Nonparametric Estimation of Demand Elasticities Using Panel Data*

Stefan Hoderlein† Matthew Shum‡

Boston College Caltech

April 7, 2014

Abstract

In this paper, we propose and implement an estimator for price elasticities in demand models that makes use of Panel data. Our underlying demand model is nonparametric, and accommodates general distributions of product-specific unobservables which can lead to endogeneity of price. Our approach allows these unobservables to vary over time while, at the same time, not requiring the availability of instruments which are orthogonal to these unobservables. Monte Carlo simulations demonstrate that our estimator works remarkably well, even with modest sample sizes. We provide an illustrative application to estimating the cross-price elasticity matrix for carbonated soft drinks.

Keywords: demand elasticities, nonparametric estimation.

1 Introduction

The estimation of demand models occupies a large part of the empirical literature in industrial organization. Demand models are estimated in order to obtain

*We thank Khai Chiong for research assistance.

†Stefan Hoderlein: Boston College, Department of Economics, 140 Commonwealth Ave, Chestnut Hill, MA 02467, USA, email: stefan_hoderlein@yahoo.com.

‡Matthew Shum, Caltech, Division of Humanities and Social Sciences, MC 228-77, 1288 E. California Blvd., Pasadena, CA 91125, USA. email: mshum@caltech.edu.

values for the various own- and cross-price elasticities among a set of goods. In turn, these elasticities are crucial inputs into many policy evaluations of interest, including merger analysis (Nevo (2000)) and the welfare measurements of new goods (Hausman (1997), Petrin (2002)).

In this paper, we propose an estimator for price elasticities in demand models. Since our goal is to estimate demand elasticities in as flexible a manner as possible, we avoid making parametric restrictions on the underlying demand model, letting it be an arbitrary function of observed variables as well as unobserved variables. We emphasize in particular that we place no restrictions on individual level heterogeneity, as well as objects like unobserved product characteristics. The main insight in our approach is that demand elasticities are derivatives of the (log-) demand functions; we take a cue from recent developments in the econometrics of nonlinear, nonseparable models, which has shown that (average) derivatives of these models can be identified and estimated, even when the full underlying model is not. Hence, we dispense with estimating the underlying demand model, but rather focus on estimating its average derivatives.

As we mentioned above, the empirical literature on demand estimation is voluminous. There is a large literature on the use of flexible functional forms for the estimation of demand systems; perhaps the most well-known instances of these are the Translog (Jorgenson et. al. (1982)) and Almost Ideal Demand System (Deaton and Muellbauer (1980)) specifications. More recently, a large number of papers in empirical industrial organization has explored the estimation of aggregate demand models based on discrete choice models of individual behavior (Berry (1994), Berry, Levinsohn, Pakes (1995; hereafter "BLP")). At the same time, there are also recent papers exploring the nonparametric identification of these models (eg. Berry and Haile (2008), Chiappori and Komunjer (2009)) but these papers have not explored estimation. As far as we are aware, this is one of the first papers to consider estimation of demand elasticities from a fully nonparametric demand system. At the same time, our estimators for these demand elasticities are very easy to compute, in one form involving little more than regression techniques, and are readily implementable using standard statistical or econometric software packages. We are also among the first to apply tools from the recent literature on nonlinear panel data models to a demand estimation setting.

Methodologically, as alluded to above, our estimator is related to literature on nonlinear models in which the observed and unobserved variables do not enter in a separable manner, and where they are correlated with the observables. This literature dates back to work by Chamberlain (1982), and interest was recently revived by important papers of Altonji and Matzkin (2005) and Graham and Powell (2012). Our estimation strategy follows, in particular, the paper of Chernozhukov, Fernandez-Val, Hoderlein, Holzmann and Newey (2013, CFHHN)), who consider the estimation of average derivatives of a general nonseparable panel data model, in which the observed variables can be arbitrarily correlated with time-varying unobserved components (thus generalizing the notion of "fixed effects" in linear panel models. Importantly, a benefit of adapting this nonlinear panel approach to demand estimation is that it does not require the availability of instruments for endogenous prices, as is needed in most other econometric demand models.

There are several features of the CFHHN framework which make it particularly natural for estimation in a demand context. Compared to other panel data models, CFHHN allow several unobserved components to enter the model in an arbitrary fashion; in contrast, both Chamberlain (1982) and Graham and Powell (2012) assume a linear correlated random coefficients structure, which rules out demand models (such as BLP) in which the market level model is obtained as an aggregate of individual-level multivariate choice models. Altonji and Matzkin (2005; AM) consider a nonseparable model whose general structure is compatible with an individual level multivariate choice model. However, AM's framework cannot allow for arbitrary correlation between the unobserved product characteristics and the observed characteristics in this setup.[1] Moreover, both AM and Hoderlein and White (2012; HW) assume that the correlated unobservable is time invariant, but the CFHHN approach we follow allows for time-varying correlated unobservables. On the downside, however, this approach only allows us to estimate demand elasticities for a certain subpopulation of markets – namely, the markets for which the observables (including prices) change only little between two time periods. (Taking a cue from

[1] More specifically, AM need to impose restrictions on the time dependence of these quantities; an example of such a restriction would be that the current observables are independent of the unobserved characteristics, conditional on past observables.

the program evaluation literature, we will refer to these markets as a subpopulation of "stayers".)

Finally, this paper also contributes to the small but growing literature on panel data analysis of multinomial choice models. This includes Hausman (1996), Nevo (2001), Moon, Shum and Weidner (2010), and Pakes and Porter (2013). One difference of the present paper relative to these others lies in how the dimensions of the panel are defined. In the above papers, the "cross-sectional" units are products, and the "time periods" can be either markets or explicit time periods. In this paper, however, the cross-sectional unit is a market which is assumed to be observed across different time periods. The reason we use this definition here is because we are interested in estimating the whole matrix of cross-product elasticities, and this cannot be done if we were to treat each product as a separate cross-sectional observation.

In the next section we introduce the model and describe our estimator. Section 3 presents Monte Carlo simulation results, and Section 4 contains an empirical application. Section 5 concludes.

2 Model and Estimator

Our environment is one in which the researcher has data on quantities (or market shares), prices, and product characteristics. We consider a panel setting, in which the quantities sold and prices for each product are observed for a small number of periods, and across many geographic markets. That is, the two dimensions in our panel are time and geographic markets, with time being the "short" dimension, and the number of markets being the "long" dimension. For what follows then, we will just assume that the number of periods $T = 2$, while the number of markets $M \to \infty$, i.e. we consider a population of markets.

Specifically, consider a given product market, consisting of J products (indexed by $j = 1, \ldots, J$), along with corresponding J-vectors Y, P and $X \equiv (X^1, \ldots, X^K)$ denoting, respectively, the log-quantities, prices and K characteristics for the products. The panel data $\{Y_{mt}, P_{mt}, X_{mt}\}$ are observed, for two time periods $t = 1, 2$ and a large number of markets $m = 1, \ldots, M$. We consider a general demand system linking prices and characteristics of the products to the quantity of the product sold:

$$Y_{mt} = \phi(P_{mt}, X_{mt}, V_{mt}) \quad t = 1, 2; \quad m = 1, \ldots, M. \tag{1}$$

In the above, Y, X, and P are observed, while the vector V is unobserved. One could consider V_t to contain objects such as a classical fixed effect or time invariant random coefficients, but also a time varying variables which may contain correlated time varying unobservables like product specific charactersitics, but also traditional exogenous shocks. It is important to notice that the correlated unobservables can vary across both markets m and time periods t. We want to emphasize here that in theory there is no restriction to the dimensionality of the unobservable; there may be arbitrarily many such variables. Also, as we will see below, these unobservables may be arbitrarily correlated with the observables.

In what follows, we will typically omit the m subscript for convenience; that is, we now discuss identification of the cross-price elasticities of demand population of markets. To do so, we introduce the following notation: Let ∂_p denote the $J \times J$ Jacobian matrix of the $\phi(\cdots)$ function with respect to the vector P. Let Δ denote the time-difference of a variable between periods $t = 1$ and $t = 2$, e.g., $\Delta X = X_2 - X_1$. Throughout, we assume that P is continuously distributed. The main identifying assumption in CFHHN (2012) is the following:

Assumption A1 *V_t is conditionally stationary:*

$$F_{V_1|\Delta X=0, X_1, \Delta P=0, P_1} = F_{V_2|\Delta X=0, X_2, \Delta P=0, P_2}.$$

This assumption restricts the vector valued V_t process to have a stationary marginal distribution; this means that the period-specific vector of shock is always drawn from the same distribution. This allows for many correlated product characteristics across markets to change arbitrarily over time, as long as their conditional distribution would stay the same across time. However, it allows for the unconditional distribution to change across time. If for instance marketing campaigns are being combined with price changes across time, this can be well accommodated in our model.[2] On the other hand, this assumption essentially rules out major period-

[2]This assumption relaxes the setup in HW, which assumed correlated market- and product specific variables to be time-invariant. Such an assumption may be considered restrictive, especially as there may be unobserved variables (such as advertising campaigns) which vary over time, and can be chosen by firms in conjunction with their pricing decisions.

specific aggregate shocks, but allows for time series correlation in the idiosyncratic shocks. It is best satisfied, if the two periods are close by, so that the macroeconomic environment is comparable.

The second set of assumptions are differentiability assumptions:

Assumption A2 *$P \in \mathcal{P}$ where $\mathcal{P}$ is an open convex set, and for each $(x, v) \in \mathcal{X} \times \mathcal{V}$, $\phi(\cdot, x, v)$ is twice continuously differentiable on $\mathcal{P}$ with bounded derivatives up to order two.*

This assumption requires the structural function ϕ to be differentiable. Moreover, we assume that there is positive density around zero changes, i.e., there is a substantial part of the population that experience zero or very small price changes across time, and - by the differentiability assumption - responds smoothly to those.

To state the main result, we define the *local average response (LAR)* of the $\phi(\cdots)$ function. Letting Δ , the LAR is

$$\mathbb{E}\left[\partial_p \phi(P_1, X_1, V_1) | \Delta P = 0, \Delta X = 0, p_1, X_1\right].$$

Obviously, the LAR here corresponds to the matrix of cross-price semi-elasticities of the demand system $\phi(\cdots)$, averaged over all unobserved components conditional on $(\Delta P = 0, \Delta X = 0, P_1, X_1)$. (By analogy with the program evaluation literature, the subpopulation of markets for which $(\Delta P = 0, \Delta X = 0, P_1, X_1)$ are "stayers" for which prices and characteristics in the second period remained the same as in the first period.) The corresponding matrix of cross-price elasticities, then, can be obtained by multiplying through by the vector of prices in the first period:

$$\mathbb{E}\left[\partial_p \phi(P_1, X_1, V_1) | \Delta P = 0, \Delta X = 0, p_1, X_1\right] \otimes P_1'.$$

Under the previous assumptions,[3] CFHHN (2013) show the identification of the LAR matrix:

Proposition 1. (CFHHN (2013)): The LAR is equivalent to a derivative of the first-differenced regression of ΔY on $(\Delta P, P_1)$:

$$\mathbb{E}\left[\partial_p \phi(P_1, X_1, V_1) | \Delta P = 0, \Delta X = 0, P_1, X_1\right] = \partial_{\Delta P} \mathbb{E}\left[\Delta Y | \Delta P, \Delta X, P_1, X_1\right] |_{\Delta P = 0}.$$

[3] See CFHHN (2013) for details. There are some additional technical assumptions underlying the results, but we do not list them here for convenience.

Discussion of Proposition 1. Obviously, the quantity of the right-hand side is estimable straight from the data, as it involves only the observed variables Y and P. The left hand side object is the structural object of interest: Let ξ_1 denote time and market specific correlated product characteristics, a subvector of the vector $V_1 = v_1$ in period 1. Then, $\partial_P \phi(p_1, x_1, v_1)$ defines the structural elasticity for a person or market characterized by a certain set of (possibly time varying) preferences who experiencies a certain market environment in period 1, including a subvector ξ_1 that described the market specific product characteristic in this period 1, when faced with prices $P_1 = p_1$ and characterized by other observable covariates $X_1 = x$. We obtain the average of these structural elasticities for the subpopulation which faces the same price and and other variables (including observed time varying product characteristics as part of X_t, time invariant product characteristics can simply be omitted as time invariant variables automatically satisfy the stationarity property), but for which prices and these other variables do not change dramatically between the periods. For typical applications in industrial organization, this can be a large subsample, especially in mature product markets were prices and product characteristics are not very volatile across time, where price variation across markets is more prominent than price variation over time.[4]

We provide an informal sketch of the proof for this result, but refer to CFHHN and HW for the general result. Consider the special case in which there are no X variables, and where there is only a single good ($J = 1$). Then, we can write $\Delta Y = \phi(P_1 + \Delta P, V_2) - \phi(P_1, V_1)$. Using a linearization of $\phi(P_1 + \Delta P, V_2)$ around P_1, and considering small price changes ΔP, we get that by A2

$$\Delta Y \cong \phi(P_1, V_2) + \Delta P \cdot \partial_p \phi(P_1, V_2) - \phi(P_1, V_1)$$

Now, A1 implies that $\mathbb{E}[\phi(P_1, V_2)|\Delta P, P_1] - \mathbb{E}[\phi(P_1, V_1)|\Delta P, P_1] = 0$, so that

$$\begin{aligned} \mathbb{E}[\Delta Y|\Delta P, P_1] &\cong \Delta P \cdot \mathbb{E}[\partial_p \phi(P_1, V_2)|\Delta P, P_1] \\ &= \Delta P \cdot \mathbb{E}[\partial_p \phi(P_1, V_1)|\Delta P, P_1] \end{aligned}$$

implying that $\frac{\partial}{\partial \Delta P} \mathbb{E}[\Delta Y|\Delta P, P_1] \cong \mathbb{E}[\partial_p \phi(P_1, V_1)|\Delta P, P_1]$ for ΔP small. ■

[4] Moreover, under additional conditions, the proposition above can be extended to identify the average elasticities across the whole population, but we do not elaborate on this here, see CFHHN and HW for details.

Example: random-coefficients logit demand Next, we illustrate the scope of our nonparametric estimator by considering a flexible multinomial choice model similar to the random-coefficients logit model of Berry, Levinsohn, and Pakes (1995). This is also the example used in our Monte Carlo experiments below. Consider a market for J products. We observe the aggregate market shares S_{jt}, for products $j = 1, \ldots, J$ across markets $m = 1, \ldots, M$ and time periods $t = 1, \ldots, T$. The market share function for product j in market m and period t takes the form

$$S^j_{mt} = \int \frac{\exp(-\alpha p^j_{mt} + X^j_{mt}\beta + \xi^j_{mt} + \eta^j_{mt})}{1 + \sum_{j'=1}^{J} \exp(\alpha p^{j'}_{mt} + X^{j'}_{mt}\beta + \xi^{j'}_{mt} + \eta^{j'}_{mt})} dG(\alpha, \beta; \theta_m). \tag{2}$$

Here, the price coefficient α and the taste parameter β vary across individual consumers according to the distribution G, with market specific parameter θ_m, say the mean and variance. In principle, we can let this distribution vary with time, but we desist from this greater generality here as it is not matched in the literature.

Applying the notation in Eq. (1) to Eq. (2) above, the time-varying correlated effects V_{mt} (which we allow to vary arbitrarily across markets, and which generalize the classical notion of fixed effects) contains both the time-invariant but market-specific preference distribution parameters θ_m, as well as the preference shocks $\left\{\eta^j_{mt}, j = 1, \ldots, J\right\}$ and unobserved product characteristics $\xi^1_{mt}, \ldots, \xi^J_{m}$, both of which can move market shares across time periods.[5] ■

2.1 Estimation

The identification results above imply that $\partial_{\Delta P}\mathbb{E}\left[\Delta Y|\Delta P, \Delta X, P_1, X_1\right]|_{\Delta P=0}$ is the reduced form object of interest which is to be estimated. HW propose an estimator for this quantity in which the conditional expectation on the RHS of the statement of the proposition is estimated by a local quadratic regression. Accordingly, the partial derivative of this conditional expectation can be computed from the coefficients of the local quadratic regression. Details of this regression procedure are provided in the next section. HW also derive the asymptotic theory for the estimated average partial derivatives; since the limiting distributions involve components which are

[5] Note also that our model, in Eq. (1), allows for interactions between X_j and ξ_{jt}, thus addressing on important critique in this literature (cf. Gandhi, Kim, Petrin (2011)).

tedious and difficult to compute, we approximate the standard errors in the estimates using bootstrap resampling.

3 Monte Carlo experiments

In order to gauge the small-sample performance of our estimator, and also consider other design aspects in the implementation of our estimator, we perform a Monte Carlo exercise based on the simple multinomial logit demand model, as in Eq. (2). We assume that $J = 2$, so there are only two available products (in addition to outside good). Then the market shares are given by (for $j = 1, 2$):

$$S^j_{mt} = \frac{\exp(-\alpha p^j_{mt} + \xi^j_{mt} + \eta^j_{mt})}{1 + \sum_{j'=1}^{2} \exp(\alpha p^{j'}_{mt} + \xi^{j'}_{mt} + \eta^{j'}_{mt})}. \tag{3}$$

There are M markets (indexed by m), and two time periods ($t = 1, 2$).

In the first set of experiments, we assume that the fixed effects are time-invariant; specifically, the fixed effects are generated by $\xi^{j=1}_m \sim U[0.5, 1.5]$ and $\xi^{j=2}_m \sim U[1, 2]$. The idiosyncratic shocks are generated as $\eta^j_{mt} \sim N(0, 0.5)$, i.i.d. across (j, m, t). Prices are generated in several steps. Define $y^j_{mt} = \xi^j_m + \nu^j_{mt}$, with $\nu^j_{mt} \sim N(0, \sigma^2)$. Then define the prices

$$p^j_{mt} = \begin{cases} y^j_{mt} & \text{if } y^j_{mt} > 0 \\ 0.1 & \text{otherwise.} \end{cases}$$

So, on average across markets and over time, good 1 is more expensive than good 2.

The model parameters are set as $\alpha = -1$, $\sigma = 0.5$. In the exercises below, we consider dataset size of $J = 2$, $T = 2$, $M = (100, 500, 1000)$ for each replication, and consider 100 replications for each exercise.

In Table 1 we report the results for the average elasticities. Proposition 1 implies that we can estimate an average cross-price elasticity for product j with respect to a price change in product i by the expression

$$\epsilon_{i,j,t=1} \equiv \mathbb{E}\left\{p^i * \mathbb{E}\left[\Delta Y^j | \Delta p^1 = 0, \Delta p^2 = 0, p^1_{t=1} = \bar{p}^1, p^2_{t=1} = \bar{p}^2\right]\right\}.$$

We approximated the above by:

$$\tilde{\epsilon}_{i,j,t=1} \approx \frac{1}{M} \sum_{m=1}^{M} p^i_{m,t=1} \cdot \left(\mathbb{E}\left[\Delta Y^i_m | \Delta p^1 = 0, \Delta p^2 = 0, p^1_{m,t=1}, p^2_{m,t=1}\right]\right).$$

For the estimates reported in Table 1, we use a local quadratic regression (as proposed in HW (2012)) for estimating the conditional expectation $\mathbb{E}[\Delta Y^j|\Delta p^1 = 0, \Delta p^2 = 0, p^1_{t=1} = \bar{p}^1, p^2_{t=1} = \bar{p}^2]$ in the above expression. Specifically, we computed a local quadratic regression of $\Delta Y^j = Y^j_{m,t=2} - Y^j_{m,t=1}$ on a constant and linear and quadratic terms in the components $(\Delta p^1_m, \Delta p^2_m)$, $(p^1_{m,t=1}, p^2_{m,t=1})$ using kernel weights equal to

$$w_m = K\left(\frac{\Delta p^1_m}{h_1}\right) K\left(\frac{\Delta p^2_m}{h_1}\right) K\left(\frac{p^1_{m,t=1} - \bar{p}^1}{h_2}\right) K\left(\frac{p^1_{m,t=2} - \bar{p}^2}{h_2}\right). \tag{4}$$

We used a standard Gaussian kernel $K(x) = (\pi)^{-0.5}\exp(-0.5 * X^2)$. We also explored different values for the bandwidths (h_1, h_2). In Table 1, we report the average cross-price elasticities for each of the four pairs of products (i, j), with $i, j \in \{1, 2\}$. As we would expect, the own-price elasticities $\epsilon_{1,1}$ and $\epsilon_{2,2}$ are negative in sign because, in the logit model, the two products are substitutes; correspondingly, the cross-price elasticities $\epsilon_{1,2}$ and $\epsilon_{2,1}$ are positive in sign, and smaller in magnitude than the own-price elasticities.

In Table 1, we also report the root-mean-squared error (RMSE) between the estimated elasticities and the true values for the elasticities. For the true values of the elasticities, we note that in the multinomial logit model, the formula for the average cross-price elasticity for product j with respect to a price change in product i $(i, j \in \{1, 2\})$:

$$\tilde{\epsilon}_{i,j,t=1} = \mathbb{E}_{p^1,p^2} p^i * \mathbb{E}\left[\alpha * (1 - S^j_{m,1})|\Delta p^1 = 0, \Delta p^2 = 0, p^1_{t=1} = \bar{p}^1, p^2_{t=1} = \bar{p}^2\right]$$

where the partial mean $\mathbb{E}\left[\alpha * p^i_{m,1} * (1 - S^i_{m,1})|\Delta p^1 = 0, \Delta p^2 = 0, p^1_{m,t=1} = \bar{p}^1, p^2_{m,t=1} = \bar{p}^2\right]$ was, again, computed using a local quadratic regression of $\alpha * p^i_{m,1} * (1 - S^i_{m,1})$ on a constant and $(\Delta p^1_m, \Delta p^2_m, p^1_{m,t=1}, p^2_{m,t=1})$ using kernel weights as in (4) above. In the results reported below, for each set of exercises we used the same bandwidth (h_1, h_2) in computing both the estimates of the average elasticities, as well as computing their true counterparts. We see that, uniformly across different sample sizes, the root-mean-squared errors are small, indicating that our elasticity estimates are quite accurate. There is a mild deterioration in accuracy in the smallest sample size $(M = 100)$; but even in this case, we see that the average elasticity estimator performs very well for bandwidths of $h = 0.5, 1.0$.

In Table 2 we consider a second specification, in which the fixed-effects $\vec{\xi}_{mt} = (\xi^1_{mt}, \xi^2_{mt})$ are allowed to be time-varying, corresponding to the main specification put forward in this paper and analyzed in CFHHN (2013). Note that in this specification, there are no time-invariant observables at all. The stationarity condition (Assumption A1) in this case becomes

$$F_{\vec{\xi}_{m,1}}(\cdot|\vec{p}_{m,1}, \vec{p}_{m,2}) = F_{\vec{\xi}_{m,2}}(\cdot|\vec{p}_{m,1}, \vec{p}_{m,2}).$$

That is, the distribution of $\vec{\xi}_{m,t}$, conditional on both period's prices $\vec{p}_{m,1}, \vec{p}_{m,2}$, is invariant for $t = 1, 2$.

In the simulation, we generate the prices in several steps. Define $y^1_{mt} \sim \mathcal{N}(1, \sigma^2)$ and $y^2_{mt} \sim \mathcal{N}(1.5, \sigma^2)$. Then define the prices

$$p^j_{mt} = \begin{cases} y^j_{mt} & \text{if } y^j_{mt} > 0 \\ 0.1 & \text{otherwise.} \end{cases}$$

Then we generate the fixed effects as

$$\xi^1_{m,t} = 0.5 * (p^1_{m,1} + p^1_{m,2}) + z^1_{m,t}; \quad z^1_{m,t} \sim \mathcal{N}(0, 1)$$

and

$$\xi^2_{m,t} = 0.5 * (p^2_{m,1} + p^2_{m,2}) + z^2_{m,t}; \quad z^2_{m,t} \sim \mathcal{N}(0, 1).$$

Thereupon, the design of the Monte Carlo experiment proceeds as in the case described previously. The results are reported in Table 2. Not unexpectedly, we see that the RMSE in these specifications are typically at least one order of magnitude higher than in the corresponding results in Table 1. Also, there is a more marked deterioration in performance for the smallest sample size ($M = 100$), especially for the bandwidth of $h = 0.25$, which was the smallest which we considered. However, even with such a small sample size, the RMSE stays remarkably small when the bandwidth is increased (to $h = 0.5, 1.0$). One lesson from these Monte Carlos appears to be that we should not use very small bandwidths.

4 Empirical application

4.1 Data description

We consider an application to scanner data from the carbonated soft drink market. The data are drawn from the IRI Marketing Dataset, which is a large-scale scanner panel data set which is ongoing since 2000, see Bronnenberg, Kruger, and Mela (2008) for a description. This dataset contains weekly-level sales and prices for all soft drinks for a large sample of 1025 stores across the United States. Using this dataset, each cross-sectional unit is a store, while each time period is a week. We consider two weeks of data, corresponding to the two consecutive weeks beginning from July 16-22 and July 23-29, 2001.[6]

We aggregate up to the six largest brands: (i) Coke; (ii) Pepsi; (iii) Sprite – regular and diet; (iv) Mountain Dew – regular and citrus-flavored; (v) Diet Coke; (vi) Diet Pepsi. Of these six beverages, the Coca-Cola company produces Coke, Diet Coke, and Sprite, whereas PepsiCo produces Pepsi, Diet Pepsi, and Mountain Dew. The estimated matrix of cross-price elasticities is presented in Table 3. In these results, the various bandwidths used in the local quadratic regression were set proportional to the standard deviations of the variables; for the proportionality constant we tried both 1.0 and 0.75.

4.2 Computational details for estimation

As discussed above, the identification result in CFHHN (2013) and HW (2012) has the advantage that it is constructive, in the sense that it lends itself to straightforward sample counterparts estimation. In particular, following HW (2012), we employ local polynomial estimators (more precisely, locally quadratic) estimators to estimate the derivative of the mean regression[7]. More specifically, we use a standard Gaussian Kernel. Instead of choosing a separate bandwidth for every regressor, we choose the bandwidth to be proportional to a baseline bandwidth times the stan-

[6]For previous empirical work on the carbonated soft drinks market, see Gasmi, Laffont, and Vuong (1992) and Dube (2004).

[7]The (standard) convergence behavior, including asymptotic normality at a nonparametric rate (recall that we are working with the subpopulation for which $\Delta P = 0$) is established in HW (2012).

dard deviation of the respective regressor, and instead of choosing a single value for the bandwidth, we have experimented with various bandwidths to see whether the results are robust, at least in a qualitative sense. Since we consider estimating a derivative of a regression function, there is no clear guidance on how to do this in a data driven way, and we have thus opted to present results for several values of the bandwidth.

4.3 Empirical results

Overall, we see that the estimates of the cross-price elasticities are generally less precisely estimated than the own-price elasticities. However, the point estimates indicate that some products are substitutes, while others are complements and, in addition, that signs of the cross-price elasticity matrix is not always symmetric. Note in this respect that there is no need for symmetry to hold, even if individual rationality were to hold, as the equations are market level aggregates over unobserved heterogeneity. For instance, focusing on the $h = 1.0 * \sigma$ results (the top panel of Table 3), we see that the demand for Pepsi responds positively to a price rise in Coke (cross-price elasticity is 1.2406); however, the demand for Coke responds negatively to price increases in Pepsi (cross-price elasticity is -0.5628). Results like these suggest a broader range of own- and cross-price elasticities than would be allowed for in typical discrete-choice models, and support the case for nonparametric regression. For instance, in logit-based multinomial models, substitution between all products is imposed from the outset as a parametric restriction, and this assumption seems to be violated for some beverages.

5 Conclusions

In this paper we have proposed a new estimator for the matrix of cross-price elasticities in demand models, utilizing panel data in product-level quantities and prices observed across a large number of markets in a small number of time periods. We allow the underlying demand model to be nonparametric, and allow the product-specific unobservables to be arbitrarily complex an enter in arbitrary nonlinear fashion. Monte Carlo simulations demonstrate that our estimator works remarkably well, even with modest sample sizes. An illustrative empirical application to the carbonated soft drink market reveals some patterns of both complementarity and substitutability across different soft drink products, which suggest that a typically logit-based approach (which imposes substitution across all products) may be too restrictive for this market.

References

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

[2] S. Berry, J. Levinsohn, and A. Pakes (1995), "Automobile Prices in Market Equilibrium," *Econometrica*, 63: 841-890.

[3] S. Berry, J. Levinsohn, and A. Pakes (2004), "Differentiated Products Demand Systems from a Combination of Micro and Macro Data: the New Car Market," *Journal of Political Economy*, 112: 68-105.

[4] S. Berry and P. Haile (2009), "Nonparametric Identification of Multinomial Choice Demand Models with Heterogeneous Consumers," NBER working paper 15276.

[5] B. Bronnenberg, M. Kruger, and C. Mela (2008), "The IRI Marketing Dataset," *Marketing Science*, 27: 745-748.

[6] G. Chamberlain (1982), "Multivariate Regression Models for Panel Data," *Journal of Econometrics,* 18(1), 5 - 46.

[7] V. Chernozhukov, I. Fernandez-Val, S. Hoderlein, S. Holzmann, W. Newey (2013), "Nonparametric Identification in Panels using Quantiles", MIT, working paper.

[8] P.-A. Chiappori and I. Komunjer (2009), "On the Nonparametric Identification of Multiple Choice Models," manuscript, Columbia University.

[9] A. Deaton and J. Muellbauer (1980), *Economics and Consumer Behavior*, Cambridge University Press.

[10] J.-P. Dube (2004), "Multiple Discreteness and Product Choice: Demand for Carbonated Soft Drinks," *Marketing Science*, 23: 66-81.

[11] A. Gandhi, K. Kim, and A. Petrin (2011), "Identification and Estimation in Discrete Choice Demand Models when Endogenous Variables Interact with the Error," NBER working paper 16894.

(h_1, h_2)		$\epsilon_{1,1,t=1}$	$\epsilon_{1,2,t=1}$	$\epsilon_{2,1,t=1}$	$\epsilon_{2,2,t=1}$
M=100					
(1.0, 1.0)	avg.[a] estimate	-0.6760	0.4969	0.3233	-0.9894
	RMSE[b]	0.0121	0.0115	0.0133	0.0098
(0.5, 0.5)	avg. estimate	-0.6643	0.5034	0.3010	-1.0182
	RMSE	0.0303	0.0317	0.0333	0.0308
(0.25, 0.25)	avg. estimate	-0.6407	0.5104	0.2489	-1.0448
	RMSE	0.2414	0.2739	0.1842	0.2561
M=500					
(1.0, 1.0)	avg. estimate	-0.6807	0.4923	0.3352	-0.9863
	RMSE	0.0029	0.0027	0.0028	0.0026
(0.5, 0.5)	avg. estimate	-0.6813	0.4753	0.3193	-0.9957
	RMSE	0.0055	0.0067	0.0057	0.0056
(0.25, 0.25)	avg. estimate	-0.6903	0.4603	0.2979	-1.0050
	RMSE	0.0376	0.0373	0.0409	0.0308
M=1000					
(1.0, 1.0)	avg. estimate	-0.6746	0.5138	0.3276	-1.0021
	RMSE	0.0013	0.0013	0.0013	0.0015
(0.5, 0.5)	avg. estimate	-0.6822	0.4993	0.3227	-1.0152
	RMSE	0.0028	0.0028	0.0030	0.0029
(0.25, 0.25)	avg. estimate	-0.7035	0.4939	0.3160	-1.0394
	RMSE	0.0169	0.0184	0.0179	0.0175

$$\mathbb{E}_{\vec{P}_{t=1}} \vec{P}_{t=1} * \mathbb{E}\left[\partial_{\vec{p}} \log S_{t=1} | \Delta \vec{P} = 0, \vec{P}_{t=1}\right]$$

[a]: averaged across 100 Monte Carlo replications

[b]: taken across 100 Monte Carlo replications

Table 1: Monte Carlo Results: Average Elasticities

(h_1, h_2)		$\tilde{\epsilon}_{1,1,t=1}$	$\tilde{\epsilon}_{1,2,t=1}$	$\tilde{\epsilon}_{2,1,t=1}$	$\tilde{\epsilon}_{2,2,t=1}$
M=100					
(1.0, 1.0)	avg. estimate	-0.7473	0.6375	0.3381	-0.8575
	RMSE	0.0661	0.0678	0.0368	0.0387
(0.5, 0.5)	avg. estimate	-0.7536	0.5956	0.3354	-0.8404
	RMSE	0.1568	0.1730	0.0729	0.0950
(0.25, 0.25)	avg. estimate	-0.8131	0.4068	0.2953	-0.7842
	RMSE	0.8994	0.9890	0.3745	0.4599
M=500					
(1.0, 1.0)	avg. estimate	-0.6896	0.6379	0.3112	-0.8457
	RMSE	0.0133	0.0131	0.0047	0.0049
(0.5, 0.5)	avg. estimate	-0.6948	0.6202	0.3117	-0.8360
	RMSE	0.0310	0.0285	0.0111	0.0133
(0.25, 0.25)	avg. estimate	-0.6578	0.6200	0.2897	-0.8218
	RMSE	0.1936	0.2595	0.0672	0.0971
M=1000					
(1.0, 1.0)	avg. estimate	-0.7005	0.6376	0.3206	-0.8617
	RMSE	0.0074	0.0075	0.0028	0.0026
(0.5, 0.5)	avg. estimate	-0.6935	0.6443	0.3194	-0.8626
	RMSE	0.0181	0.0158	0.0066	0.0052
(0.25, 0.25)	avg. estimate	-0.6724	0.7039	0.3226	-0.8931
	RMSE	0.0968	0.0875	0.0358	0.0394

$$\mathbb{E}_{\vec{P}_{t=1}} \vec{P}_{t=1} * \mathbb{E}\left[\partial_{\vec{p}} \log S_{t=1} | \Delta \vec{P} = 0, \vec{P}_{t=1}\right]$$

[a]: averaged across 100 Monte Carlo replications

[b]: taken across 100 Monte Carlo replications

Table 2: Monte Carlo Results: Average Elasticities, with "time-varying" fixed effects

	Coke	Pepsi	Sprite	MDew	DCoke	DPepsi
Coke	-1.0830	-0.5628	1.8413	0.5399	-0.1575	-0.2852
	(0.1791)	(0.2062)	(0.2448)	(0.1071)	(0.2341)	(0.1606)
Pepsi	1.2406	-0.5740	-1.4945	-0.1959	-1.5974	0.6834
	(0.2464)	(0.2645)	(0.2649)	(0.1530)	(0.2869)	(0.1634)
Sprite	-0.4123	-0.1491	-1.6711	1.6983	0.1389	-0.8239
	(0.3281)	(0.3893)	(0.4385)	(0.2616)	(0.3934)	(0.2535)
MDew	-1.0440	2.0596	0.3331	-0.7966	-1.6627	1.2473
	(0.4439)	(0.4637)	(0.5120)	(0.2480)	(0.5942)	(0.3157)
DCoke	-0.2532	-1.2477	2.1616	-0.0683	-0.5688	-0.4286
	(0.2269)	(0.2359)	(0.2368)	(0.1377)	(0.2542)	(0.1553)
DPepsi	1.5480	0.5649	0.7368	-1.0669	-1.8764	-0.5909
	(0.4366)	(0.4667)	(0.4894)	(0.3510)	(0.5352)	(0.2752)

Bandwidth: $h_1, h_2 = 1.0 * \sigma$

Coke	-1.1517	-0.8629	2.0175	0.6848	0.3724	-0.5123
	(0.2077)	(0.2449)	(0.2770)	(0.1326)	(0.2703)	(0.1818)
Pepsi	1.7444	-0.6680	-1.3743	-0.2956	-2.2893	0.7207
	(0.3035)	(0.3126)	(0.3261)	(0.1928)	(0.3414)	(0.1960)
Sprite	0.2026	0.1586	-2.5461	0.7669	0.3615	-0.8426
	(0.4255)	(0.4604)	(0.5452)	(0.2893)	(0.5141)	(0.3313)
MDew	-0.3363	2.7415	-0.3518	-0.7147	-1.0706	0.5454
	(0.5497)	(0.5497)	(0.6662)	(0.3114)	(0.9168)	(0.3678)
DCoke	-0.4702	-1.5552	2.4967	0.0321	-0.2299	-0.4458
	(0.2807)	(0.2921)	(0.2994)	(0.1666)	(0.3014)	(0.1826)
DPepsi	1.7962	0.1132	1.7968	-0.0917	-2.4599	-1.0308
	(0.5426)	(0.5878)	(0.6446)	(0.4134)	(0.7112)	(0.3632)

Bandwidth: $h_1, h_2 = 0.75 * \sigma$

Elasticity of demand for row brand, with respect to price change in column brand.
Bootstrapped standard errors in parentheses.

Table 3: Matrix of Cross-price elasticities

[12] F. Gasmi, J. Laffont, and Q. Vuong (1992), "Econometric Analysis of Collusive Behavior in a Soft-Drink Market," *Journal of Economics and Management Strategy*, 1: 277-311.

[13] B. Graham and J. Powell (2008): "Identification and Estimation of 'Irregular' Correlated Random Coefficient Models," NBER Working Paper 14469.

[14] J. Hausman (1997), "Valuation of New Goods under Perfect and Imperfect Competition," in *The Economics of New Goods*, University of Chicago Press.

[15] S. Hoderlein and H. White (2012), "Nonparametric Identification in Nonseparable Panel Data Models with Generalized Fixed Effects," *Journal of Econometrics*, 168: 300-314.

[16] D. Jorgenson, T. Stoker, and L. Lau (1982), "The Transcendental Logarithmic Model of Aggregate Consumer Behavior," in *Advances in Econometrics*, Volume 1 (pp. 97-238), JAI Press.

[17] R. Moon, M. Shum, and M. Weidner (2010), "Estimation of Random Coefficients Logit Demand Models with Interactive Fixed Effects," working paper.

[18] A. Nevo (2000), "Mergers with differentiated products: The case of the ready-to-eat cereal industry," *RAND Journal of Economics*, pp. 395-421.

[19] A. Nevo (2001), "Estimating Market Power in the Ready-to-Eat Cereal Industry," *Econometrica*, 69: 307-342.

[20] A. Pakes and J. Porter (2013), "Moment Inequalities for Semiparametric Multinomial Choice with Fixed Effects," working paper.

[21] A. Petrin (2002), "Quantifying the Benefits of New Products: the Case of the Minivan," *Journal of Political Economy*, 110: 705-729.