Consumer demands with nonlinear budgets”. *Journal of Urban Economics* 22:35–52.

McFadden, Daniel. 1978. “MODELING THE CHOICE OF RESIDENTIAL LOCATION”. *Transportation Research Record.*

— . 1977. “Modelling the choice of residential location”.

Miller, Nathan H, and Matthew C Weinberg. 2017. “Understanding the price effects of the MillerCoors joint venture”. *Econometrica* 85:1763–1791.

Milunovich, George, and Minxian Yang. 2013. “On identifying structural VAR models via ARCH effects”. *Journal of Time Series Econometrics* 5:117–131.

Miravete, Eugenio J, Katja Seim, and Jeff Thurk. 2020. “One markup to rule them all: Taxation by liquor pricing regulation”. *American Economic Journal: Microeconomics* 12:1–41.

— . 2022. *Robust pass-through estimation in discrete choice models*. Tech. rep. Working paper, Department of Economics, University of Pennsylvania.

Moran, Alyssa J, et al. 2019. “Characteristics Associated with Household Purchases of Sugar-Sweetened Beverages in US Restaurants”. *Obesity* 27:339–348.

Navarro, Salvador, and Jin Zhou. 2017. “Identifying agent’s information sets: An application to a lifecycle model of schooling, consumption and labor supply”. *Review of Economic Dynamics* 25:58–92.

Nevo, Aviv. 2000. “A practitioner’s guide to estimation of random-coefficients logit models of demand”. *Journal of economics & management strategy* 9:513–548.

— . 2001. “Measuring market power in the ready-to-eat cereal industry”. *Econometrica* 69:307–342.

Nevo, Aviv, John L Turner, and Jonathan W Williams. 2016. “Usage-based pricing and demand for residential broadband”. *Econometrica* 84:411–443.

Newey, Whitney K, and James L Powell. 2003. “Instrumental variable estimation of nonparametric models”. *Econometrica* 71:1565–1578.

Newmark, Craig M. 2004. "Price-concentration studies: there you go again". *Antitrust Policy Issues*: 9–42.

O'Connell, Martin, and Kate Smith. 2020. "Corrective tax design and market power".

Okrent, Abigail, and Julian Alston. 2012. "The demand for disaggregated food-away-from-home and food-at-home products in the United States". *USDA-ERS Economic Research Report*.

Pakes, Ariel. 2017. "Empirical tools and competition analysis: Past progress and current problems". *International Journal of Industrial Organization* 53:241–266.

Peters, Jonas, Dominik Janzing, and Bernhard Schölkopf. 2017. *Elements of causal inference: foundations and learning algorithms*. The MIT Press.

Petrin, Amil. 2002. "Quantifying the benefits of new products: The case of the minivan". *Journal of political Economy* 110:705–729.

Petrin, Amil, and Kenneth Train. 2010. "A control function approach to endogeneity in consumer choice models". *Journal of marketing research* 47:3–13.

Pissarides, Christopher A, and Guglielmo Weber. 1989. "An expenditure-based estimate of Britain's black economy". *Journal of public economics* 39:17–32.

Qu, Zhongjun, and Denis Tkachenko. 2012. "Identification and frequency domain quasi-maximum likelihood estimation of linearized dynamic stochastic general equilibrium models". *Quantitative Economics* 3:95–132.

Ramsey, Frank P. 1927. "A Contribution to the Theory of Taxation". *The economic journal* 37:47–61.

Rao, C Radhakrishna. 1966. "Characterisation of the distribution of random variables in linear structural relations". *Sankhyā: The Indian Journal of Statistics, Series A*: 251–260.

— . 1971. "Characterization of probability laws by linear functions". *Sankhyā: The Indian Journal of Statistics, Series A*: 265–270.

Reiersøl, Olav. 1950. “Identifiability of a linear relation between variables which are subject to error”. *Econometrica: Journal of the Econometric Society*: 375–389.

Reynaert, Mathias, and Frank Verboven. 2014. “Improving the performance of random coefficients demand models: The role of optimal instruments”. *Journal of Econometrics* 179:83–98.

Rigobon, Roberto. 2003. “Identification through heteroskedasticity”. *Review of Economics and Statistics* 85:777–792.

Robinson, Peter M. 1988. “Root-N-consistent semiparametric regression”. *Econometrica: Journal of the Econometric Society*: 931–954.

Rojas, Christian, and Emily Yucai Wang. 2017. “Do taxes for soda and sugary drinks work? Scanner data evidence from Berkeley and Washington”. *Scanner Data Evidence from Berkeley and Washington (September 23, 2017).*

Rothenberg, Thomas J. 1971. “Identification in parametric models”. *Econometrica: Journal of the Econometric Society*: 577–591.

Rozema, Kyle, and Alex Rees-Jones. 2018. “Price Isn’t Everything: Behavioral Response Around Changes in Sin Taxes”. In *111th Annual Conference on Taxation*. NTA.

Rysman, Marc. 2004. “Competition between networks: A study of the market for yellow pages”. *The Review of Economic Studies* 71:483–512.

Saksena, Michelle J, et al. 2018. *America’s eating habits: food away from home*. Tech. rep. United States Department of Agriculture, Economic Research Service.

Schennach, Susanne M. 2019. “Convolution without independence”. *Journal of econometrics* 211:308–318.

— . 2004. “Estimation of nonlinear models with measurement error”. *Econometrica* 72:33–75.

— . 2013. “Regressions with Berkson errors in covariates—a nonparametric approach”. *The Annals of Statistics*: 1642–1668.

Schennach, Susanne M, and Yingyao Hu. 2013. "Nonparametric identification and semiparametric estimation of classical measurement error models without side information". *Journal of the American Statistical Association* 108:177–186.

Seiler, Stephan, Anna Tuchman, and Song Yao. 2021. "The impact of soda taxes: Pass-through, tax avoidance, and nutritional effects". *Journal of Marketing Research* 58:22–49.

Small, Kenneth A, and Harvey S Rosen. 1981. "Applied welfare economics with discrete choice models". *Econometrica: Journal of the Econometric Society*: 105–130.

Steiner, Robert L. 2004. "The nature and benefits of national brand/private label competition". *Review of Industrial Organization* 24:105–127.

Sweeting, Andrew, James W Roberts, and Chris Gedge. 2020. "A model of dynamic limit pricing with an application to the airline industry". *Journal of Political Economy* 128:1148–1193.

Sydsæter, Knut, et al. 2008. *Further mathematics for economic analysis*. Pearson education.

Székely, GJ, and CR Rao. 2000. "Identifiability of distributions of independent random variables by linear combinations and moments". *Sankhyā: The Indian Journal of Statistics, Series A*: 193–202.

Thompson, T Scott. 1989. "Identification of semiparametric discrete choice models".

Wang, Ao. 2022. "Sieve BLP: A semi-nonparametric model of demand for differentiated products". *Journal of Econometrics*.

Wang, Emily Yucai. 2015. "The impact of soda taxes on consumer welfare: implications of storability and taste heterogeneity". *The RAND Journal of Economics* 46:409–441.

Weinberg, Matthew C, and Daniel Hosken. 2013. "Evidence on the accuracy of merger simulations". *Review of Economics and Statistics* 95:1584–1600.

Well, David N. 2007. “Accounting for the effect of health on economic growth”. *The quarterly journal of economics* 122:1265–1306.

White, Halbert. 1980. “Using least squares to approximate unknown regression functions”. *International economic review*: 149–170.

Wollmann, Thomas G. 2018. “Trucks without bailouts: Equilibrium product characteristics for commercial vehicles”. *American Economic Review* 108:1364–1406.

Wright, Jonathan H. 2003. “Detecting lack of identification in GMM”. *Econometric theory*: 322–330.

Zhen, Chen, et al. 2023. “Do obese and nonobese consumers respond differently to price changes? Implications of preference heterogeneity for obesity-oriented food taxes and subsidies”. *American Journal of Agricultural Economics*.

Zheng, Hualu, Lu Huang, and William Ross. 2019. “Reducing obesity by taxing soft drinks: tax salience and firms’ strategic responses”. *Journal of Public Policy & Marketing* 38:297–315.