Identification and Estimation of Market Size in Discrete Choice Demand Models

[Click here for the latest version.]

Linqi Zhang*

October 2023

Abstract

Within the framework of Berry (1994) and Berry, Levinsohn, and Pakes (1995), I prove that market size can be point identified along with all demand parameters in a random coefficients logit (BLP) model. I require no additional data beyond what is needed to estimate standard BLP models. 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 market in the US, and find that assuming a market size larger than the true estimated size would underestimate merger price increases by 31% on average.

JEL classification: C35, L00, C51, L41, L66

Keywords: Demand estimation, Market size, Discrete choice model, Merger simulation

*Department of Economics, Boston College (email: linqi.zhang@bc.edu). I am grateful to my advisors Arthur Lewbel and Charles Murry for extensive advice and comments. I also thank Richard Sweeney, Frank Verboven, Julie Mortimer, Michael Grubb, Joanna Venator, David Hughes, Shakeeb Khan, Christopher Conlon, Hiroaki Kaido, Philip Haile, Takuya Ura, Alon Eizenberg, Aureo de Paula, Ryan Westphal and seminar participants at Boston College Dissertation Workshop, BU-BC Econometrics Workshop, IIOC 2023, CEA 2023, EARIE 2023 for helpful discussions and suggestions. 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.

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 2 for details and examples). Yet how to choose market size in demand models has received limited formal 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

[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.

[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 (2015), among others.

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 3, I show that the log of product 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 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 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 2 for examples).

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

[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.

[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."

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 3, I start with a multinomial logit demand model to illustrate the problem of mismeasured market size and provide identification results. In section 4, the results are generalized to the random coefficients logit model. Section 5 provides extensions. Section 6 presents an empirical application. Section 7 summarizes additional results provided in the supplemental appendix, and section 8 concludes.

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, 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.

[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.

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 E, Theorem 1 is extended to a nested logit model, highlighting the discrepancies and connections between a nest structure and the market size model.

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 $\mathcal{J}_t = (1, \cdots, J_t)$ 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:

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

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

$$\ln(\pi_{jt}/\pi_{0t}) = X_{jt}'\beta + \xi_{jt} \quad \forall j \in \mathcal{J}_t. \tag{1}$$

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.

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

$$r_{jt}^* = \frac{\pi_{jt}}{\sum_{k=1}^{J_t} \pi_{kt}} \quad \forall j \in \mathcal{J}_t, \tag{2}$$

which is the true conditional choice probability of choosing product j, conditional on purchasing any inside goods. Using equations (1) and (2), we have

$$\ln\left(r_{jt}^*\right) = \ln\left(\frac{\pi_{0t}}{1-\pi_{0t}}\right) + X_{jt}'\beta + \xi_{jt} \quad \forall j \in \mathcal{J}_t. \tag{3}$$

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 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), the model of equation (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

$$\ln\left(\frac{\pi_{0t}}{1-\pi_{0t}}\right) = \ln\left(\gamma W_t - 1\right) \tag{4}$$

for some constant γ. In Appendix C, I relax equation (4) by introducing a random noise term v_t, so that this relationship is approximate rather than exact. In section 4, I further generalize the model by allowing π_{0t} to depend on multiple γ's[7].

The model of equation (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 $\ln\left(\pi_{0t}/(1-\pi_{0t})\right)$ would equal $\ln\left(\gamma W_t - 1\right)$. Equation (4) treats the usual constant γ as unknown rather than known. Furthermore, equation (4) is consistent with a deeper economic model, which I elaborate on in section 5.1.

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

$$\ln\left(r_{jt}\right) = \ln\left(\gamma W_t - 1\right) + X_{jt}'\beta + \xi_{jt} \quad \forall j \in \mathcal{J}_t \tag{5}$$

In Appendix B, I demonstrate the bias introduced in estimating β when employing the conventional approach of equation (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

[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.

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.

3.3 Identification

Here I provide identification of model (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 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. *$E(\xi_{jt} \mid Q_t, X_{1t}, \ldots, X_{J_t t}) = 0$, 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 (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 (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*

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

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 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 (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.

3.3.1 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.

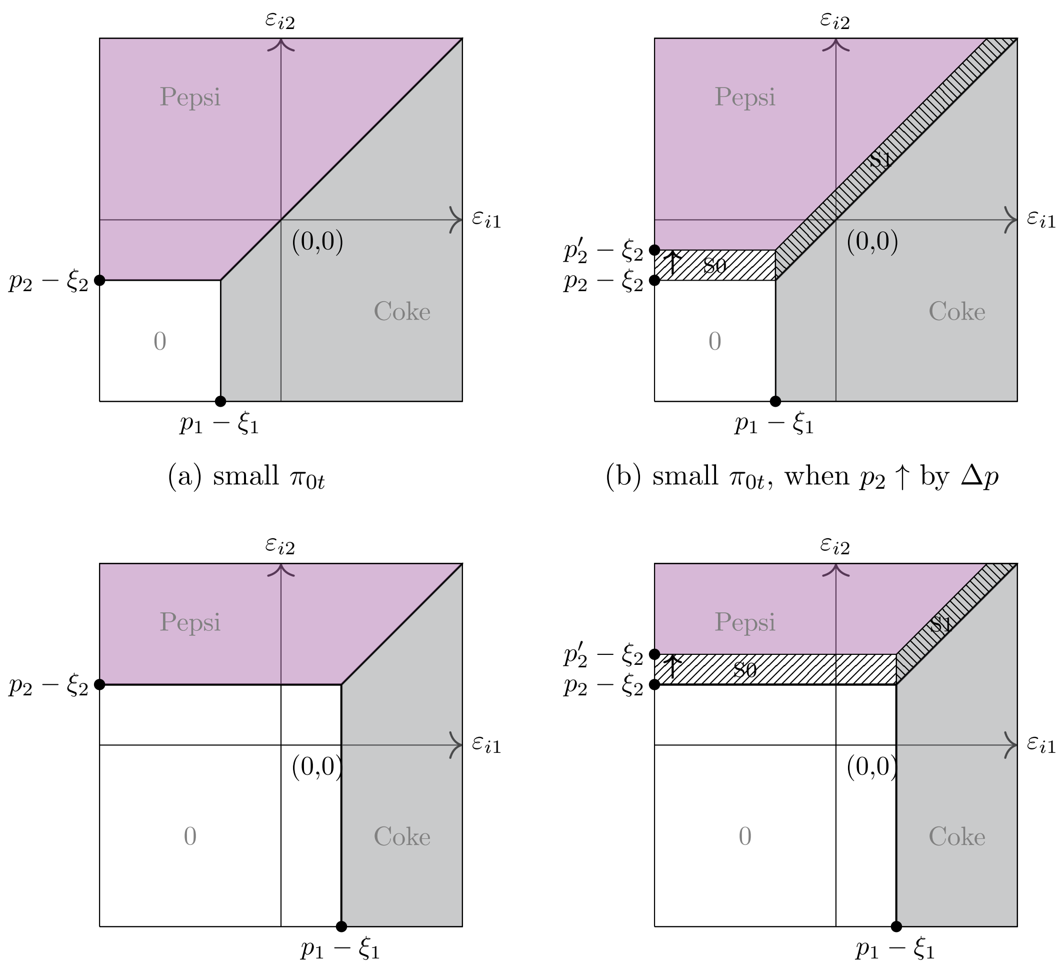
Figure 1 illustrates the aforementioned intuition. In a simplified model where $\delta_{jt} = -p_{jt} + \xi_{jt}$, with two goods ($j = 1$ Coke and $j = 2$ Pepsi), the space of ϵ_{ij} is partitioned into three regions, each corresponding to the choice of $j = 0, 1, 2$ (Berry and Haile 2014 and Thompson 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, 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 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 4, the simplified example depicted in Figure 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: Intuition for Identification in Multinomial Logit Demand Model

3.4 The Nested Logit Demand Model

In Appendix 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.

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.

4.1 Demand Model and Market Size

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

$$U_{ijt} = X'_{jt}\beta_i + \xi_{jt} + \varepsilon_{ijt}, \tag{6}$$

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 3, let δ_{jt} denote the mean utility $X'_{jt}\beta + \xi_{jt}$. Combining equations we have

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

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

$$\pi_{jt}\left(\delta_t, X_t^{(2)}; \sigma\right) = \int \frac{exp(\delta_{jt} + \sum_l \sigma_l x_{jtl}^{(2)} \nu_{il})}{1 + \sum_{k=1}^{J_t} exp(\delta_{kt} + \sum_l \sigma_l x_{ktl}^{(2)} \nu_{il})} f_\nu(\nu) d\nu, \tag{7}$$

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

$$\frac{N_t}{\sum_{k=1}^{K} \gamma_{k1} M_{kt}^{\gamma_{k2}}} = \pi_t \left(\delta_t, X_t^{(2)}; \sigma \right), \tag{8}$$

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*.

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 (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 (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 (7) and (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

[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.

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

$$\xi_{jt}\left(N_t, M_t, X_t; \theta\right) = \delta_{jt}\left(N_t, M_t, X_t^{(2)}; \gamma, \sigma\right) - X'_{jt}\beta. \tag{9}$$

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 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,*

$$E\left(h_{jt}(\tilde{\theta})\right) = 0 \iff \tilde{\theta} = \theta_0, \text{ for all } \tilde{\theta} \in \Theta \tag{10}$$

θ_0 is locally identified if (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 (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).

4.2.1 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 (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

$$
\begin{aligned}
J_1 &= \frac{\partial h_{jt}(\theta)}{\partial \gamma_{k1}} = \begin{vmatrix} \frac{\partial \pi_{1t}}{\partial \delta_{1t}} & \cdots & \frac{\partial \pi_{1t}}{\partial \delta_{Jt}} \\ \vdots & \ddots & \vdots \\ \frac{\partial \pi_{Jt}}{\partial \delta_{1t}} & \cdots & \frac{\partial \pi_{Jt}}{\partial \delta_{Jt}} \end{vmatrix}^{-1} \underbrace{\begin{vmatrix} \frac{\partial \pi_{1t}}{\partial \delta_{1t}} & \cdots & \pi_{1t} & \cdots & \frac{\partial \pi_{1t}}{\partial \delta_{Jt}} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \frac{\partial \pi_{Jt}}{\partial \delta_{1t}} & \cdots & \pi_{Jt} & \cdots & \frac{\partial \pi_{Jt}}{\partial \delta_{Jt}} \end{vmatrix}}_{(\pi_{1t},\cdots,\pi_{Jt})' \text{ is in the } j\text{-th column}} \frac{M_{kt}^{\gamma_{k2}}}{\sum_k \gamma_{k1} M_{kt}^{\gamma_{k2}}} \phi_j(Z_t), \\
&= \Psi_{jt}\left(\delta_t, X_t^{(2)}; \sigma\right) \frac{M_{kt}^{\gamma_{k2}}}{\sum_k \gamma_{k1} M_{kt}^{\gamma_{k2}}} \phi_j(Z_t), \qquad (11)
\end{aligned}
$$

and

$$
J_2 = \frac{\partial h_{jt}(\theta)}{\partial \gamma_{k2}} = \Psi_{jt}\left(\delta_t, X_t^{(2)}; \sigma\right) \frac{\gamma_{k1} \ln(M_{kt}) M_{kt}^{\gamma_{k2}}}{\sum_k \gamma_{k1} M_{kt}^{\gamma_{k2}}} \phi_j(Z_t) \qquad (12)
$$

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 (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 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

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

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 nonsingular, 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*

$$
Cov\left(\Psi_{jt}\left(\delta_t, X_t^{(2)}; \sigma\right), \phi_j^{(2)}(Z_t)\right) \neq 0, \tag{13}
$$

and

$$
Cov\left(\Phi_{jt}\left(\delta_t, X_t^{(2)}; \sigma\right), \phi_j^{(3)}(Z_t)\right) \neq 0, \tag{14}
$$

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

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

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

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

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 (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 (13) and (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.

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 (9) as

$$\xi_{jt}\left(N_t, M_t, X_t; \theta\right) = \delta_{jt}\left(N_t, M_t, X_t^{(2)}; \lambda\right) - X_{jt}'\beta. \tag{15}$$

Equation (15) encompasses equation (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.

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 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 (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 (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).

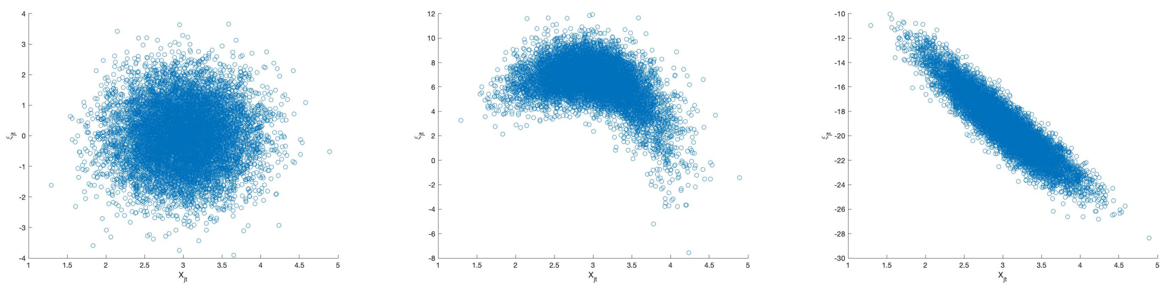
4.4.1 A Numerical Illustration

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 (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 2, I visually illustrate the intuition for identification and why one can distinguish γ and σ.

Figure 2 plots X_{jt} against the implied residual function $\xi_{jt}(\sigma, \gamma)$ for different values of (σ, γ). As depicted in Figure 2(a), there is no correlation between ξ and the X at the true parameter values. Figure 2(b) shows that when σ is different from the truth, it exhibits a hump-shaped correlation and Figure 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 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) $\sigma = \sigma^0 = 5, \gamma = 4$.

4.5 Market Fixed Effects

In Appendix 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 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.

5 Extensions

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 (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 (2016) (Example 1 in their paper), using a sieve space approximation to the distribution of random coefficients, we can write

$$\pi_{jt}(\delta_{jt};\sigma) = \sum_{r=1}^{R} \sigma_r \frac{\exp\left(\delta_{jt} + \sum_l \eta_l^r x_{jtl}\right)}{1 + \sum_{k=1}^{J_t} \exp(\delta_{kt} + \sum_l \eta_l^r x_{ktl})} \tag{16}$$

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

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

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 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 (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 (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 (2016), and the choice probability cannot be expressed in the form of equation (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.

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

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 (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

$$\frac{N_{jt}}{s(M_t)} = \pi_j\left(\delta_t, X_t^{(2)}\right), \quad j = 1, \cdots, J. \tag{17}$$

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 $g(m) = \partial E(\ln(N_{jt}) \mid m)/\partial m$.*

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

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

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

from which we can identify γ_2 by construction: that is, $\gamma_2 = \partial E(\ln(N_{jt}) \mid M_t = m)/\partial \ln(m)$. 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 $\gamma_1 = (\partial E(\ln(N_{jt}) \mid M_{1t} = m_1, M_{2t} = m_2)/\partial m_2) / (\partial E(\ln(N_{jt}) \mid M_{1t} = m_1, M_{2t} = m_2)/\partial m_1)$.

After establishing point identification of market size, the empirical shares on the left hand side of equation (17) are identified. It would suffice to impose assumptions made in Berry and Haile (2014) to obtain nonparametric identification of the demand model.

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 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 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 K, I have a second merger analysis in the Ready-to-Eat Cereal market showing that our method works in different empirical contexts.

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 2015 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 2012, 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

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

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.

6.2 Data

I use a panel of weekly scanner data from Nielsen for our analysis. The Nielsen 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 week, resulting in a total of 10504 DMA-week markets[13].

In addition to the Nielsen 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 (2015), 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 (2005), 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

[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.

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

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 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.

6.3 Demand Model

As in section 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 (2015). Similarly, in the airline industry, *carrier presence* is often considered as an exogenous attribute. The economic interpretation

[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.

of in-store presence in the present context aligns closely with carrier presence in the airline market. Just as carrier presence may raise concerns of endogeneity, it has typically been addressed through via fixed effects.

Table 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 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.

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 (2015) 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 (2012), 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 (2015) 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 (2015) 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 rep-

[15] Another distinction between Eizenberg and Salvo (2015) and other papers that use the US data is that the market size in Eizenberg and Salvo (2015) 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.

resents 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.

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 2015; 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 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

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

[17] If stores make assortment decisions after the realization of all demand shocks (as assumed in Ciliberto,

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.

6.6 Results

Table 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 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 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,

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.

[18]To verify that the estimated γ achieves global minimum for the random coefficients logit model, in Appendix 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.

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 market 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 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 4 are based on estimates from columns 3 to 5 of Table 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 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 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 J, which allow one to assess the impact of hypothetical soda taxes.

Finally, I simulate a merger between the largest manufacturer and private label man-

[20] For example, the estimated own-price elasticities in Dubé (2005) 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.

ufacturers. The merger simulation abstracts away from cost reduction, or changes in the model of competition (e.g. coordination between other firms). Table 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 (2005), 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.

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.

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).

References

Ai, Chunrong, and Xiaohong Chen. 2003. "Efficient estimation of models with conditional moment restrictions containing unknown functions". *Econometrica* 71 (6): 1795–1843.

Amemiya, Takeshi. 1974. "The nonlinear two-stage least-squares estimator". *Journal of econometrics* 2 (2): 105–110.

Andrews, Donald WK. 2017. "Examples of L2-complete and boundedly-complete distributions". *Journal of econometrics* 199 (2): 213–220.

Armstrong, Timothy B. 2016. "Large market asymptotics for differentiated product demand estimators with economic models of supply". *Econometrica* 84 (5): 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.

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 (3): 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 (5): 2087–2111.

Berry, Steven, and Philip A Haile. 2014. "Identification in differentiated products markets using market level data". *Econometrica* 82 (5): 1749–1797.

Berry, Steven, and Panle Jia. 2010. "Tracing the woes: An empirical analysis of the airline industry". *American Economic Journal: Microeconomics* 2 (3): 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 (2): 625–652.

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.

Bonnet, Céline, and Vincent Réquillart. 2013. "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 (5): 1088–1108.

Cardell, N Scott. 1997. "Variance components structures for the extreme-value and logistic distributions with application to models of heterogeneity". *Econometric Theory* 13 (2): 185–213.

Chen, Xiaohong, and Demian Pouzo. 2009. "Efficient estimation of semiparametric conditional moment models with possibly nonsmooth residuals". *Journal of Econometrics* 152 (1): 46–60.

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 (11): 2995–3038.

Conlon, Christopher, and Jeff Gortmaker. 2020. "Best practices for differentiated products demand estimation with pyblp". *The RAND Journal of Economics* 51 (4): 1108–1161.

Conlon, Christopher, and Julie Holland Mortimer. 2021. "Empirical properties of diversion ratios". *The RAND Journal of Economics* 52 (4): 693–726.

Dubé, Jean-Pierre. 2005. “Product differentiation and mergers in the carbonated soft drink industry”. *Journal of Economics & Management Strategy* 14 (4): 879–904.

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. 2015. “The rise of fringe competitors in the wake of an emerging middle class: An empirical analysis”. *American Economic Journal: Applied Economics* 7 (3): 85–122.

Fisher, Franklin M. 1966. *The identification problem in econometrics*. McGraw-Hill.

Fox, Jeremy T, Kyoo il Kim, and Chenyu Yang. 2016. “A simple nonparametric approach to estimating the distribution of random coefficients in structural models”. *Journal of Econometrics* 195 (2): 236–254.

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 (3): 493–520.

Gospodinov, Nikolay, and Serena Ng. 2015. “Minimum distance estimation of possibly non-invertible moving average models”. *Journal of Business & Economic Statistics* 33 (3): 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.

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 (49): 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 (2): 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): 1–36.

Iizuka, Toshiaki. 2007. "Experts' agency problems: evidence from the prescription drug market in Japan". *The Rand journal of economics* 38 (3): 844–862.

Iskrev, Nikolay. 2010. "Local identification in DSGE models". *Journal of Monetary Economics* 57 (2): 189–202.

Ivaldi, Marc, and Frank Verboven. 2005. "Quantifying the effects from horizontal mergers in European competition policy". *International Journal of Industrial Organization* 23 (9-10): 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 (3): 743–777.

Lehmann, Erich Leo, and Joseph P Romano. 2005. *Testing statistical hypotheses.* Vol. 3. Springer.

Li, Qi, and Jeffrey Scott Racine. 2007. *Nonparametric econometrics: theory and practice.* Princeton University Press.

Li, Sophia, et al. 2022. "Repositioning and market power after airline mergers". *The RAND Journal of Economics.*

Liu, Yizao, and Rigoberto A Lopez. 2016. "The impact of social media conversations on consumer brand choices". *Marketing Letters* 27 (1): 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 (1): 53–68.

Lopez, Rigoberto A, and Kristen L Fantuzzi. 2012. "Demand for carbonated soft drinks: implications for obesity policy". *Applied Economics* 44 (22): 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 (25): 2563–2576.

Lu, Zhentong, Xiaoxia Shi, and Jing Tao. 2021. "Semi-nonparametric estimation of random coefficient logit model for aggregate demand". *Available at SSRN 3503560.*

Marshall, Guillermo. 2015. "Hassle costs and price discrimination: An empirical welfare analysis". *American Economic Journal: Applied Economics* 7 (3): 123–46.

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 (1): 35–52.

McFadden, Daniel. 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 (6): 1763–1791.

Milunovich, George, and Minxian Yang. 2013. "On identifying structural VAR models via ARCH effects". *Journal of Time Series Econometrics* 5 (2): 117–131.

Nevo, Aviv. 2000. "A practitioner's guide to estimation of random-coefficients logit models of demand". *Journal of economics & management strategy* 9 (4): 513–548.

— . 2001. "Measuring market power in the ready-to-eat cereal industry". *Econometrica* 69 (2): 307–342.

Newey, Whitney K, and James L Powell. 2003. "Instrumental variable estimation of non-parametric models". *Econometrica* 71 (5): 1565–1578.

Newmark, Craig M. 2004. "Price-concentration studies: there you go again". *Antitrust Policy Issues*: 9–42.

Pakes, Ariel. 2017. "Empirical tools and competition analysis: Past progress and current problems". *International Journal of Industrial Organization* 53:241–266.

Petrin, Amil. 2002. "Quantifying the benefits of new products: The case of the minivan". *Journal of political Economy* 110 (4): 705–729.

Petrin, Amil, and Kenneth Train. 2010. "A control function approach to endogeneity in consumer choice models". *Journal of marketing research* 47 (1): 3–13.

Pissarides, Christopher A, and Guglielmo Weber. 1989. "An expenditure-based estimate of Britain's black economy". *Journal of public economics* 39 (1): 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 (1): 95–132.

Reynaert, Mathias, and Frank Verboven. 2014. "Improving the performance of random coefficients demand models: The role of optimal instruments". *Journal of Econometrics* 179 (1): 83–98.

Robinson, Peter M. 1988. "Root-N-consistent semiparametric regression". *Econometrica: Journal of the Econometric Society*: 931–954.

Rothenberg, Thomas J. 1971. "Identification in parametric models". *Econometrica: Journal of the Econometric Society*: 577–591.

Rysman, Marc. 2004. "Competition between networks: A study of the market for yellow pages". *The Review of Economic Studies* 71 (2): 483–512.

Steiner, Robert L. 2004. "The nature and benefits of national brand/private label competition". *Review of Industrial Organization* 24 (2): 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 (3): 1148–1193.

Sydsæter, Knut, et al. 2008. *Further mathematics for economic analysis.* Pearson education.

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.*

Weinberg, Matthew C, and Daniel Hosken. 2013. "Evidence on the accuracy of merger simulations". *Review of Economics and Statistics* 95 (5): 1584–1600.

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 (6): 1364–1406.

Wright, Jonathan H. 2003. "Detecting lack of identification in GMM". *Econometric theory*: 322–330.

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 (3): 297–315.

Tables

Table 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 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 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 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 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.

Identification and Estimation of Market Size in Discrete Choice Demand Models - Supplemental Appendix

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

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

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 $q \in \mathrm{Supp}(Q_t)$ and $x \in \mathrm{Supp}(X_{jt})$. Taking the derivative of $g(c, q, x)$ with respect to c, we have

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

The identification then follows from the assumption that there exists (q, x) on the support of (Q_t, X_{jt}) such that $\partial E\left(-\frac{W_t}{c W_t - 1} \mid Q_t = q, X_{jt} = x\right) / \partial q$ 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. $\square$

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 $g_k(\delta) < \delta_k$:

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

□

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

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

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 $(\kappa_1, \cdots, \kappa_T)$ 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

$$\frac{\partial E\left(\ln\left(r_{jt}\right) \mid Q_t = q, X_{jt} = x\right)}{\partial q} = \frac{1}{1-\sigma}\frac{\partial E\left(\ln\left(\gamma W_t - 1\right) \mid Q_t = q, X_{jt} = x\right)}{\partial q}. \tag{18}$$

Taking second-order derivative with respect to q yields

$$\frac{\partial^2 E\left(\ln\left(r_{jt}\right) \mid Q_t = q, X_{jt} = x\right)}{\partial q^2} = \frac{1}{1-\sigma} \frac{\partial^2 E\left(\ln\left(\gamma W_t - 1\right) \mid Q_t = q, X_{jt} = x\right)}{\partial q^2}. \tag{19}$$

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 (19) by (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 $q \in \mathrm{Supp}(Q_t)$ and $x \in \mathrm{Supp}(X_{jt})$. Taking the derivative of $f(c,q,x)$ with respect to c, we have

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

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. $\square$

B Bias Caused by Mismeasured Market Size

I show that the usual approach that estimates demand based on equation (1) with a mismeasured market size will lead to biased estimates of β. To see this, suppose the true model is given by equation (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

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

with

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

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) and (5). The third equality follows from equation (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.

C Extension of the Simple Logit Case

r_{jt} and r^*_{jt} are defined as in section 3. Now we assume

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

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

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

where

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

To achieve identification as in section 3, we only need to modify the mean independence assumption such that $E\left(u_{jt} \mid Q_t, X_{1t}, \ldots, X_{J_t t}\right) = 0$, where everything else is defined as in section 3.

D Market Fixed Effects Approach for Simple Logit

Returning to equation (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 (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 4.5).

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), with an additional regressor $\ln(\pi_{j|gt})$,

$$\ln(\pi_{jt}/\pi_{0t}) = X'_{jt}\beta + \rho \ln(\pi_{j|gt}) + \xi_{jt}, \tag{20}$$

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 3.2. Then we can rewrite (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 3.2, we assume equation (4) hold. Combining above equations and assumptions we get the estimating equation for the nested logit model

$$\ln\left(r_{jt}\right) = \frac{1}{1-\rho}\ln\left(\gamma W_t - 1\right) + X'_{jt}\frac{\beta}{1-\rho} + \frac{\xi_{jt}}{1-\rho}. \tag{21}$$

Theorem 6. *Given Assumption 1 and equation (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*

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

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 (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 (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 (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.

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

$$E\begin{bmatrix}\delta_t - \beta_0 \\ (\delta_t-\beta_0)Z_{1t} \\ (\delta_t-\beta_0)Z_{2t}\end{bmatrix} = 0. \tag{22}$$

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 (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

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

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 (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.

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

$$E\left[\left(\delta_{jt}\left(N_{jt}, M_t, X_t^{(2)}; \tilde{\gamma}, \sigma_0\right) - X'_{jt}\beta_0 + \delta_{jt}\left(N_{jt}, M_t, X_t^{(2)}; \gamma_0, \sigma_0\right) \right.\right.$$
$$\left.\left. - \delta_{jt}\left(N_{jt}, M_t, X_t^{(2)}; \tilde{\gamma}, \sigma_0\right)\right)\phi_j(Z_t)\right] = 0, \quad (23)$$

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 (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 (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

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

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

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.

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 3

$$U_{ijt} = \beta_0 + (\beta_p + \sigma\nu_i)P_{jt} + \beta_1 X_{1,jt} + \xi_{jt} + \varepsilon_{ijt}, \tag{24}$$

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 (7). By equations (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

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

where product j is produced by firm f and $\mathcal{J}_f$ 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).

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 (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 H.1 to H.3 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 H.1, 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 H.2 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 H.1 (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 H.3, 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 H.2.

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 H.4, we report the estimates from our proposed method of equation (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 2. 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 H.1.

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 H.1: 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 H.2: 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 H.3: Monte Carlo Results: Random Coefficient on Price, True $\gamma = 1$

		$\tilde{\gamma} = 1$			$\tilde{\gamma} = 2$			$\tilde{\gamma} = 4$		
	TRUE	MEAN	SD	MED	MEAN	SD	MED	MEAN	SD	MED
σ	0	0.013	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 H.4: 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.

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 H.5 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 H.3. 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 H.5: Sensitivity to Market Size Assumptions in Random Coefficients Logit, True $\gamma = 1$

	TRUE	$\tilde{\gamma} = 1$ MEAN	SD	MED	$\tilde{\gamma} = 2$ MEAN	SD	MED	$\tilde{\gamma} = 4$ MEAN	SD	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 $\tilde{\gamma} = 1, 2, 4, 8$.

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 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 H.6 and H.7 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 H.6 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 H.7, 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 H.6: 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 H.7: 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.

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 $\mathcal{J}_f$ 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

$$\sum_{k \in \mathcal{J}_f} \left(\frac{p_k - mc_k}{p_k} \right) \eta_{k,j} \pi_k + \pi_j = 0, \tag{25}$$

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:

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

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 (25), giving

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

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.

J Additional Results for the CSD Application

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 J.1 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.

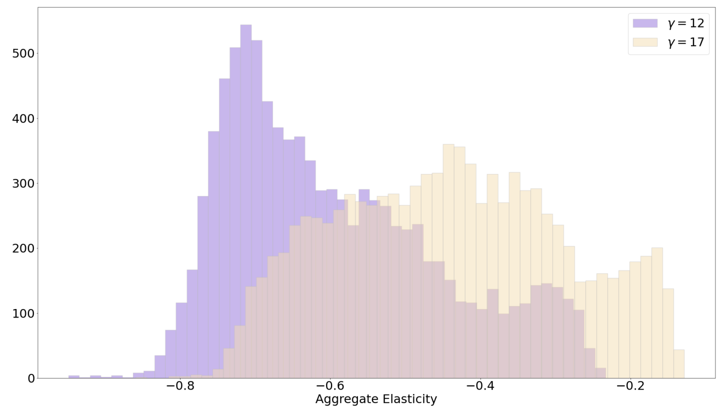


Figure J.1: Distribution of Aggregate Elasticities across Markets
Notes: The figure shows the aggregate elasticities of demand across markets for $\gamma = 12$ and 17.

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.

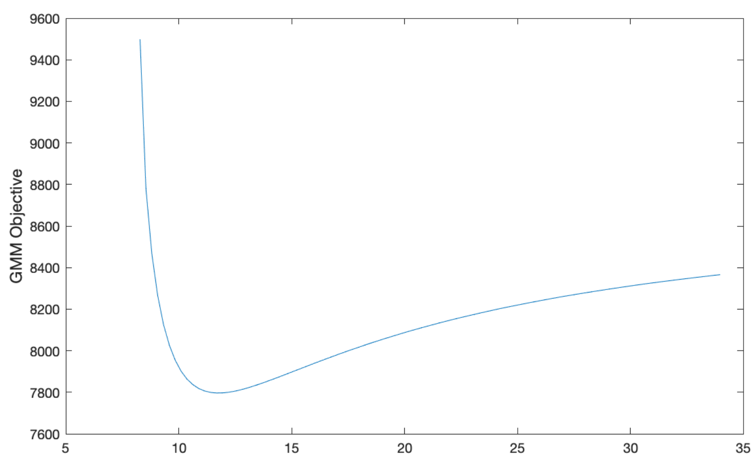


Figure J.2: Profiled GMM Objective

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

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, ν_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 K.1 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 K.1 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

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

Table K.1: 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.

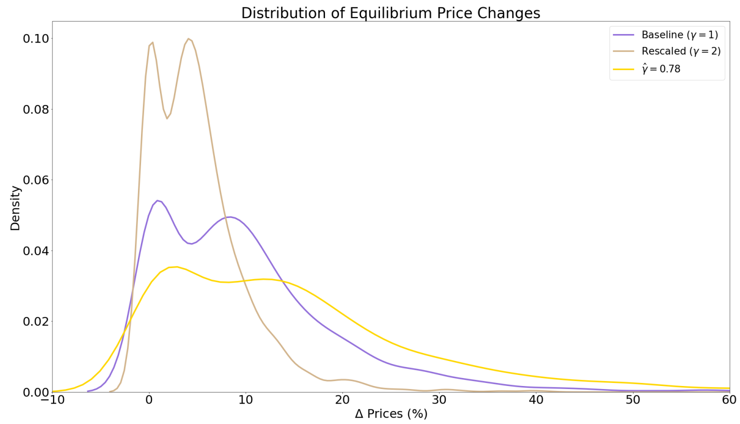


Figure K.1: Equilibrium Price Changes

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

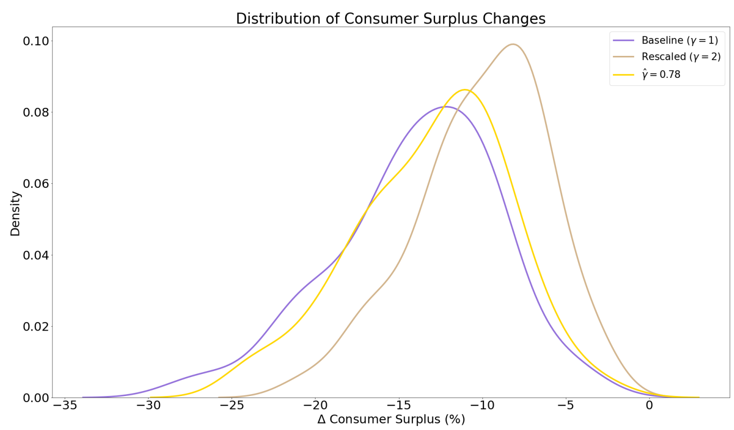


Figure K.2: Consumer Surplus Changes

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

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:

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

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 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 (9) can be rewritten as

$$\xi_{jt}\left(\frac{N_t}{\sum_k(\lambda_{\gamma_{k1}}+1)M_t^{\lambda_{\gamma_{k2}}}}, X_t; \lambda_\sigma, \beta\right) = \delta_{jt}\left(\frac{N_t}{\sum_k(\lambda_{\gamma_{k1}}+1)M_t^{\lambda_{\gamma_{k2}}}}, X_t^{(2)}; \lambda_\sigma\right) - X_{jt}'\beta. \quad (26)$$

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 4.3 can also be obtained from linearizing the inverse demand function around the true λ_0

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

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