Identification and Estimation of Market Size in Discrete Choice Demand Models

[Preliminary Draft. Please do not circulate.]

Linqi Zhang*

June 7, 2023

Abstract

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

*I am grateful to my advisors Arthur Lewbel and Charles Murry for their invaluable support. I also thank Richard Sweeney, Julie Mortimer, Michael Grubb, Joanna Venator, Shakeeb Khan, Frank Verboven, 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 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

Models of differentiated product demand systems are essential for the analysis of economic topics such as market power, mergers, firm entry and tax policy in a wide range of industries. Most modern applied work using differentiated product demand follows the estimation approach developed by Berry (1994) and Berry, Levinsohn, and Pakes (1995) (hereafter referred to as BLP). This method involves using empirically observed aggregate market shares to estimate choice probabilities predicted by discrete choice models.

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). Potential purchases are generally unobservable and are therefore a source of possible mismeasurement of market size. For example, when estimating demand in the airline industry, a market is typically defined as an origin-destination pair of cities. How then does one properly determine the number of potential flyers in that pair? Is it just people who have chosen to travel by other means between the two cities? What about those who might have decided to travel if prices were lower? Is the market the entire population in the end-point cities, which likely includes some who would never travel to the destination?

Many empirical results are sensitive to market size (see section 1.1 for details and examples). Yet how to choose market size in demand models has received limited attention in the literature. A few researchers have commented on this problem [1], 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[2]. For example, Eizenberg and Salvo (2015) assume the soft drink market size, in liters, is six times the population. This constant is not observed or estimated in general but is chosen in an ad hoc fashion by the researchers, justified based on industrial background or consumer behavior. This constant is not a free normalization as it affects the estimates of demand parameters that govern elasticities 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, we 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 equals a general function of observed variables and unknown parameters. So, for example, in the airline demand model, market size can be a function of the population in the origin city, population in the destination

[1]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".

[2]Well-known examples include Nevo (2001), Petrin (2002), Rysman (2004), Berto Villas-Boas (2007), Berry and Jia (2010) and Eizenberg and Salvo (2015).

city, other variables such as whether each city is a hub or not, and a vector of unknown parameters that are identified and estimated along with the rest of the BLP model.

Our identification exploits two important features: exogenous variation that shifts quantities across markets and the nonlinearity of demand model. We do not require side information such as micro-moments or additional data beyond those normally used in standard BLP applications. A key insight is that, when we observe any exogenous changes in product characteristics, it alters the total sales of inside goods, and the extent to which total sales respond to this exogenous variation depends on the size of the outside market. Why this variation has extra identifying power for parameters beyond ordinary demand coefficients? Let us consider the plain multinomial logit model. We will show, in section 2, that the log of product share is linear in covariates but nonlinear in market size parameters, making identification possible.

Other theoretical results in this paper include: (a) Calculating the bias caused by mismeasured market size. (b) Providing low level assumptions on instruments required for identification of the standard random coefficient BLP model, both with and without identifying market size as well. (c) Showing identification for models where market size is an unknown function of observed variables. (d) Providing stronger conditions that permit point identification and estimation of market size even when the demand model is not known or is nonparametric (e.g., in Berry and Haile (2014)'s nonparametric BLP framework). This result can also be used to test market size specification without estimating the demand model at all. (e) Providing simpler identification results for the plain multinomial logit model using, e.g., market fixed effects.

In addition to the identification results, we demonstrate how our proposed method is related to but different from commonly used approaches (e.g., implement a nested logit model) to reduce biases from unknown market size. We formally show that the intuition in prior literature and their attempts have some theoretical basis. Nevertheless, they are not equivalent to our approach and cannot eliminate all biases. We also show that a special case of nonparametric estimation of random coefficients is equivalent to estimating the market size, but it has to impose particular assumptions on the distribution of random coefficients.

On the basis of these identification results, we apply our method to a merger simulation of carbonated soft drink companies. We choose the soda market for two reasons. First, the soda market is frequently studied using structural methods that entail estimating random coefficients logit models. Second, this market satisfies the conditions for strong identification, which we will state in the Monte Carlo simulation section. In the merger analysis, we use both our 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, we predict a price effect that is 45% higher than assuming 17 servings per week as in the literature. In Appendix F, we provide a second merger analysis using the constructed cereal data from Nevo (2000). In these counterfactual simulations, the gains from our proposed correction are large.

Furthermore, using Monte Carlo simulations, we show what parameters are most sensitive to errors in market size measurements[3] and whether adding random coefficients mitigates the bias. We also show that our proposed approach performs well especially when the true share of the outside option is not too large, and so the method will generally be useful in applications.

Our proposed method is transparent and simple to implement. It requires only estimating a few extra nonlinear parameters in addition to what people normally estimate using the standard BLP. Practitioners may have tried to estimate market size, but the lack of identification theorem and the unsatisfactory empirical performance or numerical problem of the estimator is the reason why it is not widely seen in applied work. We provide conditions under which the market size is identified, discuss what variation in the data helps identify the market size, and how to test the relevance of these instruments. We hope this paper can partly eliminate practitioners' uncertainty about the market size. Moreover, whenever the market size itself is important to practitioners or regulators, this can be a potential method to infer the size of the market. Note, that although the solution is simple, it is more than just adding a regressor or adding market fixed-effects. Our identification theorems focus on market-level aggregate data, but we believe it can extend to random coefficients logit with individual-level purchase data (or micro-BLP).

1.1 Why Market Size Matters An argument for not correcting this issue is the intuition that a random coefficient on price (or on the constant term[4]), or a nesting parameter in the nested logit model, can absorb some of the effects of mismeasured market size. For example, Miller and Weinberg (2017) states that the nesting parameter ensures that estimates are not too sensitive to the market size measure. In some empirical applications where random coefficients are used, varying the assumptions of market size sometimes yield similar results for certain calculations such as own- and cross-price elasticities (see, e.g., Rysman 2004; Iizuka 2007; and Duch-Brown et al. 2017). However, biases are not fully eliminated by implementing more flexible demand models or including a large number of random coefficients. The extent to which these methods fix the problem is applications specific and depends on the underlying data generating process.

Market size assumptions have a much larger effect on some types of calculations than

[3]The direction and magnitude of the bias in demand parameters vary with dgp, but the overall effect should be that as the true market size increases, inside goods become less attractive, so consumers are either more sensitive to price or they have a lower constant term in utility.

[4]Heterogeneity in the constant term is believed to govern the substitution between inside and outside goods, which market size also affects.

others. Evidence from our Monte Carlo simulations suggests that getting the market size wrong can greatly bias estimates of outside good elasticities, outside good diversion ratios, and choice probabilities[5]. Therefore, for certain types of counterfactual simulations, particularly those related to the outside market size, different market size assumptions lead to substantially different results. An example is the effects of new goods entry or exit. The value of including a particular good in the choice set (i.e., Willingness-To-Pay) can be relevant because WTP is essentially a measure of change in outside good choice probabilities (see discussion in Conlon and Mortimer 2021). Another prominent example is effects of mergers. In Table 1, we briefly review a list of merger analysis papers where some of their calculations are sensitive to assumptions of market size. The papers included in the table cover different logit-based demand models with various specifications. Their robustness checks suggest that market size impacts aggregate elasticities, simulated price changes, and consumer welfare.

In addition to the existing evidence, we conduct two merger simulations, one in the soft drink market and one in the ready-to-eat cereal market. The extent to which consumers substitute to alternative products vs the outside option greatly affects the impact of a merger on prices. Predicting price change is crucial for correctly assessing the welfare effects of a merger. We show that wrong assumptions regarding market size yield substantially biased estimates of the loss (or gain) to society of a merger, by under- or overestimating likely price changes. 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[6]. This suggests that obtaining a consistent estimate of the true market size is important by itself, in addition to its use in removing model estimate biases.

1.2 Literature Review In the empirical industrial organization (IO) literature, a market size is often assumed rather than observed or estimated. Leading examples are previous works that use the number of households in the US as the market size to study the automobile industry (see BLP and Petrin 2002). More examples include Nevo (2001) (ready-to-eat cereal market), Rysman (2004) (yellow pages market), Berto Villas-Boas (2007) (yogurt market), Berry and Jia (2010) (airline market), Ho, Ho, and Mortimer (2012) (video rental

[5]Similarly, see, for example, Conlon and Mortimer (2021) Table 4, which shows that diversion ratios to other goods are not too sensitive to market size in the BLP automobile application, and own elasticities are not overly sensitive to market size in Nevo's cereal application, but outside diversion ratios and aggregate elasticities change a lot.

[6]For instance, in a talk given in the DOJ/FTC merger workshop, Newmark (2004) argues that market size/population could play an important role in the price-concentration study in merger cases. Additionally, on the supply side, firms predict product quantities on the basis of potential market size. In the Comments of DOJ on Joint Application Of American Airlines Et Al., it states that "To model the benefits of an alliance, airlines typically use QSI models to forecast traffic changes ... Given a fixed market size, passengers are assigned based on relative attractiveness of different airline offerings."

market), Ghose, Ipeirotis, and Li (2012) (hotel market), and Eizenberg and Salvo (2015) (soft drink market), among others, where the total market size is assumed to be observed as the population measure multiplied by a constant.

Several empirical articles recognize the issue and explicitly incorporate market size estimation into demand models (e.g., Greenstein 1996; Berry, Carnall, and Spiller 2006; Chu, Leslie, and Sorensen 2011; Sweeting, Roberts, and Gedge 2020; and Li et al. 2022). Greenstein (1996) makes a functional form assumption that the market size is a linear function of the total sales. The main distinction between Greenstein (1996) and the present paper is that his demand system is a vertical model similar to that used by Bresnahan (1987), which abstracts away from individual heterogeneity in preferences. Chu, Leslie, and Sorensen (2011) use supply side pricing conditions as additional moments to estimate market size. While this method does not impose any functional form assumptions on the market size, it requires one to observe marginal costs of firms. Berry, Carnall, and Spiller (2006) employ an idea similar to ours in estimating market size in the BLP framework. Their estimation approach, where the market size is restricted to be proportional to a single variable, can be interpreted as a special case of the more general model in the present paper. Although the authors estimate the market size, unlike us they do not discuss identification. Sweeting, Roberts, and Gedge (2020) and Li et al. (2022) both exploit a variety of market-level measures in some way to help proxying the market size. Specifically, they estimate a generalized gravity equation and obtain the predicted value from a regression of the number of passengers on variables such as city-pair distance and lagged passengers flow. The authors define market size as proportional to the expected total passengers predicted from the gravity equation, but still leave the factor of proportionality as a choice of the researcher. Hortaçsu, Oery, and Williams (2022) exploit search data to estimate a Poisson arrival process and the arrival rate is used as a proxy measure of market size. Their method is applied in a setting where individual choice data is observed, whereas we consider settings where one only has access to market level data.

The closest result to the present paper 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 some market characteristics (Chamberlain's device). They estimate demand models using the control function method to handle the price endogeneity problem 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 approach, making market size enter the moment restrictions in a nonlinear way. Their method largely relies on the linear additivity and thus can not extend directly to the BLP framework[7]. Further,

[7]Petrin and Train (2010)'s control function approach is an alternative to the BLP approach in dealing

the main focus of their paper is to remove the bias, while we also identify and estimate the market size.

Another strand of literature looks at issues arising in estimating market shares from a sample of consumers. Gandhi, Lu, and Shi (2020) are particularly concerned with zeros in market share data. They propose a moment-inequality-based approach to estimate demand in the presence of zero sales. Berry, Linton, and Pakes (2004) derive the asymptotic theory for the BLP model taking into account sampling errors in estimating shares. In comparison, the problem our paper aims to solve is rooted in the model rather than features of the data sample. The goal of this paper is to address the more fundamental problem that the share of the outside option is unobserved and that all shares will be inconsistent in the limit. The errors do not vanish like sampling errors do as the sample size grows larger.

Our identification strategy is deeply connected with the literature on point identification in nonlinear models defined by a collection of equations. Identification of parametric BLP models can be attained by making exclusion restrictions. The corresponding conditional moments are essentially a system of nonlinear equalities. The required rank conditions for identifying models of this type are discussed in Fisher (1966), Rothenberg (1971) and Hsiao (1983). The study of estimation in the general nonlinear simultaneous equations models goes back to Amemiya (1974), Jorgenson and Laffont (1974) and Amemiya (1977).

More recently, literature on the identification and estimation of random coefficients aggregate demand model has been growing. As pointed out by Berry and Haile (2014) and Gandhi and Houde (2019), identification of BLP demand models requires instruments for not only endogenous prices but also endogenous market shares. Other results on the role of instruments in BLP models are Reynaert and Verboven (2014), Armstrong (2016) and Conlon and Gortmaker (2020). We contribute to this literature by providing low level conditions on instruments required for identification of random coefficients in the standard BLP model, both with and without identifying market size as well.

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. Some identification theorems in these papers are derived from the classical nonparametric instrumental variable literature (e.g., Newey and Powell 2003; Ai and Chen 2003; Chen and Pouzo 2009; and Andrews 2017). Our paper also provides conditions for identification of market size in nonparametric specified demand models.

Our empirical application looks at price effects of mergers. Consideration of market

with the price endogeneity issue; which method to use will be application-specific. This discussion is outside the scope of the present paper.

definition has a long history in the merger evaluation literature. The use of traditional merger screening tools, such as Herfindahl-Hirschman Index, will turn on the definition of markets. See, Capps, Dranove, and Satterthwaite (2003), Garmon (2017), and Ellickson, Grieco, and Khvastunov (2020), among others, for examples and discussions. We note that this is a different issue than what we study. The main difference is that we include an outside option in the measure of market size, while market definitions in the traditional merger analysis exclude the outside option and only consider products in the competitive set. The problem we look at can arise when using structural models to predict lost competition and counterfactual price effects, that is, in a merger simulation. Merger simulation as a screening tool was newly added to the 2010 Horizontal Merger Guideline, and several papers have assessed its effectiveness and robustness (see Weinberg and Hosken 2013; Grigolon and Verboven 2014; Friberg and Romahn 2015; Bjornerstedt and Verboven 2016; and Houde 2012). In line with this literature but focusing on a different aspect, we evaluate to what extent merger simulation is sensitive to market size assumptions and the implications for antitrust policies.

The remainder of the paper is organized as follows. In section 2, we start with a multinomial logit model to illustrate the problem of mismeasured market size and provide identification strategies. In section 3, the results are generalized to the random coefficients logit model. Section 4 provides extensions. Section 5 demonstrates that our method works well in Monte Carlo simulations, while section 6 presents an empirical application.

2 The Multinomial Logit Demand Model

In this section, we first briefly review the setup of a plain multinomial logit model. We start with a model without random coefficients or individual-level covariates for illustration purposes. Prices are taken to be exogenous throughout the context of the plain logit model to ease the exposition, and to focus more on identification of market size. This assumption is relaxed in later sections. We then propose a simple model of market size. The model contains assumptions on the unobserved outside shares. We then combine both models and provide, in theorem 1, assumptions under which demand parameters and market size can be identified. Finally, we extend theorem 1 to a nested logit model, and we use it to demonstrate the discrepancies and connections between a nest structure and the market size model.

2.1 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.

Let $j = 0$ denote the outside option. As in Berry (1994), we 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. In this case, unobserved preference heterogeneity enters only through ε_{ijt}.

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 s_{ijt} be a binary variable that equals to 1 if consumer i chooses product j in market t, and 0 otherwise. Let $\pi_{jt} = Pr(s_{ijt} = 1 \mid X_{jt}, \xi_{jt})$ denote the true conditional choice probability of 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

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

If probabilities π_{jt} and π_{0t} were observed, parameters β can be consistently estimated by regressing $\ln(\pi_{jt}/\pi_{0t})$ on X_{jt}. GMM estimators would generally be constructed based on the mean independence condition $E(\xi_{jt} \mid X_{jt}) = 0$. The conditions we 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.

2.2 Market Size Model

In this subsection we 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. We then combine these assumptions with the demand system to obtain a new model which we 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 that

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

Let N_{jt} be the observed sales of good j in market t, and $N_t^{total} = \sum_{j=1}^{J_t} N_{jt}$ be the total observed sales of all goods. Suppose that the true r_{jt}^* is unobservable, so we only observe r_{jt}, where $r_{jt} = N_{jt}/N_t^{total}$ is the fraction of all purchases that are spent on good j in market t, and therefore does not depend on the outside option or the size of the total market. We call these r_{jt} *relative shares*, and we assume

$$\ln\left(r_{jt}\right) = \ln\left(r_{jt}^*\right) + e_{jt}. \tag{4}$$

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.

Think of r_{jt} for all $j \in \mathcal{J}_t$ being observable along with N_t^{total}. In most empirical applications, we might directly observe N_{jt} for all $j \in \mathcal{J}_t$ (not including the outside option). 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, then the market share of the outside option π_{0t} can be calculated from the observed N_t^{total} and the market size, and vice versa. However, observing only the relative shares r_{jt} for all J_t goods is not enough to obtain π_{0t}, and we therefore need to specify a model for it. Compared to equation (1), the model of equation (3) has the advantage that only the first term of the right side of this equation depends on the outside share, and consequently it is easier and more natural to impose assumptions on the unobserved outside option.

Let M_t be some observed population or quantity measure of market t that we believe is related to the true total market size. If a market is defined to be a city, M_t could be the size of the population of the city (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 predicted sales of all products or predicted number of passengers on 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*. If W_t were correctly measured for the population (i.e., M_t is the true market size), rather than being a possible mismeasurement of a sample, then $1/W_t$ would equal $1 - \pi_{0t}$, which is the fraction of the market that makes any purchase. As noted earlier, it is both natural and necessary to place assumptions on π_{0t}. For now we assume that the mismeasurement in W_t relative to π_{0t} takes the form

$$\ln\left(\frac{\pi_{0t}}{1-\pi_{0t}}\right) = \ln\left(\gamma W_t - 1\right) + v_t \tag{5}$$

for some constant γ and some random mean zero noise v_t, although later we will generalize it to a more flexible model[8].

To see why the model of equation (5) makes sense, consider the usual approach seen in the literature where market size is assumed to be some known constant γ multiplies an observed population measure M_t. Then $1 - \pi_{0t}$ would approximately equal $1/\left(\gamma W_t\right)$, and therefore $\ln\left(\pi_{0t}/(1-\pi_{0t})\right)$ would approximately equal $\ln\left(\gamma W_t - 1\right)$. Equation (5) treats the usual constant γ as unknown rather than known, and adds the error term v_t to account for this relationship being approximate rather than exact. Furthermore, equation (5) can be derived from a deeper economic model that provides more intuitive explanation. We defer discussion of this micro foundation to section 4.1.

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

$$\ln\left(r_{jt}\right) = \ln\left(\gamma W_t - 1\right) + X_{jt}'\beta + u_{jt} \quad \forall j \in \mathcal{J}_t \tag{6}$$

where

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

Before proving identification, we first 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 (6) 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

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

[8] An alternative way to relax this modeling assumption, although not explored in the present paper, is to allow γ to be a function of observed market-level covariates that affect preferences, and we leave it for future work.

with

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

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 (6). The third equality follows from equation (5). Although we can assume that (e_{jt}, v_t) are independent of X_{jt}, 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.

2.3 Identification

Here we provide identification of model (6). Unknown parameters in this model include the market size parameter γ and demand coefficients β. We provide an approach that allows identifying the true market size along with demand parameters. Identification relies on variation across markets, and thus would need to observe data from many markets. Our main analysis in later section will be a generalization of this approach. In Appendix B, we show an alternative approach that uses market fixed effects; this more naive approach identifies demand parameters but not the market size.

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

Assumption 1 assumes that the additive error u_{jt} is mean independent of product characteristics and some instrument Q_t, and the regressors are continuously distributed. Note that the nonlinear variable W_t in equation (6) is endogenous because it is a function of quantities. The instrument Q_t can be a vector or a scalar. For convenience, theorem 1 uses a scalar Q_t. We need the large T assumption because the theorem is based on the conditional expectation conditioning on Q_t, and the derivatives of the conditional expectation. These derivatives would be estimated by a nonparametric regression such as kernel regression or local polynomials (Li and Racine 2007). Assuming Q_t is continuous, we asymptotically need to observe all values of Q_t on its support, and so need T to go to infinity. Additional it implies that Q_t can not be a binary variable.

Theorem 1. *Given Assumption 1 and equation (6), 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. *(relevance condition) 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 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 entails, 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, failing to satisfy the condition. Generally, the second condition in theorem 1 is a nonlinear analog of the traditional relevance restriction needed in classical linear IV model, requiring W_t varies with Q_t in a certain way.

Identification requires an instrument Q_t that varies with the market total sales, and uncorrelated with the error u_{jt}. A simple candidate satisfying these conditions is the sum of exogenous characteristics over all products in market t. Product characteristics affect utilities consumers get and therefore drive variation in quantities across regions or time periods, so the relevance condition is in general satisfied. The exogeneity condition is satisfied because the error u_{jt} is mean independent of not only characteristics of product j, but also all other products of market t, so is mean independent of sum of all products. This is similar to the standard BLP instrument and we defer discussion of this "functions of inside regressors" type of instruments to next section.

If we have any exogenous changes in characteristics across markets, we observe how many consumers enter or exit the outside option to help identify γ. This exogenous variation is usually already embodied in the data. For example, consider a market with only two goods, Coke and Pepsi. When characteristics of Pepsi get worse, total quantities decrease and so W_t rises. Identification of γ would follow from the relative increase in Coke shares. If γ was large, a lot of Pepsi users might divert to the outside option, and therefore we would observe little diversion to Coke. If γ was small, then it means not many consumers are on the margin, and so when Pepsi becomes worse, more Pepsi users divert to coke rather than outside option. Figure 1 in next section illustrates this intuition.

Another special case that might also satisfy the required assumptions is when there is an exogenous change in, for instance, alcohol tax or soda tax. We would observe how much quantity falls in the market and that is the fraction of consumers who switch to the outside option. Generally, from the nature of a logit model, substitutions are proportional to the true market shares, so the true market size affects outside diversion. This suggests that if we observe any changes in the outside diversion (due to variation in Q_t), we can back out the true market size. Thus, we exploit variation in W_t and Q_t to aid in the identification of market size.

Estimation of the model of equation (6) 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.

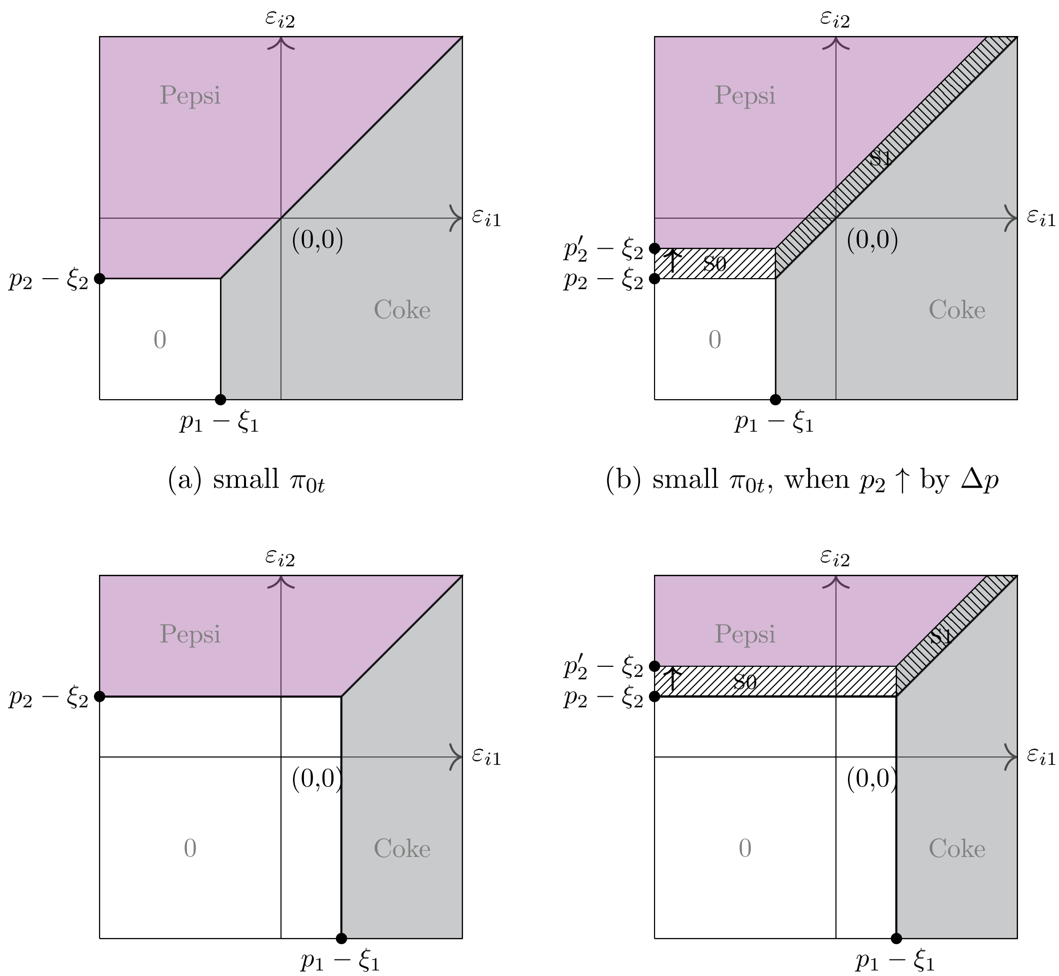
2.3.1 Visual Intuition

After our formal identification results, we give some visual intuition. Consider a simplified model where $\delta_{jt} = -p_{jt} + \xi_{jt}$. With two goods, the space of ϵ_{ij} is partitioned into three regions (Berry and Haile 2014 and Thompson 1989), with each region corresponding to choosing one of $j = 0, 1, 2$. Measure of consumers in each region, i.e. integral of ε over the region, reflects choice probabilities. For example, $Pr(j = 1 \mid p, \xi) = Pr(\varepsilon_{i1} > p_1 - \xi_1; \varepsilon_{i1} > \varepsilon_{i2} + (p_1 - \xi_1) - (p_2 - \xi_2))$. For a fixed known density function of ε_{ij}, for example, panel (c) of Figure 1 shows a larger probability of choosing outside option than panel (a). 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 dgp is that of 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 big true π_{0t}. Following a price increase of good 2, changes in choice probabilities π_{0t} and π_{1t} are captured by shaded boundaries S_0 and S_1. In panel (b), an increase in the price of good 2 induces more consumers switch to good 1, while in panel (d), the same price change makes more consumers switch to outside option. The relative diversion to outside option versus good 1, which is known, depends on how big each region originally was, which is unknown, and this relation provides identification.

By the definition of a logit model, we have $\partial \pi_{1t} / \partial p_{2t} = -\pi_{1t}$ and $\partial \pi_{0t} / \partial p_{2t} = -\pi_{0t}$.

Thus, $(\partial \pi_{1t}/\partial p_{2t})/(\partial \pi_{0t}/\partial p_{2t}) = \pi_{1t}/\pi_{0t}$. Since we are looking at ratios, the unobserved choice probabilities can be converted to 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 as we know the total sales decrease (increase) is the increase (decrease) in N_{0t}. 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}.



(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

With the introduction of random coefficients in section 3, the example showed in Figure 1 might not work because when we observe more substitution to good 1, it could be the result of good 1 and good 2 being closer substitutes. However, we could gain some intuition from this example. Even without the IIA property, cross-product substitutions are still functions of true choice probabilities, and thus the substitution to 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.

2.4 The Nested Logit Demand 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 $\ln(\pi_{j|gt})$,

$$\ln(\pi_{jt}/\pi_{0t}) = X'_{jt}\beta + \rho \ln(\pi_{j|gt}) + \xi_{jt}, \tag{7}$$

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 2.2. Then we can rewrite (7) 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 2.2, we assume equations (4) and (5) hold. Combining above equations and assumptions we get the estimating equation for the nested logit model

$$\ln\left(r_{jt}\right) = \frac{1}{1-\rho}\ln\left(\gamma W_t - 1\right) + X'_{jt}\frac{\beta}{1-\rho} + u_{jt}, \tag{8}$$

where

$$u_{jt} = \frac{\xi_{jt}}{1-\rho} + \frac{v_t}{1-\rho} + e_{jt}.$$

Theorem 2. *Given Assumption 1 and equation (8), let Γ be the set of all possible values of γ, if*

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

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

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

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

Then γ*,* β *and* ρ *are identified.*

The proof of theorem 2 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.

We 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 (8) 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. We 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 also 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 with wrong market sizes is provided in the Appendix D.

One might be concerned that the identification result of theorem 2 relies on the functional form assumption we made in equation (5). There might exist some different functional form assumption of market size which would make γ and ρ not identified. 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 (8) reduces to $\ln\left(r_{jt}\right) = 1/(1-\rho)\gamma\tilde{W}_t + X_{jt}'\beta/(1-\rho) + u_{jt}$. However, a market size model of this form is odd and lack of economic meaning.

3 The Random Coefficients Logit Demand Model

In this section, we generalize our previous results to the more flexible random coefficients demand model. We first introduce the notation and model assumptions, and then provide sufficient conditions under which the model is identified and provide guidance on types of instruments that suffice. We then discuss testing of instrument relevance and intuition for separately identifying parameters. Later, we derive some results for market fixed effects and show that fixed effects would not be a valid fix for unknown market size.

3.1 Demand Model and Market Size

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

$$U_{ijt} \quad = \quad X'_{jt}\beta_i + \xi_{jt} + \varepsilon_{ijt}, \tag{9}$$

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. By including the interaction terms between consumer characteristics ν_{il} and product characteristics X_{jt}, individuals are allowed to have heterogeneous tastes for each different observed product characteristics. For example, passengers with different income levels might generate different disutilities from the flight delay. The consumer characteristics could be either observed individual characteristics or unobserved characteristics. When estimating demand models, what econometricians usually have are aggregate product-level data, where no observed individual characteristics are available. Therefore, in our analysis, we assume ν_{il} are some unobserved characteristics with a known distribution (e.g., the distribution of individuals' income can be nonparametrically estimated from external data sources). The extension to including observed consumer characteristics will be straightforward if there are individual-level data (see micro-BLP).

Combining equations we have

$$\begin{aligned} U_{ijt} \quad &= \quad X'_{jt}\beta + \xi_{jt} + \sum_l \sigma_l x^{(2)}_{jtl} \nu_{il} + \varepsilon_{ijt} \\ &= \quad \delta_{jt} + \sum_l \sigma_l x^{(2)}_{jtl} \nu_{il} + \varepsilon_{ijt}, \end{aligned}$$

where $X_{jt} = (x_{jt1}, \cdots, x_{jtL})$ is a $L \times 1$ vector with the l-th characteristic being x_{jtl}, and $X^{(2)}_{jt} = (x^{(2)}_{jt1}, \cdots, x^{(2)}_{jtL})$ is a subvector of X_{jt} that is the nonlinear components of the indirect utility function.

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

$$\pi_{jt}\left(\delta_t, X^{(2)}_t; \sigma\right) = \int \frac{exp(\delta_{jt} + \sum_l \sigma_l x^{(2)}_{jtl} \nu_{il})}{1 + \sum_{k=1}^{J_t} exp(\delta_{kt} + \sum_l \sigma_l x^{(2)}_{ktl} \nu_{il})} f_\nu(\nu) d\nu, \tag{10}$$

where the arguments in the choice probability function are mean utilities $\delta_t = (\delta_{1t}, \cdots, \delta_{J_t t})$, nonlinear attributes $X^{(2)}_t = (X^{(2)}_{1t}, \cdots, X^{(2)}_{J_t t})$ and taste parameters $\sigma = (\sigma_1, \cdots, \sigma_L)$. The choice probability is written as a function of δ_t and σ in order to highlight its dependence on the mean utilities and parameters of the model. We suppress the dependence of the choice probability function on ν_i for brevity. The mean utility of outside good is normalized to $\delta_{0t} = 0$.

We 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 $\gamma = (\gamma_1, \gamma_2)$ be a vector of market size parameters, where $\gamma_1 = (\gamma_{11}, \cdots, \gamma_{K1})$ and $\gamma_2 = (\gamma_{12}, \cdots, \gamma_{K2})$. To ease the exposition, we assume here and now that $e_{jt} = \upsilon_t = 0$, so $r^*_{jt} = r_{jt} = N_{jt}/N^{total}_t$. Observational errors in r_{jt} and other disturbances in the mismeasurement are therefore assumed away. Assumption 2 formalizes the modeling assumption.

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

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

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

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

This particular formula of true market size has several appealing features. For example, in the airline market, 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 the two end-point cities could be $M^2_{1t} + 3M^2_{2t}$. First, it allows for different coefficients for the two terms. In this case, it could be that city B has a larger coefficient because it is a major hub. Second, it allows for nonlinearity in M_t. In the airline example, the larger the metro area is, the more likely it may be for outside substitutes to be available, as it has high speed rail or highways in more directions. Under this assumption, we have

$$\begin{aligned}
\pi_{jt}\left(\delta_t, X^{(2)}_t; \sigma\right) &= \frac{\pi_{jt}\left(\delta_t, X^{(2)}_t; \sigma\right)}{\sum_{k=1}^{J_t} \pi_{kt}\left(\delta_t, X^{(2)}_t; \sigma\right)} \sum_{k=1}^{J_t} \pi_{kt}\left(\delta_t, X^{(2)}_t; \sigma\right) \\
&= \frac{N_{jt}}{N^{total}_t} \frac{N^{total}_t}{\sum_{k=1}^K \gamma_{k1} M_{kt}^{\gamma_{k2}}} = \frac{N_{jt}}{\sum_{k=1}^K \gamma_{k1} M_{kt}^{\gamma_{k2}}},
\end{aligned}$$

The implicit system of demand equations in a given market t is given by

$$\frac{N_t}{\sum_{k=1}^K \gamma_{k1} M_{kt}^{\gamma_{k2}}} = \pi_t\left(\delta_t, X^{(2)}_t; \sigma\right), \tag{11}$$

where $N_t = (N_{1t}, \cdots, N_{J_t t})$ and $\pi_t(\cdot) = (\pi_{1t}(\cdot), \cdots, \pi_{J_t t}(\cdot))$ are vectors of observed *quantities* and *choice probability functions*. When $\gamma_2 = 0$ and γ_1 is a scalar, γ becomes a scalar factor of proportionality and the demand system reduces to the simple case $N_t/\gamma M_t = \pi_t(\delta_t, X^{(2)}_t; \sigma)$.

3.2 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 essential. 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 "inverting" the system of demand equations. Our method builds on the same form of demand inversion, while replacing observed market shares with the ratio of observed quantities to unobserved true market size.

Formally, the identification argument consists of two parts: First, we will show that for any given parameters (γ, σ) and data (N_t, M_t, X_{jt}), the implicit system of equations (11) has a unique solution δ_t for each market[9]. We show in Proposition 1 that the existence and uniqueness of demand inversion shown in Berry (1994) and Berry, Levinsohn, and Pakes (1995) apply in our setting (see also Berry and Haile (2014) for demand inversion in nonparametric models). Second, given a unique sequence of *inverse demand function* $\delta_{jt}(N_t, M_t, X_t^{(2)}; \gamma, \sigma)$, we can construct a unique sequence of *residual function* $\xi_{jt}(N_t, M_t, X_t; \gamma, \sigma, \beta)$, which we will define later. The identification is then based on conditional moment restrictions, and we will need the associated unconditional moment conditions to have a unique solution at θ_0.

Proposition 1. *Let equations (10) and (11) 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 implicit system of equations is solved for each market, therefore we drop the t subscript in the proof to simplify the notations. The proof of Proposition 1 follows closely to the proof of the contraction mapping argument in Berry, Levinsohn, and Pakes (1995). We show that all the conditions in the contraction mapping theorem are satisfied in our setting with an extra vector of γ, and therefore the function $g(\delta)$ is a contraction mapping.

In Proposition 1, we have shown 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 $\delta_{jt}(N_t, M_t, X_t^{(2)}; \gamma, \sigma)$, we define the residual function as

$$\xi_{jt}\left(N_t, M_t, X_t; \theta\right) = \delta_{jt}\left(N_t, M_t, X_t^{(2)}; \gamma, \sigma\right) - X_{jt}'\beta. \tag{12}$$

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

[9]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 it serves as a technical device, see Berry, Gandhi, and Haile (2013) for a discussion.

(2019), we will assume that the unobserved structural error term is mean independent of a set of exogenous instruments Z_t, based off which we will later construct unconditional moment conditions. Specifically, we replace the exogenous restriction in section 2 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 vector of instruments Z_t typically includes a subvector of X_t that contains all exogenous characteristics and the excluded price instruments such as cost shifters. The assumption states that the structural error is mean independent of not only the exogenous covariates of product j, but also all the other products. As with standard BLP models, we need two classes of instruments: (i) price instruments; (ii) instruments that identify nonlinear parameters (σ and γ). We 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,*

$$E\left(h_{jt}(\tilde{\theta})\right) = 0 \iff \tilde{\theta} = \theta_0, \text{ for all } \tilde{\theta} \in \Theta \tag{13}$$

θ_0 is locally identified if (13) holds only for $\tilde{\theta}$ in an open neighborhood of θ_0.

Assumption 4 in Berry and Haile (2014) and equation (5) in Gandhi and Houde (2019) both impose a high-level identification assumption similar to (13). 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 we 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 proving identification in nonlinear simultaneous equations model relying on full rank conditions dates back to Fisher (1966) and Rothenberg (1971). See Hsiao (1983) for a review. The use of full rank conditions to obtain local identification is seen in a variety of studies (McConnell and Phipps 1987; Iskrev 2010; Qu and Tkachenko 2012; Milunovich and Yang 2013; and Gospodinov and Ng 2015). Using Proposition 2, we can now establish identification theorem for random coefficients demand model with unobserved market size.

Theorem 3. *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 needs a rank condition similar to this one, 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 does not have a closed-form expression and therefore verifying full column rank can be challenging. Below we 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).

3.2.1 Sufficient Conditions for Identification

We here replace the high-level rank condition with some low-level conditions on instruments. Our 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 (11) with respect to parameters can be shown to have independent variation. In order to check the rank of the Jacobian matrix, we calculate the derivatives of $h_{jt}(\theta)$. The Jacobian matrix consists of four groups of derivatives: derivatives with respect to γ_1, γ_2, σ and β, respectively. By the implicit function theorem for a system of equations (Sydsæter et al. 2008) and applying the Cramer's rule, the first two groups of derivatives can be computed explicitly as

$$\begin{aligned}
J_1 &= \frac{\partial h_{jt}(\theta)}{\partial\gamma_{k1}} = \frac{\partial\delta_{jt}\left(N_t,M_t,X_t^{(2)};\gamma,\sigma\right)\phi_j(Z_t)}{\partial\gamma_{k1}} \\
&= \begin{vmatrix} \frac{\partial\pi_{1t}}{\partial\delta_{1t}} & \cdots & \frac{\partial\pi_{1t}}{\partial\delta_{Jt}} \\ \vdots & \ddots & \vdots \\ \frac{\partial\pi_{Jt}}{\partial\delta_{1t}} & \cdots & \frac{\partial\pi_{Jt}}{\partial\delta_{Jt}} \end{vmatrix}^{-1} \underbrace{\begin{vmatrix} \frac{\partial\pi_{1t}}{\partial\delta_{1t}} & \cdots & \pi_{1t} & \cdots & \frac{\partial\pi_{1t}}{\partial\delta_{Jt}} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \frac{\partial\pi_{Jt}}{\partial\delta_{1t}} & \cdots & \pi_{Jt} & \cdots & \frac{\partial\pi_{Jt}}{\partial\delta_{Jt}} \end{vmatrix}}_{(\pi_{1t},\cdots,\pi_{Jt})' \text{ is in the } j\text{-th column}} \frac{M_{kt}^{\gamma_{k2}}}{\sum_k\gamma_{k1}M_{kt}^{\gamma_{k2}}}\phi_j(Z_t), \\
&= \Psi_{jt}\left(\delta_t,X_t^{(2)};\sigma\right)\frac{M_{kt}^{\gamma_{k2}}}{\sum_k\gamma_{k1}M_{kt}^{\gamma_{k2}}}\phi_j(Z_t),
\end{aligned} \tag{14}$$

and

$$J_2 = \frac{\partial h_{jt}(\theta)}{\partial \gamma_{k2}} = \frac{\partial \delta_{jt}\left(N_t, M_t, X_t^{(2)}; \gamma, \sigma\right) \phi_j(Z_t)}{\partial \gamma_{k2}}$$
$$= \Psi_{jt}\left(\delta_t, X_t^{(2)}; \sigma\right) \frac{\gamma_{k1} \ln(M_{kt}) M_{kt}^{\gamma_{k2}}}{\sum_k \gamma_{k1} M_{kt}^{\gamma_{k2}}} \phi_j(Z_t) \tag{15}$$

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 (14) (and equation (15)). We 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 utility levels and characteristics of all products[10]. The Jacobian determinant of $(\pi_{1t}, \cdots, \pi_{Jt})'$ with respect to $(\delta_{1t}, \cdots, \delta_{Jt})$ is different from zero by the model definition. 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 the parameters. In such a case, the associated columns of the Jacobian matrix will be linearly dependent. For example: $E(\partial h_{jt}/\partial \gamma_{k1}) = cE(\partial h_{jt}/\partial \gamma_{k2})$ or $E(\partial h_{jt}/\partial \gamma_{k1}) = cE(\partial h_{jt}/\partial \sigma_l)$ for some non-zero constant c.

Note, that if M_t were independent of $\phi_j(Z_t)$ and of everything in the demand model, we have

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

and

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

It would be impossible to separately identify γ_{k1} and γ_{k2}, neither could we distinguish γ_{k1} and γ_{j1} for $j \neq k$. In order to disentangle the γ vector, we require some elements of instruments to satisfy

$$Cov\left(\frac{M_{kt}^{\gamma_{k2}}}{\sum_k \gamma_{k1} M_{kt}^{\gamma_{k2}}}, \phi_j(Z_t)\right) \neq 0, \quad Cov\left(\frac{\gamma_{k1} \ln(M_{kt}) M_{kt}^{\gamma_{k2}}}{\sum_k \gamma_{k1} M_{kt}^{\gamma_{k2}}}, \phi_j(Z_t)\right) \neq 0, \tag{16}$$

which suggests that we could use any outside variables that change M_t exogenously as instruments, for example, expansions of highway in a city.

---

[10]

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

where

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

The third group of derivatives is

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

where we 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 collinear[11], which implies 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 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*

$$
Cov\left(\Psi_{jt}\left(\delta_t, X_t^{(2)}; \sigma\right), \phi_j^{(2)}(Z_t)\right) \neq 0, \tag{17}
$$

and

$$
Cov\left(\Phi_{jt}\left(\delta_t, X_t^{(2)}; \sigma\right), \phi_j^{(3)}(Z_t)\right) \neq 0, \tag{18}
$$

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

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

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

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

Collectively, all that is required for identification of market size parameters are two sets of instruments: (1) shifters of market size measures M_t; (2) variables that provide exogenous variations in quantities (or shares). 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, we would require assumptions similar to the ones provided in Lemma 1, but only need instruments that satisfy condition (18). When γ is a vector of dimension ≥ 2, we need an additional source of variation to identify elements of the γ vector (i.e., variation in measures of market size).

Valid potential instruments satisfying (17) and (18) are functions of exogenous product characteristics. Therefore our proposed method can be used without requiring a new class of outside instruments over and above those commonly used in BLP models and without any additional independent variations in data. Commonly used instruments of this type are (i) sums of product characteristic of other products produced by the same firm, and the sums of product characteristics offered by rival firms, which are often referred to as BLP instruments (Berry, Levinsohn, and Pakes 1995) and (ii) sums of differences of products in characteristics space, which are referred to as differentiation instruments (Gandhi and Houde 2019). However, the validity of Gandhi and Houde (2019)'s differentiation instruments depends on the symmetry property of demand function that has not been shown in our model. By introducing γ, the symmetry property that was used to derive these instruments may not hold, as we can no longer treat the outside option same as inside goods. Therefore, in the empirical section, we use BLP type instruments to obtain main results and use differentiation instruments as robust check. Another valid set of instruments is Chamberlain's (1987) optimal instrument that has been implemented in BLP by Reynaert and Verboven (2014). The optimal instrument is the expected value of the Jacobian of inverse demand function, which in our context is to use $E(\Psi_{jt}(\delta_t, X_t^{(2)}; \sigma) \mid Z_t)$ and $E(\Phi_{jt}(\delta_t, X_t^{(2)}; \sigma) \mid Z_t)$ as instruments.

3.3 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. In this section we redefine our parameters and show that the same test of instrument relevance can be applied here, 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 we 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. The legitimacy of treating λ_γ and λ_σ alike is shown below. We first recognize that for any given (N_t, M_t, X_t) and model parameters, the residual function in equation (12) can be rewritten as

$$\xi_{jt}\left(\frac{N_t}{\sum_k(\lambda_{\gamma_{k1}}+1)M_t^{\lambda_{\gamma_{k2}}}}, X_t; \lambda_\sigma, \beta\right) = \delta_{jt}\left(\frac{N_t}{\sum_k(\lambda_{\gamma_{k1}}+1)M_t^{\lambda_{\gamma_{k2}}}}, X_t^{(2)}; \lambda_\sigma\right) - X_{jt}'\beta. \quad (19)$$

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. By absorbing λ_γ into the conditioning parameter vector, we rewrite equation (19) as

$$\xi_{jt}\left(N_t, M_t, X_t; \theta\right) = \delta_{jt}\left(N_t, M_t, X_t^{(2)}; \lambda\right) - X_{jt}'\beta. \quad (20)$$

Equation (20) encompasses equation (19) and is similar to equation (4) in Gandhi and Houde (2019). Here we 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 a linear regression of endogenous variables on excluded instruments, for nonlinear models, how to detect weak instruments is not obvious.

We use the method as in Gandhi and Houde (2019) to test the relevance of instruments for identifying λ_σ and λ_γ, which we 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[12]:

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

[12] This approximation 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 $X_t^{(2)}$.

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

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

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 on the basis of 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. We 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.

To gain more intuition, we can relate equation (7) of Gandhi and Houde (2019) to our relevance conditions (17) and (18). The coefficient of the instrument function (γ_2 in their notation) measures the correlation between the instrument $\phi_j(Z_t)$ (or $A_j(\mathbf{x}_t, \mathbf{w}_t)$ in their notation) and the derivatives of the inverse demand function, Φ_{jt} and Ψ_{jt} (or the quality gap $\Delta_{jt}(\lambda)$ in their notation).

3.4 Intuition for Identification

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

In section 2, we 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 (6) 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 (11) as

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

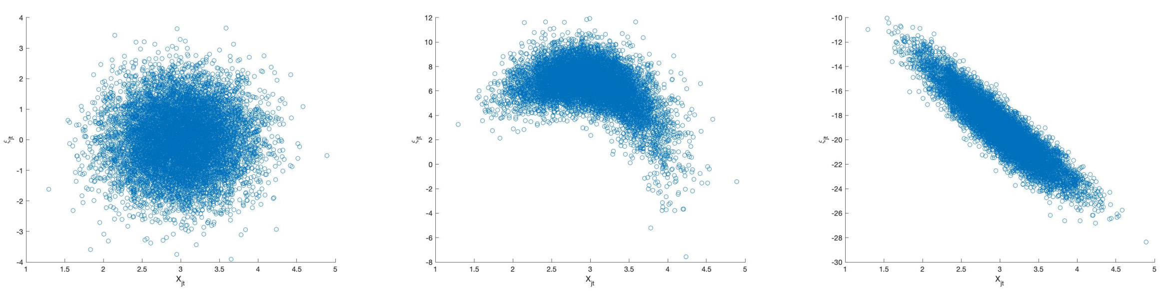
which is again linear in γ, but nonlinear in β and σ. We can exploit the same nonlinearity as in the simple logit case to distinguish γ and (β, σ).

3.4.1 A Numerical Illustration

For the numerical illustration, we 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 (11) 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 it 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, we visually illustrate the intuition for identification and why we can distinguish γ and σ.



(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) $\sigma = \sigma^0 = 5, \gamma = 4$.

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. In order for the wrong σ or γ to fit the data, ξ would have to be correlated with the instruments. Therefore once we impose the restriction 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.

3.5 Market Fixed Effects

In Appendix B we 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. We briefly discuss why the same approach cannot be taken in the random coefficients case.

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

$$E\Big[\Big(\delta_{jt}\left(N_{jt}, M_t, X_t^{(2)}; \tilde{\gamma}, \sigma_0\right) - X_{jt}'\beta_0 + \delta_{jt}\left(N_{jt}, M_t, X_t^{(2)}; \gamma_0, \sigma_0\right) \\ -\delta_{jt}\left(N_{jt}, M_t, X_t^{(2)}; \tilde{\gamma}, \sigma_0\right)\Big)\phi_j(Z_t)\Big] = 0, \quad (21)$$

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 (21), $\delta_{jt}(N_{jt}, M_t, X_t^{(2)}; \gamma_0, \sigma_0) - \delta_{jt}(N_{jt}, M_t, X_t^{(2)}; \tilde{\gamma}, \sigma_0)$, 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 (21) 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.

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

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

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

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

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

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

respectively. The denominators are identical for $j = 1, 2$. When $j = 1$, the determinant in the numerator is $\frac{1}{\gamma}\left(\int \pi_{1ti} f_\nu(\nu) d\nu\right)\left(\int \pi_{2ti}(1-\pi_{2ti}) f_\nu(\nu) d\nu\right) + \frac{1}{\gamma}\left(\int \pi_{2ti} f_\nu(\nu) d\nu\right)\left(\int \pi_{1ti}\pi_{2ti} f_\nu(\nu) d\nu\right)$. Similarly, when $j = 2$, the determinant in the numerator is $\frac{1}{\gamma}\left(\int \pi_{2ti} f_\nu(\nu) d\nu\right)\left(\int \pi_{1ti}(1-\pi_{1ti}) f_\nu(\nu) d\nu\right) + \frac{1}{\gamma}\left(\int \pi_{1ti} f_\nu(\nu) d\nu\right)\left(\int \pi_{1ti}\pi_{2ti} f_\nu(\nu) d\nu\right)$. The two are equivalent only when ν is not random and the individual choice probabilities are identical. We 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.

4 Extensions

4.1 Nonparametric Random Coefficients

We recognize that a special case of nonparametric estimation of distribution of random coefficients is equivalent to estimating market size of the naive form γM_t. More specifically, consider a model with indirect utility given by equation (9) and $\beta_i \sim F(\beta)$ follows an unknown distribution. The identification and estimation of $F(\beta)$ would be nonparametric. In the spirit 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

$$\pi_{jt}(\delta_{jt}; \sigma) = \sum_{r=1}^{R} \sigma_r \frac{\exp\left(\delta_{jt} + \sum_l \eta_l^r x_{jtl}\right)}{1 + \sum_{k=1}^{J_t} \exp(\delta_{kt} + \sum_l \eta_l^r x_{ktl})} \tag{22}$$

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 MLE estimator was originally proposed for estimation with individual choice data. Here instead we apply this approach in a BLP setting where only aggregate level data is available and prices are endogenous.

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

$$\begin{aligned}\pi_{jt}(\delta_{jt};\sigma) &= \sigma_1 \frac{\exp\left(\delta_{jt}+\eta_1\right)}{1+\sum_{k=1}^{J_t}\exp(\delta_{kt}+\eta_1)} + \sigma_2 \frac{\exp\left(\delta_{jt}+\eta_2\right)}{1+\sum_{k=1}^{J_t}\exp(\delta_{kt}+\eta_2)} \\ &= \sigma_2 \frac{\exp\left(\delta_{jt}\right)}{1+\sum_{k=1}^{J_t}\exp(\delta_{kt})},\end{aligned}$$

where the second line follows from $\eta_1 = -\infty$ and $\eta_2 = 0$. Recognize that σ_2 serves the same role as the scalar γ in section 2 when we discussed the simple logit model. The result can be extended to $R > 2$. If one element of η is $-\infty$, it implies that one type of consumers will never purchase any inside goods under any circumstances. This is the fraction of consumers that should not be considered "potential" consumers and therefore should be excluded from the measure of market size. Generally speaking, the most flexible model of this type can be approximated by a distribution with a probability mass at negative infinity. Estimating random coefficients in this way allows for the most flexible consumer tastes and also accounts for the unobserved market size (when the true market size is of the form γM_t).

Nonparametric random coefficients could fix the unknown market size problem if the distribution took the particular form above. This might be where the intuition that random coefficients can partly fix the problem is coming from. The same intuition remains relevant in the original BLP model (10) with the commonly used distributional assumption $\nu \sim N(0, 1)$, as the normal distribution has unbounded support. In practice, if the estimate of $\hat{\sigma}$ in model (10) is large, then a random draw ν_i from the standard normal distribution might have $\hat{\sigma}\nu_i \to -\infty$, which is 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. More similar to the approach in Fox and Gandhi (2016) and Fox, Kim, and Yang (2016), Wang (2022) proposes a sieve BLP estimation for aggregate demand. However, the implementation of his sieve estimation is very different from Fox, Kim, and Yang (2016) and the choice probability cannot be written in the form of equation (22). Furthermore, in order to implement sieve BLP, we need the number of instruments to be at least the number of parameters, which is of 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 a large amount of instruments to be needed.

4.2 Nonparametric Identification of Market Size

The parametric model of market size 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, we consider just the plain logit setting. We replace model (5) with

$$\ln\left(\frac{\pi_{0t}}{1-\pi_{0t}}\right) = \ln\left(\frac{s(M_t)}{N_t^{total}} - 1\right) + v_t, \tag{23}$$

where $s(\cdot)$ is an unknown function. This is approximately equivalent to assuming the true market size is given by $s(M_t)$. Under this assumption, the estimating equation becomes

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

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 a exogenous nonparametric part). Implicitly, we allow market size measures to be endogenous in the sense that $E(M_t u_{jt}) \neq 0$. Identification of β and $s(\cdot)$ can be achieved by imposing assumptions similar to those in Ai and Chen (2003). We summarize it in the following theorem.

Theorem 4. *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 $Y_{2t} = (N_t^{total}, M_t)$, $Z_{jt} = (X_{jt}, Z_{2t})$, and $dim(Z_{2t}) = dim(Y_{2t}) = K+1$. Suppose the following hold: (i) $E(u_{jt} \mid Z_{jt}) = 0$; (ii) The conditional distribution of Y_{2t} given Z_{jt} is complete; (iii) $s(\cdot) \in \Lambda_c^b(\mathbb{R}^K)$ with $b > K/2$; (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.*

Proof follows from Newey and Powell (2003) and Ai and Chen (2003), relying on the completeness of the conditional distribution[13]. 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, one can approximate it using a wide range of sieve basis.

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

4.3 Identification in Nonparametric Demand Model

The identification and estimation in sections 2 and 3 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, we generalize our results to a fully nonparametric model of BLP in the spirit of Berry and Haile (2014) to accommodate less restrictive consumer preferences. The system of demand equations is as equation (11), 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

$$\frac{N_{jt}}{s(M_t)} = \pi_j\left(\delta_t, X_t^{(2)}\right), \quad j = 1, \cdots, J. \tag{24}$$

Our result is 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 5. *Assume that M_t is continuously distributed, and is independent of (ξ_t, X_t). Assume that $s(m)$ is differentiable in m. Then $s(m) = e^{\int g(m)}\tilde{c}$ is identified up to a constant $\tilde{c}$, where $g(m) = \partial E(\ln(N_{jt}) \mid m)/\partial m$.*

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

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

from which we can identify γ_2 by construction: that is, $\gamma_2 = \partial E\left(\ln(N_{jt}) \mid M_t = m\right)/\partial \ln(m)$. When taking derivative with respect to $\ln(M_t)$, the demand function term drops out because we assume 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 in an increasing rate in M_t. Moreover, we can use these estimates to test the specification of market size model, e.g., testing if a linear model of market size holds, before estimating the whole BLP model.

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

After establishing point identification of market size, the empirical shares on the left hand side of equation (24) is identified. It would suffice to impose assumptions made in Berry and Haile (2014) to obtain nonparametric identification of the demand model.

5 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. We consider a relatively small sample size of $T = 100$. We use $R = 1000$ replications of each design.

Consumer utility is given by the random coefficients model described in Section 3

$$U_{ijt} = \beta_0 + (\beta_p + \sigma\nu_i)P_{jt} + \beta_1 X_{1,jt} + \xi_{jt} + \varepsilon_{ijt}, \tag{25}$$

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.

We compute the true choice probabilities π_{jt} in accordance with equation (10). By equations (4) and (5) and assuming $v_{jt} = e_{jt} = 0$, 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. We 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).

5.1 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. We generate data from a plain logit model ($\sigma = 0$ in the model of equation (25)). 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. We assume that the true demand coefficients are $\beta = (2, -1, 2)$.

Tables 2 to 4 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 $\tilde{\gamma}$. We 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 2, 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 3 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 2 (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 4, 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 3.

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 5, we report the estimates from our proposed method of equation (6). 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 2.

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.

5.2 Sensitivity to Market Size Assumption

The second experiment complements the first experiment. We 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 6 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 4. 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.

5.3 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 $\beta = (2, -2, 2)$. We 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 7 and 8 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 7 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 8, 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.

6 Empirical Application: A Merger Analysis

Market size has an important role especially in the context of merger analysis. The pivotal question in merger analysis is how large will the unilateral effects be. The answer to this question depends on whether an increase in the price of one product will cause the consumer to purchase an alternative product in the market; also important is whether the consumer will divert to an outside option, which is, in the soft drinks market for instance, buying juice instead. The measure of market size affects both cross-price substitutions and substitution to outside options. As we will see below, market shares (or market sizes) used in estimation not only have impacts on estimates of marginal effects ($\beta' s$) but also enter firms' first-order condition for pricing. From firms' perspective of view, the profit function is largely of two things: own market shares and price elasticities. Thus, we should expect firms' markup and consumer surplus to be sensitive to different assumptions of market size that yield different estimates of substitution patterns.

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

$$\sum_{k \in \mathcal{J}_f} \left(\frac{p_k - mc_k}{p_k} \right) \eta_{k,j} \pi_k + \pi_j = 0, \tag{26}$$

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. A merger changes the prices because products that were substitutes pre-merger are now owned by a single firm, the merged firm has less incentive to maintain a low price to steal sales from its competing products.

To demonstrate how a wrong market size can undermine the conclusion of a merger analysis, we substitute the formula of price elasticities into equation (26), 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 affect three things: the estimate of random coefficient on price β_{pi}, the estimated individual choice probability π_{ji} and π_{ki}, and the share π_j itself. Intuitively, when the market size used in estimation is larger than the true size, the outside option diversion tends to be overstated. After a merger, overestimating the outside option diversion would imply that the merged firm will keep the price relatively low in order to avoid consumers switching to outside option. For example, Weinberg and Hosken (2013) consider the breakfast syrup and motor oil industries, and estimate a plain logit model. They show that the size of simulated price changes decreases with the potential market size.

In the rest of this section, we apply our method to analyze the price effects of a hypothetical merger in the Carbonated Soft Drink (CSD) market. In Appendix F, we 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 been extensively studied in the literature, primarily due to health and regulatory concerns. The standard discrete choice model is still a popular method in modeling consumers' purchase behavior in this strand of literature. Recent studies apply structural approach by combining the random coefficients logit demand model with supply-side models to evaluate the impact of taxation and alternative policies on soft drink consumption and effects on obesity (e.g., Bonnet and Réquillart 2013b; Lopez and Fantuzzi 2012; Liu, Lopez, and Zhu 2014; Zheng, Huang, and Ross 2019; Bonnet and Requillart 2011; Zhang and Palma 2021), the effect of advertising, social media conversations and the emergence of online shopping on soft drink sales (e.g., Bonnet and Etcheverry 2021; Lopez, Liu, and Zhu 2015; Liu and Lopez 2016; Chen et al. 2019), and test behavioral theories such as compromise effect (Sharpe, Staelin, and Huber 2008). Some papers focus more on the firm conduct such as the cost pass-through (Bonnet and Réquillart 2013a and Bonnet and Villas-Boas 2016), pricing and obfuscation strategies (Richards et al. 2020), as well as how market power is linked with firm size in the soft drink industry (Mariuzzo, Walsh, and Whelan 2003).

On the quantitative industrial organization side, we are aware of many works examine the questions of interest in the context of the soft drink industry, including price discrimination (Hendel and Nevo 2013 and Marshall 2015), competition between branded products and fringe competitors (Eizenberg and Salvo 2015), and modeling demand for durable goods accounting for the storability (Wang 2015). Some papers (Dubé 2005 and Ershov et al. 2021) consider hypothetical mergers of the soft drink companies in particular.

Three factors make the soft drink market ideal for the purpose of this study. First, current literature does not reach a consensus on the market size definition. Market size is measured as either (1) the product of the population and per capita consumption of nonalcoholic beverages (draw on Beverage Digest (2013)) or the maximum amount of soft drinks can be potentially consumed (Eizenberg and Salvo 2015 assumed the constant is six liters per week), or (2) the sum of quantities of soft drinks and quantities of alternative beverages (e.g., fruit juice, milk, and water).[14]

Second, this industry is one where we generally believe the outside option is not too large. The simulation results presented in last section suggests that our method would achieve stronger identification in applications where the true choice probability of outside option is not extremely large. Although we do not observe the true outside share ex ante, we argue that goods that are frequently purchased tend to have a relatively small outside market. To see why, consider an extreme case where the prices of all soft drink products drop to zero. Consumers who never drink soda will not enter the market even if the products are free, while soda-drinkers are already buying it quite often. Therefore, we would not expect to see a huge increase in the total sales, this implies that there is not an extremely large amount of potential consumption in the market. On the other hand, an example where the outside market tend to be as large as 99% is the airline market. Suppose all airline tickets become free, then the market is likely to respond by a surge in the demand for airline flights.

The third reason why we look at soda market is that a number of horizontal mergers took place in the soft drink industry in recent years. For example, the Coca-Cola Company acquired Costa Coffee in 2018 and PepsiCo acquired SodaStream in the same year.

6.1.1 Data

We observe a panel of weekly scanner data from Nielsen, consisting of total quantities and prices for soft drink brands. The Nielsen scanner data contains detailed information on prices and sales, as well as product attributes such as package size, flavor, nutritional contents. The data covers 202 DMAs in the US and 52 weeks, ranging from January 2019 to December 2019. We aggregate the dataset from retailer level to the market level. In line with the literature, we define a market as a combination of DMA and week. This gives us $202 \times 52 = 10504$ DMA-week markets in total[15].

Additionally, we combine Nielsen data with input price data, which is used as price instruments. The constructed dataset includes raw sugar price from US Department of Agriculture, Economic Research Service; local wage from U.S. Bureau of Labor Statistics; electricity price and fuel price from US Department of Energy, Energy Information Administration.

As in Eizenberg and Salvo (2015), we aggregate flavors and products in different sized packages into $j = 1, \cdots, 15$ brand-groups (e.g., Coca-Cola Cherry 12-oz and Coca-Cola Original 16.9-oz are treated as the same brand). Following Dubé (2005), we treat diet and

[14] 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. Or when we observe in the data consumers choose alternative beverages (e.g., Bonnet and Réquillart 2013a).

[15] We drop markets with extremely large or small sales (relative to the population in those markets). After doing so, we are left with $9,658$ markets.

regular drinks as separate brands because in the industry, they are targeted at different demographic groups and therefore have been advertised and promoted separately. These brand categories include 11 brands owned by the three leading companies in the industry. The 12th and 13th brand categories are aggregates of private label brands of regular drinks and diet drinks. There are a large number of small brands (each with a volume share below 1 percent). We aggregate these niche brands into the 14th and 15th brand categories, for regular and diet drinks respectively. We justify this aggregation by the fact that product differentiation among these small brands is not of importance in the context of our study. We consider soft drinks sold in package types that have significant sales. Specifically, we include in our analysis 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. The volume sales of these five sizes dominate those of other package types.

Table 9 shows volume shares of the carbonated soft drink category for each firm averaged across DMAs. Volume shares are the volume sold of brands produced by a specific manufacture divided by the total volume sold of the entire carbonated soft drink category. Brands of the largest manufacture account for a 35.07 percent share.

6.1.2 Demand Model

As in section 3, 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} contains in-store presence, brand fixed effects, seasonal effects and region fixed effects. In-store presence is measured by the proportions of stores with a given brand in stock within a market. Brand fixed effects capture the time invariant unobserved product characteristics. Seasonal effects capture time demand shocks. P_{jt} represents the price of each brand, and ξ_{jt} denotes demand shock for a specific brand-market that is observed by consumers, but unobserved by the econometrician. The second term $\sigma\nu_i P_{jt}$ introduces consumer heterogeneity. ν_i denotes the unobserved consumer characteristics following a standard normal distribution. It shifts the price sensitivity of consumers around its mean level α with a standard deviation of σ. The last term in the utility function, ε_{ijt}, represents consumer and brand-specific shocks that follow the Extreme Value Type I distribution and are independently and identically distributed across consumers, brands, and markets[16].

[16]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,

One issue is that in-store presence could be endogenous due to correlation with the unobservables ξ_{jt}. We address this potential endogeneity problem 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. We assume retailers/firms do not observe these demand shocks when making product assortment decisions. Also, in-store presence has been used as an exogenous covariate in Eizenberg and Salvo (2015). Other papers that study the airline industry commonly include "carrier presence" as one important exogenous attribute. Carrier presence in the airline market and the in-store presence in our context have similar economic meaning. For the same reason, carrier presence could raise endogeneity concern, and it has mostly either been neglected or dealt with by controlling for fixed effects.

Table 10 provides some summary statistics of the price and in-store presence in the data. Prices and in-store presence are averaged across all UPCs in each brand, weighted by the UPC volume sales. The last three columns of Table 10 show the percentage of variance explained by brand, DMA, and month dummy variables. It clearly indicates that most of the variation in prices and in-store presence is due to differences between brands. After controlling for the variation across brands, the remaining variation is mostly contributed by differences between geographic areas.

6.1.3 Market Size Definition

We define 1 serving of soft drink to be 12 ounces. To define the market share of the outside good, Eizenberg and Salvo (2015) assumed that each household has a potential consumption for soft drinks consumed of 6 liters, which is equivalent to around 17 servings, per week. Similarly, Zheng, Huang, and Ross (2019) use documented average per capita consumption of non-alcoholic beverages as γ, which is about 30 ounces per capita per day, equivalent to 17.5 servings per week. These choices of total market size may be justified, but they still possess a degree of subjectivity. Eizenberg and Salvo (2015) have shown that their results are not qualitatively sensitive to the assumption on market size, however, it might not be the case if we considerd a different counterfactual exercise (i.e., merger simulation) where the market size assumption indeed play a larger role. We should also note that Eizenberg and Salvo (2015) use scanner data in Brazil while we use the US data, and so the assumed $\gamma = 17$ could be the correct choice with their dataset.

The market size assumptions can be written, using our notation, as γM_t, where M_t is the household population in a DMA area. For the rest of this section, all comparisons will be made with regard to assuming $\gamma = 17$ servings[17]. Specifically, we estimate γ along with

in our analysis, we will not deal with correlation in ε.

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

other demand parameters. After obtaining $\hat{\gamma}$, we calculate elasticities and diversion ratios and simulate the merger using two potential market sizes, one is 17 servings per week and the other one is $\hat{\gamma}$ servings per week.

6.1.4 Instruments

Following the literature, we seek instruments that would address the likely correlation of the demand errors ξ_{jt} with prices, and that would identify the random coefficients and market size parameters. We use three sets of instruments. The first two sets are standard excluded instruments suggested by Berry and Haile (2014) and have been used in many applications (e.g. Eizenberg and Salvo 2015; Petrin and Train 2010; and Nevo 2001).

The first class of price instruments is in the spirit of Hausman, Leonard, and Zona (1994). 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. It provides variation across brands and DMAs. This set of instruments is valid because prices across geographic regions are correlated 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). In order to lessen this concern, we control for DMA-specific, brand-level in-store presence, partially absorbing common demand shocks.

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

The third class serves as instruments that identify random coefficients and market size parameters. Here we use the traditional "sums of characteristics" BLP type instruments. Specifically, it consists of sums over exogenous characteristics of brands produced by the same company and sums over rival brands. In addition, we use the differentiation instruments in the spirit of Gandhi and Houde (2019) as a robustness check. Those are sums over functions of differences in characteristics between non-rival brands and sums over rival brands. The intuition behind Gandhi and Houde instruments is that demand for products is mostly influenced by other products that are very similar in the characteristics space. We construct this class of instruments based on in-store presence and fitted values of prices. 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 helps identify the parameters associated with heterogeneity in price sensitivity.

6.1.5 Results

In Table 11, we report five sets of estimates from the demand model. Columns 1 and 2 are estimates of the plain logit and random coefficients logit respectively, with γ estimated along with other demand parameters. Columns 3 to 5 are standard BLP estimates that assume $\gamma = 17$. In column 3, we use the same specification as in column 2. In column 4, we include an additional random coefficient on the constant term that is intended to capture unobserved preference for the outside good. In column 5, we control for DMA-week specific fixed effects. The F-statistic of an IIA-test, which we discussed in section 3.3, of detecting the strength of instruments is 2819 and the associated p-value is .00, rejecting the null hypothesis of weak instruments.

In columns 1 and 2, the estimates of price sensitivity are -8.748 and -9.86. The estimate of random coefficient parameter σ in column 2 is 1.952 and is statistically significant, suggesting that it rejects the plain logit model. The estimates of γ are 12.478 and 11.767 respectively, both are lower than the assumed $\gamma = 17$ in the literature. The third column, which assumes $\gamma = 17$, gives higher price sensitivity (-13.033) and larger standard deviation (4.395) in the preference over price. This is what one would expect to see if we assume a larger potential market size. The fourth column, which includes a second random coefficient on the constant term, gives estimates that are comparable to column 3. The estimate of σ for the constant term has a small magnitude -0.09 and is statistically insignificant. In the last column, by including market fixed effects, the estimate of price sensitivity is much lower. σ appears to be quite difficult to precisely estimate, with extremely large standard errors. This noisy result is not surprising, because with market fixed effects, we are essentially including near $10,000$ dummy variables in the GMM estimation, leaving us with limited exogenous variation to identify the random coefficient.

Table 12 provides estimated own-price elasticities and outside-good diversion ratios. Column 1 gives the elasticities that are estimated based on our estimate of $\hat{\gamma} = 12$. The own-price elasticities ranges from -3.651 to -1.887, virtually the same range as estimates in previous literature[18]. Own-price elasticities for PLs are lower than other brands. One explanation is that PL is a composite brand that consists of a large number of 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) among others, find that PLs face relative inelastic demand due to limited interbrand substitution within a store. The outside-good diversion ratios are over 60% for all brands, with PLs having the highest diversion ratio. This suggests that when facing a price increase, instead of switching to the

[18]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.

branded counterparts, consumers are more likely to stop buying, which is what one would expect to see if there exists a high degree of store loyalty.

The remaining columns in Table 12 use estimates in columns 3 to 5 of Table 11. If one assumed $\gamma = 17$ as in previous literature when the true value should be $\gamma = 12$, the biases in estimated own-price elasticities are small, whereas biases in estimated outside diversion ratios are much larger, 9 percentage points for PLs and roughly 3 to 4 percentage points for other brands, implying even less substitutions across brands. Adding a second random coefficient on the constant term leads to results close to those of column 2. The main reason is that the estimated σ for the constant term is not significantly different from zero. Adding market fixed effects gives slightly lower own-price elasticities and higher outside diversion ratios. Although results with market fixed effects are comparable to those of our estimates, the standard error of the random coefficient estimate is so large that we can not conclude any statistically significant results. Our main takeaway from Table 12 is that none of these commonly used solutions produce elasticities and diversion ratios that are close to the ones using our estimated market size.

We simulate a merger between the largest manufacture and private label manufacturers. Our merger simulation here abstracts away from cost reduction, change in the model of competition (e.g. coordination between other firms). Table 13 shows the percentage change in prices of the merging products. In column 1, these estimates (approximately 2.22% to 8.41% price increases) are reasonably comparable to Dubé (2005), who estimated the price effect after a simulated merger between two leading manufactures[19]. We see the merger simulations predict larger price increases for the PLs than products of the leading manufacture. This results from the relatively lower own-price elasticities of PLs, and is consistent with previous findings that PLs enjoy higher margins in pricing.

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

In sum, 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,

[19] The predicted price increases change with different and merging parties (e.g., combine ownership of manufacture A and private label manufacturer usually generate larger price effects than a manufactures A and B merger).

the simulated price changes display a monotonic decrease.

7 Conclusion

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. We 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, we 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 merger evaluation. Apart from the merger application, our 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 our application, we 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 our model in an industry where the true market size is known, such as the pharmaceutical market, where researchers generally have knowledge of 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. 1977. "The maximum likelihood and the nonlinear three-stage least squares estimator in the general nonlinear simultaneous equation model". *Econometrica: Journal of the Econometric Society*: 955–968.

— . 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.

Bjornerstedt, Jonas, and Frank Verboven. 2016. “Does merger simulation work? Evidence from the Swedish analgesics market”. *American Economic Journal: Applied Economics* 8 (3): 125–64.

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 Clara Etcheverry. 2021. “The Impact of online grocery shopping on retail competition and profit sharing: an empirical evidence of the french soft drink market”.

Bonnet, Céline, and Vincent Requillart. 2011. “Does the EU sugar policy reform increase added sugar consumption? An empirical evidence on the soft drink market”. *Health economics* 20 (9): 1012–1024.

Bonnet, Céline, and Vincent Réquillart. 2013a. “Impact of Cost Shocks on Consumer Prices in Vertically-Related Markets: The Case of The French Soft Drink Market”. *American Journal of Agricultural Economics* 95 (5): 1088–1108.

— . 2013b. “Tax incidence with strategic firms in the soft drink market”. *Journal of Public Economics* 106:77–88.

Bonnet, Céline, and Sofia B Villas-Boas. 2016. “An analysis of asymmetric consumer price responses and asymmetric cost pass-through in the French coffee market”. *European Review of Agricultural Economics* 43 (5): 781–804.

Bresnahan, Timothy F. 1987. “Competition and collusion in the American automobile industry: The 1955 price war”. *The Journal of Industrial Economics*: 457–482.

Capps, Cory, David Dranove, and Mark Satterthwaite. 2003. “Competition and market power in option demand markets”. *RAND Journal of Economics*: 737–763.

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.

Chen, Xuan, et al. 2019. “New concerns on caffeine consumption and the impact of potential regulations: The case of energy drinks”. *Food Policy* 87:101746.

Chu, Chenghuan Sean, Phillip Leslie, and Alan Sorensen. 2011. “Bundle-size pricing as an approximation to mixed bundling”. *The American Economic Review*: 263–303.

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.

Ellickson, Paul B, Paul LE Grieco, and Oleksii Khvastunov. 2020. "Measuring competition in spatial retail". *The RAND Journal of Economics* 51 (1): 189–232.

Ershov, Daniel, et al. 2021. "Estimating Complementarity With Large Choice Sets: An Application to Mergers".

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

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

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.

Friberg, Richard, and André Romahn. 2015. "Divestiture requirements as a tool for competition policy: A case from the Swedish beer market". *International journal of industrial organization* 42:1–18.

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.

Garmon, Christopher. 2017. "The accuracy of hospital merger screening methods". *The RAND Journal of Economics* 48 (4): 1068–1102.

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 noninvertible 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.

Grigolon, Laura, and Frank Verboven. 2014. "Nested logit or random coefficients logit? A comparison of alternative discrete choice models of product differentiation". *Review of Economics and Statistics* 96 (5): 916–935.

Hausman, Jerry, Gregory Leonard, and J Douglas Zona. 1994. "Competitive analysis with differenciated products". *Annales d'Economie et de Statistique*: 159–180.

Hendel, Igal, and Aviv Nevo. 2013. "Intertemporal price discrimination in storable goods markets". *American Economic Review* 103 (7): 2722–51.

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.

Houde, Jean-François. 2012. "Spatial differentiation and vertical mergers in retail markets for gasoline". *American Economic Review* 102 (5): 2147–82.

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*.

Mariuzzo, Franco, Patrick Paul Walsh, and Ciara Whelan. 2003. "Firm size and market power in carbonated soft drinks". *Review of Industrial Organization* 23 (3): 283–299.

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 nonparametric models". *Econometrica* 71 (5): 1565–1578.

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

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.

Richards, Timothy J, et al. 2020. "Strategic obfuscation and retail pricing". *Review of Industrial Organization* 57 (4): 859–889.

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.

Sharpe, Kathryn M, Richard Staelin, and Joel Huber. 2008. "Using extremeness aversion to fight obesity: Policy implications of context dependent demand". *Journal of Consumer Research* 35 (3): 406–422.

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.*

Wang, Emily Yucai. 2015. “The impact of soda taxes on consumer welfare: implications of storability and taste heterogeneity”. *The RAND Journal of Economics* 46 (2): 409–441.

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.

Zhang, Yunjunjie, and Marco A Palma. 2021. “Revisiting Sugar Taxes and Sugary Drink Consumption: Evidence from the Random-Coefficient Demand Model”. *Journal of Agricultural and Resource Economics* 46 (1): 37–55.

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: Sensitivity Analysis to Assumptions of Market Size in Published Articles

Article	Demand Model	Market FE?	Context	Sensitivity Analysis
Ivaldi and Verboven (2005)	Nested Logit	Yes	Horizontal merger	When the potential market size is larger, the aggregate elasticities of market demand increase.
Weinberg and Hosken (2013)	Plain Logit	No	Horizontal merger	Simulated price changes are monotonically decreasing in the potential market size.
Bokhari and Mariuzzo (2018)	BLP with RC on price	Yes	Horizontal merger	Merger simulations show considerably different impact on price changes for competitors' products.
Wollmann (2018)	BLP with RC on product characteristics	No	Horizontal merger with repositioning	The total output changes and compensating variation after a merger are substantially affected by different measures of market size.

Table 2: 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 3: 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 4: 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 5: 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.

Table 6: 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 $\tilde{\gamma} = 1, 2, 4, 8$.

Table 7: 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 8: 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.

Table 9: Manufacture-Level Volume Shares of Carbonated Soft Drink

	Regular	Diet	Total
Manufacture A	22.19%	12.88%	35.07%
Manufacture B	12.25%	6.87%	19.12%
Manufacture 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 manufacture divided by the total volume sold of the carbonated soft drink category.

Table 10: 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 11: 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: Standard errors in parentheses. Constant terms are omitted due to collinearity with product fixed effects.

Table 12: 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: To save space, only top-3 regular drink products are reported in the table.

Table 13: 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 $\gamma = 17$
Manufacture 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: To save space, only merging firms' brands are reported in the table.

Appendices

A Proofs

Proof of Theorem 6. 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. □

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

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

Taking derivative with respect to q yields

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

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

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

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

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

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

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

Proof of Theorem 2. By the mean independence condition given in Assumption **??**, 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

$$\frac{\partial E\left(\ln\left(r_{jt}\right) \mid Q_t = q, X_{jt} = x\right)}{\partial q} = \frac{1}{1-\sigma} \frac{\partial E\left(\ln\left(\gamma W_t - 1\right) \mid Q_t = q, X_{jt} = x\right)}{\partial q}. \tag{27}$$

Taking second-order derivative with respect to q yields

$$\frac{\partial^2 E\left(\ln\left(r_{jt}\right) \mid Q_t = q, X_{jt} = x\right)}{\partial q^2} = \frac{1}{1-\sigma} \frac{\partial^2 E\left(\ln\left(\gamma W_t - 1\right) \mid Q_t = q, X_{jt} = x\right)}{\partial q^2}. \tag{28}$$

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 (28) by (27) yields

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

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

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

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

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

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

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

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

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

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. 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, $\delta_j \geq \bar{\delta}$, then there is at least one element k with $\pi_k(\delta;\sigma) > s_k/\gamma$.

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

Now if there is any element of δ with $\delta_j > \bar{\delta}$, 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 $g_k(\delta) < \delta_k$:

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

$\square$

Proof of Theorem 3. 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. $\square$

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. □

Proof of Theorem 5. 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.

□

B Multinomial Logit: Market Fixed-Effects Approach

Returning to equation (6), observe that the term with the unknown π_{0t} shows up additively, and it varies by market, not by product. We 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 (6) 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. We will relax this assumption in the next section. Assumption 4 also imposes no multicollinearity requirements on X_{jt}.

Theorem 6. *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 6 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 3.4).

C Nested Logit: Elasticities and Diversion Ratios

As in the prior section, we consider a simple nested logit model where all inside goods are in one group and outside good is the only alternative in the second group. For the ease of exposition, suppose $v_t = e_t = 0$ in this section. In this case, $\pi_{jt} = N_{jt}/\gamma M_t$. Without loss of generality, we assume the true value of γ is one. Suppose we choose a possibly wrong $\tilde{\gamma}$ and thus $\tilde{\gamma} M_t$ is used in place of $1 \cdot M_t$. Let $s_{jt} = N_{jt}/\tilde{\gamma} M_t$ denote the mismeasured market shares calculated based on $\tilde{\gamma} M_t$, so we have $s_{jt} = \pi_{jt}/\tilde{\gamma}$. The choice probability of product j conditional on any inside goods being selected is r^*_{jt} and it is observed without error, so $r^*_{jt} = r_{jt}$.

The estimate of own-price elasticity of product j is given by

$$\frac{\partial s_{jt}}{\partial p_{jt}} \frac{p_{jt}}{s_{jt}} = \frac{\hat{\beta}_p}{1 - \hat{\sigma}} p_{jt} \left(1 - \hat{\sigma} r_{jt} - (1 - \hat{\sigma}) s_{jt}\right),$$

expressing it in terms of true choice probabilities we have

$$\frac{\partial s_{jt}}{\partial p_{jt}}\frac{p_{jt}}{s_{jt}} = \frac{\hat{\beta}_p}{1-\hat{\sigma}}p_{jt}\left(1-\hat{\sigma}r^*_{jt}-(1-\hat{\sigma})\frac{\pi_{jt}}{\tilde{\gamma}}\right).$$

The estimate of cross-price elasticity of product j with respect to the price of product k is given by

$$\begin{aligned}\frac{\partial s_{jt}}{\partial p_{kt}}\frac{p_{kt}}{s_{jt}} &= \hat{\beta}_p s_{kt}p_{kt}\left(1+\frac{\hat{\sigma}}{1-\hat{\sigma}}\frac{1}{1-s_{0t}}\right)\\ &= \hat{\beta}_p \pi_{kt}p_{kt}\left(\frac{1}{\tilde{\gamma}}+\frac{\hat{\sigma}}{1-\hat{\sigma}}\frac{1}{1-\pi_{0t}}\right)\end{aligned}$$

The estimate of cross-price elasticity of outside good with respect to the price of product j is given by

$$\begin{aligned}\frac{\partial s_{0t}}{\partial p_{jt}}\frac{p_{jt}}{s_{0t}} &= -\hat{\beta}_p s_{jt}p_{jt}\\ &= -\hat{\beta}_p\frac{\pi_{jt}}{\tilde{\gamma}}p_{jt}.\end{aligned}$$

The diversion ratio to outside good with respect to the price of product j is given by

$$\begin{aligned}D_{0j} = -\frac{\partial s_{0t}}{\partial p_{jt}}/\frac{\partial s_{jt}}{\partial p_{jt}} &= \frac{(1-\hat{\sigma})s_{0t}}{1-\hat{\sigma}r_{jt}-(1-\hat{\sigma})s_{jt}}\\ &= \frac{(1-\hat{\sigma})(1-\frac{1}{\tilde{\gamma}}(1-\pi_{0t}))}{1-\hat{\sigma}r^*_{jt}-(1-\hat{\sigma})\frac{1}{\tilde{\gamma}}\pi_{jt}}\\ &= \frac{(1-\hat{\sigma})(\tilde{\gamma}-(1-\pi_{0t}))}{\tilde{\gamma}-\tilde{\gamma}\hat{\sigma}r^*_{jt}-(1-\hat{\sigma})\pi_{jt}}\end{aligned}$$

The impact of the choice of $\tilde{\gamma}$ on the diversion ratio has a direct effect through choice probabilities and also an indirect effect through the estimate of $\hat{\sigma}$. The total derivative of D_{0j} with respect to $\tilde{\gamma}$ is given by

$$\frac{dD_{0j}}{d\tilde{\gamma}} = \frac{\partial D_{0j}}{\partial\hat{\sigma}}\frac{\partial\hat{\sigma}}{\partial\tilde{\gamma}}+\frac{\partial D_{0j}}{\partial\tilde{\gamma}}$$

Taking derivative of D_{0j} with respect to $\hat{\sigma}$ we have

$$\begin{aligned}\frac{\partial D_{0j}}{\partial\hat{\sigma}} &= \frac{-s_{0t}(1-\hat{\sigma}r_{jt}-(1-\hat{\sigma})s_{jt})-(1-\hat{\sigma})s_{0t}(-r_{jt}+s_{jt})}{(1-\hat{\sigma}r_{jt}-(1-\hat{\sigma})s_{jt})^2}\\ &= \frac{-s_{0t}(1-r_{jt})}{(1-\hat{\sigma}r_{jt}-(1-\hat{\sigma})s_{jt})^2}\end{aligned}$$

Observe that $1-r_{jt}>0$ when $J\geq 2$ and the denominator is strictly positive, hence $\frac{\partial D_{0j}}{\partial\hat{\sigma}}<0$.

Taking partial derivative of D_{0j} with respect to $\tilde{\gamma}$ we have

$$\begin{aligned}
\frac{\partial D_{0j}}{\partial \tilde{\gamma}} &= \frac{-(1-\hat{\sigma})^2\pi_{jt} + (1-\hat{\sigma})(1-\hat{\sigma}r^*_{jt})(1-\pi_{0t})}{\left(\tilde{\gamma} - \tilde{\gamma}\sigma r^*_{jt} - (1-\sigma)\pi_{jt}\right)^2} \\
&> \frac{-(1-\hat{\sigma})^2\pi_{jt} + (1-\hat{\sigma})^2(1-\pi_{0t})}{\left(\tilde{\gamma} - \tilde{\gamma}\sigma r^*_{jt} - (1-\sigma)\pi_{jt}\right)^2} \\
&= \frac{(1-\hat{\sigma})^2(1-\pi_{0t}-\pi_{jt})}{\left(\tilde{\gamma} - \tilde{\gamma}\sigma r^*_{jt} - (1-\sigma)\pi_{jt}\right)^2},
\end{aligned}$$

which is strictly positive when $J \geq 2$.

Since $\frac{\partial D_{0j}}{\partial \hat{\sigma}} < 0$, $\frac{\partial \hat{\sigma}}{\partial \tilde{\gamma}} > 0$ and $\frac{\partial D_{0j}}{\partial \tilde{\gamma}} > 0$, the two terms of the total derivative have opposite signs. Hence the sign of $\frac{dD_{0j}}{d\tilde{\gamma}}$ is ambiguous, depending on which of the two opposite effects is larger in magnitude.

D Nested Logit: Monte Carlo Simulations

The data generating process for the simulation datasets follows that in Armstrong (2016). Prices are endogenously generated from a demand and supply model, where firms compete a la Bertrand in the market[20]. In the base case of the Monte Carlo study, the number of products varies across markets. 1/3 of markets have 20 products per market, 1/3 of markets have 60 products and the remaining 1/3 of markets have 100 products in the market. Each firm have 10 products. Other choices of number of products per market and number of products per firm do not significantly alter the results. $X_{1,jt}$ is drawn from a uniform $(0,1)$ distribution. The excluded cost shifter $X_{S,jt}$ is also drawn from a uniform $(0,1)$ distribution, which is independent of $X_{1,jt}$. Consumer utility is given by the nested logit model described above $U_{ijt} = \beta_0 + \beta_p p_{jt} + \beta_1 X_{1,jt} + \xi_{jt} + \zeta_{igt} + (1-\sigma)\varepsilon_{ijt}$. 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 as follows:

$$\xi_{jt} = u_{1,jt} + u_{2,jt} - 1, \text{ and } \eta_{jt} = u_{1,jt} + u_{3,jt} - 1,$$

with $u_{1,jt}, u_{2,jt}$ and $u_{3,jt}$ are drawn from three independent uniform $(0,1)$ distributions. All random variables are independent across products t and markets t. The true values of cost parameters are $(\alpha_0, \alpha_1, \alpha_2) = (2, 1, 1)$, demand parameters are $(\beta_0, \beta_p, \beta_1) = (2, -1, 2)$, and the nesting parameter is set to 0.3.

[20]We have also considered a DGP where prices are exogenous and drawn from the empirical distribution of real data such that the moments of the random variable match the moments of real data. Using either DGP does not change our main conclusions.

Instead of conducting a full Monte Carlo, I carry out a single simulation using an extremely large dataset having $T = 4,000$, so the total number of observations is approximately $240,000$.

Sensitivity to Market Size Assumption

Design 1: 1/3 of markets have 20 products per market, 1/3 of markets have 60 products and the remaining 1/3 of markets have 100 products in the market. Design 2: all markets have 20 products per market. Design 3: 2/3 of markets have 20 products per market and the rest of markets have 100 products in the market. Design 4: 2/3 of markets have 100 products per market and the rest of markets have 20 products in the market.

The true choice probabilities π_{jt} are calculated from the choice probability function of nested logit model. By equations (4) and (5) and assume $v_{jt} = e_{jt} = 0$, we can calculate $N_{jt}/M_t = \gamma\pi_{jt}$, where the true value is given by $\gamma = 1$. In 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, with $\tilde{\gamma}$ taking different values ($\tilde{\gamma} = 1, 2$).

Table D.1 shows how parameter estimates and relevant statistics vary with different values of $\tilde{\gamma}$. In the first panel, the observed market shares are the same as true choice probabilities. In the second panel, we assume market size that is two times as large as the true market size. When we double the market size, the nesting parameter σ is overestimated in all four designs.

Cross-elasticity and diversion ratios in Table D.1 are average values across all products and markets, and are calculated separately for different types of markets, which differ by the number of products per market. The impact on elasticities and diversion ratios varies across designs depending on market structures. For example, in Designs 1, 3 and 4, when we assume a market size that is larger than the true size, outside-good diversion ratios are biased downward for markets with 20 products per market, whereas it is biased upward for large markets that have 100 products per market. In Design 2, when all markets have 20 products per market, the mean of elasticities and diversion ratios is not sensitive to the market size assumption, but the variance is off. The standard deviation of outside-good diversion ratio across all markets is 0.022 when $\tilde{\gamma} = 1$, while it is 0.007 when $\tilde{\gamma} = 2$, implying that the distribution of outside-good diversion ratio is now more narrowly concentrated around the mean.

Table D.1: Sensitivity to $\tilde{\gamma}$

		TRUE	Design 1	Design 2	Design 3	Design 4
$\tilde{\gamma}$ =1						
σ		0.3	0.302	0.323	0.301	0.302
β_0		2	1.993	1.921	1.999	1.979
β_p		-1	-0.997	-0.962	-0.999	-0.993
β_1		2	1.997	1.929	2.001	1.99
Mean own-elasticity			-6.913	-7.071	-7.011	-6.876
Mean cross-elasticity						
		Small market	0.246	0.248	0.246	0.245
		Medium market	0.095			
		Large market	0.059		0.059	0.059
Mean diversion ratio						
		Small market	0.036	0.036	0.036	0.036
		Medium market	0.014			
		Large market	0.009		0.009	0.009
Mean outside-good diversion						
		Small market	0.321	0.312	0.322	0.321
		Medium market	0.156			
		Large market	0.111		0.111	0.111
$\tilde{\gamma}$ =2						
σ		0.3	0.712	0.584	0.693	0.695
β_0		2	-0.178	0.087	-0.173	-0.188
β_p		-1	-0.412	-0.591	-0.439	-0.434
β_1		2	0.825	1.185	0.88	0.87
Mean own-elasticity			-6.905	-7.073	-7.005	-6.858
Mean cross-elasticity						
		Small market	0.282	0.249	0.279	0.277
		Medium market	0.093			
		Large market	0.056		0.055	0.055
Mean diversion ratio						
		Small market	0.041	0.036	0.040	0.041
		Medium market	0.014			
		Large market	0.008		0.008	0.008
Mean outside-good diversion						
		Small market	0.217	0.311	0.231	0.230
		Medium market	0.178			
		Large market	0.168		0.179	0.178

Notes: Mean own-elasticity is calculated by taking average across all products and across all markets. Mean cross-elasticity and diversion ratios are averaged across products for different types of markets. Small markets refer to markets with 20 products per market. Medium markets are markets with 60 products per market. Large markets are markets with 100 products per market.

Market Size Estimation

The instruments we use in the GMM estimation are

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

where product j is produced by firm f and $\mathcal{J}_f$ is the set of all products produced by firm f. We include BLP-type characteristics instruments as well as functions of excluded cost instruments.

Table D.2 shows the results of a single estimation with a large sample size for Design 1. The optimization algorithm we use for the GMM estimation is the gradient-based quasi-Newton algorithm (fminunc in MATLAB). Since we found multiple local minimum, we iterate over a grid of starting values in order to find the global minimum value of the objective function. Since the system now have two endogenous variables, the estimates would perform better when instruments for prices are strong enough. In an extreme case where prices are exogenous, the GMM estimation can more easily find the global minimum.

Table D.2: γ Estimation, Design 1

	TRUE	ESTIMATE	SE
β_0	2	1.854	0.568
β_p	-1	-0.937	0.005
β_1	2	1.878	0.006
σ_1	0.3	0.343	0.578
γ	1	1.021	0.315

E Identification of Random Coefficients Logit Model, Bernoulli Distribution

Suppose there is only one product in a market, so $J = 1$. 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

$$\beta_{0i} = \beta_0 + \sigma_0 \nu_i$$
$$\beta_{1i} = \beta_1 + \sigma_1 \nu_i.$$

In order 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

$$\begin{aligned}
\pi_t(\delta_t; \sigma) &= E\left[\frac{\exp(\beta_{0i} + X_t\beta_{1i} + \xi_t)}{1 + \exp(\beta_{0i} + X_t\beta_{1i} + \xi_t)} \mid X_t, \xi_t\right] \\
&= \frac{1}{2} \cdot \frac{\exp(\beta_0 + X_t\beta_1 + \xi_t)}{1 + \exp(\beta_0 + X_t\beta_1 + \xi_t)} + \frac{1}{2} \cdot \frac{\exp(\beta_0 + \sigma_0 + X_t(\beta_1 + \sigma_1) + \xi_t)}{1 + \exp(\beta_0 + \sigma_0 + X_t(\beta_1 + \sigma_1) + \xi_t)} \\
&= \frac{1}{2} \cdot \frac{\exp(\delta_t)}{1 + \exp(\delta_t)} + \frac{1}{2} \cdot \frac{\exp(\delta_t + \sigma_0 + X_t\sigma_1)}{1 + \exp(\delta_t + \sigma_0 + X_t\sigma_1)},
\end{aligned}$$

Now suppose that the true market size is $M_t^* = \gamma P_t$, where P_t is the population of market t, and the observed market share is $s_t = \sum_{j=1}^{J_t} N_{jt}/P_t$, where $\sum_{j=1}^{J_t} N_{jt}$ is the sale in market t observed without error. Then the observed and true market share would be linked by $s_t = \gamma\pi_t$. Following the BLP practice, we can implicitly solve for δ_t by equating

$$\frac{s_t}{\gamma} = \pi_t(\delta_t; \sigma) \tag{29}$$

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 make things simple 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

$$E\begin{bmatrix} \delta_t - \beta_0 \\ (\delta_t - \beta_0)Z_{1t} \\ (\delta_t - \beta_0)Z_{2t} \end{bmatrix} = 0. \tag{30}$$

Note that δ_t is solved from equation (29), 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 (30). A sufficient condition is that the Jacobian matrix with respect to $(\beta_0, \sigma_0, \gamma)$ is non-singular.

To save space, 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:

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

When the relevance assumptions are satisfied, the Jacobian matrix is non-singular and therefore the moment conditions (30) 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.

F Merger Analysis: Ready-to-Eat Cereal Market

Data and Demand Specification

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). We 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. We 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. We 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.

Results

Table F.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. Our 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 F.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

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

and $V_{ijt} \equiv U_{ijt} - \varepsilon_{ijt}$.

Table F.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.

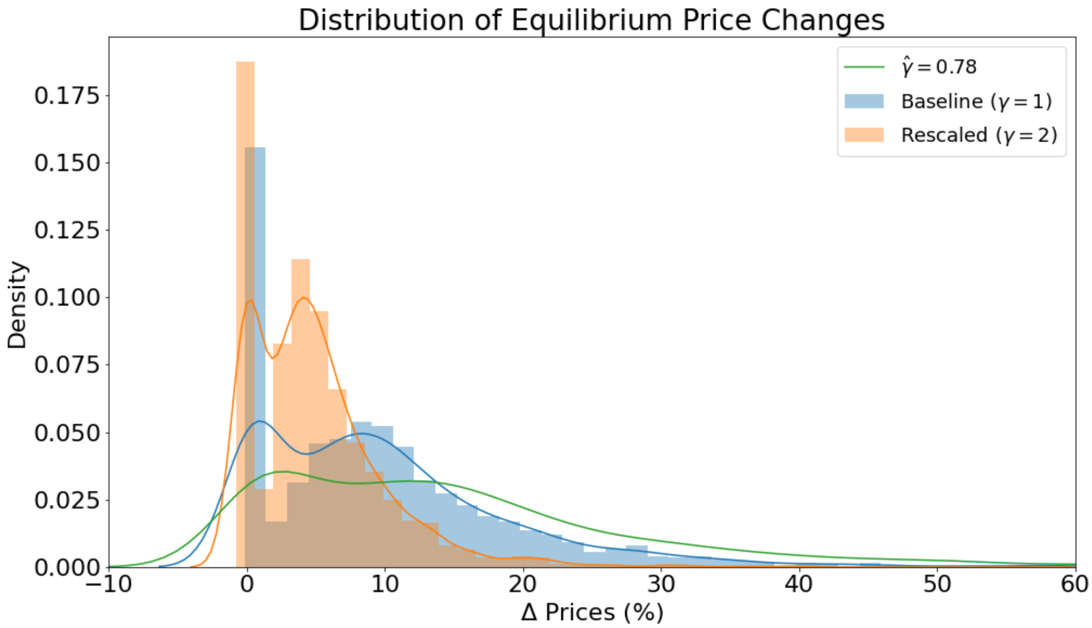


Figure F.1: Equilibrium Price Changes

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

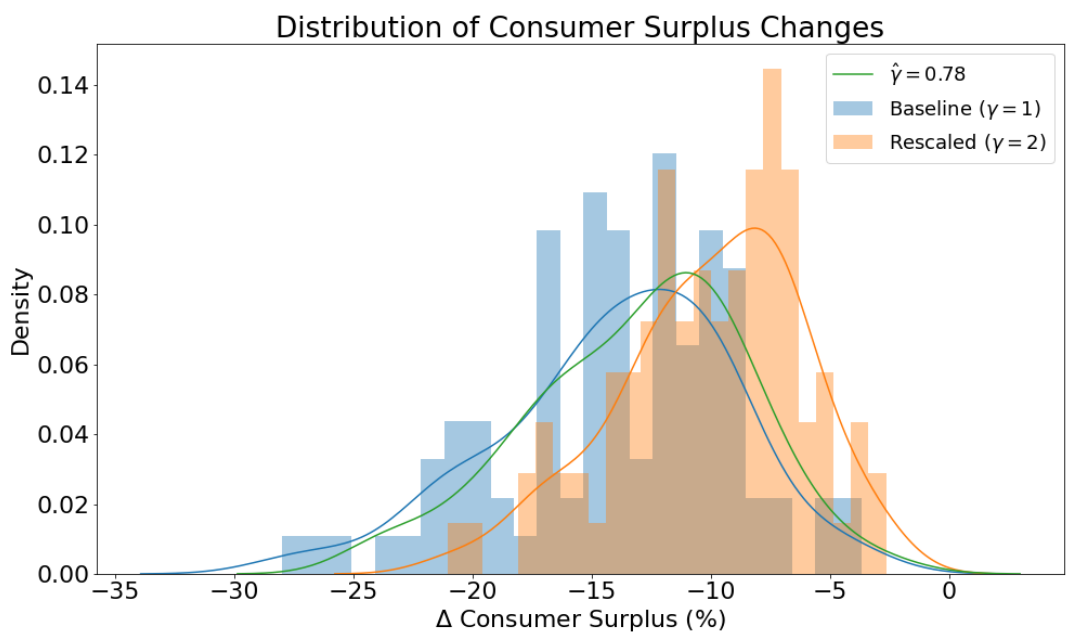


Figure F.2: Consumer Surplus Changes

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