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

Ali Hortaçsu†
University of Chicago and NBER

Joonhwi Joo‡
University of Chicago

First Version: June 14, 2015
Current Version: October 29, 2015

Abstract

We develop a demand estimation framework with observed and unobserved product characteristics based on CES preferences. We show that our demand system can nest the logit demand system with observed and unobserved product characteristics, which has been widely used since Berry (1994); Berry et al. (1995). Furthermore, the demand system we develop can directly accommodate zero market shares by separating the extensive and the intensive margins. We apply our framework to the scanner data of cola sales, which shows that the estimated demand curves can even be upward sloping if zero market shares are not properly accommodated.

JEL classification: C51, D11, D12

*We thank Jean-Pierre Dubé, Chad Syverson, Frank Wolak and seminar participants at the University of Chicago for their valuable comments and suggestions.
†hortacsu@uchicago.edu
‡joonhwi@uchicago.edu
1 Introduction

The Constant Elasticity of Substitution (henceforth CES) preferences, often referred to as the Dixit-Stiglitz-Spence preferences, has long been used to analyze the markets with product differentiation in the macroeconomics and international trade literature since Spence (1976); Dixit and Stiglitz (1977); Krugman (1980). However, recent analyses on the demand of differentiated products in the empirical industrial organization literature since Berry (1994); Berry et al. (1995) have been based on a completely different microfoundation: the discrete choice random utility approach in the characteristics space. We reconcile these two different approaches in this paper by adding the “quality kernel” to the CES preferences. Quality kernel is a nonnegative function which maps the observed and unobserved product characteristics to the marginal utility multiplier of consuming one unit of a specific product. It allows us to (i) incorporate the observed and unobserved product characteristics in the CES preferences, (ii) derive the same share equation with that of Berry (1994); Berry et al. (1995), and (iii) accommodate the zero market share directly. We further show how we can semiparametrically estimate the product differentiated demand model with the data which has a multitude of zero market shares.

Demand estimation has been regarded as a central problem in the industrial organization. The recent empirical industrial organization literature has taken the characteristics space approach in demand estimation, which was initiated by the seminal work of Berry (1994); Berry et al. (1995). In the characteristic space approach, a product is defined as a bundle of observed and unobserved product characteristics. In the baseline model of Berry (1994); Berry et al. (1995); Nevo (2001), a consumer can choose up to one product which yields the highest utility, or can decide not to buy anything. A consumer’s utility from consuming a product is composed of the utility from income net of price, observed product characteristics, unobserved product characteristics, and the idiosyncratic utility shock. The individual choice probability equation is derived from the statistical property of idiosyncratic utility shock, which is assumed to be distributed as Type-I extreme value distribution. We refer to the demand models based on these microfoundations as the discrete choice demand models henceforth. The discrete choice demand models provided a
tractable way to estimate differentiated products demands system by reducing the dimension of parameters to be estimated.

As a drawback of such a tractability, discrete choice demand models have imposed some strong assumptions that are often unrealistic. To be specific, the following features of the discrete choice demand models have been often criticized: single choice assumption, lack of wealth effects in the utility specification, impossibility of accommodating zero shares, and ignoring dynamics (See, Ackerberg et al. (2007); Nevo (2000); Reiss and Wolak (2007) among others). We resolve the first three problems by developing a new microfoundation for the existing discrete choice demand estimation frameworks, based on the CES preferences.

Our first main contribution to the literature is that we develop a novel and tractable method to incorporate the observed and unobserved product characteristics in the CES demand system. Starting from Feenstra (1994), CES preferences and its variations have been widely used in the macroeconomics and international trade literature to estimate the welfare effects from product varieties, allowing free trade, and so on. Among recent works, Broda and Weinstein (2006) estimate the elasticities of substitution for a vast number of goods based on the multi-stage CES preferences, Handbury (2013); Handbury and Weinstein (2014) calculate the city-specific price indices based on the estimation of the preference parameters of multi-stage CES preferences, and Bronnenberg (2015); Li (2013) estimate the utility from product varieties using the CES preferences. Although many works in the literature are closely related to, and have a room to incorporate the product characteristics, to our knowledge none of them directly modeled the utility from the product characteristics.

Our second main contribution is that we provide a concrete linkage between the CES demand system and the discrete choice demand system. CES demand system and discrete choice demand system have been considered to be fundamentally different. First of all, they are based on starkly different microfoundations. CES demand system is based on the CES preferences with infinitely divisible products. By solving the constrained utility maximization problem, the consumer’s demand function is derived as a form of the consumer’s optimal quantity choice. Then the quantity shares
can be calculated, which can be regarded as the individual choice probability of the representative consumer. In the demand system based on discrete choice framework, a consumer can choose up to one product which yields the maximal indirect utility, without being budget constrained. Then, by assuming the independent and identically distributed (i.i.d. henceforth) Type-I extreme value shocks on the preferences, the individual choice probability equations are derived from the statistical property of the i.i.d. Type-I extreme value random variables. Another difference lies in the product characteristics. Discrete choice demand system is regarded to be able to directly incorporate the product characteristics to reduce the dimensionality of parameter space. Although some early works such as [Anderson et al. (1987, 1988, 1989, 1992); Verboven (1996)] point out the similarities with CES demand system and discrete choice demand system, none of them considered the product characteristics as a direct argument of the consumer’s utility function. Among the international trade literature, works such as [Feenstra (1994); Li (2013)] incorporate so-called taste parameter in the CES preferences, but it is not directly related to product characteristics either.

We provide the linkage between the CES demand system and the discrete choice demand system by deriving the individual choice probability equation of discrete choice demand system from the CES preferences, using what we denote as “quality kernel.” This equivalence result of CES demand system and the logit/BLP demand system is remarkable for both empirical industrial organization literature and macroeconomics/international trade literature. For the empirical industrial organization literature, it provides an additional appealing microfoundation for the existing demand estimation frameworks based on discrete choice. For macroeconomics/international trade literature, it provides a strong justification that allows for the demand system based on the CES preferences to use the estimation methods developed in the context of discrete choice.

Our third main contribution is that we provide a direct method to accommodate the zero individual choice probabilities using the quality kernel. We embed both the extensive and the intensive margins on the quality kernel. Thus, it allows us to model the representative consumer’s choice as a two-stage decision process. In the first stage, the consumer chooses the choice set. In the second stage, for each product in the choice set, the consumer chooses how much to buy.
The idea to model both the extensive and the intensive margins in a single utility maximization problem was first pioneered by Hanemann (1984), and followed by Chiang (1991); Chintagunta (1993); Nair et al. (2005). More recently, Handbury (2013) employs this idea in the multi-stage CES utility maximization problem. However, all of these previous works rely on the single-choice assumption, in the sense that the representative consumer chooses up to one brand/product. Our framework overcomes the single-choice assumption, to allow the representative consumer purchase multiple brands/products.

As a result of separately modeling the extensive margin and the intensive margin, we provide an empirically tractable estimation framework that can accommodate the zero individual choice probabilities. Accommodating the zero individual choice probabilities has been a central problem in the demand estimation literature for the last two decades since Berry (1994); Berry et al. (1995). In fact, the discrete choice framework with an additive idiosyncratic error that has unrestricted support inherently rules out the zero choice probabilities. This is exactly the case with the widely used additive Type-I extreme value error assumption. For the additive Type-I extreme value errors, exponential function in the numerator of the individual choice probabilities are inevitable. This implies that provided a product yields any utility higher than negative infinity, the product must have a strictly positive individual choice probabilities. But, the zero market shares, which is equalized with the the choice probabilities in the estimation, are very often observed in the data. Hence, in practice, researchers just dropped the samples with zero market shares or added a small arbitrary number on the zero shares. Both measures suffer from causing the bias in the estimates. By modeling the extensive and intensive margins separately, we argue that the selection process on the consumer’s choice set must be considered during the estimation. This choice set selection drives the conditional expectation of the unobserved product characteristics conditioned on all the instruments to be nonzero, and it is highly likely to be positive. The usual GMM estimation procedure will yield biased estimates when the choice set selection is ignored.

There is one remarkable recent work of Gandhi et al. (2013), which can rationalize the zero market shares. They take the zero market share as a “measurement error” and provides a partial
identification result. The critical difference between theirs and ours is that ours rationalize the zero individual choice probabilities as well as zero market shares, while they allow only the market share to be zero. Nonetheless, their Monte-Carlo simulation and empirical applications show the result which has a very similar implication to ours: when the data with the zero shares are dropped, the price coefficient estimates will be biased upward.

In a different but related context, zero trade flow has also been a serious problem in the gravity models in the international trade literature. As a simple solution, Santos Silva and Tenreyro (2006) argue that in the model that should be estimated is \( E[y_i|x_i] = \exp(x'_i\beta) \), not \( E[\ln y_i|x_i] = x'_i\beta \) which has been a convention in the literature. However, this method cannot be applied to the discrete choice demand models, because the unobservables in the discrete choice demand models should enter inside the exponential functions. Another reason why this seemingly appealing approach loses its justification is that, if \( E[y_i|x_i] = 0 \), there must be negative realizations of \( y_i \) given \( x_i \) unless \( y_i \) is constant. But, choice probabilities are bounded below by zero. A more closely related to our method is developed by Helpman et al. (2008). They use a gravity model with endogenous censoring in the trade volumes. Their structural approach with censoring is similar to ours. Yet, they assumed the normality in the error term while we do not specify the distribution of unobservables. In our empirical example, we also provide an evidence that the distribution of the unobservable product characteristics is far from Gaussian.

Our work can be viewed as a “hedonic model” or “pure characteristics model” of demand, in the sense that we do not require an individual level i.i.d. random utility shocks. Hedonic demand estimation frameworks have been developed by Bajari and Benkard (2005); Berry and Pakes (2007). The former is more closely related to our work. They study the general hedonic model of demand, and focus on the local identification and estimation of the coefficients in the product characteristics. For the global identification when the product space is continuous, they specify the Cobb-Douglas preferences. Our work can be viewed as an extension to CES preferences with product characteristics, which can also accommodate the zero shares.

The rest of the paper is organized as follows. In Section 2, we present the general microfoun-
dation of the CES demand system with product characteristics. Section 3 provides the linkage between CES demand system and logit demand system. Section 4 presents the semiparametric estimation method when the data contains a multitude of zero shares. Section 5 presents simulation evidences. In Section 6, we implement our estimation method to the scanner data of cola sales. Section 7 concludes.

2 CES Demand System with Observed and Unobserved Product Characteristics

2.1 Specification of the CES Demand System

We consider a differentiated product market denoted by subscript $t$ composed of homogenous consumers with a CES preference. The utility from a product category is given by:

$$u\left(\{q_{j,t}, x_{j,t}, \xi_{j,t}, w_{j,t}, \eta_{j,t}\}_{j \in J_t}\right) := \left(\sum_{j \in J_t} \left\{\chi(x_{j,t}, \xi_{j,t}, w_{j,t}, \eta_{j,t})\right\}^{\frac{1}{\sigma}} q_{j,t}^{\frac{\sigma-1}{\sigma}}\right)^{-\frac{1}{\sigma-1}}. \quad (2.1)$$

The set $J_t$ is a set of alternatives in the category, which may or may not include the numeraire that represents the outside option. $q_{j,t}$ is the quantity of product $j$ consumed in market $t$. $\chi(x_{j,t}, \xi_{j,t}, w_{j,t}, \eta_{j,t})$, what we define by the “quality kernel,” is a nonnegative function of observed and unobserved product characteristics. $x_{j,t}$ and $w_{j,t}$ 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 of product $j$’s characteristics which are unobservable to the econometrician. $w_{j,t}$ and $\eta_{j,t}$ are the extensive margin shifters that a consumer considers whether to buy the product or not. $x_{j,t}$ and $\xi_{j,t}$ are the intensive margin shifters that determine the level of utility when a consumer buys the product. $w_{j,t}$ and $x_{j,t}$ may have common components, but we require the exclusion restriction on $w_{j,t}$ for identification when the extensive margin actually matters; $w_{j,t}$ has to contain at least one component that is not in $x_{j,t}$ when zero market shares are observed in data. We will explain further about the identification conditions later in Section 4. As an important remark, we note that the observed extensive
margin shifters $w_{j,t}$ may contain the prices $p_{j,t}$ or the logarithm of prices $\ln p_{j,t}$.

The quality kernel $\chi(x_{j,t}, \xi_{j,t}, w_{j,t}, \eta_{j,t})$ in (2.1) plays the key role in our analysis. Conventionally, researchers have put some taste parameters or utility weights in the place where we put the quality kernel. They commonly have the interpretation of a multiplier to the (marginal) utility of consuming a specific product. The quality kernel is a straightforward extension of such conventions, which allows to directly incorporate the observed and unobserved product characteristics in consumer’s utility. Furthermore, we allow the possibility of explicitly separating the extensive margin and the intensive margin. This feature will play a key role to accommodate the zero shares in the estimation.

The representative consumer’s utility maximization problem is as follows:

$$\max_{\{q_{j,t}\}_{j \in J_t}} \left( \sum_{j \in J_t} \left\{ \chi(x_{j,t}, \xi_{j,t}, w_{j,t}, \eta_{j,t}) \right\} \right) \frac{1}{\sigma} q_{j,t}^{\sigma-1} \text{ s.t. } \sum_{j \in J_t} p_{j,t} q_{j,t} = w_t$$

(2.2)

The Marshallian demand function is given by

$$q_{j,t} = w_t \left\{ \frac{\chi(x_{j,t}, \xi_{j,t}, w_{j,t}, \eta_{j,t}) p_{j,t}^{-\sigma}}{\sum_{k \in J_t} \chi(x_{k,t}, \xi_{k,t}, w_{k,t}, \eta_{k,t}) p_{k,t}^{1-\sigma}} \right\},$$

(2.3)

which leads to the following quantity share equation:

$$\pi_{j,t} = \frac{q_{j,t}}{\sum_{k \in J_t} q_{k,t}} = \frac{\chi(x_{j,t}, \xi_{j,t}, w_{j,t}, \eta_{j,t}) p_{j,t}^{-\sigma}}{\sum_{k \in J_t} \chi(x_{k,t}, \xi_{k,t}, w_{k,t}, \eta_{k,t}) p_{k,t}^{1-\sigma}}.$$

(2.4)

(2.4) can be regarded as the individual choice probability of the representative consumer. Either (2.3) or (2.4) specifies the CES demand system with observed and unobserved product characteristics.

For the invertibility of the demand system, we consider the subset $J_t^+ (\subseteq J_t)$ such that $\pi_{j,t} > 0$ for all $j \in J_t^+$. The demand system specified by $\{q_{j,t}\}_{j \in J_t^+}$ or $\{\pi_{j,t}\}_{j \in J_t^+}$ satisfies the “connected substitutes” conditions by Berry et al. (2013), and thus invertible. It implies $\sigma$, the elasticity of
substitution, is globally identified. Furthermore, if we impose suitable structures on \( \chi(\cdot) \) such as monotonicity with index restriction, the structural parameters of \( \chi(\cdot) \) is also identified. We shall investigate further on the specific functional forms of \( \chi(\cdot) \) in Section 3.

### 2.2 Properties of the CES demand system and Comparison to the Discrete Choice Demand System

In this subsection, we explain the properties of the demand system specified by the Marshallian demand function (2.3). The demand system that we propose, which is derived from the constrained optimization problem, has a few more desirable properties over the existing demand estimation frameworks based on the discrete choice random utility models.

We begin with the Marshallian and Hicksian own and cross price elasticities of the demand system. Let \( b_j \) be the budget share of product \( j \). If \( w_{j,t} \) does not include the prices \( p_{j,t} \) or its function, we have the following simple closed-form formula for the Marshallian and Hicksian own and cross price elasticities:

\[
\begin{align*}
\varepsilon_{jj,t}^M &= -\sigma + (\sigma - 1) b_{j,t} \\
\varepsilon_{jc,t}^M &= (\sigma - 1) b_{c,t} \\
\varepsilon_{jj,t}^H &= -\sigma (1 - b_{j,t}) \\
\varepsilon_{jc,t}^H &= \sigma b_{c,t},
\end{align*}
\]

and the income elasticity is 1. Because the budget shares are observed in data, these elasticities can be calculated given that \( \sigma \) is identified. From these elasticity expressions, it is immediate that a version of independence of irrelevant alternatives property holds. That is, the substitution pattern depends solely on the budget share of the corresponding products. It is also noteworthy that the price elasticities should not be derived based on the individual choice probabilities as in
the discrete choice frameworks.\footnote{1}

If \( w_{j,t} \) includes the prices \( p_{j,t} \) or its function so that the extensive margin is affected by the price changes, there is no simple closed-form expression for the own and cross elasticities. In practice, the corresponding price elasticities can be calculated using simulation.

Because we have derived the demand system from the constrained optimization problem, the duality between the Marshallian demand function and the Hicksian demand function holds. The Slutsky equation also holds, so we can decompose the substitution effect and the income effect in a more natural way. The Slutsky equation in the elasticity form is given by

\[
\varepsilon_{j,c,t}^M = \varepsilon_{j,c,t}^H - \varepsilon_{j,t}^I b_{c,t},
\]

where \( \varepsilon_{j,t}^I = 1 \) in the CES demand system. The income effect depends only on the budget shares, which is a considerable limitation. However, there are at least two advantages over the discrete choice counterpart. First, the income effect depends on the budget shares, not on the quantity shares. In the discrete choice demand frameworks, income effect of a product with a tiny budget share and large quantity shares will be large, which is even more unrealistic. Next, although the numeraire can be included in the consumer’s choice set \( J_t \), it is not necessary in our framework. In the demand system derived from the quasilinear utility where budget constraint never binds, the inclusion of numeraire in the choice set is necessary. If the numeraire is not included in the choice set, an increase of the price will only lead to switching to other products in the choice set. In order

\[
\frac{\partial \ln \pi_{j,t}}{\partial \ln p_{j,t}} = \frac{\partial \ln q_{j,t} - \ln (\sum_k q_{k,t})}{\partial \ln \pi_{j,t}} = \frac{\partial \ln \pi_{j,t}}{\partial \ln p_{j,t}} - \frac{\partial \ln (\sum_k q_{k,t})}{\partial \ln p_{j,t}} = \varepsilon_{j,t}^M - \frac{\partial \ln (\sum_k q_{k,t})}{\partial \ln p_{j,t}} \neq \frac{\partial \ln q_{j,t}}{\partial \ln p_{j,t}}
\]

The term \( \frac{\partial \ln \pi_{j,t}}{\partial \ln p_{j,t}} \) is the Marshallian price elasticity only when \( \sum_k q_{k,t} \) is constant, which is the case for the discrete choice demand models.
to include the numeraire in the choice set, the share of the numeraire or the potential market size has to be taken as known in the discrete choice demand models. In a sense the magnitude of the income effect is predetermined, because it solely depends on the size of the outside share.

3 The Exponential Quality Kernel

We let \( \chi(x_{j,t}, \xi_{j,t}, w_{j,t}, \eta_{j,t}) \) flexible so far. In principle, \( \chi(x_{j,t}, \xi_{j,t}, w_{j,t}, \eta_{j,t}) \) can be any nonnegative function. Under this weak restriction, the demand system specified by the individual choice probability equation (2.4) may be locally identified, as studied in Bajari and Benkard (2005). However, nonparametric estimation of the locally identified demand system places considerable burden on the data and computational power, which is often impractical. Furthermore, the locally identified parameter values are often uninformative for the counterfactual analyses. Alternatively, we can impose further structures on the consumer utility from the product characteristics.

We focus on the exponential quality kernel with the index restriction. This specific functional form is remarkable for two prominent reasons. First, by using the functional form, we can derive the same individual choice probability equation of the homogenous and random coefficients logit model of demand from the CES demand system. Second, it simplifies the estimation problem substantially, because the demand system reduces to the log-linear form. We use the exponential quality kernel to suggest a tractable semiparametric estimation method which can also accommodate the zero individual choice probabilities.

3.1 Nesting Homogenous and Random Coefficients Logit Model of Demand

In this subsection, we show that the individual choice probability equation of the homogenous and random coefficients logit models of demand can be derived from (2.4) by choosing a specific functional form of the quality kernel \( \chi(x_{j,t}, \xi_{j,t}, w_{j,t}, \eta_{j,t}) \).

Suppose \( x_{j,t} = w_{j,t}, \xi_{j,t} = \eta_{j,t}, \chi(x_{j,t}, \xi_{j,t}) > 0, \pi_{j,t} > 0 \) and let \( x_{j,t} \) be exogenous for all \( j, t \). We
do not require an exclusion restriction in this setup because the individual choice probabilities are positive for every alternative. Let \( J_t \) contain the numeraire denoted by product 0. Taking ratios of product \( j \) and product 0, and then taking logarithm on \((2.4)\) yields:

\[
\ln \left( \frac{\pi_{jt}}{\pi_{0,t}} \right) = -\sigma \ln (p_{jt}) + \ln \chi (x_{jt}, \xi_{jt}) - \ln \chi (x_{0,t}, \xi_{0,t}).
\] (3.1)

It is straightforward that if we normalize \( x_{0,t} = 0, \xi_{0,t} = 0 \) and let \( \chi (x_{jt}, \xi_{jt}) = \exp \left( x'_{jt} \beta + \xi_{jt} \right) \), \((3.1)\) becomes

\[
\ln \left( \frac{\pi_{jt}}{\pi_{0,t}} \right) = -\sigma \ln (p_{jt}) + x'_{jt} \beta + \xi_{jt}
\] (3.2)

which is the estimation equation of the homogeneous logit model of demand, except that in \((3.2)\) the \( \ln (p_{jt}) \) enters in place of \( p_{jt} \), which has been a convention in the literature. The logarithm of price should enter \((3.2)\) because it is inherited from the consumer’s budget constraint. However, we observe that \( \ln (p_{jt}) \) can also enter in place of \( p_{jt} \) in the utility specification of the logit demand system. That is, by substituting \( \ln (p_{jt}) \) with \( p_{jt} \) in the linear utility specification in the logit demand model, the demand system we propose exactly lines up with the demand system of the logit model. We take the substitution with the logarithm of prices as a simple scale adjustment in the linear utility specification of the logit model.

\footnote{Berry (1994), Berry et al. (1995) specifies the linear utility as:}

\[
u_{i,j,t} = \alpha_i (w_i - p_{jt}) + x'_{jt} \beta_i + \xi_{jt} + \epsilon_{i,j,t} = \alpha_i w_i - (\alpha + \Pi_i q_i + \Sigma_\alpha v_{\alpha,i}) p_{jt} + x_{jt} \beta (\alpha + \Pi_i q_i + \Sigma_\beta v_{\beta,i}) + \xi_{jt} + \epsilon_{i,j,t}
\] = \alpha_i w_i + (\alpha p_{jt} + x'_{jt} \beta + \xi_{jt}) - (\Pi_i q_i + \Sigma_\alpha v_{\alpha,i}) p_{jt} + x'_{jt} (\Pi_i q_i + \Sigma_\beta v_{\beta,i}) + \epsilon_{i,j,t}
\] = \alpha_i w_i + (\alpha p_{jt} + x'_{jt} \beta + \xi_{jt}) + (\alpha p_{jt} x'_{jt}) (\Pi_i q_i + \Sigma_\beta v_{\beta,i}) + \epsilon_{i,j,t}
\] = \alpha_i w_i + \delta_{j,t} + \mu_{i,j,t} + \epsilon_{i,j,t},

where \( q_i \) is demographic variables and \( v_i \) is a unit normal shock. Given \( F (z_i), F (v_i) \) and \( \epsilon_{i,j,t} \sim i.i.d. T1EV \), the predicted market share of good \( j \) becomes:

\[
s_{j,t} := \int \int \int_{\{\epsilon_{i,j} \mid u_{i,j} \geq u_{i,k} \forall k\}} dF (\epsilon_{i,j,t}) dF (z_i) dF (v_i) = \int \int \frac{\exp (\delta_{j,t} + \mu_{i,j,t})}{\sum_{k \in J_t} \exp (\delta_{k,t} + \mu_{i,k,t})} dF (z_i) dF (v_i).
\] (3.3)

If \( \beta_i = \beta \) so there is no heterogeneity of preferences across individuals, the model reduces to a homogenous logit demand model.
The random coefficients logit model of demand by Berry et al. (1995) (henceforth BLP) can be nested in the same way. Let $i$ denote an individual, and let us suppress the market subscript $t$ for a while. For notational simplicity, let $\phi_j := \ln p_j$. Suppose we specify the linear utility of the BLP model as

$$u_{i,j} = \alpha_i (\ln w_i - \phi_j) + \mathbf{x}_j' \beta_i + \xi_j + \epsilon_{i,j}.$$ 

On the other hand, the individual choice probability equation (2.4) from CES preferences is

$$\pi_{i,j} = \frac{\chi_i (\mathbf{x}_j, \xi_j) \exp (-\sigma_i \phi_j)}{\sum_{k=0}^{J} \chi_i (\mathbf{x}_k, \xi_k) \exp (-\sigma_i \phi_k)}$$

$$= \frac{\exp (-\sigma_i \phi_j + \mathbf{x}_j' \beta_i + \xi_j)}{\sum_{k=0}^{J} \exp (-\sigma_i \phi_k + \mathbf{x}_k' \beta_i + \xi_k)}$$

where the second equality follows by specifying $\chi_i (\mathbf{x}_j, \xi_j) = \exp (\mathbf{x}_j' \beta_i + \xi_j)$ and $\beta_i = \beta + \Pi_{\beta z_i} + \Sigma_{\beta v_{\beta, i}}$. Notice that (3.5) is almost exactly same as the individual choice probability equation obtained in the BLP model. The market share equation is obtained by aggregating these individual choice probabilities over $i$.

Discussion in the current subsection provides the microfoundation and justification for international trade or macroeconomics literature based on CES preferences to use the differentiated products demand estimation methods developed in empirical industrial organization literature since Berry (1994); Berry et al. (1995); Nevo (2001). However, the discrete choice differentiated products demand estimation literature has imposed one critical restriction which has been regarded as necessary in inverting the individual choice probabilities: $\pi_{j,t} > 0$ for all $j, t$. The restriction is

3The only structural difference is the correlation structure of the individual heterogeneity. That is, we have to assume $Cov(\sigma_i, \beta_i) = 0$. Because those cross-correlations are often assumed to be zero in practice when estimating BLP model (See, e.g. Dube et al. (2012)), we do not take the restriction as a serious limitation.

Although the individual choice probability equations are the same, the price elasticities are slightly different. There is no distinction between Marshallian and Hicksian own and cross price elasticities in the logit demand models. All the price elasticities of the logit models are in a sense compensated, which are:

$$\varepsilon_{jj,t} = -\sigma (1 - \pi_{j,t})$$

$$\varepsilon_{jc,t} = -\sigma \pi_{c,t}.$$ 

Notice that the only difference to the Hicksian price elasticities derived from CES preferences is that the multiplied term to $-\sigma$ is composed of the individual choice probabilities, not the budget shares.

4For a detailed discussion on share inversion, see Berry et al. (2013).
inevitable in the discrete choice demand models which assume the additive idiosyncratic shocks on
the preferences distributed with the unrestricted support. The most important example in the lit-
erature has been the additive i.i.d. Type-I extreme value distributed shocks. The individual choice
probabilities derived from such assumption must have exponential functions in the numerators of
choice probabilities.

Zero quantity shares, which are equalized with the individual choice probabilities in the discrete
choice demand models, are very often observed in data. Within the discrete choice frameworks,
it has often been regarded impossible to rationalize the zero shares. In contrast, the flexibility
of quality kernel \( \chi(x_{j,t}, \xi_{j,t}, w_{j,t}, \eta_{j,t}) \) in our model allows us to accommodate the zero individual
choice probabilities by embedding the buy-or-not decision of the consumer, which determines the
extensive margins. In the following subsection, we illustrate how we can directly accommodate the
zeros in the market shares.

3.2 Accommodating Zero Market Shares: Separating the Extensive
and Intensive Margins

From here we restrict our attention on the homogenous consumers, and let \( x_{j,t} \neq w_{j,t}, \eta_{j,t} \neq \xi_{j,t} \).
Let \( J_t \) contain the numeraire. Recall the individual choice probability equation of the CES demand
system:

\[
\pi_{j,t} = \frac{\chi(x_{j,t}, \xi_{j,t}, w_{j,t}, \eta_{j,t}) \exp(-\sigma \phi_{j,t})}{\sum_{k \in J_t} \chi(x_{j,t}, \xi_{j,t}, w_{j,t}, \eta_{j,t}) \exp(-\sigma \phi_{k,t})}.
\]  

(3.6)

The individual choice probability equation (3.6) allows the zero shares, by letting \( \chi(x_{j,t}, \xi_{j,t}, w_{j,t}, \eta_{j,t}) = 0 \) for some subset of the product characteristic space where \( (w_{j,t}, \eta_{j,t}) \) lives on. We can even ra-
tionalize the data where all products in a particular market have zero shares. By taking the ratio
we obtain a reduced form of the demand system (3.6) as
\[
\frac{\pi_{j,t}}{\pi_{0,t}} = p^{-\sigma} \chi (x_{j,t}, \xi_{j,t}, w_{j,t}, \eta_{j,t}) \chi (x_{0,t}, \xi_{0,t}, w_{0,t}, \eta_{0,t}).
\] (3.7)

For tractability in identification and estimation, again we consider the following functional form with the index restriction:
\[
\chi (x_{j,t}, \xi_{j,t}, w_{j,t}, \eta_{j,t}) = \begin{cases} 1 & \{\gamma + w_{j,t}' \delta + \eta_{j,t} > 0\} \\
 h (\alpha + x_{j,t}' \beta + \xi_{j,t}) \end{cases},
\] (3.8)

where \(\mathbb{1} (\cdot)\) is an indicator function and \(h\) is a nonnegative, strictly increasing function on \(\mathbb{R}_+\).

Employing this quality kernel is equivalent to assuming a certain structure on the consumer’s choice. The consumer first considers the utility from product characteristics represented by \(w_{j,t}' \delta + \eta_{j,t}\). If it exceeds the threshold \(-\gamma\), the consumer decides to buy it. Then \((\phi_{j,t}, x_{j,t}, \xi_{j,t})\) is taken into consideration, which affects the amount of consumption \(q_{j,t}\). On the other hand, if the utility does not exceed the threshold \(-\gamma\), the consumer decides not to buy, thus \(q_{j,t} = \pi_{j,t} = 0\). We again note that \(w_{j,t}\) can contain the current prices \(p_{j,t}\) or other endogenous variables provided that the corresponding instruments are available to the researcher.

The choice \(h (y) = \exp (y)\), which we will focus on throughout the rest of the paper, deserves a special interest. As in Section 3.1, the demand system reduces to a log-linear form under this choice of the quality kernel. It simplifies the estimation problem significantly because the error term becomes additively separable. Furthermore, our demand system becomes a direct extension of the homogenous logit demand system, which also allows the accommodation of the zero choice probabilities. Our demand system derived from the CES preferences can provide with the microfoundation required to rationalize the zero choice probabilities by separating the intensive margin and the extensive margin.

In Appendix A we provide a two-stage modeling of consumer behavior within the discrete

---

\(^5\)If \(\mathcal{J}_t\) does not include the numeraire, any product with a strictly positive market share can be taken as product 0. All the arguments of the current and the next section remain valid provided that we assume the independence of observable and unobservable product characteristics across products.
choice random utility framework. Both ways lead to the same estimation equation, which is presented in the following Section. Although we find that sticking to the discrete choice random utility framework is less appealing because the extensive margins and the intensive margins cannot be conceptually distinguished, there might be instances in which the single choice assumption is more adequate. The result presented in Appendix A can be useful under such circumstances.

4 Semiparametric Estimation Framework with Exponential Quality Kernel and Zero Shares

In this section, we provide a semiparametric estimation framework for the exponential quality kernel which accommodates the zero market shares. The estimation method we provide is in two stages. In the first stage, we estimate the parameters that determine the extensive margin, using the efficient semiparametric estimators developed by Klein and Spady (1993). In the second stage, we estimate the parameters that determine the intensive margin, correcting for both the endogeneity in prices and the selectivity bias caused from the consumers’ choice set selection. The second-stage estimator we use is developed by Ahn and Powell (1993); Powell (2001). Note that when there is no observed zero shares, one can proceed with the existing demand estimation frameworks including BLP.

The first-stage estimation framework illustrated in this section allows only the exogenous covariates for the observed extensive margin shifters $w_{j,t}$. We chose this partly due to our availability of the data, and partly due to the efficiency concern.\footnote{Although our data has an information about the availability even when the sales in the corresponding week/store pair is zero, it does not record the prices of the week when the sales is zero. Thus, we could not contain $p_{j,t}$ in $w_{j,t}$ in the first-stage estimation. The characteristics of the data used in our empirical application will be discussed further in details in Section.} If the researcher wants to include the endogenous variables such as prices in the extensive margin shifters $w_{j,t}$, one can proceed with the method developed by Blundell and Powell (2003, 2004) in the first-stage. They provide the semiparametric estimation framework for the binary choice models with endogenous covariates.

We assume the existence of instruments for prices such that $E[\xi_{j,t}|z_{j,t}] = 0$, where $z_{j,t}$ may in-
clude \( x_{j,t} \). It is very well documented in the literature that \( E [ \xi_{j,t} | \phi_{j,t}, x_{j,t} ] \neq 0 \), and it is highly likely to be positive. As a result, when the prices are not instrumented, very often the upward sloping demand curves are estimated. The same intuition applies when the consumers’ choice set selection is ignored and zero shares are simply dropped during the estimation. That is, even after instrumenting for the prices, \( E [ \xi_{j,t} | z_{j,t}, \pi_{j,t} > 0 ] = 0 \) does not imply \( E [ \xi_{j,t} | z_{j,t}, \pi_{j,t} > 0 ] \) is likely to be positive, because consumers will select the products with high \( \eta_{j,t} \) in the first-stage choice set decision, and \( \eta_{j,t} \) is likely to be positively correlated with \( \xi_{j,t} \). Thus, dropping zero shares in the estimation will bias the price coefficient upward, which can even yield the positive price coefficients. Imputing zero shares with some small positive numbers during the estimation can cause an even more serious problem; the direction of the bias is unpredictable.

We normalize \( \phi_0 = \ln p_0 = 0, \xi_0 = \eta_0 = 0, w_0 = 0, \) and \( x_0 = 0 \). Under the choice of \( \chi \) specified in (3.8), (3.7) simplifies to:

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

(4.1)

which is the econometric model that we are going to identify and estimate in this section. A consumer buys product \( j \) if \( \gamma + w'_{j,t} \delta + \eta_{j,t} > 0 \). For the sample with \( \pi_{j,t} > 0 \), the demand system (4.1) further reduces to:

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

However, conditional expectation \( E [ \xi_{j,t} | z_{j,t}, w_{j,t}, \pi_{j,t} > 0 ] \) does not hold anymore. It boils down to a classic example of the selection problem.

Several methods to estimate the model parameters of the censored selection model have been suggested in the literature under different assumptions\footnote{For example, Powell (1984, 1986); Blundell and Powell (2007) suggested the LAD type estimator under the conditional quantile restriction, and Honoré et al. (1997) suggested the symmetric trimming under symmetry assumption of the error terms.}. We follow the spirit of Heckman (1979) here, which imposes the conditional mean restriction. By taking the conditional expectation, we...
have:

$$E \left[ \ln \left( \frac{\pi_{j,t}}{\pi_{0,t}} \right) | z_{j,t}, w_{j,t}, \pi_{j,t} > 0 \right] = -\sigma \phi_{j,t} + x'_{j,t} \beta + E \left[ \xi_{j,t} | z_{j,t}, w_{j,t}, \pi_{j,t} > 0 \right]. \quad (4.2)$$

Ahn and Powell (1993); Powell (2001); Newey (2009) proposed two-stage $\sqrt{N}$-consistent estimators of the model parameters of (4.2). In particular, we use Ahn and Powell (1993); Powell (2001)'s pairwise differenced weighted least squares estimator, which allows for identification using an instrument in the second stage.

In the first stage, $\delta$ needs to be estimated. There are a few estimators for this semiparametric binary choice model, among which we use the method by Klein and Spady (1993) which achieves the asymptotic efficiency. In the second stage, the parameters $(\sigma, \beta)$ from the following linear equation is estimated:

$$E \left[ \ln \left( \frac{\pi_{j,t}}{\pi_{0,t}} \right) | z_{j,t}, w_{j,t}, \pi_{j,t} > 0 \right] = -\sigma \phi_{j,t} + x'_{j,t} \beta + \lambda \left( 1 - G \eta \left( -w'_{j,t} \delta \right) \right) \quad (4.3)$$

where $\lambda(\cdot)$ is an unknown function.

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 $w_{j,t}$ is exogenous.

**Assumption 2.** $\eta_{j,t}$ is independent of $w_{j,t}$ with $E[\eta_{j,t}|w_{j,t}] = 0$, and i.i.d. over $j$.

**Assumption 3.** There exists a set of instrument $z_{j,t}$ such that $\xi_{j,t} \perp p_{j,t}|z_{j,t}$, $E[\xi_{j,t}|z_{j,t}] = 0$, and $\text{dim}(z_{j,t}) \geq \text{dim}(p_{j,t}, x_{j,t})$.

**Assumption 4.** The parameter vector $(\sigma, \alpha, \beta, \gamma, \delta)$ lies in a compact parameter space, with the true parameter value lying in the interior.

**Assumption 5.** $w_{j,t}$ contains at least one component which is not included in $x_{j,t}$.

**Assumption 6.** Data generating process of $(\pi_{j,t}, p_{j,t}, w_{j,t}, x_{j,t}, z_{j,t})$ satisfies (C.6) of Klein and Spady (1993) and Assumption 5.7 of Powell (2001).
Assumption 7. The conditional distribution of $\eta_{j,t}$ given $w_{j,t}$ satisfies (C.4a), (C.4b) and (C.9) of Klein and Spady (1993).

In Assumptions 1 and 2, we are imposing the independence of the observed and unobserved product characteristics and homoskedasticity of unobservable product characteristics $\eta_{j,t}$, which is related to the extensive margins. However, we do not assume that the unobserved product characteristics and the prices are independent. We allow for the endogeneity in prices, which should be considered in the identification and estimation; the prices can be some function of the observed and unobserved product characteristics. Assumption 3 is the standard instrument condition to correct for the endogeneity in prices. For the discussion about suitable instruments in practice, see Nevo (2001).

The remaining assumptions are imposed for the $\sqrt{N}$-consistency and asymptotic normality of the distribution-free efficient semiparametric estimator developed by Klein and Spady (1993). Klein-Spady estimator allows to semiparametrically estimate the parameters of the binary choice models without having to specify the distribution of the error term. The key insight is to replace the likelihood with its uniformly consistent estimates, and to run the pseudo maximum likelihood. Assumption 5 is the exclusion restriction, which is required for identification in the second-stage. Assumptions 6 and 7 are regularity conditions for the $\sqrt{N}$-consistency and asymptotic normality of Klein and Spady (1993) and Powell (2001) estimator. Under Assumptions 1-7, conditions (C.1)-(C.9) of Klein and Spady (1993) and Assumptions 5.1-5.7 of Powell (2001) are satisfied.

We describe our first and second stage estimators. In the first stage, we estimate $\delta$ using Klein-Spady estimator. The estimator is defined by:

$$\hat{\delta} := \arg \max_\delta \sum_{j,t} \left\{ 1 (\pi_{j,t} > 0) \ln \left( 1 - \hat{G}_\eta (-w_{j,t}' \delta) \right) + 1 (\pi_{j,t} = 0) \ln \left( \hat{G}_\eta (-w_{j,t}' \delta) \right) \right\}, \quad (4.4)$$

where

$$\hat{G}_\eta (-w_{j,t}' \delta) = \frac{\sum_{k \neq j,t} \kappa \left( \frac{1}{h_n} (w_k - w_{j,t})' \delta + \tau_0 (\delta) \right) (1 - 1 (\pi_{j,t} > 0))}{\sum_{k \neq j,t} \kappa \left( \frac{1}{h_n} (w_k - w_{j,t})' \delta + \tau (\delta) \right)}.$$

$\kappa (\cdot)$ is a local smoothing kernel, $h_n$ is the bandwidth, and $\hat{\tau}_{j,t}, \tau_0 (\delta), \tau (\delta)$ are trimming sequences.
for small estimated densities. In the second stage, we follow Powell (2001). With an abuse of notation by suppressing the market index $t$ and letting $r_j := (\phi_j, x_j)'$, the estimator is defined by

\[
(-\hat{\sigma}, \hat{\beta}) = \left( \frac{1}{n} \sum_{i=1}^{n} \sum_{j=i+1}^{n} \hat{\omega}_{i,j} (z_i - z_j) (r_i - r_j)' \right)^{-1} \left( \frac{1}{n} \sum_{i=1}^{n} \sum_{j=i+1}^{n} \hat{\omega}_{i,j} (z_i - z_j) \left[ \ln \left( \frac{\pi_i}{\pi_0} \right) - \ln \left( \frac{\pi_j}{\pi_0} \right) \right] \right),
\]

(4.5)

where $\hat{\omega}_{i,j} = \frac{1}{\hat{\kappa}} \kappa \left( \frac{1}{\hat{\kappa}} (w_i - w_j)' \delta \right)$. The intuition for the estimator is to cancel out the bias correction term $\lambda \left( 1 - G_\eta \left( -w_j' \delta \right) \right)$; if $w_i$ is same as $w_j$, the term $\lambda \left( 1 - G_\eta \left( -w_j' \delta \right) \right)$ in (4.3) will be canceled out when we take the difference. Thus, we place more weight on the differenced term which are “close” to each other, and then run the least square estimation. 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 the discussion made in this subsection so far.

**Theorem 4.1.** Under Assumptions 1-7, $(\hat{\sigma}, \hat{\beta})$ defined in (4.5) is $\sqrt{N}$-consistent and asymptotically normal.

The log-linear estimation illustrated in this subsection has an advantage of allowing for the consumer’s buy-or-not decision to be completely separate from how-much-to-buy decision. A downside is that we need an exclusion restriction on $w_{j,t}$ to identify $(-\sigma, \beta)$; $w_{j,t}$ cannot be a linear combination of $(\phi_{j,t}, x_{j,t})$. This exclusion restriction can be circumvented by adding an interaction term or a nonlinear transformation of nonbinary variables contained in $(\phi_{j,t}, x_{j,t})$. For instance, if one proceeds with the method suggested by Blundell and Powell (2003, 2004) that can accommodate the prices in the first-stage estimation, including the raw prices $p_{j,t}$ in $w_{j,t}$ may be a possible choice. But ideally, it would be the best to find additional exogenous variables which only affect the consumer’s buy-or-not decision.

\footnote{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.}


5 Monte-Carlo Simulation

In this section, we simulate the market data and try to back out the model parameters to examine the finite-sample performance of the estimator we proposed in the previous section. We compare the estimation result using our model to the estimation result of the logit model which either drops the sample with zero shares or imputes the zero shares with a small positive number. It turns out that the estimator we proposed works well when the model is correctly specified. The price coefficient estimates are quite robust on the model misspecification.

We first describe the data generating process which satisfies the exclusion restriction. There are 2-5 products in each market $t$. The exact number of products by market is randomly drawn. The observed product characteristics vector $w_{j,t}$ includes three continuous components, one discrete component, and three brand dummies. One of the continuous component is excluded in $x_{j,t}$. The first component $w_{j,t}^{(1)}$ follows lognormal $(0, 1)$, the second component $w_{j,t}^{(2)}$ uniform $(1, 5)$, the third component $w_{j,t}^{(3)}$ Poisson $(3)$, the fourth component $w_{j,t}^{(4)}$ $N(0, 1)$. We let only $w_{j,t}^{(4)}$ is not included in $x_{j,t}$. $\eta_{j,t}|w_{j,t}$ follows Type-I extreme value distribution with mean zero.

There exists two instruments for the prices, which are the proxies of the cost shock. The prices $p_{j,t}$, which is an endogenous variable, is determined by $p_{j,t} = \psi(x_{j,t}, \xi_{j,t})$, where $\psi$ is some (possibly) nonlinear function which is strictly monotonic in $\xi_{j,t}$. To be specific, we specify $\psi$ as:

$$
\psi(z_{j,t}, \xi_{j,t}) = 2 + \frac{1}{50} \left( 2z_{j,t}^{(1)} + 4z_{j,t}^{(2)} + 2x_{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 the common circumstances in practice. The individual choice probability is determined by (3.6).

We calibrate the parameters as $\sigma = 2$, $\alpha = 1$, $\beta = (1, -2, 1.5, 0.3, 0.2, 0.4)'$, $\gamma = \alpha$, $\delta = \frac{1}{4} \times (\beta, 0.1)'$.

Table 1 presents the estimation result of the simulated data. A few remarks are in order. First, the performance of the estimator we proposed is good if the model is correctly specified. Even if the quality kernel is misspecified, the price coefficient estimates are close to the true parameter. Next, dropping the sample with zero shares or imputing small positive numbers in place of the
zero shares biases the estimators substantially. Lastly, we also generated and estimated several other specifications such as different error term distributions, different functional forms of quality kernels, different variables, different pricing functions, and so on. Although we did not present all of them here for brevity, we note that the results and implications presented in this section were robust.

Table 1: Estimation Result of Simulated Data

| DGP $h(y)$ | exp($y$) | y | $y^2$ | exp($y$) | Logit, Drop 0 | Logit, Impute $10^{-8}$ |
|------------|----------|---|-------|----------|---------------|--------------------------|
|            | Estimation | Our Model, K/S |                      |                      |               |                          |
| Log prices (-2) | -1.988 | -2.097 | -1.764 | -1.287 | -5.444 |
|              | (0.189) | (0.236) | (0.679) | (0.074) | (0.372) |
| $x_{j,t}^{(1)}$ (1) | 0.983 | 0.095 | 0.144 | 0.886 | 1.536 |
|              | (0.018) | (0.009) | (0.095) | (0.007) | (0.041) |
| $x_{j,t}^{(2)}$ (-2) | -1.622 | -0.471 | -0.901 | -1.574 | -3.884 |
|              | (0.029) | (0.019) | (0.064) | (0.013) | (0.050) |
| $x_{j,t}^{(3)}$ (1.5) | 1.253 | 0.310 | 0.648 | 1.339 | 1.821 |
|              | (0.026) | (0.013) | (0.039) | (0.008) | (0.040) |
| $x_{j,t}^{(4)}$ (0.3) | 0.212 | 0.036 | 0.164 | 0.366 | -0.468 |
|              | (0.115) | (0.058) | (0.139) | (0.046) | (0.219) |
| $x_{j,t}^{(5)}$ (0.2) | 0.163 | 0.061 | 0.034 | 0.266 | -0.673 |
|              | (0.108) | (0.075) | (0.148) | (0.046) | (0.220) |
| $x_{j,t}^{(6)}$ (0.4) | 0.339 | 0.070 | 0.028 | 0.458 | -0.099 |
|              | (0.104) | (0.077) | (0.256) | (0.047) | (0.225) |
| D | 5059 | 4680 | 4593 | 5059 | 10500 |
| N | 10500 | 10557 | 10482 | 5059 | 10500 |

Note. (i) Target values are in the parentheses of the corresponding item in the first column. (ii) The “DGP $h(y)$” row (row 1) specifies the quality kernel used in the data generating process. (iii) The “Estimation” row (row 2) specifies the estimation method in the estimation. “Our Model, K/S” denotes our model with the first stage propensity score is estimated using Klein-Spady estimator. (iv) Asymptotic standard error estimates are in the parentheses. (v) $D$ is the number of noncensored sample, $N$ is the effective sample size.

6 Empirical Example: Scanner Data with a Multitude of Zero Shares

In this section, we implement our proposed demand estimation method to Dominick’s supermarket cola sales scanner data. The data is obtained through James M. Kilts Center for Marketing,
University of Chicago Booth School of Business. The data contains weekly pricing and sales information of the Dominick’s chain stores during 1989-1997, of every UPC-level product in 29 product categories. Promotion statuses and profitability by each unit sold are also recorded in the data. One shortcoming of the data is that there is no systemic record of the product characteristics, which we overcome by choosing cola sales data and hand-coding the product characteristics.

6.1 Data

We choose the Dominick’s data since it is ideal for our purpose of illustration, for the following two reasons. First, Dominick’s data contains the information on which products were displayed in the shelves, even if the product was not sold at all. This feature is necessary for our purpose because we want the exact information of products which were in the consumer’s choice set but not chosen at all. Indeed, as presented in Figure 6.1, about 1/4 of the observations exhibit zero shares. Second, Dominick’s data contains the information of average profit per each unit sold. Combined with the price data, we can back out the average cost per unit. The cost information is very useful, because an ideal instrument for prices in estimating the consumer demand should be a proxy of cost shocks. We can avoid constructing instruments using indirect proxies for the cost, which has been one of the prevailing problems in the demand estimation literature.

We restrict our interest to the cola sales for a few reasons. Cola market is a typical market of product differentiation in which many brands are competing with different tastes and packages. Among them two prominent brands, Coke and Pepsi, take the majority of market shares. Next, the product characteristics are not separately coded in Dominick’s data, but only the category information such as “soft drinks” or “bottled juices” are available. We had to extract the information from product descriptions truncated at 30 characters, for which cola is ideal because it has clearly labeled product characteristics. Another reason is that the companies producing cola and the product characteristics of cola have not been changed much for the last few decades. Coke and Pepsi have been two leaders in the market. Diet, cherry-flavored, caffeine-free colas are still sold in the market with considerable market shares in 2015, as well as in 1995. This feature makes our
analysis not only convenient, but also the implication of the analysis more realistic. In Appendix B.3 we also present the estimation results for laundry detergent demand as a robustness check.

Dominick’s data covers 100 chain stores around 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 use the cross-section data of a week because we wanted to avoid the following potential problems. First, stockpiling is very common for the products lasting for more than weeks. The same consumer will respond more sensitively in terms of the purchasing behavior than in terms of the consumption behavior. Thus, if we use the data across time series, the estimated elasticities can be larger than it actually is. Next, the demand of soft drinks fluctuate in a week with holiday or events such as Super-bowl, and varies considerably by seasons. Thus we chose one week in March, without any close holidays. Because Dominick’s experimented with the prices and sales across different chain stores even for a same product, we still have enough price variations after choosing a cross-section of data. Even after restricting the sample to a cross-section of one week, the sample size is as large as 4300. We present the summary statistics in Table 2.

We define the individual product and the individual market in a natural way. An individual product is defined by UPC, and a market is defined by store-week pair. It is the finest way to define the product and the market that our data allows, which produces a multitude of zero shares. As illustrated in Figure 6.1 around one third of the products in the shelves were not sold at all.
Figure 6.1: Histogram of Observed Market Shares
Table 2: Descriptive Statistics

(a) Summary of Product Characteristics in 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 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| 160.202| -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 Non-censored 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) The data is the cross-section of week 391 (03/06/1997-03/12/1997) in the Dominick’s scanner data. (ii) Dominick’s recorded the price and cost as zero if the sales of the product is zero in the corresponding week. The mean and median in Sub-Table (b) are calculated including those zeros.

A few remarks on the data are in order. First, we converted the unit price and unit costs to per-ounce price and costs, respectively. Next, Dominick’s did not record the price and cost of the week if the sales of the product is zero in the corresponding week. It is the reason why we cannot include the prices in $w_{j,t}$, and proceed only with other exogenous product characteristics. Also, in estimating the logit model while substituting the observed zero shares with small numbers on the observed zero shares, we had to impute the missing prices using other chain store’s prices and
profits with the same product and promotion status. Lastly, we had to compute the market share of the outside option, both for our model and logit model.\footnote{Although it is not necessary for our model to include the numeraire in the consumer’s choice set, we included it because we wanted to compare the estimation results of our model to those of the logit model in the exactly same setup.} It boils down to estimate the market size. We assume that an average person consumes 100 ounces of soft drinks a week\footnote{On average, Americans consume around 45 gallons of soft drinks a year. Source: http://adage.com/article/news/consumers-drink-soft-drinks-water-beer/228422/} and compute the size of the market using the daily customer count data of each stores.

### 6.2 Estimation

We estimate the model using the method we proposed in Section \[4\]. We also estimate our model with using Probit as the first-stage estimator for comparison. As a benchmark, we estimate the logit model, with different ways to handle the observed zero shares: (i) dropping the sample with zero shares, and (ii) substituting zero shares with small numbers.

We estimate two models with different specifications. In the baseline model (Model 1), \(x_{j,t}\) includes the following product characteristics: bottle size, number of bottles per bundle, diet, caffeine free, cherry flavor, Coke/Pepsi brand dummies, and the promotion status. As an instrument for per-ounce price, we used the per-ounce cost calculated from the profitability variable. For Model 1, we used the store-level demographics for the variables satisfying the exclusion restriction in the customer’s buy-or-not decision in \(w_{j,t}\): % blacks and Hispanics, % college graduates, and log of median income. The exclusion assumption on these variables is that a consumer who never buys a certain product will not become an inframarginal consumer no matter what the other product characteristics are. For Model 2, we excluded promotion status from \(x_{j,t}\), and used it as the variable satisfying the exclusion restriction. The exclusion assumption is that the promotion affects only the consumers’ information about the choice set, not the level of utility by consuming a certain product.

To compare the magnitude of coefficients, we use the logarithm of prices in the logit model as well. As emphasized before, we can take it as a scale adjustment to use the logarithm of prices in
estimating the discrete choice logit model of demand. In Appendix B.1, we provide the robustness checks using the data of different weeks.

The first-stage parameter estimation result for $\hat{\delta}$ is presented in Table 3. Model 1 is the baseline model with store-level demographics in the first-stage. Model 2 uses the promotion statuses as the excluded variables in the second stage estimation. For both Model 1 and Model 2, we estimated the Probit model for benchmark and for setting the initialization value for the optimizer to estimate the Klein-Spady model. The coefficients for the Bottle size are normalized to one. We find that the 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 6.2. The estimated densities of $\eta_{j,t}$ given $\mathbf{w}_{j,t}$ is not even unimodal, which we suggest as a strong evidence that the unobservable product characteristics $\eta_{j,t}$ does not follow the Gaussian distribution.

The main estimation result is presented in Table 4, which is striking. In logit models, the coefficients of the logarithm of prices are positive, economically and statistically very significant, even after instrumenting for the prices using the supplier side cost information. However, log-linear estimation of our model with Klein-Spady first stage estimator returns the expected signs and magnitudes for the coefficients of the logarithm of prices. It is also surprising that the estimator with Probit first-stage estimator performs well, even though the distribution of the unobservables are far from the Gaussian distribution. However, we note that we are not sure whether this result can be generalized.

The result provides a very strong evidence of the selection process which has been ignored in the demand estimation literature thus far. Recall the estimation equation (4.2) under the exponential kernel:

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

Except for the term $E [\xi_{j,t} \mid \mathbf{z}_{j,t}, \mathbf{w}_{j,t}, \pi_{j,t} > 0]$, the estimation equation is the same with the logit model when we drop the sample with zero observed shares. That is, Column 1 (Our Model, K/S, Model 1) and Column 5 (Logit Model, Drop 0) should exactly coincide when the term $E [\xi_{j,t} \mid \mathbf{z}_{j,t}, \mathbf{w}_{j,t}, \pi_{j,t} > 0]$ is zero. Yet, it turns out that it is not the case. As shown in Table 4, ig-
noring the consumer’s selection process can bias the estimates, even resulting in an upward-sloping demand curve. It lines up with the intuition of [Heckman (1979)]: when selection is ignored the whole estimates from a regression can be misleading. To be specific, in our case $E[\xi_{j,t}|z_{j,t}, w_{j,t}, \pi_{j,t} > 0]$ is likely to be positive. This is because consumers will select on unobservables $\eta_{j,t}$ as well as observables $w_{j,t}$, and $\eta_{j,t}$ is highly correlated with $\xi_{j,t}$. Hence, even after instrumenting for the prices, the price coefficient estimates are likely to be biased upward when the samples with zero shares are just dropped. Imputing some small numbers on zero shares may cause a more serious problem, that the direction of the bias cannot be predicted. In contrast to Table 1 in the previous section, we observe that imputing zeros cause the upward bias in the price coefficient estimates in Table [4]. Indeed, we really do not have an explanation on the direction of the bias when zeros are imputed.

In Appendix [B] we also present the estimation results for the cola data from a different week, and for the laundry detergent data. All the results in Appendix [B] exhibit the same pattern as in Table 4, which suggest that our findings are robust.

Figure 6.2: Estimated Densities of $\eta_{j,t}|w_{j,t}$

Note. (i) Klein-Spady model only identifies the distribution of unobservables up to location and scale. Thus, we normalized the $E[\eta_{j,t}|w_{j,t}] = 0$, and $Var(\eta_{j,t}|w_{j,t}) = 4$. (ii) Standard normal density is plotted for comparison.
Table 3: First-stage Parameter Estimates $\hat{\delta}$

| $w_{j,t}$       | Model 1 Probit | Klein-Spady | Model 2 Probit | Klein-Spady |
|-----------------|----------------|-------------|----------------|-------------|
| Bottle Size     | 1 (0.136)      | 1 (0.112)   | 1 (0.136)      | 1 (0.385)   |
| # Bottles per Bundle | -1.451 (0.339) | -6.651 (0.452) | -1.431 (0.339) | -60.463 (15.740) |
| Diet            | 18.188 (4.655) | 3.461 (2.457) | 18.342 (4.659) | 6.272 (19.005) |
| Caffeine Free   | -13.618 (5.597) | -13.877 (2.703) | -13.619 (5.598) | -16.736 (21.370) |
| Cherry          | 53.508 (11.972) | -8.519 (4.396) | 53.620 (11.985) | 9.272 (28.527) |
| Diet            | 41.798 (9.177) | 42.069 (5.124) | 41.948 (9.173) | 111.836 (24.275) |
| Coke            | 79.788 (5.463) | 68.677 (5.768) | 79.989 (5.467) | 142.978 (21.801) |
| Pepsi           | 135.292 (6.587) | 123.745 (9.117) | 135.448 (6.590) | 160.924 (14.946) |
| Promo           | % Blacks and Hispanics | -26.631 (16.784) | -1.326 (9.578) | - |
| % College Graduates | -3.585 (26.250) | 1.940 (10.052) | - |
| Log Median Income | 7.311 (13.962) | 1.659 (43.387) | - |
| $D$             | 3226           | 3226        | 3226           | 3226        |
| $N$             | 4356           | 4356        | 4356           | 4356        |

Note. (i) $D$ is the number of nonzero share observations and $N$ is the sample size. (ii) Asymptotic standard error estimates are in the parentheses. (iii) Unit of bottle size is liquid ounces. (iv) We normalized the coefficients of Bottle Size to one.
the estimates of the price coefficients can be misleading not only in their magnitudes but also in
set, which has been ignored in the literature so far. If the choice set selection problem is ignored,
we have introduced. The key breakthrough we made was to account for the selection of the choice
the intensive margin and extensive margin, and embedding them both in the quality kernel that
it can accommodate the zero choice probabilities with a reasonable microfoundation by separating
preferences with observed and unobserved product characteristics. Our framework is compatible
In this paper, we have developed a semiparametric demand estimation framework based on CES
preferences with observed and unobserved product characteristics. Our framework is compatible
with the widely used logit model of demand based on the discrete choice framework. Furthermore,
it can accommodate the zero choice probabilities with a reasonable microfoundation by separating
the intensive margin and extensive margin, and embedding them both in the quality kernel that
we have introduced. The key breakthrough we made was to account for the selection of the choice
set, which has been ignored in the literature so far. If the choice set selection problem is ignored,
the estimates of the price coefficients can be misleading not only in their magnitudes but also in

### Table 4: Second-stage Parameter Estimates ($\hat{\delta}, \hat{\beta}$)

| $(\phi_{j,t}, x_{i,t})$ | Our Model, K/S | Our Model, Probit | Logit Model |
|-------------------------|----------------|------------------|-------------|
|                         | Model 1 | Model 2 | Model 1 | Model 2 | Drop 0 | $10^{-8}$ | $10^{-4}$ |
| Log Price ($-\sigma$)   | -1.528 | -2.414 | -1.499 | -2.359 | 4.656 | 7.042 | 4.519 |
|                         | (0.051) | (0.198) | (0.039) | (0.079) | (0.094) | (0.143) | (0.069) |
| Bottle Size             | 0.012  | 0.015  | 0.013  | 0.018  | 0.110 | 0.174 | 0.103 |
|                         | (0.001) | (0.002) | (0.001) | (0.002) | (0.004) | (0.006) | (0.003) |
| # Bottles per Bundle    | 0.085  | 0.127  | 0.092  | 0.119  | 0.336 | 0.423 | 0.302 |
|                         | (0.003) | (0.005) | (0.005) | (0.007) | (0.012) | (0.019) | (0.009) |
| Diet                    | -0.079 | -0.130 | -0.041 | 0.024  | 0.878 | 1.955 | 0.844 |
|                         | (0.047) | (0.052) | (0.055) | (0.067) | (0.113) | (0.187) | (0.091) |
| Caffeine Free           | -1.077 | -0.944 | -1.136 | -0.868 | -2.235 | -3.228 | -1.961 |
|                         | (0.054) | (0.062) | (0.082) | (0.091) | (0.126) | (0.209) | (0.101) |
| Cherry                  | -2.240 | -2.340 | -2.059 | -1.983 | -3.531 | -5.451 | -2.943 |
|                         | (0.074) | (0.065) | (0.099) | (0.089) | (0.329) | (0.485) | (0.235) |
| Coke                    | 0.216  | 0.409  | 0.134  | 0.348  | 0.533 | 0.087 | 0.118 |
|                         | (0.059) | (0.051) | (0.066) | (0.076) | (0.242) | (0.361) | (0.175) |
| Pepsi                   | 1.822  | 1.583  | 1.785  | 1.838  | 4.397 | 8.500 | 4.177 |
|                         | (0.060) | (0.089) | (0.072) | (0.107) | (0.164) | (0.244) | (0.118) |
| Promo                   | 1.150  | -      | 1.086  | -      | 0.886 | 3.373 | 1.122 |
|                         | (0.060) | (0.072) | -      | (0.072) | (0.129) | (0.208) | (0.101) |
| $D$                     | 3226   | 3226   | 3226   | 3226   | 3226  | 4090  | 4090   |
| $N$                     | 4356   | 4356   | 4356   | 4356   | 3226  | 4090  | 4090   |

Note. (i) For Columns “Our Model, K/S” and “Our Model, Probit,” results from Klein-Spady and Probit estimators in Table 3 are used for the first-stage estimator, respectively. (ii) $D$ is the effective number of nonzero share observations, $N$ is the effective sample size. (iii) Asymptotic standard error estimates are in the parentheses. (iv) Unit of bottle size is liquid ounces. (v) Because Dominick’s did not record the price and cost when sales is zero, in estimating “$10^{-8}$” column and “$10^{-4}$” column, we had to use the average prices and costs of the same product with same promotion statuses of other stores.

### 7 Conclusion

In this paper, we have developed a semiparametric demand estimation framework based on CES
preferences with observed and unobserved product characteristics. Our framework is compatible
with the widely used logit model of demand based on the discrete choice framework. Furthermore,
it can accommodate the zero choice probabilities with a reasonable microfoundation by separating
the intensive margin and extensive margin, and embedding them both in the quality kernel that
we have introduced. The key breakthrough we made was to account for the selection of the choice
set, which has been ignored in the literature so far. If the choice set selection problem is ignored,
the estimates of the price coefficients can be misleading not only in their magnitudes but also in
their signs. We have shown that it can even result in an upward sloping demand curve.

A direct extension of our work will be random coefficients demand model which can accommodate the zero choice probabilities. As in the homogenous logit model, the own and cross price elasticities derived from our model also exhibit the unrealistic substitution pattern. Overcoming such unrealistic substitution patterns was one of the most important motivations on the development of the random coefficients logit model of demand by Berry et al. (1995). Although we provide the microfoundation for the random coefficients, we did not develop the estimation method with random coefficients accommodating the zero individual choice probabilities. We leave such an extension as a future research agenda.
References

ACKERBERG, D., C. L. BENKARD, S. BERRY, AND A. PAKES (2007): *Econometric tools for analyzing market outcomes*, Elsevier, vol. 6A of Handbook of Econometrics, chap. 63, 4171–4276.

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

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.

——— (1989): “Demand for differentiated products, discrete choice models, and the characteristics approach,” *Review of Economic Studies*, 56, 21–35.

——— (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.

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.

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.

Broda, C. and D. E. Weinstein (2006): “Globalization and the gains from variety,” *Quarterly Journal of Economics*, 121, 541–585.

Bronnenberg, B. J. (2015): “The provision of convenience and variety by the market,” *RAND Journal of Economics*, 46, 480–498.

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 Science*, 12, 184–208.

Dixit, A. K. and J. E. Stiglitz (1977): “Monopolistic competition and optimum product diversity,” *American Economic Review*, 67, 297–308.

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.

Feenstra, R. C. (1994): “New product varieties and the measurement of international prices,” *American Economic Review*, 84, 157–177.
GANDHI, A., Z. LU, AND X. SHI (2013): “Estimating demand for differentiated products with error in market shares,” Working Paper.

HANDBURY, J. (2013): “Are poor cities cheap for everyone?: Non-homotheticity and the cost of living across U.S. cities,” Working Paper.

HANDBURY, J. AND D. E. WEINSTEIN (2014): “Goods prices and availability in cities,” *Review of Economic Studies*, 82, 258–296.

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 Economics*, 123, 441–487.

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.

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.

LI, N. (2013): “An Engel curve for variety,” Working Paper.

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

NEVO, A. (2000): “A practitioner’s guide to estimation of random-coefficients logit models of demand,” *Journal of Economics and Management Strategy*, 9, 513–548.
—— (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.

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.

Reiss, P. C. and F. A. Wolak (2007): *Structural econometric modeling: Rationales and examples from industrial organization*, Elsevier, vol. 6A of *Handbook of Econometrics*, chap. 64, 4277–4415.

Santos Silva, J. M. C. and S. Tenreyro (2006): “The log of gravity,” *Review of Economics and Statistics*, 88, 641–658.

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

Verboven, F. (1996): “The nested logit model and representative consumer theory,” *Economics Letters*, 50, 57–73.
A Derivation of Selection-Correction Estimation Equation from the Two-stage Discrete Choice Models

In this section, we show that the estimation equation (4.2) can be derived from a two-stage decision process within a discrete choice random utility framework. Consider a representative consumer with two-stage decision process. In the first stage, the consumer makes a search over $\mathcal{J}_t$, which is all the possible alternatives. The consumer’s choice set $\mathcal{J}_t^+$ is determined as a result of the search. In the second stage, the consumer encounters the usual discrete-choice decision problem over $\mathcal{J}_t^+$, i.e. purchase the product that yields the highest utility.

Let $(w_{j,t}, \eta_{j,t})$ be the variables that affect the first-stage choice set search, $(x_{j,t}, \xi_{j,t})$ be the variables that affect the second-stage discrete choice unconstrained utility maximization problem. Notice that these form an analogue of the notations we used in Sections 3-4. Let us model the threshold decision rule of the search for the choice set as $d_{j,t} = 1 \left\{ \gamma + w'_{j,t} \delta + \eta_{j,t} > 0 \right\}$, where $d_{j,t} = 1$ denotes the product $j$ in market $t$ is contained in the consumer’s choice set $\mathcal{J}_t^+$. The second stage utility of the consumer is modeled as

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

The representative consumer solves

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

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

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

In the estimation, $\Pr (i \to j | t)$ is equalized with the market shares $s_{j,t}$. The inversion theorem of Berry (1994); Berry et al. (1995) applies. The only difference is again the moment condition; $E \left[ \xi_{j,t} | z_{j,t}, \mathcal{J}_t^+ \right]$ is not zero, and it is highly likely to be positive. Thus, there has to be a correction
term for this selection on the choice set, which leads us to the following estimation equation

\[
E \left[ \ln \left( \frac{s_{j,t}}{s_{0,t}} \right) | z_{j,t}, w_{j,t}, J_t^+ \right] = -\alpha \ln p_{j,t} + x_j' \beta \theta + E \left[ \xi_{j,t} | z_{j,t}, w_{j,t}, J_t^+ \right]. \tag{A.1}
\]

(A.1) coincides with (4.2).
B Robustness Checks

B.1 Cola Demand for Week 382

In this subsection, we estimate the same models as in Section 6, with the data from a different week. We use the data of week 382 (from January 1 to 9, 1997). In Tables 5 and 6, we repeat the estimation procedure for Tables 3 and 4.

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

| $w_{j,t}$                  | Model 1          |               | Model 2          |               |
|---------------------------|------------------|---------------|------------------|---------------|
|                           | Probit Klein-Spady| Probit Klein-Spady | Probit Klein-Spady | Probit Klein-Spady |
| Bottle Size               | 1                | (0.151)       | 1                | (0.153)       |
|                           |                 | (0.420)       |                 | (0.403)       |
| # Bottles per Bundle      |                  | (0.404)       |                 | (0.410)       |
|                           |                 | (21.313)      |                 | (19.527)      |
| Diet                      |                  | (4.823)       |                 | (4.894)       |
|                           |                 | (21.925)      |                 | (21.039)      |
| Caffeine Free             |                  | (5.787)       |                 | (5.870)       |
|                           |                 | (26.388)      |                 | (25.227)      |
| Cherry                    |                  | (12.395)      |                 | (12.544)      |
|                           |                 | (51.599)      |                 | (48.965)      |
| Coke                      |                  | (70.200)      |                 | (70.672)      |
|                           |                 | 45.037        |                 | 40.009        |
| Pepsi                     |                  | (47.380)      |                 | (48.250)      |
|                           |                 | 54.267        |                 | 56.447        |
| Promo                     |                  | (5.486)       |                 | (5.568)       |
|                           |                 | (23.366)      |                 | (22.543)      |
|                           |                  | (123.783)     |                 | (126.415)     |
|                           |                 | 34.220        |                 | 36.507        |
| % Blacks and Hispanics    |                  | (6.643)       |                 | (6.741)       |
|                           |                 | (19.433)      |                 | (18.391)      |
|                           | 16.697           | 7.835         |                 |               |
| % College Graduates       |                  | (17.689)      |                 | (100.061)     |
|                           |                 | (100.061)     |                 | -             |
|                           |                  | (27.114)      |                 | (97.722)      |
| Log Median Income         |                  | (50.849)      |                 | (38.996)      |
|                           |                  | (14.406)      |                 | (439.842)     |
|                           |                  |               |                 |               |
| $D$                       | 3226             | 3226          | 3226             | 3226          |
| $N$                       | 4337             | 4337          | 4337             | 4337          |

Note. (i) $D$ is the number of nonzero share observations and $N$ is the sample size. (ii) Asymptotic standard error estimates are in the parentheses. (iii) Unit of bottle size is liquid ounces. (iv) We normalized the coefficients of Bottle Size to one.
Table 6: Second-stage Parameter Estimates ($\hat{\sigma}, \hat{\beta}$)

| ($\phi_{j,t}, x_{j,t}$) | Our Model, K/S | Our Model, Probit | Logit Model |
|--------------------------|----------------|------------------|-------------|
|                          | Model 1        | Model 2          |              | Drop 0 | $10^{-8}$ | $10^{-4}$ |
| **Log Price ($-\sigma$)** | -1.394         | -1.120           | -1.416       | -1.082 | 4.384     | 5.299     | 4.040     |
|                          | (0.072)        | (0.085)          | (0.073)      | (0.102) | (0.131)   | (0.166)   | (0.085)   |
| **Bottle Size**          | 0.003          | 0.012            | 0.002        | 0.017  | 0.104     | 0.122     | 0.091     |
|                          | (0.001)        | (0.002)          | (0.001)      | (0.002) | (0.005)   | (0.007)   | (0.004)   |
| **# Bottles per Bundle** | -0.013         | 0.032            | -0.012       | 0.026  | 0.226     | 0.114     | 0.170     |
|                          | (0.005)        | (0.005)          | (0.004)      | (0.003) | (0.017)   | (0.022)   | (0.011)   |
| **Diet**                 | -0.202         | -0.167           | -0.227       | -0.160 | 1.164     | 1.457     | 0.921     |
|                          | (0.047)        | (0.057)          | (0.050)      | (0.067) | (0.135)   | (0.191)   | (0.098)   |
| **Caffeine Free**        | -1.142         | -0.998           | -1.107       | -1.005 | -1.805    | -2.990    | -1.682    |
|                          | (0.050)        | (0.059)          | (0.048)      | (0.060) | (0.137)   | (0.200)   | (0.102)   |
| **Cherry**               | -1.704         | -1.938           | -1.484       | -1.952 | -3.056    | -6.164    | -2.887    |
|                          | (0.040)        | (0.039)          | (0.069)      | (0.079) | (0.412)   | (0.478)   | (0.244)   |
| **Coke**                 | 1.158          | 1.288            | 0.962        | 1.301  | 3.244     | 5.440     | 2.778     |
|                          | (0.053)        | (0.060)          | (0.085)      | (0.098) | (0.260)   | (0.357)   | (0.183)   |
| **Pepsi**                | 0.795          | 0.807            | 0.702        | 0.838  | 3.518     | 5.470     | 2.987     |
|                          | (0.063)        | (0.077)          | (0.075)      | (0.098) | (0.199)   | (0.262)   | (0.134)   |
| **Promo**                | 1.170          |                 | 1.138        |        | 0.643     | 2.414     | 0.736     |
|                          | (0.070)        |                 | (0.065)      |        | (0.141)   | (0.207)   | (0.106)   |
| **D**                    | 3226           | 3226             | 3226         | 3226   | 3226      | 4118      | 4118      |
| **N**                    | 4337           | 4337             | 4337         | 4337   | 3226      | 4118      | 4118      |

Note. (i) For Columns “Our Model, K/S” and “Our Model, Probit,” results from Klein-Spady and Probit estimators in Table 5 are used for the first-stage estimator, respectively. (ii) $D$ is the effective number of nonzero share observations, $N$ is the effective sample size. (iii) Asymptotic standard error estimates are in the parentheses. (iv) Unit of bottle size is liquid ounces. (v) Because Dominick’s did not record the price and cost when sales is zero, in estimating “$10^{-8}$” column and “$10^{-4}$” column, we had to use the average prices and costs of the same product with same promotion statuses of other stores.
We repeat the estimation procedure for the cola data of week 278 (from January 5 to 11, 1995).

It turns out that all the results show the same pattern as in the previous sections. Although not tabulated here, we also examined the data from many other weeks and the results were robust.

### Table 7: First-stage Parameter Estimates \( \hat{\delta} \)

| \( w_{j,t} \) | \multicolumn{2}{c}{Model 1} | \multicolumn{2}{c}{Model 2} |
|---------------|-----------------|-----------------|-----------------|-----------------|
| \multicolumn{2}{c}{Probit} | \multicolumn{2}{c}{Klein-Spady} | \multicolumn{2}{c}{Probit} | \multicolumn{2}{c}{Klein-Spady} |
| Bottle Size   | 1               | 1               | 1               | 1               |
| \( j \)       | (0.133)         | (0.109)         | (0.134)         | (0.108)         |
| \( t \)       | -5.266          | -10.812         | -5.334          | -10.733         |
| \# Bottles per Bundle | (0.459) | (0.738) | (0.463) | (0.731) |
| Diet          | -75.269         | -0.857          | -76.148         | -0.821          |
| \( j \)       | (6.087)         | (2.455)         | (6.152)         | (2.436)         |
| \( t \)       | 74.971          | 6.079           | 74.187          | 6.033           |
| Caffeine Free | (8.504)         | (3.415)         | (8.529)         | (3.366)         |
| Cherry        | -222.735        | -16.707         | -223.705        | -16.529         |
| \( j \)       | (39.064)        | (7.919)         | (39.390)        | (7.857)         |
| \( t \)       | 151.601         | 12.249          | 151.950         | 12.323          |
| Coke          | (19.153)        | (5.315)         | (19.269)        | (5.270)         |
| \( j \)       | 17.957          | 7.703           | 18.408          | 7.791           |
| \( t \)       | (6.146)         | (2.620)         | (6.213)         | (2.598)         |
| Pepsi         | 302.445         | 23.821          | 305.655         | 23.798          |
| Promo         | (8.172)         | (1.651)         | (8.242)         | (1.628)         |
| % Blacks and Hispanics | 3.354 | -2.569 | - | - |
| \( j \)       | (21.057)        | (10.593)        | -               | -               |
| % College Graduates | -96.761 | -1.464 | - | - |
| \( j \)       | (32.885)        | (11.688)        | -               | -               |
| \( t \)       | 46.202          | 4.689           | -               | -               |
| Log Median Income | (17.394) | (45.986) | - | - |

Note. (i) \( D \) is the number of nonzero share observations and \( N \) is the sample size. (ii) Asymptotic standard error estimates are in the parentheses. (iii) Unit of bottle size is liquid ounces. (iv) We normalized the coefficients of Bottle Size to one.
Table 8: Second-stage Parameter Estimates \( (\hat{\sigma}, \hat{\beta}) \)

| \( (\phi_{j,t}, x_{j,t}) \) | Our Model, K/S | Our Model, Probit | Logit Model |
|---|---|---|---|
| | Model 1 | Model 2 | Model 1 | Model 2 | Drop 0 | \(10^{-8}\) | \(10^{-4}\) |
| Log Price \( (-\sigma) \) | -1.106 | -1.088 | -1.013 | -0.997 | 23.669 | 22.774 | 18.042 |
| | (0.068) | (0.066) | (0.059) | (0.056) | (5.078) | (3.272) | (2.592) |
| Bottle Size | 0.007 | 0.007 | 0.013 | 0.013 | 0.869 | 0.857 | 0.681 |
| | (0.002) | (0.002) | (0.001) | (0.001) | (0.197) | (0.135) | (0.107) |
| # Bottles per Bundle | 0.043 | 0.041 | 0.055 | 0.053 | 2.271 | 2.200 | 1.795 |
| | (0.004) | (0.004) | (0.003) | (0.003) | (0.512) | (0.357) | (0.283) |
| Diet | -0.072 | -0.129 | -0.066 | -0.130 | 4.408 | 4.422 | 3.243 |
| | (0.043) | (0.051) | (0.053) | (0.040) | (1.336) | (0.979) | (0.775) |
| Caffeine Free | -1.306 | -1.327 | -1.247 | -1.301 | -5.282 | -6.027 | -5.408 |
| | (0.041) | (0.040) | (0.066) | (0.060) | (1.025) | (0.972) | (0.770) |
| Cherry | -3.216 | -3.250 | -3.051 | -3.085 | -3.918 | -3.949 | -3.887 |
| | (0.090) | (0.091) | (0.356) | (0.337) | (7.265) | (6.981) | (5.529) |
| Coke | 1.794 | 1.783 | 1.716 | 1.679 | 3.923 | 6.317 | 4.372 |
| | (0.044) | (0.046) | (0.160) | (0.165) | (1.437) | (1.325) | (1.050) |
| Pepsi | 1.595 | 1.591 | 1.541 | 1.542 | 13.673 | 16.056 | 12.862 |
| | (0.065) | (0.065) | (0.045) | (0.045) | (3.222) | (2.693) | (2.133) |
| Promo | -0.171 | -0.134 | - | - | 18.601 | 14.647 | 10.021 |
| | (0.084) | (0.059) | - | - | (4.562) | (2.177) | (1.724) |
| \( D \) | 3667 | 3667 | 3667 | 3667 | 3667 | 4069 | 4069 |
| \( N \) | 5185 | 5185 | 5185 | 5185 | 3667 | 4069 | 4069 |

Note. (i) For Columns “Our Model, K/S” and “Our Model, Probit,” results from Klein-Spady and Probit estimators in Table 7 are used for the first-stage estimator, respectively. (ii) \( D \) is the effective number of nonzero share observations, \( N \) is the effective sample size. (iii) Asymptotic standard error estimates are in the parentheses. (iv) Unit of bottle size is liquid ounces. (v) Because Dominick’s did not record the price and cost when sales is zero, in estimating “\(10^{-8}\)” column and “\(10^{-4}\)” column, we had to use the average prices and costs of the same product with same promotion statuses of other stores.
B.3 Laundry Detergent Demand for Week 375

In this subsection, we estimate the demand for laundry detergent using the same Dominick’s data as in Section 6. We randomly picked a cross section of week 375, which is the third week of November 1996. The product is defined by UPC, and the market is defined by the week-store pair.

We computed the market shares by loads. There are two types of laundry detergents: liquid and powder. We converted the size of canister by the following criteria. For liquid laundry detergent, 1.6 ounces are counted as 1 load. For powder laundry detergent, 2.3 ounces are counted as 1 load. Some of the powder detergent used pound instead of ounces. For such product, 0.08262 pound is counted as 1 load. We also converted the size of a package. Since the density of powder detergents are around 0.65g/cm³ and 1g/cm³ = 0.065198lb/oz, 1 pound of powder detergent is approximately 23.6 ounces. The market size is computed by assuming each consumer visiting the store consumes 6 loads of laundry detergent a week. Other details about the data are similar to what we described in Section 6.

Table 9 presents the first-stage estimators, and Table 10 presents the second-stage estimators. We find the same pattern as in Table 4: even after instrumenting for the prices, the estimated demand curve is upward sloping when the selection is not considered in the estimation.

\[^{12}\text{It happened only for Arms and Hammers powder detergent.}\]
Table 9: First-stage Parameter Estimates $\hat{\delta}$

| $w_{j,t}$                      | Probit  | Klein-Spady |
|--------------------------------|---------|-------------|
| Package Size                   | $(0.092)$ | $(0.037)$   |
| Liquid                         | $379.843$ | $-23.448$   |
|                                | $(15.881)$ | $(1.709)$   |
| Heavy duty / Concentrated / Double | $114.863$ | $-7.821$    |
|                                | $(47.615)$ | $(4.610)$   |
|                                | $-6.143$  | $-4.429$    |
| Bleach                         | $(18.979)$ | $(1.813)$   |
|                                | $230.570$ | $-2.277$    |
|                                | $(20.329)$ | $(1.829)$   |
|                                | $31.894$  | $-6.435$    |
| Wisk                           | $(27.071)$ | $(2.124)$   |
| Ajax / Arm&Hammer / Surf / Purex | $-240.999$ | $6.229$    |
|                                | $(22.170)$ | $(1.883)$   |
|                                | $18.464$  | $-3.177$    |
| % Blacks and Hispanics         | $(58.225)$ | $(5.564)$   |
| % College Graduates            | $242.358$  | $2.795$     |
|                                | $(91.135)$ | $(9.021)$   |
| Log Median Income              | $-9.221$  | $-1.292$    |
|                                | $(48.219)$ | $(4.762)$   |

$D$                                | $7177$  | $7177$    |
$N$                                | $14999$ | $14999$  |

Note. (i) $D$ is the number of nonzero share observations and $N$ is the sample size. (ii) Asymptotic standard error estimates are in the parentheses. (iii) Unit of package size is converted to liquid ounces. (iv) We normalized the coefficients of Package Size to one.
Table 10: Second-stage Parameter Estimates \((\hat{\sigma}, \hat{\beta})\)

| \((\phi_{j,t}, x_{j,t})\) | Our Model | Logit Model |
|-----------------------------|-----------|-------------|
|                             | First Stage: K/S | Probit | Zero: Drop 0 | \(10^{-8}\) | \(10^{-4}\) |
| Log Price \((-\sigma)\)     | -1.122 (0.050) | -0.960 (0.045) | 6.455 (0.095) | 9.755 (0.138) | 6.371 (0.064) |
| Package Size                | 0.001 (0.000) | 0.002 (0.000) | 0.030 (0.001) | 0.038 (0.001) | 0.026 (0.001) |
| Liquid                      | 0.398 (0.049) | 0.545 (0.039) | 6.551 (0.153) | 9.990 (0.228) | 6.020 (0.106) |
| Heavy duty / Concentrated / Double | -0.378 (0.050) | -0.380 (0.090) | 4.218 (0.249) | 7.803 (0.467) | 4.108 (0.217) |
| Bleach                      | 0.085 (0.108) | -0.088 (0.075) | 0.575 (0.101) | 0.837 (0.172) | 0.720 (0.080) |
| Tide                        | 1.049 (0.045) | 1.111 (0.098) | 0.796 (0.099) | 2.055 (0.175) | 0.578 (0.081) |
| Wisk                        | 0.014 (0.025) | -0.105 (0.050) | -0.613 (0.137) | -1.118 (0.229) | -0.742 (0.107) |
| Ajax / ArmHammer / Surf / Purex | -0.021 (0.072) | 0.041 (0.025) | 1.715 (0.133) | 3.180 (0.226) | 2.034 (0.105) |
| \(D\)                      | 7177        | 7177        | 7177        | 10639       | 10639       |
| \(N\)                      | 14999       | 14999       | 7177        | 10639       | 10639       |

Note. (i) For Columns “Our Model, K/S” and “Our Model, Probit,” results from Klein-Spady and Probit estimators in Table 9 are used for the first-stage estimator, respectively. (ii) \(D\) is the effective number of nonzero share observations, \(N\) is the effective sample size. (iii) Asymptotic standard error estimates are in the parentheses. (iv) Unit of package size is converted to liquid ounces. (v) Because Dominick’s did not provide with the price and cost data when sales is zero, in estimating “\(10^{-8}\)” column and “\(10^{-4}\)” column, we had to use the average prices and costs of the same product with same promotion statuses of other stores.