Control Function Corrections for Unobserved Factors in Differentiated Product Models

Kyoo il Kim
Michigan State University

Amil Petrin*
University of Minnesota and NBER

March, 2019

Abstract

Unobserved factors in differentiated product models can generate bias in price elasticities. We provide a set of assumptions under which pricing functions can be inverted to obtain controls that condition out the problematic part of the demand error. Our approach extends the literature by addressing endogenous prices while simultaneously allowing for non-separability between observed and unobserved factors or complementarities in demand, cases for which the Berry (1994) correction will not work. Our approach, in practice, is a simple two-step procedure of OLS followed by Maximum Likelihood. We show how to implement our approach on three data sets and demand specifications estimated elsewhere that span a range of markets and levels of aggregation, including automobiles (the original Berry, Levinsohn, and Pakes (1995) application), cable television, and margarine.

Keywords: Differentiated products, Discrete choice demand, Consumer choice, Endogeneity, Control function

*Corresponding authors: Amil Petrin, Department of Economics, University of Minnesota, email: petrin@umn.edu; Kyoo il Kim, Department of Economics, Michigan State University, email: kyookim@msu.edu
1 Introduction

Models of differentiated products are widely used for estimating demand elasticities and substitution patterns. In applications of these models it is rare that all relevant factors are observed by the econometrician. When some factors are unobserved, price will typically be correlated with these unobserved factors through the equilibrating mechanism in the market. For example, products that display desirable attributes observed by consumers and producers but not measured by the econometrician will often have prices that are positively correlated with the demand error. Alternatively, if advertising or other promotional activities are omitted from the specification, and if prices are set simultaneously with these promotional levels, then price will be correlated with the demand error. The problem has arisen in both aggregate (i.e. market-level) data and disaggregate (i.e., customer-level) data, and empirically has tended to bias estimates of price elasticities in a positive direction.

Since demand in differentiated product settings is not linear in price, standard linear methods for correcting this endogeneity problem are not immediately applicable. In this paper we expand on the approach from Petrin and Train (2010), exploiting the information that prices contain on unobserved demand factors. The intuition underlying the correction in non-linear settings is similar to that in a linear environment, where a new control variable is included in the regression to condition out the part of the error that is correlated with the endogenous regressor.\footnote{See the ideas discussed in Telser (1964), which are more formally developed in Heckman (1976), Heckman (1978), and Hausman (1978). It has been applied to a Tobit model by Smith and Blundell (1986) and binary probit by Rivers and Vuong (1988).}

Our main contribution is to provide a set of conditions on demand and supply such that equilibrium pricing functions can be inverted to recover controls which are one-to-one functions with the problematic part of the demand error. Estimation using these variables as controls in the demand equation conditions out the dependence of price on the demand error. Finally, we provide a simple portmanteau specification test which rejects if any of the conditions required for consistency of this control function approach fail to hold.

Our approach has several attractive features. The setup can be used in environments where goods are either substitutes, complements, or both, and in which consumers can choose more than one unit of the good. Estimation is straightforward and can be done in standard programming packages. Finally, our framework does not require additive separability between observed and unobserved factors in consumer utility, so (e.g.) unobserved product characteristics or unobserved promotional activities (like advertising) can affect the marginal impact of price on utility.

We are not the first to look to the information that prices contain on demand unobservables (see e.g. Trajtenberg (1989), Villas-Boas and Winer (1999)). Our approach is perhaps most closely related to Bajari and Benkard (2005). They extend the differentiated products setting of Rosen (1974), providing restrictions on demand and supply such that price is only a function of own-product observed and unobserved demand characteristics. Our approach loosens these restrictions, allowing for supply side factors that are excluded from the demand system to influence the equilibrium pricing function, and permitting consumers to have idiosyncratic tastes for products, so the widely used generalized extreme value (e.g. logit and nested logit) and (multivariate) normal models are
not ruled out.

Our approach also provides an alternative to the solution from Berry (1994). He works in a setting where each good can have an unobserved demand and supply factor. He shows if consumers choose only one good, if goods are all strict substitutes, and if additive separability holds between observed and unobserved demand factors, then “mean utilities” can be inverted out from the observed market shares. Standard instrumental variable (IV) techniques can then be applied to estimate price elasticities because the separability assumption implies that the unobserved demand factor is additive in the mean utility term.

Consistency of his IV approach depends critically upon the separability of price with the unobserved demand factor, an assumption that is hard to motivate with economic theory. When price is not separable from the demand error, the instrumented price is correlated with the error term because it contains price (interacted with the unobserved factor). While our setup is more restrictive than his in that we can allow for only one unobserved factor for each good, we can allow price and the unobserved demand factor to enter consumer utility in a non-separable manner. We can also evaluate all of the maintained assumptions of our framework jointly using a simple \( \chi^2 \) portmanteau specification test.

We present three empirical demand applications that replicate specifications from earlier works - all of which use the Berry (1994) approach - including Berry, Levinsohn, and Pakes (1995) (BLP hereafter), who look at automobiles, Goolsbee and Petrin (2004), who look at cable television, and Chintagunta, Dubé, and Goh (2005), who look at margarine. We choose these applications because uncorrected price elasticities have been shown to suffer from severe bias, and because they span a range of markets and potential competitive pricing behaviors, and include three types of different data: aggregate (market-level) data, household-level cross-sectional data, and household-level panel data. We use them to show in practice how one implements the control function approach.

We choose simple specifications for demand and pricing functions, maintaining additive separability of observed and unobserved factors in both equations. Our simplest formulations for the control function approach are very easy to implement in standard programming packages, as the first stage is ordinary least squares and the second stage is maximum likelihood. We find for these simple specifications that the elasticities are very similar to those obtained using the Berry (1994) correction, and that they differ significantly from the uncorrected elasticity estimates, which are biased in the positive direction.

The paper proceeds as follows. Section 2 sets up a structural demand model and describes omitted variable bias. Section 3 develops our control function approach. Section 4 establishes conditions under which the reduced form pricing function exists and is invertible, and discusses estimation of the controls. Section 5 formally develops our estimator as a sieve estimator, to allow for more flexible specifications, and shows its consistency. Section 6 reports the results from the three applications using real data, and Section 7 concludes. Other details are gathered in the Appendix.
2 Demand with Unobserved Factors

In this section we describe a non-separable model of discrete choice demand, how bias arises from omitted factors, and the implications of non-separable omitted factors for Berry (1994) and BLP (1995). Readers interested in the invertibility results can skip directly to Section 4.

2.1 The Demand Model

We assume there are \( j = 0, 1, \ldots, J_t \) goods in market \( t \) and \( j = 0 \) denotes the outside good. We let \( J_t = \{0, 1, \ldots, J_t\} \) and to simplify exposition we assume \( J_t = J \). A consumer \( i \) purchases good \( j \) at market \( t \) if \( u_{ijt} > u_{ikt} \forall k \neq j \in J_t \). Let utility consumer \( i \) derives from good \( j \) be

\[
u_{ijt} = u(x_{jt}, p_{jt}, \xi_{jt}, \omega_{it}) \text{ for } j = 1, \ldots, J,
\]

with \((x_{jt}, p_{jt})\) denoting observed product characteristics and price and with \( \xi_{jt} \) reflecting unobserved demand factor that affects utility, including unobserved physical characteristics, advertising and promotional activities of which the researcher is unaware, and where \( \omega_{it} \) (possibly a vector) denotes idiosyncratic consumer-specific tastes for observed and unobserved factors. We assume \( \omega_{it} \) is independent of \((x_{jt}, p_{jt}, \xi_{jt})\).

Then the choice probability or predicted market share for good \( j \) at market \( t \) is given by

\[
s_{jt} = \int \{u_{ijt} \geq u_{ikt} \forall k \neq j\} dF_{\omega_{it}} = s_j(\{x_{kt}, p_{kt}, \xi_{kt}\}_{k \in J}),
\]

where \( F_{\omega_{it}} \) denotes the distribution of \( \omega_{it} \) (we abstract from observed individual characteristics here).\(^2\) We will write \( x_t = (x_{1t}, \ldots, x_{Jt}) \), \( p_t = (p_{1t}, \ldots, p_{Jt}) \), \( \xi_t = (\xi_{1t}, \ldots, \xi_{Jt}) \), and \( s_t = (s_{1t}, \ldots, s_{Jt}) \).

Our utility specification in (1) can be very general and include flexible demand models as (but not restricted to) e.g.,

\[
u_{ijt} = \beta'_{it} x_{jt} - \alpha_{it} p_{jt} + \gamma^x_{it} x_{jt} \xi_{jt} + \gamma^p_{it} p_{jt} \xi_{jt} + \varepsilon_{ijt}
\]

where we allow the interactions of observed characteristics (including price) with the unobserved demand factor as well as random coefficients on the unobserved factor and its interactions with observed demand factors. In our notation we let \( \omega_{it} = (\beta_{it}, \alpha_{it}, \gamma^x_{it}, \gamma^p_{it}, \varepsilon_{i0t}, \ldots, \varepsilon_{iJt}) \) and we assume the distribution of \( \omega_{it} \) is known up to unknown parameters, (e.g.,) the random coefficients are distributed as independent normal errors (up to some normalizations if necessary) and \( \varepsilon_{ijt} \) denotes the idiosyncratic logit demand errors (type I extreme value errors).

We, therefore, extend classical differentiated product models (e.g., Berry, Levinsohn, and Pakes (1995)) in two important dimensions. First we can include the interactions (and also higher order terms of price and the unobserved demand factor) because, without them, underlying preferences would have to be strongly separable in the arguments of the utility function, requiring marginal

\(^2\)This is without loss of generality as adding a vector of individual characteristics to the specification simply entails rewriting the utility and share equations.
rates of substitution between observed and unobserved factors to be independent of the levels of consumption of each factor. Second we allow for random coefficients on the unobserved demand factor as well as on its interactions with other factors.

We discuss below that our extension is not trivial and the flexible demand model we study cannot be consistently estimated using other approaches including Berry (1994), BLP (1995), and Berry and Haile (2014).

2.2 Bias from Unobserved Factors

If \( \xi_t \) in part reflects unobserved factors that are not measured by the econometrician, and if sellers charge prices based on these unobserved factors, then \( \xi_{jt} \) and \( p_{jt} \) will not be mean independent. In this case parameter estimates will not be consistent if the choice model maintains that \( \xi_{jt} \) is mean independent of \( (x_{jt}, p_{jt}) \). For example, if consumers’ willingness to pay increases in the unobserved factor, and producers charge more for products with more of the unobserved factor, then the price elasticity bias is likely to go in the positive direction; consumers look less price sensitive than they actually are because they are getting “more” for paying the observed price they pay than the econometrician has taken into account.

Since Trajtenberg (1989)’s finding of upward sloping demand curves for CT scanners, numerous empirical applications have shown that \( \xi_{jt} \) and \( p_{jt} \) can be highly positively correlated. Other examples where the presence of this correlation is important empirically include automobiles (Berry, Levinsohn, and Pakes (1995) and Petrin (2002)), cable television choices (Goolsbee and Petrin (2004) and Crawford (2000)), supermarket goods like cereals (Nevo (2001)), yogurt and ketchup (Villas-Boas and Winer (1999)), and margarine and orange juice (Chintagunta, Dubé, and Goh (2005)), to name a few.

2.3 Berry, Levinsohn, and Pakes (1995)

Even assuming away all the random coefficients the Berry(1994) solution to endogenous prices in the face of non-separable demand is not consistent. To see this rewrite (2) in the spirit of Berry (1994),

\[
u_{ijt} = \beta' x_{jt} - \alpha p_{jt} + \xi_{jt} + \gamma' x_{jt} \xi_{jt} + \gamma' p_{jt} \xi_{jt} + \varepsilon_{ijt} \equiv \delta_{jt} + \varepsilon_{ijt}.
\]

Berry (1994) provides conditions under which there exists a unique vector \( \delta_t = (\delta_{1t}, \ldots, \delta_{Jt}) \) such that observed and predicted market shares are exactly equal. The first step of the Berry (1994) estimation approach is to use the uniqueness result and observed market shares to invert and recover \( \delta_t \). While the setup above differs from Berry’s setup in that \( (\gamma', \gamma'^P) \neq 0 \), Berry’s inversion result continues to hold.

The problem arises in the final step of his approach, where he uses an instrumental variables estimator to recover the parameters that are contained in \( \delta_t \), which are necessary to calculate price elasticities. Under the assumption that \( E[\xi_{jt}|z_{jt}] = 0 \) - with \( x_{jt} \) and instruments included in \( z_{jt} \) - he suggests regressing \( \delta_{jt} \) on \( x_{jt} \) (including an intercept term), and \( p_{jt} \) while treating price as endogenous. While \( \gamma' \neq 0 \) raises no econometric problems in this non-separable setup under the conditional mean restriction, \( \gamma'^P \neq 0 \) renders this approach inconsistent. From the formulation for
\( \delta_{jt} \) above, when \( \gamma^p \neq 0 \) one has \( p_{jt} \xi_{jt} \) in the error, and the instrumented price is correlated with \( p_{jt} \) and thus correlated with the error.

It is possible to sign the bias in the case where \( \gamma^p > 0 \), that is, when consumers become less price sensitive as the unobserved demand factor increases. Price will be positively correlated with its instrumented value, and price is typically positively correlated with \( \xi_{jt} \) (from previous empirical studies). Thus, consumers will appear less price sensitive than they are in reality. To summarize, when price and the unobserved demand factor interact in consumer utility, the Berry (1994) inversion continues to hold, but instrumental variables estimates of price elasticities are likely to be positively biased when price is not separable from the unobserved demand characteristic.

When we allow for random coefficients on observed characteristics (including price) only as

\[
\begin{align*}
    u_{ijt} = \beta^t_{it} x_{jt} - \alpha_{it} p_{jt} + \xi_{jt} + \gamma^t x_{jt} \xi_{jt} + \gamma^p p_{jt} \xi_{jt} + \varepsilon_{ijt} \equiv \delta_{jt} + \mu_{ijt} + \varepsilon_{ijt},
\end{align*}
\]

using BLP (1995)’s inversion (that matches the observed market shares and the predicted market shares) one can still invert the mean utility term \( \delta_{jt} = \delta_{jt}(\sigma) \) (profiled or indexed by the parameters, \( \sigma \) that determine the distribution of random coefficients) that does not depend on the idiosyncratic consumer tastes (\( \mu_{ijt} \)) as

\[
\begin{align*}
    \delta_{jt} = \beta^t x_{jt} - \alpha p_{jt} + \xi_{jt} + \gamma^t x_{jt} \xi_{jt} + \gamma^p p_{jt} \xi_{jt}.
\end{align*}
\]

However, a similar argument above shows that the conditional moment restriction \( E[\xi_{jt}|z_{jt}] = 0 \) does not suffice for identification of demand parameters, including the coefficients on the observed characteristics (\( \beta, \alpha \)) unless \( \gamma^p = 0 \). Gandhi, Kim, and Petrin (2013) formalize this argument and propose an alternative identification and estimation strategy for the demand models with interactions of observed and unobserved factors.\(^3\)

### 2.4 Berry and Haile (2014)’s Non-parametric Demand

Berry and Haile (2014) show identification of demand in general non-parametric setup. They, however, require that a special type of characteristic exists and show how to use it in conjunction with conditional moment restrictions to achieve identification in differentiated products models with market level data. This special characteristic - call it \( x_{jt}^{(1)} \) - must be perfectly substitutable with \( \xi_{jt} \), and the coefficient on the special characteristic must be normalized. The approach allows for non-parametric identification of demand in the variables \( (x_{jt}^{(1)} + \xi_{jt}, x_{jt}^{(2)}, p_{jt}) \). Thus the utility specification is restricted to

\[
\begin{align*}
    u_{ijt} = u(x_{jt}^{(2)}, p_{jt}, x_{jt}^{(1)} + \xi_{jt}, \omega_{it}) \quad (3)
\end{align*}
\]

and this does not allow a utility specification like (2).

\(^3\)Gandhi, Kim, and Petrin (2013), however, do not allow for random coefficients on the unobserved factor nor allow for general non-separable demand models as (1).
3 A Control Function Approximation to Demand

In section 4 we show that the unobserved demand factor $\xi_{jt}$ enters the reduced form pricing function as an argument. Under conditions outlined there, a one-to-one function of $\xi_{jt}$ is identified - defined as $\tilde{\xi}_{jt} = v_j(\xi_t)$ - and since $\omega_{it}$ is independent of $(x_{jt}, p_{jt}, \xi_{jt})$, it is independent of $(x_{jt}, p_{jt}, \tilde{\xi}_{jt})$ as well.

Our control function approach proceeds in two stages. In the first stage we recover $\tilde{\xi}_t = (\tilde{\xi}_1, \ldots, \tilde{\xi}_J)$ by inverting the pricing function. In the second stage by plugging $\xi_{jt} = \phi_j(\tilde{\xi})$, the inverse function to obtain $\xi_{jt}$, in the utility we obtain

$$u_{ijt} = u(x_{jt}, p_{jt}, \xi_{jt}, \omega_{it}) = u(x_{jt}, p_{jt}, \phi_j(\tilde{\xi}), \omega_{it}),$$

and approximate $\phi_j(\cdot)$ using parametric or non-parametric models for estimation. For the example of the demand model (2) we can write

$$u_{ijt} = \beta_j'x_{jt} - \alpha_jp_{jt} + \gamma_\xi\phi_j(\tilde{\xi}_t) + \gamma^x_\xi x_{jt}\phi_j(\tilde{\xi}_t) + \gamma^p_\xi p_{jt}\phi_j(\tilde{\xi}_t) + \epsilon_{ijt}$$

$$\approx \beta_j'x_{jt} - \alpha_jp_{jt} + \gamma_\xi\phi_j(\Lambda_\xi\tilde{\xi}_t) + \gamma^x_\xi x_{jt}(\Lambda_\xi\tilde{\xi}_t) + \gamma^p_\xi p_{jt}(\Lambda_\xi\tilde{\xi}_t) + \epsilon_{ijt}$$

where we approximate $\phi_j(\cdot)$ using a linear function of $\tilde{\xi}_t$. The approach addresses price endogeneity, allows for heterogeneity in tastes, does not impose additive separability between observed and unobserved factors, and is easily extended to allow for more flexible specifications.

Following the aforementioned control function approach, estimation proceeds in two stages. In the first stage, we estimate the controls $\tilde{\xi}_t$ from the pricing equations as described in Section 4. In the second stage, we specify distributions of $\omega_{it}$ (e.g. type I extreme value distributions and/or normal distributions for idiosyncratic errors, and independent normal distributions for random coefficients), and estimate corresponding choice models using the maximum likelihood method treating controls as additional regressors. We then formalize our approach as semiparametric sieve estimation in Section 5. Examples of estimation method are illustrated with the three empirical applications later in Section 6, and we also provide formulas for the asymptotic standard errors for a few different data settings in Appendix F.

A direct test of specification is to evaluate whether the predicted shares from the estimated model are statistically different from the observed shares in the data using a standard $\chi^2$ test. Indeed, BLP argue that models that do not have product-specific errors often “overfit” because only sampling error can explain the difference between the data and the model, and there is often insufficient variance to account for the discrepancy. This same test is applicable here. If the test rejects the model without any controls and does not reject the control function specification, there is evidence that the main symptom of endogenous prices has been addressed by the addition of $\tilde{\xi}_t$ to the demand specification. If the control function specification is still rejected, then either a more flexible formulation for the utility (4) is required, or the control variable $\tilde{\xi}_t$ does not provide a good approximation to $\xi_t$.

4In this setup we need to normalize one coefficient in $\lambda_j$ or the mean of the random coefficient $\gamma_\xi^\epsilon$.

5In principle a standard argument for non-parametric estimation would require that higher-order terms of $\tilde{\xi}_t$ be added as the sample size increases.
4 Inverting Prices to Recover Control Variates

We consider the reduced form prices that allow for demand errors as well as cost side unobservables as

\[ p_{jt} = p_j(x_t, z_t, \eta_t, \xi_t), \]

with \( p_j(\cdot) \) determined by demand shifters \( x_t, \) cost-side factors \( z_t = (z_{1t}, \ldots, z_{Jt}) \), unobserved demand factors \( \xi_t \), and unobserved cost shocks \( \eta_t = (\eta_{1t}, \ldots, \eta_{Jt}) \). First we utilize a supply side model of Berry and Haile (2014), and argue that \( \eta_t \) is nonparametrically identified given observed shares, prices, and demand and cost factors. Then given \( (x_t, z_t, \eta_t) \) we establish sufficient conditions for the existence of a reduced form price equation that is invertible in \( \xi_t \). We start with the single-product monopolist and then show sufficient conditions for invertibility for the differentiated product setting with multi-product vendors, where there is a system of \( J \) price equations and the vector of demand errors \( \xi_t \). Finally, we provide estimators for \( \tilde{\xi}_t \), the one-to-one function of \( \xi_t \).

4.1 Non-parametric Identification of the Unobserved Cost Shock

To invert the unobserved cost shock \( (\eta_t) \) we consider a supply side model of Berry and Haile (2014) (BH hereafter) that generalizes the parametric models of static oligopoly, in which one can obtain marginal costs in terms of demand parameters by utilizing first-order conditions of firms profit maximization. While we impose the same index restriction as BH in the marginal cost specification below, we do not need to impose the index restriction (3) in the demand.

Write the market shares as

\[ s_{jt} = \int \{ u_{ijt} \geq u_{ikt} \forall k \neq j \} dF_{\omega_{jt}} \equiv \sigma_j(x_t, p_t, \xi_t). \]

Then, as BH, we assume the marginal cost of producing product \( j \) depends on its output quantity \( M_t s_{jt} = M_t \sigma_j(x_t, p_t, \xi_t) \) with \( M_t \) being the market size, a cost index that includes the unobservable, and other cost shifters:

\[ mc_j(M_t s_{jt}, \zeta_{jt}, z_{jt}^{(2)}) \]

where \( z_{jt}^{(1)} \) is a cost shifter excluded from demand, and other cost factors \( z_{jt}^{(2)} \) may include observable demand factors. Here the key restriction imposed is perfect substitution between the unobserved cost shock \( \eta_{jt} \) and the cost shifter \( z_{jt}^{(1)} \) inside the marginal cost function. However, as BH argue this restriction is satisfied in many standard supply side models. For example, in a linear marginal cost function the index restriction is equivalent to not allowing random coefficient on a particular cost factor \( z_{jt}^{(1)} \) (which is standard in the literature). We assume the following conditions from BH.

**Assumption 4.1 (BH1)** \( \sigma_j(x_t, p_t, \xi_t) \) is continuously differentiable w.r.t. price \( p_{kt} \) \( \forall j, k \in J \).

**Assumption 4.2 (BH2)** (i) \( mc_j(M_t s_{jt}, \zeta_{jt}, z_{jt}^{(2)}) \) is strictly monotonic in \( \zeta_{jt} \equiv z_{jt}^{(1)} + \eta_{jt} \);

(ii) \( u(x_{jt}, p_{jt}, \xi_{jt}, \omega_{jt}) \) is strictly decreasing in \( p_{jt} \);

(iii) there exists a function \( \psi_j \) (possibly unknown) such that for any equilibrium value of \( (s_t, p_t) \)

\[ mc_j(M_t s_{jt}, \zeta_{jt}, z_{jt}^{(2)}) = \psi_j(s_t, M_t, D_t(s_t, p_t, x_t), p_t) \]
where \( D_t(s_t, p_t, x_t) \) is the \( J \times J \) matrix of partial derivatives \( \left\{ \frac{\partial \sigma_j(x_t, p_t, \sigma_{-1}^{-1}(x_t, p_t, s_t), ..., \sigma_{-1}^{-1}(x_t, p_t, s_t))}{\partial p_{kt}} \right\}_{k,j\in J} \).

BH1 is required in using first-order conditions. BH2 (i) permits the invertibility of the cost index \( \zeta_t \) from the marginal cost. BH2 (ii) would hold trivially with fixed price coefficient while it may restrict the support of random coefficient on price. BH2 (iii) means that one can utilize first-order conditions to express marginal cost for each product as a function of equilibrium quantities, prices, and derivatives of demand at these prices and quantities. Here the existence and uniqueness of inversion \( \xi_{jt} = \sigma_{-1}^{-1}(x_t, p_t, s_t) \) \( \forall j \in J \) from share equations is implicit, which holds true under the “connected substitutes” assumption in BH.

Assumption BH2 (Lemma 2 in BH) ensures that for any market size \( M_t \) and any given \((s_t, p_t, x_t, z_t^{(2)})\) one can invert unique \( \zeta_t \) such that \( \zeta_{jt} = mc^{-1}(M_t, s_t, \psi_j(s_t, M_t, D_t(s_t, p_t, x_t), p_t), z_t^{(2)}) \). We then can rewrite this, with a fixed market size \( M_t \), as

\[
\frac{\zeta_{jt}^{(1)} + \eta_{jt}}{\zeta_{jt}} = \varphi_j^{-1}(s_t, p_t, x_t, z_t^{(2)})
\]

and therefore we can consistently estimate \( \varphi_j^{-1}(s_t, p_t, x_t, z_t^{(2)}) \) using the non-parametric IV approach under the following conditional moment restriction and the completeness condition (see Theorem 3 in BH)

**Assumption 4.3 (BH3)** \( E[\eta_{jt}|z_t, x_t] = 0 \) almost surely for \( \forall j \in J \).

**Assumption 4.4 (BH4)** For all functions \( B(s_t, p_t) \) with finite expectation, if \( E[B(s_t, p_t)|z_t, x_t] = 0 \) almost surely then \( B(s_t, p_t) = 0 \) almost surely.

The non-parametric IV method utilizes the conditional moment restriction

\[
E[\eta_{jt}|z_t, x_t] = E[\varphi_j^{-1}(s_t, p_t, x_t, z_t^{(2)}) - z_{jt}^{(1)}|z_t, x_t] = 0,
\]

and the estimation proceeds first by approximating \( \varphi_j^{-1}(s_t, p_t, x_t, z_t^{(2)}) \) as a flexible function of \((s_t, p_t, x_t, z_t^{(2)})\) using approximating basis functions (e.g., power series or splines), and second by estimating the approximated function using the IV estimation as we would do for a parametric model (e.g., see Newey and Powell (2003), Hall and Horowitz (2005), and Darolles, Fan, Florens, and Renault (2011)). Since \( \eta_t \) is non-parametrically identified and can be consistently estimated using the non-parametric IV approach under BH1-BH4, in what follows we treat \( \eta_t \) (also \( \zeta_t = z_{jt}^{(1)} + \eta_{jt} \)) as known unless we otherwise note it is to be estimated.

### 4.2 Invertibility of the Unobserved Demand Factor

We provide a set of sufficient conditions for the invertibility of the unobserved demand factor from the pricing function. We start with a single-product setting and extend the result to a multi-product setting.
4.2.1 Single-Product Setting

We start with the single-product monopolist. We broadly define product characteristics to include physical or perceived physical characteristic of the good (like style), and advertising or promotional activities undertaken by firms to encourage demand for the good. Let \( x_t \) be this vector of product characteristics, where the \( \xi_t \) might be one of the characteristics observed to consumers and producers but unobserved to the practitioner. Let \( z_t \) be other potential cost shifters, \( q_t(p_t, x_t, \xi_t) \) be quantity demanded at price \( p_t \) given \((x_t, \xi_t)\), and \( mc(q_t, \zeta_t, z_{t}^{(2)}) \) be the marginal cost of production. Profits are given as

\[
\Pi = (p_t - mc(q_t, \zeta_t, z_{t}^{(2)}))q_t(p_t, x_t, \xi_t),
\]

and the optimal (static) price \( p_t^* \) solves (given the others)

\[
\frac{\partial \Pi(p_t^*, x_t, z_{t}^{(2)}, \zeta_t)}{\partial p_t} = 0.
\] (5)

Lemma 4.1 provides a set of sufficient conditions that relate changes in the unobserved demand factor to changes in price conditional on the others.

**Lemma 4.1** Assume \( \frac{\partial q_t}{\partial p_t} < 0 \), \( \frac{\partial mc}{\partial \xi_t} \geq 0 \), \( \frac{\partial q_t}{\partial \xi_t} \geq 0 \), and \( -\frac{\partial^2 \Pi(p_t^*, \cdot)}{\partial p_t^2} \) is strictly positive. Then \( \frac{\partial p_t^*}{\partial \xi_t} > 0 \) if

\[
\frac{\partial q_t}{\partial \xi_t} - \frac{\partial mc}{\partial \xi_t} \frac{\partial q_t}{\partial p_t} > -(p_t^* - mc) \frac{\partial^2 q_t}{\partial \xi_t \partial p_t},
\] (6)

holds for all optimal price \( p_t^* \) that solves (5).

**Proof.** Total differentiation of the first-order condition with respect to \( p_t \) and \( \xi_t \) coupled with the implicit function theorem leads to the comparative static:

\[
\frac{\partial p_t}{\partial \xi_t} = -\left( \frac{\partial^2 \Pi(p_t^*, \cdot)}{\partial p_t^2} \right)^{-1} \frac{\partial^2 \Pi(p_t^*, \cdot)}{\partial p_t \partial \xi_t} = -\left( \frac{\partial^2 \Pi(p_t^*, \cdot)}{\partial p_t^2} \right)^{-1} \left( \frac{\partial q_t}{\partial \xi_t} - \frac{\partial mc}{\partial \xi_t} \frac{\partial q_t}{\partial p_t} + (p_t^* - mc) \frac{\partial^2 q_t}{\partial \xi_t \partial p_t} \right).
\]

The result follows because \( -\frac{\partial^2 \Pi(p_t^*, \cdot)}{\partial p_t^2} \) is strictly positive, and the aforementioned conditions hold for all optimal price \( p_t^* \). ■

The first three assumptions are weak, requiring respectively that demand is downward sloping, marginal cost is non-decreasing in the unobserved demand factor, and demand is non-decreasing in the unobserved demand factor (that consumers value). Also \( -\frac{\partial^2 \Pi(p_t^*, \cdot)}{\partial p_t^2} \geq 0 \) is a (second-order) necessary condition for profit maximization. We only assume that this condition holds as strictly positive.

The left hand side of (6) is positive under these assumptions, so the result depends on \( \frac{\partial^2 q_t}{\partial \xi_t \partial p_t} \), which characterizes how the price elasticity of demand changes as the unobserved demand factor increases. If demand becomes less price elastic, then the result holds. If demand becomes more price elastic as the unobserved demand factor increases, then the left hand side of the inequality must exceed the right hand side. When these sufficient conditions hold, price is monotonically increasing in \( \xi_t \), which implies both that price can be written as a function of the unobserved demand factor.
given other observables and the cost shock as \( p_t = p(x_t, z_t(2), \zeta_t, \xi_t) \), and that price is invertible in the unobserved demand factor \( \xi_t \).

### 4.2.2 Multi-Product Setting

We focus on the static Bertrand-Nash price competition, which is perhaps the most popular maintained competitive assumption in the applied discrete choice literature. Recall \( \xi_t = (\xi_{1t}, \ldots, \xi_{Jt})' \) denotes the vector of unobserved factors, one for each product. For the Bertrand-Nash setup, define \( \Pi_f(p_t, x_t, z_t(2), \zeta_t, \xi_t) \) as the profits for firm \( f, f = 1, \ldots, F \), which produces a subset of the goods \( j \in J_f \) and chooses prices to maximize static profits. Let \( p_t \) enter as the first \( J \) arguments and \( \xi_t \) enter as the last \( J \) arguments in every profit function. For fixed \((x_t, z_t(2), \zeta_t)\), let \( p^*_t = (p^*_{1t}, \ldots, p^*_{Jt})' \) denote a vector that satisfies

\[
\frac{\partial \Pi_f(p^*_t, x_t, z_t(2), \zeta_t, \xi_t)}{\partial p_{jt}} = 0, \quad j \in J_f \text{ and } f = 1, \ldots, F. \tag{7}
\]

Without loss of generality, let \( J_f \) contain a subset of the goods with indexes higher than those in \( J_{f'} \) if \( f > f' \), so \( J_1 \) must contain the good "1" and \( J_F \) must contain the good "J". Then define two \( J \times J \) matrices

\[
D_{pp}\Pi(p^*_t, x_t, z_t(2), \zeta_t, \xi_t) = \begin{bmatrix}
\frac{\partial^2 \Pi_1(p^*_t, x_t, z_t(2), \zeta_t)}{\partial p_{1t}\partial p_{1t}} & \cdots & \frac{\partial^2 \Pi_1(p^*_t, x_t, z_t(2), \zeta_t)}{\partial p_{1t}\partial p_{Jt}} \\
\vdots & \ddots & \vdots \\
\frac{\partial^2 \Pi_F(p^*_t, x_t, z_t(2), \zeta_t)}{\partial p_{Jt}\partial p_{Jt}} & \cdots & \frac{\partial^2 \Pi_F(p^*_t, x_t, z_t(2), \zeta_t)}{\partial p_{Jt}\partial p_{Jt}}
\end{bmatrix}, \quad D_{\xi p}\Pi(\cdot, \zeta_t, \xi_t) = \begin{bmatrix}
\frac{\partial^2 \Pi_1(p^*_t, x_t, z_t(2), \zeta_t)}{\partial \xi_{1t}\partial p_{1t}} & \cdots & \frac{\partial^2 \Pi_1(p^*_t, x_t, z_t(2), \zeta_t)}{\partial \xi_{1t}\partial p_{Jt}} \\
\vdots & \ddots & \vdots \\
\frac{\partial^2 \Pi_F(p^*_t, x_t, z_t(2), \zeta_t)}{\partial \xi_{Jt}\partial p_{Jt}} & \cdots & \frac{\partial^2 \Pi_F(p^*_t, x_t, z_t(2), \zeta_t)}{\partial \xi_{Jt}\partial p_{Jt}}
\end{bmatrix}.
\]

**Lemma 4.2** Given \((x_t, z_t(2), \zeta_t)\), suppose each of the first order conditions in (7) is continuously differentiable in its first and last \( J \) arguments and suppose \( D_{pp}\Pi(p^*_t, x_t, z_t(2), \zeta_t, \xi_t) \) has the full rank, then \( p^*_t \) can be expressed as an implicit function of \( \xi_t \) given \((x_t, z_t(2), \zeta_t)\),

\[
p^*_t(x_t, z_t(2), \zeta_t, \xi_t) = (p^*_{1t}(x_t, z_t(2), \zeta_t, \xi_t), \ldots, p^*_J(x_t, z_t(2), \zeta_t, \xi_t))'.
\]

in a neighborhood of \( \xi_t \) (given the others) that satisfies (7). If

\[
D_{\xi p}\Pi(p^*_t, x_t, z_t(2), \zeta_t, \xi_t)
\]

has full rank for all equilibrium price \( p^*_t \) that solves (7), then the matrix of derivatives \( \frac{\partial p_t}{\partial \xi_t} \) is invertible.

**Proof.** Total differentiations of the first order conditions from (7) yields

\[
D_{pp}\Pi(p^*_t, x_t, z_t(2), \zeta_t, \xi_t)\partial p_t = -D_{\xi p}\Pi(p^*_t, x_t, z_t(2), \zeta_t, \xi_t)\partial \xi_t.
\]

The matrix

\[
D_{pp}\Pi(p^*_t, x_t, z_t(2), \zeta_t, \xi_t)
\]

...
is full rank, so the first claim follows directly from the implicit function theorem. Solving for the \( J \times J \) matrix of derivatives \( \frac{\partial p_t}{\partial \xi_t} \) yields

\[
\frac{\partial p_t}{\partial \xi_t} = - \left( D_{pp} \Pi(p_t^*, x_t, z_t^{(2)}, \zeta_t, \xi_t) \right)^{-1} D_{\xi p} \Pi(p_t^*, x_t, z_t^{(2)}, \zeta_t, \xi_t).
\]

\( \frac{\partial p_t}{\partial \xi_t} \) is equal to the product of two full rank matrices, which is also full rank and thus invertible. The result follows since the aforementioned conditions hold for all equilibrium price \( p_t^* \).

Lemma 4.2 shows that the conditions under which prices can be written as a one-to-one function of the vector of unobserved demand factors given the other factors. The matrix (9) is negative semi-definite due to a (second-order) necessary condition for profit maximization. We only assume that it is also negative definite and so has the full rank. Thus, the invertibility turns on the full rank of (8). This requires the vector of first order conditions to vary in \( J \) independent directions when differentiated with respect to the vector \( \xi_t \).

### 4.3 Estimation of Control Variable

Given invertibility, which we now maintain throughout, the final step that remains is to show the conditions under which a one-to-one function of \( \xi_{jt} \) is identified from the pricing function - defined as \( \xi_{jt} = v_j(\xi_t) \) for all \( j \in J \), which becomes the control variable to use for estimation. Again we first consider the single-product market case, and then turn to multi-product markets.

In our empirical approaches we will be using repeated observations on markets to recover the control variables, and we will be explicit by adding a market index \( t \) to the pricing equation when we propose the estimators. \( t \) can serve as an index for the same market observed repeatedly over time, a cross-section of markets at a given point in time, or a combination of the two.

#### 4.3.1 Single Product Markets

In the univariate case, we extend the estimator proposed in Imbens and Newey (2009) to demand systems to identify \( \xi_t \) up to a normalization. Theorem D.1 in Appendix D provides the details, showing that if \( p_t(x_t, z_t^{(2)}, \zeta_t, \xi_t) \) is strictly monotonic in \( \xi_t \), and \( \xi_t \) is independent of \( (x_t, z_t^{(2)}, \zeta_t) \), then the conditional distribution of \( p_t \) given \( (x_t, z_t^{(2)}, \zeta_t) \) is equal to the marginal distribution of \( \xi_t \) up to a normalization:

\[
F_{p_t|x_t, z_t^{(2)}, \zeta_t}(p_t|x_t, z_t^{(2)}, \zeta_t) = F_{\xi_t}(\xi_t)
\]

where \( F_{p_t|x_t, z_t^{(2)}, \zeta_t} \) denotes the conditional CDF (cumulative distribution function) of \( p_t \) conditional on \( (x_t, z_t^{(2)}, \zeta_t) \) and \( F_{\xi_t}(\xi_t) \) denotes the CDF of \( \xi_t \).

The condition states that two markets \( a \) and \( b \) with the same observed demand \( x_t \) and supply factors \( (z_t^{(2)}, \zeta_t) \) but different prices differ in their unobserved demand factors, with \( p_a > p_b \) if and only if \( \xi_a > \xi_b \). Specifically, the condition implies that a set of markets with the same observed factors has a distribution of prices that has the exact same ordering as the distribution of unobserved factors.
The result points directly to the empirical cumulative distribution function for $F_{p_t|z_t,\xi_t^{(2)},\zeta_t}$ as an estimator for $F_{\xi_t}$. Specifically, if we normalize the distribution of $\xi_t$ as uniform on $[0,1]$ the control function is defined as $\tilde{\xi}_t = F_{\xi_t}(\xi_t) = F_{p_t|z_t,\xi_t^{(2)},\zeta_t}(p_t|z_t,\xi_t^{(2)},\zeta_t)$, a random variable that is one-to-one with $\xi_t$ and uniformly distributed over the unit interval. If there are $T$ total markets, one can estimate the value of the control for market $t$ as

$$\tilde{\xi}_t = \frac{\int_{\tilde{\xi} \in \tilde{\mathcal{P}}} \tilde{f}(p, x_t, \xi_t^{(2)}, \tilde{\xi}_t) dp}{\int_{\tilde{\xi} \in \tilde{\mathcal{P}}} \tilde{f}(p, x_t, \xi_t^{(2)}, \tilde{\xi}_t) dp}$$  \hspace{1cm} (10)

where $[\tilde{\mathcal{P}}, \tilde{\mathcal{P}}]$ denotes the support of price $p$ and $\tilde{f}(p, x_t, \xi_t^{(2)}, \tilde{\xi}_t)$ denotes a kernel density estimator of the density of $(p, x_t, \xi_t^{(2)}, \tilde{\xi}_t)$ where we explicitly note that we need to estimate the cost factor $\tilde{\xi}_t = \xi_t^{(1)} + \tilde{\eta}_t$ at the first stage using the non-parametric IV method of Berry and Haile (2014) as described in Section 4.1.

### 4.3.2 Multi-Product Markets

In the multi-product setting, where we have $J$ products, there are $J$ pricing equations and $J$ unobserved demand factors to invert out. Note that all observed demand characteristics $x_t$ and cost factors $(\xi_t^{(2)}, \zeta_t)$ in market $t$ relevant for the determination of prices in equilibrium are included in the pricing functions.\(^6\) The price for any product $j$ in any market $t$ is then given by the equilibrium pricing function

$$p_{jt} = p_j(x_t, \xi_t^{(2)}, \zeta_t, \xi_t), \quad j = 1, \ldots, J.$$ \hspace{1cm} (11)

The system of equations in (11) is a non-linear simultaneous equations system. We need to invert this system to recover the vector $\tilde{\xi}_t$, a one-to-one (vector) function of $\xi_t$ for each market $t$. Generic identification results of some structural objects for this kind of setting are given in Matzkin (2008). In what follows we show, similar to the univariate case, if invertibility and independence hold, and there exists a (unknown) real-valued function $v_j(\xi_t)$ such that $p_{jt} = p_j(x_t, \xi_t^{(2)}, \zeta_t, v_j(\xi_t))$, with $p_j(x_t, \xi_t^{(2)}, \zeta_t, v_j(\xi_t))$ strictly increasing in $v_j(\xi_t)$ for every $j$, then a similar non-parametric estimator to the one used in the single-product case as (10) for each $\tilde{\xi}_{jt} = v_j(\xi_t)$ can be used to recover the unique set of controls.

We present our main theorem below that shows $\tilde{\xi}_t = (v_1(\xi_t), \ldots, v_J(\xi_t))'$ is one-to-one with $\xi_t$.

**Assumption 4.5** With probability one the $J \times J$ matrix of derivatives $\frac{\partial p_{jt}}{\partial \xi_t}$ is invertible.

**Assumption 4.6** $\xi_t$ are jointly independent of $(x_t, \xi_t^{(2)}, \zeta_t)$, and $\xi_{jt}$ and $\xi_{kt}$ are independent for $\forall k \neq j \in J$.

**Assumption 4.7** There exists a scalar (unknown) real-valued function $v_j(\xi_t)$ such that $p_{jt} = p_j(x_t, \xi_t^{(2)}, \zeta_t, \xi_t) = p_j(x_t, \xi_t^{(2)}, \zeta_t, v_j(\xi_t))$.\(^6\)

---

\(^6\)In a setting with multi-product sellers we also need to include indicators for sellers of each product because the same set of products divided up differently across the set of producers will generally result in different equilibrium prices.
Assumption 4.8 For all \((x_t, z_t^{(2)}, \zeta_t), p_j(x_t, z_t^{(2)}, \zeta_t, v_j(\xi_t))\) is strictly increasing in \(v_j(\xi_t)\) for \(\forall j \in J\).

A set of sufficient conditions for the invertibility in Assumption 4.5 is provided in Section 4.2.2. The very assumption in Assumptions 4.7-4.8 that we can write the pricing function as \(p_{jt} = p_j(x_t, z_t^{(2)}, \zeta_t, v_{jt})\) for some variable \(v_{jt}\), with \(p_j(\cdot, v_{jt})\) strictly increasing in \(v_{jt}\), is not a material assumption (this is just rewriting of the function). This is because, assuming \(p_{jt}\) is continuous, we can always rewrite \(p_j\) as a function of \((x_t, z_t^{(2)}, \zeta_t)\) and a continuous single error term \(v_{jt}\) - \(p_{jt} = p_j(x_t, z_t^{(2)}, \zeta_t, v_{jt})\) - such that \(v_{jt}\) is independent of \((x_t, z_t^{(2)}, \zeta_t)\) and \(p_j(\cdot, v_{jt})\) is strictly increasing in \(v_{jt}\) as shown below (see Matzkin (2003)). To see this subtle fact, normalize the error term \(v_{jt}\) to be uniform over the unit interval \([0, 1]\). Then we have \(v_{jt} = F_{p_j|x_t,z_t^{(2)},\zeta_t}(p_{jt}|x_t,z_t^{(2)},\zeta_t)\) where \(F_{p_j|x_t,z_t^{(2)},\zeta_t}\) denotes the conditional CDF of \(p_{jt}\) given \((x_t, z_t^{(2)}, \zeta_t)\). It follows that

\[
p_{jt} = F_{p_j|x_t,z_t^{(2)},\zeta_t}^{-1}(v_{jt}|x_t,z_t^{(2)},\zeta_t) = p_j(x_t, z_t^{(2)}, \zeta_t, v_{jt})
\]

which is also strictly increasing in \(v_{jt}\) by definition of the conditional CDF. Therefore, the above result only assumes \(p_{jt}\) is continuous. Then the only material assumption in Assumptions 4.7-4.8 is that this variable \(v_{jt}\) is given by a function of \(\xi_t\) only as \(v_{jt} = v_j(\xi_t)\), and so it does not depend on any other factors. The theorem below shows that the control variable given by \(\hat{\xi}_t = (v_1(\xi_t), \ldots, v_J(\xi_t))'\) is one-to-one with the true demand error \(\xi_t\).

Theorem 4.1 Let \(\hat{\xi}_t = v_j(\xi_t) = p_j^{-1}(x_t, z_t^{(2)}, \zeta_t, p_{jt})\) for \(\forall j \in J\). If Assumptions 4.5-4.8 hold, then \(\hat{\xi}_t = (\hat{\xi}_1, \ldots, \hat{\xi}_J)'\) is a one-to-one function of \(\xi_t\).

See Appendix D for the proof. In particular, Corollary D.1 in Appendix D applies to the case when prices are additively separable in the unobserved demand factors. In this case \(p_{jt}\) can be written as

\[
p_{jt} = g_j(x_t, z_t^{(2)}, \zeta_t) + v_j(\xi_t)
\]

where w.l.o.g. we can write \(g_j(x_t, z_t^{(2)}, \zeta_t) = E[p_{jt}|x_t, z_t^{(2)}, \zeta_t]\). In this setting the inversion of the equations from (11) can be done using OLS for each of the \(J\) equations. The estimate of the vector of controls \(\hat{\xi}_t = (v_1(\xi_t), \ldots, v_J(\xi_t))'\), which is one-to-one with \(\xi_t\), is then obtained as the vector of regression residuals of the pricing equations,

\[
\hat{\xi}_t = (\hat{v}_1t, \hat{v}_2t, \ldots, \hat{v}_Jt), \quad \hat{v}_{jt} = v_j(\hat{\xi}_t) = p_{jt} - \hat{g}_j(x_t, z_t^{(2)}, \hat{\xi}_t) \text{ for } \forall j \in J
\]

where \(\hat{g}_j(\cdot)\) denotes the estimated regression function, and \(\hat{\xi}_t\) is the estimated cost factor.

4.4 Control Function Approach with Demand Side Only

Although the general approach can allow for supply side unobservables, it also requires a supply side shifter that is fully excluded from the demand side. If this supply side variable is not available, one
may use the control function approach assuming there is only one unobservable (product specific shock) to each product. In this case we write the reduced form for prices as

\[ p_{jt} = p_j(x_t, z_t, \xi_t). \]

Here the maintained assumption is that either all cost side factors are observed or the unobservable is fully captured in \( \xi_t \), a strong but testable assumption. In this case the approach to estimate the control variates, proposed in Section 4.2-4.3, can be still used simply by dropping the cost shock \( \eta_t \) (and the first step to recover the shock).

### 4.5 Discussion

If in the data the number of observed markets is large relative to the dimension of observed demand and supply factors, variation in prices and observed characteristics are typically sufficient to identify and estimate the controls. This is the setting for both the cable television and margarine cases, where there are four product types, and variants of them are observed in a cross-section (U.S. cable franchise markets in 2001) and a time-series (margarine sales at a supermarket over 117 weeks).

When the dimension of observed demand and supply factors is large relative to the number of observed markets, there may be “too many” regressors to use in the estimation of the residuals. We suggest using an approximation from Pakes (1994), who provides a parsimonious basis for equilibrium pricing functions that are partially exchangeable (that is, when one can change the order in which some of the arguments enter the function without changing the value of the function). This result has wide applicability to differentiated product markets, and we discuss its usefulness further in the context of the control function approach in the BLP automobile case-study.

### 5 Estimation of the Demand Models

In this section we formally develop the second stage ML estimator of demand parameters in our control function approach, discussed in Section 3, as a semiparametric sieve estimator and show consistency of the sieve ML estimator that uses the estimated controls as regressors. Let \( \theta \) collect the mean utility parameters and parameters in the distribution of \( \omega_{jt} \) (random coefficients and the idiosyncratic demand errors), and let \( \theta_0 \) denote its true value. Recall \( \xi_{jt} = \phi_j(\tilde{\xi}_{jt}) \) where \( \phi_j(\cdot) \) denotes the control function for product \( j \) and let \( \phi_{0j}(\cdot) \) and \( \tilde{\xi}_{0jt} \) denote the true \( \phi_j(\cdot) \) and \( \tilde{\xi}_{jt} \) for each \( j \in J \), respectively. In the first stage we estimate the unobserved cost shocks (if the model allows), and also estimate the demand error control variates. We can estimate the demand error control \( \tilde{\xi}_{0jt} \) and obtain the estimate \( \tilde{\xi}_{jt} \) for \( j \in J \) as described in Sections 4.3-4.4. Then, we construct approximating basis functions using \( \tilde{\xi}_{jt} \), and in the second stage we estimate the demand parameters \( \theta_0 \) as well as the control functions \( \phi_0(\cdot) = (\phi_{01}(\cdot), \ldots, \phi_{0J}(\cdot)) \) using a sieve ML method.

Let \( \Psi_j \) denote a space of functions that includes the true function \( \phi_{0j}(\cdot) \), endowed with \( \| \cdot \|_s \) a pseudo-metric on \( \Psi_j \) for all \( j \in J \). We write the infeasible basis functions (and we replace them with their estimates below) to approximate \( \phi_j(\cdot) \), when \( \tilde{\xi}_{0jt} = (\tilde{\xi}_{01jt}, \ldots, \tilde{\xi}_{0Jjt}) \) is known, as...
which is a sequence of approximating basis functions of \( \xi_{0t} \) such as power series or splines. With the sample size \( T \) define the sieve space \( \Psi_j^T \) as the collection of functions

\[
\Psi_j^T = \{ \phi_j : \phi_j = \sum_{r=1}^{R(T)} \lambda_j^r \varphi_j^r(\cdot), \| \phi_j \|_s < \bar{C} \} \tag{12}
\]

for some bounded positive constant \( \bar{C} \) and coefficients \( (\lambda_j^1, \ldots, \lambda_j^{R(T)}) \), with \( R(T) \to \infty \) and \( R(T)/T \to 0 \) such that \( \Psi_j^T \subseteq \Psi_j^{T+1} \subseteq \ldots \subseteq \Psi_j \), so we use more flexible approximations as the sample size grows. For estimation we then replace the sequence of the infeasible basis functions \( \varphi_j^r(\xi_{0t}) \) with feasible ones as \( \varphi_j^r(\hat{\xi}_t) \) in (12). Note that as long as \( \varphi_j^r(\cdot) \) is continuous \( \varphi_j^r(\hat{\xi}_t) \to_p \varphi_j^r(\xi_{0t}) \) as \( \hat{\xi}_t \to_p \xi_{0t} \) and therefore for any \( \phi_j(\cdot) = \sum_{r=1}^{R(T)} \lambda_j^r \varphi_j^r(\cdot) \in \Psi_j^T \) we also have \( \phi_j(\hat{\xi}_t) \to_p \phi_j(\xi_{0t}) \) as \( \hat{\xi}_t \to_p \xi_{0t} \).

Define \( \phi(\cdot) = (\phi_1(\cdot), \ldots, \phi_J(\cdot)) \) and for \( (\theta, \phi) \in \Theta \times \Psi_1^T \times \ldots \times \Psi_J^T \) let \( L_T(\theta, \phi(\cdot)) \) denote a sample criterion function for the ML estimation in the second stage. Also define the corresponding population criterion function as \( L^0(\theta, \phi(\cdot)) \) for \( (\theta, \phi) \in \Theta \times \Psi_1 \times \ldots \times \Psi_J \). Then we obtain the ML estimator as

\[
(\hat{\theta}, \hat{\phi}) = \arg\max_{(\theta, \phi) \in \Theta \times \Psi_1^T \times \ldots \times \Psi_J^T} L_T(\theta, \phi(\hat{\xi}_t)). \tag{13}
\]

We derive the consistency of the estimator under the following assumptions based on the results in Chen (2007).\(^7\) Here we state all the conditions we use to show consistency as assumptions for transparency. However, assumptions below either hold trivially or are satisfied under standard regularity conditions.

We first assume identification:

**Assumption 5.1 (A1)** \((\theta_0, \phi_0) \in \Theta \times \Psi_1 \times \ldots \times \Psi_J \) is the only \((\theta, \phi)\) that maximizes \( L^0(\theta, \phi(\cdot)) \).

Noting that \( \xi_{jt} = \phi_{0j}^{\varphi_j}(\xi_{0t}) \) Assumption A1 means the utility parameters and the unobserved demand shock \( \xi_{jt} \) are identified in the demand model.

Next we assume that the ML estimator solves (13) up to a vanishing tolerance in computation.

**Assumption 5.2 (A2)** \( L_T(\hat{\theta}, \hat{\phi}(\hat{\xi}_t)) \geq \arg\max_{(\theta, \phi) \in \Theta \times \Psi_1^T \times \ldots \times \Psi_J^T} L_T(\theta, \phi(\hat{\xi}_t)) - o_p(1) \).

Next we assume the estimated controls in Section 4.3 are consistent.

**Assumption 5.3 (A3)** \( \hat{\xi}_t \to_p \xi_{0t} \).

When the cost-side unobservables are absent, then this consistency result in Assumption 5.3 is standard in the non-parametric regression literature (additive case), and also in the control function

\(^7\)Here we abstract from the sampling error in the market shares when the aggregate level data are used for estimation. See BLP and Berry, Linton, and Pakes (2004) for explicit treatments of this sampling error.
estimation literature (non-separable case, e.g. Matzkin 2003 and Imbens and Newey 2009). When
the cost-side unobservables are present, we can first consistently estimate the cost errors using non-
parametric IV estimation (e.g. Newey and Powell 2003), as hinted in Berry and Haile (2014), and
then we estimate the demand errors using the estimated cost errors as generated regressors. In this
case too the consistency result can be shown straightforwardly, given the continuity of the regression
function.

Next we require the sieve space become dense as the sample size increases, and assume any
function in $\Psi_j$ is well approximated by a function in the sieve space $\Psi_j^T$ for all $j \in J$. It is well
known that the latter is satisfied when $\Psi_j$ is a collection of bounded and smooth functions (e.g.
Hölder class).

**Assumption 5.4 (A4)** The sieve space $\Psi_j^T$ satisfies $\Psi_j^T \subseteq \Psi_j^{T+1} \subseteq \cdots \subseteq \Psi_j$ for all $T \geq 1$ and for
all $j \in J$; and for any $\phi_j \in \Psi_j$ there exists $\pi_{T}\phi_j \in \Psi_j^T$ such that $\|\phi_j - \pi_{T}\phi_j\|_s \to 0$ as $T \to \infty$ for
all $j \in J$.

We maintain the following continuity conditions. The first one is easy to show for the ML
objective function, and the second one is satisfied when we use polynomials or splines for the
approximating basis functions, for example.

**Assumption 5.5 (A5)** $L^0(\theta, \phi(\cdot))$ is continuous in $(\theta, \phi) \in \Theta \times \Psi_1 \times \cdots \times \Psi_J$.

**Assumption 5.6 (A6)** All $\phi_j(\cdot) \in \Psi_j^T$ is continuous in $\tilde{\xi}_t$ for all $j \in J$.

Next we impose compactness on the sieve space. This is also satisfied when the sieve space is
based on polynomials or splines, for example.

**Assumption 5.7 (A7)** The parameter space $\Theta$ is compact and the sieve space, $\Psi_1^T \times \cdots \times \Psi_J^T$, is
compact under the pseudo-metric $\| \cdot \|_s$.

The last condition we add is that in the neighborhoods of $\tilde{\xi}_0t$ and $\phi_0j(\cdot)$ for any $j \in J$, the
difference between the sample criterion function and the population criterion function is small
enough when $T$ is large. When $L_T(\theta, \phi(\cdot))$ is Lipschitz in $(\theta, \phi)$ this assumption is satisfied with
standard regularity conditions under a proper law of large numbers (e.g., Chebychev’s weak LLN)
(see e.g. Chen 2007).

**Assumption 5.8 (A8)** For all positive sequences $\epsilon_T = o(1)$, we have

$$\sup_{(\theta, \phi) \in \Theta \times \Psi_1^T \times \cdots \times \Psi_J^T, \|\tilde{\xi}_t - \tilde{\xi}_0\| \leq \epsilon_T, \|\phi_j(\cdot) - \phi_0j(\cdot)\|_s \leq \epsilon_T} |L_T(\theta, \phi(\cdot)) - L^0(\theta, \phi(\cdot))| = o_p(1).$$

**Theorem 5.1** Suppose Assumptions A1-A8 are satisfied. Then $\hat{\theta} \overset{p}{\to} \theta_0$ and $\|\hat{\phi}_j(\cdot) - \phi_0j(\cdot)\|_s \overset{p}{\to} 0$
for all $j \in J$. 

See appendix for the proof.

In Appendix F we provide the standard errors that account for the multi-step estimations in the control function approach for given parametric specifications of the first and the second stages. Even though the asymptotics will be different under two different scenarios: parametric model with fixed number of approximating basis functions, and corresponding semiparametric model with increasing number of approximating basis functions, the computed standard errors under two different cases can be numerically identical or equivalent as discussed in Ackerberg, Chen, and Hahn (2012). Therefore, we may ignore the semiparametric nature of the model, and proceed both estimation and inference as if the parametric specification is the true model in practice. See Appendix F for further details.

5.1 How to Choose Specifications and Approximations: A Cross Validation

To implement our control function approach for estimation we need to make choices on several specifications in the utility specification, in the approximation (i.e. specification) of the control function $\phi_j(\cdot)$, and also in the approximation of pricing functions to recover the control $\tilde{\xi}_{jt}$. Although one would prefer richer specifications with abundant data, one needs to sacrifice the flexibility in finite samples due to sampling errors. We propose a cross-validation approach to select the best specification among a set of alternative specifications and approximations.

Cross-validation is a practical tool often employed in econometrics when a researcher needs to select a model specification among alternatives. The idea is to minimize the mean squared error of the final estimator by comparing the out-of-sample fit (on hold-out samples) of candidate models estimated using training samples. The mean squared error accounts for both bias (of choosing more parsimonious specifications) and variance (from sampling error that increases with richer specifications). Cross-validation seeks to balance bias and variance in a particular dataset.

Given a candidate specification $c \in \mathcal{C}$ where $\mathcal{C}$ is the collection of alternative specifications to be considered (variations in specifications include pricing functions, utility functions, and approximations of the unobserved demand factor), the mean squared error (MSE) of the fit to the true market shares is given by

$$\text{MSE}_T(c) \equiv E \left[ \sum_{j=1}^{J} \left( s_{jt}^0 - s_{c,jt}^T \right)^2 \right]$$  \hspace{1cm} (14)

where $s_{jt}^0$ denotes true market shares (i.e. data) and $s_{c,jt}^T$ denotes the predicted market shares of good $j$ at market $t$ from the candidate model $c$.

We need to approximate the MSE in (14) in order to implement cross validation. We propose a variant of cross-validation called “leave one partition out” or “$I$-fold” cross-validation in our setting. We first partition the data into $I$ subsets $T_i$, so that $T_i \cap T_{i'} = \emptyset$ for $i \neq i'$ and $\bigcup_{i=1}^{I} T_i = \{1, 2, \ldots, T\}$. To evaluate the performance of a model specification $c$, we estimate the demands (and generate predicted shares) using each of the $I$ samples where one of the partitions is left out; in other words we use each of the samples $\{1, 2, \ldots, T\} \setminus T_i$ as training samples and the corresponding $T_i$ as hold-out samples. Let $s_{c,jt}^{i,T}$ be the predicted shares leaving out the $i$th sample $T_i$; there are $I$ such estimates.

Let $s_{jt}$ be available unbiased estimates of shares in the sense that $E[s_{jt} - s_{jt}^0|p_t, x_t, z_t, \xi_t, \eta_t] = 0$ by definition of the conditional market shares. Then for $t \in T_i$, the estimate $s_{c,jt}^{i,T}$ which is not using
the observations in $T_i$ yields, under the assumption that the observations are i.i.d. across markets,

$$E \left[ \sum_{j=1}^{J} \left\{ s_{jt}^0 - s_{c,jt}^{-i,T} \right\}^2 \right]$$

$$= \sum_{j=1}^{J} E \left[ \left\{ s_{jt}^0 - s_{jt} \right\}^2 + 2 s_{jt} (s_{jt}^0 - s_{jt}) - 2 (s_{jt}^0 - s_{jt}) s_{c,jt}^{-i,T} \right] + \sum_{j=1}^{J} E \left[ \left\{ s_{jt} - s_{c,jt}^{-i,T} \right\}^2 \right]$$

$$= \sum_{j=1}^{J} E \left[ \left\{ s_{jt}^0 - s_{jt} \right\}^2 + 2 s_{jt} (s_{jt}^0 - s_{jt}) \right] + \sum_{j=1}^{J} E \left[ \left\{ s_{jt} - s_{c,jt}^{-i,T} \right\}^2 \right],$$

where we have $E \left[ 2 \left( s_{jt}^0 - s_{jt} \right) s_{c,jt}^{-i,T} \right] = 0$ by the law of iterated expectations, because $s_{c,jt}^{-i,T}$ does not depend on $s_{jt}$ (i.e., $t \notin T \setminus T_i$) and because $E[s_{jt} - s_{jt}^0| p_t, x_t \cup z_t, \xi_t, \eta_t] = 0$. Note that the first term in (15) does not depend on the model, and that we can consistently estimate the second term in (15) using the sample mean of the observations in the holdout sample. Therefore, one can select an optimal candidate model specification from the cross-validation as

$$c_{opt} = \arg \min_{c \in C} \frac{1}{T} \sum_{i=1}^{I} \sum_{j=1}^{J} \left\{ s_{jt} - s_{c,jt}^{-i,T} \right\}^2$$

because any candidate model $c$ that minimizes the above criterion will also approximately minimize the MSE in (14).

Note that the computational cost of our cross-validation procedure is proportional to the number of partitions $I$ since the estimator of each candidate model to generate predicated shares must be computed $I$ times, one for each partition. Researchers may pick a smaller or a larger $I$ in empirical work based on the computation time for each run.

### 6 Three Empirical Applications

Our empirical applications span three commonly used data types: aggregate (market-level) data, household-level cross-sectional data, and household-level panel data. For each case we briefly describe the data, the demand side model, the instruments, and different options for control functions. Our goal in these applications is to see whether demand and pricing specifications that are assumed to be separable in both observed and unobserved factors yield similar corrected estimates of elasticities for our approach and the Berry (1994) approach.

#### 6.1 Automobiles (the BLP case study)

Our first example is the original BLP (1995) example: price endogeneity in the automobile market. The application is identical to the reported BLP case in almost every respect: data, demand specification, instruments, and estimation. The only difference is that we do not use a supply side model when we estimate the demand side model (so our point estimates only exactly match their
estimates for the cases they examine without the supply side).\(^8\)

### 6.1.1 Data and Demand Specification

The application uses the same 2217 market-level observations on prices, quantities, and characteristics of automobiles sold in the 20 U.S. automobile markets beginning in 1971 and continuing annually to 1990. The utility function used in BLP is\(^9\)

\[
    u_{ijt} = \alpha \ln(y_i - p_{jt}) + \delta_{jt} + \sum_k \sigma_k \nu_{ik} x_{jkt} + \epsilon_{ijt},
\]

where

\[
    \delta_{jt} = \beta_0' x_{jt} + \tilde{\xi}_{jt}, \quad (16)
\]

with \(\tilde{\xi}_{jt}\) the unobserved factor, which we write with the overscore because it is chosen such that

\[
    s(\theta, \delta(\theta)) = s^{Data},
\]

where \(\theta\) includes all parameters but the product-market control, and \(\delta_{jt}\) is obtained from matching observed to predicted shares. \(\alpha\) is the marginal utility of income parameter and income is assumed to follow a log-normal distribution.\(^{10}\) Characteristics \(x_{jt}\) include a constant term, the ratio of horsepower to weight, interior space (length times width), whether air-conditioning is standard (a proxy for luxury), and miles per dollar. The random coefficients on characteristics are assumed to be normally distributed and independent across characteristics with the mean \((\beta_{0k})\) and variance \((\sigma_k)\) (so \(\nu_{ik}\) are the mean zero standard normal deviates).\(^{11}\) \(\epsilon_{ijt}\) is i.i.d. extreme value, so the differenced errors - what’s relevant for choice data - are distributed i.i.d. logit.

The control function specification is similar, and is given by

\[
    u_{ijt} = \alpha \ln(y_i - p_{jt}) + \beta_0' x_{jt} + \sum_k \sigma_k \nu_{ik} x_{jkt} + \lambda_j' \tilde{\xi}_t + \epsilon_{ijt}.
\]

The only difference is, without the \(\delta\)'s, \(\beta_0' x_{jt}\) is included directly, along with the control function.

### 6.1.2 Instruments

Important for both BLP and the control function approach is the determination of prices in the automobile market. BLP consider an equilibrium pricing function of the general form from (11). They follow the literature and assume that observed product characteristics (except price) are uncorrelated with unobserved characteristics \(\tilde{\xi}_{jt}\). (11) implies that in any market \(t\) every product characteristic affects every price in the market, so any product characteristic is a valid instrument for any price. This leads to an abundance of instruments, most of which are likely to be very weak. Pakes (1994) derives the first order basis for the optimal instruments, which amounts to

---

\(^8\)We focus on the demand side for three reasons: it makes the comparison more transparent, most researchers do not impose a supply side model when estimating demands, and the results are easier to replicate.

\(^9\)Consumer \(i\) is in one and only one market \(t\), and \(t(i)\) is a function of \(i\) (we do not explicitly write \(t\) in \(i\)'s presence below).

\(^{10}\)The mean varies annually and the variance assumed to be constant across the twenty years.

\(^{11}\)A variance term is included for the constant to allow for heterogeneity in taste for the outside good.
three instruments for each demand characteristic: the characteristic itself (because characteristics are endogenous), the sum of the characteristic across own-firm products (excluding that product), and the sum of the characteristic across rival firm products. The intuition comes from the first order conditions of the oligopoly pricing equilibrium (from BLP, pg. 855):

“products that face good substitutes will tend to have low markups, whereas other products will have high markups and thus high prices relative to cost. Similarly, because Nash markups will respond differently to own and rival products, the optimal instruments will distinguish between the characteristics of products produced by the same multi-product firm versus the characteristics of products produced by rival firms.”

With 5 characteristics per vehicle, this yields 15 instruments for each product, and we denote this vector $\tilde{Z}_{jt}$.

6.1.3 The Control Functions

We construct an estimate of the expected price for each product conditional on all exogenous factors observed by the econometrician. With the automobile data, very few observations are available on the same nameplate (i.e. the same product) over time, because cars change characteristics and/or exit. This means some restrictions on $E[p_{jt} | \tilde{Z}_{1t}, \ldots, \tilde{Z}_{Jt}]$ across vehicles will be necessary. Some possibilities include assuming that the expected price function is the same across vehicles in the same year, or across similar vehicles, or both. We make a stronger assumption, imposing that the parameters of this function are the same across all products and all years. This yields 2217 observations on this one function.

A second issue arises because of the abundance of arguments in this function (similar to the abundance of instruments in BLP). We follow the logic outlined in Pakes (1994) - described above - and use as arguments for each product $j$ the 15 regressors given by $\tilde{Z}_{jt}$, which reflect both demand and cost factors relevant for each product. The only demographic variable is average annual income, and it has little effect on the predicted values for price, so we define

$$\tilde{\xi}_{jt} \equiv v_j(\xi_t) = p_{jt} - E[p_{jt} | \tilde{Z}_{jt}]$$

and estimate the expectation using ordinary least squares.\footnote{A second order approximation yielded nearly identical results.} Because the expectation is estimated with error, an additional source of error arises. We describe the correction for the standard errors in Appendix A.

The dimensionality problem also arises with the approximation of $\xi_t$. We consider two parsimonious specifications that are based on the assumption that the own-product residual is principally a function of the own-product unobserved factor. For the first specification, only the own-product residual $\tilde{\xi}_{jt}$ from the pricing function enters utility for product $j$. $\lambda_t$, the parameter scaling the price residual, is allowed to vary by year, so

$$\xi_{jt} = \lambda_t \tilde{\xi}_{jt}.$$
The second control function specification we use has three terms in the control function and adds only three new parameters. It is given by

$$
\xi_{jt} = \lambda_1 \tilde{\xi}_{jt} + \lambda_2 \left( \sum_{k \neq j, k \in J_f} \tilde{\xi}_{kt} \right) + \lambda_3 \left( \sum_{k \notin J_f} \tilde{\xi}_{kt} \right).
$$

The motivation for this control function is similar to the motivation for the instruments in Pakes (1994). The first term is the own-product residual (here with one parameter common across years given as \(\lambda_1\)). The sum of other products’ price residuals may also contain information on the magnitude of the own-product’s unobserved demand factor (conditional on all observed factors). We use the same two sums that are proposed for pricing instruments; the sum of all of the other residuals of the products made by the same firm, or \(\sum_{k \neq j, k \in J_f} \tilde{\xi}_{kt}\), where \(J_f\) is the set of products produced by the firm that produces the product \(j\), and the sum of all the residuals of all the products made by other firms, or \(\sum_{k \notin J_f} \tilde{\xi}_{kt}\).\(^{13}\)

### 6.1.4 Estimation and Results

The estimation approach for BLP starts with candidate values of parameters \((\alpha, \sigma)\), where \(\sigma\) is the vector of \(\sigma_k\)’s. The contraction algorithm locates the \(\delta\)’s that match observed to predicted market shares for all 2217 automobiles (the logit error ensures that it converges). These product-market controls are then used in an instrumental variables regression for equation (16) to obtain estimates of \(\beta_0\). The residuals from the estimated equation (16) are then interacted with the instruments to generate the moments that enter the GMM objective function. Minimizing over \((\alpha, \sigma)\) is achieved by iterating over these steps.

Estimation for the control function approach proceeds in two steps. In the first step estimates of \(\tilde{\xi}_{jt} = v_j(\xi_t)\)’s are obtained. In the second step the likelihood function is maximized. Common parameters for both of the control function specifications are \((\alpha, \beta_0, \sigma)\). Specification one includes 20 additional parameters, \((\lambda_71, \ldots, \lambda_{90})\), each indexed by the year of the data. The second specification, motivated by Pakes (1994), includes three additional parameters \((\lambda_1, \lambda_2, \lambda_3)\).

The point estimates and standard errors from these specifications are reported Table A (in Appendix A), and Table 1 translates these estimates into elasticities. The first column uses the uncorrected logit specification from Column 1 of Table III in BLP (1995); because the data sets are the same, these are the same elasticities that result from the coefficients of their Table III. As they report, ignoring price endogeneity severely biases price elasticities towards zero; overall, 67% of them are inelastic.

Columns 2, 3, and 4 report, respectively, specifications one and two of the control function approach and the BLP approach. Column 2, which uses only the own-product price residual with coefficients that vary by year, is very similar to the Column 3 results, which use the three functions of the price residuals and three coefficients (common across years). Both are very similar in almost every respect to the BLP results in Column 4. Across the corrected specifications no automobile price elasticity is inelastic, and the median elasticity is -2.16 for the BLP case, and either -2.08 or

\(^{13}\)Other specifications (for example) might allow the residuals of cars “close” in product space to \(j\) to enter the utility for \(j\).
-2.23, depending on which control function specification is examined. The one difference is that the spread of elasticities is slightly larger for BLP, with a one standard deviation spread of 0.19 vs. 0.10 for the control function approaches. All of the results from Columns 2-4 strongly contrast with the uncorrected results from Column 1; for example, at -0.77, the median uncorrected elasticity is only one-third that from the corrected approaches.

Table 1
Automobile Elasticities: Uncorrected, Control Functions, and BLP

<table>
<thead>
<tr>
<th></th>
<th>No Correction¹</th>
<th>Control Function (1)</th>
<th>Control Function (2)</th>
<th>BLP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Results for 1971-1990</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td>-0.77</td>
<td>-2.08</td>
<td>-2.23</td>
<td>-2.16</td>
</tr>
<tr>
<td>Mean</td>
<td>-1.04</td>
<td>-2.08</td>
<td>-2.22</td>
<td>-2.17</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>0.76</td>
<td>0.10</td>
<td>0.10</td>
<td>0.19</td>
</tr>
<tr>
<td>No. of Inelastic Demands</td>
<td>67%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>Elasticities from 1990²</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>-1.24</td>
<td>-2.11</td>
<td>-2.24</td>
<td>-2.22</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>0.83</td>
<td>0.14</td>
<td>0.14</td>
<td>0.20</td>
</tr>
<tr>
<td>No. of Inelastic Demands</td>
<td>53%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>1990 Models (from BLP, Table VI):</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mazda 323</td>
<td>-0.44</td>
<td>-1.82</td>
<td>-1.94</td>
<td>-1.92</td>
</tr>
<tr>
<td>Honda Accord</td>
<td>-0.82</td>
<td>-2.10</td>
<td>-2.27</td>
<td>-2.17</td>
</tr>
<tr>
<td>Acura Legend</td>
<td>-1.67</td>
<td>-2.25</td>
<td>-2.37</td>
<td>-2.42</td>
</tr>
<tr>
<td>BMW 735i</td>
<td>-3.32</td>
<td>-2.06</td>
<td>-2.21</td>
<td>-2.24</td>
</tr>
</tbody>
</table>

Notes: The uncorrected specification is that from Table III of BLP (1995). 1990 is the year BLP focus on for the individual models; we choose every fourth automobile from their Table VI (the other elasticities were also very similar). The first control function specification allows \(\lambda\) to vary by year; the second specification follows Pakes (1994) (as defined in the text).

BLP report elasticities for selected automobiles from 1990, so we do the same, choosing every fourth automobile from their Table III, in which vehicles are sorted in order of ascending price (the overall average elasticities for 1990 are again very similar between BLP and the control function specifications, and substantially different from the uncorrected approach). The discrepancies between the individual elasticities across the three approaches are small; the absolute value of the difference between BLP and the second control function specification for the Mazda 323, Honda Accord, Acura Legend, and BMW 735i are respectively 0.02, 0.10, 0.05, and 0.03. The discrepancy in the spread of elasticities across all vehicles is also smaller for 1990, as the standard deviations are now 0.14 for the control function approaches vs. 0.2 for the BLP approach. Overall, the corrected
approaches in this application yield very similar elasticity estimates and reject the “no correction” results from Column 1.\textsuperscript{14}

6.2 Multi-Channel Video (Television)

Our second application is only slightly different from Petrin and Train (2010), and it applies the uncorrected and both correction methods to households’ choice among television reception options in 2001. We review the most basic elements here and refer the reader to Petrin and Train (2010) for a more complete discussion.

6.2.1 Data and Demand Specification

The specification is similar to Goolsbee and Petrin (2004). Four alternatives are available to households: (1) antenna only, (2) expanded basic service, (3) expanded basic cable with a premium service added, such as HBO, and (4) satellite dish. The data used is a sample of 11,810 households in 172 geographically distinct markets, where each market contains only one cable franchise. Utility for the control function approach is specified as:

\[ u_{ijt} = \alpha p_{jt} + \sum_{g=2}^{5} \theta_g p_{jt} 1_{ig} + \beta'_0 x_{jt} + \gamma'_d d_i + \sigma \nu_i c_j + \lambda'_j \xi_t + \epsilon_{ijt}. \]  

\textsuperscript{17}x_{jt} are the observed characteristics of the product (including a product intercept term).\textsuperscript{15} The price effect varies across five income groups, with the lowest income group taken as the base and the binary variable \(1_{ig}\) indicating whether household \(i\) is in income group \(g\).\textsuperscript{16} \(\beta'_0 x_{jt}\) denotes the base utility derived from observed product characteristics. Demographic variables for household \(i\) are given by \(d_i\) and enter each choice \(j\) with a separate coefficient vector \(\gamma_j\).\textsuperscript{17} A random coefficient is included to allow for correlation in unobserved utility over the three non-antenna alternatives. In particular, \(c_j = 1\) if \(j\) is one of the three non-antenna alternatives and \(c_j = 0\) otherwise, and \(\nu_i\) is an i.i.d. standard normal deviate. The coefficient \(\sigma\) is the standard deviation of the random coefficient, reflecting the degree of correlation among the non-antenna alternatives. Finally, \(\xi_t = \lambda'_j \xi_t\), with \(\lambda_j\) the vector of parameters associated with the controls, and \(\epsilon_{ijt}\) is i.i.d. extreme value.\textsuperscript{18}

The utility specification with product-market controls is given as:

\[ u_{ijt} = \delta_{jt} + \sum_{g=2}^{5} \theta_g p_{jt} 1_{ig} + \gamma'_d d_i + \sigma \nu_i c_j + \epsilon_{ijt}, \]

\textsuperscript{14}The control function specification nests the uncorrected specification, so one formal test asks whether the \(\lambda's\) from the control function approach enter significantly.

\textsuperscript{15}The attributes we include, which vary over markets, are the channel capacity of a cable system, the number of pay channels available, whether pay per view is available from that cable franchise, the price of expanded basic service, the price of premium service, and the number of over-the-air channels available. Many of the cable operators are owned by multiple system operators (MSO’s) like AT+T and Time-Warner, and we include MSO dummy variables.

\textsuperscript{16}The price coefficient for a household in the lowest income group is \(\alpha\) while that for a household in group \(g > 1\) is \(\alpha + \theta_g\).

\textsuperscript{17}These include family income, household size, education, and type of living accommodations.

\textsuperscript{18}Our approach here differs from Petrin and Train (2010) in the specification of the error distribution (see that paper for details).
Table 2
Television Choice Elasticities: Uncorrected, Control Function, and BLP

<table>
<thead>
<tr>
<th></th>
<th>No Correction</th>
<th>Control Function</th>
<th>BLP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Price of expanded-basic cable</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Antenna-only share</td>
<td>Upward</td>
<td>0.96</td>
<td>0.79</td>
</tr>
<tr>
<td>Expanded-basic cable share</td>
<td>Sloping</td>
<td>-1.18</td>
<td>-0.97</td>
</tr>
<tr>
<td>Premium cable share</td>
<td>Demand</td>
<td>0.99</td>
<td>0.88</td>
</tr>
<tr>
<td>Satellite share</td>
<td></td>
<td>0.95</td>
<td>0.87</td>
</tr>
<tr>
<td>Price of premium cable</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Antenna-only share</td>
<td></td>
<td>0.60</td>
<td>0.52</td>
</tr>
<tr>
<td>Expanded-basic cable share</td>
<td></td>
<td>0.65</td>
<td>0.57</td>
</tr>
<tr>
<td>Premium cable share</td>
<td></td>
<td>-2.36</td>
<td>-2.04</td>
</tr>
<tr>
<td>Satellite share</td>
<td></td>
<td>0.64</td>
<td>0.58</td>
</tr>
<tr>
<td>Price of satellite</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Antenna-only share</td>
<td></td>
<td>0.43</td>
<td>0.42</td>
</tr>
<tr>
<td>Expanded-basic cable share</td>
<td></td>
<td>0.48</td>
<td>0.43</td>
</tr>
<tr>
<td>Premium cable share</td>
<td></td>
<td>0.48</td>
<td>0.45</td>
</tr>
<tr>
<td>Satellite share</td>
<td></td>
<td>-3.79</td>
<td>-3.59</td>
</tr>
</tbody>
</table>

where all of the elements of utility that do not vary within a market are subsumed into the product-market controls, which are a function of price and other observed attributes:

$$\delta_{jt} = \alpha p_{jt} + \beta' x_{jt} + \xi_{jt},$$

with $\xi_{jt}$ the unobserved factor estimated to match observed to predicted market shares for the BLP approach.

For both approaches we use Hausman (1997)-type price instruments. The price instrument for market $t$ is calculated as the average price in other markets that are served by the same multiple system operator as market $t$. A separate instrument is created for the price of expanded-basic cable and the price of premium cable.

6.2.2 The Control Functions

For the base specification we construct the price residuals for expanded basic by regressing the expanded-basic price on all the product attributes listed above for both choices plus both Hausman (1997)-type price instruments (the expanded basic and the premium one).\(^{19}\)

\(^{19}\)Premium residuals are constructed in a similar manner.
Estimation of the control function approach proceeds first by obtaining estimates of the price residuals (as described above). Then, the likelihood function is maximized using the equation for utility from (17). Parameter estimates are reported in Appendix B in Table B.

For the BLP approach we estimate the model with product-market controls using the contraction procedure to solve for the 516 (172*3) additional parameters (conditional on parameters \( \theta \) not captured in the \( \delta_j \)'s). The value of the objective function to estimate the mixed logit model is then computed at this value of \( (\theta, \delta(\theta)) \), and the function is maximized over \( \theta \). After the function is maximized, the resulting \( \hat{\delta}_j \)'s are regressed on the product attributes using 3SLS to estimate the mean utility parameter \( (\alpha, \beta_0) \). These parameter estimates are reported in Petrin and Train (2010).

Table 2 gives price elasticities from the models for each approach. Without any correction for price endogeneity the correlation between price and the unobserved characteristics is so strong that demands are upward sloping (consumers like to pay more). Parameter estimates and standard errors from both the control function approach and the BLP approach reject the uncorrected model. The elasticities from these approaches are very similar, with expanded basic at either -0.97 or -1.18, premium at -2.04 or -2.36, and satellite at -3.59 or -3.79.

6.3 Margarine

Our third application uses household-level panel data to estimate the demand for margarine. The framework and data exactly follows that outlined in Chintagunta, Dubé, and Goh (2005), who use product-market controls to demonstrate that unobserved brand characteristics result in a price endogeneity problem for margarine.\(^\text{20}\)

6.3.1 Data and Demand Specification

The data are weekly purchase histories of 992 households between January 1993 and March 1995 and were collected by Nielsen for the Denver area using checkout-counter scanners. The data for margarine are restricted to the 16 oz. category and the four observed products are Blue Bonnet, I Can’t Believe It’s Not Butter (ICBINB), Parkay, and Shedd’s. Weekly prices and marketing mix variables - including whether the product is on display and whether it is featured - are recorded for every product available in these categories for all 117 weeks. Posted prices may respond to changes in shelf-space, the availability of in-store coupons, or promotions in complementary or substitute categories, all of which are unobserved by the econometrician. Omitted inventories, if correlated across households because of persistence in prices, can also lead to a correlation in price and the unobserved demand shock.

The utility specification for the control function approach is given as

\[
    u_{ijt} = (\alpha_0 + \alpha_i)p_{jt} + (\beta_0 + \beta_i)d_{jt} + (\beta_0F + \beta_iF)F_{jt} + (\beta_0D + \beta_iD)D_{jt} + \lambda_j^\prime \tilde{\xi}_t + \epsilon_{ijt},
\]

\(\text{20}\) They also show this to be true for orange juice. We thank JP Dubé for running the control function specification with his data.
where $p_{jt}$ is the posted price for brand $j$ at time $t$, $F_{jt}$ and $D_{jt}$ are indicators that are one if the brand is on feature or display respectively at time $t$, and $\tilde{\xi}_t$ is the vector of control variates. The common-across-consumers price sensitivity term is given by $\alpha_0$, and similarly the brand specific intercepts and feature and display intercepts are given by $\beta_{0j}$, $j = 1, \ldots, 4$, $\beta_{0F}$, and $\beta_{0D}$ respectively. Consumer specific tastes vary around these mean taste parameters and are given by $\alpha_i$, $\beta_{ij}$, $j = 1, \ldots, 4$, $\beta_{iF}$, and $\beta_{iD}$, which are mean zero multivariate normal draws. These random taste coefficients freely vary and covary across price, feature, display, and the brand intercept terms (seven factors), adding a total of 28 additional parameters that summarize the variance covariance matrix of unobserved taste heterogeneity. Finally, $\epsilon_{ijt}$ is i.i.d. logit.

The utility specification for the fixed effects model is given as:

$$u_{ijt} = \delta_{jt} + \alpha_i p_{jt} + \beta_{ij} + \beta_{iF} F_{jt} + \beta_{iD} D_{jt} + \epsilon_{ijt}.$$  

All of the elements of utility that do not vary for product $j$ in week $t$ are subsumed into the fixed effects, so

$$\delta_{jt} = \alpha_0 p_{jt} + \beta_{0j} + \beta_{0F} F_{jt} + \beta_{0D} D_{jt} + \tilde{\xi}_{jt}.$$  

Wholesale prices are the instruments for the reported shelf price. The price instruments vary weekly for each brand of margarine. These instruments are appropriate if, for example, the unobserved promotional activities at the retail level are uncorrelated with the wholesale price. For the control function specification, the estimator for $\tilde{\xi}_t$ is the (vector of) residuals from the regression of each product’s retail price at time $t$ on an intercept, the list price at the wholesale level, and the discount off list price (at wholesale). Since the number of products is small relative to the number of markets, we are again able to enter all of the residuals into each product’s expression for utility. However, we found again that elasticity estimates were very similar when only the own-product residual entered, and it was the only residual entering significantly. For the results we report the specification where each product has its own coefficient for its residual, so $\xi_{jt} = \lambda_j \tilde{\xi}_{jt}$.

### Table 3

Margarine Own-Price Elasticities: Uncorrected, Control Function, and Fixed Effects

<table>
<thead>
<tr>
<th></th>
<th>No Control Correction</th>
<th>Control Function</th>
<th>Fixed Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blue Bonnet</td>
<td>-1.74</td>
<td>-2.09</td>
<td>-2.05</td>
</tr>
<tr>
<td>ICan’tBINB</td>
<td>-4.64</td>
<td>-5.33</td>
<td>-5.44</td>
</tr>
<tr>
<td>Parkay</td>
<td>-2.69</td>
<td>-3.31</td>
<td>-3.34</td>
</tr>
<tr>
<td>Shedd’s</td>
<td>-3.32</td>
<td>-4.23</td>
<td>-4.20</td>
</tr>
</tbody>
</table>

$^{21}$Wholesale prices of the other products at time $t$ did not enter significantly.
6.3.2 Results

Table 3 gives price elasticities across the three models. Appendix C reports the point estimates and standard errors for each of the three specifications. Without any correction for price endogeneity the correlation between price and the unobserved brand characteristics is strong enough for margarine such that the own-price elasticities are substantially underestimated with no correction. Across brands they increase between 20-35% with either correction. In each of these cases the control function and the fixed effects approach provide elasticity estimates that are very similar: -2.09 vs. -2.05, -5.33 vs. -5.44, -3.31 vs -3.34, and -4.23 vs. -4.20.22

7 Conclusion

In applications of differentiated product models it is rare that all the relevant factors are observed by the econometrician. When some demand factors are omitted, price will typically be correlated with these unobserved factors through the equilibrating mechanism in the market, and this correlation will bias estimated price elasticities.

In this paper we develop a control function method that inverts prices to recover a random variable that is a one-to-one function of the unobserved product attribute, and then uses this control to condition out the dependence of the demand error on price. Our approach has several attractive features. It is available in environments where goods are substitutes or complements, and when more than one unit of the good may be purchased at a time. The approach does not require additive separability between observed and unobserved factors in consumer utility, so (e.g.) unobserved product characteristics or unobserved promotional activities (like advertising) can affect the marginal impact of price on utility.

We present three empirical demand applications that replicate specifications from earlier works - all of which use product-market controls - including Berry, Levinsohn, and Pakes (1995), who look at automobiles, Goolsbee and Petrin (2004), who look at cable television, and Chintagunta, Dubé, and Goh (2005), who look at margarine. These applications are chosen to span a range of markets and potential competitive pricing behaviors, and include three types of different data: aggregate (market-level) data, household-level cross-sectional data, and household-level panel data. We use them to show in practice how one implements the control function approach for these different types of products and aggregation levels. For simple demand and pricing specifications that assume the unobserved factor enters additively, the estimated elasticities are almost identical across the BLP and control function approaches and differ significantly from the uncorrected elasticity estimates, which are biased upward in every case.

22Dubé reported to us similar findings using orange juice, that is, without a correction, results are biased down, but either correction produces similar elasticities (see early versions of their paper for exact details of their orange juice specification, which is similar in flexibility to the margarine specification described here).
Appendix

A Automobile Case-Study Details

Table A contains the estimated demand parameters and standard errors for the automobile data. These parameters yield the reported elasticities in Table 1. The first column of estimates is the specification reported in column one of Table III in BLP (1995), where the dependent variable is the log of good $j$’s market share minus the log of the outside good’s market share. This log-odds ratio is regressed on price and characteristics to estimate the parameters of the utility function (this specification has no random coefficients). The price parameter is sufficiently biased towards zero to result in 67% of the estimated price elasticities being inelastic, which is inconsistent with profit maximizing behavior.\textsuperscript{23} We emphasize again that these parameter estimates and inelastic elasticities have already been reported in BLP (1995) (these results are virtually identical to the results reported there because the data set has almost been perfectly replicated).

The parameter estimates from the control function approach and BLP approach respectively are reported next. The demand specification and data are identical to BLP (1995). The specifications include random coefficients on the characteristics, and price and income enter as \(\ln(y_i - p_{jt})\). The price parameter for BLP we obtain here is similar to that reported in their second specification in Table IV. Again, the BLP results reported here do not impose the supply side model during estimation, and are thus not identical to their point estimates in Table IV.\textsuperscript{24}

\textsuperscript{23}These price (and other) parameters are not directly comparable to the parameters from the control function and BLP specifications.

\textsuperscript{24}For the control function approach, the control variable $\tilde{\xi}_{jt} = v_j(\xi_t)$ is estimated, and used as a regressor in the second stage estimation routine, so the standard errors from the traditional formulas (and output by standard estimation routines) are biased downward. To approximate this additional source of variance in the control function approach, we bootstrap the price regressions. Specifically, we reestimate the expected price with a bootstrapped sample, calculate the implied residuals, and re-estimate the model with these new residuals (otherwise using the original data). We then repeat this exercise over many bootstrapped samples. The variance in the parameter estimates across the bootstrapped samples is then added to the variance from the traditional formulas (which are appropriate when $\xi_{jt}$ is observed without error).
### Table A

Estimated Parameters for Automobile Demand: Uncorrected, Control Function, and BLP

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Variable</th>
<th>No Correction</th>
<th>Control Function</th>
<th>BLP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Term on Price (α)</td>
<td>price</td>
<td>-0.088</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.004)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ln(y - p)</td>
<td></td>
<td>29.743</td>
<td>23.565</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.828)</td>
<td>(0.341)</td>
<td></td>
</tr>
<tr>
<td>Means (β’s)</td>
<td>Constant</td>
<td>-10.071</td>
<td>-4.319</td>
<td>-6.768</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.252)</td>
<td>(0.115)</td>
<td>(27.781)</td>
</tr>
<tr>
<td></td>
<td>HP/Weight</td>
<td>-0.122</td>
<td>1.851</td>
<td>-1.157</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.277)</td>
<td>(0.032)</td>
<td>(3.076)</td>
</tr>
<tr>
<td></td>
<td>Air</td>
<td>-0.034</td>
<td>0.548</td>
<td>-0.067</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.072)</td>
<td>(0.033)</td>
<td>(2.657)</td>
</tr>
<tr>
<td></td>
<td>MP$</td>
<td>0.265</td>
<td>-0.150</td>
<td>0.260</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.043)</td>
<td>(0.004)</td>
<td>(18.624)</td>
</tr>
<tr>
<td></td>
<td>Size</td>
<td>2.342</td>
<td>2.100</td>
<td>3.272</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.125)</td>
<td>(0.009)</td>
<td>(37.989)</td>
</tr>
<tr>
<td>Std. Deviations (σβ’s)</td>
<td>Constant</td>
<td>0.022</td>
<td>0.003</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.005)</td>
<td>(0.322)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>HP/Weight</td>
<td>0.048</td>
<td>3.817</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.020)</td>
<td>(0.173)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Air</td>
<td>0.001</td>
<td>1.233</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.069)</td>
<td>(0.059)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>MP$</td>
<td>0.001</td>
<td>0.001</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.001)</td>
<td>(6.794)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Size</td>
<td>0.008</td>
<td>0.033</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.002)</td>
<td>(0.081)</td>
<td></td>
</tr>
<tr>
<td>Control Function (λ’s)</td>
<td>λ₁</td>
<td>0.065</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.003)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>λ₂</td>
<td>-0.002</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.001)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>λ₃</td>
<td>0.001</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.001)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The demand specification and data are identical to BLP (1995). Column 1 is virtually identical to results reported in Table III. We do not impose the supply side model, so column 3 is not identical to their results reported in Table IV, although some coefficients (including price) are very similar.
B  Television Case-Study Details

Table B gives the estimated parameters and standard errors. The first column of Table B gives the model without any correction for the correlation between price and omitted attributes. The second column applies the control function approach. These results are almost identical to the results from Petrin and Train (2010) and we refer the reader to that paper for a more complete discussion of data and results, including the results from the BLP specification (which is identical).

\footnote{We bootstrap standard errors in a manner similar to that done in Petrin and Train (2010).}
### Table B

Control Function Approach to Modeling TV Reception Choice


Variables enter alternatives in parentheses and zero in other alternatives.

<table>
<thead>
<tr>
<th>Explanatory variable</th>
<th>Uncorrected (Standard errors in parentheses)</th>
<th>With control functions (Standard errors in parentheses)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Price, in dollars per month (1-4)</td>
<td>-.0202 (.0047)</td>
<td>-.0969 (.0400)</td>
</tr>
<tr>
<td>Price for income group 2 (1-4)</td>
<td>.0149 (.0024)</td>
<td>.0150 (.0025)</td>
</tr>
<tr>
<td>Price for income group 3 (1-4)</td>
<td>.0246 (.0030)</td>
<td>.0247 (.0031)</td>
</tr>
<tr>
<td>Price for income group 4 (1-4)</td>
<td>.0269 (.0034)</td>
<td>.0269 (.0035)</td>
</tr>
<tr>
<td>Price for income group 5 (1-4)</td>
<td>.0308 (.0036)</td>
<td>.0308 (.0038)</td>
</tr>
<tr>
<td>Number of cable channels (2,3)</td>
<td>-.0023 (.0011)</td>
<td>.0026 (.0029)</td>
</tr>
<tr>
<td>Number of premium channels (3)</td>
<td>.0375 (.0163)</td>
<td>.0448 (.0233)</td>
</tr>
<tr>
<td>Number of over-the-air channels (1)</td>
<td>.0265 (.0090)</td>
<td>.0222 (.0111)</td>
</tr>
<tr>
<td>Whether pay per view is offered (2,3)</td>
<td>.4315 (.0666)</td>
<td>.5813 (.1104)</td>
</tr>
<tr>
<td>Indicator: ATT is cable company (2)</td>
<td>.1279 (.1195)</td>
<td>-.1949 (.1845)</td>
</tr>
<tr>
<td>Indicator: ATT is cable company (3)</td>
<td>.0993 (.1195)</td>
<td>-.2370 (.1944)</td>
</tr>
<tr>
<td>Indicator: Adelphia Comm is cable company (2)</td>
<td>.3304 (.1224)</td>
<td>.3425 (.1898)</td>
</tr>
<tr>
<td>Indicator: Adelphia Comm is cable company (3)</td>
<td>.2817 (.1511)</td>
<td>.2392 (.2246)</td>
</tr>
<tr>
<td>Indicator: Cablevision is cable company (2)</td>
<td>.6923 (.2243)</td>
<td>.1342 (.3677)</td>
</tr>
<tr>
<td>Indicator: Cablevision is cable company (3)</td>
<td>1.328 (.2448)</td>
<td>.7350 (.3856)</td>
</tr>
<tr>
<td>Indicator: Charter Comm is cable company (2)</td>
<td>.0279 (.1010)</td>
<td>.0580 (.1441)</td>
</tr>
<tr>
<td>Indicator: Charter Comm is cable company (3)</td>
<td>-.0618 (.1310)</td>
<td>-.1757 (.1825)</td>
</tr>
<tr>
<td>Indicator: Comcast is cable company (2)</td>
<td>.2325 (.1107)</td>
<td>.0938 (.2072)</td>
</tr>
<tr>
<td>Indicator: Comcast is cable company (3)</td>
<td>.5010 (.1325)</td>
<td>.1656 (.2262)</td>
</tr>
<tr>
<td>Indicator: Cox Comm is cable company (2)</td>
<td>.2907 (.1386)</td>
<td>-.0577 (.2496)</td>
</tr>
<tr>
<td>Indicator: Cox Comm is cable company (3)</td>
<td>.5258 (.1637)</td>
<td>.0874 (.2954)</td>
</tr>
<tr>
<td>Indicator: Time-Warner is cable company (2)</td>
<td>.1393 (.0974)</td>
<td>-.0817 (.1507)</td>
</tr>
<tr>
<td>Indicator: Time-Warner cable company (3)</td>
<td>.2294 (.1242)</td>
<td>-.0689 (.1891)</td>
</tr>
<tr>
<td>Education level of household (2)</td>
<td>-.0644 (.0220)</td>
<td>-.0619 (.0221)</td>
</tr>
<tr>
<td>Education level of household (3)</td>
<td>-.1137 (.0278)</td>
<td>-.1123 (.0280)</td>
</tr>
<tr>
<td>Education level of household (4)</td>
<td>-.1965 (.0369)</td>
<td>-.1967 (.0372)</td>
</tr>
<tr>
<td>Household size (2)</td>
<td>-.0494 (.0240)</td>
<td>-.0518 (.0241)</td>
</tr>
<tr>
<td>Household size (3)</td>
<td>.0160 (.0286)</td>
<td>.0134 (.0287)</td>
</tr>
<tr>
<td>Household size (4)</td>
<td>.0044 (.0357)</td>
<td>.0050 (.0358)</td>
</tr>
<tr>
<td>Household rents dwelling (2-3)</td>
<td>-.2471 (.0867)</td>
<td>-.2436 (.0886)</td>
</tr>
<tr>
<td>Household rents dwelling (4)</td>
<td>-.2129 (.1562)</td>
<td>-.2149 (.1569)</td>
</tr>
<tr>
<td>Single family dwelling (4)</td>
<td>.7622 (.1523)</td>
<td>.7649 (.1523)</td>
</tr>
<tr>
<td>Residual for expanded-basic cable price (2)</td>
<td>.0805 (.0416)</td>
<td>.0873 (.0418)</td>
</tr>
<tr>
<td>Residual for premium cable price (4)</td>
<td>1.119 (.2668)</td>
<td>2.972 (1.057)</td>
</tr>
<tr>
<td>Alternative specific constant (2)</td>
<td>.1683 (.3158)</td>
<td>2.903 (1.487)</td>
</tr>
<tr>
<td>Alternative specific constant (3)</td>
<td>-.2213 (.4102)</td>
<td>4.218 (2.386)</td>
</tr>
<tr>
<td>Error components, standard deviation (2-4)</td>
<td>.5087 (.6789)</td>
<td>.5553 (.8567)</td>
</tr>
<tr>
<td>Log likelihood at convergence</td>
<td>-14660.84</td>
<td>-14635.47</td>
</tr>
<tr>
<td>Number of observations: 11810</td>
<td>32</td>
<td></td>
</tr>
</tbody>
</table>
C Margarine Case-Study Details

Table C contains the estimated demand parameters and standard errors for the margarine data. These parameters yield the reported elasticities in Table 3. In addition to the parameters listed in the table, there are 28 additional parameters that are associated with the fully flexible multivariate normal taste distribution across the seven variables: price, the four brands, and the feature and display variable. The fixed effects approach has $117 \times 4 = 468$ additional parameters, or 464 more than the control function specification, which has four additional parameters relative to the uncorrected approach, one for each of the brand residuals: $\lambda_{BB}, \lambda_{IC}, \lambda_{PA}, \lambda_{SH}$.

The first column of estimates is the standard logit model with the additional taste heterogeneity, but without controls for price endogeneity. The next columns report coefficient estimates for the control function and the fixed effects approach respectively. The point estimates associated with price are very similar, at -74.55 and -73.98, and are approximately 25% larger than the price coefficient from the standard logit model.
Table C
Estimated Parameters for Margarine Demand: Uncorrected, Control Function, and Fixed Effects

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Variable</th>
<th>No Correction</th>
<th>Control Function</th>
<th>Fixed Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Term on Price ($\alpha_0$)</td>
<td>price</td>
<td>-59.88</td>
<td>-74.55</td>
<td>-73.98</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(2.30)</td>
<td>(3.48)</td>
<td>(5.48)</td>
</tr>
<tr>
<td>Brand means ($\beta_0$’s)</td>
<td>Blue Bonnet</td>
<td>-1.90</td>
<td>-1.22</td>
<td>-1.50</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.10)</td>
<td>(0.13)</td>
<td>(0.20)</td>
</tr>
<tr>
<td></td>
<td>I Can’t BINB</td>
<td>1.11</td>
<td>2.56</td>
<td>2.38</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.22)</td>
<td>(0.35)</td>
<td>(0.54)</td>
</tr>
<tr>
<td></td>
<td>Parkay</td>
<td>-0.73</td>
<td>-0.04</td>
<td>-0.29</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.13)</td>
<td>(0.18)</td>
<td>(0.29)</td>
</tr>
<tr>
<td></td>
<td>Shedds</td>
<td>-1.04</td>
<td>-0.30</td>
<td>-0.33</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.15)</td>
<td>(0.22)</td>
<td>(0.33)</td>
</tr>
<tr>
<td>Promotional controls ($\beta_0$’s)</td>
<td>Feature</td>
<td>0.17</td>
<td>0.20</td>
<td>0.23</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.06)</td>
<td>(0.06)</td>
<td>(0.06)</td>
</tr>
<tr>
<td></td>
<td>Display</td>
<td>1.38</td>
<td>1.28</td>
<td>1.59</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.27)</td>
<td>(0.29)</td>
<td>(0.30)</td>
</tr>
<tr>
<td>Control Function ($\lambda$’s)</td>
<td>$\lambda_{BB}$</td>
<td>17.94</td>
<td>(1.74)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\lambda_{IC}$</td>
<td>43.71</td>
<td>(7.14)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\lambda_{PA}$</td>
<td>2.39</td>
<td>(4.09)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\lambda_{SH}$</td>
<td>10.66</td>
<td>(1.77)</td>
<td></td>
</tr>
<tr>
<td>Log-likelihood</td>
<td></td>
<td>-23703</td>
<td>-23633</td>
<td>-23021</td>
</tr>
<tr>
<td>Total trips</td>
<td></td>
<td>56138</td>
<td>56138</td>
<td>56138</td>
</tr>
<tr>
<td>Total households</td>
<td></td>
<td>992</td>
<td>992</td>
<td>992</td>
</tr>
</tbody>
</table>

Note: All three specifications include a fully-flexible normal variance covariance matrix for taste heterogeneity across the seven variables (a total of 28 parameters): price, four brands, and feature and display.
D Identification of Control Variables

D.1 Single-Product Markets

Assumption D.1 For all \((x_t, z^{(2)}_t, \zeta_t)\), \(p_t = p(x_t, z^{(2)}_t, \zeta_t, \xi_t)\) is strictly monotonic in \(\xi_t\).

Assumption D.2 \(\xi_t\) is independent of \((x_t, z^{(2)}_t, \zeta_t)\).

Theorem D.1 is the identification result for the control variate. It adopts the approach from Matzkin (2003) and Imbens and Newey (2009) to the demand setting.

**Theorem D.1** If Assumptions D.1-D.2 hold, then

\[
F_{p_t \mid x_t, z^{(2)}_t, \zeta_t} (p_t \mid x_t, z^{(2)}_t, \zeta_t) = F_{\xi_t} (\xi_t).
\]

**Proof.** Let \(\xi_t = p^{-1}(x_t, z^{(2)}_t, \zeta_t, p_t)\) denote the inverse of \(p_t = p(x_t, z^{(2)}_t, \zeta_t, \xi_t)\) in its last argument, which exists by Assumption D.1. Then for any given realization of \((x_t, z^{(2)}_t, \zeta_t)\)

\[
F_{p_t \mid x_t, z^{(2)}_t, \zeta_t} (p_t \mid x_t, z^{(2)}_t, \zeta_t) = \Pr(p_t \leq \tilde{p}_t \mid (x_t, z^{(2)}_t, \zeta_t) = c) = \Pr(p(x_t, z^{(2)}_t, \zeta_t, \xi_t) \leq \tilde{p}_t \mid (x_t, z^{(2)}_t, \zeta_t) = c)
\]

\[
= \Pr(\xi_t \leq p^{-1}(x_t, z^{(2)}_t, \zeta_t, \tilde{p}_t) \mid (x_t, z^{(2)}_t, \zeta_t) = c) = \Pr(\xi_t \leq p^{-1}(c, \tilde{p}_t)) = F_{\xi_t} (\tilde{\xi}_t),
\]

where the third equality follows from Assumption D.1 and the fourth equality follows from Assumption D.2. ■

The proof is constructive, suggesting the empirical CDF for \(F_{p_t \mid x_t, z^{(2)}_t, \zeta_t}\) as an estimator for \(F_{\xi_t}\).

Specifically, the control variate is defined as \(\tilde{\xi}_t = F_{p_t \mid x_t, z^{(2)}_t, \zeta_t} (p_t \mid x_t, z^{(2)}_t, \zeta_t)\), a one-to-one function of \(\xi_t\).

D.2 Multi-Product Markets

Here we prove Theorem 4.1.

**Proof.** (Theorem 4.1) First, as discussed in the main text, Assumptions 4.7-4.8 imply that the pricing function can be written as \(p_{jt} = p_j(x_t, z^{(2)}_t, \zeta_t, v_j(\xi_t))\), with \(p_j(\cdot)\) strictly increasing in \(\xi_j t \equiv v_j(\xi_t)\). Then the inverse \(\tilde{\xi}_{jt} = p_j^{-1}(x_t, z^{(2)}_t, \zeta_t, p_{jt})\) exists, and we have \(\tilde{\xi}_{jt} = F_{p_{jt} \mid x_t, z^{(2)}_t, \zeta_t}\) by the similar argument to Theorem D.1. Then, under Assumptions 4.7-4.8, we have for any \((x_t, z^{(2)}_t, \zeta_t)\)

\[
\frac{\partial p_j(x_t, z^{(2)}_t, \zeta_t, v_j(\xi_t))}{\partial \xi_j} = \frac{\partial p_j(x_t, z^{(2)}_t, \zeta_t, v_j(\xi_t))}{\partial \xi_j} \frac{\partial v_j(\xi_t)}{\partial \xi_j}.
\]

Based on (18), we can construct the following equation of \(J \times J\) matrices such that

\[
\frac{\partial p(x_t, z^{(2)}_t, \zeta_t, \xi_t)}{\partial \xi_j} = \text{diag} \left( \frac{\partial p_1(x_t, z^{(2)}_t, \zeta_t, v_1(\xi_t))}{\partial v_1(\xi_t)}, \ldots, \frac{\partial p_J(x_t, z^{(2)}_t, \zeta_t, v_J(\xi_t))}{\partial v_J(\xi_t)} \right) \frac{\partial v(\xi_t)}{\partial \xi_j}
\]

where \(v(\xi_t) = (v_1(\xi_t), \ldots, v_J(\xi_t))'\). Note that diag \(\left( \frac{\partial p_1(x_t, z^{(2)}_t, \zeta_t)}{\partial v_1(\xi_t)}, \ldots, \frac{\partial p_J(x_t, z^{(2)}_t, \zeta_t)}{\partial v_J(\xi_t)} \right)\) is invertible because
of Assumption 4.8, \( \frac{\partial p_j(z_t(2), \xi_t, \xi_t)}{\partial v_j(\xi_t)} > 0 \) for all \( j \in J \). We have

\[
\frac{\partial v(\xi_t)}{\partial \xi_t} = \begin{pmatrix} \frac{\partial p_1(x_t, z_t(2), \xi_t, v_j(\xi_t))}{\partial v_1(\xi_t)} & \cdots & \frac{\partial p_J(x_t, z_t(2), \xi_t, v_J(\xi_t))}{\partial v_J(\xi_t)} \end{pmatrix}^{-1} \frac{\partial p(x_t, z_t(2), \xi_t, \xi_t)}{\partial \xi_t}.
\]

We conclude \( \frac{\partial \hat{\xi}_t}{\partial \xi_t} = \frac{\partial v(\xi_t)}{\partial \xi_t} \) is invertible because \( \left\{ \frac{\partial p_1(z_t(2), \xi_t, \xi_t)}{\partial v_1(\xi_t), \cdots, \frac{\partial p_J(z_t(2), \xi_t, \xi_t)}{\partial v_J(\xi_t)} \right\}^{-1} \) has the full rank, and \( \frac{\partial p(\xi_t)}{\partial \xi_t} \) has the full rank by Assumption 4.5 and because the product of two full rank matrices is also invertible. Therefore, \( \hat{\xi}_t \) is a one-to-one function of \( \xi_t \).

Next we consider the case that price is written as additively separable with the unobserved demand factors:

\[ p_{jt} = g_j(x_t, z_t(2), \xi_t) + v_j(\xi_t) \quad \text{for} \quad \forall j \in J. \]

Then this obviously satisfies Assumptions 4.7-4.8. Therefore we conclude

**Corollary D.1** Suppose the price equation can be written as (19). Let \( \hat{\xi}_{jt} = p_{jt} - E[p_{jt}|x_t, z_t(2), \xi_t] \). Then, \( \hat{\xi}_t = (\hat{\xi}_{1t}, \ldots, \hat{\xi}_{Jt})' \) is a one-to-one function of \( \xi_t \).

**Proof.** Invertibility of \( p(x_t, z_t(2), \xi_t, \xi_t) \) in \( \xi_t \) (holding the others \( (x_t, z_t(2), \xi_t) \) fixed) implies \( v(\xi_t) \) is a one-to-one function of \( \xi_t \), where \( v(\xi_t) = p_t - g(x_t, z_t(2), \xi_t) \) (all of these objects are \( J \times 1 \) vectors) where \( g(\cdot) = (g_1, \ldots, g_J)' \). Then

\[ \hat{\xi}_t = p_t - E[p_t|x_t, z_t(2), \xi_t] = p_t - g(x_t, z_t(2), \xi_t) - E[v(\xi_t)] = v(\xi_t) - E[v(\xi_t)] \]

is also invertible in \( \xi_t \), because subtracting off a vector of constants \( E[v(\xi_t)] \) from a function invertible in \( \xi_t \) yields a function that is still invertible in \( \xi_t \) (or simply we can normalize \( E[v(\xi_t)] = 0 \) w.o.l.g.).

**E Proof of Consistency (Theorem 5.1)**

We prove the consistency by extending Chen (2007)'s consistency theorem, allowing for estimated controls. The result shows that given continuity of control functions, the fact that we use estimated control variables instead of the true ones in the ML estimation does not complicate the consistency of the proposed ML estimator.

Let \( \varepsilon > 0 \) be any small real numbers. By Assumption A2, the ML estimator \( (\hat{\theta}, \hat{\phi}) \) satisfies that with probability approaching to one (w.p.a.1), \( L_T(\hat{\theta}, \hat{\phi}(\hat{\xi}_t)) > L_T(\theta, \phi_T(\hat{\xi}_t)) - \frac{\varepsilon}{6} \) for all \( (\theta, \phi_T) \in \Theta \times \Psi_T^1 \times \cdots \times \Psi_T^J \). From the fact that \( \theta_0 \in \Theta \) and \( \pi_T \phi_0 \in \Psi_T^1 \times \cdots \times \Psi_T^J \), it follows that w.p.a.1

\[ L_T(\hat{\theta}, \hat{\phi}(\hat{\xi}_t)) > L_T(\theta_0, \pi_T \phi_0(\hat{\xi}_t)) - \frac{\varepsilon}{6}. \]

(20)

Also by the uniform convergence (Assumption A8), we have w.p.a.1, \( L^0(\hat{\theta}, \hat{\phi}(\hat{\xi}_t)) + \frac{\varepsilon}{6} > L_T(\hat{\theta}, \hat{\phi}(\hat{\xi}_t)) \)

36
and \( L_T(\theta_0, \pi_T \phi_0(\tilde{\xi}_t)) - I^0(\theta_0, \pi_T \phi_0(\tilde{\xi}_t)) > -\frac{\varepsilon}{6} \). From (20) it follows that w.p.a.1,

\[
L^0(\hat{\theta}, \hat{\phi}(\tilde{\xi}_t)) + \frac{\varepsilon}{6} > L^0(\theta_0, \pi_T \phi_0(\tilde{\xi}_t)) - \frac{\varepsilon}{6} - \frac{\varepsilon}{6}.
\]

Next we note that, by the continuity assumptions (Assumption A5-A6) and the consistency of the estimator of controls (Assumption A3) in the first stage, we have w.p.a.1, \( L^0(\hat{\theta}, \hat{\phi}(\tilde{\xi}_t)) - L^0(\hat{\theta}, \hat{\phi}(\tilde{\xi}_t)) < \frac{\varepsilon}{6} \) and \( L^0(\theta_0, \pi_T \phi_0(\tilde{\xi}_t)) - L^0(\theta_0, \pi_T \phi_0(\tilde{\xi}_t)) > -\frac{\varepsilon}{6} \). It follows that w.p.a.1,

\[
L^0(\hat{\theta}, \hat{\phi}(\tilde{\xi}_t)) + \frac{\varepsilon}{6} > L^0(\theta_0, \pi_T \phi_0(\tilde{\xi}_t)) - \frac{\varepsilon}{6} - \frac{3\varepsilon}{6}.
\]

By Assumption A1 (identification) and A5 (continuity) and the fact that \( \|\phi_0j - \pi_T \phi_0j\|_s \to 0 \) for all \( j \in J \) as \( T \to \infty \), for all \( T > T_0 \) large enough we have \( L^0(\theta_0, \pi_T \phi_0(\tilde{\xi}_t)) > L^0(\theta_0, \pi_T \phi_0(\tilde{\xi}_t)) - \frac{\varepsilon}{6} \).

It follows that w.p.a.1,

\[
L^0(\hat{\theta}, \hat{\phi}(\tilde{\xi}_t)) > L^0(\theta_0, \phi_0(\tilde{\xi}_t)) - \varepsilon.
\]

Next note that for any \( \delta > 0 \), by Assumption A4, A5(continuity), A7 (compactness),

\[
L_0^0 \equiv \sup_{\{\theta, \phi\} \in \Theta \times \Psi_l, \ldots, \Psi_j, \|\theta - \theta_0\| + \sum_j \|\phi_j - \phi_0j\|_s \geq \delta} L^0(\theta, \phi(\tilde{\xi}_t))
\]

exists. Then by Assumption A1 (identification) and the fact that \( \Psi_1^T \times \ldots \times \Psi_J^T \subset \Psi_1 \times \ldots \times \Psi_J \), it must be that \( L^0(\theta_0, \phi_0(\tilde{\xi}_t)) > L_0^0 \). We take \( L^0(\theta_0, \phi_0(\tilde{\xi}_t)) - L_0^0 = \varepsilon \). It follows that w.p.a.1, \( L^0(\hat{\theta}, \hat{\phi}(\tilde{\xi}_t)) > \sup_{\{\theta, \phi\} \in \Theta \times \Psi_l, \ldots, \Psi_j, \|\theta - \theta_0\| + \sum_j \|\phi_j - \phi_0j\|_s \geq \delta} L^0(\theta, \phi(\tilde{\xi}_t)) \). Then by Assumption A5 (continuity) and the fact that \( (\hat{\theta}, \hat{\phi}) \in \Theta \times \Psi_l, \ldots, \Psi_j \), we conclude \( \|\hat{\theta} - \theta_0\| + \sum_j \|\hat{\phi}_j - \phi_0j\|_s < \delta \) for arbitrary small \( \delta \) w.p.a.1. This completes the consistency proof.

**F Standard Errors**

We provide standard errors that account for the multi-step estimation in the control function approach. We present a general model that can be tailored to specific applications. We focus on the additively separable setting, which would be easier to implement in applied work, as we demonstrate with the three empirical applications. Extending the result to the non-separable setting would be possible, but with additional burden of notation.

The control function approach is implemented in two stages. In the first stage we estimate the unobserved cost shock (if the supply side model is used), and we estimate controls from the pricing equations. Then, in the second stage, we estimate parameters in the utility including the parameters for the distribution of random coefficients. We collect all the parameters of the utility in \( \theta \), including the mean utility parameters and the parameters in the distribution of random coefficients. We derive the results for given parametric specifications of all stages and later discuss how the results can extend to semiparametric models.

We first estimate \( \eta_l \) (so \( \zeta_t = \tilde{z}_t^{(1)} + \eta_t \)) using the non-parametric IV as described in Section 4.1. Write the estimate as \( \hat{\eta}_t = -\tilde{z}_t^{(1)} + \tilde{\phi}_j^{(1)}(s_t, p_t, x_t, \tilde{z}_t^{(2)}) \) and \( \hat{\zeta}_t = \tilde{z}_t^{(1)} + \hat{\eta}_t \). Then the estimates of the
controls $\tilde{\xi}_t$ are obtained from regressions of $p_{jt}$ on $(x_t, z_t^{(2)}, \hat{\zeta}_t)$ as

$$\hat{\xi}_{jt} = p_{jt} - \tilde{W}_t' \tilde{\pi}_j, \forall j \in J$$

where $\tilde{W}_t = (x_t, z_t^{(2)}, \hat{\zeta}_t)$ or $\tilde{W}_t$ can also include (e.g.) higher order polynomials of $(x_t, z_t^{(2)}, \hat{\zeta}_t)$. We let $T$ be the sample size (e.g. number of markets or number of time periods) in the first stage estimation, $N$ be the sample size (e.g. number of markets or number of households) for the second stage estimation, and $n$ be the size of consumer sample from which the market shares are calculated. We denote the estimate of the true $\xi_{0t}$ as $\hat{\xi}_t$, and observed market shares as $s_t^n = (s_{1t}^n, \ldots, s_{Jt}^n)$. With some abuse of notation let $\bar{\theta} = (\theta', \lambda_1', \ldots, \lambda_J')'$ denote all the parameters to be estimated in the second-stage ML estimation including the control functions $\phi_j(\hat{\xi}_t) = \sum_{r=1}^R \lambda_j^r \varphi_j^r(\hat{\xi}_t)$ for all $j \in J$ with some approximating basis functions $\varphi_j^r(\hat{\xi}_t)$. Similarly let $\hat{\theta} = (\bar{\theta}', \bar{\lambda}_1', \ldots, \bar{\lambda}_J')'$ be the second-stage ML estimator of the true $\theta_0$. The inference for the utility parameter estimate $\bar{\theta}$ is then obtained as a subset inference from the asymptotic distribution of $\hat{\theta}$.

We start with the aggregate (market-level) data case. In this case we have $T = N$. The utility parameters are estimated using MLE and we write the sample ML objective function as

$$L_T(\bar{\theta}, \hat{\xi}_t, s_t) = \sum_{t=1}^T \sum_{j=1}^J s_{jt} \ln G_j(\delta_t(\hat{\xi}_t); \bar{\theta}) / T$$

where we let $\delta_t$ denote the mean utility (the utility term that does not depend on the idiosyncratic consumer tastes) of the good $j$ in market $t$, $\delta_t = (\delta_{1t}, \ldots, \delta_{Jt})$, and $G_j(\cdot)$ denotes the predicted choice probability of taking the choice $j$ such as the logit or mixed logit choice probability. Here and in what follows, we suppress other demand factors $(p_t, x_t)$ in $\delta_t$ to ease notation. We further write $G_{jt} = G_j(\delta_t(\xi_{0t}); \theta_0)$.

The ML estimator solves the first order condition $\frac{\partial L_T(\hat{\xi}_t, s_t)}{\partial \theta} = 0$ (up to some asymptotically negligible term). Define $\Gamma_{T,n} = -\frac{\partial L_T(\theta, \hat{\xi}_t, s_t^n)}{\partial \theta \partial \theta'}$ and $\Gamma_0 = \text{plim}_{T,n \to \infty} \Gamma_{T,n}$. The asymptotic distribution of $\sqrt{T}(\hat{\theta} - \theta_0)$ is obtained using the asymptotic expansion (obtained from the mean-value expansion of the FOC around $\theta_0$):

$$\sqrt{T}(\hat{\theta} - \theta_0) = \Gamma_{T,n}^{-1} \sqrt{T} \left( \frac{\partial L_T(\bar{\theta}, \hat{\xi}_t, s_t^n)}{\partial \theta} \right) + o_p(1) = \Gamma_{T,n}^{-1} \sqrt{T} \left( \frac{\partial L_T(\theta_0, \hat{\xi}_{0t}, s_t)}{\partial \theta} \right) + \left( \frac{\partial L_T(\theta_0, \hat{\xi}_t, s_t^n)}{\partial \theta} - \frac{\partial L_T(\theta_0, \hat{\xi}_t, s_t)}{\partial \theta} \right) + o_p(1).$$

Therefore the asymptotic variance of $\sqrt{T}(\hat{\theta} - \theta_0)$ contains three variance terms that arise from (i) the ML estimation, (ii) the first stage estimation, and (iii) using the observed market shares instead of the true market shares. We write these three terms as

$$\sqrt{T}(\hat{\theta} - \theta_0) \rightarrow_d \mathcal{N}(0, \Gamma_0^{-1}(\Gamma_0 + V_1 + V_2)\Gamma_0^{-1}).$$

To analyze the second term we define infeasible estimates of controls, obtained from regressions of
\[ p_{jt} \text{ on } (x_t, z_t^{(2)}, \zeta_t), \]
\[ \hat{\xi}_{jt} = p_{jt} - W_t^o \pi_j^o, \quad \forall j \in J \]  
(23)

where \( W_t \) is constructed by replacing \( \zeta_t \) with \( \tilde{\zeta}_t \) in \( \tilde{W}_t \). Also define another infeasible estimates of controls as \( \hat{\xi}_{jt}^o = p_{jt} - W_t^o \pi_{0j}, \quad \forall j \in J \) which are obtained when \( \zeta_t \) is estimated in the first step but the true parameter \( \pi_{0j} \) is known where \( \hat{\xi}_{j0t} = p_{jt} - W_t^o \pi_{0j} \).

Note that for the second term in (21), we can further write
\[
\sqrt{T} \left( \frac{\partial L_T(\hat{\theta}_0, \hat{\xi}_t, s_t)}{\partial \theta} - \frac{\partial L_T(\hat{\theta}_0, \hat{\xi}_0t, s_t)}{\partial \theta} \right)
= \sqrt{T} \left( \frac{\partial L_T(\hat{\theta}_0, \hat{\xi}_t, s_t)}{\partial \theta} - \frac{\partial L_T(\hat{\theta}_0, \hat{\xi}_0t, s_t)}{\partial \theta} \right) + \sqrt{T} \left( \frac{\partial L_T(\hat{\theta}_0, \hat{\xi}_t, s_t)}{\partial \theta} - \frac{\partial L_T(\hat{\theta}_0, \hat{\xi}_0t, s_t)}{\partial \theta} \right)
= \sqrt{T} \left( \frac{\partial L_T(\hat{\theta}_0, \hat{\xi}_t, s_t)}{\partial \theta} - \frac{\partial L_T(\hat{\theta}_0, \hat{\xi}_0t, s_t)}{\partial \theta} \right) + o_p(1)
\]  
(24)

(25)

In (25) the first term is about the contribution to the variance by estimating the unobservable \( \hat{\xi}_{0t} \) in the second step of the first stage, and the second term is about the contribution to the variance by estimating the cost shock \( \eta_t \) in the first step of the first stage. The important result is obtained in (26) following Hahn and Ridder (2013) who show that the two terms in (25) reduce to the one term in (26). This term in (26) is what one would obtain if one uses the infeasible estimates of the controls \( \hat{\xi}_t^o \) defined in (23) instead of the actual estimates \( \hat{\xi}_t \). This implies that the first step estimation of \( \eta_t \) does not contribute to the asymptotic variance of the second stage estimator \( \hat{\theta} \).

Now we further expand the term in (26) to obtain the influence function
\[
\sqrt{T} \left( \frac{\partial L_T(\hat{\theta}_0, \hat{\xi}_t, s_t)}{\partial \theta} - \frac{\partial L_T(\hat{\theta}_0, \hat{\xi}_0t, s_t)}{\partial \theta} \right)
= \frac{1}{T} \sum_{t=1}^{T} \sum_{j=1}^{J} s_{jt} \sum_{j' = 1}^{J} \sum_{j'' = 1}^{J} \sum_{j'''}^{J} \frac{\partial^2 \ln G_{jt}}{\partial \hat{\theta} \partial \delta_{j't}} \frac{\partial \delta_{j''t}(\hat{\xi}_0t)}{\partial \delta_{j''''t}} \sqrt{T} (\hat{\pi}_{j'''}^o - \pi_{j''''o}) + o_p(1)
\]  
(27)

where \( \sqrt{T} (\hat{\pi}_{j'''}^o - \pi_{j''''o}) = (\frac{1}{T} \sum_{t=1}^{T} W_t W_t')^{-1} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} W_t^o \hat{\xi}_{j0t} \). Let \( \sqrt{T} (\hat{\pi}_{j''}^o - \pi_{j''o}) \to_d \omega_j \) where \((\omega_1', \ldots, \omega_J')'\) follow a joint normal distribution and define
\[
\Lambda_{j''}(\tilde{\theta}, \tilde{\xi}_t, s_t) = \frac{1}{T} \sum_{t=1}^{T} \sum_{j=1}^{J} s_{jt} \sum_{j' = 1}^{J} \sum_{j'' = 1}^{J} \sum_{j'''}^{J} \frac{\partial^2 \ln G_{jt}(\delta_{j''t}(\hat{\xi}_0t); \tilde{\theta})}{\partial \hat{\theta} \partial \delta_{j't}} \frac{\partial \delta_{j''t}(\hat{\xi}_0t)}{\partial \delta_{j''''t}} \sqrt{T} (\hat{\pi}_{j''''t} - \pi_{j''''t})
\]

Then lastly we can derive the limit distribution of (24) as
\[
\sqrt{T} \left( \frac{\partial L_T(\hat{\theta}_0, \hat{\xi}_t, s_t)}{\partial \theta} - \frac{\partial L_T(\hat{\theta}_0, \hat{\xi}_0t, s_t)}{\partial \theta} \right)
= \sum_{j'' = 1}^{J} \Lambda_{j''}(\tilde{\theta}, \tilde{\xi}_0t, s_t) \omega_{j''} + o_p(1) \to_d \mathcal{N}(0, \sum_{j,k} \Lambda_{0j} \text{Cov}(\omega_j, \omega_k) \Lambda_{0k}^t) = \mathcal{N}(0, V_1)
\]
where \( \Lambda_{0j} = \lim_{T \to \infty} A^T_j(\tilde{\theta}_0, \tilde{\xi}_0, s_t) \) for \( \forall j \in J \). We can consistently estimate \( \Gamma_0 \) by \( \hat{\Gamma} = -\frac{\partial L_T(\hat{\theta}, \hat{\xi}, s_t)}{\partial \theta \partial \vartheta} \) and \( V_1 \) by \( \hat{V}_1 = \sum_{j,k} A^T_j(\hat{\theta}, \hat{\xi}_t, s_t) \hat{\text{Cov}}(\bar{w}_j, \bar{w}_k) A^T_k(\hat{\theta}, \hat{\xi}_t, s_t)' \) where one can also estimate \( \text{Cov}(\bar{w}_j, \bar{w}_k) \) as

\[
\hat{\text{Cov}}(\bar{w}_j, \bar{w}_k) = \left( \sum_{t=1}^{T} \hat{W}_t \hat{W}_t' / T \right)^{-1} \left( \sum_{t=1}^{T} \hat{\xi}_jt \hat{\xi}_kt \hat{W}_t \hat{W}_t' / T \right) \left( \sum_{t=1}^{T} \hat{W}_t \hat{W}_t' / T \right)^{-1}.
\]  

(28)

Also note that the derivative terms in \( A^T_j(\hat{\theta}, \hat{\xi}_t, s_t) \) are easy to calculate for a given specification of the model \( (G_j(\cdot) \) and \( \delta_{jt}(\hat{\xi}_t) \). When one assumes homoskedasticity and zero correlations of \( \hat{\xi}_{jt0} \) across \( j \neq k \), \( V_1 \) simplifies to \( V_1 = \sum_{j=1}^{J} \Lambda_{0j} E[W_t W_t']^{-1} E[\hat{\xi}_{jt0}^2] A^T_{0j} \), and its consistent estimator is similarly obtained. We can also approximate \( V_2 = \lim_{T,n \to \infty} \frac{T}{n} \text{Var} \left\{ \sqrt{n} \left( \frac{\partial L_T(\theta, \hat{\xi}_t, s_t)}{\partial \theta} - \frac{\partial L_T(\theta_0, \hat{\xi}_t, s_t)}{\partial \theta} \right) \right\} \)

to allow for sampling errors of the market shares, but this term will be negligible when \( n \) is large enough. When we ignore the term \( V_2 \) and use the same data in the first and the second stage estimation, the asymptotic variance-covariance matrix in (22) would become identical to the one that one can obtain following the approach in Murphy and Topel (1985). Therefore we can approximate the variance matrix of \( \hat{\theta} \) reasonably well with \( \hat{\Gamma}^{-1}(\hat{\Gamma} + \hat{V}_1) \hat{\Gamma}^{-1} \). Then, the estimator of the asymptotic variance of \( \hat{\theta} \) is obtained as a sub-matrix of \( \hat{\Gamma}^{-1}(\hat{\Gamma} + \hat{V}_1) \hat{\Gamma}^{-1} \).

Now we consider the case that household-level cross-sectional data (e.g. Television case) or household-level panel data (Margarine case) is used in the second stage estimation. In this case we write the sample ML objective function as

\[
L_N(\hat{\theta}, \hat{\xi}_t) = \frac{1}{N} \sum_{t=1}^{T} \sum_{i=1}^{N_t} \sum_{j=1}^{J} d_{ijt} \ln G_{ij}(\delta_i(\hat{\xi}_t); \hat{\theta})
\]

where \( N = \sum_{t=1}^{T} N_t \) and \( d_{ijt} \) denotes the choice of the household \( i \), which equals to one if the household \( i \) chooses \( j \) and equals to zero otherwise, and \( G_{ij}(\cdot) \) denotes the predicted choice probability of the alternative \( j \). We write \( G_{ij} = G_{ij}(\delta_i(\hat{\xi}_t); \hat{\theta}) \). Here the ML estimator solves the first order condition \( \frac{\partial L_N(\hat{\xi}_t)}{\partial \vartheta} = 0 \) (up to some asymptotically negligible term). Define \( \Gamma_{N,T} = -\frac{\partial L_N(\theta, \hat{\xi}_t)}{\partial \theta \partial \vartheta} \) and \( \Gamma_0 = \text{plim}_{N,T \to \infty} \Gamma_{N,T} \). The asymptotic distribution of \( \sqrt{N}(\hat{\theta} - \theta_0) \) is obtained using the asymptotic expansion:

\[
\sqrt{N}(\hat{\theta} - \theta_0) = \Gamma_{N,T}^{-1} \left\{ \sqrt{N} \left( \frac{\partial L_N(\theta_0, \hat{\xi}_t)}{\partial \theta} + \sqrt{N} \left( \frac{\partial L_N(\hat{\theta}, \hat{\xi}_t)}{\partial \theta} - \frac{\partial L_N(\theta_0, \hat{\xi}_t)}{\partial \theta} \right) \right) \right\} + o_p(1).
\]  

(29)

Therefore the asymptotic variance of \( \sqrt{N}(\hat{\theta} - \theta_0) \) contains two variance terms that are from (i) the ML estimation and (ii) the first stage estimation. We write these two terms as

\[
\sqrt{N}(\hat{\theta} - \theta_0) \to_d N(0, \Gamma_0^{-1}(\Gamma_0 + V_1) \Gamma_0^{-1}).
\]

To derive the variance term \( V_1 \), we approximate the second term in (29), similar to (24)-(26) and
distribution of the estimator can view our estimator as a semiparametric multi-stage sieve estimator. Even though the asymptotic or regressors, one can view them as non-parametric approximations (under the promise that we ϕ for the approximation of J j

been studied in Ackerberg, Chen, and Hahn (2012). if the parametric model is true when we calculate the standard errors. This equivalence result has number of approximation terms), the computed standard errors can be made numerically identical with fixed number of approximating terms, and corresponding semiparametric model with growing approximation terms. One is for the approximation of

ϕ J j

parametric approximations. One is for the approximation of

s J j

E[ p_J t | x_t, z_l T2], E[ p_J t | x_t, z_l T2], and Φ_j(ξ_t) with fixed number of approximating basis functions or regressors, one can view them as non-parametric approximations (under the promise that we would use more flexible specifications for the three objects as the sample size grows). Therefore, one can view our estimator as a semiparametric multi-stage sieve estimator. Even though the asymptotic distribution of the estimator ̂θ would be different under two asymptotic scenarios (parametric model with fixed number of approximating terms, and corresponding semiparametric model with growing number of approximation terms), the computed standard errors can be made numerically identical or equivalent. This means we can ignore the semiparametric nature of the model, and proceed as if the parametric model is true when we calculate the standard errors. This equivalence result has been studied in Ackerberg, Chen, and Hahn (2012).

\[
\sqrt{N} \left( \frac{\partial L_N(\hat{\theta}_0, \hat{\xi}_t)}{\partial \theta} - \frac{\partial L_N(\hat{\theta}_0, \hat{\xi}_0)}{\partial \theta} \right) = \frac{\sqrt{N}}{\sqrt{T}} \sum_{t=1}^{T} \sum_{i=1}^{N_t} \sum_{j=1}^{J} d_{ijt} \sum_{j'=1}^{J} \frac{\partial^2 \ln G_{ijt}}{\partial \theta \partial \delta_{j't}} \sum_{j''=1}^{J} \frac{\partial \delta_{j''t}(\hat{\xi}_0)}{\partial \xi_{j''t}} \sqrt{T} \left( \pi_{j''t} - \pi_{j''0} \right) + o_p(1). \tag{30}
\]

Define

\[
\Lambda_{j''}^{N,T}(\hat{\theta}, \hat{\xi}_t) = \frac{1}{N} \sum_{t=1}^{T} \sum_{i=1}^{N_t} \sum_{j=1}^{J} d_{ijt} \sum_{j'=1}^{J} \frac{\partial^2 \ln G_{ijt}(\xi_t; \hat{\theta})}{\partial \theta \partial \delta_{j't}} \frac{\partial \delta_{j''t}(\hat{\xi}_0)}{\partial \xi_{j''t}} \pi_{j''t}.
\]

Then using similar notation with the aggregate data case, we can rewrite (30) as

\[
\sqrt{N} \left( \frac{\partial L_N(\hat{\theta}_0, \hat{\xi}_t)}{\partial \theta} - \frac{\partial L_N(\hat{\theta}_0, \hat{\xi}_0)}{\partial \theta} \right) \rightarrow_d N(0, \lim_{N,T \to \infty} \frac{N}{T} \sum_{j,k}^{J} \Lambda_{0j} \text{Cov}(\bar{\pi}_j, \bar{\pi}_k) \Lambda_{0k}) = N(0, V_1)
\]

where \( \Lambda_{0j} = \lim_{N,T \to \infty} \Lambda_{j}^{N,T}(\hat{\theta}_0, \hat{\xi}_0) \) for \( \forall j \in J \). To obtain a consistent estimator of the variance matrix, we use \( \hat{\Gamma} = -\frac{\partial L_N(\hat{\theta}_0, \hat{\xi}_0)}{\partial \theta \partial \theta'} \) and \( \hat{V}_1 = \frac{N}{T} \sum_{j,k}^{J} \Lambda_{j}^{N,T}(\hat{\theta}, \hat{\xi}_t) \text{Cov}(\bar{\pi}_j, \bar{\pi}_k) \Lambda_{j}^{N,T}(\hat{\theta}, \hat{\xi}_t)' \) where \( \text{Cov}(\bar{\pi}_j, \bar{\pi}_k) \) is obtained from (28). The variance matrix of \( \hat{\theta} \) is approximated by \( \frac{1}{N} \hat{\Gamma}^{-1}(\hat{\Gamma} + \hat{V}_1) \hat{\Gamma}^{-1} \). Then, the estimator of the asymptotic variance of \( \hat{\theta} \) is obtained as a sub-matrix of \( \frac{1}{N} \hat{\Gamma}^{-1}(\hat{\Gamma} + \hat{V}_1) \hat{\Gamma}^{-1} \).

Finally we discuss implications of our result for the semiparametric inference. When we derive the asymptotic distributions of the control function approach estimators above, we have used three parametric approximations. One is for the approximation of \( \varphi_j^{-1}(s_t, p_t, x_t, z_l T2) \) to estimate \( \zeta_jt \) for all \( j \in J \) (if the supply side model is used), the second is to estimate \( E[p_{jt} | x_t, z_l T2, \zeta_t] \), and the third is for the approximation of \( \zeta_jt = \phi_j(\xi_t) \) for all \( j \in J \). Although we use parametric specifications for all \( \varphi_j^{-1}(s_t, p_t, x_t, z_l T2) \), \( E[p_{jt} | x_t, z_l T2, \zeta_t] \), and \( \phi_j(\xi_t) \) with fixed number of approximating basis functions or regressors, one can view them as non-parametric approximations (under the promise that we would use more flexible specifications for the three objects as the sample size grows). Therefore, one can view our estimator as a semiparametric multi-stage sieve estimator. Even though the asymptotic distribution of the estimator \( \hat{\theta} \) would be different under two asymptotic scenarios (parametric model with fixed number of approximating terms, and corresponding semiparametric model with growing number of approximation terms), the computed standard errors can be made numerically identical or equivalent. This means we can ignore the semiparametric nature of the model, and proceed as if the parametric model is true when we calculate the standard errors. This equivalence result has been studied in Ackerberg, Chen, and Hahn (2012).
References


