The Interaction of Observed and Unobserved Factors in Non-Separable Demand Models

Amit Gandhi
University of Wisconsin-Madison

Kyoo il Kim
University of Minnesota, Twin Cities

Amil Petrin
University of Minnesota, Twin Cities and NBER

March 17, 2010

Abstract

Many papers estimate demand in discrete choice settings using the inversion from Berry (1994) to control for unobserved product quality and its correlation with price. The inversion requires utility to be additively separable in the unobserved factor, a restriction hard to motivate theoretically. In this paper we show how to model the interactions between observed and unobserved product attributes. We use a weaker invertibility result from Gandhi (2009), which allows for some forms of non-separability of preferences. Given invertibility, we show that the standard conditional moment restrictions do not generally suffice for identification. We develop a non-linear least squares estimator for our demand setting that generalizes the literature on non-separable estimation. We use price controls originally proposed in Petrin and Train (2010) and Kim and Petrin (2010), and show they suffice for identification in our setting even when the controls are not one-to-one functions with unobserved factors. We run several monte carlos that show the approach works when the standard IV approaches fail because of non-separability. We also test and reject additive separability in the original Berry, Levinsohn, and Pakes (1995) automobile data, and we show that demand becomes significantly more elastic when the correction is applied.

1 Introduction

Demand estimation is a critical issue in many policy problems and correlation between unobserved demand factors and prices from the equilibrating market mechanism can confound estimation. If unobserved factors enter aggregate demand additively, Instrumental Variables (IV) approach allows economists to use instruments to control for the endogeneity problem and consistently estimate demand. Berry (1994) shows that when individual preferences are separable in the unobserved factor a transformation exists that converts the demand model into the separable form addressable by the IV approach.

However, there is no compelling economic motivation to treat the demand unobservables as separable in preferences. For example separability does not allow unobserved advertising to affect marginal utility derived for observed characteristics or the marginal utility of income, even though this is often a purpose of advertising. It also rules out unobserved physical characteristics impacting the marginal utility of observed factors.

We show how to model the interactions between observed and unobserved product attributes. We appeal to the results from Gandhi (2009), who extends the Berry inversion result to some settings with non-separable preferences. Once we add the interaction terms we show that conditional mean restrictions (CMR) that are usually used for identification are no longer sufficient by themselves to recover demand parameters.

We supplement the standard CMRs from Berry, Levinsohn, and Pakes (1995) with a control function setup to achieve identification. Our results extend Kim and Petrin (2009) and Petrin and Train (2010), as we use the same controls that they propose for identification, but we do not require that the controls are one-to-one with the unobserved factors, as do other non-separable control function approaches. We develop a semiparametric sieve estimator that can estimate the non-separable demand model parameters.

We run several Monte Carlos and we reestimate demand using the same automobile data as was used in BLP (1995). The Monte Carlos suggest interaction terms can significantly bias demand elasticities, and estimated automobile demands become substantially more elastic once we allow for interactions between the demand error and price and characteristics in utility.

2 Utility Specification

We use the standard discrete choice utility formulation. Consumer i chooses the product j from the J choices and the outside good to maximize her utility. Utility consumer i derives from product j is given as

$$u_{ij} \quad = \quad \alpha(y_i - p_{mj}) + \tilde{\delta}_{mj} + \epsilon_{ij},$$

where y_i is income for household i and α is the marginal utility of income, p_{mj} denotes the price of good j in market m, $\tilde{\delta}_{mj}$ is mean utility for product j in market m, and ϵ_{ij} is the independent and identically distributed (i.i.d.) extreme value taste draw for consumer i and product j.[1]

Mean utility decomposes as

$$\tilde{\delta}_{mj} = \beta' x_{mj} + \xi_{mj},$$

with x_{mj} denoting observed product characteristics and the marginal utility from each characteristic x_{mjk} is given as β_k, an element of β. ξ_{mj} is the characteristic observed to consumer and producers but unobserved to the econometrician, and it might include other physical attributes or advertising

[1] The differenced errors - what is relevant for choice - are then distributed i.i.d. logit.

that are not observed by the econometrician. We let

$$\delta_{mj} = \tilde{\delta}_{mj} - \alpha p_{mj} = \beta' x_{mj} - \alpha p_{mj} + \xi_{mj}.$$

The market share functions are given by

$$s_{mj} = \frac{e^{\delta_{mj}}}{1 + \sum_{j'} e^{\delta_{mj'}}}$$

where the sum is taken over the J_m products in market m. The Berry (1994) inversion is then

$$\delta_{mj} = \ln(s_{mj}) - \ln(s_{m0}).$$

where the unique vector $\delta_m(\theta)$ matches observed to predicted shares in a market conditional on θ, which includes all parameters but the product-market controls. Alternatively, this is the $\delta_m(\theta)$ that solves the system of equations

$$s(\theta, \delta(\theta)) = s^{Data}.$$

We denote the difference in the sample shares between good j and outside good as $\hat{\delta}_{mj} = \ln(s^n_{mj}) - \ln(s^n_{m0})$, where the n denotes the sample size. With endogenous prices one can consistently estimate parameters by regressing the δ's on characteristics and instrumented prices as the equation is now linear in price (BLP follow this approach).

We extend this setup to allow for a non-separable demand error and potentially endogenous prices. For the latter we posit utility as $u_{ij} = \delta_{mj} + \epsilon_{ij}$, with

$$\delta_{mj} = c + \beta' x_{mj} - \alpha p_{mj} + \xi_{mj} + \sum_{k}^{K} \gamma_k x_{mjk} \xi_{mj} + \gamma_p p_{mj} \xi_{mj} \tag{1}$$

and (γ, γ_p) is the vector of parameters associated with the interaction terms. For example, this specification allows advertising to potentially affect demand for observed characteristics or the marginal utility of income. Similarly, the specification allows for any unobserved physical characteristic to have consumption levels that impact the marginal utility derived from either income or an included characteristic. We will carry out most of our discussion in the model with logit heterogeneity to maintain transparency, but we can allow for random coefficients on characteristics and price, and we will do so in the empirical section.

3 Identification with Conditional Moment Restrictions and Endogenous Variables

We consider identification for equation (1) using the standard BLP conditional moment restrictions. We collect the model parameters into $\theta = (c, \beta', \alpha, \gamma', \gamma_p)'$ and denote its true value by θ_0. A set of instruments Z is presumed to exist such that $E[\xi_j(\theta_0)|Z] = 0$. We suppress the market index m for transparency and we let $\xi_j = \xi_j(\theta_0)$. To achieve identification of parameters in this non-separable setting this conditional moment restriction (CMR) suggests the possibility of using

$$E[\xi_j|Z] = E[\delta_j - (c_0 + \beta_0' x_j - \alpha_0 p_j + \xi_j(\gamma_0' x_j + \gamma_{p0}\, p_j))|Z] = 0. \tag{2}$$

These moments are the analog to the separable case, where BLP use for identification

$$E[\xi_j|Z] = E[\delta_j - (c_0 + \beta_0' x_j - \alpha_0 p_j)|Z] = 0.$$

For the CMR moments alone to identify parameters in the non-separable setting, in constructing the moment condition the practitioner must have a way to control for the interaction term between the error and price conditional on Z, given as $E[p_j\xi_j|Z]$. Specifically, one can use $E[p_j|Z]$ for the p_j term, and assuming x_j is included in Z (the standard discrete choice assumption), x_j is known and $E[x_j\xi_j|Z] = x_j E[\xi_j|Z] = 0$. However, p_j is not generally known given Z, so $E[p_j\xi_j|Z] \neq p_j E[\xi_j|Z]$. In the extreme case one must fully specify how p_j is determined, as it is a function of ξ_j, and this can be difficult to do as p_j can be complicated because it equilibrates demand and supply. [2]

The standard BLP two stage least squares (2SLS) estimator is inconsistent in the non-separable case because $p_j\xi_j$ will generally be correlated with the instrumented value of p_j, that is, $E[p_j|Z]$ is correlated with $p_j\xi_j$. Ruling out non-separability between prices and unobserved product characteristics in demand models, as in Berry (1994) and BLP (1995), is thus one way to achieve identification, although it is hard to motivate theoretically and it turns out not to be necessary, as we show in Section 4.

The CMR approach would in principle use the moment condition obtained by solving for ξ_j as a function of the other arguments

$$E[\xi_j|Z] = E\left[\frac{\delta_j - c_0 - \beta_0' x_j + \alpha_0 p_j}{1 + \gamma_0' x_{jk} + \gamma_{p0} p_j}|Z\right] = 0, \tag{3}$$

but this does not identify the model parameters in general unless we restrict the parameter space (γ, γ_p) to ensure identification. If the parameter space for (γ, γ_p) is unrestricted, then any (γ, γ_p) making the denominator arbitrarily large will satisfy the moment condition exactly (ξ degenerate at zero).

As with past control function approaches our solution will be based in part on conditioning directly upon price so we do not have to resolve the exact relationship between p_j and ξ_j.

[2] This problem is at the heart of the impossibility of identification for the nonseparable models using the classical CMR approach as discussed in Blundell and Powell (2003) and Hahn and Ridder (2008).

4 A New Control Function

In this section we develop a control function approach that can identify the demand parameters for the non-separable demand models from equation (1) (or with random coefficients added to that specification). For any set of controls $\mathbf{V}$ we let the conditional expectation of the error given Z and $\mathbf{V}$ be given as $f_j(Z, \mathbf{V}) = E[\xi_j | Z, \mathbf{V}]$.[3] We look for controls $\mathbf{V}$ such that $p_j \notin \mathbf{V}$ with probability one, and $\mathbf{V}$ satisfying

Condition 1. *(CF) Any bounded function of* (Z, p_j) *is uncorrelated with* ξ_j *given* $f_j(Z, \mathbf{V})$ *for* $j = 1, \ldots, J$*.*

The control function $f_j(Z, \mathbf{V})$ is sufficient to condition out the dependence between p_j and ξ_j so that all remaining variation of p_j is causal. [4]

Theorem 1. *If there exists control(s)* $\mathbf{V}$ *such that* $p_j \notin \mathbf{V}$ *and* p_j *is known conditional on* $(Z, \mathbf{V})$ *for* $j = 1, \ldots, J$*, then the condition CF is satisfied.*

Proof. For any bounded function of (Z, p_j), say $h(Z, p_j)$, we have $E[h(Z, p_j)(\xi_j - f_j(Z, \mathbf{V}))] = 0$ due to the law of iterated expectation, because $E[h(Z, p_j)(\xi_j - f_j(Z, \mathbf{V}))|Z, \mathbf{V}] = h(Z, p_j)E[\xi_j - f_j(Z, \mathbf{V})|Z, \mathbf{V}] = 0$ by the definition of $f_j(Z, \mathbf{V})$ and the fact that p_j is known given $(Z, \mathbf{V})$. □

We can then exploit the moment condition:

$$0 = E\left[\delta_j - \{c_0 + \beta_0 x_j - \alpha_0 p_j + f_j(Z, \mathbf{V})(1 + \gamma_0 x_j + \gamma_{p0}\, p_j)\} | Z, \mathbf{V}\right], \tag{4}$$

where we let x_j be scalar for simplicity, and $f_j(Z, \mathbf{V})$ controls for the problematic part of the error term (see, e.g. Heckman (1978)). Letting $\widetilde{\xi_j} = (1 + \gamma x_j + \gamma_p p_j)\, \xi_j$ we now obtain

$$\begin{aligned} E[\widetilde{\xi_j}|Z, \mathbf{V}] &= E[\xi_j|Z, \mathbf{V}] + \gamma E[x_j \xi_j|Z, \mathbf{V}] + \gamma_p E[p_j \xi_j|Z, \mathbf{V}] \\ &= E[\xi_j|Z, \mathbf{V}](1 + \gamma x_j + \gamma_p p_j) \\ &= f_j(Z, \mathbf{V})(1 + \gamma x_j + \gamma_p p_j), \end{aligned}$$

assuming $x_j \in Z$, and because p_j is also known conditional on Z and $\mathbf{V}$, so

$$E[p_j \xi_j|Z, \mathbf{V}] = p_j E[\xi_j|Z, \mathbf{V}] = p_j f_j(Z, \mathbf{V}).$$

The choice of the control function coupled with (4) thus allows us to circumvent the problem of specifying the exact relationship between p_j and ξ_j. The structural parameters are identified if no linear functional relationship exists between 1, x_j, p_j, $f_j(Z, \mathbf{V})$, $f_j(Z, \mathbf{V})x_j$, and $f_j(Z, \mathbf{V})p_j$.

---

[3] $f_j(Z, \mathbf{V})$ is well-defined and (almost surely) unique as long as the unconditional expectation $E[\xi_j]$ exists.

[4] p_j does not need to be independent of ξ_j given the controls $\mathbf{V}$, which is a key departure from other control function approaches.

In general (4) will not be sufficient for identification because one will not be able to separate the coefficients (c_0, β_0, α_0) from the function $f_j(Z, \mathbf{V})$ ($f_j(Z, \mathbf{V})$ may contain linear functions of x_j or be collinear with p_j). We introduce

Condition 2 (CMR)**.** $E[\xi_j|Z] = 0$.

The CMR condition imposes

$$0 = E[\xi_j|Z] = E[E[\xi_j|Z, \mathbf{V}]|Z] = E[f_j(Z, \mathbf{V})|Z]. \tag{5}$$

In many cases the condition CF combined with the implied shape restrictions from CMR on $f_j(Z, \mathbf{V})$ will suffice for the identification of the structural parameters θ_0.

We start with the usual controls originally proposed in Theil (1953), in our setting the mean projection residuals of p_j's on Z_m,

$$p_j \quad = \quad E[p_j|Z] + V_j, j = 1, \ldots, J, \tag{6}$$

with $V_j \in \mathbf{V}$. They satisfy the necessary control function condition CF by Theorem 1. Together CF and CMR can be written as a set of moment conditions and restrictions as (6) and

$$0 = E\left[\delta_j - \{c_0 + \beta_0 x_j - \alpha_0 p_j + f_j(Z, \mathbf{V})(1 + \gamma_0 x_j + \gamma_{p0}\, p_j)\}|Z, \mathbf{V}\right], j = 1, \ldots, J \tag{7}$$

where $f_j(Z, \mathbf{V})$ is now restricted to satisfy

$$E[f_j(Z, \mathbf{V})|Z] = 0. \tag{8}$$

We consider a simple example to illustrate, where we approximate $f_j(Z, \mathbf{V})$ with

$$f_j(Z, \mathbf{V}) = \pi_0 + \pi_1' Z + \pi_2 V_j + \pi_3' Z\, V_j.$$

Then (5) (CMR) implies

$$\begin{aligned} f_j(Z, \mathbf{V}) &= f_j(Z, \mathbf{V}) - E[f_j(Z, \mathbf{V})|Z] \\ &= (\pi_0 + \pi_1' Z + \pi_2 V_j + \pi_3' Z\, V_j) - (\pi_0 + \pi_1' Z + \pi_2 E[V_j|Z] + \pi_3' Z E[V_j|Z]) \\ &= \pi_2 V_j + \pi_3' Z\, V_j, \end{aligned} \tag{9}$$

because $V_j = p_j - E[p_j|Z]$ so by construction $E[V_j|Z] = 0$. CMR imposes that $f_j(Z, \mathbf{V})$ is a function of the residual V_j and its interaction with Z, but conditional on these terms is not a function of p_j nor functions of only Z (so functions of x_j are also ruled out). In this example from (7) the structural parameter θ_0 (and also the function $f_j(Z, \mathbf{V})$) is identified unless $1, x_j, p_j, V_j, Z\, V_j, V_j x_j, Z\, V_j x_j, V_j p_j$ and $Z\, V_j p_j$ are "multicollinear". We formalize this identification result in Section 5.

4.1 Discussion\Estimation Outline

Our results extend Kim and Petrin (2009) and Petrin and Train (2010) in two important ways. First, we do not require that ξ_j is a function of $V_1, \ldots, V_J$ only. That is, we do not require the controls $\mathbf{V} = (V_1, \ldots, V_J)$ to be one-to-one with ξ conditional on Z, because conditional mean decomposition $p_j = E[p_j|Z] + V_j$ does not impose the additive separability of the price equations as $p_j = g_{j1}(Z) + g_{j2}(\xi_1, \ldots, \xi_J)$.[5] Second, we allow the controls $V_1, \ldots, V_J$ to depend on Z_m.[6]

Based on these moment conditions we estimate the utility parameters together with the non-parametric function $f_j(Z, \mathbf{V})$ using a sieve multi-step least squares method, approximating $f_j(Z, \mathbf{V})$ with sieves. In the first-step we obtain consistent estimates of $\mathbf{V}$ from $V_j = p_j - E[p_j|Z]$. In the second step we construct the approximation of $f_j(Z, \mathbf{V})$ such that it satisfies the condition (8). In the final step we estimate θ_0 and $f_j(Z, \mathbf{V})$ simultaneously. Kim and Petrin (2010) provide the consistency and the asymptotic normality of this sieve multi-step estimator.

When we implement the estimation using a sieve method, we construct basis functions that approximate $f_j(Z, \mathbf{V})$ such that the restriction (8) (i.e., (5)) is embedded. To illustrate this point, let both Z and $\mathbf{V}$ be scalars. Then for example, we can approximate $f_j(Z, \mathbf{V})$ using polynomials of $(Z, \mathbf{V})$ as

$$f_j(Z, \mathbf{V}) = \sum_{l=1}^{\infty} \pi_l^j (\mathbf{V}^l - E[\mathbf{V}^l|Z]) + \sum_{l=2}^{\infty} \sum_{l_1 \geq 1, l_2 \geq 1 \text{ s.t. } l_1 + l_2 = l} \pi_{l_1,l_2}^j Z^{l_1} (\mathbf{V}^{l_2} - E[\mathbf{V}^{l_2}|Z])). \tag{10}$$

Then by construction, we have $E[f_j(Z, \mathbf{V})|Z] = 0$ and $f_j(Z, \mathbf{V})$ does not contain basis functions that are polynomials of Z only. Alternatively one can start with the unconstrained approximation, $\tilde{f}_j(Z, \mathbf{V}) = \sum_{l=1}^{\infty} \pi_l^j \mathbf{V}^l + \sum_{l=2}^{\infty} \sum_{l_1 \geq 1, l_2 \geq 1 \text{ s.t. } l_1 + l_2 = l} \pi_{l_1,l_2}^j Z^{l_1} \mathbf{V}^{l_2} + \sum_{l=1}^{\infty} \pi_{z,l}^j Z^l$ and then obtain

$$f_j(Z, \mathbf{V}) = \tilde{f}_j(Z, \mathbf{V}) - E[\tilde{f}_j(Z, \mathbf{V})|Z].$$

We develop the formal identification results and conditions under which the sieve estimator is consistent next.

5 Identification

We first present an identification theorem using the moment condition (7) and the controls $\mathbf{V}$ that satisfy the CF condition. Then we show how the CMR condition helps to achieve the identification. Because $\mathbf{V} = (V_1, \ldots, V_J)$ is identified from the first step regression of (6), we treat $(E[p_1|Z], \ldots, E[p_J|Z])$ and $\mathbf{V}$ as known. Then, from the moment condition (7), the identification of

[5] Kim and Petrin (2009)develop a control function estimator for the case when ξ_j is obtained from the reduced form equation $\xi_j = r_j(Z, \mathbf{V}) = f_j(Z, \mathbf{V})$ (i.e., ξ is invertible from the pricing functions). In that setting the conditional expectation $E[\xi_j|Z, \mathbf{V}]$ is deterministic in the sense that $\xi_j = f_j(Z, \mathbf{V})$.

[6] If we compare our control function approach with other control function approaches in the literature (e.g., Newey, Powell, and Vella (1999), Imbens and Newey (2003), and Altonji and Matzkin (2005)), we relax their strong independence conditions between the instruments and the controls with weaker conditional mean restrictions, $E[\mathbf{V}_m|Z_m] = 0$. See Kim and Petrin (2010) for this paper's contribution in terms of the control function approach in general.

the key structural parameter θ_0 is equivalent to the nonexistence of any additive functional relationship between 1, x_j, p_j, $f_j(Z,\mathbf{V})$, $x_j f_j(Z,\mathbf{V})$, and $p_j f_j(Z,\mathbf{V})$. Here we will maintain the assumption $Var[f_j(Z,\mathbf{V})|Z] > 0$ - price p_j is endogenous - but we can allow for settings where the practitioner does not know whether the variable is exogenous or endogenous (Kim and Petrin (2010)). We formalize this identification result in Theorem 2. Then we also derive a set of sufficient conditions for the identification assuming the differentiability of $E[p_j|Z]$, $\mathbf{V}$, and $f_j(Z,\mathbf{V})$ with respect to Z and $\mathbf{V}$. Define $\Pi_j(Z) = E[p_j|Z]$. Let $Z = (X', Z_2')'$ and let subscripts with 1, 2, and v_j denote partial differentiation with respect to X, Z_2, and V_j.

Theorem 2 (Identification)**.** *Suppose the CF condition holds. If*

$$\Pr\left\{\Psi_j(x_j, p_j, f_j(Z,\mathbf{V}); \psi) = 0\right\} < 1 \tag{11}$$

where $\Psi_j(x_j, p_j, f_j(Z,\mathbf{V}); \psi) = \psi_0 + \psi_1' x_j + \psi_2 p_j + \psi_3 f_j(Z,\mathbf{V}) + \psi_4' x_j f_j(Z,\mathbf{V}) + \psi_5 p_j f_j(Z,\mathbf{V})$, *then the structural parameter,* $\theta_0 = (c_0, \beta_0', \alpha_0, \gamma_0', \gamma_{p0})'$ *is identified.*

Further suppose $\Pi_j(Z)$, $\mathbf{V}$, *and* $f_j(Z,\mathbf{V})$ *are differentiable and suppose* $\Pi_{j,2}(Z)$ *has the full rank (i.e.,* $\Pi_{j,2}(Z) \neq 0$*). Then,* θ_0 *is identified (i) if* $f_{j,v_k}(Z,\mathbf{V})$ *is not equal to zero for at least one* $k \neq j$ *with probability one or (ii) if*

$$\Pr\left\{(\Pi_{j,2}(Z) + V_{j,2}) f_{j,v_j}(Z,\mathbf{V}) - f_{j,2}(Z,\mathbf{V}) = 0\right\} < 1. \tag{12}$$

Proof. Suppose there exists an additive functional relationship between them. Then we must find a vector of constants $\psi \neq 0$ such that

$$\Pr\left\{\Psi_j(x_j, p_j, f_j(Z,\mathbf{V}); \psi) = 0\right\} = 1. \tag{13}$$

Given differentiability, we have

$$\begin{aligned}
\Psi_{j,2}(\cdot;\psi) &= (\Pi_{j,2}(Z) + V_{j,2})(\psi_2 + \psi_5 f_j(Z,\mathbf{V})) + (\psi_3 + \psi_4' x_j + \psi_5 p_j) f_{j,2}(Z,\mathbf{V}) \\
\Psi_{j,1}(\cdot;\psi) &= \psi_1 + \psi_2(\Pi_{j,1}(Z) + V_{j,1}) + (\psi_3 + \psi_4' x_j + \psi_5 p_j) f_{j,1}(Z,\mathbf{V}) + \psi_4 f_j(Z,\mathbf{V}) \\
\Psi_{j,v_j}(\cdot;\psi) &= \psi_2 + (\psi_3 + \psi_4' x_j + \psi_5 p_j) f_{j,v_j}(Z,\mathbf{V}) + \psi_5 f_j(Z,\mathbf{V}) \\
\Psi_{j,v_k}(\cdot;\psi) &= (\psi_3 + \psi_4' x_j + \psi_5 p_j) f_{j,v_k}(Z,\mathbf{V}) = 0 \text{ for } k \neq j.
\end{aligned}$$

If there is an additive functional relationship as $\Psi_j(\cdot;\psi) = 0$ with probability one, then we must also have

$$\Psi_{j,2}(\cdot;\psi) = 0, \Psi_{j,1}(\cdot;\psi) = 0, \text{ and } \Psi_{j,v_k}(\cdot;\psi) \text{ for } k = 1, \ldots, J$$

with probability one.

If $f_{j,v_k}(Z,\mathbf{V})$ is not equal to zero for at least one $k \neq j$, $\Psi_{j,v_k}(\cdot;\psi) = 0$ implies $\psi_3 + \psi_4' x_j + \psi_5 p_j = 0$. This is not possible unless there is an additive functional relationship between 1, x_j, and p_j, which is ruled out by the condition that $\Pi_{j,2}(Z)$ has the full rank (i.e., $\Pi_{j,2}(Z) \neq 0$). Therefore, as long as $f_{j,v_k}(Z,\mathbf{V})$ is not equal to zero for at least one $k \neq j$, identification holds.

Next we consider the case that $f_{j,v_k}(Z, \mathbf{V})$ is equal to zero for all k (this includes the binary choice model). Combining $\Psi_{j,2}(\cdot;\psi) = 0$ and $\Psi_{j,v_j}(\cdot;\psi) = 0$, we obtain

$$\{(\Pi_{j,2}(Z) + V_{j,2})f_{j,v_j}(Z, \mathbf{V}) - f_{j,2}(Z, \mathbf{V})\}(\psi_3 + \psi_4' x_j + \psi_5 p_j) = 0 \tag{14}$$

with probability one. This implies $(\Pi_{j,2}(Z) + V_{j,2})f_{j,v_j}(Z, \mathbf{V}) - f_{j,2}(Z, \mathbf{V}) = 0$ with probability one because $\psi_3 + \psi_4' x_j + \psi_5 p_j \neq 0$ with probability one. Therefore, unless $(\Pi_{j,2}(Z) + V_{j,2})f_{j,v_j}(Z, \mathbf{V}) - f_{j,2}(Z, \mathbf{V}) = 0$ with probability one, the identification holds. This completes the proof. □

Next we note that the identification condition (12) in Theorem 2 is much simplified under the independence condition.

Corollary 1 (Independence)**.** *Suppose $E[\xi_j|Z, \mathbf{V}] = f_j(\mathbf{V})$ and $\mathbf{V}$ is jointly independent of Z and suppose $\Pi_j(Z)$, $\mathbf{V}$, and $f_j(\mathbf{V})$ are differentiable. If $\Pi_{j,2}(Z)$ has the full rank (i.e., $\Pi_{j,2}(Z) \neq 0$) and $f_{j,v_j}(\mathbf{V}) \neq 0$, (12) is satisfied.*

Proof. Under the independence condition, we have $V_{j,2} = 0$ and $f_{j,2}(\mathbf{V}) = 0$ (i.e., $f_{j,2}(Z, \mathbf{V}) = 0$) then from (14) it follows that $\Pi_{j,2}(Z)f_{j,v_j}(\mathbf{V}) = 0$ with probability one but this is not possible because $\Pi_{j,2}(Z)$ has the full rank (i.e., at least one element of $\Pi_{j,2}(Z)$ is not equal to zero) and $f_{j,v_j}(\mathbf{V}) \neq 0$. □

The conditional moment restrictions are weaker than this Independence condition, and we show that they can be used for identification. Theorem 3 formalizes this idea.

Theorem 3 (Identification with CMR)**.** *Suppose conditions CF and CMR hold. Then, θ_0 is identified.*

Proof. We resort to Theorem 2. Define $\Psi_j(x_j, p_j, f_j(Z, \mathbf{V}); \psi) = \psi_0 + \psi_1' x_j + \psi_2 p_j + \psi_3 f_j(Z, \mathbf{V}) + \psi_4' x_j f_j(Z, \mathbf{V}) + \psi_5 p_j f_j(Z, \mathbf{V})$. First we show that there does not exist an additive functional relationship between 1, x_j, p_j, and $f_j(Z, \mathbf{V})$ or between 1, x_j, p_j, and $x_j f_j(Z, \mathbf{V})$. Suppose there is an additive functional relationship between 1,x_j, p_j, and $f_j(Z, \mathbf{V})$, it must be that $\psi_0 + \psi_1' x_j + \psi_2 p_j + \psi_3 f_j(Z, \mathbf{V}) = 0$ with probability one. By applying the conditional expectation on Z to this relationship, we find

$$\begin{aligned} 0 &= E[\psi_0 + \psi_1' x_j + \psi_2 p_j + \psi_3 f_j(Z, \mathbf{V})|Z] \\ &= \psi_0 + \psi_1' x_j + \psi_2 E[p_j|Z] \end{aligned}$$

but this cannot be true if (e.g.) Z includes at least one variable other than x_j (an instrument) so $E[p_j|Z]$ is not collinear with x_j. Similarly we can show 1, x_j, p_j, and $x_j f_j(Z, \mathbf{V})$ do not have an additive functional relationship. It follows that 1, x_j, p_j, $f_j(Z, \mathbf{V})$, and $x_j f_j(Z, \mathbf{V})$ do not have an additive functional relationship either because $f_j(Z, \mathbf{V})$ and $x_j f_j(Z, \mathbf{V})$ are not collinear. It also implies that if (13) holds, it must be that $\psi_5 \neq 0$. Similarly we can conclude 1, x_j, p_j, $f_j(Z, \mathbf{V})$, $x_j f_j(Z, \mathbf{V})$, and $p_j f_j(Z, \mathbf{V})$ do not have an additive functional relationship because $p_j f_j(Z, \mathbf{V})$

cannot be written as a linear combination of the others. Therefore, the identification is achieved by Theorem 2. □

6 Sieve Estimation and Consistency

The estimation proceeds in multiple steps. In the first stage we estimate $\Pi_j(z)$ and obtain $\hat{V}_j = p_j - \hat{\Pi}_j(z)$ for $j = 1, \ldots, J$. In the second step, we construct approximating basis functions using $\hat{V}_j$ and Z, subtracting out the conditional mean of $\hat{V}_j$ given Z. In the final step we estimate θ_0 and $f_j(\cdot)$ using a sieve method.

We write these approximating basis functions for $f_j(\cdot)$ as

$$\tilde{\varphi}_l(\mathbf{V}, Z) = \varphi_l(\mathbf{V}, Z) - \bar{\varphi}_l(Z)$$

where we write $\bar{\varphi}_l(Z) = E[\varphi_l(\mathbf{V}, Z)|Z]$ and their estimates as $\hat{\tilde{\varphi}}_l(\hat{\mathbf{V}}, Z) = \varphi_l(\hat{\mathbf{V}}, Z) - \hat{\bar{\varphi}}_l(Z)$. We will suppress the subscript j in $f_j(\cdot)$ for simplicity. Then we define the sieve space $\mathcal{F}_M$ as the collection of functions

$$\mathcal{F}_M = \{f : f = \sum_{l \le L(M)} a_l \tilde{\varphi}_l(v, z), \|f\|_{\mathcal{F}} < \bar{C}\}$$

for some bounded positive constant $\bar{C}$ with $L(M) \to \infty$ and $L(M)/M \to 0$ such that $\mathcal{F}_M \subseteq \mathcal{F}_{M+1} \subseteq \ldots \subseteq \mathcal{F}$ and $\mathcal{F}$ denotes a space of functions that includes f_0. We let $\|\cdot\|_{\mathcal{F}}$ be a pseudo-metric on $\mathcal{F}$. We define the sieve space constructed using the estimated controls and $\hat{\bar{\varphi}}_l(Z)$ as

$$\hat{\mathcal{F}}_M = \{f : f = \sum_{l \le L(M)} a_l \hat{\tilde{\varphi}}_l(v, z), \|f\|_{\mathcal{F}} < \bar{C}\}. \tag{15}$$

Under weak regularity conditions $\hat{\mathcal{F}}_M \to \mathcal{F}_M$ as $\hat{\Pi}(z) \to \Pi_0(z)$ and $\hat{\bar{\varphi}}_l(z) \to \bar{\varphi}_l(z)$.

Denote a criterion function based on the moment conditions of (7) as

$$Q(\delta_m, Z_m, p_m, \mathbf{V}_m; \theta, f)$$

and its sample analogue as $Q_M(\delta_m, Z_m, p_m, \hat{V}_m; \theta, f)$. If one uses a nonlinear sieve least squares estimator, then the sample criterion function becomes (e.g.)

$$Q_M(\delta_m, Z_m, p_m, \hat{V}_m; \theta, f) = \frac{1}{\sum_{m=1}^M J_m} \sum_{m=1}^{M} \sum_{j=1}^{J_m} \{\delta_{mj} - (c + \beta' x_{mj} - \alpha p_{mj} + f_j(\cdot)(1 + \gamma' x_{mj} + \gamma_p p_{mj}))\}^2$$

subject to $(\theta, f) \in \Theta \times \hat{\mathcal{F}}_M$.

We define our control function approach estimator by

$$(\hat{\theta}, \hat{f}) = \text{arginf}_{(\theta, f) \in \Theta \times \hat{\mathcal{F}}_M} Q_M(\delta_m, Z_m, p_m, \hat{\mathbf{V}}_m; \theta, f) \tag{16}$$

where $\hat{V}_{mj} = p_{mj} - \hat{\Pi}_j(Z_m)$.

We obtain the consistency of our estimator under the following assumptions based on the results in Newey and Powell (2003), Chen, Linton, and van Keilegom (2003), and Chen (2006). [7]

We first assume the identification that we have discussed in Section 5.

Assumption 1 (A1). *$(\theta_0, f_0) \in \Theta \times \mathcal{F}$ is the only $(\theta, f) \in \Theta \times \mathcal{F}$ satisfying the moment condition (4) and $Q(\delta_m, Z_m, p_m, \mathbf{V}_m; \theta_0, f_0) < \infty$.*

Next we assume that our estimator is an extremum estimator that solves (16).

Assumption 2 (A2). $Q_M(\delta_m, Z_m, p_m, \hat{\mathbf{V}}_m; \hat{\theta}, \hat{f}) \leq \textit{arginf}_{(\theta,f)\in\Theta\times\hat{\mathcal{F}}_M} Q_M(\delta_m, Z_m, p_m, \hat{\mathbf{V}}_m; \theta, f) + o_p(1)$

Denote the true functions of $\Pi(Z)$ and $\bar{\varphi}_l(Z)$ as $\Pi_0(Z)$ and $\bar{\varphi}_{0l}(Z)$, respectively. Next we assume that both $\Pi_0(Z)$ and $\bar{\varphi}_{0l}(Z)$ can be approximated by the first stage and the middle stage series approximations. For example, this is known to be satisfied for power series and splines approximation if $\Pi_{0j}(Z)$'s and $\bar{\varphi}_{0l}(Z)$'s are smooth and their derivatives are bounded (e.g., belong to a Hölder class of functions). We also assume $\Pi(Z)$ and $\bar{\varphi}_l(Z)$ are endowed with a pseudo-metric $\|\cdot\|_s$.

Assumption 3 (A3). $\left\|\hat{\Pi}_j(z) - \Pi_{0j}(Z)\right\|_s = o_p(1)$ *for all* $j = 1, \ldots, J$; *and* $\left\|\hat{\bar{\varphi}}_l(Z) - \bar{\varphi}_{0l}(Z)\right\|_s = o_p(1)$ *for all* l.

Assumption 4 (A4). *The sieve space $\mathcal{F}_M$ satisfies $\mathcal{F}_M \subseteq \mathcal{F}_{M+1} \subseteq \ldots \subseteq \mathcal{F}$ for all $M \geq 1$; and for any $f \in \mathcal{F}$ there exists $\pi_M f \in \mathcal{F}_M$ such that $\|f - \pi_M f\|_{\mathcal{F}} \to 0$ as $M \to \infty$.*

Although it is obvious in our problem, we assume the following continuity conditions for completeness.

Assumption 5 (A5). *$Q(\delta_m, Z_m, p_m, \mathbf{V}_m; \theta, f)$ is continuous in $(\theta, f) \in \Theta \times \mathcal{F}$.*

Assumption 6 (A6). *$Q(\delta_m, Z_m, p_m, \mathbf{V}_m; \theta, f_M)$ is continuous in $\Pi(Z)$ and $\bar{\varphi}_l(Z)$ uniformly for all $(\theta, f_M) \in \Theta \times \mathcal{F}_M$ where $V_{mj} = p_{mj} - \Pi(Z_m)$.*

Next we impose compactness on the sieve space. This compactness is also satisfied when the sieve space is based on power series or splines.

Assumption 7 (A7). *The sieve spaces, $\mathcal{F}_M$, are compact under the pseudo-metric $||\cdot||_{\mathcal{F}}$.*

The last condition we assume is that in the neighborhood of $\Pi_0(Z)$ and $\bar{\varphi}_{0l}(Z)$, the sample criterion function $Q_M(\delta_m, Z_m, p_m, \mathbf{V}_m; \theta, f)$ uniformly converges to the population criterion function.

[7] Our problem is different from Newey and Powell (2003)'s Theorem 4.1 because we use estimated regressors (functions, $\hat{\Pi}$ and $\hat{\bar{\varphi}}_l(Z)$) in the main estimation. Our problem is also different from Chen, Linton, and van Keilegom (2003) because we estimate the parametric component (θ_0) and the nonparametric component (f_0) simultaneously in the main estimation.

Assumption 8 (A8)**.** *For all positive sequences* $\epsilon_M = o(1)$*, we have*

$$\sup_{(\theta,f)\in\Theta\times\mathcal{F}_M,\|\Pi-\Pi_0\|_s\leq\epsilon_M,\|\bar{\varphi}_l-\bar{\varphi}_{0l}\|_s\leq\epsilon_M\forall l} |Q_M(\delta_m, Z_m, p_m, \mathbf{V}_m; \theta, f) - Q(\delta_m, Z_m, p_m, \mathbf{V}_m; \theta, f)| = o_p(1)$$

where $V_{mj} = p_{mj} - \Pi_j(Z_m)$.

Theorem 4. *Suppose Assumptions A1-A8 are satisfied. Then* $\hat{\theta} \to_p \theta_0$.

See Kim and Petrin (2010) for the proof. They also develop the asymptotic distribution of $\hat{\theta}$ in the context of the sieve estimation where both $\Pi(Z)$ and $\bar{\varphi}_l(Z)$ are nonparametrically estimated. The formulas for standard errors of $\hat{\theta}$ developed there can be also used for the parametric estimation case when the truncated semiparametric model (i.e., with fixed length of sieves) is the true parametric model.

7 Monte Carlo Evidence

We demonstrate the performance of our estimator in terms of a simultaneous two equations model. Our simulations include cases where our control function approach works reasonably well while 2SLS is inconsistent and has sizable biases. We consider a few different designs for how the endogenous variable is generated and find that our approach is robust while 2SLS is sensitive to different specifications of the reduced form equations.

We consider the following outcome equation where the endogenous regressor p is not additively separable with the error term ξ:

$$q = \alpha - \beta p + \gamma p\xi + \xi$$

and the five different designs of the reduced form equatio

$$\begin{aligned}
[1]\ p &= Z + 3\xi + \varsigma \\
[2]\ p &= Z + Z^2 + (\xi + 1)^3 + \varsigma \\
[3]\ p &= Z + (Z^2 + 5Z)\xi + \varsigma \\
[4]\ p &= Z + (5 + 5Z + \varsigma)\xi \\
[5]\ p &= Z + Z^2 + 5(\xi + 1/3)^2 + \varsigma
\end{aligned}$$

We call ξ the structural error term because it appears in the outcome equation and ς the reduced form error term. In the first design [1] the structural error is additively separable both from instruments and the reduced form error. In the designs [2] and [5], they are additively separable but the structural error term enters the reduced form equation nonlinearly. In the third design [3], the instruments and the structural error are not additively separable. In the fourth design [4] the structural error is not additively separable from the instrument nor the reduced form error.

We generate a simulation data based on the five different designs with the following distributions: $\xi \sim U_{[-1/2,1/2]}$, $\varsigma \sim U_{[-1/2,1/2]}$, $Z = 2 + 2U_{[-1/2,1/2]}$, and they are independent where $U_{[-1/2,1/2]}$ denotes the uniform distribution supported on $[-1/2, 1/2]$. In these designs, the control $V = p - E[p|Z]$ is independent of Z in [1], [2], [5], but it is not independent of Z in [3] and [4]. We set the true parameter values $(\alpha_0, \beta_0, \gamma_0) = (1, 1, 0.5)$. The data is generated with the sample sizes: $M = 1,000$ and $M = 10,000$. We take one reasonable sample size and one large sample size because we are interested both in a finite sample performance and the consistency of our proposed estimator.

We estimate the models using three methods: OLS, 2SLS, and our control function estimator. The control function approach is implemented in three steps. First we estimate $\hat{V} = p - (\hat{\pi}_0 + \hat{\pi}_1 Z + \hat{\pi}_2 Z^2)$ using OLS and construct approximating functions $\tilde{V}_1 = \hat{V}$, $\tilde{V}_2 = \hat{V}^2 - \hat{E}[\hat{V}^2|Z]$, $\tilde{V}_3 = \hat{V}^3 - \hat{E}[\hat{V}^3|Z]$, and others are defined similarly where $\hat{E}[\cdot|Z]$ is implemented by the OLS estimation on $(1, Z, Z^2)$. In the last step we estimate the model parameters using nonlinear least squares:

$$(\hat{\alpha}, \hat{\beta}, \hat{\gamma}, \hat{a}) = \text{argmin} \sum_{m=1}^{M} \{q_m - (\alpha - \beta p_m + \gamma p_m (\sum_{l=1}^{L_M} a_l \tilde{V}_{ml}) + \sum_{l=1}^{L_M} a_l \tilde{V}_{ml})\}^2 / M.$$

In the designs [1],[2], we select the controls: $\tilde{V}_1, \tilde{V}_2$, and $\tilde{V}_3$, in the design [3] we use the controls: $\tilde{V}_1, Z\tilde{V}_1, Z^2\tilde{V}_1, Z^3\tilde{V}_1, Z^4\tilde{V}_1, Z^5\tilde{V}_1, \tilde{V}_2, Z\tilde{V}_2$, and $Z^2\tilde{V}_2$, in the design [4] we use $\tilde{V}_1, \tilde{V}_2$, $Z\tilde{V}_1$, and $Z^2\tilde{V}_1$ for estimations, and finally we use $\tilde{V}_1, \tilde{V}_2$, $\tilde{V}_3$, and $\tilde{V}_4$ in the design [5]. [8]

We report the biases and the RMSE based on 100 repetitions of the estimations: OLS, 2SLS, and our control function estimator. The simulation results (Tables I-V) clearly show that OLS is biased in all designs. 2SLS is also biased. Our estimator is robust regardless of the reduced form equations.

In the designs [1]-[5], 2SLS estimates for the constant term (α) are biased (12%, 13%, -15%, 21%, 14% respectively). In the designs [3]-[4] the 2SLS estimates for the coefficient on the endogenous regressor (β) are severely biased (-37% and -21%). From other Monte Carlos (not reported here) we find higher coefficients on ξ in the reduced form equation create larger biases for the 2SLS estimates of α and higher coefficients on the interaction term $Z\xi$ in the reduced form equation generate larger biases for the 2SLS estimates of β.

Table I: Design [1], $\alpha_0 = 1, \beta_0 = 1, \gamma_0 = 0.5$**, Controls:** $\tilde{V}_1, \tilde{V}_2, \tilde{V}_3$

[8] In the design [3] it was harder to estimate the interaction parameter (γ) correctly with a smaller set of controls. One can choose an *optimal* set of controls among alternatives based on the cross validation (CV) criterion, although the validity of CV may be compromised due to the presence of the first and the second step in our estimation.

		mean	bias	RMSE	mean	bias	RMSE
			$M = 1,000$			$M = 10,000$	
OLS	α	0.2669	-0.7331	0.7335	0.2670	-0.7330	0.7330
	β	0.5708	-0.4292	0.4293	0.5709	-0.4291	0.4291
2SLS	α	**1.1255**	**0.1255**	0.1445	**1.1227**	**0.1227**	0.1243
	β	1.0001	0.0001	0.0373	0.9988	-0.0012	0.0105
CF	α	1.0012	0.0012	0.0797	0.9977	-0.0023	0.0220
	β	1.0003	0.0003	0.0379	0.9990	-0.0010	0.0103
	γ	0.5198	0.0198	0.1298	0.5064	0.0064	0.0378

Table II: Design [2], $\alpha_0 = 1, \beta_0 = 1, \gamma_0 = 0.5$**, Controls:** $\tilde{V}_1, \tilde{V}_2, \tilde{V}_3$

		mean	bias	RMSE	mean	bias	RMSE
			$M = 1,000$			$M = 10,000$	
OLS	α	0.0745	-0.9255	0.9325	0.0674	-0.9326	0.9333
	β	0.8606	-0.1394	0.1404	0.8596	-0.1404	0.1405
2SLS	α	**1.1332**	**0.1332**	0.1912	**1.1293**	**0.1293**	0.1346
	β	1.0002	0.0002	0.0201	0.9997	-0.0003	0.0056
CF	α	1.0022	0.0022	0.1274	0.9985	-0.0015	0.0350
	β	1.0002	0.0002	0.0193	0.9997	-0.0003	0.0053
	γ	0.5027	0.0027	0.0809	0.4960	-0.0040	0.0231

Table III: Design [3], Controls: $\tilde{V}_1, Z\tilde{V}_1, Z^2\tilde{V}_1, Z^3\tilde{V}_1, Z^4\tilde{V}_1, Z^5\tilde{V}_1, \tilde{V}_2, Z\tilde{V}_2, Z^2\tilde{V}_2$

		mean	bias	RMSE	mean	bias	RMSE
			$M = 1,000$			$M = 10,000$	
OLS	α	1.3327	0.3327	0.3335	1.3303	0.3303	0.3303
	β	0.8666	-0.1334	0.1336	0.8663	-0.1337	0.1337
2SLS	α	**0.8522**	**-0.1478**	0.1843	**0.8464**	**-0.1536**	0.1574
	β	**0.6251**	**-0.3749**	0.3803	**0.6244**	**-0.3756**	0.3761
CF	α	1.0095	0.0095	0.0699	0.9966	-0.0034	0.0090
	β	1.0049	0.0049	0.0302	0.9995	-0.0005	0.0044
	γ	0.5067	0.0067	0.1435	0.5008	0.0008	0.0282

Table IV: Design [4], $\alpha_0 = 1, \beta_0 = 1, \gamma_0 = 0.5$**, Controls:** $\tilde{V}_1, \tilde{V}_2, Z\tilde{V}_1, Z^2\tilde{V}_1$

		mean	bias	RMSE	mean	bias	RMSE
			$M = 1,000$			$M = 10,000$	
OLS	α	1.3599	0.3599	0.3606	1.3582	0.3582	0.3582
	β	0.8666	-0.1334	0.1335	0.8663	-0.1337	0.1337
2SLS	α	**1.2140**	**0.2140**	0.2274	**1.2070**	**0.2070**	0.2081
	β	**0.7932**	**-0.2068**	0.2108	**0.7907**	**-0.2093**	0.2095
CF	α	1.0140	0.0140	0.0459	1.0017	0.0017	0.0161
	β	1.0056	0.0056	0.0218	1.0004	0.0004	0.0083
	γ	0.4999	-0.0001	0.1510	0.5018	0.0018	0.0623

Table V: Design [5], $\alpha_0 = 1, \beta_0 = 1, \gamma_0 = 0.5$**, Controls:** $\tilde{V}_1, \tilde{V}_2, \tilde{V}_3, \tilde{V}_4$

		mean	bias	RMSE	mean	bias	RMSE
			$M = 1,000$			$M = 10,000$	
OLS	α	0.0835	-0.9165	0.9229	0.0779	-0.9221	0.9227
	β	0.8555	-0.1445	0.1455	0.8547	-0.1453	0.1454
2SLS	α	**1.1402**	**0.1402**	0.1914	**1.1371**	**0.1371**	0.1416
	β	1.0001	0.0001	0.0199	0.9997	-0.0003	0.0055
CF	α	1.0030	0.0030	0.1202	0.9992	-0.0008	0.0328
	β	1.0003	0.0003	0.0189	0.9998	-0.0002	0.0052
	γ	0.5026	0.0026	0.0890	0.4906	-0.0094	0.0249

8 The BLP Automobile Application

We revisit the original Berry, Levinsohn, and Pakes (1995) automobile application to investigate whether interaction terms are important for own- and cross-price elasticities. The application uses the same data from BLP so results are easy to replicate. There are 2217 market-level observations on prices, quantities, and characteristics of automobiles sold in the 20 U.S. automobile markets indexed m beginning in 1971 and continuing annually to 1990. We do not use a supply side model when we estimate the demand side model so our point estimates only exactly match their estimated specifications for the cases they examine without the supply side.[9] We discuss the various specifications, the instruments, and the results.

Recall we posit $u_{ij} = \delta_{mj} + \epsilon_{ij}$ as

$$\delta_{mj} = c + \beta' x_{mj} - \alpha p_{mj} + \xi_{mj} + \gamma' x_{mj} \xi_{mj} + \gamma_p p_{mj} \xi_{mj}, \tag{17}$$

and (γ_0, γ_p) is the vector of parameters associated with the interaction terms. Automobile characteristics x_{mj} 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. For

[9]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.

example, this specification allows automobile advertising to potentially affect demand for observed characteristics or the marginal utility of income. Similarly, the specification allows for any unobserved physical characteristic to have consumption levels that impact the marginal utility derived from either income or an included characteristic.

8.1 Estimation

We estimate the demand model (17) following the control function approach as described in Sections 4 and 6. We start by constructing 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 product (or nameplate) over time, because cars change characteristics and/or exit. This means some restrictions on $E[\,p_j \,|\, Z_m]$ across vehicles will be necessary in order to estimate this function, where Z_m denotes all observed and exogenous demand and cost factors in market m.

We follow the logic outlined in Pakes (1996) and described above, and use as arguments for each product j the 15 regressors given by $\tilde{Z}_{mj}$, which reflect both demand and cost factors relevant for each product. Our first basis control is defined as

$$\tilde{\xi}_{mj} = p_{mj} - E[\,p_j \,|\, \tilde{Z}_{mj}],$$

and we estimate the expectation using ordinary least squares. Two additional basis controls are given as

$$\tilde{\xi}_{(1)mj} = \sum_{k \neq j, k \in J_f} \tilde{\xi}_{mk}$$

and

$$\tilde{\xi}_{(2)mj} = \sum_{k \notin J_f} \tilde{\xi}_{mk},$$

where J_f is the set of products produced by the firm that produces the product j. These controls are respectively the sum of all of the other residuals of the products made by the same firm, given by $\tilde{\xi}_{(1)mj}$, and the sum of all the residuals of all the products made by other firms, given by $\tilde{\xi}_{(2)mj}$.

Based on these $\tilde{\xi}_{mj}$, $\tilde{\xi}_{(1)mj}$, and $\tilde{\xi}_{(2)mj}$, we generate the following nine controls that we use for our estimation:

$$\begin{aligned}
V_{1mj} &= \tilde{\xi}_{mj}, V_{2mj} = \tilde{\xi}^2_{mj} - E[\tilde{\xi}^2_{mj}|\tilde{Z}_{mj}], V_{3mj} = \tilde{\xi}^3_{mj} - E[\tilde{\xi}^3_{mj}|\tilde{Z}_{mj}] \\
V_{4mj} &= \tilde{\xi}_{(1)mj}, V_{5mj} = \tilde{\xi}^2_{(1)mj} - E[\tilde{\xi}^2_{(1)mj}|\tilde{Z}_{mj}], V_{6mj} = \tilde{\xi}^3_{(1)mj} - E[\tilde{\xi}^3_{(1)mj}|\tilde{Z}_{mj}] \\
V_{7mj} &= \tilde{\xi}_{(2)mj}, V_{8mj} = \tilde{\xi}^2_{(2)mj} - E[\tilde{\xi}^2_{(2)mj}|\tilde{Z}_{mj}], V_{9mj} = \tilde{\xi}^3_{(2)mj} - E[\tilde{\xi}^3_{(2)mj}|\tilde{Z}_{mj}].
\end{aligned}$$

Then we estimate the following equation using nonlinear least squares

$$\delta_{mj} = c + \beta' x_{mj} - \alpha p_{mj} + f_j(Z_m, \hat{\mathbf{V}}_m)(1 + \gamma' x_{mj} + \gamma_p p_{mj}) + \epsilon_{mj},$$

where we approximate $f_j(Z_m, \hat{\mathbf{V}}_m) = \sum_{l=1}^{9} \pi_l \hat{V}_{lmj}$ using the estimated controls.

8.2 Results

The results for the separable error and exogenous price case are in Column 1 of Table 1 and they replicate those results from the first column of Table III in BLP. The price coefficient increases from -0.08 to -0.13 when we move from OLS to 2SLS, suggesting prices are endogenous. Without additive separability, that is, when $(\gamma, \gamma_p) \neq 0$, the standard two-stage least squares approach that is used in BLP is no longer consistent (column 2). Column 3 includes our control function results, which nest the specifications from both columns 1 and columns 2. The price coefficient increases from -0.13 to -0.23 and is significantly different from the coefficient from 2SLS. The separable BLP specification is rejected using a simple goodness-of-fit test.

Table 2 translates these estimates into elasticities. 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 first column uses the uncorrected logit specification from Column 1 of Table III in BLP (1995). [10]Ignoring price endogeneity severely biases price elasticities towards zero. As we control the endogeneity using the 2SLS the price elasticities change significantly and become more elastic, as the median elasticity moves from -0.77 to -1.18. However, biggest change comes when we move from 2SLS to the control function approach, as the median elasticity increases from -1.18 to -2.06, and the mean elasticity increases from -1.6 to -2.6.[11]Thus ignoring non-separability in the BLP data results in much less elastic demand estimates that are significantly different from our non-separable control function approach.

9 Conclusion

Many empirical studies in discrete choice have shown that unobserved factors have significant explanatory power and are typically highly correlated with price. The unobserved factor can include omitted physical product characteristics or unobserved advertising.

The inversion from Berry (1994) controls for unobserved product quality and its correlation with price and is widely used in empirical work. However, it requires that additively separability between the observed and unobserved factors holds, which is hard to motivate economically. For example it does not allow unobserved advertising to affect marginal utility derived for observed characteristics or the marginal utility of income, even though this is often a purpose of advertising. It also rules out unobserved physical characteristics impacting the marginal utility of observed factors.

We show how to model the interactions between observed and unobserved product attributes. We develop a new control function approach for some non-separable settings. It is based on the

[10]Because the data sets are the same, these are the same elasticities that result from the coefficients of their Table III.

[11]Preliminary estimates from the specifications with random coefficients on characteristics actually yield slightly more elastic demand estimates.

Table 1
Estimated Parameters for Automobile Demand: No Correction,
2SLS (without Interactions), and Control Function approach
Dependent Variable: $\ln(s_j^n) - \ln(s_0^n)$

Parameter	Variable	No Correction*	2SLS (w/o Interactions)	Control Function (with Interactions)
Term on Price (α)	price	-0.088	-0.136	-0.233
		(0.004)	(0.011)	(0.016)
Mean Parameters	Constant	-10.071	-9.915	-9.657
		(0.252)	(0.263)	(.253)
	HP/Weight	-0.122	1.226	2.803
		(0.277)	(0.404)	(0.421)
	Air	-0.034	0.486	1.385
		(0.072)	(0.133)	(0.148)
	MP$	0.265	0.172	0.106
		(0.043)	(0.049)	(0.047)
	Size	2.342	2.292	2.367
		(0.125)	(0.129)	(0.128)
Interaction Parameters	price$\cdot\xi$			0.112
				(0.248)
	HP/Weight$\cdot\xi$			2.340
				(6.137)
	Air$\cdot\xi$			1.107
				(2.482)
	MP$$\cdot\xi$			-0.360
				(0.614)
	Size$\cdot\xi$			0.489
				(2.181)
Control Functions	V_1			1.071
				(2.337)
	V_2			-0.414
				(1.015)
	V_3			0.067
				(0.371)
	V_4			-0.220
				(0.478)
	V_5			0.021
				(0.068)
	V_6			0.328
				(0.717)
	V_7			-0.028
				(0.068)
	V_8			0.089
				(0.198)
	V_9			-0.032
				(0.100)

The data are identical to BLP (1995). Column 1 replicates estimates for the model of their first column of results in their Table III. The second column uses the same instruments from BLP and estimates 2SLS for the characteristics used in Column 1. The third column estimates our control function approach using the price controls from Kim and Petrin (2010). We do not impose a supply side model during estimation. Standard errors do not yet reflect the "first-stage estimates."

Table 2
Automobile Elasticities: No Correction,
2SLS (without Interactions), and Control Function approach

	No Correction[1]	2SLS (w/o Interactions)	Control Function (with Interactions)
Results for 1971-1990			
Median	-0.77	-1.18	-2.06
Mean	-0.75	-1.60	-2.66
Standard Deviation	0.34	1.17	1.68
No. of Inelastic Demands	68%	21%	1%
Elasticities from 1990			
Median	-0.93	-1.43	-2.81
Mean	-0.91	-1.90	-3.24
Standard Deviation	0.46	1.28	1.84
No. of Inelastic Demands	53%	12%	0%
1990 Models (from BLP, Table VI):			
Mazda 323	-0.44	-0.69	-1.64
Honda Accord	-0.81	-1.26	-1.40
Acura Legend	-1.67	-2.57	-4.17
BMW 735i	-3.39	-5.09	-7.09

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).

price controls proposed in Petrin and Train (2010) and Kim and Petrin (2009). We extend the non-separable control function literature as our approach does not require that our controls be one-to-one with the unobserved factors. We develop a sieve semiparametric estimator for the nonseparable demand models. Monte Carlos suggest IV estimators in the non-separable setting perform poorly, while our control function approach is consistent. Using the same automobile data as was used in BLP (1995), our initial estimates reveal that these interactions terms are significant and make demand elasticities substantially more elastic.

10 Appendix

10.1 Do Higher Moments Help Identification?

Here we show that conditional moment restrictions using higher moments do not render identification either when p_j is endogenous. Suppose $E[\xi_j^2|Z] = \sigma_j^2$ (i.e., conditional homoskedasticity). Define $\tilde{\xi}_j^{(2)} \equiv \xi_j^2 - \sigma_j^2$ and by construction, we have

$$E[\tilde{\xi_j}^{(2)}|Z] = 0.$$

Define $g(\delta_j, x_j, p_j; c_0, \beta_0, \alpha_0) = \delta_j - c_0 - \beta_0' x_j + \alpha_0 p_j$ and $h(x_j, p_j; \gamma_0, \gamma_{p0}) = \gamma_0' x_j + \gamma_{p0} p_j$ for simplicity and also for generality of our claim.

Then to use the second moment, consider

$$\xi_j^2(\theta_0) = \frac{g^2(\delta_j, x_j, p_j; c_0, \beta_0, \alpha_0)}{(1 + h(x_j, p_j; \gamma_0, \gamma_{p0}))^2} \text{ or } \xi_j^2(\theta_0)(1+h(x_j, p_j; \gamma_0, \gamma_{p0}))^2 - g^2(\delta_j, x_j, p_j; c_0, \beta_0, \alpha_0) = 0. \quad (18)$$

It follows that replacing ξ_j^2 with $\tilde{\xi}_j^{(2)} + \sigma_j^2$,

$$\begin{aligned} 0 &= E[\xi_j^2(1 + h(x_j, p_j; \gamma_0, \gamma_{p0}))^2 - g^2(\delta_j, x_j, p_j; c_0, \beta_0, \alpha_0)|Z] \\ &= E[(\tilde{\xi}_j^{(2)} + \sigma_j^2)(1 + h(x_j, p_j; \gamma_0, \gamma_{p0}))^2 - g^2(\delta_j, x_j, p_j; c_0, \beta_0, \alpha_0)|Z] \\ &= E[\tilde{\xi}_j^{(2)} + \tilde{\xi}_j^{(2)}(2h(x_j, p_j; \cdot) + h^2(x_j, p_j; \cdot)) + \sigma_j^2(1 + h(x_j, p_j; \cdot))^2 - g^2(\delta_j, x_j, p_j; c_0, \beta_0, \alpha_0)|Z] \\ &= E[\tilde{\xi}_j^{(2)}(2h(x_j, p_j; \cdot) + h^2(x_j, p_j; \cdot)) + \sigma_j^2(1 + h(x_j, p_j; \cdot))^2 - g^2(\delta_j, x_j, p_j; c_0, \beta_0, \alpha_0)|Z] \end{aligned}$$

where the last equality holds by $E[\tilde{\xi}_j^{(2)}|Z] = 0$. Note that the above conditional moment condition based on the second moment does not help identification because it still contains the conditional expectations of interaction terms with $\tilde{\xi}_j^{(2)}$. Again the conditional moment restriction $E[\tilde{\xi}_j^{(2)}|Z] = 0$ alone cannot determine $E[\tilde{\xi}_j^{(2)} h(x_j, p_j; \gamma_0, \gamma_{p0})|Z]$ or $E[\tilde{\xi}_j^{(2)} h^2(x_j, p_j; \gamma_0, \gamma_{p0})|Z]$. This creates the essentially same non-identification problem with (2) because of the nonseparability. Specifically we cannot handle the term $E[\tilde{\xi}_j^{(2)} p_j|Z]$, $E[\tilde{\xi}_j^{(2)} p_j^2|Z]$, and other interaction terms containing p_j and $\tilde{\xi}_j^{(2)}$.

We can make similar arguments for moment restrictions based on other higher moments of ξ_j. Therefore using higher moments does not help identification in the CMR approach.

10.2 Identification with Exogenous Prices

We note that non-additive separability itself does not render non-identification result in the IV approach but it is the non-additive separability combined with endogeneity that makes the IV approach fail. To see this point, now suppose that prices are exogenous (so Z includes prices too). Then, we have

$$\begin{aligned} 0 &= E[\xi_j + \xi_j h(x_j, p_j; \gamma_0, \gamma_{p0}) - g(\delta_j, x_j, p_j; c_0, \beta_0, \alpha_0)|Z] \\ &= E[\xi_j h(x_j, p_j; \gamma_0, \gamma_{p0}) - g(\delta_j, x_j, p_j; c_0, \beta_0, \alpha_0)|Z] \\ &= E[-g(\delta_j, x_j, p_j; c_0, \beta_0, \alpha_0)|Z] \end{aligned} \quad (19)$$

where the last equation is obtained applying the law of iterated expectation. This moment condition alone identifies c_0, β_0, and α_0. Then by imposing additional restriction on the second moment of ξ_j, we can further identify the interaction parameters γ_0 and γ_{p0}. For example we may assume the conditional homoskedasticity as $E[\xi_j^2|Z] = \sigma_j^2$. Note that this is weaker than the full independence condition. We have

$$\begin{aligned} 0 &= E[\xi_j^2(1 + h(x_j, p_j; \gamma_0, \gamma_{p0}))^2 - g^2(\delta_j, x_j, p_j; c_0, \beta_0, \alpha_0)|Z] \\ &= E[\sigma_j^2(1 + h(x_j, p_j; \gamma_0, \gamma_{p0}))^2 - g^2(\delta_j, x_j, p_j; c_0, \beta_0, \alpha_0)|Z] \end{aligned}$$

and this moment condition further identifies γ_0, γ_{p0}, and σ_j^2 because we can treat c_0, β_0, and α_0 known since they are identified from (19) with valid IV's.

References

Altonji, J., and R. Matzkin (2005): "Cross Section and Panel Data Estimators for Nonseparable Models with Endogenous Regressors," *Econometrica*, 73, 1053–1102.

Berry, S. (1994): "Estimating Discrete-Choice Models of Product Differentiation," *The RAND Journal of Economics*, 25(2), 242–262.

Berry, S., J. Levinsohn, and A. Pakes (1995): "Automobile Prices in Market Equilibrium," *Econometrica*, 60(4), 889–917.

Blundell, R., and J. Powell (2003): "Endogeneity in Nonparametric and Semiparametric Regression Models," *Advances in Economics and Econometrics:Theory and Applications, Eighth World Congress.*

Chen, X. (2006): "Large Sample Sieve Estimation of Semi-Nonparametric Models," *forthcoming in Handbook of Econometrics, Volume 6.*

Chen, X., O. Linton, and I. van Keilegom (2003): "Estimation of Semiparametric Models When the Criterion Function is Not Smooth," *Econometrica*, 71, 1591–1608.

Gandhi, A. (2009): "Inverting Demand Systems," Working paper.

Hahn, J., and G. Ridder (2008): "Conditional Moment Restrictions and Triangular Simultaneous Equations," Working Paper.

Imbens, G., and W. Newey (2003): "Identification and Estimation of Triangular Simultaneous Equations Models without Additivity," Working Paper.

Kim, K., and A. Petrin (2009): "Control Function Corrections for Unobserved Factors in Differentiated Product Models," Working Paper.

——— (2010): "Control Functions for a Class of Non-Separable Models without Independence and Monotonicity Conditions," Working paper.

Newey, W., and J. Powell (2003): "Instrumental Variable Estimation of Nonparametric Models," *Econometrica*, 71, 1565–1578.

Newey, W., J. Powell, and F. Vella (1999): "Nonparametric Estimation of Triangular Simultaneous Equaitons Models," *Econometrica*, 67, 565–603.

Pakes, A. (1996): "Dynamic structural models, problems and prospects: mixed continuous discrete controls and market interaction," in *Advances in Econometrics, Sixth World Congress, Volume II*, ed. by C. Sims, pp. 171–259, New York. Cambridge.

Petrin, A., and K. Train (2010): "A Control Function Approach to Endogeneity in Consumer Choice Models," *Journal of Marketing Research*, 47.

Theil, H. (1953): “Repeated Least Squares Applied to Complete Equation Systems,” The Hague, Netherlands: Central Planning Bureau.