# Semiparametric Estimation of a CES Demand System with Observed and Unobserved Product Characteristics

**Abstract**

We develop a characteristics based demand estimation framework for the Marshallian demand system obtained by solving a budget-constrained constant elasticity of substitution (CES) utility maximization problem. From our Marshallian CES demand system, we derive the same market share equation of Berry (1994); Berry, Levinsohn, and Pakes (1995)'s characteristics based logit demand system. Our CES demand estimation framework can accommodate zero predicted and observed market shares by conceptually separating the whether-to-buy decision and how-much-to-buy decision. Furthermore, the estimator we suggest allows a tractable semiparametric estimation strategy that is flexible regarding the distribution of unobservable product characteristics. We apply our framework to scanner data on cola sales, where we show estimated demand curves can be upward sloping if zero market shares are not accommodated properly.

JEL classification: C51, D11, D12

# 1 Introduction

Constant elasticity of substitution (CES) preferences, often called Dixit-Stiglitz-Spence preferences, have been used extensively to analyze markets with product differentiation since Spence (1976); Dixit and Stiglitz (1977); Anderson (1979); Krugman (1980). In marketing, demand systems derived from CES preference or its variants have been used extensively in combination with mostly individual- or purchase-level data (see, e.g., Kim et al., 2002; Allenby et al., 2004; Dubé, 2004; Kim et al., 2007; Lee et al., 2013; Lee and Allenby, 2014; Howell et al., 2016 among others). However, when only aggregate market-level data are available to a researcher, Berry (1994); Berry, Levinsohn, and Pakes (1995)'s demand estimation framework has become the *de facto* standard method, which is based on a different microfoundation – the discrete choice random utility model in the product characteristic space. We reconcile these approaches of differentiated products demand estimation using aggregate market-level data, showing that *the direct utility approach* based on CES preferences can also be just as rich and flexible as Berry (1994); Berry, Levinsohn, and Pakes (1995)'s demand estimation framework. We thereby shed light on an important connection between *the direct utility approach* and *the indirect utility approach* (Section 3.1 of Chintagunta and Nair, 2011), or, stated differently, between *the neoclassical model* and *the pure discrete choice model* (Section 3 of Dubé, 2018), in consumer demand estimation.

In this paper, we provide a general CES demand estimation framework that can accommodate zero observed and predicted market shares. To accommodate the zero shares, we develop a two-stage model of discrete-continuous choice based on a budget-constrained CES utility maximization problem. Our CES demand estimation framework is attractive for the following reasons. First, we show that the identical market share equation of Berry (1994); Berry et al. (1995) can be derived from the budget-constrained CES utility maximization problem. It allows the identification results and estimation strategy developed for Berry (1994); Berry et al. (1995) to be directly applied to the CES demand system when zero market shares are not present in the data. Second, when zero market shares are present in the data, we explicitly introduce the exclusion restriction on whether-to-buy decision of consumers, providing a conceptually clean identification argument. Third, we employ a tractable semiparametric estimation strategy that is flexible regarding the distribution of unobservable product characteristics.

The current *de facto* standard framework for differentiated products demand estimation using aggregate market data was developed by Berry (1994); Berry et al. (1995), which made breakthroughs in the demand estimation literature in several aspects.[1] One of the breakthroughs was to (re)introduce the characteristic space approach, which dates back to Lancaster (1966), in demand estimation. The characteristic space approach can be very useful in predicting the demand for a new product and evaluating its effects on the market (Petrin, 2002). In Berry (1994); Berry et al.

[1] The breakthroughs include explicitly recognizing the correlation of unobservable characteristics with the prices, market share inversion, and simulation methods to estimate the random coefficients.

(1995)'s characteristics based demand estimation framework, a product is defined as a bundle of observed and unobserved product characteristics. A consumer can choose up to one product that yields the highest utility among her finite choice set, or can decide to buy nothing. A consumer's (dis)utility of consuming a product consists of the utility from price, observed product characteristics, unobserved product characteristics, and idiosyncratic utility shock. The individual choice probability equation is derived from the distributional properties of the idiosyncratic utility shock, which is assumed to follow the Type-I extreme value distribution. Individual choice probabilities are taken as equal to the predicted quantity shares of the individual demand, the aggregation of which is taken as the predicted quantity market shares. We refer to demand models based on these microfoundations as logit demand models, which provide a tractable method of estimating differentiated product demand systems by reducing the dimension of the parameters to be estimated.

Our first contribution to the literature is provision of a concrete link between the CES demand system and Berry (1994); Berry et al. (1995)'s homogeneous/random coefficients logit demand system. We do so by deriving the identical market share equation of Berry (1994); Berry et al. (1995)'s characteristics based logit demand system from the CES demand system. A nonnegative function in CES preferences that we refer to as the *quality kernel,* which is often referred to as the "taste parameter" in the literature, plays a key role in directly incorporating observed and unobserved product characteristics into the CES demand system. Incorporating the taste parameter in CES preferences dates back to at least Spence (1976); Anderson (1979). To name just a few, Kim et al. (2002); Dubé (2004) incorporated the idiosyncratic preference shocks on the taste parameter. Einav et al. (2014) incorporated a sales tax indicator and distance from the seller, which are the seller-consumer specific characteristics, into the taste parameter of the CES preferences. However, to the best of our knowledge, none of the literature models the taste parameter of the CES preferences directly as a mapping from the observed and unobserved product characteristics to develop a general empirical framework for demand estimation. Adding the quality kernel allows us to derive the identical, predicted quantity individual/market share equation of Berry (1994); Berry et al. (1995) from the resulting Marshallian CES demand system. Early studies by Anderson et al. (1987, 1992) point out *similarities* between the CES and logit demand systems without product characteristics. Our market share equation equivalence result is an extension of Anderson et al. (1987, 1992), in the context of Berry (1994); Berry et al. (1995)'s characteristics based demand estimation framework.

Our second contribution is the development of a direct method that accommodates zero predicted and observed market shares. Accommodating zero market shares in demand estimation has been a major difficulty in the literature for decades, dating back to at least Deaton and Muellbauer (1980). Berry (1994); Berry et al. (1995)'s logit demand model is not an exception. Discrete choice frameworks with additive idiosyncratic errors with unrestricted support inherently do not allow for zero individual choice probabilities. Individual choice probabilities are treated as predicted individual quantity shares, aggregated over homogeneous or heterogeneous individuals and equated with observed market shares for identification and estimation of model parameters. In logit demand models, additive idiosyncratic shocks are distributed as an i.i.d. Type-I extreme value. In such a case, the numerator of the individual choice probability is the exponential of the alternative utility's deterministic part. Provided that an alternative yields any utility higher than negative infinity, the alternative must have a strictly positive predicted market share. However, zero observed market shares are often observed in data. Thus, in practice, researchers simply drop samples with zero observed market shares or add a small, arbitrary number to zero observed market shares. These *ad hoc* measures cause biases in estimates. We argue that selection in a consumer's consideration set must be taken into account for identification and estimation of model parameters. The consideration set selection drives the conditional expectation of unobserved product characteristics that are conditioned on instruments being non-zero and likely positive. The usual generalized method of moments estimation yields price coefficient estimates that are biased upward when this consideration set selection process is ignored.

To accommodate the zero predicted and observed market shares, we provide a microfoundation for the selection-correction estimation equation *à la* Heckman (1979), by embedding both extensive and intensive margins on the quality kernel. Dubin and McFadden (1984); Hanemann (1984); Chiang (1991); Chintagunta (1993); Kim et al. (2002); Nair et al. (2005) introduced and developed modeling both margins in a utility maximization problem to model the demand for a variety products. We extend the idea to the Marshallian CES demand system and model a consumer's choice as a two-stage decision process. In addition, we explicitly introduce the exclusion restriction to the consideration set stage, which provides a clean identification argument that separates the whether-to-buy stage and how-much-to-buy stage.

Although we employ the direct utility approach, how we model the demand for variety is different from most existing direct utility maximization models in marketing. We model the demand for variety as a two-stage decision, whether-to-buy and how-much-to-buy, in contrast to the one-step decision on which the marketing literature has focused (see Chintagunta and Nair, 2011; Dubé, 2018 for surveys). We interpret the first-stage whether-to-buy decision as the consideration set formulation, which has a long tradition in the marketing literature (see, e.g., Roberts and Lattin (1991); Ben-Akiva and Boccara (1995); Jedidi et al. (1996); Mehta et al. (2003); Gilbride and Allenby (2004) among many others). It follows naturally that an item that never sold in a market was not in the consideration set of consumers, which motivates our use of the exclusion restriction and the selection-correction estimation equation that accounts for the zero market shares. We employ the Klein and Spady (1993) estimator for the whether-to-buy stage, demonstrating how the distribution-free efficient semiparametric estimator for the binary response model can be easily applied to the demand estimation problem with a multitude of zero predicted and observed market shares. Furthermore, in our empirical example, we provide evidence that the distribution of the unobservable product characteristics is far from Gaussian. As for the contexts in which the single choice assumption is more plausible, we also provide the microfoundation for the same selection-correction estimation equation in the context of the logit demand model.

In our empirical example in which our proposed estimation framework is applied to scanner data that recorded the items on the shelves that never sold during a given week, we demonstrate that dropping zero market shares or imputing them with small positive numbers can cause serious biases in the price-coefficient estimates. In particular, if zero market shares are simply dropped, the price coefficient estimates will be biased upward, even resulting in upward-sloping demand curves. Our results have important implications for estimating demand elasticities using scanner data: items that were on the shelves but not sold at all must be considered and accommodated properly during estimation of demand functions. Such information, however, is not included in the majority of scanner datasets. We suggest that collecting such information during the data collection procedure would be beneficial for correctly estimating demand elasticities.

The remainder of the article is organized as follows. Section 2 briefly summarizes the related literature. Section 3 introduces our suggested CES demand system, and Section 4 derives the market-share equation equivalence result. Section 5 outlines the distribution-free semiparametric estimation framework for our CES demand system, and Section 6 provides a discussion relating our CES demand system to the pure discrete choice logit demand system. Section 7 provides Monte Carlo results, and Section 8 presents an empirical example in which our framework is applied to scanner data with a multitude of zero shares. Section 9 concludes.

## 2 Related Literature

This paper relates to the long tradition of consumer demand modeling in the marketing literature, hedonic demand models in the economics literature, and the economics literature accommodating zero observed shares in multiplicative models.

In marketing, the direct utility approach was developed to accommodate the purchase of multiple categories/brands. The most popular direct utility specification would be translog preferences, which date back to Christensen et al. (1975). The richness of translog preferences comes from the second-order term that allows for the possibility of complements. In practice, however, second-order terms of translog preferences are often omitted for the sake of tractability, in which case the preferences can be nested as a variant of CES preference (Bhat, 2008). Kim et al. (2002); Dubé (2004); Kim et al. (2007) used CES preference or its variants in modeling the demand for a variety of consumers. Allenby et al. (2004); Howell et al. (2016) used CES preference in the context of non-linear budget constraints associated with quantity discount or price promotion, respectively, and Lee et al. (2013); Lee and Allenby (2014) used it in the context of asymmetric complements and indivisibility of demand.

Marketing literature has a long tradition of multiple discrete choice and discrete-continuous models of demand that dates back to Hanemann (1984). Chiang (1991); Chintagunta (1993); Mehta (2007); Song and Chintagunta (2007) model quantity choice and brand choice simultaneously, which entails the extensive margins and intensive margins, respectively. Contrary to their focus on the single brand choice, however, we focus on the zero observed quantity shares and provide a tractable estimation method accommodating the zero shares. The econometric model we derive is conceptually similar to Gilbride and Allenby (2004); Nair et al. (2005), where the screening and choice components are modeled simultaneously. However, neither of them had an exclusion restriction for discrete and quantity choice for identification, and they do not account for zero market shares.

Logit models of consumer demand in marketing date back to at least Guadagni and Little (1983). When market-level data are available to the researcher, the demand estimation framework developed by Berry (1994); Berry et al. (1995) has become the *de facto* standard method to estimate the demand, both in economics and marketing (see, e.g., Besanko et al. 1998; Sudhir 2001; Chintagunta et al. 2002; Chintagunta and Desiraju 2005; Draganska and Jain 2006; Hitsch 2006; Wilbur 2008; Albuquerque and Bronnenberg 2009; Goldfarb et al. 2009; Ghose et al. 2012 among many others). The marketing literature suggests a few variants to Berry (1994); Berry et al. (1995) as well. Chintagunta (2001) developed an estimation method for when the idiosyncratic error term is Gaussian, and Bruno and Vilcassim (2008) developed one for when the product availability is varying. When individual-level data are available, Villas-Boas and Winer (1999); Chintagunta et al. (2005) used random coefficients demand estimation with instrumental variables, to examine the effect of correcting for endogeneity in brand choice.

Our approach to construct an empirical demand system without an idiosyncratic preference shock can be viewed as a hedonic, or pure characteristics, model of demand. Recent developments on hedonic demand estimation frameworks were made by Bajari and Benkard (2005); Berry and Pakes (2007), the former of which relates more closely to our study. Bajari and Benkard investigate a general hedonic model of demand with product characteristics, focusing on local identification and estimation of model parameters. For global identification when a product space is continuous, they specify Cobb-Douglas preferences. Our study extends their Cobb-Douglas specification to the more flexible CES preferences specification that can also accommodate zero predicted and observed market shares.

Two papers in economics literature have tried to accommodate the zero shares in the multiplicative models. Gandhi et al. (2013) rationalize zero observed market shares differently, regarding such shares as measurement errors of strictly positive predicted market shares, and provide a partial identification result of model parameters. The difference between their research and ours is that we rationalize zero predicted and observed market shares, whereas they allow only observed market shares to be zero. Nevertheless, their Monte Carlo simulations and empirical applications suggest an implication similar to ours: when samples with zero market shares are dropped, price coefficient estimates are biased upward. In the international trade literature, Helpman et al. (2008) developed another method that relates closely to ours in the context of gravity models. They used a gravity model with endogenous censoring of trade volumes, and their structural approach to handling zero trade flows is similar to ours. However, their approach is fully parametric in that they assume the Gaussian error term, whereas our approach is semiparametric because we do not specify the distribution of unobservables in our preferred specification.

## 3 CES Demand System with Observed and Unobserved Product Characteristics

### 3.1 Specification of the CES Demand System

We consider a differentiated product market denoted by subscript $t$, composed of homogeneous consumers with a CES preference. We begin by focusing on homogeneous consumers. The extension to product markets comprised of heterogeneous consumers, with each consumer allowed to have disparate utility parameters, is considered in Section 4.1. The utility from a product category is:

$$u\left(\left\{q_{j,t}, \mathbf{x}_{j,t}, \xi_{j,t}, \mathbf{w}_{j,t}, \eta_{j,t}\right\}_{j\in\mathcal{J}_t}\right) := \left(\sum_{j\in\mathcal{J}_t} \left\{\chi\left(\mathbf{x}_{j,t}, \xi_{j,t}, \mathbf{w}_{j,t}, \eta_{j,t}\right)\right\}^{\frac{1}{\sigma}} q_{j,t}^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}}. \tag{3.1}$$

Set $\mathcal{J}_t$ is a set of alternatives in the category, which might include the numeraire that represents the outside option. $q_{j,t}$ is the quantity of product $j$ consumed in market $t$. $\chi\left(\mathbf{x}_{j,t}, \xi_{j,t}, \mathbf{w}_{j,t}, \eta_{j,t}\right)$, defined by the quality kernel, is a non-negative function of observed and unobserved product characteristics. $\mathbf{x}_{j,t}$ and $\mathbf{w}_{j,t}$ are vectors of product $j$'s characteristics in market $t$, which are observable to the econometrician. $\xi_{j,t}$ and $\eta_{j,t}$ are scalars that represent utility from product $j$'s characteristics that are unobservable to the econometrician. $\mathbf{w}_{j,t}$ and $\eta_{j,t}$ are extensive margin shifters that a consumer considers whether to buy the product. $\mathbf{x}_{j,t}$ and $\xi_{j,t}$ are intensive margin shifters that determine the level of utility when a consumer buys a product. $\mathbf{w}_{j,t}$ and $\mathbf{x}_{j,t}$ might have common components, but we can require exclusion restriction on $\mathbf{w}_{j,t}$ for semiparametric identification when the extensive margin matters. In such a case, $\mathbf{w}_{j,t}$ must contain at least one component that is not in $\mathbf{x}_{j,t}$. We explain identification conditions further in Section 5. The observed extensive margin shifter, $\mathbf{w}_{j,t}$, might contain the price $p_{j,t}$ or a nonlinear function of $p_{j,t}$.

The quality kernel $\chi\left(\mathbf{x}_{j,t}, \xi_{j,t}, \mathbf{w}_{j,t}, \eta_{j,t}\right)$, introduced in equation (3.1), is critical to our framework. Researchers conventionally employ taste parameters or utility weights in places we put the quality kernel. The quality kernel, taste parameters, and utility weights can be commonly interpreted as multipliers on the (marginal) utility of consuming a product. The quality kernel is a straightforward extension of such conventions that allows us to incorporate observed and unobserved product characteristics directly into a consumer's utility. The quality kernel also allows the possibility of explicitly separating intensive and extensive margins. This feature accommodates zero predicted and observed market shares in model parameters.

The representative consumer's budget-constrained utility maximization problem, the solution of which is the Marshallian demand system, is:

$$\max_{\{q_{j,t}\}_{j\in\mathcal{J}_t}} \left(\sum_{j\in\mathcal{J}_t} \{\chi\left(\mathbf{x}_{j,t}, \xi_{j,t}, \mathbf{w}_{j,t}, \eta_{j,t}\right)\}^{\frac{1}{\sigma}} q_{j,t}^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}} \qquad s.t. \qquad \sum_{j\in\mathcal{J}_t} p_{j,t}q_{j,t} = y_t. \tag{3.2}$$

The Marshallian demand system is:

$$q_{j,t} = y_t \left\{\frac{\chi\left(\mathbf{x}_{j,t}, \xi_{j,t}, \mathbf{w}_{j,t}, \eta_{j,t}\right) p_{j,t}^{-\sigma}}{\sum_{k\in\mathcal{J}_t} \chi\left(\mathbf{x}_{k,t}, \xi_{k,t}, \mathbf{w}_{k,t}, \eta_{k,t}\right) p_{k,t}^{1-\sigma}}\right\} \qquad \forall j \in \mathcal{J}_t, \tag{3.3}$$

which leads to the predicted quantity market shares expression:

$$\begin{aligned} \pi_{j,t} &\equiv \frac{q_{j,t}}{\sum_{k\in\mathcal{J}_t} q_{k,t}} \\ &= \frac{\chi\left(\mathbf{x}_{j,t}, \xi_{j,t}, \mathbf{w}_{j,t}, \eta_{j,t}\right) p_{j,t}^{-\sigma}}{\sum_{k\in\mathcal{J}_t} \chi\left(\mathbf{x}_{k,t}, \xi_{k,t}, \mathbf{w}_{k,t}, \eta_{k,t}\right) p_{k,t}^{-\sigma}}. \end{aligned} \tag{3.4}$$

Equation (3.4) is what we call the CES demand system with observed and unobserved product characteristics. The demand system (3.4), which is in the form of the predicted quantity market shares, is our primary interest because the same predicted quantity market share expression from Berry (1994); Berry et al. (1995) can be derived by imposing a further structure on the quality kernel, $\chi(\cdot)$. (3.4), a system of predicted quantity market shares, imposes only $\#(\mathcal{J}_t) - 1$ constraints on the Marshallian demand system, $\mathbf{q}_t$, in (3.3). Only when combined with the budget constraint $\sum_{j\in\mathcal{J}_t} p_{j,t}q_{j,t} = y_t$ can the Marshallian demand quantities, $\mathbf{q}_t$, be uniquely pinned down for a given price vector, $(\mathbf{p}_t, y_t) \in \mathbb{R}^{\#(\mathcal{J}_t)+1}$.

For the invertibility of the demand system, (3.4), we consider the subset $\mathcal{J}_t^+ (\subseteq \mathcal{J}_t)$, such that $\pi_{j,t} > 0$ for all $j \in \mathcal{J}_t^+$. The demand system specified by $\{\pi_{j,t}\}_{j\in\mathcal{J}_t^+}$ satisfies the connected substitutes conditions from Berry et al. (2013), and is thus invertible. Invertibility of the demand system implies that $\sigma$, the elasticity of substitution, is identified. If we impose suitable structures on $\chi(\cdot)$, such as monotonicity with index restriction, the structural parameters of $\chi(\cdot)$ are also identified. We investigate the specific functional forms of $\chi(\cdot)$ in Section 4.

## 3.2 Properties of the CES Demand System and Comparison with the Logit Demand System

We now explain the properties of the CES demand system (3.4). We begin with the Marshallian and Hicksian own and cross price elasticities of the demand system. Let $b_{j,t}$ and $\pi_{j,t}$ be the budget and quantity share of product $j$ in market $t$, respectively.[2] Denote $\varepsilon_{jc,t}^{M}$ and $\varepsilon_{jc,t}^{H}$ by the Marshallian and Hicksian cross price elasticities between alternatives $j$ and $c$, respectively. If $\mathbf{w}_{j,t}$ does not include

[2]We use the term budget share and expenditure share exchangeably.

the prices or function of the prices as its component, we have the following simple closed-form formulas for the Marshallian and the Hicksian own and cross price elasticities:

$$\begin{aligned}
\varepsilon^{M}_{jj,t} &= -\sigma + (\sigma - 1)\, b_{j,t} \\
\varepsilon^{M}_{jc,t} &= (\sigma - 1)\, b_{c,t} \\
\varepsilon^{H}_{jj,t} &= -\sigma \left(1 - b_{j,t}\right) \\
\varepsilon^{H}_{jc,t} &= \sigma b_{c,t},
\end{aligned} \tag{3.5}$$

and the income elasticity is 1.[3][4] These elasticities can be easily calculated given that $\sigma$ is identified. From these elasticity expressions, it can be immediately noticed that a version of the independence of irrelevant alternatives (IIA) property holds; the substitution pattern depends solely on the budget shares of corresponding products. The price elasticities of the CES demand system should not be derived based on the quantity market shares as in the logit demand models.[5] Price elasticities in the logit demand models when the mean utility is log-linear in prices are given by:

$$\begin{aligned}
\varepsilon^{L}_{jj,t} &= -\alpha \left(1 - \pi_{j,t}\right) \\
\varepsilon^{L}_{jc,t} &= \alpha \pi_{c,t}.
\end{aligned} \tag{3.6}$$

[3] In calculating the elasticities in practice, observed budget shares can be used in place of $b_{j,t}$ and $b_{c,t}$.

[4] If $\mathbf{w}_{j,t}$ includes the prices or a function of the prices so that the extensive margin is affected by the price changes, then the simple closed-form expressions for the own and cross elasticities cannot be derived. In practice, the corresponding price elasticities can be calculated using simulations.

[5]

$$\begin{aligned}
\frac{\partial \ln \pi_{j,t}}{\partial \ln p_{j,t}} &= \frac{\partial \left(\ln q_{j,t} - \ln \left(\sum_{k \in \mathcal{J}_t} q_{k,t}\right)\right)}{\partial \ln p_{j,t}} \\
&= \frac{\partial \ln q_{j,t}}{\partial \ln p_{j,t}} - \frac{\partial \ln \left(\sum_{k \in \mathcal{J}_t} q_{k,t}\right)}{\partial \ln p_{j,t}} \\
&= \varepsilon^{M}_{jj,t} - \frac{\partial \ln \left(\sum_{k \in \mathcal{J}_t} q_{k,t}\right)}{\partial \ln p_{j,t}} \\
&\neq \frac{\partial \ln q_{j,t}}{\partial \ln p_{j,t}}
\end{aligned}$$

The term $\frac{\partial \ln \pi_{j,t}}{\partial \ln p_{j,t}}$ is the Marshallian price elasticity only when $\sum_{k \in \mathcal{J}_t} q_{k,t}$ is constant, which is the case for the logit demand models. See Appendix A for the details.

[6][7] The expressions (3.6) parallel the Hicksian price elasticities of the CES demand system. The only difference to the Hicksian price elasticities (3.5), derived from the CES demand system, is that the multiplied terms of the log-price coefficient, $\alpha$, are comprised of quantity market shares, not budget shares.

Because we derive the demand system from the budget-constrained CES utility maximization problem, the duality between the Marshallian and Hicksian demand functions holds. The Slutsky equation follows, and thus we can decompose the substitution and income effect more naturally. The Slutsky equation in the elasticity form is:

$$\varepsilon^{M}_{jc,t} = \varepsilon^{H}_{jc,t} - \varepsilon^{I}_{j,t} b_{c,t}.$$

Because $\varepsilon^{I}_{j,t} = 1$ in the CES demand system, the income effect depends solely on budget shares, which is a considerable limitation. However, there are at least two advantages over the discrete choice counterpart. First, although the numeraire can be included in the consumer's consideration set, $\mathcal{J}_t$, it is unnecessary in our CES demand system. In contrast, inclusion of the numeraire in the consideration set is necessary in the logit demand system to induce an income effect, especially when the income is not a direct argument of the discrete choice utility or it is canceled out.[8] In such a case, the price increase of an alternative leads to consumers switching to only other alternatives in the consideration set. The magnitude of the income effect is in a sense determined *a priori* by the researcher in logit demand models because the income effect depends primarily on quantity market shares of the numeraire. The size of the share of the numeraire is often arbitrarily assumed or imposed by a researcher in practice. Second, the income effect depends on budget shares, not quantity shares, in the Marshallian CES demand system. In logit demand models, the income effect of a product with a small budget share and a large quantity share is large, which is even more unrealistic.

# 4 The Exponential Quality Kernel

So far, we have not restricted the quality kernel, $\chi\left(\mathbf{x}_{j,t}, \xi_{j,t}, \mathbf{w}_{j,t}, \eta_{j,t}\right)$. In principle, $\chi\left(\mathbf{x}_{j,t}, \xi_{j,t}, \mathbf{w}_{j,t}, \eta_{j,t}\right)$ can be any non-negative function. Under this weak restriction, the demand system specified by predicted market shares (3.4) can be identified locally, as investigated by Bajari and Benkard (2005). However, nonparametric estimation of a locally identified demand system places a considerable

[6] When the mean utility is linear in prices, the elasticity expression becomes:

$$\begin{aligned}\varepsilon^{L}_{jj,t} &= -\alpha p_{j,t}\left(1-\pi_{j,t}\right)\\ \varepsilon^{L}_{jc,t} &= \alpha p_{c,t}\pi_{c,t}.\end{aligned}$$

See Section 4.1 for a further discussion on the functional form of the mean utilities in the logit demand model.

[7] In calculating the elasticities in practice, observed quantity shares can be used in place of $\pi_{j,t}$ and $\pi_{c,t}$.

[8] For example, the case when the mean utility is linear in income net of price.

burden on the data and computational power, which is often impractical. Locally identified parameter values are often uninformative regarding counterfactual analyses, and alternatively, we can impose further structures on the consumer utility from product characteristics. We focus on the exponential quality kernel with an index restriction. This specific functional form deserves a special attention for two reasons. First, by using this functional form, we can derive the same individual choice probability equation of the homogeneous and random coefficient logit models of demand from the CES demand system developed in the previous section. Second, this functional form simplifies the estimation problem substantially because the estimation equation reduces to a log-linear form. We use the exponential quality kernel to propose a tractable, semiparametric estimation method that accommodates zero predicted and observed market shares.

### 4.1 Nesting the Homogeneous and Random Coefficient Logit Models of Demand

We show that the predicted quantity market share expressions of the homogeneous and random coefficient logit models of demand can be derived from (3.4) by choosing a functional form of the quality kernel, $\chi\left(\mathbf{x}_{j,t}, \xi_{j,t}, \mathbf{w}_{j,t}, \eta_{j,t}\right)$. Suppose that $\mathbf{x}_{j,t} = \mathbf{w}_{j,t}$, $\xi_{j,t} = \eta_{j,t}$, $\chi\left(\mathbf{x}_{j,t}, \xi_{j,t}\right) > 0$, $\pi_{j,t} > 0$, and let $\mathbf{x}_{j,t}$ be exogenous for all $j, t$. We do not require an exclusion restriction in this setup because the predicted quantity shares are positive for every alternative. Let $\mathcal{J}_t$ contain the numeraire, denoted by product 0, and normalize $p_{0,t} = 1$.[9] Taking the ratios of products $j$ and 0, and taking the logarithm of equation (3.4), yields:

$$\ln\left(\frac{\pi_{j,t}}{\pi_{0,t}}\right) = -\sigma \ln\left(p_{j,t}\right) + \ln \chi\left(\mathbf{x}_{j,t}, \xi_{j,t}\right) - \ln \chi\left(\mathbf{x}_{0,t}, \xi_{0,t}\right). \tag{4.1}$$

We normalize $\mathbf{x}_{0,t} = \mathbf{0}$, $\xi_{0,t} = 0$, and let $\chi\left(\mathbf{x}_{j,t}, \xi_{j,t}\right) = \exp\left(\mathbf{x}_{j,t}'\boldsymbol{\beta} + \xi_{j,t}\right)$. (4.1) then becomes:

$$\ln\left(\frac{\pi_{j,t}}{\pi_{0,t}}\right) = -\sigma \ln\left(p_{j,t}\right) + \mathbf{x}_{j,t}'\boldsymbol{\beta} + \xi_{j,t}. \tag{4.2}$$

(4.2) coincides with the estimation equation of the homogeneous logit model of demand, except that in (4.2), $\ln\left(p_{j,t}\right)$ is used in place of $p_{j,t}$, which is a convention in the literature.[10] The log of price should be used in (4.2) because it is inherited from the consumer's budget constraint. In contrast, we observe that $\ln\left(p_{j,t}\right)$ can also be used in place of $p_{j,t}$ in the utility specification of the logit demand system; by substituting $\ln\left(p_{j,t}\right)$ with $p_{j,t}$ in the linear utility specification in the logit demand model, the estimation equation of the proposed CES demand system lines up exactly with that of the homogeneous logit demand system. We take this substitution with the log of prices as a

[9] We emphasize that $\mathcal{J}_t$ might not contain a numeraire for our CES demand system. In such a case, product 0 can be considered any alternative in $\mathcal{J}_t$, and all estimation equations that follow should be adjusted in terms of differences between $j$ and 0.

[10] See Appendix A for the derivation of homogeneous and random coefficient logit models.

simple scale adjustment in the linear utility specification of the logit demand model. The predicted market share equation of the random coefficient logit model of demand developed by Berry et al. (1995) can be derived similarly. Let $i$ denote an individual, and suppress the market subscript $t$ temporarily. For the sake of notational simplicity, let $\phi_j := \ln p_j$. We specify the quasi-linear utility of the random coefficient logit model of demand as:

$$u_{i,j} = -\alpha_i \phi_j + \mathbf{x}_j' \boldsymbol{\beta}_i + \xi_j + \epsilon_{i,j}.$$

In contrast, the individual quantity share expressions of the CES demand system (3.4) become:

$$\begin{aligned} \pi_{i,j} &= \frac{\chi_i \left(\mathbf{x}_j, \xi_j\right) \exp\left(-\sigma_i \phi_j\right)}{\sum_{k=0}^{J} \chi_i \left(\mathbf{x}_k, \xi_k\right) \exp\left(-\sigma_i \phi_k\right)} && (4.3) \\ &= \frac{\exp\left(-\sigma_i \phi_j + \mathbf{x}_j' \boldsymbol{\beta}_i + \xi_j\right)}{\sum_{k=0}^{J} \exp\left(-\sigma_i \phi_k + \mathbf{x}_k' \boldsymbol{\beta}_i + \xi_k\right)}, && (4.4) \end{aligned}$$

where the second equality follows by specifying $\chi_i \left(\mathbf{x}_j, \xi_j\right) = \exp\left(\mathbf{x}_j' \boldsymbol{\beta}_i + \xi_j\right)$. Note that (4.4) is nearly identical to the individual choice probability equation obtained by Berry et al. (1995).[11] The predicted market share equation is obtained by aggregating these individual quantity shares over $i$.

Discussions in the current subsection provide the microfoundation and justification for international trade and macroeconomics literature, based on the CES demand system, to use differentiated products demand estimation methods developed in empirical industrial organizational literature since Berry (1994); Berry et al. (1995); Nevo (2001). After model parameters are estimated, price and income elasticities can be calculated according to equation (3.5), and the welfare analyses can be conducted correspondingly.

However, discrete choice differentiated product demand estimation literature imposes a critical restriction, which is necessary when inverting the individual quantity share, $\pi_{j,t} > 0$, for all $j, t$.[12] The restriction is inevitable in logit demand models, which assume additive idiosyncratic shocks on preferences distributed with unrestricted support. The most important example in the literature is additive i.i.d. Type-I extreme value distributed shocks. Individual choice probabilities derived from the assumption must have exponential functions in the numerators of choice probabilities. Zero quantity market shares are often observed in data, which are equated with predicted market shares for identification and estimation of model parameters. The flexibility of the quality kernel, $\chi\left(\mathbf{x}_{j,t}, \xi_{j,t}, \mathbf{w}_{j,t}, \eta_{j,t}\right)$, in our model allows us to accommodate zero predicted market shares by embedding a buy-or-not decision of the consumer, which determines extensive margins. We

[11] The only structural difference is the correlation structure of the individual heterogeneity; we must assume that $Cov\left(\sigma_i, \boldsymbol{\beta}_i\right) = \mathbf{0}$. As those cross-correlations are often assumed to be zero in practice when estimating the random coefficient logit model of demand (see Dube et al. (2012)), we do not consider the restriction a serious limitation.

[12] For a detailed discussion on share inversion, see Berry et al. (2013).

now illustrate how to accommodate zero predicted and observed market shares directly.

### 4.2 Accommodating Zero Predicted and Observed Market Shares: Separating Intensive and Extensive Margins

We restrict attention to homogeneous consumers again, and let $\mathbf{x}_{j,t} \neq \mathbf{w}_{j,t}$, $\eta_{j,t} \neq \xi_{j,t}$. We let $\mathcal{J}_t$ contain the numeraire for convenience of illustration, and normalize $p_{0,t} = 1$. The predicted market shares equation of the proposed CES demand system is:

$$\pi_{j,t} = \frac{\chi\left(\mathbf{x}_{j,t}, \xi_{j,t}, \mathbf{w}_{j,t}, \eta_{j,t}\right) \exp\left(-\sigma \phi_{j,t}\right)}{\sum_{k \in \mathcal{J}_t} \chi\left(\mathbf{x}_{j,t}, \xi_{j,t}, \mathbf{w}_{j,t}, \eta_{j,t}\right) \exp\left(-\sigma \phi_{k,t}\right)}. \tag{4.5}$$

The expression (4.5) allows zero predicted market shares of product $j$ by letting $\chi\left(\mathbf{x}_{j,t}, \xi_{j,t}, \mathbf{w}_{j,t}, \eta_{j,t}\right) = 0$ for some subset of the product characteristic space where $\left(\mathbf{w}_{j,t}, \eta_{j,t}\right)$ lives on. By taking the ratio $\pi_{j,t}/\pi_{0,t}$, we obtain a reduced form of the demand system (4.5) as:

$$\frac{\pi_{j,t}}{\pi_{0,t}} = p_{j,t}^{-\sigma} \frac{\chi\left(\mathbf{x}_{j,t}, \xi_{j,t}, \mathbf{w}_{j,t}, \eta_{j,t}\right)}{\chi\left(\mathbf{x}_{0,t}, \xi_{0,t}, \mathbf{w}_{0,t}, \eta_{0,t}\right)}. \tag{4.6}$$

If $\mathcal{J}_t$ does not include the numeraire, any product with a strictly positive market share can be considered a reference product, denoted by product 0. All arguments in the current and subsequent sections remain valid provided that statistical independence of the observable and unobservable product characteristics across products can be assumed. This assumption implies that product characteristics are uncorrelated across products, which is consistent with many extant demand estimation frameworks, including Berry (1994); Berry et al. (1995). For tractability during identification and estimation, we consider the following functional form with an index restriction:

$$\chi\left(\mathbf{x}_{j,t}, \xi_{j,t}, \mathbf{w}_{j,t}, \eta_{j,t}\right) = \mathbf{1}\left(\left\{\gamma + \mathbf{w}_{j,t}'\delta + \eta_{j,t} > 0\right\}\right) \exp\left(\alpha + \mathbf{x}_{j,t}'\boldsymbol{\beta} + \xi_{j,t}\right), \tag{4.7}$$

where $\mathbf{1}\left(\cdot\right)$ is an indicator function. Employing this quality kernel is equivalent to assuming a certain structure on the consumer's choice. A consumer initially considers the utility from product characteristics represented by $\mathbf{w}_{j,t}'\delta + \eta_{j,t}$. If the utility exceeds the threshold $-\gamma$, the consumer decides to buy the product. Then $\left(\phi_{j,t}, \mathbf{x}_{j,t}, \xi_{j,t}\right)$ is considered, which affects the amount of consumption $q_{j,t}$. In contrast, if the utility does not exceed the threshold $-\gamma$, the consumer decides not to buy the product, and thus, $q_{j,t} = \pi_{j,t} = 0$. We emphasize that $\mathbf{w}_{j,t}$ can contain the raw price, $p_{j,t}$, or other endogenous variables provided that the corresponding instruments are available to the researcher.

# 5 A Semiparametric Estimation Framework with Exponential Quality Kernel and Zero Market Shares

We provide a semiparametric estimation framework for the CES demand system with exponential quality kernel that accommodates zero predicted and observed market shares. The estimation method we provide includes two stages. During the first stage, parameters that determine extensive margins are estimated using the efficient semiparametric estimator developed by Klein and Spady (1993), and during the second, parameters that determine intensive margins are estimated, correcting for price endogeneity and selectivity bias caused by a consumer's consideration set selection. The second-stage estimator that we use was developed by Ahn and Powell (1993); Powell (2001). When zero market shares are not observed in the data, one can proceed with existing demand estimation frameworks developed by Berry (1994); Berry et al. (1995) to estimate model parameters. The first-stage estimation framework illustrated in this section allows only exogenous covariates for the observed extensive margin shifter, $\mathbf{w}_{j,t}$. We chose this framework because of the availability of data and efficiency.[13] If a researcher wants to include endogenous variables such as prices in the extensive margin shifters, $\mathbf{w}_{j,t}$, the researcher can proceed with the method developed by Blundell and Powell (2003, 2004) or Rothe (2009) during the first stage. They provide semiparametric estimation frameworks for binary choice models with endogenous covariates.

We assume the existence of instruments for prices, such that $E\left[\xi_{j,t}|\mathbf{z}_{j,t}\right] = 0$, where $\mathbf{z}_{j,t}$ might include $\mathbf{x}_{j,t}$. Let $d_{j,t} = \mathbf{1}\left(\left\{\gamma + \mathbf{w}'_{j,t}\delta + \eta_{j,t} > 0\right\}\right)$. It is well documented in the literature that $E\left[\xi_{j,t}|\phi_{j,t}, \mathbf{x}_{j,t}\right] \neq 0$, and it is highly likely to be positive. Consequently, when prices are not instrumented, upward-sloping demand curves are often estimated. The same intuition applies when a consumer's consideration set selection is ignored and samples with zero observed market shares are simply dropped during estimation. Even after instrumenting for prices, $E\left[\xi_{j,t}|\mathbf{z}_{j,t}\right] = 0$ does not imply that $E\left[\xi_{j,t}|\mathbf{z}_{j,t}, d_{j,t} = 1\right]$ is zero. $E\left[\xi_{j,t}|\mathbf{z}_{j,t}, d_{j,t} = 1\right]$ is likely to be positive because consumers select products with high $\eta_{j,t}$ during the first-stage consideration set decision, and $\eta_{j,t}$ is likely to be positively correlated with $\xi_{j,t}$. Thus, dropping samples with zero observed market shares during estimation biases price coefficients upward, which can even yield positive price coefficients. Imputing zero observed market shares with some small positive numbers during estimation can cause an even more serious problem in that the direction of the bias is unpredictable.

We normalize $\phi_{0,t} \equiv \ln p_{0,t} = 0$, $\xi_{0,t} = \eta_{0,t} = 0$, $\mathbf{w}_{0,t} = \mathbf{0}$, and $\mathbf{x}_{0,t} = \mathbf{0}$. Under the choice of $\chi(\cdot)$ specified in (4.7), (4.6) simplifies to:

$$\frac{\pi_{j,t}}{\pi_{0,t}} = \mathbf{1}\left(\left\{\gamma + \mathbf{w}'_{j,t}\delta + \eta_{j,t} > 0\right\}\right)\exp\left(-\sigma\phi_{j,t} + \mathbf{x}'_{j,t}\boldsymbol{\beta} + \xi_{j,t}\right), \tag{5.1}$$

[13] Although our data include information on product availability to consumers, even when sales in a corresponding week/store pair were zero, they do not include prices in corresponding weeks/stores without sales. Thus, we could not contain the endogenous variable $p_{j,t}$ in $\mathbf{w}_{j,t}$ during first-stage estimation. Characteristics of data used in our empirical application are discussed in Section 8.

which is the econometric model that we identify and estimate in this section. A consumer buys product $j$ if $\gamma + \mathbf{w}'_{j,t}\delta + \eta_{j,t} > 0$. For the sample with $d_{j,t} = 1$, demand system (5.1) further reduces to:

$$\ln\left(\frac{\pi_{j,t}}{\pi_{0,t}}\right) = -\sigma\phi_{j,t} + \mathbf{x}'_{j,t}\boldsymbol{\beta} + \xi_{j,t}.$$

However, the conditional expectation $E\left[\xi_{j,t}|\mathbf{z}_{j,t}, \mathbf{w}_{j,t}, d_{j,t} = 1\right]$ is not zero anymore, which leads to the sample selection problem. Several methods to estimate parameters of the sample selection models have been proposed in the literature under different assumptions.[14] We follow Heckman (1979), who imposes a conditional mean restriction. By taking the conditional expectation, we have:

$$E\left[\ln\left(\frac{\pi_{j,t}}{\pi_{0,t}}\right)|\mathbf{z}_{j,t}, \mathbf{w}_{j,t}, d_{j,t} = 1\right] = -\sigma\phi_{j,t} + \mathbf{x}'_{j,t}\boldsymbol{\beta} + E\left[\xi_{j,t}|\mathbf{z}_{j,t}, \mathbf{w}_{j,t}, d_{j,t} = 1\right]. \tag{5.2}$$

Ahn and Powell (1993); Powell (2001); Newey (2009) propose two-stage $\sqrt{N}$-consistent estimators for the model parameters of (5.2). We use the pairwise differenced weighted least squares estimator from Ahn and Powell (1993); Powell (2001), which corrects for the endogeneity of $\phi_{j,t}$ using instruments during the second stage. During the first stage, $\delta$ should be estimated. A few estimators are available for this semiparametric binary choice model, among which we use the method from Klein and Spady (1993), which achieves asymptotic efficiency. During the second stage, parameters $(\sigma, \boldsymbol{\beta})$ from the following linear equation are estimated:

$$E\left[\ln\left(\frac{\pi_{j,t}}{\pi_{0,t}}\right)|\mathbf{z}_{j,t}, \mathbf{w}_{j,t}, d_{j,t} = 1\right] = -\sigma\phi_{j,t} + \mathbf{x}'_{j,t}\boldsymbol{\beta} + \lambda\left(1 - G_\eta\left(-\mathbf{w}'_{j,t}\delta\right)\right), \tag{5.3}$$

where $\lambda(\cdot)$ is an unknown smooth function. For semiparametric identification of $(\sigma, \boldsymbol{\beta})$, term $\lambda\left(1 - G_\eta\left(-\mathbf{w}'_{j,t}\delta\right)\right)$ must not be a linear combination of $\left(\phi_{j,t}, \mathbf{x}_{j,t}\right)$; some component of $\mathbf{w}_{j,t}$ must be excluded from $\left(\phi_{j,t}, \mathbf{x}_{j,t}\right)$. We impose the following assumptions on the data-generating process for the $\sqrt{N}$-consistency and asymptotic normality of our proposed estimator.

**Assumption 1.** *The vector of observed product characteristics $\mathbf{w}_{j,t}$ is exogenous.*

**Assumption 2.** *$\mathbf{w}_{j,t}$ contains at least one component that is not included in $\mathbf{x}_{j,t}$.*

**Assumption 3.** *$\eta_{j,t}$ is independent of $\mathbf{w}_{j,t}$ with $E\left[\eta_{j,t}|\mathbf{w}_{j,t}\right] = 0$, $\eta_{j,t}$ is i.i.d. over $j$ and over $t$, and the conditional distribution of $\eta_{j,t}$ given $\mathbf{w}_{j,t}$ has the full support over $\mathbb{R}$ with bounded first derivatives.*

**Assumption 4.** *$\mathbf{w}_{j,t}$ and $\mathbf{x}_{j,t}$ contain at least one continuous variable. Furthermore, the conditional distribution of the continuous variable conditioned on $d_{j,t}$ and other exogenous variables is sufficiently smooth.*

[14]For example, Powell (1984, 1986); Blundell and Powell (2007) propose a least absolute deviation type estimator under the conditional quantile restriction, and Honoré et al. (1997) propose symmetric trimming under the symmetricity assumption of error terms.

**Assumption 5.** *Denote* $\mathbf{r}_j := (\phi_j, \mathbf{x}_j)'$. *Denote* $g_{(\cdot)}(w) := E\left[\cdot|\mathbf{w}_{j,t}'\delta = w\right]$ *and* $f(w)$ *be the density of* $\mathbf{w}_{j,t}'\delta$. *Then,* $f(w)$, $g_{\mathbf{r}_{j,t}}(w)$, $g_{\mathbf{z}_{j,t}}(w)$, $g_{\eta_{j,t}}(w)$, $g_{\mathbf{z}_{j,t}\mathbf{w}_{j,t}'}(w)$, *and* $g_{\mathbf{z}_{j,t}\mathbf{r}_{j,t}'}(w)$ *are sufficiently smooth on their supports.*

**Assumption 6.** *There exists a set of instruments* $\mathbf{z}_{j,t}$ *such that* $\xi_{j,t} \perp\!\!\!\perp \phi_{j,t}|\mathbf{z}_{j,t}$, $E\left[\xi_{j,t}|\mathbf{z}_{j,t}\right] = 0$, *and* $\dim\left(\mathbf{z}_{j,t}\right) \geq \dim\left(\phi_{j,t}, \mathbf{x}_{j,t}\right)$.[15]

**Assumption 7.** *The parameter vector* $(\sigma, \alpha, \boldsymbol{\beta}, \gamma, \delta)$ *lies in a compact parameter space, with the true parameter value lying in the interior.*

In Assumptions 1 through 3, we impose the independence of observed and unobserved product characteristics, and homoskedasticity of unobservable product characteristics, $\eta_{j,t}$, that relate to extensive margins. However, we do not assume that unobserved product characteristics and prices are independent. We allow for endogeneity in prices, which should be considered during identification and estimation; prices can be a function of observed and unobserved product characteristics. Assumption 2 is the exclusion restriction, which is required for identification during the second stage. Note that a sufficient condition for Assumption 3 to hold is that $\eta_{j,t} \perp\!\!\!\perp \mathbf{w}_{j,t}$ and $\eta_{j,t}$ has a bounded and continuous density over the real line. Assumptions 4 and 5 are smoothness conditions, imposed for the suggested estimators to be well-behaved.[16] Assumption 6 is the standard instrument condition to correct for price endogeneity.[17] Assumption 7 is the usual compactness assumption.

We now describe first- and second-stage estimators. During the first stage, we estimate $\delta$ using the efficient semiparametric estimator developed by Klein and Spady (1993). The estimator allows us to estimate parameters of binary choice models without having to specify the distribution of the unobservables. The insight is to replace the likelihood with its uniformly consistent estimates, and run the pseudo-maximum likelihood. The estimator is defined as:

$$\hat{\delta} := \arg\max_{\delta} \sum_{j,t} \left\{ \mathbf{1}\left(\pi_{j,t} > 0\right) \ln\left(1 - \hat{G}_\eta\left(-\mathbf{w}_{j,t}'\delta\right)\right) + \mathbf{1}\left(\pi_{j,t} = 0\right) \ln\left(\hat{G}_\eta\left(-\mathbf{w}_{j,t}'\delta\right)\right)\right\}, \tag{5.4}$$

where

$$\hat{G}_\eta\left(-\mathbf{w}_{j,t}'\delta\right) = \hat{\tau}_{j,t} \frac{\sum_{k\neq j,t} \kappa\left(\frac{1}{h_n}\left(\mathbf{w}_k - \mathbf{w}_{j,t}\right)'\delta + \iota_0(\delta)\right)\left(1 - \mathbf{1}\left(\pi_{j,t} > 0\right)\right)}{\sum_{k\neq j,t} \kappa\left(\frac{1}{h_n}\left(\mathbf{w}_k - \mathbf{w}_{j,t}\right)'\delta + \iota(\delta)\right)}.$$

$\kappa(\cdot)$ is a fourth-order kernel, $h_n$ is the bandwidth, and $\hat{\tau}_{j,t}, \iota_0(\delta), \iota(\delta)$ are trimming sequences for small estimated densities.[18] During the second stage, we follow the method illustrated by Powell

[15] $\mathbf{z}_{j,t}$ may contain the exogenous components of $\mathbf{x}_{j,t}$.

[16] Roughly, they require the existence of the higher-order derivatives for the respective conditional distribution and conditional expectation functions. See (C.6) of Klein and Spady (1993) and Assumption 5.7 of Powell (2001) for the exact conditions.

[17] See, e.g., Nevo (2001) for a discussion of suitable instruments in practice.

[18] We ignore these trimming sequences for technical and notational convenience from now on. Klein and Spady (1993) also note that the trimming does not affect the estimates in practice.

(2001). With an abuse of notation by suppressing the market index $t$ and letting $\mathbf{r}_j = (\phi_j, \mathbf{x}_j)'$, the estimator is defined by the following weighted instrumental variable estimator:

$$\left(-\hat{\sigma}, \hat{\boldsymbol{\beta}}\right) = \left(\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \hat{\omega}_{i,j} \left(\mathbf{z}_i - \mathbf{z}_j\right) \left(\mathbf{r}_i - \mathbf{r}_j\right)'\right)^{-1} \left(\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \hat{\omega}_{i,j} \left(\mathbf{z}_i - \mathbf{z}_j\right) \left(\ln\left(\frac{\pi_i}{\pi_0}\right) - \ln\left(\frac{\pi_j}{\pi_0}\right)\right)\right), \tag{5.5}$$

where $\hat{\omega}_{i,j} = \frac{1}{h_n} \kappa \left(\frac{1}{h_n} \left(\mathbf{w}_i - \mathbf{w}_j\right)' \hat{\delta}\right)$.[19][20] Intuition regarding the estimator suggests canceling out the bias correction term, $\lambda \left(1 - G_\eta \left(-\mathbf{w}_j'\delta\right)\right)$; if $\mathbf{w}_i$ is the same as $\mathbf{w}_j$, term $\lambda \left(1 - G_\eta \left(-\mathbf{w}_j'\delta\right)\right)$ in (5.3) cancels out when differences are taken. Thus, more weights are placed on differenced terms that are close. The estimator is $\sqrt{N}$-consistent and asymptotically normal. For the closed-form covariance matrix formula and its consistent estimator, see Powell (2001). The following theorem summarizes discussions in this subsection thus far.

**Theorem 5.1.** *Under Assumptions 1 through 7, $\left(-\hat{\sigma}, \hat{\boldsymbol{\beta}}\right)$, defined in (5.5), is $\sqrt{N}$-consistent and asymptotically normal.*

The semiparametric, log-linear estimation illustrated in this subsection requires an exclusion restriction on $\mathbf{w}_{j,t}$ to identify $(-\sigma, \boldsymbol{\beta})$; $\mathbf{w}_{j,t}$ cannot be a linear combination of $(\phi_{j,t}, \mathbf{x}_{j,t})$. This exclusion restriction can be circumvented by adding an interaction term or nonlinear transformation of a non-binary variable contained in $(\phi_{j,t}, \mathbf{x}_{j,t})$. For example, if one employs the method proposed by Blundell and Powell (2003, 2004), which accommodates endogenous variables during first-stage estimation, including raw prices, $p_{j,t}$, in $\mathbf{w}_{j,t}$ can be a viable choice. However, finding additional exogenous variables that affect only a consumer's buy-or-not decision is ideal. If one is willing to assume that $\eta_{j,t}$ is distributed as standard Gaussian, the classic Heckman correction estimator with instruments can be used, in which the inverse Mills ratio is added as an additional regressor. In that case, identification of model parameters is achieved by the non-linearity of the inverse Mills ratio, and therefore the exclusion restriction is unnecessary.

# 6 Excursus: Derivation of the Selection-Correction Estimation Equation for the Logit Demand Model

In this section, we provide a two-stage model of consumer choice within the logit demand frameworks when zero market shares are present. The logit demand model with two-stage decision process can also lead to the same estimation equation derived from our proposed CES demand

[19]When the number of instruments is larger than that of explanatory variables, the projection matrix can be calculated beforehand to find the $\mathbf{z}_j$ vector. Efficiency loss might occur, but the estimator will be still $\sqrt{N}$-consistent and asymptotically normal.

[20]The bandwidth sequence $h_n$ should be such that $h_n \to 0$, $nh_n^6 \to \infty$, and $nh_h^8 \to 0$ as $n \to \infty$ in both the first and the second stage.

system, which is presented in Section 5. Although sticking to the logit demand frameworks might be less appealing because the intensive and extensive margins cannot be distinguished conceptually, the single-choice assumption can be more adequate in some contexts. In such contexts, the first-stage decision obtains an interpretation akin to the consideration set selection in the consideration set literature.

We show that the estimation equation (5.2) can be derived from the two-stage decision process from the logit demand model. We consider a representative consumer with a two-stage decision process. During the first stage, the consumer searches $\mathcal{J}_t$, which includes all possible alternatives. The consumer's consideration set $\mathcal{J}_t^+$ is determined from the search. During the second stage, the consumer encounters the usual discrete choice decision problem over $\mathcal{J}_t^+$, that is, purchase the product that yields the highest utility. Let $\left(\mathbf{w}_{j,t}, \eta_{j,t}\right)$ be the variables that affect the first-stage consideration set search, and $\left(\mathbf{x}_{j,t}, \xi_{j,t}\right)$ the variables that affect the second-stage discrete choice unconstrained utility maximization problem. Notice that these variables form an analogue of the notations used in Sections 4 and 5. The second-stage utility of the consumer is modeled as:

$$u_{i,j,t} = -\alpha \ln p_{j,t} + \mathbf{x}_{j,t}'\boldsymbol{\beta} + \xi_{j,t} + \epsilon_{i,j,t}.$$

[21][22] The representative consumer solves:

$$\max_{j \in \mathcal{J}_t^+} \left\{ u_{i,j,t} \right\}.$$

With the i.i.d. Type-I extreme value assumption on $\epsilon_{i,j,t}$'s, the individual choice probability becomes:

$$\Pr\left(i \to j | t\right) = \frac{\exp\left(-\alpha \ln p_{j,t} + \mathbf{x}_{j,t}'\boldsymbol{\beta} + \xi_{j,t}\right)}{\sum_{k \in \mathcal{J}_t^+} \exp\left(-\alpha \ln p_{k,t} + \mathbf{x}_{k,t}'\boldsymbol{\beta} + \xi_{k,t}\right)}.$$

$\Pr\left(i \to j | t\right)$ is the predicted quantity market share, $\pi_{j,t}$. During estimation, $\pi_{j,t}$ is equated with the observed market share, $s_{j,t}$. The inversion theorem of Berry (1994); Berry et al. (1995) applies.

[21] By not including the income $y_i$, we disregard the "indirect utility" interpretation of the alternative choice utility $u_{i,j,t}$ here. Early literature on the discrete choice consumer demand, which dates back to McFadden (1974, 1978, 1981); Dubin and McFadden (1984); Anderson et al. (1987, 1988, 1992), stick to the indirect utility interpretation of an alternative utility. To our understanding, the main intention of interpreting $u_{i,j,t}$ as an indirect utility was to place the discrete choice demand systems in the context of the Walrasian demand, especially because the price should not be a direct argument of the Walrasian utility function. However, discrete choice modeling has gained greater popularity and has been applied to a much wider context since McFadden's original works. The "modern" approach tend more to interpret $u_{i,j,t}$ as a direct utility of an alternative. Many recent research using the discrete choice demand estimation framework specify the mean utility as either linear in $-p_{j,t}$ or in $y_i - p_{j,t}$ so the income $y_i$ cancels out over alternatives. See, for example, Berry (1994); Nevo (2001). Berry et al. (1995) suggest using $-p_{j,t}$ in their homogeneous coefficients utility specification and $\ln\left(y_i - p_{j,t}\right)$ in their random coefficients utility specification. We note that $y_i$ does not cancel out over the alternatives in the latter case.

[22] In the context of the logit demand models, the utility can be linear, not log-linear, in prices. If one does not want to interpret the estimated parameters to be originating from a CES demand system, one can replace $\ln p_{j,t}$ with $p_{j,t}$.

Again, the only difference is the moment condition; $E\left[\xi_{j,t}|\mathbf{z}_{j,t}, d_{j,t}=1\right]$ is not zero and is highly likely to be positive. Thus, a correction term is needed for the selection of the consideration set, which leads to the estimation equation:

$$E\left[\ln\left(\frac{\pi_{j,t}}{\pi_{0,t}}\right)|\mathbf{z}_{j,t}, \mathbf{w}_{j,t}, d_{j,t}=1\right] = -\alpha \ln p_{j,t} + \mathbf{x}'_{j,t}\boldsymbol{\beta} + E\left[\xi_{j,t}|\mathbf{z}_{j,t}, \mathbf{w}_{j,t}, d_{j,t}=1\right]. \tag{6.1}$$

(6.1) coincides with (5.2).

# 7 Monte-Carlo Simulations

We simulate market data and back out model parameters to examine the finite-sample performance of the estimator that we proposed in the previous section. We compare the estimation result using our model to the estimation result of the logit demand model, which either drops the sample with zero observed market shares or imputes the zero observed market shares with a small positive number. The estimator we proposed in the previous section works well when the model is specified correctly.

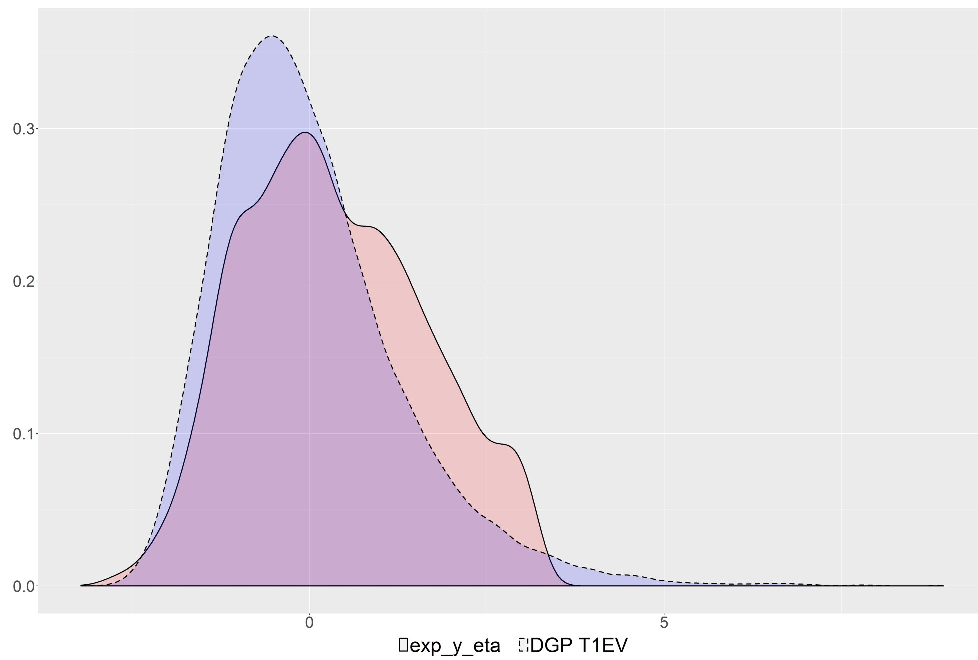
We first describe our data-generating process that satisfies the exclusion restriction. Each market, $t$, has two to five products, with the exact number of products in each market drawn randomly. The observed product characteristic vector, $\mathbf{w}_{j,t}$, includes three continuous components, one discrete component, and three brand dummies. One of the continuous components is excluded in $\mathbf{x}_{j,t}$. The first component, $w_{j,t}^{(1)}$, follows lognormal $(0,1)$, the second component, $w_{j,t}^{(2)}$, follows uniform $(1,5)$, the third, $w_{j,t}^{(3)}$, Poisson $(3)$, and the fourth, $w_{j,t}^{(4)}$, $\mathcal{N}(0,1)$. $w_{j,t}^{(4)}$ is excluded from $\mathbf{x}_{j,t}$. $\eta_{j,t}|\mathbf{w}_{j,t}$ follows the Type-I extreme value distribution with mean zero. Two instruments are employed for prices, which are proxies for cost shocks. Prices, $p_{j,t}$, which is an endogenous variable, is determined by $p_{j,t} = \psi\left(\mathbf{x}_{j,t}, \xi_{j,t}\right)$, where $\psi$ is some (possibly) nonlinear function that is strictly monotonic in $\xi_{j,t}$. We specify $\psi$ as:

$$\psi\left(\mathbf{z}_{j,t}, \xi_{j,t}\right) = 2 + \frac{1}{50}\left(2z_{j,t}^{(1)} + 4z_{j,t}^{(2)} + 2x_{j,t}^{(1)} + x_{j,t}^{(1)}x_{j,t}^{(2)} - x_{j,t}^{(2)}x_{j,t}^{(3)} + 5x_{j,t}^{(4)} + 7x_{j,t}^{(5)} + 9x_{j,t}^{(6)} + 8\xi_{j,t}\right).$$

We intentionally let the influence of the cost proxies, $z_{j,t}^{(1)}$ and $z_{j,t}^{(2)}$, to be fairly weak, which reflects common circumstances in practice. We calibrate the parameters as $\sigma = 2$, $\alpha = 1$, $\boldsymbol{\beta} = (1, -2, 1.5, 0.3, 0.2, 0.4)'$, $\gamma = \alpha$, and $\boldsymbol{\delta} = \frac{1}{4} \times (\boldsymbol{\beta}, 0.1)'$, and market shares are determined by (4.5).

Figure 7.1 depicts the estimated density of $\eta_{j,t}|\mathbf{w}_{j,t}$ from the first stage, and compares it with the distribution used for generating the data. Although the estimated density does not coincide perfectly with the exact density of the Type-I extreme value distribution, it preserves the approximate shape of the distribution. A larger sample is needed for the estimated densities to fit exactly with the distribution used during data generation.

Figure 7.1: Estimated Densities of $\eta_{j,t}|\mathbf{w}_{j,t}$




Note. (i) "exp_y_eta," the pink solid density, is the estimated density of $\eta_{j,t}|\mathbf{w}_{j,t}$ from the Klein-Spady model. "DGP T1EV," the blue dotted density, is the Type-I Extreme Value density that is used to generate the data. (ii) 10,000 sample draws are taken and plotted from the estimated density of the Klein-Spady model and the true Type-I Extreme Value density, respectively. (iii) The Klein-Spady model identifies the distribution of unobservables up to location and scale. Thus, we made the location and scale adjustment.

Table 1: Estimation Result of the Simulated Data

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Estimation | Our Model, K/S | Our Model, Heckman | Logit, Drop 0 | Logit, Impute $10^{-8}$ |
| Log prices $(-2)$ | $-1.969$ | $-1.972$ | $-1.287$ | $-5.444$ |
| | (0.128) | (0.119) | (0.074) | (0.372) |
| $x_{j,t}^{(1)}$ (1) | 0.807 | 1.017 | 0.886 | 1.536 |
| | (0.014) | (0.011) | (0.007) | (0.041) |
| $x_{j,t}^{(2)}$ $(-2)$ | $-1.635$ | $-2.045$ | $-1.574$ | $-3.884$ |
| | (0.028) | (0.035) | (0.013) | (0.050) |
| $x_{j,t}^{(3)}$ (1.5) | 1.222 | 1.530 | 1.339 | 1.821 |
| | (0.020) | (0.023) | (0.008) | (0.040) |
| $x_{j,t}^{(4)}$ (0.3) | 0.251 | 0.290 | 0.366 | $-0.468$ |
| | (0.051) | (0.071) | (0.046) | (0.219) |
| $x_{j,t}^{(5)}$ (0.2) | 0.140 | 0.169 | 0.266 | $-0.673$ |
| | (0.051) | (0.071) | (0.046) | (0.220) |
| $x_{j,t}^{(6)}$ (0.4) | 0.338 | 0.435 | 0.458 | $-0.099$ |
| | (0.053) | (0.073) | (0.047) | (0.225) |
| $D$ | 5059 | 5059 | 5059 | 10500 |
| $N$ | 10500 | 10500 | 5059 | 10500 |

Note. (i) Target values are in parentheses of corresponding items in the first column. (ii) The Estimation row specifies the method used during estimation. Column (1) is our proposed estimator, in which the first-stage propensity score was estimated using the Klein-Spady estimator. For Column (2), Probit was used for the first-stage propensity score estimation, and the inverse Mills ratio is added as an additional regressor during the second stage. Column (3) is the logit estimator with dropping the samples with zero observed market shares, and Column (4) is the logit estimator with imputing $10^{-8}$ in place of the zero observed market shares. (iii) Asymptotic standard error estimates appear in parentheses. (iv) $D$ is the number of non-censored samples, and $N$ is the effective sample size.

Table 1 shows the estimation results of the simulated data. The "Estimation" row indicates the estimation method used. Column (1) is the correct quality kernel specification with our semi-parametric estimator, and Column (2) is the correct quality kernel specification with the classical Heckman correction estimator assuming Gaussian error term in the first stage. Our estimator is successful in recovering the true parameters if the model is specified correctly. The estimator continues to be successful when we estimated the model using the classical Heckman correction estimator that assumes the joint normality of the error term distribution. Column (3) is the logit estimator where we drop the sample with zero observed market shares, and Column (4) is the logit estimator where we impute small positive numbers in place of zero observed market shares. Both dropping zeros and imputing small numbers in place of zeros bias the estimators substantially. The price coefficient is biased upward when the zero shares are dropped, whereas it is biased downward when a small number is imputed in place of zero shares. We also generated and estimated several other specifications, such as different error term distributions, functional forms of quality kernels, variables, pricing functions, etc. For brevity, we do not present all specifications here, but we note that results and implications presented in this section remain robust to these alternative specifications. Details on the estimation procedure appear in Appendix B.

# 8 Empirical Example: Scanner Data with a Multitude of Zero Shares

We implement our proposed demand estimation framework using Dominick's supermarket cola sales scanner data. Data were obtained from the James M. Kilts Center for Marketing, University of Chicago Booth School of Business. The data contained weekly pricing and sales information for the Dominick's chain of stores from 1989 to 1997 for every universal product code (UPC) level product in 29 product categories. Promotion statuses and profitability of each unit sold were also recorded in the data. One shortcoming was that systemic records of product characteristics were unavailable, which we overcame by choosing cola sales data and hand-coding the product characteristics.

## 8.1 Data

We chose Dominick's data because they were ideal for illustrating the application of our framework for two reasons. First, Dominick's data contained information on which products were displayed on the shelves, even if a product did not sell in the corresponding week and store. This feature was necessary because we wanted the exact information on products that were in a consumer's consideration set but were not chosen. Presented in Figure 8.1, approximately one-fourth of observations exhibited zero observed market shares. Second, Dominick's data contained information on average profit per unit sold. Combined with price data, we could back out the average cost per unit. Cost information is useful because an ideal instrument for prices when estimating consumer demand should proxy cost shocks. We avoided constructing instruments using indirect proxies for cost, which has been a major difficulty in demand estimation literature.

We focus on cola sales for several reasons. First, the cola market is a typical market of product differentiation, in which many brands with disparate tastes and packages competes. Among them, Coke and Pepsi, the two prominent brands, take the majority of market shares. Second, product characteristics were not coded separately in Dominick's data, but only category information such as "soft drinks" or "bottled juices." We had to extract the information from product descriptions truncated at 30 characters, for which cola was ideal because it had clearly labeled product characteristics. Finally, companies producing cola, and product characteristics of cola, have not changed much during the past few decades. Coke and Pepsi have been two leaders in the market. Diet, cherry-flavored, and caffeine-free colas are still sold in the market with considerable market shares in 2016, and in 1996. This feature made our analysis convenient, and the implications of analysis more realistic. In Appendices C.3 and C.4, we present estimation results for laundry detergent demand as a robustness check.

Dominick's data covered 100 chain stores in the Chicago area for 400 weeks, from September 1989 to May 1997. We chose the cross-section of week 391, which is the second week of March 1997. We used the cross-section data of a week because demand for soft drinks fluctuates in weeks with holidays or events such as the Super Bowl, and varies considerably by season. Therefore, we chose a week in March without any close holidays. As Dominick's experimented with prices across chain stores for the same product during the same week, we still have sufficient price variations after choosing a cross-section of data. Even after restricting the sample to a cross-section of one week, the sample size was as large as 4,300. We present summary statistics in Table 2.

We define individual products and markets naturally. An individual product was defined by its UPC, and a market by a store-week pair. This was the finest manner of defining a product and market that the data allowed, which resulted in a multitude of zero observed market shares. Illustrated in Figure 8.1, approximately one-fourth of products that were displayed on shelves did not sell.

We converted package prices and costs to per-ounce prices and costs. Dominick's did not record the price and cost of the week if sales of a product were zero in a corresponding week. Therefore, we could not include prices in $\mathbf{w}_{j,t}$, and proceeded only with other exogenous variables during first-stage estimation. When estimating the logit model while substituting the zero observed market shares with small numbers, we imputed missing prices and costs using other chain stores' prices and profits with the same product and promotion status. We had to compute market shares of outside options for both our model and the logit demand model.[23] When estimating market size, we assumed that an average person consumed 100 ounces of soft drinks a week,[24]

[23] Although including the numeraire in a consumer's consideration set was unnecessary in our model, we included it because we wanted to compare estimation results of our model with those of the logit model using the same setup.

[24] On average, Americans consume about 45 gallons of soft drinks a year. Source:

Table 2: Descriptive Statistics

(a) Summary of the Product Characteristics of the Sample

| | Frequency | Mean | Std |
|---|---|---|---|
| Diet | 2163 | 0.497 | 0.500 |
| Caffeine Free | 1085 | 0.249 | 0.433 |
| Cherry | 151 | 0.035 | 0.183 |
| Coke | 365 | 0.084 | 0.277 |
| Pepsi | 2644 | 0.607 | 0.488 |
| Promo | 1751 | 0.402 | 0.490 |
| Bottle Size | - | 26.592 | 29.696 |
| # Bottles per Bundle | - | 12.436 | 9.667 |
| # Stores | 73 | - | - |
| Uncensored Obs ($D$) | 3226 | - | - |
| Sample Size ($N$) | 4356 | - | - |

(b) Per-ounce Price, Cost, Profitability, and Market Shares of Products in the Full Sample

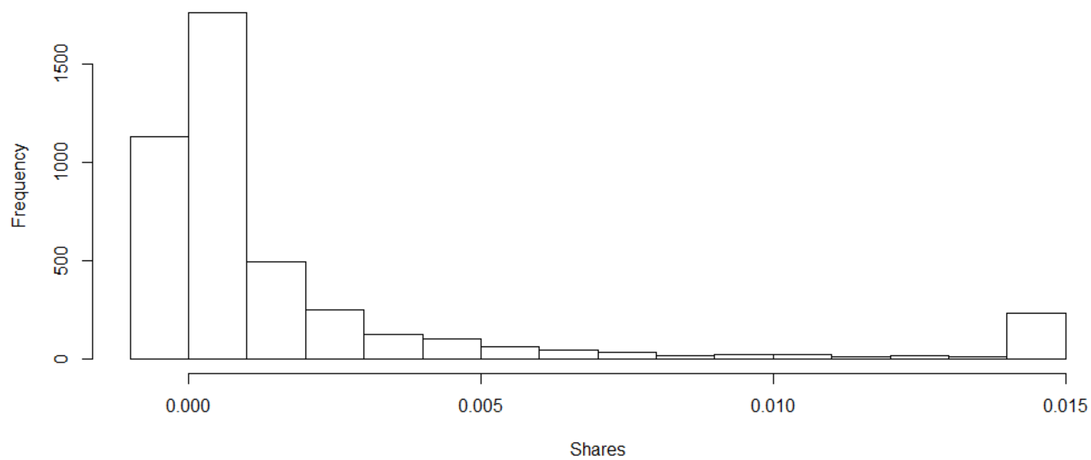
| | Mean | Median | Std | Min | Max |
|---|---|---|---|---|---|
| Per-ounce Prices ($) | 0.020 | 0.024 | 0.014 | 0 | 0.042 |
| Per-ounce Cost ($) | 0.014 | 0.017 | 0.010 | 0 | 0.028 |
| Profitability (%) | 20.470 | 28.380 | 16.020 | −98.550 | 58.620 |
| Shares (%) | 0.586 | 0.038 | 2.879 | 0 | 42.967 |

(c) Per-ounce Price, Cost, Profitability, and Market Shares of Products in the Noncensored Sample

| | Mean | Median | Std | Min | Max |
|---|---|---|---|---|---|
| Per-ounce Prices ($) | 0.027 | 0.026 | 0.008 | 0.005 | 0.042 |
| Per-ounce Cost ($) | 0.019 | 0.019 | 0.006 | 0.003 | 0.028 |
| Profitability (%) | 27.640 | 29.180 | 12.499 | −98.550 | 58.620 |
| Shares (%) | 0.791 | 0.084 | 3.321 | 0.001 | 42.967 |

Note. (i) Data are the cross-section of week 391 (03/06/1997 to 03/12/1997) in Dominick's scanner data. (ii) Dominick's recorded the price and cost as zero if sales of a product were zero in a corresponding week. The mean and median of price and cost in Table 2-(b) were calculated including those zeros.

Figure 8.1: Histogram of the Observed Market Shares




Note. (i) This figure plots the histogram of the observed quantity market shares for cola sales of week 391 (03/06/1997 to 03/12/1997) in Dominick's scanner data. (ii) Sample points larger than 0.015 is top-coded as 0.015. 216 out of 4356 (4.96%) sample points are top-coded. (iii) 1130 out of 4356 samples (25.94%) have zero market shares.

and computed the size of the market using daily customer count data for each store in the chain.

## 8.2 Estimation, Result, and Discussion

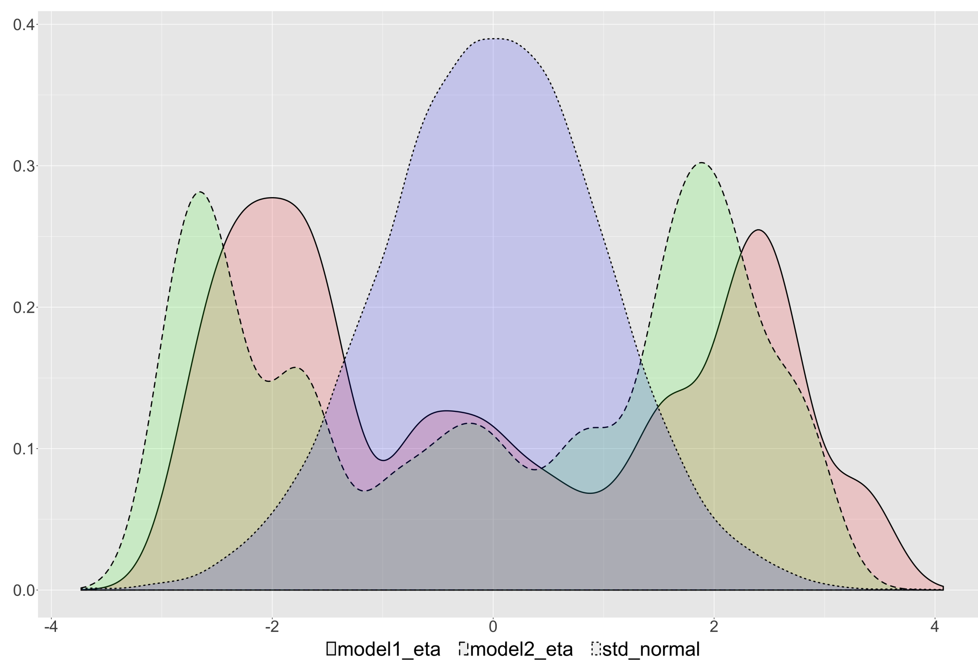
We estimate our model using the method proposed in Section 5. We also estimate the model correcting for a consumer's consideration set selection using the Probit as a first-stage estimator, with the Powell (2001) estimator and the simple Heckman selection correction estimator during the second stage. The simple Heckman estimator was implemented using the inverse Mills ratio as an additional regressor as usual. As a benchmark, we estimated the homogeneous logit model of demand, with different ways of handling the zero observed market shares: (i) dropping samples with zero observed market shares, and (ii) substituting zero observed market shares with small numbers. We also used the log of prices in the logit model to compare the magnitudes of coefficients. Mentioned previously, using the log of prices instead of raw prices represents a scale adjustment in the utility specification of the logit demand model.

We estimated two models with different specifications. In the baseline model (Model 1), $\mathbf{x}_{j,t}$ includes several product characteristics: bottle size, number of bottles per bundle, diet, caffeine-free, cherry flavor, Coke/Pepsi brand dummies. As an instrument of the per-ounce price, we used the per-ounce cost calculated from the profitability variable. For Model 1, we excluded promotion status from $\mathbf{x}_{j,t}$, and use it as a variable that satisfies the exclusion restriction. The exclusion

http://adage.com/article/news/consumers-drink-soft-drinks-water-beer/228422/.

assumption in this case reflects the informational hypothesis: promotions affect only consumers' information about a consideration set, not the level of utility associated with consuming a certain product. For Model 2, we included the promotion statuses in $\mathbf{x}_{j,t}$, and used store-level demographics for variables included in $\mathbf{w}_{j,t}$ that were not included in $\mathbf{x}_{j,t}$: % Blacks and Hispanics, % college graduates, and log of the median income. The exclusion assumption of these variables reflects the preferential hypothesis of the extensive margin: a consumer who never buys a certain product will not become an inframarginal consumer regardless of other product characteristics.

Figure 8.2: Estimated Densities of $\eta_{j,t}|\mathbf{w}_{j,t}$




Note. (i) "model1_eta," the pink solid density, is the estimated density of $\eta_{j,t}|\mathbf{w}_{j,t}$ from Model 1. "model2_eta," the green dotted density, is the estimated density of $\eta_{j,t}|\mathbf{w}_{j,t}$ from Model 2. "std_normal," the blue dotted density, is the standard normal density plotted for benchmark. (ii) The Klein-Spady model identifies the distribution of unobservables up to location and scale, and thus we made a location and scale adjustment of $E\left[\eta_{j,t}|\mathbf{w}_{j,t}\right] = 0$ and $Var\left(\eta_{j,t}|\mathbf{w}_{j,t}\right) = 4$. (iii) For Model 1 and 2, 10,000 sample draws were taken from the density estimates of the Klein-Spady model, and the density of the drawn sample was then plotted. (iv) The density of 10,000 sample draws from the standard Gaussian distribution is plotted for comparison.

Table 3: First-stage Parameter Estimates $\hat{\delta}$

| | Model 1 | | Model 2 | |
|---|---|---|---|---|
| | Probit | Klein-Spady | Probit | Klein-Spady |
| $\mathbf{w}_{j,t}$ | (1) | (2) | (3) | (4) |
| Bottle Size | 1 | 1 | 1 | 1 |
| | (0.195) | (−) | (0.196) | (−) |
| # Bottles per Bundle | −1.431 | −91.991 | −1.453 | −133.500 |
| | (0.329) | (3.780) | (0.330) | (12.535) |
| Diet | 18.342 | −277.300 | 18.222 | −515.187 |
| | (4.818) | (12.238) | (4.822) | (50.981) |
| Caffeine Free | −13.619 | 149.733 | −13.663 | 118.946 |
| | (6.158) | (10.108) | (6.171) | (4.446) |
| Cherry | −53.619 | 19.536 | −53.582 | −22.960 |
| | (13.821) | (4.700) | (13.805) | (4.620) |
| Coke | −41.949 | −20.901 | −41.915 | 91.351 |
| | (10.192) | (3.607) | (10.209) | (4.797) |
| Pepsi | 79.989 | −70.255 | 79.904 | −170.295 |
| | (5.917) | (2.492) | (5.910) | (8.323) |
| Promo | 135.449 | 396.440 | 135.534 | 601.361 |
| | (8.679) | (18.234) | (8.686) | (53.123) |
| % Blacks and Hispanics | - | - | −29.891 | −6.454 |
| | | | (21.430) | (13.069) |
| % College Graduates | - | - | 3.850 | 4.550 |
| | | | (20.978) | (12.624) |
| Log Median Income | - | - | −16.888 | 26.399 |
| | | | (93.441) | (43.820) |
| $D$ | 3226 | 3226 | 3226 | 3226 |
| $N$ | 4356 | 4356 | 4356 | 4356 |

Note. (i) $D$ is the number of non-zero market share observations, and $N$ is the sample size. (ii) Asymptotic standard error estimates appear in parentheses. (iii) The unit of bottle size is liquid ounces. (iv) We normalized the coefficients of the Bottle Size variable to one.

Table 4: Implied Own Price Elasticities and Second-stage Parameter Estimates $\left(\hat{\sigma}, \hat{\boldsymbol{\beta}}\right)$

| | Our Model, K/S | | Our Model, Probit | | Heckman Correction | | Logit Model | | |
|---|---|---|---|---|---|---|---|---|---|
| | Model 1 | Model 2 | Model 1 | Model 2 | Model 1 | Model 2 | Drop 0 | $10^{-8}$ | $10^{-4}$ |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
| Mean Price Elasticity (CES) | −1.326 | −1.369 | −1.138 | −1.139 | −1.299 | −1.290 | - | - | - |
| Mean Price Elasticity (Logit) | −1.324 | −1.367 | −1.133 | −1.135 | −1.297 | −1.287 | 4.777 | 7.703 | 4.726 |
| Log Price ($-\sigma$ / $-\alpha$) | −1.331 | −1.375 | −1.140 | −1.142 | −1.304 | −1.295 | 4.805 | 7.748 | 4.754 |
| | (0.041) | (0.040) | (0.058) | (0.061) | (0.040) | (0.041) | (0.099) | (0.158) | (0.074) |
| Bottle Size | 0.011 | 0.009 | 0.020 | 0.021 | 0.013 | 0.011 | 0.121 | 0.219 | 0.118 |
| | (0.002) | (0.001) | (0.010) | (0.019) | (0.001) | (0.002) | (0.004) | (0.007) | (0.003) |
| # Bottles per Bundle | 0.167 | 0.060 | 0.176 | 0.171 | 0.130 | 0.136 | 0.385 | 0.589 | 0.357 |
| | (0.010) | (0.025) | (0.019) | (0.010) | (0.003) | (0.006) | (0.012) | (0.020) | (0.009) |
| Diet | 0.361 | 0.207 | −0.350 | −0.316 | −0.186 | −0.219 | 0.943 | 2.241 | 0.940 |
| | (0.072) | (0.124) | (0.160) | (0.112) | (0.045) | (0.056) | (0.118) | (0.206) | (0.097) |
| Caffeine Free | −1.369 | −1.198 | −1.104 | −1.104 | −1.072 | −1.052 | −2.157 | −2.924 | −1.859 |
| | (0.058) | (0.054) | (0.178) | (0.116) | (0.056) | (0.063) | (0.131) | (0.229) | (0.108) |
| Cherry | −2.990 | −2.927 | −3.086 | −3.085 | −2.232 | −2.139 | −3.452 | −5.367 | −2.915 |
| | (0.123) | (0.132) | (0.835) | (1.071) | (0.176) | (0.185) | (0.343) | (0.535) | (0.252) |
| Coke | −0.375 | −0.181 | −0.828 | −0.729 | 0.063 | 0.071 | 0.593 | 1.151 | 0.472 |
| | (0.103) | (0.115) | (0.255) | (0.549) | (0.102) | (0.113) | (0.252) | (0.393) | (0.186) |
| Pepsi | 1.644 | 1.699 | 0.977 | 1.120 | 1.068 | 0.868 | 4.398 | 9.090 | 4.374 |
| | (0.069) | (0.066) | (0.191) | (0.962) | (0.062) | (0.191) | (0.172) | (0.269) | (0.127) |
| Promo | - | 0.671 | - | 0.168 | - | −0.260 | - | - | - |
| | | (0.130) | | (1.446) | | (0.233) | | | |
| $D$ | 3226 | 3226 | 3226 | 3226 | 3226 | 3226 | 3226 | 4090 | 4090 |
| $N$ | 4356 | 4356 | 4356 | 4356 | 4356 | 4356 | 3226 | 4090 | 4090 |

Note. (i) For columns "Our Model, K/S" and "Our Model, Probit," results from the Klein-Spady and Probit estimators in Table 3 were used for the first-stage estimator, respectively. Then, the pairwise differenced weighted instrumental variable estimator was used during the second stage. For the "Heckman Correction" columns, Probit was used during the first stage, and Heckman's selection correction estimator with the inverse Mills ratio as an additional regressor was used during the second stage. (ii) Row "Mean Price Elasticity (CES)" is the mean of the implied Marshallian own price elasticity over the sample for the CES demand system. Row "Mean Price Elasticity (Logit)" is the mean of the implied own price elasticity over the sample for the Logit demand system when the alternative utility is log-linear in prices. (iii) $D$ is the number of non-zero market share observations, and $N$ is the effective sample size. (iv) Asymptotic standard error estimates are in parentheses. (v) The unit of bottle size is liquid ounces. (vi) Because Dominick's did not record the price and cost when sales were zero, when estimating the $10^{-8}$ and $10^{-4}$ columns, we used average prices and costs of the same product with the same promotion statuses from other stores.

The first-stage parameter estimation result for $\hat{\delta}$ is shown in Table 3. Model 1 is the baseline model, with promotion statuses as excluded variables during second-stage estimation. Model 2 can be considered an additional robustness check, which uses the store-level demographics in the first stage. For Models 1 and 2, we estimated the Probit model for a benchmark, and for setting an initial value for the nonlinear optimizer to estimate the Klein-Spady model. The coefficient for bottle size was normalized to 1. We find that coefficient estimates from Probit estimation and Klein-Spady estimation are considerably different. We also plot the estimated conditional density of $\eta_{j,t}$ given $\mathbf{w}_{j,t}$ from each model in Figure 8.2. The estimated density of $\eta_{j,t}$ given $\mathbf{w}_{j,t}$ is not even unimodal, which is strong evidence that the unobservable product characteristic, $\eta_{j,t}$, does not follow a Gaussian distribution.

The primary estimation result is shown in Table 4. In the first two rows, we present the mean of the implied own price elasticities from the Marshallian CES and logit demand systems, respectively. In the logit demand models, coefficients of the log of prices were positive, and economically and statistically significant, even after instrumenting for prices using supplier side cost information. As a result, the own price elasticities are positive, meaning that an upward-sloping demand curve is estimated. In contrast, the log-linear estimation of our model with Klein-Spady first-stage estimator returned the expected signs and magnitudes for coefficients of the log of prices, and thus, the own price elasticities are negative and the estimated demand curves are downward-sloping. Estimators assuming a standard Gaussian distribution on unobservables performed well, despite estimated distributions of unobservables being far from Gaussian. We argue that such good performance of models assuming normality is due to the fact that estimated propensity scores from the Klein-Spady and Probit models correlate highly, with a correlation coefficient of about 0.7 for Models 1 and 2. Although this pattern was consistent in all robustness checks (Appendix C), we are unsure whether it can be generalized to a different dataset or market.[25]

Results provide strong evidence of a consideration set selection process that has been ignored in demand estimation literature. Ignoring the consideration set selection process of consumers biases the estimates, even resulting in an upward-sloping demand curve. Recall the estimation equation (5.2) under the exponential quality kernel:

$$E\left[\ln\left(\frac{\pi_{j,t}}{\pi_{0,t}}\right)|\mathbf{z}_{j,t},\mathbf{w}_{j,t},d_{j,t}=1\right]=-\sigma\phi_{j,t}+\mathbf{x}_{j,t}'\boldsymbol{\beta}+E\left[\xi_{j,t}|\mathbf{z}_{j,t},\mathbf{w}_{j,t},d_{j,t}=1\right].$$

Except for term $E\left[\xi_{j,t}|\mathbf{z}_{j,t},\mathbf{w}_{j,t},d_{j,t}=1\right]$, the estimation equation is the same as that of the logit demand model when we dropped samples with zero observed market shares. Columns (1) (Our Model, K/S, Model 1) and (7) (Logit Model, Drop 0) should coincide exactly when term $E\left[\xi_{j,t}|\mathbf{z}_{j,t},\mathbf{w}_{j,t},d_{j,t}=1\right]$

[25] Gandhi et al. (2013) also uses bath tissue data from Dominick's database, using a time series variation of a single chain store. Their estimated demand function is much more elastic than ours. For example, price coefficient estimates from simply dropping samples with zero market shares, which should be biased upward, remain negative. However, the implication they draw – that samples with zero observed market shares should not be simply dropped – is similar to ours.

is zero, yet this was not the case. $E\left[\xi_{j,t}|\mathbf{z}_{j,t}, \mathbf{w}_{j,t}, d_{j,t} = 1\right]$ is likely positive in our case because consumers select unobservables $\eta_{j,t}$ and observables $\mathbf{w}_{j,t}$, and $\eta_{j,t}$ correlates highly with $\xi_{j,t}$. Even after instrumenting for prices, price coefficient estimates are likely to be biased upward when samples with zero observed market shares are simply dropped. Imputing small numbers on zero observed market shares might cause a more serious problem – the direction of the bias cannot be predicted. In contrast to Table 1 in the previous section, Table 4 shows that imputing zero observed market shares with small positive numbers causes upward bias in price coefficient estimates. We cannot explain the direction of the bias when zeros are imputed. In Online Appendix C, we present estimation results for cola data from different weeks, and for laundry detergent data, with all results demonstrating the same pattern as that in Table 4, suggesting our findings are robust.

# 9 Conclusion

We develop a semiparametric demand estimation framework based on the Marshallian demand function derived from the budget-constrained CES utility maximization problem. Our framework is sufficiently flexible to incorporate observed and unobserved product characteristics, and is compatible with the widely used homogeneous and random coefficient logit models of demand. The framework accommodates zero predicted and observed market shares with a reasonable microfoundation by separating intensive and extensive margins, and embedding both margins in a quality kernel. We account for selection of a consumer's consideration set, which is unrecognized in the literature. If the consideration set selection stage is ignored, estimates of price coefficients can be misleading not only regarding their magnitudes, but also their signs. We demonstrate that ignoring consideration set selection can even result in upward-sloping demand curves. A direct extension of our study is a random coefficient demand estimation framework that can accommodate zero predicted and observed market shares. When a representative agent is assumed, the own and cross price elasticities derived from our model exhibited unrealistic substitution patterns, as in the homogeneous logit demand model of Berry (1994). Overcoming such unrealistic substitution patterns was one of the most important motivations for development of a random coefficient logit model of demand by Berry et al. (1995). Although we provide the microfoundation for a random coefficient CES demand estimation framework, we do not develop identification and estimation of model parameters with random coefficients that can accommodate zero market shares. We leave that extension to future research.

## References

AHN, H. AND J. L. POWELL (1993): "Semiparametric estimation of censored selection models with a nonparametric selection mechanism," *Journal of Econometrics*, 58, 3–29.

ALBUQUERQUE, P. AND B. J. BRONNENBERG (2009): "Estimating demand heterogeneity using aggregated data: An application to the frozen pizza category," *Marketing Science*, 28, 356–372.

ALLENBY, G. M., T. S. SHIVELY, S. YANG, AND M. J. GARRATT (2004): "A choice model for packaged goods: Dealing with discrete quantities and quantity discounts," *Marketing Science*, 23, 95–108.

ANDERSON, J. E. (1979): "A theoretical foundation for the gravity equation," *American Economic Review*, 69, 106–116.

ANDERSON, S. P., A. DE PALMA, AND J.-F. THISSE (1987): "The CES is a discrete choice model?" *Economics Letters*, 24, 139–140.

——— (1988): "A representative consumer theory of the logit model," *International Economic Review*, 29, 461–466.

——— (1992): *Discrete choice theory of product differentiation*, Cambridge, Massachusetts: The MIT Press.

BAJARI, P. AND C. L. BENKARD (2005): "Demand estimation with heterogeneous consumers and unobserved product characteristics: A hedonic approach," *Journal of Political Economy*, 113, 1239–1276.

BEN-AKIVA, M. AND B. BOCCARA (1995): "Discrete choice models with latent choice sets," *International Journal of Research in Marketing*, 12, 9–24.

BERRY, S. (1994): "Estimating discrete-choice models of product differentiation," *RAND Journal of Economics*, 25, 242–262.

BERRY, S., A. GANDHI, AND P. HAILE (2013): "Connected substitutes and invertibility of demand," *Econometrica*, 81, 2087–2111.

BERRY, S., J. LEVINSOHN, AND A. PAKES (1995): "Automobile prices in market equilibrium," *Econometrica*, 63, 841–890.

BERRY, S. AND A. PAKES (2007): "The pure characteristics demand model," *International Economic Review*, 48, 1193–1225.

BESANKO, D., S. GUPTA, AND D. JAIN (1998): "Logit demand estimation under competitive pricing behavior: An equilibrium framework," *Management Science*, 44, 1533–1547.

BHAT, C. R. (2008): "The multiple discrete-continuous extreme value (MDCEV) model: Role of utility function parameters, identification considerations, and model extensions," *Transportation Research Part B*, 42, 274–303.

BLUNDELL, R. AND J. L. POWELL (2007): "Censored regression quantiles with endogenous regressors," *Journal of Econometrics*, 141, 65–83.

BLUNDELL, R. W. AND J. L. POWELL (2003): *Endogeneity in nonparametric and semiparametric regression models*, Cambridge: Cambridge University Press, vol. 2 of *Advances in Economics and Econometrics: Theory and Applications, Eighth World Congress*, chap. 8, 312–357.

——— (2004): "Endogeneity in semiparametric binary response models," *Review of Economic Studies*, 71, 655–679.

BRUNO, H. A. AND N. J. VILCASSIM (2008): "Structural demand estimation with varying product availability," *Marketing Science*, 27, 1126–1131.

CHIANG, J. (1991): "A simultaneous approach to the whether, what and how much to buy questions," *Marketing Science*, 10, 297–315.

CHINTAGUNTA, P. K. (1993): "Investigating purchase incidence, brand choice and purchase quantity decisions of households," *Marketing Scinece*, 12, 184–208.

——— (2001): "Endogeneity and heterogeneity in a Probit demand model: Estimation using aggregate data," *Marketing Science*, 20, 442–456.

CHINTAGUNTA, P. K., A. BONFRER, AND I. SONG (2002): "Investigating the effects of store-brand introduction on retailer demand and pricing behavior," *Management Science*, 48, 1242–1267.

CHINTAGUNTA, P. K. AND R. DESIRAJU (2005): "Strategic pricing and detailing behavior in international markets," *Marketing Science*, 24, 67–80.

CHINTAGUNTA, P. K., J.-P. DUBÉ, AND K. Y. GOH (2005): "Beyond the endogeneity bias: The effect of unmeasured brand characteristics on household-level brand choice models," *Management Science*, 51, 832–849.

CHINTAGUNTA, P. K. AND H. S. NAIR (2011): "Discrete-choice models of consumer demand in marketing," *Marketing Science*, 30, 977–996.

CHRISTENSEN, L. R., D. W. JORGENSON, AND L. J. LAU (1975): "Transcendental logarithmic utility functions," *American Economic Review*, 65, 367–383.

DEATON, A. AND J. MUELLBAUER (1980): "An almost ideal demand system," *American Economic Review*, 70, 312–326.

DIXIT, A. K. AND J. E. STIGLITZ (1977): "Monopolistic competition and optimum product diversity," *American Economic Review*, 67, 297–308.

DRAGANSKA, M. AND D. C. JAIN (2006): "Consumer preferences and product-line pricing strategies: An empirical analysis," *Marketing Science*, 25, 164–174.

DUBÉ, J.-P. (2004): "Multiple discreteness and product differentiation: Demand for carbonated soft drinks," *Marketing Science*, 23, 66–81.

——— (2018): "Microeconometric models of consumer demand," *Working Paper*.

DUBE, J.-P., J. T. FOX, AND C.-L. SU (2012): "Improving the numerical performance of static and dynamic aggregate discrete choice random coefficients demand estimation," *Econometrica*, 80, 2231–2267.

DUBIN, J. A. AND D. MCFADDEN (1984): "An econometric analysis of residential electric appliance holdings and consumption," *Econometrica*, 52, 345–362.

EINAV, L., D. KNOEPFLE, J. LEVIN, AND N. SUNDARESAN (2014): "Sales taxes and internet commerce," *American Economic Review*, 104, 1–26.

GANDHI, A., Z. LU, AND X. SHI (2013): "Estimating demand for differentiated products with error in market shares," Working Paper.

GHOSE, A., P. G. IPEIROTIS, AND B. LI (2012): "Designing ranking systems for hotels on travel search engines by mining user-generated and crowdsourced content," *Marketing Science*, 31, 493–520.

GILBRIDE, T. J. AND G. M. ALLENBY (2004): "A choice model with conjunctive, disjunctive, and compensatory screening rules," *Marketing Science*, 23, 391–406.

GOLDFARB, A., Q. LU, AND S. MOORTHY (2009): "Measuring brand value in an equilibrium framework," *Marketing Science*, 28, 69–86.

GUADAGNI, P. M. AND J. D. C. LITTLE (1983): "A logit model of brand choice calibrated on scanner data," *Marketing Science*, 2, 203–238.

HANEMANN, W. M. (1984): "Discrete/continuous models of consumer demand," *Econometrica*, 52, 541–561.

HECKMAN, J. J. (1979): "Sample selection bias as a specification error," *Econometrica*, 47, 153–62.

HELPMAN, E., M. MELITZ, AND Y. RUBINSTEIN (2008): "Estimating trade flows: Trading partners and trading volumes," *Quarterly Journal of Eonomics*, 123, 441–487.

HITSCH, G. J. (2006): "An empirical model of optimal dynamic product launch and exit under demand uncertainty," *Marketing Science*, 25, 25–50.

HONORÉ, B. E., E. KYRIAZIDOU, AND C. UDRY (1997): "Estimation of Type 3 Tobit models using symmetric trimming and pairwise comparisons," *Journal of Econometrics*, 76, 107–128.

HOWELL, J. R., S. LEE, AND G. M. ALLENBY (2016): "Price promotions in choice models," *Marketing Science*, 35, 319–334.

JEDIDI, K., R. KOHLI, AND W. S. DESARBO (1996): "Consideration sets in conjoint analysis," *Journal of Marketing Research*, 33, 364–372.

KIM, J., G. M. ALLENBY, AND P. E. ROSSI (2002): "Modeling consumer demand for variety," *Marketing Science*, 21, 229–250.

——— (2007): "Product attributes and models of multiple discreteness," *Journal of Econometrics*, 138, 208–230.

KLEIN, R. W. AND R. H. SPADY (1993): "An efficient semiparametric estimator for binary response models," *Econometrica*, 61, 387–421.

KRUGMAN, P. (1980): "Scale economies, product differentiation, and the pattern of trade," *American Economic Review*, 70, 950–959.

LANCASTER, K. J. (1966): "A new approach to consumer theory," *Journal of Political Economy*, 74, 132–157.

LEE, S. AND G. M. ALLENBY (2014): "Modeling indivisible demand," *Marketing Science*, 33, 364–381.

LEE, S., J. KIM, AND G. M. ALLENBY (2013): "A direct utility model for asymmetric complements," *Marketing Science*, 32, 454–470.

MCFADDEN, D. (1974): *Conditional logit analysis of qualitative choice behavior*, Academic Press, chap. 4, 105–142.

——— (1978): *Modelling the choice of residential location*, North-Holland Publishing Company, vol. 3 of *Spatial interaction theory and planning models*, chap. 3, 75–96.

——— (1981): *Econometric models of probabilistic choice*, The MIT Press, chap. 5, 198–272, Structural analysis of discrete data with econometric applications.

MEHTA, N. (2007): "Investigating consumers' purchase incidence and brand choice decisions across multiple product categories: A theoretical and empirical analysis," *Marketing Science*, 26, 196–217.

MEHTA, N., S. RAJIV, AND K. SRINIVASAN (2003): "Price uncertainty and consumer search: A structural model of consideration set formation," *Marketing Science*, 22, 58–84.

NAIR, H., J.-P. DUBÉ, AND P. CHINTAGUNTA (2005): "Accounting for primary and secondary demand effects with aggregate data," *Marketing Science*, 24, 444–460.

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

NEWEY, W. K. (2009): "Two-step series estimation of sample selection models," *Econometrics Journal*, 12, S217–S229.

NEWEY, W. K. AND D. MCFADDEN (1994): *Large sample estimation and hypothesis testing*, Amsterdam: Elsevier, vol. 4 of *Handbook of Econometrics*, chap. 36, 2111–2245.

PETRIN, A. (2002): "Quantifying the benefits of new products: The case of the minivan," *Journal of Political Economy*, 110, 705–729.

POWELL, J. L. (1984): "Least absolute deviations estimation for the censored regression model," *Journal of Econometrics*, 25, 303–325.

——— (1986): "Censored regression quantiles," *Journal of Econometrics*, 32, 143–155.

——— (2001): *Semiparametric estimation of censored selection models*, Cambridge: Cambridge University Press, chap. 6, 165–196.

ROBERTS, J. H. AND J. M. LATTIN (1991): "Development and testing of a model of consideration set composition," *Journal of Marketing Research*, 28, 429–440.

ROTHE, C. (2009): "Semiparametric estimation of binary response models with endogenous regressors," *Journal of Econometrics*, 153, 51–64.

SONG, I. AND P. K. CHINTAGUNTA (2007): "A discrete-continuous model for multicategory purchase behavior of households," *Journal of Marketing Research*, 44, 595–612.

SPENCE, M. (1976): "Product selection, fixed costs, and monopolistic competition," *Review of Economic Studies*, 43, 217–235.

SUDHIR, K. (2001): "Competitive pricing behavior in the auto market: A structural analysis," *Marketing Science*, 20, 42–60.

VILLAS-BOAS, J. M. AND R. S. WINER (1999): "Endogeneity in brand choice models," *Management Science*, 45, 1324–1338.

WILBUR, K. C. (2008): "A two-sided, empirical model of television advertising and viewing markets," *Marketing Science*, 27, 356–378.

# A Derivation of the Logit Demand System

We illustrate the derivation of a homogeneous and random coefficient logit demand systems for completeness. The illustration in this section largely follows the original presentation of Berry (1994); Berry et al. (1995).

Let $j \in \mathcal{J}_t$, where $\mathcal{J}_t$ is a finite set of alternatives that must contain the numeraire. Individual $i$ in market $t$ solves the following discrete choice utility maximization problem:

$$\max_{j \in \mathcal{J}_t} \left\{ u_{i,j,t} \right\},$$

where the (indirect) utility of choosing alternative $j$ in market $t$ is:

$$u_{i,j,t} = \alpha_i \left( y_i - p_{j,t} \right) + \mathbf{x}'_{j,t} \boldsymbol{\beta}_i + \xi_{j,t} + \epsilon_{i,j,t}. \tag{A.1}$$

$\epsilon_{i,j,t}$ follows the i.i.d. Type-I extreme value distribution. Note that it is also legitimate to specify the utility as:

$$u_{i,j,t} = -\alpha_i \ln p_{j,t} + \mathbf{x}'_{j,t} \boldsymbol{\beta}_i + \xi_{j,t} + \epsilon_{i,j,t},$$

given that we stick to the direct utility interpretation of $u_{i,j,t}$ as individual $i$ choosing alternative $j$ in market $t$. The logarithm can be regarded as a scale adjustment on the level of disutility from prices.

The coefficients $(\alpha_i, \boldsymbol{\beta}_i)$ might vary over individuals, and are specified as:

$$\begin{aligned} \alpha_i &:= \alpha + \Pi_\alpha \mathbf{q}_i + \Sigma_\alpha v_{\alpha,i} \\ \boldsymbol{\beta}_i &:= \boldsymbol{\beta} + \boldsymbol{\Pi}_\beta \mathbf{q}_i + \boldsymbol{\Sigma}_\beta \mathbf{v}_{\beta,i}, \end{aligned}$$

where $\mathbf{q}_i$ is the demographic variable, $\mathbf{v}_i$ is the vector of a unit normal shock, $\left(\Pi_\alpha, \boldsymbol{\Pi}_\beta\right)$ is the correlation component between demographic variables and the corresponding coefficients, and $\left(\Sigma_\alpha, \boldsymbol{\Sigma}_\beta\right)$ represents the covariance structure of the shocks on the coefficients. The linear utility specification (A.1) becomes:

$$\begin{aligned} u_{i,j,t} &= \alpha_i \left( y_i - p_{j,t} \right) + \mathbf{x}'_{j,t} \boldsymbol{\beta}_i + \xi_{j,t} + \epsilon_{i,j,t} \\ &= \alpha_i y_i - \left( \alpha + \Pi_\alpha \mathbf{q}_i + \Sigma_\alpha v_{\alpha,i} \right) p_{j,t} + \mathbf{x}_{j,t} \left( \boldsymbol{\beta} + \boldsymbol{\Pi}_\beta \mathbf{q}_i + \boldsymbol{\Sigma}_\beta \mathbf{v}_{\beta,i} \right) + \xi_{j,t} + \epsilon_{i,j,t} \\ &= \alpha_i y_i + \left( -\alpha p_{j,t} + \mathbf{x}'_{j,t} \boldsymbol{\beta} + \xi_{j,t} \right) - \left( \Pi_\alpha \mathbf{q}_i + \Sigma_\alpha v_{\alpha,i} \right) p_{j,t} + \mathbf{x}'_{j,t} \left( \boldsymbol{\Pi}_\beta \mathbf{q}_i + \boldsymbol{\Sigma}_\beta \mathbf{v}_{\beta,i} \right) + \epsilon_{i,j,t} \\ &= \alpha_i y_i + \left( -\alpha p_{j,t} + \mathbf{x}'_{j,t} \boldsymbol{\beta} + \xi_{j,t} \right) + \begin{pmatrix} -p_{j,t} & \mathbf{x}'_{j,t} \end{pmatrix} \left( \boldsymbol{\Pi} \mathbf{q}_i + \boldsymbol{\Sigma} \mathbf{v}_i \right) + \epsilon_{i,j,t} \\ &=: \alpha_i y_i + \delta_{j,t} + \mu_{i,j,t} + \epsilon_{i,j,t}, \end{aligned}$$

where $\delta_{j,t}$ is the mean utility of alternative $j$ that is common to every individual in market $t$, and $\mu_{i,j,t}$ is the individual specific structural utility component. For the log-linear specification, one can simply replace the term $p_{j,t}$ with $\ln p_{j,t}$.

Given the assumption that $\epsilon_{i,j,t}$ follows i.i.d. Type-I extreme value distribution, the individual choice probability $\Pr(i \to j|t)$ becomes:

$$\Pr(i \to j|t) = \frac{\exp\left(\delta_{j,t} + \mu_{i,j,t}\right)}{\sum_{k \in \mathcal{J}_t} \exp\left(\delta_{k,t} + \mu_{i,k,t}\right)}.$$

This individual choice probability is taken as the individual predicted quantity share $\pi_{i,j,t}$. Given distributions of the demographics $F(\mathbf{z}_i)$ and of shocks on the preference parameter $F(\mathbf{v}_i)$, the predicted quantity market share of good $j$ is aggregated as:

$$\begin{aligned} \pi_{j,t} &= \int\int \pi_{i,j,t} dF(\mathbf{z}_i)\, dF(\mathbf{v}_i) \\ &= \int\int \frac{\exp\left(\delta_{j,t} + \mu_{i,j,t}\right)}{\sum_{k \in \mathcal{J}_t} \exp\left(\delta_{k,t} + \mu_{i,k,t}\right)} dF(\mathbf{z}_i)\, dF(\mathbf{v}_i). \end{aligned} \tag{A.2}$$

If $\alpha_i = \alpha$ and $\boldsymbol{\beta}_i = \boldsymbol{\beta}$, which implies that the preference is homogeneous across individuals, the model reduces to the homogeneous logit demand model.

By definition, the predicted market share $\pi_{j,t}$ is:

$$\pi_{j,t} := \frac{q_{j,t}}{\sum_{k \in \mathcal{J}_t} q_{k,t}}.$$

This system of predicted quantity market shares for $\#(\mathcal{J}_t)$ alternatives in a market $t$ provides only $\#(\mathcal{J}_t) - 1$ restrictions on the system of quantity demand $\mathbf{q}_t$. An additional restriction is required, and Berry (1994); Berry et al. (1995) impose a fixed market size assumption to derive the quantity demand; denominator $\sum_{k \in \mathcal{J}_t} q_{k,t}$ is regarded as fixed at some level $M$.

# B Implementation Details

We use the Gaussian kernel during first- and second-stage estimation. For tractability, higher-order kernels are not used. The bandwidth, $h_n$, for the Klein-Spady estimator is $h_n = \text{std}\left(\mathbf{w}_j' \hat{\delta}_{\text{Probit}}\right) C_1 n^{-\frac{1}{7}}$. The rate $n^{-\frac{1}{7}}$ follows the original suggestion from Klein and Spady (1993). Bandwidth for the second-stage Powell (2001) estimator is $h_n = \text{std}\left(\mathbf{w}_j' \hat{\delta}_{\text{KS}}\right) C_2 n^{-\frac{1}{7}}$, where $\hat{\delta}_{\text{KS}}$ is the Klein-Spady estimator from the first stage. We use the tuning parameter $C_1 = C_2 = 1$ in Section 7 and $C_1 = C_2 = 0.5$ in Section 8.

We tried 100 randomly generated starting values in the first stage Klein-Spady estimation to guard against the argument that the optimization routine stopped at the local minima. The randomly generated initial values follow the distribution $\mathcal{N}\left(\hat{\delta}_{\text{Probit}}, \frac{1}{5}\text{diag}\left(\sqrt{\left|\hat{\delta}_{\text{Probit}}\right|}\right)\right)$. We also tried several tuning parameters to assess robustness. First-stage parameter estimates varied considerably regarding the choice of tuning parameters and bandwidth, whereas second-stage parameter estimates, which are our primary interest, were robust to the choice of bandwidth and initial values for nonlinear optimization.

For the simple Heckman estimator with endogeneity, we computed standard errors that account for the fact that the inverse Mills ratio is a generated regressor. Details on the covariance formula of the estimator can be found in Newey and McFadden (1994). Because finding the inverse Mills ratio is fast, one can also consider the bootstrapped standard errors for the Heckman estimator. Finally, we tried both IPOPT and KNITRO, which are state-of-the-art, derivatives-based, nonlinear optimizers for nonlinear optimization. Results were robust to choice of optimizer.

# C For Online Publication: Robustness Checks

## C.1 Cola Demand for Week 382

We estimate the same models as in Section 8 using cola data from a different week. We use data from week 382 (January 1 through 9, 1997). In Tables 5 and 6, we repeat the estimation procedure from Tables 3 and 4.

Table 5: First-stage Parameter Estimates $\hat{\delta}$

| | Model 1 | | Model 2 | |
|---|---|---|---|---|
| | Probit | Klein-Spady | Probit | Klein-Spady |
| $\mathbf{w}_{j,t}$ | (1) | (2) | (3) | (4) |
| Bottle Size | 1 | 1 | 1 | 1 |
| | (0.188) | (−) | (0.193) | (−) |
| # Bottles per Bundle | −5.867 | −46.247 | −5.825 | −157.181 |
| | (0.429) | (12.736) | (0.426) | (23.221) |
| Diet | 14.892 | 7.834 | 14.674 | 5.802 |
| | (5.016) | (3.170) | (4.986) | (5.447) |
| Caffeine Free | −22.949 | 43.032 | −22.501 | −335.901 |
| | (5.971) | (4.613) | (5.938) | (44.885) |
| Cherry | −68.882 | −231.403 | −69.030 | −188.603 |
| | (11.761) | (71.908) | (11.635) | (23.457) |
| Coke | 70.670 | 122.727 | 70.572 | 262.402 |
| | (9.636) | (17.578) | (9.577) | (33.91) |
| Pepsi | 48.249 | −150.090 | 47.706 | 377.354 |
| | (5.888) | (56.156) | (5.839) | (45.846) |
| Promo | 126.413 | 866.065 | 124.981 | 1007.944 |
| | (7.699) | (229.737) | (7.633) | (139.527) |
| % Blacks and Hispanics | - | - | 24.100 | 30.682 |
| | | | (22.881) | (23.885) |
| % College Graduates | - | - | −45.023 | −66.658 |
| | | | (22.069) | (21.641) |
| Log Median Income | - | - | −284.144 | −226.55 |
| | | | (100.267) | (100.23) |
| $D$ | 3226 | 3226 | 3226 | 3226 |
| $N$ | 4337 | 4337 | 4337 | 4337 |

Note. (i) $D$ is the number of non-zero market share observations, and $N$ is the sample size. (ii) The asymptotic standard error estimates appear in parentheses. (iii) The unit of bottle size is liquid ounces. (iv) We normalized the coefficients of the Bottle Size variable to one.

Table 6: Second-stage Parameter Estimates $\left(\hat{\sigma}, \hat{\boldsymbol{\beta}}\right)$

| | Our Model, K/S | | Our Model, Probit | | Heckman Correction | | Logit Model | | |
|---|---|---|---|---|---|---|---|---|---|
| | Model 1 | Model 2 | Model 1 | Model 2 | Model 1 | Model 2 | Drop 0 | $10^{-8}$ | $10^{-4}$ |
| $\left(\phi_{j,t}, \mathbf{x}_{j,t}\right)$ | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
| Log Price $(-\sigma)$ | −1.175 | −1.254 | −1.266 | −1.318 | −1.272 | −1.339 | 4.367 | 5.348 | 4.055 |
| | (0.039) | (0.035) | (0.045) | (0.038) | (0.041) | (0.038) | (0.132) | (0.168) | (0.085) |
| Bottle Size | 0.010 | 0.009 | −0.003 | −0.002 | −0.002 | 0.002 | 0.107 | 0.136 | 0.095 |
| | (0.004) | (0.002) | (0.006) | (0.003) | (0.001) | (0.001) | (0.005) | (0.007) | (0.004) |
| # Bottles per Bundle | 0.053 | −0.001 | 0.046 | 0.011 | 0.036 | −0.018 | 0.246 | 0.192 | 0.193 |
| | (0.003) | (0.009) | (0.004) | (0.014) | (0.003) | (0.004) | (0.015) | (0.019) | (0.010) |
| Diet | −0.114 | −0.127 | −0.113 | −0.119 | −0.275 | −0.155 | 1.132 | 1.360 | 0.891 |
| | (0.036) | (0.032) | (0.087) | (0.042) | (0.048) | (0.035) | (0.136) | (0.197) | (0.099) |
| Caffeine Free | −1.108 | −1.215 | −1.148 | −1.111 | −0.851 | −1.146 | −1.699 | −2.740 | −1.605 |
| | (0.053) | (0.041) | (0.114) | (0.054) | (0.053) | (0.040) | (0.135) | (0.204) | (0.103) |
| Cherry | −2.222 | −1.889 | −1.772 | −1.749 | −0.725 | −1.696 | −3.153 | −6.635 | −3.031 |
| | (0.272) | (0.105) | (0.324) | (0.132) | (0.218) | (0.120) | (0.410) | (0.485) | (0.245) |
| Coke | 1.670 | 1.461 | 1.415 | 1.271 | 0.445 | 1.178 | 3.270 | 5.801 | 2.888 |
| | (0.136) | (0.113) | (0.392) | (0.135) | (0.123) | (0.098) | (0.259) | (0.361) | (0.182) |
| Pepsi | 0.844 | 0.942 | 1.045 | 0.917 | 0.289 | 0.798 | 3.476 | 5.612 | 3.030 |
| | (0.126) | (0.062) | (0.237) | (0.079) | (0.069) | (0.052) | (0.202) | (0.264) | (0.134) |
| Promo | - | 1.510 | - | 0.645 | - | 1.256 | - | - | - |
| | | (0.063) | | (0.257) | | (0.066) | | | |
| $D$ | 3226 | 3226 | 3226 | 3226 | 3226 | 3226 | 3226 | 4118 | 4118 |
| $N$ | 4337 | 4337 | 4337 | 4337 | 4337 | 4337 | 3226 | 4118 | 4118 |

Note. (i) For columns Our Model, K/S and Our Model, Probit, results from the Klein-Spady and Probit estimators in Table 5 were used for the first-stage estimator, respectively. A pairwise differenced weighted instrumental variable estimator was then used during the second stage. For the Heckman Correction column, the Probit was used during the first stage, and the Heckman's selection correction estimator with the inverse Mills ratio as an additional regressor was used during the second stage. (ii) $D$ is the number of non-zero market share observations, and $N$ is the effective sample size. (iii) Asymptotic standard error estimates are in parentheses. (iv) The unit of bottle size is liquid ounces. (v) Because Dominick's did not record the price and cost when sales were zero, when estimating the $10^{-8}$ and $10^{-4}$ columns, we used average prices and costs of the same product with the same promotion statuses from other stores.

## C.2 Cola Demand for Week 278

We repeat estimation of cola data from week 278 (January 5 through 11, 1995). All results demonstrated the same pattern as in previous sections. Although not tabulated here, we also examined data from many other weeks, and results were robust.

Table 7: First-stage Parameter Estimates $\hat{\delta}$

| | Model 1 | | Model 2 | |
|---|---|---|---|---|
| | Probit | Klein-Spady | Probit | Klein-Spady |
| $\mathbf{w}_{j,t}$ | (1) | (2) | (3) | (4) |
| Bottle Size | 1 | 1 | 1 | 1 |
| | (0.170) | (−) | (0.169) | (−) |
| # Bottles per Bundle | −5.334 | −166.569 | −5.289 | −68.799 |
| | (0.489) | (140.200) | (0.487) | (5.218) |
| Diet | −76.147 | −1520.618 | −75.673 | −207.218 |
| | (6.761) | (1289.560) | (6.774) | (15.787) |
| Caffeine Free | 74.187 | 1565.247 | 75.038 | 252.376 |
| | (8.674) | (1292.587) | (8.743) | (18.212) |
| Cherry | −223.710 | −255.074 | −224.050 | −1406.267 |
| | (108.580) | (12145.276) | (107.821) | (106.027) |
| Coke | 151.949 | 83.965 | 152.293 | 1460.142 |
| | (17.472) | (6.953) | (17.382) | (106.708) |
| Pepsi | 18.408 | −751.111 | 18.103 | 92.893 |
| | (7.076) | (643.282) | (7.046) | (5.724) |
| Promo | 305.656 | 4718.847 | 303.695 | 2118.076 |
| | (7.716) | (3869.202) | (7.697) | (156.271) |
| % Blacks and Hispanics | - | - | 1.897 | 20.684 |
| | | | (25.054) | (22.897) |
| % College Graduates | - | - | −60.347 | −44.537 |
| | | | (27.667) | (21.267) |
| Log Median Income | - | - | −213.245 | −103.183 |
| | | | (112.231) | (92.389) |
| $D$ | 3667 | 3667 | 3667 | 3667 |
| $N$ | 5185 | 5185 | 5185 | 5185 |

Note. (i) $D$ is the number of non-zero market shares observations, and $N$ is the sample size. (ii) Asymptotic standard error estimates appear in parentheses. (iii) The unit of bottle size is liquid ounces. (iv) We normalized the coefficients of the Bottle Size variable to one.

Table 8: Second-stage Parameter Estimates $\left(\hat{\sigma}, \hat{\boldsymbol{\beta}}\right)$

| | Our Model, K/S | | Our Model, Probit | | Heckman Correction | | Logit Model | | |
|---|---|---|---|---|---|---|---|---|---|
| | Model 1 | Model 2 | Model 1 | Model 2 | Model 1 | Model 2 | Drop 0 | $10^{-8}$ | $10^{-4}$ |
| $\left(\phi_{j,t}, \mathbf{x}_{j,t}\right)$ | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
| Log Price ($-\sigma$) | −0.918 | −1.225 | −0.261 | −0.261 | −0.969 | −0.991 | 18.789 | 20.461 | 16.460 |
| | (0.176) | (0.068) | (0.110) | (0.117) | (0.040) | (0.043) | (3.213) | (2.736) | (2.174) |
| Bottle Size | 0.008 | 0.000 | 0.028 | 0.017 | 0.013 | 0.012 | 0.742 | 0.817 | 0.654 |
| | (0.001) | (0.002) | (0.003) | (0.005) | (0.001) | (0.001) | (0.137) | (0.120) | (0.095) |
| # Bottles per Bundle | 0.047 | −0.018 | 0.097 | 0.153 | 0.054 | 0.057 | 2.169 | 2.310 | 1.871 |
| | (0.006) | (0.019) | (0.012) | (0.026) | (0.003) | (0.003) | (0.401) | (0.345) | (0.274) |
| Diet | −0.213 | 0.088 | 0.007 | 0.826 | −0.127 | −0.070 | 9.715 | 10.045 | 7.090 |
| | (0.094) | (0.064) | (0.124) | (0.404) | (0.037) | (0.042) | (2.054) | (1.502) | (1.193) |
| Caffeine Free | −1.437 | −1.536 | −1.554 | −2.364 | −1.281 | −1.319 | −2.958 | −3.842 | −3.912 |
| | (0.106) | (0.070) | (0.108) | (0.417) | (0.043) | (0.045) | (0.597) | (0.714) | (0.567) |
| Cherry | −2.820 | −4.157 | −2.672 | −0.268 | −3.018 | −2.892 | −0.743 | −0.779 | −1.718 |
| | (18.229) | (0.472) | (2.073) | (3.573) | (0.312) | (0.351) | (5.991) | (6.469) | (5.142) |
| Coke | 1.802 | 2.631 | 1.833 | 0.252 | 1.684 | 1.633 | 4.432 | 6.665 | 4.610 |
| | (0.509) | (0.377) | (0.186) | (0.600) | (0.106) | (0.109) | (1.233) | (1.249) | (0.993) |
| Pepsi | 1.366 | 1.456 | 1.446 | 1.275 | 1.464 | 1.471 | 12.259 | 15.105 | 12.212 |
| | (0.081) | (0.054) | (0.082) | (0.130) | (0.044) | (0.044) | (2.377) | (2.376) | (1.889) |
| Promo | - | 0.294 | - | −3.216 | - | −0.331 | - | - | - |
| | | (0.466) | | (1.264) | | (0.128) | | | |
| $D$ | 3667 | 3667 | 3667 | 3667 | 3667 | 3667 | 3667 | 4069 | 4069 |
| $N$ | 5185 | 5185 | 5185 | 5185 | 5185 | 5185 | 3667 | 4069 | 4069 |

Note. (i) For columns Our Model, K/S and Our Model, Probit, results from the Klein-Spady and Probit estimators in Table 7 were used during the first-stage estimator, respectively. A pairwise differenced weighted instrumental variable estimator was then used during the second stage. For the Heckman Correction column, the Probit was used during the first stage, and the Heckman's selection correction estimator with the inverse Mills ratio as an additional regressor was used during the second stage. (ii) $D$ is the number of non-zero market share observations, and $N$ is the effective sample size. (iii) Asymptotic standard error estimates appear in parentheses. (iv) The unit of bottle size is liquid ounces. (v) Because Dominick's did not record the price and cost when sales were zero, when estimating the $10^{-8}$ and $10^{-4}$ columns, we used average prices and costs of the same product with the same promotion statuses from other stores.

### C.3 Laundry Detergent Demand for Week 375

We estimate demand for laundry detergent using the same data from Dominick's as in Section 8. We chose a cross-section of week 375 randomly, which is the third week of November 1996. The product was defined by its UPC, and the market was defined as the week-store pair. We compute market shares by loads. There are two types of laundry detergents–liquid and powder. We convert the size of a canister by the following criteria. For liquid laundry detergent, 1.6 ounces were counted as one load. For powder laundry detergent, 2.3 ounces were counted as one load. Some powder detergents used pounds instead of ounces as a unit of package size.[26] For such products, 0.08262 pounds was counted as one load. As the density of powder detergents are approximately $0.65 \mathrm{g}/cm^3$ and $1\mathrm{g}/cm^3 = 0.065198 \mathrm{lb}/\mathrm{oz}$, one pound of powdered detergent is approximately 23.6 ounces. Market size was calculated assuming that each consumer who visited the store consumed 6 loads of laundry detergent each week. Other details on the data were similar to what we describe in Section 8.

Table 9 shows first-stage estimates, and Table 10 shows those for the second stage. We find the same pattern as in Table 4 in Section 8; even after instrumenting for prices, the estimated demand curve was upward-sloping when consideration set selection was not considered during estimation.

[26] Arm & Hammer powdered detergent.

Table 9: First-stage Parameter Estimates $\hat{\delta}$

| $\mathbf{w}_{j,t}$ | Probit (1) | Klein-Spady (2) |
|---|---|---|
| Package Size | 1 | 1 |
| | (0.091) | $(-)$ |
| Liquid | 379.843 | 320.578 |
| | (16.031) | (5.676) |
| Heavy duty / Concentrated / Double | 114.862 | $-18.295$ |
| | (42.456) | (5.571) |
| Bleach | $-6.144$ | $-0.279$ |
| | (18.997) | (2.714) |
| Tide | 230.570 | 266.716 |
| | (20.483) | (5.676) |
| Wisk | 31.895 | $-94.849$ |
| | (27.169) | (4.451) |
| Ajax / Arm&Hammer / Surf / Purex | $-240.998$ | $-152.440$ |
| | (22.464) | (4.440) |
| % Blacks and Hispanics | 18.464 | $-4.242$ |
| | (57.478) | (8.147) |
| % College Graduates | 242.359 | 1.831 |
| | (91.294) | (13.016) |
| Log Median Income | $-9.221$ | $-2.217$ |
| | (48.458) | (6.950) |
| $D$ | 7177 | 7177 |
| $N$ | 14999 | 14999 |

Note. (i) $D$ is the number of non-zero market share observations, and $N$ is the sample size. (ii) Asymptotic standard error estimates appear in parentheses. (iii) The unit of bottle size is liquid ounces. (iv) We normalized the coefficients of Package Size variable to one.

Table 10: Second-stage Parameter Estimates $\left(\hat{\sigma}, \hat{\boldsymbol{\beta}}\right)$

| | | Our Model | | | | Logit Model | | |
|---|---|---|---|---|---|---|---|---|
| | first-stage: | K/S | Probit | Heckman | Zero: | Drop 0 | $10^{-8}$ | $10^{-4}$ |
| $\left(\phi_{j,t}, \mathbf{x}_{j,t}\right)$ | | (1) | (2) | (3) | | (4) | (5) | (6) |
| Log Price ($-\sigma$) | | −0.991 | −0.823 | −1.059 | | 6.455 | 9.755 | 6.371 |
| | | (0.069) | (0.194) | (0.044) | | (0.095) | (0.138) | (0.064) |
| Package Size | | 0.009 | 0.004 | 0.000 | | 0.030 | 0.038 | 0.026 |
| | | (0.003) | (0.004) | (0.001) | | (0.001) | (0.001) | (0.001) |
| Liquid | | 3.077 | 1.585 | −0.032 | | 6.551 | 9.990 | 6.020 |
| | | (0.985) | (1.538) | (0.296) | | (0.153) | (0.228) | (0.106) |
| Heavy duty / | | −0.397 | 0.318 | −0.435 | | 4.218 | 7.803 | 4.108 |
| Concentrated / Double | | (0.104) | (0.843) | (0.143) | | (0.249) | (0.467) | (0.217) |
| Bleach | | −0.016 | −0.016 | −0.035 | | 0.575 | 0.837 | 0.720 |
| | | (0.039) | (0.053) | (0.054) | | (0.101) | (0.172) | (0.080) |
| Tide | | 3.510 | 1.503 | 0.826 | | 0.796 | 2.055 | 0.578 |
| | | (0.872) | (0.793) | (0.183) | | (0.099) | (0.175) | (0.081) |
| Wisk | | −0.491 | 0.038 | −0.033 | | −0.613 | −1.118 | −0.742 |
| | | (0.257) | (0.103) | (0.071) | | (0.137) | (0.229) | (0.107) |
| Ajax / ArmHammer / | | −0.722 | −0.490 | 0.323 | | 1.715 | 3.180 | 2.034 |
| Surf / Purex | | (0.437) | (0.806) | (0.199) | | (0.133) | (0.226) | (0.105) |
| $D$ | | 7177 | 7177 | 7177 | | 7177 | 10639 | 10639 |
| $N$ | | 14999 | 14999 | 14999 | | 7177 | 10639 | 10639 |

Note. (i) For columns Our Model, K/S and Our Model, Probit, results from the Klein-Spady and Probit estimators in Table 9 were used for the first-stage estimator, respectively. A pairwise differenced weighted instrumental variable estimator was then used during the second stage. For the Heckman Correction column, the Probit was used during the first stage, and the Heckman's selection correction estimator with the inverse Mills ratio as an additional regressor was used during the second stage. (ii) $D$ is the number of non-zero market share observations, and $N$ is the effective sample size. (iii) Asymptotic standard error estimates appear in parentheses. (iv) The unit of package size was converted to liquid ounces. (v) Because Dominick's did not record the price and cost when sales were zero, when estimating the $10^{-8}$ and $10^{-4}$ columns, we used average prices and costs of the same product with the same promotion statuses from other stores.

## C.4 Laundry Detergent Demand for Week 398

We estimate demand for laundry detergent using the same data from Dominick's as in the previous subsection. We selected a cross-section of week 398, which is the last week of April 1997. Other details on data handling were the same as in the previous subsection.

Table 11: First-stage Parameter Estimates $\hat{\delta}$

| | Probit | Klein-Spady |
|---|---|---|
| $\mathbf{w}_{j,t}$ | (1) | (2) |
| Package Size | 1 | 1 |
| | (0.101) | (−) |
| Liquid | 345.336 | 449.144 |
| | (17.758) | (6.124) |
| Heavy duty / Concentrated / Double | 111.494 | 123.172 |
| | (48.905) | (5.447) |
| Bleach | −1.811 | 6.195 |
| | (20.737) | (2.008) |
| Tide | 341.478 | 387.581 |
| | (23.082) | (7.019) |
| Wisk | 191.418 | 242.408 |
| | (31.135) | (7.153) |
| Ajax / Arm&Hammer / Surf / Purex | −94.144 | −143.853 |
| | (24.119) | (3.401) |
| % Blacks and Hispanics | 101.921 | −3.192 |
| | (63.318) | (6.717) |
| % College Graduates | 141.812 | −3.712 |
| | (101.556) | (11.025) |
| Log Median Income | 21.230 | −0.033 |
| | (52.890) | (5.735) |
| $D$ | 7177 | 7177 |
| $N$ | 14999 | 14999 |

Note. (i) $D$ is the number of non-zero market share observations, and $N$ is the sample size. (ii) Asymptotic standard error estimates appear in parentheses. (iii) The unit of bottle size was liquid ounces. (iv) We normalized the coefficients of Package Size variable to one.

Table 12: Second-stage Parameter Estimates $\left(\hat{\sigma}, \hat{\boldsymbol{\beta}}\right)$

| | Our Model | | | Logit Model | | |
|---|---|---|---|---|---|---|
| first-stage: / Zero: | K/S | Probit | Heckman | Drop 0 | $10^{-8}$ | $10^{-4}$ |
| $\left(\phi_{j,t}, \mathbf{x}_{j,t}\right)$ | (1) | (2) | (3) | (4) | (5) | (6) |
| Log Price ($-\sigma$) | −0.664 | −0.779 | −0.938 | 3.287 | 9.632 | 6.617 |
| | (0.080) | (0.135) | (0.051) | (0.036) | (0.157) | (0.077) |
| Package Size | 0.012 | 0.007 | 0.000 | 0.009 | 0.037 | 0.028 |
| | (0.001) | (0.002) | (0.001) | (0.000) | (0.001) | (0.001) |
| Liquid | 6.189 | 2.554 | −0.070 | 1.861 | 9.688 | 6.287 |
| | (0.629) | (1.074) | (0.390) | (0.065) | (0.254) | (0.124) |
| Heavy duty / Concentrated / Double | 0.660 | −0.156 | −0.699 | 0.780 | 0.087 | 4.105 |
| | (0.198) | (0.246) | (0.207) | (0.160) | (0.488) | (0.238) |
| Bleach | 0.084 | 0.044 | 0.040 | −0.066 | 1.388 | 0.803 |
| | (0.042) | (0.083) | (0.071) | (0.063) | (0.189) | (0.092) |
| Tide | 5.678 | 2.620 | −0.028 | −0.262 | 1.579 | −0.171 |
| | (0.538) | (1.045) | (0.371) | (0.062) | (0.190) | (0.092) |
| Wisk | 3.657 | 1.186 | −0.305 | −1.196 | 1.025 | 0.097 |
| | (0.441) | (0.562) | (0.231) | (0.088) | (0.258) | (0.126) |
| Ajax / ArmHammer / Surf / Purex | −1.003 | −0.105 | 0.191 | 0.601 | 2.719 | 1.681 |
| | (0.178) | (0.332) | (0.144) | (0.081) | (0.227) | (0.111) |
| $D$ | 6339 | 6339 | 6339 | 6339 | 8587 | 8587 |
| $N$ | 13625 | 13625 | 13625 | 13625 | 8587 | 8587 |

Note. (i) For columns Our Model, K/S and Our Model, Probit, results from the Klein-Spady and Probit estimators in Table 11 were used for the first-stage estimator, respectively. A pairwise differenced weighted instrumental variable estimator was then used during the second stage. For the Heckman Correction column, the Probit was used during the first stage, and the Heckman's selection correction estimator with the inverse Mills ratio as an additional regressor was used during the second stage. (ii) $D$ is the number of non-zero market share observations, and $N$ is the effective sample size. (iii) Asymptotic standard error estimates appear in parentheses. (iv) The unit of package size was converted to liquid ounces. (v) Because Dominick's did not record the price and cost when sales were zero, when estimating the $10^{-8}$ and $10^{-4}$ columns, we used average prices and costs of the same product with the same promotion statuses from other stores.