Nonparametric Demand Estimation in Differentiated Products Markets *

Giovanni Compiani †

November 2018

Abstract

I propose a nonparametric approach to estimate demand in differentiated products markets. Demand estimation is key to addressing many questions in economics that hinge on the shape—and notably the curvature—of market demand functions. My approach allows for a broad range of consumer behaviors and preferences, including standard discrete choice, complementarities across goods, consumer inattention and loss aversion. Further, no distributional assumptions are made on the unobservables and only limited functional form restrictions are imposed on utility. Using California grocery store data, I apply my approach to perform two counterfactual exercises: quantifying the pass-through of a tax and assessing how much the multi-product nature of sellers contributes to markups. I find that estimating demand flexibly has a significant impact on the results relative to standard mixed logit models, I highlight how the outcomes relate to the estimated shape of the demand functions, and show how the nonparametric approach may be used to guide the choice among parametric specifications.

Keywords: Nonparametric demand estimation, Incomplete tax passthrough, Multi-product firm

JEL codes: L1, L66

*I am very grateful to Steven Berry, Philip Haile, Penny Goldberg and Xiaohong Chen for their guidance and support, as well as Daniel Ackerberg, Nikhil Agarwal, Judith Chevalier, Jan De Loecker, Ying Fan, Sharat Ganapati, Mitsuru Igami, Adam Kapor, Yuichi Kitamura, Lorenzo Magnolfi, Camilla Roncoroni, Marcelo Sant'Anna, Paul Scott, and participants of the Yale IO Prospectus Workshop, EARIE 2017, and seminars at MIT, UCL, EIEF, NYU for helpful comments and conversations. The empirical results in this paper were calculated (or derived) based on data from The Nielsen Company (US), LLC and marketing databases provided by the Kilts Center for Marketing Data Center at the University of Chicago Booth School of Business. All errors and omissions are my own.

†email: gcompiani@berkeley.edu

1 Introduction

Many areas of economics study questions that hinge on the shape of the demand functions for given products. Examples include investigating the sources of market power,[1] evaluating the effect of a tax or subsidy,[2] merger analysis,[3] assessing the impact of a new product being introduced into the market,[4] understanding the drivers of the well-documented incomplete pass-through of cost shocks and exchange-rate shocks to downstream prices,[5] and determining whether firms play a game with strategic complements or substitutes.[6] Given a model of supply, the answers to these questions crucially depend on the level, the slope, and often the curvature of the demand functions. Therefore, if the chosen demand model is not rich enough, the results could turn out to be driven by the convenient, but often arbitrary, restrictions embedded in the model, rather than by the true underlying economic forces. This motivates using demand estimation methods that rely on minimal parametric assumptions.

In this paper, I propose a nonparametric approach to estimate demand in differentiated products markets based on aggregate data.[7] In such settings, a standard practice is to posit a random coefficients discrete choice logit model[8] and estimate it using the methodology developed by Berry, Levinsohn, and Pakes (1995) (henceforth BLP). While the approach in BLP accomplishes the crucial goals of generating reasonable substitution patterns and allowing for price endogeneity, it relies on a number of parametric restrictions for tractability. A recent paper by Berry and Haile (2014) (henceforth BH) shows that most of these restrictions are not needed for identification of the demand functions, i.e. that these restrictions are not necessary to uniquely pin down the demand functions in the hypothetical scenario in which the researcher has access to data on the entire relevant population. While this type

[1] E.g., Berry, Levinsohn, and Pakes (1995) and Nevo (2001).

[2] See, e.g., Bulow and Pfleiderer (1983) and Weyl and Fabinger (2013).

[3] E.g., Nevo (2000) and Capps, Dranove, and Satterthwaite (2003).

[4] E.g., Petrin (2002).

[5] E.g., Nakamura and Zerom (2010) and Goldberg and Hellerstein (2013).

[6] See Bulow, Geanakoplos, and Klemperer (1985).

[7] By differentiated products markets, I mean markets in which consumers face a range of options that are differentiated in ways that are both observed and unobserved to the researcher. In the presence of unobserved heterogeneity, all the variables that are chosen by firms after observing consumer preferences—e.g. prices in many models—are endogenous. The need to deal with endogeneity is a key feature of the estimation approach developed in this paper.

[8] Throughout the paper, we use the terms "random coefficients (discrete choice) logit model" and "mixed logit model" interchangeably.

of results constitutes a necessary first step for any estimation approach, a natural question is whether it is possible to translate the identification arguments in BH into an estimation method that can be implemented on the type of (finite) datasets available to economists. To the best of my knowledge, this paper is the first to do so.[9]

There are two main dimensions in which my approach is more flexible than existing methods. First, as mentioned above, I relax most parametric restrictions on the utility function. For instance, one does not need to assume that the idiosyncratic taste shocks or the random coefficients on product characteristics belong to a parametric family of distributions. Instead, I leverage a range of constraints—such as monotonicity of demand in certain variables and properties of the derivatives of demand—that are grounded in economic theory. Second, by directly targeting the demand functions as opposed to the underlying utility parameters, my approach relaxes several assumptions on consumer behavior and preferences that are embedded in BLP-type models. The latter models assume that each consumer picks the product yielding the highest (indirect) utility among all the available options. This implies, among other things, that the goods are substitutes to each other,[10] that consumers are aware of all products and their characteristics,[11] and that each consumer buys at most one unit of a single product.[12] In contrast to this, the approach I develop allows for a much broader range of consumer behaviors and preferences, including complementarities across goods, consumer inattention, and consumer loss aversion.

In practice, I propose approximating the demand functions using Bernstein polynomials, which make it easy to enforce a number of economic constraints in the estimation routine. In order to obtain valid standard errors, I rely on advances in the econometrics literature on

[9]Souza-Rodrigues (2014) proposes a nonparametric estimation approach for a class of models that includes binary demand. However, extension to the case with multiple inside goods does not appear to be trivial.

[10]Gentzkow (2007) develops a parametric demand model that allows for complementarities across goods and applies it to the market for news. Given the relatively small number of options available to consumers, pursuing a nonparametric approach seems feasible in this industry and we view this as a promising avenue for future research.

[11]Goeree (2008) uses a combination of market-level and micro data to estimate a BLP-type model where consumers are allowed to ignore some of the available products. The model specifies the inattention probability as a parametric function of advertising and other variables. Relative to Goeree (2008), this paper allows for more general forms on inattention. Specifically, any model that satisfies the connected substitutes conditions in Assumption 2 (as well as the index restriction in Assumption 1) is permitted. Section 4.2 presents simulation results from one such model. A recent paper by Abaluck and Adams (2017) obtains identification of both utility and consideration probabilities in a class of models with inattentive consumers facing exogenous prices.

[12]A few studies, including Hendel (1999) and Dubé (2004), estimate models of "multiple discreteness", where agents buy multiple units of multiple products. However, these papers typically rely on individual-level data rather than aggregate data. The same applies to papers that model discrete/continuous choices, such as Dubin and McFadden (1984) for the case of electric appliances and electricity.

nonparametric instrumental variables regression—specifically Chen and Pouzo (2015) and Chen and Christensen (forthcoming) (henceforth, CC)—and I provide primitive conditions for the case where the objects of interest are price elasticities and counterfactual prices.

A typical limitation of nonparametric estimators is that the number of parameters tends to increase quickly with the number of goods and/or covariates. While my approach is no exception, I show that one can mitigate this curse of dimensionality by assuming an exchangeability restriction that is easy to impose in estimation.

I illustrate the applicability and the feasibility of the approach through various simulations in Section 4. The results suggest that the method works well in moderate samples and that it is able to capture the shape of the own- and cross-price elasticities in a number of settings where standard random coefficients logit models fail.

I then apply the approach to data on strawberry sales from California grocery stores. While this is a small product category, it has features that make it especially suitable for the first application of a nonparametric—and thus data-intensive—method. Advantages include the ability to abstract from dynamic considerations and the availability of a large dataset relative to the number of products. I perform two counterfactual exercises on this data. The first is to quantify the pass-through of a tax into retail prices. The second concerns the role played by the multi-product nature of retailers in driving up markups (the "portfolio effect" in the terminology of Nevo (2001)). I compare the results obtained using my approach to those given by a standard random coefficients logit model and find substantial differences. Notably, mixed logit significantly over-estimates the pass-through of a per-unit tax on organic strawberries relative to the nonparametric approach. This reflects the fact that the mixed logit own-price elasticity grows (in absolute value) with price more slowly than the nonparametric elasticity. A retailer finds it optimal to increase prices by a greater extent when faced with a more slowly-growing elasticity function. On the other hand, in the portfolio effect counterfactual, a flexible enough mixed logit model matches the nonparametric results very closely. This is not the case for less flexible mixed logit models, suggesting that the approach proposed in the paper may be used to guide the choice among competing parametric models.

This paper contributes to the vast literature on models of demand in differentiated products markets pioneered by BLP. As mentioned above, the estimation approach I propose builds on the recent nonparametric identification results in BH. We emphasize that the present paper, as well as BH, focus on the case where the researcher has access to market-level data, typically in the form of shares or quantities, prices, product characteristics and other market-level covariates. This is in contrast to studies that are based on consumer-level data, such as Goldberg (1995) and Berry, Levinsohn, and Pakes (2004). Recently, Berry and Haile (2010) have provided conditions for the nonparametric identification in this "micro data" setting, which opens up an interesting avenue for future research on nonparametric estimation of those models.

Second, the paper is related to the large literature on incomplete pass-through[13] and, particularly, the papers that adopt a structural approach to decompose the different sources of incompleteness. For instance, Goldberg and Hellerstein (2008), Nakamura and Zerom (2010) and Goldberg and Hellerstein (2013) estimate BLP-type models to assess the contribution of markup adjustment in generating incomplete pass-through[14] against competing explanations (i.e. nominal rigidities and the presence of costs not affected by the shocks). The present paper contributes to that literature by providing a method to evaluate markups that relaxes a number of restrictions on consumer behavior and preferences. The results I obtain suggest that estimating demand flexibly may decrease or increase the estimates of markup adjustment depending on the product. Notably, for the fresh organic strawberry market, the estimated markup adjustment is significantly higher and thus the tax pass-through is more incomplete relative to what is predicted by a more restrictive model.[15] [16]

Third, the paper relates to the literature investigating the sources of market power, notably Nevo (2001). Once again, I offer a more flexible method to disentangle different components of market power, and I quantify the role of one such component, the portfolio effect, in the California market for fresh strawberries.

The rest of the paper is organized as follows. Section 2 presents the general model and summarizes the nonparametric identification results from BH. Section 3 discusses the proposed nonparametric estimation approach, with a special emphasis on how to impose a range of constraints from economic theory. Section 4 presents the results of several Monte Carlo simulations. Section 5 applies the methodology to data from California grocery stores to assess the pass-through of a tax and evaluate the effect of multi-product retailers on markups. Section 6 concludes the main text. Appendix A provides some details on Bernstein polynomials. Appendix B discusses several economic constraints and shows how to enforce them

[13]The literature on estimating pass-through is large and we do not attempt to provide an exhaustive list of references. We just mention an interesting recent paper by Atkin and Donaldson (2015) which estimates the pass-through of wholesale prices into retail prices, and uses this to quantify how the gains from falling international trade barriers vary geographically within developing countries.

[14]Specifically, Goldberg and Hellerstein (2008) and Goldberg and Hellerstein (2013) focus on exchange rate pass-through, while Nakamura and Zerom (2010) consider cost pass-through.

[15]This comparison is performed assuming a Bertrand-Nash model for the supply side. Note that alternative supply models (e.g. Cournot) tend to deliver lower levels of pass-through. However, they do not appear to be realistic in many industries, including the one we consider in this paper.

[16]For non-organic strawberries, I find that mixed logit over-estimates markup adjustment—and thus under-estimates pass-through—relative to the nonparametric approach, but the two confidence intervals overlap.

in estimation. Appendix C contains all the assumptions and proofs for the inference results. Appendix D presents the results of additional Monte Carlo simulations. Appendix E discusses the construction of the data and contains descriptive statistics. Appendix F shows that some quantities of interest are robust to certain violations of the exogeneity restrictions maintained throughout the paper. Finally, Appendix G provides two possible micro-foundations for the demand model estimated in the empirical application.

2 Model and Identification

The general model I consider is the same as that in BH. In this section, I summarize the main features of the model as well as the key identification result. In a given market t, there is a continuum of consumers choosing from the set $\mathcal{J} \equiv \{1, ..., J\}$. Each market t is defined by the choice set $\mathcal{J}$ and by a collection of characteristics χ_t specific to the market and/or products. The set χ_t is partitioned as follows:

$$\chi_t \equiv (x_t, p_t, \xi_t),$$

where x_t is a vector of exogenous observable characteristics (e.g. exogenous product characteristics or market-level income), $p_t \equiv (p_{1t}, ..., p_{Jt})$ are observable endogenous characteristics (typically, market prices) and $\xi_t \equiv (\xi_{1t}, ..., \xi_{Jt})$ represent unobservables potentially correlated with p_t (e.g. unobserved product quality). Let $\mathcal{X}$ denote the support of χ_t.

Next, I define the structural demand system

$$\sigma : \mathcal{X} \to \Delta^J,$$

where Δ^J is the unit $J-$simplex. The function σ gives, for every market t, the vector s_t of shares for the J goods. We emphasize that this formulation of the model is general enough to allow for different interpretations of shares. The vector s_t could simply be the vector of choice probabilities (market shares) for the inside goods in a standard discrete choice model. However, s_t could also represent a vector of "artificial shares," e.g. a transformation of the vector of quantities sold in the market to the unit simplex. This case arises whenever the data does not come in the form of shares, but quantities, and the researcher does not want to take a stand on what the market size is. Indeed, the empirical application in Section 5 fits into this framework. I also define

$$\sigma_0(\chi_t) \equiv 1 - \sum_{j=1}^{J} \sigma_j(\chi_t),$$

for every market t, where $\sigma_j(\chi_t)$ is the $j-$th element of $\sigma(\chi_t)$. In a standard discrete choice setting, σ_0 corresponds to the share of the outside option, but again this interpretation is not required for the results stated below.

Following BH, I impose two key conditions on the share functions σ that ensure that demand is invertible.[17] The first is an index restriction.

Assumption 1. *Let the exogenous observables be partitioned as* $x_t = \left(x_t^{(1)}, x_t^{(2)}\right)$*, where* $x_t^{(1)} \equiv \left(x_{1t}^{(1)}, ..., x_{Jt}^{(1)}\right)$*,* $x_{jt}^{(1)} \in \mathbb{R}$ *for* $j \in \mathcal{J}$*, and define the linear indices*

$$\delta_{jt} = x_{jt}^{(1)}\beta_j + \xi_{jt}, \quad j = 1, ..., J$$

For every market t*,*

$$\sigma(\chi_t) = \sigma\left(\delta_t, p_t, x_t^{(2)}\right)$$

where $\delta_t \equiv (\delta_{1t}, ..., \delta_{Jt})$*.*

Assumption 1 requires that, for $j = 1, ..., J$, $x_{jt}^{(1)}$ and ξ_{jt} affect consumer choice only through the linear index δ_{jt}. In other words, $x_{jt}^{(1)}$ and ξ_{jt} are assumed to be perfect substitutes. On the other hand, $x_t^{(2)}$ is allowed to enter the share function in an unrestricted fashion. We emphasize that Assumption 1 is stronger than what is needed for identification of the demand system. Specifically, as shown in Appendix B of BH, both the linearity of δ_{jt} in $x_{jt}^{(1)}$ and its separability in the unobservable ξ_{jt} can be relaxed.[18] However, this stronger assumption simplifies the estimation procedure in that it leads to a *separable* nonparametric regression model. Given that this is the first attempt at estimating demand nonparametrically for this class of models, maintaining Assumption 1 appears to be a reasonable compromise.[19]

The second condition is what BH call "connected substitutes assumption."[20]

---

[17] Here I present the identification conditions for a generic demand system. More primitive sufficient conditions tailored to our empirical setting are given in Appendix G.

[18] What is critical turns out to be the strict monotonicity in ξ_{jt}.

[19] In the absence of separability of δ_{jt} in ξ_{jt}, one could think of applying existing estimation approaches for nonseparable regression models with endogeneity (e.g. Chernozhukov and Hansen (2006), Chen and Pouzo (2009), Chen and Pouzo (2012) and Chen and Pouzo (2015)).

[20] Here I use a strengthened version of this assumption that also requires differentiability of σ (see Assumption 3* in Berry, Gandhi, and Haile (2013)). While the stronger version is not needed for identification, it will be helpful because it implies properties of the Jacobian matrix of σ that can be imposed in estimation. Further, note that, unlike Assumption 2 in BH, I only require σ to satisfy the connected substitutes assumption in δ, but not in p. This is because connected substitutes in δ is sufficient for identification of the demand system. BH use connected substitutes in p to show that one can discriminate between different oligopoly models on the supply side, but I do not pursue this here.

Assumption 2. *(i)* $\frac{\partial}{\partial \delta_{jt}}\sigma_k\left(\delta_t, p_t, x_t^{(2)}\right) \leq 0$ *for all* $j > 0, k \neq j$ *and all* $\left(\delta_t, p_t, x_t^{(2)}\right)$*; (ii) for every* $\mathcal{K} \subseteq \mathcal{J}$ *and every* $\left(\delta_t, p_t, x_t^{(2)}\right)$*, there exist a* $k \in \mathcal{K}$ *and a* $j \notin \mathcal{K}$ *such that* $\frac{\partial}{\partial \delta_{kt}}\sigma_j\left(\delta_t, p_t, x_t^{(2)}\right) < 0$.

Assumption 2 requires the goods to be weak gross substitutes in δ_t in terms of the demand system σ; in addition, part (ii) requires some degree of strict substitutability. We emphasize that, because the model does not require that s_t be interpreted as a vector of market shares, Assumptions 1 and 2 must only hold under *some* transformation of the demand system. As a result, although Assumption 2 seems to rule out complementary goods, the model does accommodate some forms of complementarities, as illustrated in Section 4.3.

By Theorem 1 in Berry, Gandhi, and Haile (2013), Assumptions 1 and 2 ensure that demand is invertible, i.e. that, for any triplet of vectors s_t, p_t and $x_t^{(2)}$, there exists at most one vector δ_t such that $s_t = \sigma\left(\delta_t, p_t, x_t^{(2)}\right)$. This means that we can write

$$\delta_{jt} = \sigma_j^{-1}\left(s_t, p_t, x_t^{(2)}\right), \quad j = 1, ..., J. \tag{1}$$

To obtain identification, I impose two additional instrumental variable (IV) restrictions from BH.

Assumption 3. $\mathbb{E}\left(\xi_j | X, Z\right) = 0$ *a.s.-*(X, Z)*, for* $j \in \mathcal{J}$ *and for a random vector* $Z = (Z_1, ..., Z_J)$ *of instruments.*

Assumption 4. *For all functions* $B\left(\cdot,\cdot,\cdot\right)$ *with finite expectation, if* $\mathbb{E}\left(B\left(S, P, X^{(2)}\right) | X, Z\right) = 0$ *a.s.-*(X, Z)*, then* $B\left(S, P, X^{(2)}\right) = 0$ *a.s.-*$(S, P, X^{(2)})$.

Assumption 3 imposes exogeneity of the observed characteristics x_t. In addition, it requires a vector of (exogenous) instruments z_t excluded from the share function (e.g. cost shifters). Intuitively, the vector $x_t^{(1)}$ serves as "included" instruments for the market shares s_t, while z_t is instrumenting for prices p_t. Assumption 4 constitutes a completeness condition on the joint distribution of (s_t, p_t, x_t, z_t) with respect to (s_t, p_t). In words, it requires the exogenous variables (x_t, z_t) to shift the distribution of the endogenous variables (s_t, p_t) to a sufficient extent. It is a nonparametric analog of the standard rank condition in linear IV models.[21]

BH show that, under the maintained assumptions, the demand system is identified, which

[21] BH also show that identification of the demand model can be achieved without the completeness condition under additional structure. I do not pursue this here.

we formalize in the following result.

Theorem 1. *Under Assumptions 1, 2, 3 and 4, the structural demand system σ is point-identified.*

Theorem 1 implies that the own-price and cross-price elasticity functions are identified. Therefore, in combination with a model of supply, it allows one to address a wide range of economic questions, including evaluation of markups, predicting equilibrium responses to a policy (e.g. a tax), and testing hypotheses on consumer preferences or behavior (e.g. testing the presence of income effects).[22]

3 Nonparametric Estimation

3.1 Setup and asymptotic results

Our proposed estimation approach is based on equation (1). Consistent with the goal of avoiding arbitrary functional form and distributional restrictions, we propose a nonparametric approach. Specifically, we rewrite (1) as

$$x_{jt}^{(1)} = \sigma_j^{-1}\left(s_t, p_t, x_t^{(2)}\right) - \xi_{jt} \quad j = 1, ..., J, \tag{2}$$

where we use the normalization $\beta_j = 1$ as in BH. Equation (2), coupled with the IV exogeneity restriction in Assumption 3

$$\mathbb{E}\left(\xi_j | X, Z\right) = 0 \ \ a.s. - (X, Z), \quad j \in \mathcal{J}$$

suggests estimating σ_j^{-1} using nonparametric instrumental variables (NPIV) methods.[23]

Some additional notation is needed to formalize this. We denote by T the sample size, i.e. the number of markets in the data. Let Σ be the space of functions to which σ^{-1} belongs and let $\psi_{M_j}^{(j)}(\cdot) \equiv \left(\psi_{1,M_j}^{(j)}(\cdot), ..., \psi_{M_j,M_j}^{(j)}(\cdot)\right)'$ be the basis functions used to approximate σ_j^{-1} for

[22] As pointed out by BH (Section 4.2), one important exception is evaluation of individual consumer welfare, which can be performed with aggregate data only by committing to a parametric functional form for utility. However, identification of the demand system σ is sufficient to pin down some welfare measures, such as the aggregate change in consumer surplus due to a change in prices.

[23] The literature on NPIV methods is vast and we refer the reader to recent surveys, such as Horowitz (2011) and Chen and Qiu (2016).

$j \in \mathcal{J}$.[24] Note that, since we pursue a nonparametric approach, we will let M_j go to infinity with the sample size for all j; thus, although we suppress it in the notation, M_j is a function of T. Then, we let $\Sigma_T \equiv \left\{ \left(\tilde{\sigma}_1^{-1}, ..., \tilde{\sigma}_J^{-1}\right) : \tilde{\sigma}_j^{-1} = \pi_j' \psi_{M_j}^{(j)}(\cdot),\ j \in \mathcal{J} \right\}$ be the sieve space for Σ and note that Σ_T depends on the sample size through $\{M_j\}_{j \in \mathcal{J}}$. Next, we denote by $a_{K_j}^{(j)}(\cdot) \equiv \left(a_{1,K_j}^{(j)}(\cdot), ..., a_{K_j,K_j}^{(j)}(\cdot)\right)'$ the basis functions used to approximate the instrument space for good j's equation, and we let $A_{(j)} = \left(a_{K_j}^{(j)}(x_1, z_1), ..., a_{K_j}^{(j)}(x_T, z_T)\right)'$ for $j \in \mathcal{J}$. Again, we suppress the dependence of $\{K_j\}_{j \in \mathcal{J}}$ on the sample size. Further, we require that $K_j \geq M_j$ for all j, which corresponds to the usual requirement in parametric instrumental variable models that the number of instruments be at least as large as the number of endogenous variables. Finally, we let $r_{jt}\left(s_t, p_t, x_t, z_t; \tilde{\sigma}_j^{-1}\right) \equiv \left(x_{jt}^{(1)} - \tilde{\sigma}_j^{-1}\left(s_t, p_t, x_t^{(2)}\right)\right) \times a_{K_j}^{(j)}(x_t, z_t)$. Then, the estimation problem is as follows[25]

$$\min_{\tilde{\sigma}^{-1} \in \Sigma_T} \sum_{j=1}^{J} \left[\sum_{t=1}^{T} r_{jt}\left(s_t, p_t, x_t, z_t; \tilde{\sigma}_j^{-1}\right)\right]' \left(A_j' A_j\right)^{-} \left[\sum_{t=1}^{T} r_{jt}\left(s_t, p_t, x_t, z_t; \tilde{\sigma}_j^{-1}\right)\right], \tag{3}$$

The solution $\hat{\sigma}^{-1}$ to (3) minimizes a quadratic form in the terms $\{r_{jt}(\cdot), j \in \mathcal{J}, t = 1, ..., T\}$, i.e. the implied regression residuals interacted with the instruments. Note that the objective function in (3) is convex. Thus, if the set Σ_T is also convex, readily available algorithms are guaranteed to converge to the global minimizer.

Moreover, we can leverage recent advances in the NPIV literature to perform inference. Specifically, I rely on results in Chen and Christensen (forthcoming) (henceforth, CC) to obtain asymptotically valid standard errors for functionals of the demand system.[26] I now state a result for general scalar functionals, which is a slight modification of Theorem D.1 in CC.[27] For conciseness, the regularity conditions needed for the result, the definition of the estimator for the variance of the functional, and the proof of the theorem, are postponed to

[24] In the simulations of Section 4, as well as in the application of Section 5, I use Bernstein polynomials to approximate each of the unknown functions. However, the inference result in Theorem 2 below does not depend on this choice, hence the general notation used in the first part of this section.

[25] In equation (3), we let $\tilde{A}^{-}$ denote the Moore-Penrose inverse of a matrix $\tilde{A}$.

[26] See also Chen and Pouzo (2015).

[27] Note that CC consider inference on functionals of an *unconstrained* sieve estimator of the unknown function, whereas our model features a range of economic constraints. However, under the assumption that the true demand functions satisfy the inequality constraints strictly, the restrictions will not be binding asymptotically. On a related note, one may wonder whether the asymptotic standard errors derived here provide a good approximation to the finite sample performance of the constrained estimator. Comparing the Monte Carlo evidence in Section 4 to the asymptotic standard errors in the empirical application from Section 5, it seems that the asymptotic standard errors might be conservative (in spite of leveraging a much bigger sample, the asymptotic standard errors for the elasticity functionals in the application are slightly larger than those in the Monte Carlo simulations).

Appendix C.

Theorem 2. *Let the assumptions of Theorem 1 hold. Let f be a scalar functional of the demand system and $\hat{v}_T(f)$ be the estimator of the standard deviation of $f(\hat{\sigma}^{-1})$ defined in* (13) *in Appendix C. In addition, let Assumptions 5, 6, 7 and 8 in Appendix C hold. Then,*

$$\sqrt{T}\frac{\left(f\left(\hat{\sigma}^{-1}\right)-f\left(\sigma^{-1}\right)\right)}{\hat{v}_T(f)}\xrightarrow{d}N(0,1).$$

Proof. See Appendix C. □

Theorem 2 yields asymptotically valid standard errors for scalar functionals of the demand system. These, in turn, may be used to construct confidence intervals for the functionals and to test hypothesis on the demand functions.

We now provide more primitive sufficient conditions for Theorem 2 for two functionals of interest: price elasticities and equilibrium prices. In both cases, we assume $J = 2$, which corresponds to the model estimated in the empirical application of Section 5. Further, for simplicity, we focus on the case where there are no additional exogenous covariates $x^{(2)}$ in the inverse of the demand system (2), although $x^{(2)}$ could be included following Section 3.3 of CC. We state the results here and postpone the full presentation of the assumptions, as well as the proof, to Appendix C

Theorem 3. *Let the assumptions of Theorem 1 hold. Let f_ϵ be the own-price elasticity functional defined in* (24) *in Appendix C, let $\hat{v}_T(f_\epsilon)$ denote the estimator of the standard deviation of $f_\epsilon(\hat{\sigma}^{-1})$ based on* (13)*, and let Assumptions 5, 6(iii), 7 and 9 from Appendix C hold. Then,*

$$\sqrt{T}\frac{\left(f_\epsilon\left(\hat{\sigma}^{-1}\right)-f_\epsilon\left(\sigma^{-1}\right)\right)}{\hat{v}_T(f_\epsilon)}\xrightarrow{d}N(0,1).$$

Proof. See Appendix C. □

Theorem 3 establishes the asymptotic distribution of the own-price elasticity for good 1. An analogous argument holds for the own-price elasticity of good 2 and for the cross-price elasticities.

Next, we state a result establishing the asymptotic distribution of the equilibrium price for good 1. Again, the case of good 2 follows immediately.

Theorem 4. *Let the assumptions of Theorem 1 hold. Let f_{p_1} be the equilibrium price functional defined in* (26) *in Appendix C, let $\hat{v}_T(f_{p_1})$ denote the estimator of the standard devi-*

ation of $f_{p_1}(\hat{\sigma}^{-1})$ based on (13)*, and let Assumptions 5, 6(iii), 7 and 10 from Appendix C hold. Then,*

$$\sqrt{T}\frac{\left(f_{p_1}\left(\hat{\sigma}^{-1}\right)-f_{p_1}\left(\sigma^{-1}\right)\right)}{\hat{v}_T\left(f_{p_1}\right)} \xrightarrow{d} N\left(0,1\right).$$

Proof. See Appendix C. □

In Section 5, we apply Theorem 4 to obtain confidence intervals for equilibrium prices under two counterfactual scenarios, i.e. the levying of a tax and the switch from monopoly to duopoly.

3.2 Constraints

We conclude this section with a discussion of the curse of dimensionality that is inherent in nonparametric estimation, and of ways to tackle the issue. Note that each of the unknown functions σ_j^{-1} has $2J+n_{x^{(2)}}$ arguments, where $n_{x^{(2)}}$ denotes the number of variables included in $x^{(2)}$. Therefore, the number of parameters to estimate grows quickly with the number of goods and/or the number of characteristics included in $x^{(2)}$, and it will typically be much larger than in conventional parametric models (in the hundreds or even thousands). While this is an objective limitation of the approach, we argue that there are a number of factors alleviating the problem. First, as illustrated by the extensive simulations in Section 4, the own- and cross-price elasticity functions are precisely estimated with moderate sample sizes (3,000 observations) for the $J=2$ case. Given that many economic questions hinge on functionals of the elasticities, this suggests that it is possible to obtain informative confidence intervals for several quantities of interest. Second, very large market-level data sets are increasingly available to researchers, including the Nielsen scanner data which is used in Section 5. This provides further reassurance on the viability of data-intensive nonparametric methods. Lastly, we show below that imposing constraints from economic theory substantially aids nonparametric estimation. Some constraints (e.g. exchangeability) directly reduce the number of parameters to be estimated. This simplification is often dramatic, especially as the number of goods increases. Other restrictions (e.g. monotonicity) do not affect the number of parameters, but play a role in disciplining the estimation routine, as illustrated in Section 4.

Imposing constraints in model (2) is complicated by the fact that economic theory gives us restrictions on the demand system σ, but what is targeted by the estimation routine is σ^{-1}. Therefore, one contribution of the paper is to translate constraints on the demand system σ into constraints on its inverse σ^{-1}, and show that the latter can be enforced in a computationally feasible way. Specifically, we propose to estimate the functions σ_j^{-1} in (2) using Bernstein polynomials, which we find to be very convenient for imposing economic restrictions.[28]

In this paper, we consider a number of constraints, including exchangeability of the demand functions, monotonicity and lack of income effects. We emphasize that this is not an exhaustive list, and one may wish to impose additional constraints in a given application.[29] Conversely, not all constraints discussed in this paper need to be enforced simultaneously in order to make the approach feasible.[30]

In the remainder of this section, we focus on an exchangeability constraint, which leads to a dramatic reduction in the number of parameters to estimate and thus is very helpful in tackling the curse of dimensionality.[31] In Appendix B, we consider other constraints that one might be willing to impose and show how to enforce them in estimation. In order to define exchangeability, let $\pi : \{1, ..., J\} \to \{1, ..., J\}$ be any permutation and let $x^{(2)} = \left(x_1^{(2)}, ..., x_J^{(2)}\right)$, i.e. we assume that $x^{(2)}$ is a vector of product-specific characteristics.[32] Then, we say the structural demand system σ is exchangeable if

$$\sigma_j \left(\delta, p, x^{(2)}\right) = \sigma_{\pi(j)} \left(\delta_{\pi(1)}, ..., \delta_{\pi(J)}, p_{\pi(1)}, ..., p_{\pi(J)}, x_{\pi(1)}^{(2)}, ..., x_{\pi(J)}^{(2)}\right), \tag{4}$$

for $j = 1, ..., J$. In words, the structural demand system is exchangeable if the demand func-

[28]See Appendix A for a more formal discussion of Bernstein polynomials.

[29]Note that one restriction that could yield a substantial reduction in the number of parameters is the constraint that $x^{(2)}$ enter the demand functions through the indices δ. Specifically, each demand function goes from having $2J + n_{x^{(2)}}$ to $2J$ arguments. This, in turn, means that the number of Bernstein coefficients for each demand function goes from $m^{2J+n_{x^{(2)}}}$ to m^{2J}, where for simplicity we assume the degree m of the polynomials is the same for all arguments. Restricting the way in which characteristics enter the demand function is typically hard to motivate on economic grounds and therefore this type of constraints is somewhat at odds with the general spirit of the paper. However, such restrictions might constitute an appealing compromise in settings where the number of characteristics and/or goods is relatively high and dimension reduction becomes a necessity.

[30]For example, in the empirical application in Section 5, we do not assume lack of income effects.

[31]Exchangeability restrictions were discussed in Pakes (1994) and Berry, Levinsohn, and Pakes (1995).

[32]This need not be the case. For instance, $x^{(2)}$ could be a vector of market-level variables. In these settings, we say the demand system is exchangeable if $\sigma_j \left(\delta, p, x^{(2)}\right) = \sigma_{\pi(j)} \left(\delta_{\pi(1)}, ..., \delta_{\pi(J)}, p_{\pi(1)}, ..., p_{\pi(J)}, x^{(2)}\right)$, which requires $x^{(2)}$ to affect the demand of each good in the same way. Of course, the case where $x^{(2)}$ includes both market-level and product-specific variables can be handled similarly at the cost of additional notation.

tions do not depend on the identity of the products, but only on their attributes $\left(\delta, p, x^{(2)}\right)$.[33] For instance, for $J = 3$, exchangeability implies that $\sigma_1 = \sigma_2 = \sigma_3$, and that

$$\sigma_1\left(\delta_1, \underline{\delta}, \overline{\delta}, p_1, \underline{p}, \overline{p}, x_1^{(2)}, \underline{x}^{(2)}, \overline{x}^{(2)}\right) = \sigma_1\left(\delta_1, \overline{\delta}, \underline{\delta}, p_1, \overline{p}, \underline{p}, x_1^{(2)}, \overline{x}^{(2)}, \underline{x}^{(2)}\right)$$

for all $\left(\delta_1, \overline{\delta}, \underline{\delta}, p_1, \overline{p}, \underline{p}, x_1^{(2)}, \overline{x}^{(2)}, \underline{x}^{(2)}\right)$. One may be willing to impose exchangeability when it seems reasonable to rule out systematic discrepancies between the demands for different products. This assumption is often implicitly made in discrete choice models. For example, in a standard random coefficient logit model without brand fixed-effects, if the distribution of the random coefficients is the same across goods, then exchangeability is satisfied.[34]

Moreover, one may allow for additional flexibility by allowing the intercepts of the δ indices to vary across goods. This preserves the advantages of exchangeability in terms of dimension reduction, which we discuss below, while simultaneously allowing each unobservable to have a different mean. Relative to existing methods, this is no more restrictive than standard random coefficient logit models with brand fixed-effects and the same distribution of random coefficients across goods.

Imposing exchangeability on the demand system σ is facilitated by the following result.

Lemma 1. *If σ is exchangeable, then σ^{-1} is also exchangeable.*

Proof. See Appendix B.3. □

Lemma 1 implies that we can directly impose exchangeability of the target functions σ^{-1}. This can be achieved by simply requiring that the Bernstein coefficients be the same for all goods (up to appropriate rearrangements of the arguments of the functions) and imposing that the value of each function be invariant to certain permutations of its arguments. By the approximation properties of Bernstein polynomials (see Appendix A), the latter may be conveniently enforced through *linear* restrictions on the Bernstein coefficients.

Exchangeability is especially helpful in that it dramatically reduces the number of parameters to be estimated. To illustrate this point, we show in Table 1 how the number of parameters

[33] For simplicity, here I consider the extreme case of exchangeability across all goods $1, ..., J$. However, one could also think of imposing exchangeability only within a subset of the goods, e.g. the set of goods produced by one company. The arguments in this section would then apply to the subset of products on which the restriction is imposed.

[34] This also uses the fact that the idiosyncratic logit shocks are iid - and thus exchangeable - across goods.

Table 1: Number of parameters with and without exchangeability

J	exchangeability	no exchangeability
3	324	729
4	900	6,561
5	2,025	59,049
10	27,225	3.4bn

Note: Tensor product of univariate Bernstein polynomials of degree 2. $n_{x^{(2)}}$ is assumed to be zero.

grows with J depending on whether we do or do not impose exchangeability.[35] While the number of parameters grows large with J in both cases, the curse of dimensionality is much worse if we do not impose exchangeability—indeed to the point where estimation becomes quickly computationally intractable.

4 Monte Carlo Simulations

In this section, we present the results from a range of Monte Carlo simulations. The goal is twofold. On the one hand, we illustrate that the estimation procedure works well in moderate sample sizes. In doing this, we also highlight the importance of imposing constraints, which motivates the work in Section 3. On the other hand, we show how the general model from Section 2 may be applied to a variety of settings which include—but are not limited to—standard discrete choice. All simulations are for the case with $J = 2$ number of goods, $T = 3,000$ number of markets, and 200 Monte Carlo repetitions. In Appendix D, we present additional simulations to test how robust the results are when the sample size is lower, the number of goods is larger than two, or Assumption 1 is violated.

We compare the performance of the estimated procedure to that of standard methods. Specifically, we take as a benchmark a random coefficient logit model with normal random coefficients. We refer to this model as BLP. In order to summarize the results, we plot the own- and cross-price elasticities as a function of price, along with point-wise confidence bands. We choose to plot elasticity functions as that is what is needed for many counterfactuals of interest. In each plot, all market-level variables different from the own-price are fixed at their median values.

[35]In Table 1, the degree of the Bernstein basis is taken to be 2. Further, each structural demand function σ_j is assumed to depend on the $2J$ arguments (δ, p) only. Of course, the number of parameters would be even higher if one were to use a higher degree for the Bernstein basis and/or include additional covariates.

4.1 Correctly specified BLP model

The first simulation is a random coefficients logit model with normal random coefficients. This means that the BLP procedure is correctly specified and therefore we expect it to perform well. On the other hand, we expect the nonparametric approach to yield larger standard errors, due to the fact that it does not rely on any parametric assumptions. Thus, comparing the relative performance of the two should shed some light on how large a cost one has to pay for not committing to a parametric structure when that happens to be correct. In the simulation, the utility that consumer i derives from good j takes the form

$$u_{ij} = \alpha_i p_j + \beta x_j + \xi_j + \epsilon_{ij}$$

where ϵ_{ij} is independently and identically distributed (iid) extreme value across goods and consumers, α_i is distributed $N(-1, 0.15^2)$ iid across consumers, and we set $\beta = 1$. The exogenous shifters x_j are drawn from a uniform $[0, 2]$ distribution,[36] whereas the unobserved quality indices ξ_j are distributed normally with mean 1 and standard deviation 0.15. We draw excluded instruments z_j from a uniform [0,1] distribution and generate prices according to

$$p_j = 2(z_j + \eta_j) + \xi_j,$$

where η_j is uniform [0,0.1].[37]

Given this data generating process (dgp), we run our proposed estimation procedure. We impose the following constraints from Section 3 and Appendix B: exchangeability, diagonal dominance of $\mathbb{J}^{\delta}_{\sigma}$ and monotonicity of σ^{-1}. We then compare our results with those obtained by applying BLP. Figures 1 and 2 show the own- and cross-price elasticity functions for good 1, respectively, together with confidence bands for BLP and our proposed procedure (henceforth shortened as NPD, for "Non-Parametric Demand"). Both the NPD and the BLP confidence bands contain the true elasticity functions. As expected, the NPD confidence band is larger than the BLP one for the cross-price elasticity; however, they are still informative. On the other hand, the NPD and the BLP confidence bands for the own-price elasticity appear to be comparable. Overall, we take this as suggestive that the penalty one pays when ignoring correct parametric assumptions is not substantial.

---

[36]Note that we drop the superscript on x_j, since in the simulations we only have one scalar exogenous shifter for each good, i.e. there is no $x^{(2)}$. This applies to all the simulations in this section.

[37]Note that, while for simplicity we do not generate the prices from the supply first order conditions, the definition of prices above is such that they are positively correlated with both the excluded instruments (consistent with their interpretation as cost shifters) and the unobserved quality (consistent with what would typically happen in equilibrium).

Figure 1: BLP model: Own-price elasticity function

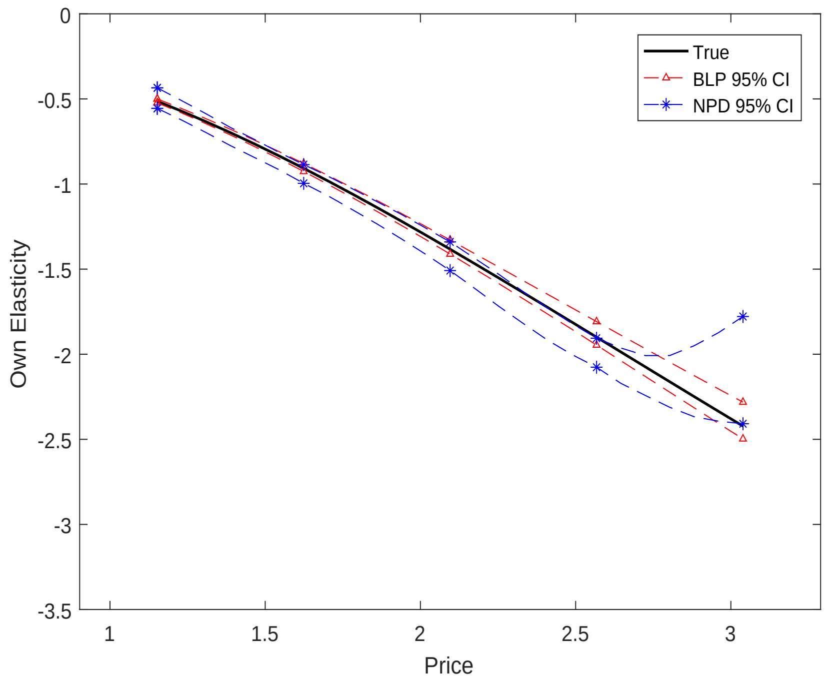
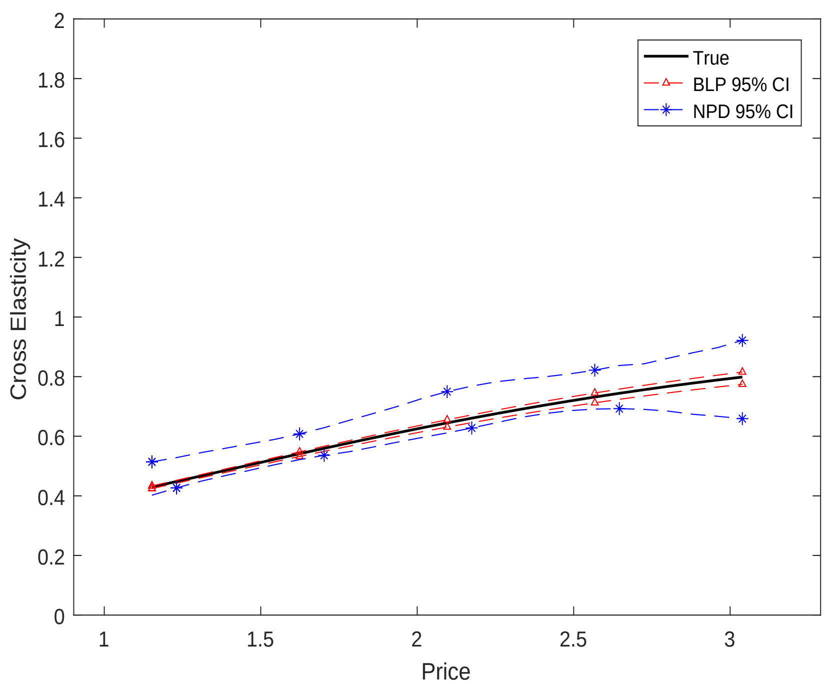


Figure 2: BLP model: Cross-price elasticity function



4.2 Inattention

Next, we consider a discrete choice model with inattention. In any given market, we assume a fraction of consumers ignore good 1 and therefore maximize their utility over good 2 and the outside option only. On the other hand, the remaining consumers consider all goods. We take the fraction of inattentive consumers to be $1 - \Phi(3 - p_1)$, where Φ is the standard normal cdf. This implies that, as the price of good 1 increases, more consumers will ignore good 1, which is consistent with the idea that consumers might pay more attention to cheaper products (e.g. if goods are filtered in on-line shopping or products that are on sale are advertised in supermarkets). Except for the presence of inattentive consumers, the simulation design is the same as in Section 4.1. In estimation, we impose the following constraints: monotonicity of σ^{-1}, and diagonal dominance and symmetry of $\mathbb{J}_{\sigma}^{\delta}$.[38] Note that we do not impose exchangeability, since the demand function for good 1 is now different from that of good 2 due to the presence of inattentive consumers. Accordingly, in the BLP procedure, we allow different constants for the two goods.

Figures 3 and 4 show the results for good 1. The nonparametric method captures the shape of both the own- and the cross-price elasticity functions, whereas BLP is off the mark, underestimating the own-price elasticity and overestimating the cross-price elasticity.

[38]See Appendix B for a discussion of these constraints.

Figure 3: Inattention: Own-price elasticity function

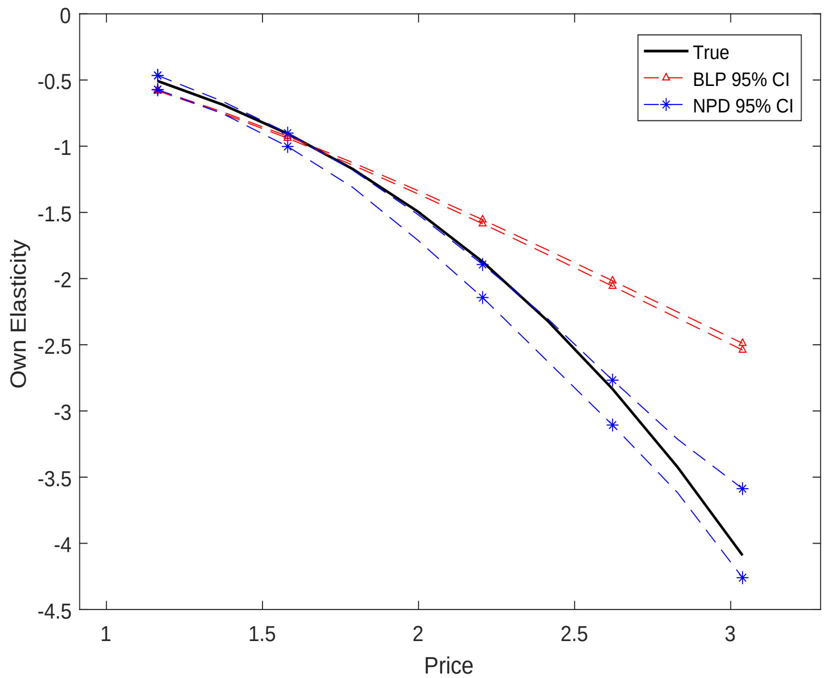
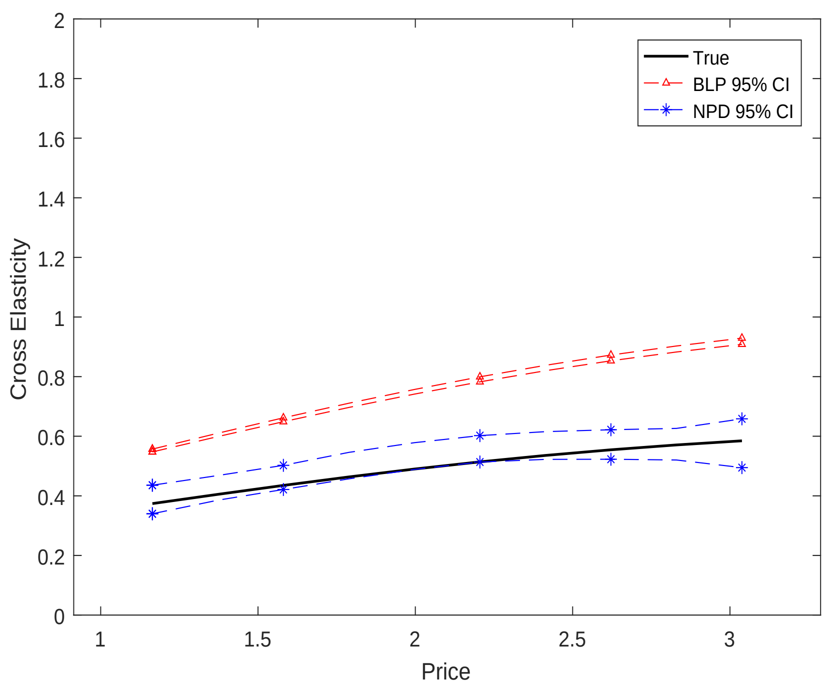


Figure 4: Inattention: Cross-price elasticity function



4.3 Complementary goods

We now consider a model where good 1 and 2 are not substitutes, but complements. We generate the exogenous covariates and prices as in the previous two simulations,[39] but we now let market *quantities* be as follows

$$q_j(\delta, p) \equiv 10\frac{\delta_j}{p_j^2 p_k} \quad j = 1, 2; \quad k \neq j.$$

Note that q_j decreases with p_k and thus the two goods are complements. Now define the function σ_j as

$$\sigma_j(\delta, p) = \frac{q_j(\delta, p)}{1 + q_1(\delta, p) + q_2(\delta, p)}$$

Unlike in standard discrete choice settings, here σ_j does not correspond to the market share function of good j. Instead, it is simply a transformation of the quantities yielding a demand system that satisfies the connected substitutes assumption.[40] In the NPD estimation, we impose the following constraints: monotonicity of σ^{-1}, diagonal dominance of $\mathbb{J}_\sigma^\delta$ and exchangeability.[41]

Figures 5 and 6 show the results for good 1. Again, NPD captures the shape of the elasticity functions well. Specifically, note that the cross-price elasticity is slightly negative given that good 1 and good 2 are complements. On the other hand, the BLP confidence bands are mostly off target, consistent with the fact that a discrete choice model is not well-suited for markets with complementarities. In particular, BLP largely over-estimates the magnitude of the own-price elasticity and forces the cross-price elasticity to be positive.

---

[39] One difference is that we now take the mean of ξ_1 and ξ_2 to be 2 instead of 1 in order to obtain shares that are not too close to zero.

[40] See also Example 1 in Berry, Gandhi, and Haile (2013).

[41] See Section 3.2 and Appendix B for a discussion of these constraints.

Figure 5: Complements: Own-price elasticity function

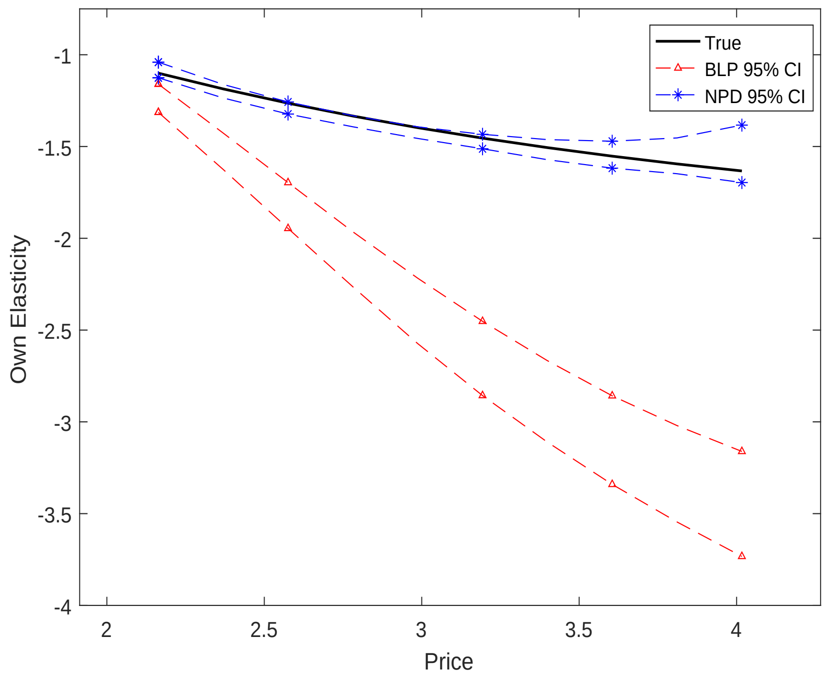
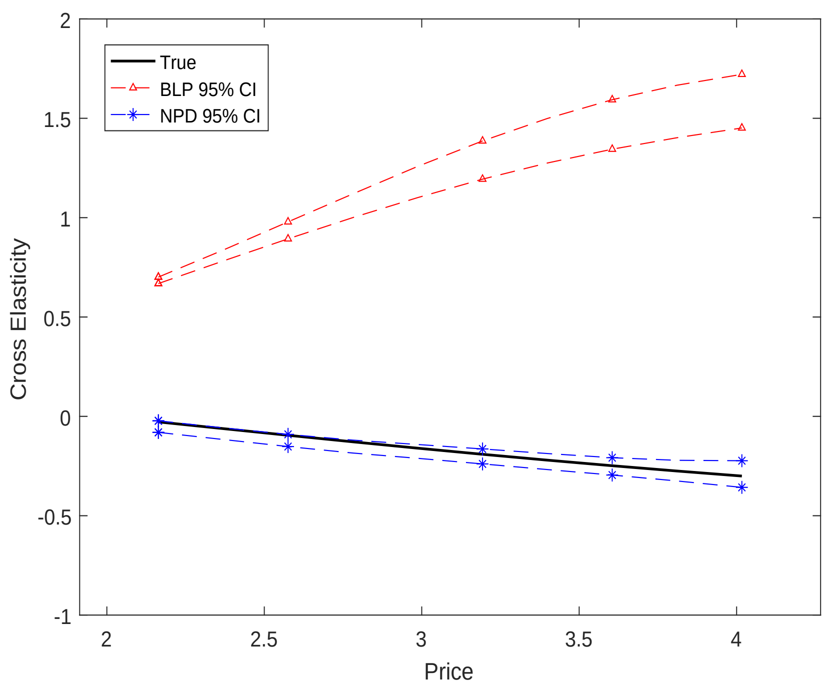


Figure 6: Complements: Cross-price elasticity function



5 Application to Tax Pass-Through and Multi-Product Firm Pricing

In this section, we apply the proposed nonparametric procedure to perform two counterfactual exercises using data from California grocery stores. The first is to quantify the pass-through of a tax into retail prices. It is well-known that the extent to which a tax is passed through to consumers hinges on the curvature of demand.[42] Therefore, flexibly capturing the shape of the demand function is crucial to accurately assessing the effect of a tax on prices and quantities, which motivates pursuing a nonparametric approach.

The second counterfactual concerns the role played by the multi-product nature of retailers in driving up markups. Specifically, a firm simultaneously pricing multiple goods is able to internalize the competition that would occur if those goods were sold by different firms, which pushes prices upwards.[43] Quantifying the magnitude of this effect is ultimately an empirical question which again depends on the shape of the demand functions.

5.1 Data

We use data on sales of fresh fruit at stores in California. Specifically, we focus on strawberries, and look at how consumers choose between organic strawberries, non-organic strawberries and other fresh fruit, which we pool together as the outside option. While this is a small product category, it has a few features that make it especially suitable for a clean comparison between different static demand estimation methods. First, given the high perishability of fresh fruit, we may reasonably abstract from dynamic considerations on both the demand and the supply side. Strawberries, in particular, belong to the category of non-climacteric fruits,[44] which means that they cannot be artificially ripened using ethylene.[45] This limits the ability of retailers as well as consumers to stockpile and further motivates ignoring dynamic considerations in the model. Second, while strawberries are harvested in California essentially year-round, other fruits—e.g. peaches—are not, which provides some arguably exogenous supply-side variation in the richness of the outside option relative to the inside

[42] See, e.g., Weyl and Fabinger (2013).

[43] This is one of the determinants of markups considered by Nevo (2001) in his analysis of the ready-to-eat cereal industry. On the other hand, when multiple products are sold by the same firm, prices might be driven down by, among other things, economies of scale and loss-leader pricing (see, e.g., Lal and Matutes (1994), Lal and Villas-Boas (1998) and Chevalier, Kashyap, and Rossi (2003)).

[44] See, e.g., Knee (2002).

[45] Unlike climacteric fruits, such as bananas.

goods. Finally, the large number of store/week observations combined with the limited number of goods provide an ideal setting for the first application of a nonparametric—and thus data-intensive—estimation approach.

We allow for price endogeneity by using Hausman IVs. In addition, for the inside goods, we also use shipping-point spot prices, as a proxy for the wholesale prices faced by retailers. Besides prices, we include the following shifters in the demand functions: (i) a proxy for the availability of non-strawberry fruits in any given week; (ii) a measure of consumer tastes for organic produce in any given store; and (iii) income.

Appendix E provides further details on the construction of the dataset, as well as some summary statistics and results for the first-stage regressions.

5.2 Model

Let 0,1 and 2 denote non-strawberry fresh fruit, non-organic strawberries and organic strawberries, respectively. We take the following model to the data

$$\begin{aligned} s_1 &= \sigma_1\left(\delta_{str}, \delta_{org}, p_0, p_1, p_2, x^{(2)}\right) \\ s_2 &= \sigma_2\left(\delta_{str}, \delta_{org}, p_0, p_1, p_2, x^{(2)}\right) \\ \delta_{str} &= \beta_{0,str} - \beta_{1,str} x^{(1)}_{str} + \xi_{str} \\ \delta_{org} &= \beta_{0,org} + \beta_{1,org} x^{(1)}_{org} + \xi_{org} \end{aligned} \tag{5}$$

In the display above, s_i denotes the share of product i, defined as the quantity of i divided by the total quantity across the three products, $x^{(1)}_{org}$ denotes the taste for organic products, $x^{(1)}_{str}$ denotes the availability of other fruit, $x^{(2)}$ denotes income, and (ξ_{str}, ξ_{org}) denote unobserved store/week level shocks for strawberries and organic produce, respectively. These unobservables could include, among other things, shocks to the quality of produce at the store/week level,[46] variation in advertising and/or display across stores and time, and taste shocks idiosyncratic to a given store's customer base (possibly varying over time). To the extent that these factors are taken into account by the store when pricing produce, the prices (p_0, p_1, p_2) will be econometrically endogenous. In contrast, we assume that the demand shifters $\left(x^{(1)}_{str}, x^{(1)}_{org}\right)$ are mean independent of (ξ_{str}, ξ_{org}). Regarding $x^{(1)}_{str}$, this is a proxy for the total supply of non-strawberry fruits in California in a given week. As such, we view this

[46]Note that we do not include time dummies. This is motivated by the fact that (i) more than 90% of all strawberries produced in the US are grown in California (United States Department of Agriculture (2017)), and (ii) strawberries are harvested in California essentially year-round. Thus, the assumption that the quality of strawberries *sold* in California does not systematically vary over time seems reasonable to a first order.

as a purely supply-side variable that shifts demand for strawberries inwards by increasing the richness of the outside option,[47] but is independent of store-level shocks.[48] As for $x^{(1)}_{org}$, this is meant to approximate the taste for organic products of a given store's customer base. One plausible violation of exogeneity for this variable would arise if consumers with a stronger preference for organic products (e.g. wealthy consumers) tended to go to stores that sell better-quality organic produce (e.g. Whole Foods). This could induce positive correlation between $x^{(1)}_{org}$ and ξ_{org}. However, we show in Appendix F that many objects of interest, including the counterfactuals in Section 5.4, are robust to certain forms of endogeneity arising through this channel.

We compare the nonparametric approach to a standard parametric model of demand. Specifically, we consider the following mixed logit model:

$$\begin{aligned} u_{i,1} &= \beta_1 + \left(\beta_{p,i} + \beta_{x^{(2)}} x^{(2)}\right) p_1 + \beta_{p,0} p_0 + \beta^{par}_{str} x^{(1)}_{str} + \xi_1 + \epsilon_{i,1} \\ u_{i,2} &= \beta_2 + \left(\beta_{p,i} + \beta_{x^{(2)}} x^{(2)}\right) p_2 + \beta_{p,0} p_0 + \beta^{par}_{str} x^{(1)}_{str} + \beta^{par}_{org} x^{(2)}_{org} + \xi_2 + \epsilon_{i,2} \end{aligned} \tag{6}$$

where $(\epsilon_{i,norg}, \epsilon_{i,org})$ are iid extreme value shocks, (ξ_1, ξ_2) represent unobserved quality of non-organic and organic strawberries, respectively, and the price coefficient $\beta_{p,i}$ is allowed to vary across consumers.[49]

Note one important difference between model (5) and model (6). The latter specifies the indirect utility from each good and thus imposes the implicit (and unrealistic) assumption that each consumer makes a discrete choice between one unit of non-organic strawberries, one unit of organic strawberries, and one unit of other fruits. On the other hand, model (5) allows for a broader range of consumer behaviors, including continuous choice, as we show in Appendix G.2. This is one of the advantages of targeting the structural demand function directly as opposed to the underlying utility parameters.

[47]For example, in the summer many fresh fruits (e.g, Georgia peaches) are in season, which tends to increase the appeal of the outside option relative to strawberries.

[48]The variable $x^{(1)}_{str}$ would be endogenous if the quality of strawberries systematically varied with the harvesting patterns of other fresh fruits. However, as motivated in footnote 46, we abstract from this.

[49]We chose this specification over one where the price coefficient is normally distributed (as in BLP) because we found that it made it much easier to impose non-negativity constraints on the marginal costs. Given that we are imposing such constraints in the nonparametric procedure, we wish to impose them in the mixed logit estimation as well for a fair comparison.

Table 2: Mixed logit estimation results

Variable	Type I	Type II
Price	-7.58	-89.85
	(0.07)	(6.53)
Price×Income	0.89	
	(0.06)	
Price other fruit	8.70	
	(0.23)	
Other fruit	-0.37	
	(0.01)	
Taste for organic	0.08	
	(0.06)	
Fraction of consumers	0.82	0.18
	(0.00)	(0.00)

Note: Model includes product dummies. Asymptotically valid standard errors in parentheses.

5.3 Estimation Results

First, we present the results from the mixed logit model in Table 2. We take the random coefficient on price to have a two-point distribution.[50] Intuitively, this means that we allow consumers to be of two different types depending on their price sensitivity. The last two columns report the coefficients for each of the two types. All coefficients have the expected signs.

We now present the nonparametric estimation results. We impose the constraints on the Jacobian of demand discussed in Section B.2, but do not impose exchangeability. Thus, we allow the organic and non-organic category to have different demand functions. Further, we choose the degree of the polynomials for the Bernstein approximation based on a two-fold cross-validation procedure.[51] Because this procedure involves a number of parameters too large to report, we instead show in Table 3 the median estimated own- and cross-price elasticities for the two inside goods.

In order to compare the fit of the nonparametric model relative to the mixed logit model, we follow the same two-fold cross-validation approach used to choose the degree for the

[50] Following the original BLP paper, we also estimated a mixed-logit model with a normal random coefficient. The coefficients—and more importantly—the counterfactuals in Section 5.4 are very similar across the two specifications. In the paper, we present the two-point distribution because it is slightly more flexible (it has one extra parameter) and is faster to compute given that it does not require simulating any integrals to compute the predicted market shares.

[51] See, e.g., Chetverikov and Wilhelm (2017). Specifically, we partition the sample into two subsamples of equal size. Then, we estimate the model using the first subsample and compute the mean squared error (MSE) for the second subsample. We repeat this procedure inverting the role of the two subsamples and use the average of the two MSEs as the criterion for choosing the polynomial degree. We let the polynomial degree vary in the set {6, 8, 10, 12, 14} an find that a polynomial of degree 10 delivers the lowest average MSE.

Table 3: Nonparametric estimation results

	Non-organic	Organic
Own-price elasticity	−1.402	−5.503
	(0.032)	(0.672)
Cross-price elasticity	0.699	1.097
	(0.044)	(0.177)

Note: Median values. Asymptotically valid standard errors in parentheses.

Table 4: Two-Fold Cross-Validation Results

	NPD	Mixed Logit
MSE	0.93	2.38

Bernstein polynomial approximation. As shown in Table 4, the greater flexibility of the NPD model translates into lower average MSE.

5.4 Counterfactuals

We use the estimates to address two counterfactual questions. First, we consider the effects of a per-unit tax on prices.[52] In each market, we compute the equilibrium prices when a tax is levied on each of the inside goods individually. We set the tax equal to 25% of the price for the product in that market. As shown in Table 5, the nonparametric approach delivers a higher median tax pass-through in the case of non-organic strawberries relative to the mixed logit model. However, the two confidence intervals overlap. On the contrary, in the case of organic strawberries, the nonparametric model yields a much lower median pass-through (33% of the tax) relative to mixed logit (91%) with no overlap in the confidence intervals. To shed some light on the drivers of this pattern, in Figure 7 we plot the own-price elasticity for the organic product as a function of its price.[53] The own-price elasticity estimated nonparametrically is much steeper than the parametric one. This is consistent with the pass-through results. A retailer facing a steeper elasticity function has a stronger incentive to contain the price increase in response to the tax relative to a retailer facing a flatter elasticity function.

[52] As argued in Weyl and Fabinger (2013), the equilibrium outcomes are not affected by whether the tax is nominally levied on the consumers or on the retailer. This is true for a variety of models of supply, including monopoly. Therefore, without loss of generality, we may assume the tax is nominally levied on consumers in the form of a sales tax.

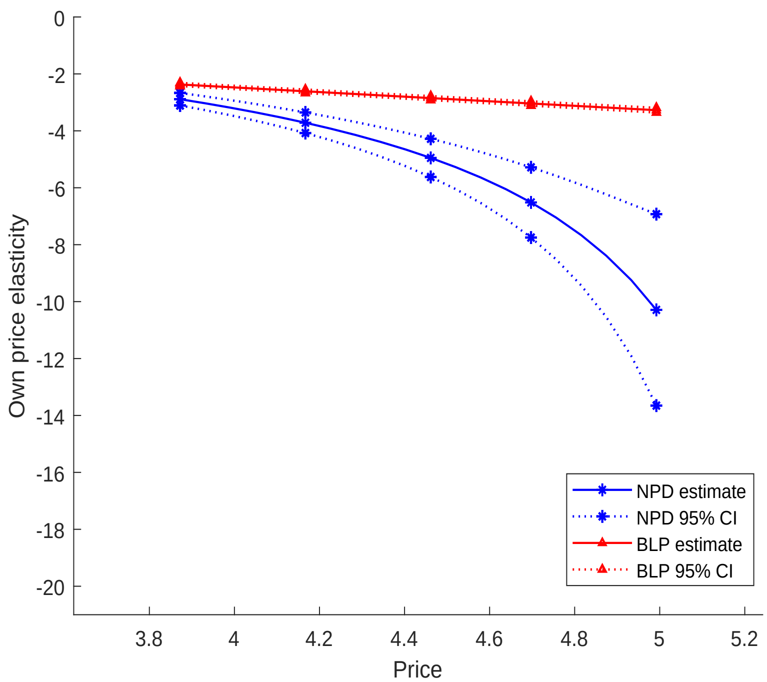
[53] Own-price on the horizontal axis varies within its interquartile range. We set all other variables at their median levels, except for δ_2 which we set at its 75% percentile. Setting it at its median delivers a similar shape for the elasticity function, but noisier estimates due to the fact that s_2—which shows up in the denominator of the elasticity—approaches zero as p_2 increases.

Table 5: Effect of a specific tax

	NPD	Mixed Logit
Non-organic	0.84	0.53
	(0.17)	$(5{\cdot}10^{-3})$
Organic	0.33	0.91
	(0.23)	$(5{\cdot}10^{-4})$

Note: Median changes in prices as a percentage of the tax. 95% confidence intervals in parentheses.

Figure 7: Organic strawberries: Own-price elasticity function



As a second counterfactual experiment, we quantify the "portfolio effect". Specifically, we ask what prices would be charged if, in each market, there were two competing retailers, one selling organic strawberries and the other selling non-organic strawberries. We assume the two retailers compete on prices, compute the resulting equilibrium and compare it to the observed prices which, according to our model, are jointly chosen by the (multi-product) retailer to maximize profits. This type of exercise helps understand the impact of big-box retailers on consumer prices. On the one hand, one would expect big players to enjoy economies of scale and thus face lower marginal costs. On the other hand, a retailer selling multiple products is able to partially internalize price competition, which might lead to higher prices all else equal. In Table 6 we report the difference between the observed prices and the prices that would arise in the counterfactual world with two single-product retailers. The parametric model in 6—labeled Mixed Logit (I) in Table 6—and the nonparametric

Table 6: Effect of multi-product pricing

	NPD	Mixed Logit (I)	Mixed Logit (II)	Mixed Logit (III)
Non-organic	0.10 $(3{\cdot}10^{-3})$	0.08 $(1{\cdot}10^{-3})$	0.20 $(8{\cdot}10^{-4})$	0.21 $(2{\cdot}10^{-3})$
Organic	0.43 $(6{\cdot}10^{-3})$	0.42 $(2{\cdot}10^{-3})$	0.54 $(9{\cdot}10^{-4})$	0.55 $(1{\cdot}10^{-3})$

Note: Median difference between the observed prices and the optimal prices chosen by two competing retailers as a percentage of markups. 95% confidence intervals in parentheses. Mixed Logit (I) refers to the model in 6. Mixed Logit (II) refers to the model in 6 with $\beta_1 = \beta_2$; Mixed Logit (III) refers to the model in 6 with $\beta_1 = \beta_2 = 0$.

approach produce very close results. In the median market, both attribute around 10% and just above 40% of markups to the portfolio effect for non-organic and organic strawberries, respectively. One may wonder how robust this result is to modifications of the parametric model. To this end, we also estimate two additional models—labeled Mixed Logit (II) and (III)—that restrict the constants in model 6 to be the same and to be zero, respectively. In other words, Mixed Logit (I) allows for product-specific dummies, Mixed Logit (II) only allows for a dummy for the inside goods jointly, and Mixed Logit (III) does not allow for any unobserved systematic differences between the inside goods or between the inside and the outside goods. The two restricted models tend to attribute much more importance to the portfolio effect relative to the unrestricted parametric model or NPD. This suggests that allowing for product specific dummies is important in this context and points to a wider use of the approach developed in this paper as a tool for selecting among different possible (parametric) models.

6 Conclusion

In this paper, we develop and apply a nonparametric approach to estimate demand in differentiated products markets. Our proposed methodology relaxes several arguably arbitrary restrictions on consumer behavior and preferences that are embedded in standard discrete choice models. Instead, we pursue a nonparametric approach that directly targets the demand functions and leverages a number of constraints from economic theory. Further, we provide primitive conditions sufficient to obtain valid standard errors for quantities of interest.

We apply our approach to quantify the pass-through of a tax and assess the effect of multi-product retailers on prices. While we find that the nonparametric method yields lower tax pass-through for one product category, the results for the second counterfactual exercise suggest that a flexible enough parametric model captures the patterns in the data well.

Appendix A: Bernstein Polynomials

For a positive integer m, the Bernstein basis function is defined as

$$b_{v,m}(u) = \binom{m}{v} u^v (1-u)^{m-v},$$

where $v = 0, 1, ..., m$ and $u \in [0,1]$. The integer m is called the degree of the Bernstein basis. In order to approximate a univariate function on the unit interval, one may take a linear combination of the Bernstein basis functions

$$\sum_{v=0}^{m} \theta_{v,m} b_{v,m}(u),$$

for some coefficients $(\theta_{v,m})_{v=0}^{m}$. Similarly, for a function of N variables living in the $[0,1]^N$ hyper-cube, one may use a polynomial of the form

$$\sum_{v_1=0}^{m} \cdots \sum_{v_N=0}^{m} \theta_{v_1,\cdots,v_N,m} b_{v_1,m}(u_1) \cdots b_{v_N,m}(u_N)$$

Note that here we are assuming that the order m is the same for each dimension $n = 1, ..., N$. This is not needed, but we only discuss this case for notational convenience.

Historically, Bernstein polynomials were introduced to approximate an arbitrary function g by a sequence of smooth functions. This is motivated by the following result.[54]

Lemma 2. *Let g be a bounded real-valued function on $[0,1]^N$ and define*

$$B_m[g] = \sum_{v_1=0}^{m} \cdots \sum_{v_N=0}^{m} g\left(\frac{v_1}{m}, \cdots, \frac{v_N}{m}\right) b_{v_1,m}(u_1) \cdots b_{v_N,m}(u_N)$$

Then,

$$\sup_{\mathbf{u} \in [0,1]^N} |B_m[g](\mathbf{u}) - g(\mathbf{u})| \to 0$$

as $m \to \infty$.

This means that, for an appropriate choice of the coefficients, the sequence of Bernstein polynomials provide a uniformly good approximation to any bounded function on the unit hyper-cube as the degree m increases. Specifically, the approximation in Lemma 2 is such that the coefficient on the $b_{v_1,m}(u_1) \cdots b_{v_N,m}(u_N)$ term corresponds to the target function evaluated at $\left[\frac{v_1}{m}, \cdots, \frac{v_N}{m}\right]$, for $v_i = 0, ..., m$ and $i = 1, ..., N$.

One important implication of this result is that, for large m, any property satisfied by the target function g at the grid points $\left\{\left\{\frac{v_1}{m}, \cdots, \frac{v_N}{m}\right\}_{v_i=0}^{m}\right\}_{i=1}^{N}$ should be inherited by the corresponding Bernstein coefficients in order for the resulting approximation to be uniformly good. This gives us *necessary* conditions on the Bernstein coefficients for large m.

To fix ideas, consider the following simple example. Suppose the target function $g : [0,1] \to \mathbb{R}$ is nondecreas-

[54]See, e.g., Chapter 2 of Gal (2008).

ing and that we approximate it using

$$\hat{g}(u) = \sum_{v=0}^{m} \theta_{v,m} b_{v,m}(u) \qquad u \in [0,1].$$

Then for large m, the coefficients $(\theta_{v,m})_{v=0}^{m}$ must satisfy $\theta_{0,m} \leq \theta_{1,m} \leq \cdots \leq \theta_{m,m}$ in order for $\hat{g}$ to be uniformly close to g. To see this, suppose by contradiction that $\theta_{j,m} > \theta_{k,m}$ for some $j < k$ and large m. Then, by Lemma 2, $\hat{g}$ is close to a function h such that $h\left(\frac{j}{m}\right) > h\left(\frac{k}{m}\right)$, i.e. a function that is *not* monotonically nondecreasing. In other words, a monotonicity restriction on the target g implies that the Bernstein coefficients must satisfy intuitive - and in this case linear - monotonicity constraints for large m. The same logic applies to any other assumptions we might be willing to impose on g. This is a powerful tool in the context of demand estimation because economic theory provides us with several restrictions on the structural demand function σ (and therefore on its inverse σ^{-1}).

Appendix B: Additional Constraints

In this appendix, we consider several constraints that one might be willing to impose in estimation besides the exchangeability restrictions discussed in Section 3.2, and we show how to enforce them in estimation in a computationally tractable way. Because these constraints are defined conditional on any given value of $x^{(2)}$, we drop this from notation for notational convenience.

B.1 Symmetry

Let $\mathbb{J}_{\sigma}^{p}(\delta, p)$ denote the Jacobian matrix of σ with respect to p:

$$\mathbb{J}_{\sigma}^{p}(\delta, p) = \begin{bmatrix} \frac{\partial}{\partial p_1}\sigma_1(\delta,p) & \cdots & \frac{\partial}{\partial p_J}\sigma_1(\delta,p) \\ \vdots & \ddots & \vdots \\ \frac{\partial}{\partial p_1}\sigma_J(\delta,p) & \cdots & \frac{\partial}{\partial p_J}\sigma_J(\delta,p) \end{bmatrix}$$

This matrix is the Jacobian of the Marshallian demand system. If we assume that there are no income effects, it coincides with the Jacobian of the Hicksian demand by Slutsky equation and therefore it must be symmetric.

Similarly, let $\mathbb{J}_{\sigma}^{\delta}(\delta, p)$ denote the Jacobian matrix of σ with respect to δ:

$$\mathbb{J}_{\sigma}^{\delta}(\delta, p) = \begin{bmatrix} \frac{\partial}{\partial \delta_1}\sigma_1(\delta,p) & \cdots & \frac{\partial}{\partial \delta_J}\sigma_1(\delta,p) \\ \vdots & \ddots & \vdots \\ \frac{\partial}{\partial \delta_1}\sigma_J(\delta,p) & \cdots & \frac{\partial}{\partial \delta_J}\sigma_J(\delta,p) \end{bmatrix}$$

In a discrete choice model where δ_j is interpreted as a quality index for good j, if one assumes that, for all j, δ_j enters the utility of good j linearly (and does not enter the utility of the other goods), then $\mathbb{J}_{\sigma}^{\delta}(\delta, p)$ must be symmetric.

Conveniently, symmetry of $\mathbb{J}_{\sigma}^{\delta}(\delta, p)$ implies linear constraints on the Bernstein coefficients. To see this, note that by the implicit function theorem, for every (δ, p) and for $s = \sigma(\delta, p)$,

$$\mathbb{J}^{s}_{\sigma^{-1}}(s, p) = \left[\mathbb{J}^{\delta}_{\sigma}(\delta, p)\right]^{-1} \tag{7}$$

Because the inverse of a symmetric matrix is symmetric, symmetry of $\mathbb{J}^{\delta}_{\sigma}(\delta, p)$ implies symmetry of $\mathbb{J}^{s}_{\sigma^{-1}}(s, p)$. This, in turn, imposes *linear* constraints on the Bernstein coefficients as the degree of the approximation goes to infinity.[55]

On the other hand, it appears that symmetry of $\mathbb{J}^{p}_{\sigma}$ requires nonlinear constraints. This is because, by the implicit function theorem, for every (δ, p) and for $s = \sigma(\delta, p)$,

$$\mathbb{J}^{p}_{\sigma}(\delta, p) = -\left[\mathbb{J}^{s}_{\sigma^{-1}}(s, p)\right]^{-1} \mathbb{J}^{p}_{\sigma^{-1}}(s, p) \tag{8}$$

which shows that $\mathbb{J}^{p}_{\sigma}$ is a nonlinear function of the derivatives of σ^{-1} and therefore of the Bernstein coefficients. In estimation, we found it convenient to rewrite (8) as

$$\mathbb{J}^{s}_{\sigma^{-1}}(s, p)\, \mathbb{J}^{p}_{\sigma}(\delta, p) = -\mathbb{J}^{p}_{\sigma^{-1}}(s, p)$$

We then express $\mathbb{J}^{s}_{\sigma^{-1}}$ and $\mathbb{J}^{p}_{\sigma^{-1}}$ as linear combinations of the Bernstein polynomials and introduce extra parameters (call them γ) for the entries of $\mathbb{J}^{p}_{\sigma}$. In this way, we obtain a set of nonlinear constraints that are linear in the Bernstein coefficients θ, given γ, and linear in γ, given θ.[56]

B.2 Additional Properties of the Jacobian of Demand

The matrix $\mathbb{J}^{\delta}_{\sigma}(\delta, p)$ has a number of additional features that we might want to impose in estimation. First, the weak substitutability in Assumption 2(i) requires the off-diagonal elements to be non-positive. Further, it follows from Remark 2 of Berry, Gandhi, and Haile (2013) that the diagonal elements must be positive.[57]

Moreover, $\mathbb{J}^{\delta}_{\sigma}(\delta, p)$ belongs to the class of M-matrices, which are the object of a vast literature in linear algebra.[58] One of the most common definitions of this class is as follows.

Definition 1. A square real matrix is called an M-matrix if (i) it of the form $A = \alpha I - P$, where all entries of P are non-negative; (ii) A is nonsingular and A^{-1} is entry-wise non-negative.

[55] More precisely, the difference between two (appropriately chosen) Bernstein coefficients approximates the change in the function σ_j^{-1} given by a change in s_k. Thus, we would in principle need to divide by the distance between the grid points associated with the two coefficients in order to obtain the derivative of σ_j^{-1} with respect to s_k. However, because we are interested in comparisons between derivatives and the grid points are equidistant, the increments in the denominator cancel out. Therefore, we are left with inequalities involving simple differences of the Bernstein coefficients.

[56] This is helpful especially when it comes to writing the analytic gradient of the constraints to input in the optimization problem.

[57] This is simply the requirement that the structural demand of product j increase in the index δ_j. While a very reasonable condition, it is not needed for identification, but rather it follows from the sufficient conditions given in Section 2.

[58] See, e.g., Plemmons (1977).

We now formalize the aforementioned result, which is a simple corollary of Theorem 2 in Berry, Gandhi, and Haile (2013).

Lemma 3. *Let Assumptions 1 and 2 hold. Then* $\mathbb{J}_{\sigma}^{\delta}(\delta,p)$ *is an M-matrix for all* (δ,p)*.*

Proof. See Section B.3. □

The linear algebra literature provides several properties of M-matrices. However, it is not *a priori* clear how to impose these properties in estimation, since we estimate σ^{-1} rather than σ itself. The Jacobian of the function we estimate is

$$\mathbb{J}_{\sigma^{-1}}^{s}(s,p)=\begin{bmatrix}\frac{\partial}{\partial s_1}\sigma_1^{-1}(s,p) & \cdots & \frac{\partial}{\partial s_J}\sigma_1^{-1}(s,p)\\ \vdots & \ddots & \vdots\\ \frac{\partial}{\partial s_1}\sigma_J^{-1}(s,p) & \cdots & \frac{\partial}{\partial s_J}\sigma_J^{-1}(s,p)\end{bmatrix}$$

Recall that, by the implicit function theorem, we have that, for every (δ,p) and for $s=\sigma(\delta,p)$,

$$\mathbb{J}_{\sigma^{-1}}^{s}(s,p)=\left[\mathbb{J}_{\sigma}^{\delta}(\delta,p)\right]^{-1}$$

Therefore, $\mathbb{J}_{\sigma^{-1}}^{s}(s,p)$ is the inverse of an M-matrix or, in the jargon used in the linear algebra literature, an *inverse M-matrix.* Fortunately, inverse M-matrices have also been widely studied.[59] Thus, we may borrow results from that literature to impose conditions on the Bernstein coefficients for σ^{-1} that must hold in order for $\mathbb{J}_{\sigma}^{\delta}(\delta,p)$ to be an M-matrix.

First, it follows from part (ii) of Definition 1 that $\mathbb{J}_{\sigma^{-1}}^{s}(s,p)$ must have non-negative elements for all (s,p). This means that, for every j, σ_j^{-1} must be increasing in s_k for all k. As discussed in Appendix A, monotonicity is very easy to impose in estimation, given that it reduces to a collection of linear inequalities on the Bernstein coefficients.

Second, under Assumption 2, $\mathbb{J}_{\sigma}^{\delta}$ satisfies a property called column diagonal dominance. The economic content of this property is that the (positive) effect of δ_j on the share of good j is larger than the combined (negative) effect of δ_j on the shares of all other goods, in absolute value. A few definitions are necessary to formalize this point.

Definition 2. An I−by−I matrix $A=(a_{ij})$ is (weakly) diagonally dominant of its rows if

$$|a_{ii}|\geq\sum_{j\neq i}|a_{ij}|,$$

for $i=1,...,I$.

Definition 3. An I−by−I matrix $A=(a_{ij})$ is (weakly) diagonally dominant of its row entries if

$$|a_{ii}|\geq|a_{ij}|,$$

for $i=1,...,I$ and $j\neq i$.

Column diagonal dominance and column entry diagonal dominance are defined analogously. By Theorem 3.2 of McDonald, Neumann, Schneider, and Tsatsomeros (1995), if an M-matrix M is weakly diagonally

[59] See, e.g., Johnson and Smith (2011).

dominant matrix of its columns, then $(M)^{-1}$ is weakly diagonally dominant of its row entries.[60] This immediately implies the following result.

Lemma 4. *Fix (δ, p) and let $s = \sigma(\delta, p)$. If $\mathbb{J}_\sigma^\delta$ is diagonally dominant of its columns, then $\frac{\partial}{\partial s_j}\sigma_j^{-1}(s,p) \geq \frac{\partial}{\partial s_k}\sigma_j^{-1}(s,p)$ for all j and all $k \neq j$.*

Lemma 4 translates the assumption that $\mathbb{J}_\sigma^\delta$ is diagonally dominant of its columns into linear inequalities involving the derivatives of σ^{-1}. Therefore it follows from the same argument used for symmetry that diagonal dominance may be imposed through linear constraints on the Bernstein coefficients.

B.3 Proofs

Proof of Lemma 1. Let $\pi : \{1, ..., J\} \to \{1, ..., J\}$ be any permutation and let $\tilde{\pi}$ denote the function that, for any J-vector y returns the reshuffled version of y obtained by permuting its subscripts according to π, i.e.

$$\tilde{\pi}([y_1, ..., y_J]) = \left[y_{\pi(1)}, ..., y_{\pi(J)}\right]$$

Then, we can rewrite the definition of exchangeability for a generic function $g(y, w)$ of $2J$ arguments as

$$\tilde{\pi}(g(y, w)) = g(\tilde{\pi}(y), \tilde{\pi}(w)).$$

Now take any (δ, p) and let $s = \sigma(\delta, p)$. We can invert the demand system at (δ, p) to obtain

$$\delta = \sigma^{-1}(s, p) \tag{9}$$

By exchangeability of σ, we also have

$$\tilde{\pi}(s) = \sigma(\tilde{\pi}(\delta), \tilde{\pi}(p))$$

Inverting this demand system, we obtain

$$\tilde{\pi}(\delta) = \sigma^{-1}(\tilde{\pi}(s), \tilde{\pi}(p)) \tag{10}$$

Combining (9) and (10),

$$\tilde{\pi}\left(\sigma^{-1}(s, p)\right) = \sigma^{-1}(\tilde{\pi}(s), \tilde{\pi}(p))$$

which shows that σ^{-1} is exchangeable.

□

Proof of Lemma 3. Under Assumptions 1 and 2, Theorem 2 in Berry, Gandhi, and Haile (2013) implies that $\mathbb{J}_\sigma^\delta(\delta, p)$ is a P-matrix for every (δ, p), where a P-matrix is a square matrix such that all of its principal minors are strictly positive. Next, by the weak substitutability imposed by Assumption 2, $\mathbb{J}_\sigma^\delta(\delta, p)$ is also a Z-matrix, where a Z-matrix is a matrix with non-positive off-diagonal entries. Finally, since a Z-matrix which is also a P-matrix is an M-matrix,[61] the result follows.

□

[60] To reconcile Theorem 3.2 of McDonald, Neumann, Schneider, and Tsatsomeros (1995) and Definition 3, recall that an inverse M-matrix has non-negative entries.

[61] See, e.g., 8.148 in Seber (2007).

Appendix C: Inference Results

This appendix contains all the assumptions and proofs for the inference results stated in the main text.

C.1 Setup

We first introduce some notation that is used throughout this appendix. We denote by $\mathcal{S}, \mathcal{P}, \mathcal{Z}, \Xi$ the support of S, P, Z, ξ, respectively. Also, we let $W \equiv (X, Z)$ denote the exogenous variables and $\mathcal{W}$ denote its support. Similarly, we let $Y \equiv \left(S, P, X^{(2)}\right)$ denote the (endogenous and exogenous) regressors and $\mathcal{Y}$ denote its support. For every $y \in \mathcal{S} \times \mathcal{P}$, let $h_0(y) \equiv [h_{0,1}(y), ..., h_{0,J}(y)]' \equiv \left[\sigma_1^{-1}(y), ...\sigma_J^{-1}(y)\right]'$, so that the estimating equations become

$$x_j = h_{0,j}(y) + \xi_j, \quad j \in \mathcal{J}. \tag{11}$$

We assume that $h_0 \in \mathcal{H}$, where $\mathcal{H}$ is the Hölder ball of smoothness r, and we endow it with the norm $||\cdot||_\infty$ defined by $||h||_\infty \equiv \max_{j\in\mathcal{J}} ||h_j||_\infty$ for a function $h = [h_1, ...h_J]$.
Further, we let $\left\{\psi^{(i)}_{1,M_i}, ..., \psi^{(i)}_{M_i,M_i}\right\}$ be the collection of basis functions used to approximate $h_{0,i}$ for $j \in \mathcal{J}$, and let $M = \sum_{j=1}^J M_j$ be the dimension of the overall sieve space for h. Similarly, we let $\left\{a^{(i)}_{1,K_i}, ..., a^{(i)}_{K_i,K_i}\right\}$ be the collection of basis functions used to approximate the instrument space for $h_{0,i}$, and let and $K = \sum_{j=1}^J K_j$.

Next, for $j \in \mathcal{J}$, letting $\text{diag}(mat_1, ..., mat_J) \equiv \begin{bmatrix} mat_1 & 0_{d_{1,r}\times d_{2,c}} & \cdots & 0_{d_{1,r}\times d_{J,c}} \\ 0_{d_{2,r}\times d_{1,c}} & mat_2 & \cdots & 0_{d_{2,r}\times d_{J,c}} \\ \vdots & \vdots & \ddots & \vdots \\ 0_{d_{J,r}\times d_{1,c}} & 0_{d_{J,r}\times d_{2,c}} & \cdots & mat_J \end{bmatrix}$ for matrices $mat_j \in$

$\mathbb{R}^{d_{j,r}\times d_{j,c}}$ with $i \in \mathcal{J}$, we define

$$\begin{aligned}
\psi_{M_i}^{(i)}(y) &= \left(\psi_{1,M_i}^{(i)}(y), ..., \psi_{M_i,M_i}^{(i)}(y)\right)' && M_i - \text{by} - 1 \\
\psi_M(y) &= \text{diag}\left(\psi_{M_1}^{(1)}(y), ..., \psi_{M_J}^{(J)}(y)\right) && M - \text{by} - J \\
\Psi_{(i)} &= \left(\psi_{M_i}^{(i)}(y_1), ..., \psi_{M_i}^{(i)}(y_T)\right)' && T - \text{by} - M_i \\
a_{K_i}^{(i)}(w) &= \left(a_{1,K_i}^{(i)}(w), ..., a_{K_i,K_i}^{(i)}(w)\right)' && K_i - \text{by} - 1 \\
a_K(w) &= \text{diag}\left(a_{K_1}^{(1)}(w), ..., a_{K_J}^{(J)}(w)\right) && K - \text{by} - J \\
A_{(i)} &= \left(a_{K_i}^{(i)}(w_1), ..., a_{K_i}^{(i)}(w_T)\right)' && T - \text{by} - K_i \\
A &= \text{diag}\left(A_{(1)}, ..., A_{(J)}\right) && JT - \text{by} - K \\
L_i &= \mathbb{E}\left(a_{K_i}^{(i)}(W_t)\,\psi_{M_i}^{(i)}(Y_t)'\right) && K_i - \text{by} - M_i \\
L &= \text{diag}\left(L_1, ..., L_J\right) && K - \text{by} - M \\
\hat{L}_i &= \frac{A_{(i)}'\Psi_{(i)}}{T} && K_i - \text{by} - M_i \\
\hat{L} &= \text{diag}\left(\hat{L}_1, ..., \hat{L}_J\right) && K - \text{by} - M \\
G_{A,i} &= \mathbb{E}\left(a_{K_i}^{(i)}(W_t)\,a_{K_i}^{(i)}(W_t)'\right) && K_i - \text{by} - K_i \\
G_A &= \text{diag}\left(G_{A,1}, ..., G_{A,J}\right) && K - \text{by} - K \\
\hat{G}_{A,i} &= \frac{A_{(i)}'A_{(i)}}{T} && K_i - \text{by} - K_i \\
\hat{G}_A &= \text{diag}\left(\hat{G}_{A,1}, ..., \hat{G}_{A,J}\right) && K - \text{by} - K
\end{aligned}$$

$$\begin{aligned}
G_{\psi,i} &= \mathbb{E}\left(\psi_{M_i}^{(i)}(Y_t)\,\psi_{M_i}^{(i)}(Y_t)'\right) && M_i - \text{by} - M_i \\
G_\psi &= \text{diag}\left(G_{\psi,1}, ..., G_{\psi,J}\right) && M - \text{by} - M \\
X_{(i)} &= \left(x_{i1}, ..., x_{iT}\right)' && T - \text{by} - 1 \\
X &= \left(X_{(1)}', ..., X_{(J)}'\right)' && JT - \text{by} - 1
\end{aligned}$$

Also, we let

$$\begin{aligned}
\Omega_{jj} &= \mathbb{E}\left(\xi_{j,t}^2 a_{K_j}^{(j)}(W_t)\,a_{K_j}^{(j)}(W_t)'\right) && K_j - \text{by} - K_j \\
\Omega_{jk} = \Omega_{kj}' &= \mathbb{E}\left(\xi_{j,t}\xi_{k,t}a_{K_j}^{(j)}(W_t)\,a_{K_k}^{(k)}(W_t)'\right) && K_j - \text{by} - K_k \\
\Omega &= \begin{bmatrix} \Omega_{11} & \Omega_{12} & \cdots & \Omega_{1J} \\ \Omega_{21} & \Omega_{22} & \cdots & \Omega_{2J} \\ \vdots & \vdots & \ddots & \cdots \\ \Omega_{J1} & \Omega_{J2} & \cdots & \Omega_{JJ} \end{bmatrix} && K - \text{by} - K
\end{aligned}$$

and, similarly,

$$\hat{\Omega}_{jj} = \frac{1}{T}\sum_{t=1}^{T}\hat{\xi}_{j,t}^{2} a_{K_j}^{(j)}(w_t)\, a_{K_j}^{(j)}(w_t)' \qquad K_j - \text{by} - K_j$$

$$\hat{\Omega}_{jk} = \hat{\Omega}_{kj}' = \frac{1}{T}\sum_{t=1}^{T}\hat{\xi}_{j,t}\hat{\xi}_{j,t} a_{K_j}^{(j)}(w_t)\, a_{K_k}^{(k)}(w_t)' \qquad K_j - \text{by} - K_k$$

$$\hat{\Omega} = \begin{bmatrix} \hat{\Omega}_{11} & \hat{\Omega}_{12} & \cdots & \hat{\Omega}_{1J} \\ \hat{\Omega}_{21} & \hat{\Omega}_{22} & \cdots & \hat{\Omega}_{2J} \\ \vdots & \vdots & \ddots & \cdots \\ \hat{\Omega}_{J1} & \hat{\Omega}_{J2} & \cdots & \hat{\Omega}_{JJ} \end{bmatrix} \qquad K - \text{by} - K$$

where $\hat{\xi}_{j,t} = x_{j,t} - \hat{h}_j(y_t)$.
For $j \in \mathcal{J}$, we define

$$\zeta_{A,i} \equiv \sup_{w\in\mathcal{W}} \left|\left| G_{A,i}^{-1/2} a_{K_i}^{(i)}(w) \right|\right| \qquad \zeta_{\psi,i} \equiv \sup_{y\in\mathcal{Y}} \left|\left| G_{\psi,i}^{-1/2} \psi_{M_i}^{(i)}(y) \right|\right| \qquad \zeta_i \equiv \zeta_{A,i} \vee \zeta_{\psi,i}$$

and let $\zeta \equiv \max_{j\in\mathcal{J}} \zeta_j$. As in CC, we use the following sieve measure of ill-posedness, for $i \in \mathcal{J}$,

$$\tau_{M_i}^{(i)} \equiv \sup_{h_i\in\Psi_{M_i}: h_i\neq 0} \left(\frac{\mathbb{E}\left[(h_i(Y))^2 \right]}{\mathbb{E}\left[(\mathbb{E}[h_i(Y)|W])^2 \right]} \right)^{1/2}$$

where $\Psi_{M_i} \equiv clsp\left\{ \psi_{M_i}^{(i)} \right\}$ and we let $\tau_M \equiv \max_{j\in\mathcal{J}} \tau_{M_j}^{(j)}$.
For every $2J-$vector of integers $\tilde{\alpha}$ and function $g : \mathcal{Y} \mapsto \mathbb{R}$, we let $|\tilde{\alpha}| \equiv \sum_{j=1}^{2J} \tilde{\alpha}_j$ and $\partial^{\tilde{\alpha}} g \equiv \frac{\partial^{|\tilde{\alpha}|} g}{\partial^{\tilde{\alpha}_1} s_1 \cdots \partial^{\tilde{\alpha}_J} s_J \partial^{\tilde{\alpha}_{J+1}} p_1 \cdots \partial^{\tilde{\alpha}_{2J}} p_J}$.
Similarly, for $h = [h_1, ..., h_J] : \mathcal{Y} \mapsto \mathbb{R}^J$, we let $\partial^{\tilde{\alpha}} h \equiv \left[\partial^{\tilde{\alpha}} h_1, \cdots, \partial^{\tilde{\alpha}} h_J \right]$.
The (unconstrained) sieve NPIV estimator $\hat{h}_i$ has the following closed form

$$\hat{h}_i(y) \quad = \quad \psi_M^{(i)}(y)' \hat{\theta}_i$$

for

$$\hat{\theta}_i \quad = \quad \left[\Psi_{(i)}' A_{(i)} \left(A_{(i)}' A_{(i)} \right)^{-} A_{(i)}' \Psi_{(i)} \right]^{-} \Psi_{(i)}' A_{(i)} \left(A_{(i)}' A_{(i)} \right)^{-} A_{(i)}' X_{(i)}$$

We write this in a more compact form as

$$\hat{\theta}_i \quad = \quad \frac{1}{T} \left[\hat{L}_i' \hat{G}_{A,i}^{-} \hat{L}_i \right]^{-} \hat{L}_i' \hat{G}_{A,i}^{-} A_{(i)}' X_{(i)}$$

Stacking the J estimators, we write

$$\hat{\theta} = \left(\hat{\theta}_1', \cdots, \hat{\theta}_J' \right)' = \frac{1}{T} \left[\hat{L}' \hat{G}_A^{-} \hat{L} \right]^{-} \hat{L}' \hat{G}_A^{-} A' X$$

and

$$\hat{h}(y) = \psi_M(y)' \hat{\theta}$$

Next, letting $H_{0,j} \equiv \left(h_{0,j}\left(y_1\right), ..., h_{0,j}\left(y_T\right)\right)'$ and $H_0 \equiv \left(H_{0,1}', \cdots, H_{0,J}'\right)'$, we define

$$\tilde{\theta} = \frac{1}{T}\left[\hat{L}'\hat{G}_A^- \hat{L}\right]^- \hat{L}'\hat{G}_A^- A' H_0 \tag{12}$$

and let

$$\tilde{h}\left(y\right) = \psi_M\left(y\right)' \tilde{\theta}$$

Given a functional $f : \mathcal{H} \mapsto \mathbb{R}$ and $h \in \mathcal{H}$, we let $vec_{g,J,j}$ be the column $J-$vector valued function that returns all zeros except for the $j-$th element, where it returns the function g. Further, we let

$$Df\left(h\right)\left[\psi_{M_j}^{(j)}\right] \equiv \left(Df\left(h\right)\left[vec_{\psi_{1,M_j}^{(j)},J,j}\right], ..., Df\left(h\right)\left[vec_{\psi_{M_j,M_j}^{(j)},J,j}\right]\right)' \qquad M_j - \text{by} - 1$$

$$Df\left(h\right)\left[\psi_M\right] \equiv \left(Df\left(h\right)\left[\psi_{M_1}^{(1)}\right]', ...Df\left(h\right)\left[\psi_{M_J}^{(J)}\right]'\right)' \qquad M - \text{by} - 1$$

and, for $(h, v) \in \mathcal{H} \times \mathcal{H}$, we let $Df\left(h\right)\left[v\right] \equiv \left.\frac{\partial f(h+\tau v)}{\partial \tau}\right|_{\tau=0}$ denote the pathwise derivative of f at h in the direction v.

Finally, we let

$$v_T^2\left(f\right) = Df\left(h_0\right)\left[\psi_M\right]' \left(S'G_A^{-1}S\right)^{-1} S'G_A^{-1}\Omega G_A^{-1}S\left(S'G_A^{-1}S\right)^{-1} Df\left(h_0\right)\left[\psi_M\right]$$

denote the sieve variance for the estimator $f\left(\hat{h}\right)$ of the functional f, and let the sieve variance estimator be

$$\hat{v}_T^2\left(f\right) = Df\left(\hat{h}\right)\left[\psi_M\right]' \left(\hat{L}'\hat{G}_A^{-1}\hat{L}\right)^{-1} \hat{L}'\hat{G}_A^{-1}\hat{\Omega}\hat{G}_A^{-1}\hat{L}\left(\hat{L}'\hat{G}_A^{-1}\hat{L}\right)^{-1} Df\left(\hat{h}\right)\left[\psi_M\right] \tag{13}$$

Because the functionals of interest are defined for fixed $(\overline{s}, \overline{p})$, they will typically be irregular, i.e. $v_T^2\left(f\right) \nearrow \infty$ as $T \to \infty$. Therefore, we tailor the proofs to this case.

C.2 Theorem 2: General Irregular Functionals

The following is largely based on the proof of Theorem D.1 in CC.

We make the following assumptions.

Assumption 5. *For all $j, k \in \mathcal{J}, j \neq k$:*
(i) $\sup_{w\in\mathcal{W}} \mathbb{E}\left(\xi_j^2|w\right) \leq \overline{\sigma}^2 < \infty$;
(ii) $\inf_{w\in\mathcal{W}} \mathbb{E}\left(\xi_j^2|w\right) \geq \underline{\sigma}^2 > 0$;
(iii) $\sup_{w\in\mathcal{W}} \mathbb{E}\left(|\xi_j\xi_k||w\right) \leq \overline{\sigma}_{cov} < \infty$;
(iv) $\sup_{w\in\mathcal{W}} \mathbb{E}\left[\xi_j^2 \mathbb{I}\left\{\sum_{i=1}^J |\xi_i| > \ell\left(T\right)\right\}|w\right] = o\left(1\right)$ *for any positive sequence* $\ell\left(T\right) \nearrow \infty$;
(v) $\mathbb{E}\left(|\xi_j|^{2+\gamma^{(1)}}\right) < \infty$ *for some* $\gamma^{(1)} > 0$;
(vi) $\mathbb{E}\left(|\xi_j\xi_k|^{1+\gamma^{(2)}}\right) < \infty$ *for some* $\gamma^{(2)} > 0$.

Assumption 6. *(i)* $\tau_M \zeta \sqrt{M(\log M)/T} = o\left(1\right)$;
(ii) $\zeta^{\left(2+\gamma^{(1)}\right)/\gamma^{(1)}}\sqrt{(\log K)/T} = o\left(1\right)$ *and* $\zeta^{\left(1+\gamma^{(2)}\right)/\gamma^{(2)}}\sqrt{(\log K)/T} = o\left(1\right)$, *where* $\gamma^{(1)}, \gamma^{(2)} > 0$ *are defined in*

Assumption 5(v)-5(vi);
(iii) $K \asymp M$.

Assumption 7. *The basis used for the instrument spaces is the same across all goods, i.e.* $K_j = K_k$ *and* $a^{(j)}_{K_j}(\cdot) = a^{(k)}_{K_k}(\cdot)$ *for all* $j, k \in \mathcal{J}$.

Assumption 8. *Let* $\mathcal{H}_T \subset \mathcal{H}$ *be a sequence of neighborhoods of* h_0 *with* $\hat{h}, \tilde{h} \in \mathcal{H}_T$ *wpa1 and assume* $v_T(f) > 0$ *for every* T. *Further, assume that:*

(i) $v \mapsto Df(h_0)[v]$ *is a linear functional and there exists* α *with* $|\alpha| \geq 0$ *s.t.* $|Df(h_0)[h - h_0]| \lesssim ||\partial^\alpha h - \partial^\alpha h_0||_\infty$ *for all* $h \in \mathcal{H}_T$*;*

There are α_1, α_2 *with* $|\alpha_1|, |\alpha_2| \geq 0$ *s.t.*

(ii) $\left| f\left(\hat{h}\right) - f(h_0) - Df(h_0)[\hat{h} - h_0] \right| \lesssim ||\partial^{\alpha_1}\hat{h} - \partial^{\alpha_1}h_0||_\infty ||\partial^{\alpha_2}\hat{h} - \partial^{\alpha_2}h_0||_\infty$*;*

(iii) $\frac{\sqrt{T}}{\sigma_T(f)}\left(||\partial^{\alpha_1}\hat{h} - \partial^{\alpha_1}h_0||_\infty ||\partial^{\alpha_2}\hat{h} - \partial^{\alpha_2}h_0||_\infty + ||\partial^\alpha\tilde{h} - \partial^\alpha h_0||_\infty\right) = O_p(\eta_T)$ *for a nonnegative sequence* η_T *such that* $\eta_T = o(1)$*;*

(iv) $\frac{1}{v_T(f)}\left|\left| \left(Df\left(\hat{h}\right)[\psi_M]' - Df(h_0)[\psi_M]'\right)\left(G_A^{-1/2}S\right)^-_l \right|\right| = o_p(1)$.

Discussion of assumptions. Assumption 5 corresponds to Assumption 2 in CC, modified to account for the fact that my model has two main equations and thus two error terms. Assumption 6 restricts the growth rate of M and K; part 6(i) corresponds to the condition imposed by CC in Theorem D.1, while part 6(ii) is similar to Assumption 3(iii) in CC. Assumption 7 is stronger than necessary but we impose for simplicity. Assumption 8 corresponds to the sufficient conditions in Remark 4.1 of CC.

We now provide a proof of Theorem 2.

Proof of Theorem 2. We prove that

$$\sqrt{T}\frac{\left(f\left(\hat{h}\right) - f(h_0)\right)}{v_T(f)} \xrightarrow{d} N(0,1) \tag{14}$$

The result then follows from Lemma 5 below. By Assumption 8(ii),

$$\sqrt{T}\frac{\left(f\left(\hat{h}\right) - f(h_0)\right)}{v_T(f)} = \sqrt{T}\frac{Df(h_0)[\hat{h} - h_0]}{v_T(f)} + o_p\left(\underbrace{\frac{\sqrt{T}}{v_T(f)}||\partial^{\alpha_1}\hat{h} - \partial^{\alpha_1}h_0||_\infty ||\partial^{\alpha_2}\hat{h} - \partial^{\alpha_2}h_0||_\infty}_{c_T}\right)$$

By Assumption 8(iii), $c_T = o_p(1)$ and therefore,

$$\sqrt{T}\frac{\left(f\left(\hat{h}\right) - f(h_0)\right)}{v_T(f)} = \sqrt{T}\frac{Df(h_0)[\hat{h} - h_0]}{v_T(f)} + o_p(1) \tag{15}$$

Further, by Assumption 8(i)

$$Df(h_0)[\hat{h} - h_0] = Df(h_0)[\hat{h} - \tilde{h}] + Df(h_0)[\tilde{h} - h_0] \tag{16}$$

and

$$Df\left(h_0\right)[\tilde{h}-h_0] \lesssim ||\partial^\alpha \tilde{h}-\partial^\alpha h_0||_\infty \tag{17}$$

By (17) and Assumption 8(iii),

$$\sqrt{T}\frac{Df\left(h_0\right)[\tilde{h}-h_0]}{v_T\left(f\right)} = o_p\left(1\right) \tag{18}$$

Combining (15), (16) and (18), we obtain

$$\sqrt{T}\frac{\left(f\left(\hat{h}\right)-f\left(h_0\right)\right)}{v_T\left(f\right)} \quad = \quad \sqrt{T}\frac{Df\left(h_0\right)[\hat{h}-\tilde{h}]}{v_T\left(f\right)}+o_p\left(1\right) \tag{19}$$

We define

$$R_T\left(w\right)=\frac{Df\left(h_0\right)[\psi_M]'\left(S'G_A^{-1}S\right)^{-1}S'G_A^{-1}a_K\left(w\right)}{v_T\left(f\right)}$$

and note that $\mathbb{E}\left[\left(R_T\left(W_t\right)\cdot[\xi_{1,t},\cdots,\xi_{J,t}]'\right)^2\right]=1$. Then,

$$\begin{aligned}\sqrt{T}\frac{Df\left(h_0\right)[\hat{h}-\tilde{h}]}{v_T\left(f\right)} &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}R_T\left(w_t\right)\cdot[\xi_{1,t},\cdots,\xi_{J,t}]' \\ &+\frac{Df\left(h_0\right)[\psi_M]'\left(\left(\hat{L}'\hat{G}_A^{-}\hat{L}\right)^{-}\hat{L}'\hat{G}_A^{-}-\left(S'G_A^{-1}S\right)^{-1}S'G_A^{-1}\right)\left(A'\xi/\sqrt{T}\right)}{v_T\left(f\right)} \\ &\equiv T_1+T_2\end{aligned}$$

where $\xi\equiv[\xi_{1,1},\cdots,\xi_{1,T},\xi_{2,1},\cdots,\xi_{2,T},\cdots,\xi_{J,1},\cdots,\xi_{J,T}]'$.
We show that $T_1\xrightarrow{d}N\left(0,1\right)$ by the Lindeberg-Feller theorem. The Lindeberg condition requires that, for every $\epsilon>0$,

$$C_{0,T}\equiv\mathbb{E}\left[\left(R_T\left(W\right)\cdot[\xi_1,\cdots,\xi_J]'\right)^2\underbrace{\mathbb{I}\left\{\left|R_T\left(W\right)\cdot[\xi_1\cdots\xi_J]'\right|>\epsilon\sqrt{T}\right\}}_{Q_T(W,\xi)}\right]=o\left(1\right) \tag{20}$$

To show that this condition holds, note that

$$\begin{aligned}R_T\left(w_t\right)\cdot[\xi_{1,t}\cdots\xi_{J,t}]' &= \sum_{i=1}^{J}\frac{Df\left(h_0\right)\left[\psi_{M_i}^{(i)}\right]'\left(L_i'G_{A,i}^{-1}L_i\right)^{-1}L_i'G_{A,i}^{-1}a_{K_i}^{(i)}\left(w_t\right)}{v_T\left(f\right)}\xi_{i,t} \\ &\equiv \sum_{i=1}^{J}R_T^{(i)}\left(w_t\right)\xi_{i,t}\end{aligned}$$

Now, for $i \in \mathcal{J}$,

$$\begin{aligned} \left| R_T^{(i)}(w_t) \right| &\leq \frac{\left\| Df(h_0)\left[\psi_{M_i}^{(i)}\right]\left(L_i' G_{A,i}^{-1} L_i\right)^{-1} L_i' G_{A,i}^{-1/2} \right\|}{v_T(f)} \times \sup_{w \in \mathcal{W}} \left\| G_{A,i}^{-1/2} a_{K_i}^{(i)}(w) \right\| \\ &\equiv \lambda_i(T) \times \zeta_{A,i} \end{aligned}$$

by the Cauchy-Schwarz inequality and thus

$$\left| \sum_{j=1}^{J} R_T^{(j)}(w_t)\, \xi_{j,t} \right| \leq \sum_{j=1}^{J} |\xi_{j,t}| \max_i \left[\lambda_i(T) \times \zeta_{A,i}\right] \tag{21}$$

Equation (21) implies

$$Q_T(w, \xi) \leq \mathbb{I}\left\{ \sum_{j=1}^{J} |\xi_j| > \frac{\epsilon\sqrt{T}}{\max_i\left[\lambda_i(T) \times \zeta_{A,i}\right]} \right\} \equiv \overline{Q}_T(\xi)$$

for all $w \in \mathcal{W}$ and all $\xi \in \Xi$. Therefore, also using Cauchy-Schwarz and the law of iterated expectations, we have

$$\begin{aligned} C_{0,T} &\leq \mathbb{E}\left[\sum_{j=1}^{J} \left(R_T^{(j)}(W)\right)^2 \times \sum_{j=1}^{J} \xi_j^2 \times \overline{Q}_T(\xi) \right] \\ &\leq \sum_{j=1}^{J} \mathbb{E}\left[\left(R_T^{(j)}(W)\right)^2 \right] \sum_{j=1}^{J} \sup_{w \in \mathcal{W}} \mathbb{E}\left[\xi_j^2 \times \overline{Q}_T(\xi) \,|w\right] \end{aligned}$$

Now, note that, for $i \in \mathcal{J}$,

$$\limsup_{T \to \infty} \mathbb{E}\left[\left(R_T^{(i)}(W)\right)^2 \right] = \limsup_{T \to \infty} \lambda_i(T) < \infty$$

where the inequality follows from Lemma 8 below. Further, $\sup_{w \in \mathcal{W}} \mathbb{E}\left[\xi_i^2 \overline{Q}_T(\xi)\,|w\right] = o(1)$ by Assumption 5(iv) and the fact that, by Assumption 6(i) and Lemma 8, $\frac{\sqrt{T}}{\max_i[\lambda_i(T)\zeta_{A,i}]} \nearrow \infty$. Therefore, $C_{0,T} = o(1)$, the Lindeberg condition is verified, and $T_1 \xrightarrow{d} N(0,1)$.

Next, for T_2, we have

$$\begin{aligned} |T_2| &\leq v_T(f)^{-1} \left\| Df(h_0)[\psi_M]' \left(G_A^{-1/2} S\right)_l^- \right\| \left\| G_A^{-1/2} S \left\{ \left(\hat{G}_A^{-1/2} \hat{L}\right)_l^- \hat{G}_A^{-1/2} G_A^{1/2} - \left(G_A^{-1/2} S\right)_l^- \right\} \right\| \left\| G_A^{-1/2} A'\xi/\sqrt{T} \right\| \\ &= \left[(\lambda_1(T))^2 + (\lambda_2(T))^2 \right]^{1/2} \left\| G_A^{-1/2} S \left\{ \left(\hat{G}_A^{-1/2} \hat{L}\right)_l^- \hat{G}_A^{-1/2} G_A^{1/2} - \left(G_A^{-1/2} S\right)_l^- \right\} \right\| \left\| G_A^{-1/2} A'\xi/\sqrt{T} \right\| \\ &\leq \left[(\lambda_1(T))^2 + (\lambda_2(T))^2 \right]^{1/2} \max_{j \in \mathcal{J}} \left\| G_{A,i}^{-1/2} L_i \left\{ \left(\hat{G}_{A,i}^{-1/2} \hat{L}_i\right)_l^- \hat{G}_{A,i}^{-1/2} G_{A,i}^{1/2} - \left(G_{A,i}^{-1/2} L_i\right)_l^- \right\} \right\| \left\| G_A^{-1/2} A'\xi/\sqrt{T} \right\| \\ &= O_p\left(\max_{j \in \mathcal{J}} \left[\tau_{M_i}^{(i)} \zeta_i \sqrt{M_i \log M_i / T} \right] \right) \\ &= o_p(1). \end{aligned}$$

The first inequality follows from some algebra and the Cauchy-Schwarz inequality, the second inequality holds by the definition of matrix norm,the second equality is by Lemmas A.1, F.8 and F.10(c) in CC, Lemma 8 below and Assumption 6(iii), and the last step follows from Assumption 6(i). This completes the proof of (14).

□

Lemma 5. *Let* $||\hat{h}-h_0||_\infty = o_p(1)$ *and let Assumptions 5(i), 5(ii), 5(iii), 5(v), 5(vi), 6, 7, 8(iv) hold. Then*

$$\left|\frac{\hat{v}_T(f)}{v_T(f)} - 1\right| = o_p(1). \tag{22}$$

Proof. Following the proof of Lemma G.2 in CC, we write

$$\frac{\hat{v}_T^2(f)}{v_T^2(f)} - 1 = \frac{(\hat{\gamma}_T - \gamma_T)' \Omega^o (\hat{\gamma}_T + \gamma_T)}{v_T^2(f)} + \frac{\hat{\gamma}_T' \left(\hat{\Omega}^o - \Omega^o\right)\hat{\gamma}_T}{v_T^2(f)} \equiv T_1 + T_2$$

where

$$\hat{\Omega}^o = G_A^{-1/2}\hat{\Omega}G_A^{-1/2} \qquad \hat{\gamma}_T = G_A^{1/2}\hat{G}_A^{-1}\hat{L}\left(\hat{L}'\hat{G}_A^{-1}\hat{L}\right)^{-1} Df\left(\hat{h}\right)[\psi_M]$$

$$\Omega^o = G_A^{-1/2}\Omega G_A^{-1/2} \qquad \gamma_T = G_A^{-1/2}S\left(S'G_A^{-1}S\right)^{-1} Df(h_0)[\psi_M]$$

and note that $\frac{||\gamma_T||^2}{v_T^2(f)} = \sum_{j=1}^J \lambda_j(T)^2$.
We consider T_1 and T_2 in turn. Note that

$$\begin{aligned}
\frac{||\hat{\gamma}_T - \gamma_T||}{v_T(f)} &= \frac{1}{v_T(f)}\left|\left|Df\left(\hat{h}\right)[\psi_M]'\left(\hat{G}_A^{-1/2}\hat{L}\right)_l^- \hat{G}_A^{-1/2}G_A^{1/2} - Df(h_0)[\psi_M]'\left(G_A^{-1/2}S\right)_l^-\right|\right| \\
&\le \frac{1}{v_T(f)}\left|\left|Df\left(\hat{h}\right)[\psi_M]'\left(G_A^{-1/2}S\right)_l^-\right|\right| \times \left|\left|G_A^{-1/2}S\left\{\left(\hat{G}_A^{-1/2}\hat{L}\right)_l^- \hat{G}_A^{-1/2}G_A^{1/2} - \left(G_A^{-1/2}S\right)_l^-\right\}\right|\right| \\
&\quad + \frac{1}{v_T(f)}\left|\left|\left(Df\left(\hat{h}\right)[\psi_M]' - Df(h_0)[\psi_M]'\right)\left(G_A^{-1/2}S\right)_l^-\right|\right| \equiv T_1^{(1)} \times T_1^{(2)} + T_1^{(3)}
\end{aligned}$$

Now,

$$\begin{aligned}
T_1^{(1)} &\le \frac{1}{v_T(f)}\left|\left|\left(Df\left(\hat{h}\right)[\psi_M]' - Df(h_0)[\psi_M]'\right)\left(G_A^{-1/2}S\right)_l^-\right|\right| + \frac{J}{v_T(f)}\max_{i\in\mathcal{J}}\left|\left|Df(h_0)\left[\psi_{M_i}^{(i)}\right]'\left(G_{A,i}^{-1/2}L_i\right)_l^-\right|\right| \\
&= O_p(1)
\end{aligned}$$

where the last step follows from Assumption 8(iv) and Lemma 8. Further, $T_1^{(2)} = o_p(1)$ by Lemmas F.10(c) and A.1 in CC and Assumption 6(i),and $T_1^{(3)} = o_p(1)$ by Assumption 8(iv). This implies that

$$\frac{||\hat{\gamma}_T - \gamma_T||}{v_T(f)} = o_p(1). \tag{23}$$

Therefore, by Cauchy-Schwarz

$$|T_1| \le \frac{||\hat{\gamma}_T - \gamma_T||}{v_T(T)} \times ||\Omega^o|| \times \frac{||\hat{\gamma}_T + \gamma_T||}{v_T(T)} \le \frac{||\hat{\gamma}_T - \gamma_T||}{v_T(T)} \times ||\Omega^o|| \times \left(\frac{||\hat{\gamma}_T - \gamma_T|| + 2||\gamma_T||}{v_T(T)}\right) = o_p(1)$$

where in the last step we also use Lemma 8 and the fact that $||\Omega^o|| < \infty$ by Assumption 5(i), 5(ii), 5(iii).

Turning to $|T_2|$, note that

$$\begin{aligned}
|T_2| &\leq \frac{||\hat{\gamma}_T||}{v_T(T)} \times ||\hat{\Omega}^o - \Omega^o|| \times \frac{||\hat{\gamma}_T||}{v_T(T)} \\
&\leq \frac{||\hat{\gamma}_T - \gamma_T|| + ||\gamma_T||}{v_T(T)} \times ||\hat{\Omega}^o - \Omega^o|| \times \frac{||\hat{\gamma}_T - \gamma_T|| + ||\gamma_T||}{v_T(T)} \\
&= O_p(1) \times ||\hat{\Omega}^o - \Omega^o|| \times O_p(1)
\end{aligned}$$

where the last step follows again from Lemma 8 and (23). We complete the proof by showing that $||\hat{\Omega}^o - \Omega^o|| = O_p(1)$. Note that

$$\Omega^o = \begin{bmatrix} \Omega^o_{11} & \Omega^o_{12} & \cdots & \Omega^o_{1J} \\ \Omega^o_{21} & \Omega^o_{22} & \cdots & \Omega^o_{2J} \\ \vdots & \vdots & \ddots & \cdots \\ \Omega^o_{J1} & \Omega^o_{J2} & \cdots & \Omega^o_{JJ} \end{bmatrix} \qquad \hat{\Omega}^o = \begin{bmatrix} \hat{\Omega_{11}}^o & \hat{\Omega_{12}}^o & \cdots & \hat{\Omega_{1J}}^o \\ \hat{\Omega_{21}}^o & \hat{\Omega_{22}}^o & \cdots & \hat{\Omega_{2J}}^o \\ \vdots & \vdots & \ddots & \cdots \\ \hat{\Omega_{J1}}^o & \hat{\Omega_{J2}}^o & \cdots & \hat{\Omega_{JJ}}^o \end{bmatrix}$$

where, for $j, k \in \mathcal{J}$,

$$\Omega^o_{jk} = G^{-1/2}_{A,j} \Omega_{jk} G^{-1/2}_{A,k} \qquad \hat{\Omega}^o_{jk} = G^{-1/2}_{A,j} \hat{\Omega}_{jk} G^{-1/2}_{A,k}$$

Using this notation, we have that, for some $v = [v_1' \cdots v_J']'$, with $v_i \in \mathbb{R}^{K_i}$, $j \in \mathcal{J}$, and $||v|| = 1$,

$$\begin{aligned}
||\hat{\Omega}^o - \Omega^o|| &= \sum_{j=1}^{J} v_j' \left(\hat{\Omega}^o_{jj} - \Omega^o_{jj}\right) v_j + 2 \sum_{j=1}^{J} \sum_{k=1}^{j-1} v_j' \left(\hat{\Omega}^o_{jk} - \Omega^o_{jk}\right) v_k \\
&\leq J \max_{j \in \mathcal{J}} ||\hat{\Omega}^o_{jj} - \Omega^o_{jj}|| + 2J^2 \max_{j,k \in \mathcal{J}, j \neq k} ||\hat{\Omega}^o_{jk} - \Omega^o_{jk}||
\end{aligned}$$

where each step uses the definition of matrix norm and the second step also uses Cauchy-Schwarz and the fact that $||v_j|| \leq 1$ for all j. Now, $||\hat{\Omega}^o_{ii} - \Omega^o_{ii}|| = O_p(1)$ for $i \in \mathcal{J}$ by Lemma G.3 in CC.For the third term, note that, by the triangle inequality, for all $j, k \in \mathcal{J}, j \neq k$,

$$\begin{aligned}
||\hat{\Omega}^o_{jk} - \Omega^o_{jk}|| \leq & \left|\left| G^{-1/2}_{A,j} \left[\frac{1}{T} \sum_{t=1}^{T} \xi_{j,t} \xi_{k,t} a^{(j)}_{K_j}(w_t) \, a^{(j)}_{K_j}(w_t)' - \mathbb{E}\left(\xi_j \xi_k a^{(j)}_{K_j}(W) \, a^{(j)}_{K_j}(Z)'\right) \right] G^{-1/2}_{A,j} \right|\right| \\
& + \left|\left| G^{-1/2}_{A,j} \frac{1}{T} \sum_{t=1}^{T} \left[\left(\hat{\xi}_{j,t} - \xi_{j,t}\right) \xi_{k,t} a^{(j)}_{K_j}(w_t) \, a^{(j)}_{K_j}(w_t)' \right] G^{-1/2}_{A,j} \right|\right| \\
& + \left|\left| G^{-1/2}_{A,j} \frac{1}{T} \sum_{t=1}^{T} \left[\left(\hat{\xi}_{j,t} - \xi_{j,t}\right) \left(\hat{\xi}_{k,t} - \xi_{k,t}\right) a^{(j)}_{K_j}(w_t) \, a^{(j)}_{K_j}(w_t)' \right] G^{-1/2}_{A,j} \right|\right| \\
& + \left|\left| G^{-1/2}_{A,j} \frac{1}{T} \sum_{t=1}^{T} \left[\xi_{1,t} \left(\hat{\xi}_{2,t} - \xi_{2,t}\right) a^{(j)}_{K_j}(w_t) \, a^{(j)}_{K_j}(w_t)' \right] G^{-1/2}_{A,j} \right|\right| \\
& \equiv ||T_{\Omega,1}|| + ||T_{\Omega,2}|| + ||T_{\Omega,3}|| + ||T_{\Omega,4}||
\end{aligned}$$

where we use the fact that $G_{A,j} = G_{A,k}$ and $a^{(j)}_{K_j} = a^{(k)}_{K_k}$ for all $j, k \in \mathcal{J}$ by Assumption 7. Using Lemma 9 below, we obtain $||T_{\Omega,1}|| = O_p(1)$. Further, $||T_{\Omega,2}|| = O_p(1)$ by $\left(\hat{\xi}_{j,t} - \xi_{j,t}\right) \xi_{k,t} \leq ||\hat{h}_j - h_{0,j}||_\infty \left(1 + \xi^2_{k,t}\right)$ and Lemma F.7 in CC.Similarly, $||T_{\Omega,4}|| = O_p(1)$. Finally, $||T_{\Omega,3}|| = O_p(1)$ by $\left(\hat{\xi}_{j,t} - \xi_{j,t}\right) \left(\hat{\xi}_{k,t} - \xi_{k,t}\right) \leq$

$||\hat{h}_j - h_{0,j}||_\infty^2$ and Lemma F.7 in CC.

□

Remark 1. *Note that I do not impose Assumption 4(i) in CC. This is because the assumption is automatically satisfied if the basis functions used for the sieve space and those used for the instrument space are both Riesz bases for the conditional expectation operator. I follow CC in assuming that this is the case.*

C.3 Theorem 3: Price elasticity functionals

We now focus on the case where the functional f is the own-price price elasticity of good 1 at a fixed $(\bar{s}, \bar{p}) \equiv (\bar{s}_1, \bar{s}_2, \bar{p}_1, \bar{p}_2)$ and Bernstein polynomials are used for both the sieve space and the instrument space. The goal is to provide sufficient, lower-level conditions for Theorem 2. Analogous arguments hold for the own-price elasticity of good 2 and for the cross price elasticities.

The functional of interest takes the form

$$f_\epsilon (h_0) = -\frac{p_1}{s_1} \frac{\frac{\partial h_{0,2}(\bar{s},\bar{p})}{\partial s_2} \frac{\partial h_{0,1}(\bar{s},\bar{p})}{\partial p_1} - \frac{\partial h_{0,1}(\bar{s},\bar{p})}{\partial s_2} \frac{\partial h_{0,2}(\bar{s},\bar{p})}{\partial p_1}}{\frac{\partial h_{0,1}(\bar{s},\bar{p})}{\partial s_1} \frac{\partial h_{0,2}(\bar{s},\bar{p})}{\partial s_2} - \frac{\partial h_{0,1}(\bar{s},\bar{p})}{\partial s_2} \frac{\partial h_{0,2}(\bar{s},\bar{p})}{\partial s_1}} \equiv -\frac{p_1}{s_1} \frac{N_1 - N_2}{D_1 - D_2} \tag{24}$$

We make the following assumptions.

Assumption 9. *(i) P has bounded support and (P, S) have densities bounded away from 0 and ∞;*
(ii) The basis used for both the sieve space and the instrument space is tensor-product Bernstein polynomials. Further, for the sieve space, the univariate Bernstein polynomials all have the same degree $M^{1/4}$;
(iii) $h_0 = [h_{0,1}, h_{0,2}]$ where $h_{0,1}$ and $h_{0,2}$ belong to the Hölder ball of smoothness $r \geq 4$ and finite radius L, and the order of the tensor-product Bernstein polynomials used for the sieve space is greater than r;
(iv) $M^{\frac{2+\gamma^{(1)}}{2\gamma^{(1)}}} \sqrt{\frac{\log T}{T}} = o(1)$ and $M^{\frac{1+\gamma^{(2)}}{2\gamma^{(2)}}} \sqrt{\frac{\log T}{T}} = o(1)$, where $\gamma^{(1)}, \gamma^{(2)} > 0$ are defined in Assumption 5(v)-5(vi);
(v) $\frac{\sqrt{T}}{v_T(f_\epsilon)} \times \left(M^{\frac{3-r}{4}} + \tau_M^2 M^{9/4} \frac{\log M}{T} \right) = o(1)$.

Discussion of Assumptions. Assumptions 9(i), 9(iii) and 9(iv) are regularity conditions sufficient to apply the sup-norm rate results in CC. 9(ii) is made for simplicity but it is not necessary. 9(v) corresponds to the second part of Assumption CS(v) in CC and is used to verify Assumption 8. More concrete sufficient conditions for Assumptions 9(iv) and 9(v) may be provided in specific settings. For example, Lemma 6 below gives sufficient conditions for the mildly ill-posed case.[62]

We now provide a proof of Theorem 3.

Proof of Theorem 3. We prove the statement by showing that the assumptions of Theorem 2 hold. Assumptions 5, 6(iii) and 7 are maintained. Assumption 6(i) is implied by Assumption 9(v), Lemma 10 and the fact that $\zeta = O_p\left(\sqrt{M}\right)$ for Bernstein polynomials.[63] Similarly, Assumption 6(ii) is implied by Assumption 9(iv) and $\zeta = O_p\left(\sqrt{M}\right)$.

We now verify Assumption 8. In what follows, unless otherwise specified, it is assumed that the arguments of all functions are $(\bar{s}, \bar{p})$ and the dependence is suppressed for notational convenience.

[62]See CC (p.15) for a formal definition of mild and severe ill-posedness.

[63]This follows from the fact that Bernstein polynomials are a special case of B-splines (see, e.g., Remark 4.7 on p.188 of Schumaker (2007).).

8(*i*) The pathwise derivative of f_ϵ in the direction $v \equiv (v_1, v_2)' \in \mathcal{H}$ is

$$Df_\epsilon(h_0)[v] \equiv \left.\frac{\partial f_\epsilon(h_0+\tau v)}{\partial \tau}\right|_{\tau=0} = \frac{p_1}{s_1}\left(C_1\frac{\partial v_2}{\partial s_2} + C_2\frac{\partial v_1}{\partial s_2} + C_3\frac{\partial v_1}{\partial p_1} + C_4\frac{\partial v_2}{\partial p_1} + C_5\frac{\partial v_2}{\partial s_1} + C_6\frac{\partial v_1}{\partial s_1}\right) \quad (25)$$

where

$$C_1 = -\frac{(D_1-D_2)\frac{\partial h_{0,1}}{\partial p_1} - (N_1-N_2)\frac{\partial h_{0,1}}{\partial s_1}}{(D_1-D_2)^2} \qquad C_2 = -\frac{-(D_1-D_2)\frac{\partial h_{0,2}}{\partial p_1} + (N_1-N_2)\frac{\partial h_{0,2}}{\partial s_1}}{(D_1-D_2)^2}$$

$$C_3 = -\frac{\frac{\partial h_{0,2}}{\partial s_2}}{(D_1-D_2)} \qquad C_4 = \frac{\frac{\partial h_{0,1}}{\partial s_2}}{(D_1-D_2)}$$

$$C_5 = -\frac{(N_1-N_2)\frac{\partial h_{0,1}}{\partial s_2}}{(D_1-D_2)^2} \qquad C_6 = \frac{(N_1-N_2)\frac{\partial h_{0,2}}{\partial s_2}}{(D_1-D_2)^2}$$

Therefore, $Df_\epsilon(h_0) : \mathcal{H} \mapsto \mathbb{R}$ is a linear functional.
Next, note that, for any $h = [h_1, h_2] \in \mathcal{H}_T$,

$$\begin{aligned}\left|\frac{\partial h_1}{\partial s_1} - \frac{\partial h_{0,1}}{\partial s_1}\right| &\leq \int_{-\infty}^{\overline{s}_2}\int_{-\infty}^{\overline{p}_1}\left|\frac{\partial^3 h_1}{\partial s_1\partial s_2\partial p_1}\left(\overline{s}_1, \underline{s}_2, \underline{p}_1, \overline{p}_2\right) - \frac{\partial^3 h_{0,1}}{\partial s_1\partial s_2\partial p_1}\left(\overline{s}_1, \underline{s}_2, \underline{p}_1, \overline{p}_2\right)\right| d\underline{s}_2 d\underline{p}_1 \\ &\leq constant\left|\left|\frac{\partial^3 h_1}{\partial s_1\partial s_2\partial p_1} - \frac{\partial^3 h_{0,1}}{\partial s_1\partial s_2\partial p_1}\right|\right|_\infty\end{aligned}$$

where the first inequality follows from the triangle inequality and the fundamental theorem of calculus, and the second inequality follows from assumption 9(i) and the fact that the support of (S_1, S_2) is the unit simplex and thus trivially bounded. By a similar argument, we can bound all the other derivatives in (25) and write

$$\begin{aligned}Df_\epsilon(h_0)[h-h_0] &\leq constant \times \max\left\{\left|\left|\frac{\partial^3 h_1}{\partial s_1\partial s_2\partial p_1} - \frac{\partial^3 h_{0,1}}{\partial s_1\partial s_2\partial p_1}\right|\right|_\infty, \left|\left|\frac{\partial^3 h_2}{\partial s_1\partial s_2\partial p_1} - \frac{\partial^3 h_{0,2}}{\partial s_1\partial s_2\partial p_1}\right|\right|_\infty\right\} \\ &\equiv constant\left|\left|\frac{\partial^3 h}{\partial s_1\partial s_2\partial p_1} - \frac{\partial^3 h_0}{\partial s_1\partial s_2\partial p_1}\right|\right|_\infty\end{aligned}$$

which shows that Assumption 8(i) holds with $\alpha = [1, 1, 1, 0]$.

8(*ii*) By the mean value theorem,

$$\begin{aligned}f_\epsilon\left(\hat{h}\right) - f_\epsilon(h_0) = \frac{p_1}{s_1}\Bigg[&\tilde{C}_1\left(\frac{\partial \hat{h}_2}{\partial s_2} - \frac{\partial h_{0,2}}{\partial s_2}\right) + \tilde{C}_2\left(\frac{\partial \hat{h}_1}{\partial s_2} - \frac{\partial h_{0,1}}{\partial s_2}\right) + \tilde{C}_3\left(\frac{\partial \hat{h}_1}{\partial p_1} - \frac{\partial h_{0,1}}{\partial p_1}\right) + \tilde{C}_4\left(\frac{\partial \hat{h}_2}{\partial p_1} - \frac{\partial h_{0,2}}{\partial p_1}\right) \\ &+ \tilde{C}_5\left(\frac{\partial \hat{h}_2}{\partial s_1} - \frac{\partial h_{0,2}}{\partial s_1}\right) + \tilde{C}_6\left(\frac{\partial \hat{h}_1}{\partial s_1} - \frac{\partial h_{0,1}}{\partial s_1}\right)\Bigg]\end{aligned}$$

$$\tilde{C}_1 = -\frac{\left(\tilde{D}_1 - \tilde{D}_2\right)\frac{\partial \tilde{h}_1}{\partial p_1} - \left(\tilde{N}_1 - \tilde{N}_2\right)\frac{\partial \tilde{h}_1}{\partial s_1}}{\left(\tilde{D}_1 - \tilde{D}_2\right)^2} \qquad \tilde{C}_2 = -\frac{-\left(\tilde{D}_1 - \tilde{D}_2\right)\frac{\partial \tilde{h}_2}{\partial p_1} + \left(\tilde{N}_1 - \tilde{N}_2\right)\frac{\partial \tilde{h}_2}{\partial s_1}}{\left(\tilde{D}_1 - \tilde{D}_2\right)^2}$$

$$\tilde{C}_3 = -\frac{\frac{\partial \tilde{h}_2}{\partial s_2}}{\left(\tilde{D}_1 - \tilde{D}_2\right)} \qquad \tilde{C}_4 = \frac{\frac{\partial \tilde{h}_1}{\partial s_2}}{\left(\tilde{D}_1 - \tilde{D}_2\right)}$$

$$\tilde{C}_5 = -\frac{\left(\tilde{N}_1 - \tilde{N}_2\right)\frac{\partial \tilde{h}_1}{\partial s_2}}{\left(\tilde{D}_1 - \tilde{D}_2\right)^2} \qquad \tilde{C}_6 = \frac{\left(\tilde{N}_1 - \tilde{N}_2\right)\frac{\partial \tilde{h}_2}{\partial s_2}}{\left(\tilde{D}_1 - \tilde{D}_2\right)^2}$$

where $\left[\frac{\partial \tilde{h}_1}{\partial p_1}, \frac{\partial \tilde{h}_1}{\partial s_1}, \frac{\partial \tilde{h}_1}{\partial s_2}, \frac{\partial \tilde{h}_2}{\partial p_1}, \frac{\partial \tilde{h}_2}{\partial s_1}, \frac{\partial \tilde{h}_2}{\partial s_2}\right]$ lies on the line segment between $\left[\frac{\partial h_{0,1}}{\partial p_1}, \frac{\partial h_{0,1}}{\partial s_1}, \frac{\partial h_{0,1}}{\partial s_2}, \frac{\partial h_{0,2}}{\partial p_1}, \frac{\partial h_{0,2}}{\partial s_1}, \frac{\partial h_{0,2}}{\partial s_2}\right]$ and
$\left[\frac{\partial \hat{h}_1}{\partial p_1}, \frac{\partial \hat{h}_1}{\partial s_1}, \frac{\partial \hat{h}_1}{\partial s_2}, \frac{\partial \hat{h}_2}{\partial p_1}, \frac{\partial \hat{h}_2}{\partial s_1}, \frac{\partial \hat{h}_2}{\partial s_2}\right]$ and $\tilde{N}_1, \tilde{N}_2, \tilde{D}_1, \tilde{D}_2$ are defined accordingly. Therefore, after some algebra, we obtain

$$\begin{aligned}\left|f_\epsilon\left(\hat{h}\right) - f_\epsilon(h_0) - Df_\epsilon(h_0)\left[\hat{h} - h_0\right]\right| \quad &\leq \quad F_1\left|\frac{\partial \hat{h}_2}{\partial s_2} - \frac{\partial h_{0,2}}{\partial s_2}\right| + F_2\left|\frac{\partial \hat{h}_1}{\partial s_2} - \frac{\partial h_{0,1}}{\partial s_2}\right| + F_3\left|\frac{\partial \hat{h}_1}{\partial p_1} - \frac{\partial h_{0,1}}{\partial p_1}\right| \\ &+ \quad F_4\left|\frac{\partial \hat{h}_2}{\partial p_1} - \frac{\partial h_{0,2}}{\partial p_1}\right| + F_5\left|\frac{\partial \hat{h}_2}{\partial s_1} - \frac{\partial h_{0,2}}{\partial s_1}\right| + F_6\left|\frac{\partial \hat{h}_1}{\partial s_1} - \frac{\partial h_{0,1}}{\partial s_1}\right|\end{aligned}$$

where $(F_i)_{i=1}^6$ are linear combinations (with finite coefficients) of $||\partial^{\tilde{\alpha}}\hat{h} - \partial^{\tilde{\alpha}}h_0||_\infty$ for vectors $\tilde{\alpha}$ with $|\tilde{\alpha}| = 1$. Thus

$$\left|f_\epsilon\left(\hat{h}\right) - f_\epsilon(h_0) - Df_\epsilon(h_0)\left[\hat{h} - h_0\right]\right| \leq constant||\partial^{\alpha_1}\hat{h} - \partial^{\alpha_1}h_0||_\infty||\partial^{\alpha_2}\hat{h} - \partial^{\alpha_2}h_0||_\infty$$

for some α_1, α_2 with $|\alpha_1| = |\alpha_2| = 1$.

8(*iii*) Given the choice of $\alpha, \alpha_1, \alpha_2$ above and by Corollary 3.1 in CC, we have

$$\begin{aligned}\left|\left|\partial^{\alpha_1}\hat{h} - \partial^{\alpha_1}h_0\right|\right|_\infty\left|\left|\partial^{\alpha_2}\hat{h} - \partial^{\alpha_2}h_0\right|\right|_\infty + \left|\left|\partial^{\alpha}\tilde{h} - \partial^{\alpha}h_0\right|\right|_\infty \quad &= \quad O_p\left(\left[M^{\frac{1-r}{4}} + \tau_M M^{3/4}\sqrt{\log M/T}\right]^2\right) \\ &+ \quad O_p\left(M^{\frac{3-r}{4}}\right)\end{aligned}$$

Thus, Assumption 8(iii) is implied by Assumption 9(v) and Lemma 10.

8(*iv*) By Remark 4.1 in CC, a sufficient condition for Assumption 8(iv) is

$$T_{iv,\epsilon} \equiv \frac{\tau_M\sqrt{\sum_{m=1}^M\left(Df_\epsilon\left(\hat{h}\right)\left[\left(G_\psi^{-1/2}\psi_M\right)_m\right] - Df_\epsilon(h_0)\left[\left(G_\psi^{-1/2}\psi_M\right)_m\right]\right)^2}}{v_T(f_\epsilon)} = O_p(1)$$

where $\left(G_\psi^{-1/2}\psi_M\right)_m$ denotes the $m-$th row of the matrix $G_\psi^{-1/2}\psi_M$. Note that, after some algebra, we can write $Df_\epsilon\left(\hat{h}\right)\left[\left(G_\psi^{-1/2}\psi_M\right)_m\right] - Df_\epsilon(h_0)\left[\left(G_\psi^{-1/2}\psi_M\right)_m\right]$ for every m as the linear combination of terms, where each term is the difference between a first-order partial derivative of $\hat{h}_i$ and the same derivative of $h_{0,i}$ for some $i \in \{1, 2\}$, and each coefficient is a first-order partial derivative of an element of $\left(G_\psi^{-1/2}\psi_M\right)_m$.

Therefore, we can write

$$\begin{aligned} T_{iv,\epsilon} &\lesssim \frac{\tau_M}{\upsilon_T(f_\epsilon)} \sqrt{M^{3/2}\zeta^2 \times \left[O_p\left(M^{(1-r)/4} + \tau_M M^{3/4}\sqrt{\log M/T}\right)\right]^2} \\ &= O_p\left(\frac{\sqrt{T}}{\upsilon_T(f_\epsilon)} \times \left[\frac{\tau_M M}{\sqrt{T}} M^{(2-r)/4} + \frac{\tau_M^2 M^2 \sqrt{\log M}}{T}\right]\right) = O_p(1) \end{aligned}$$

where the first step uses Corollary 3.1 in CC and the proof of Lemma 10, the second step uses the fact that, under the maintained assumptions, $\zeta = O\left(\sqrt{M}\right)$, and the last step follows from Assumption 9(v) (which implies $\frac{\tau_M M}{\sqrt{T}} = o(1)$).

□

Lemma 6. *Let Assumptions 9(i) and 9(iii) hold, and let* $(\upsilon_T(f_\epsilon))^2 \asymp \tau_M^2 \sum_{m=1}^M m^a$ *for* $a \leq 0$. *Further, assume that* $\tau_M \asymp M^{\varsigma/2}$ *for* $\varsigma \geq 0, a+\varsigma+1 > 0$.[64] *Then, Assumptions 9(iv) and 9(v) are satisfied if* $M \asymp T^\rho$ *with*

$$\rho \in \left(\frac{2}{r-3+2(a+\varsigma+1)}, \min\left\{\frac{2}{2\varsigma - 2a + 7}, \frac{\gamma^{(1)}}{2+\gamma^{(1)}}, \frac{\gamma^{(2)}}{1+\gamma^{(2)}}\right\}\right)$$

Further, M *may be chosen to satisfy the latter condition if* $r + 4a - 8 > 0$ *and* $\gamma^{(i)}(r + 2a + 2\varsigma - 3) - 4 > 0$ *for* $i \in \{1, 2\}$.

Proof. As shown in CC, under the maintained assumptions, $(\upsilon_T(f_\epsilon))^2 \asymp M^{a+\varsigma+1}$. The result follows by inspection.

□

C.4 Theorem 4: Equilibrium price functionals

We now specialize Theorem 2 to the case where the functional f is the equilibrium price of good 1 in a market characterized by marginal costs $\overline{mc} \equiv (\overline{mc}_1, \overline{mc}_2)$ and indices $\overline{\delta} \equiv (\overline{\delta}_1, \overline{\delta}_2)$. I let $f_p \equiv [f_{p_1}, f_{p_2}] : \mathcal{H} \mapsto \mathbb{R}^2$ denote the functional that returns the equilibrium prices, so that the goal is to obtain the asymptotic distribution of the sieve estimator $f_{p_1}\left(\hat{h}\right)$. An analogous argument holds for the price of good 2. As in the rest of the appendix, I let $h_0 = [h_{0,1}, h_{0,2}]$ denote the inverse of the demand system σ_0. Further, I use $h_0^{-1} = \left[h_{0,1}^{-1}, h_{0,2}^{-1}\right] = [\sigma_{0,1}, \sigma_{0,2}]$ to denote the demand system itself. The equilibrium prices $\overline{p} \equiv (\overline{p}_1, \overline{p}_2) \equiv [f_{p_1}(h_0), f_{p_2}(h_0)]$ solve the firm's first-order conditions

$$\begin{bmatrix} g_1\left(\overline{\delta}, \overline{p}, \overline{mc}, h_0\right) \\ g_2\left(\overline{\delta}, \overline{p}, \overline{mc}, h_0\right) \end{bmatrix} \equiv -\left[\left(\mathbb{J}_{h_0}^s\right)^{-1} \mathbb{J}_{h_0}^p\right]' \begin{bmatrix} \overline{p}_1 - \overline{mc}_1 \\ \overline{p}_2 - \overline{mc}_2 \end{bmatrix} + \begin{bmatrix} h_{0,1}^{-1}\left(\overline{\delta}, \overline{p}\right) \\ h_{0,2}^{-1}\left(\overline{\delta}, \overline{p}\right) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{26}$$

where

$$\mathbb{J}_{h_0}^s \equiv \begin{bmatrix} \frac{\partial h_{0,1}\left(h_0^{-1}(\overline{\delta},\overline{p}),\overline{p}\right)}{\partial s_1} & \frac{\partial h_{0,1}\left(h_0^{-1}(\overline{\delta},\overline{p}),\overline{p}\right)}{\partial s_2} \\ \frac{\partial h_{0,2}\left(h_0^{-1}(\overline{\delta},\overline{p}),\overline{p}\right)}{\partial s_1} & \frac{\partial h_{0,2}\left(h_0^{-1}(\overline{\delta},\overline{p}),\overline{p}\right)}{\partial s_2} \end{bmatrix} \qquad \mathbb{J}_{h_0}^p \equiv \begin{bmatrix} \frac{\partial h_{0,1}\left(h_0^{-1}(\overline{\delta},\overline{p}),\overline{p}\right)}{\partial p_1} & \frac{\partial h_{0,1}\left(h_0^{-1}(\overline{\delta},\overline{p}),\overline{p}\right)}{\partial p_2} \\ \frac{\partial h_{0,2}\left(h_0^{-1}(\overline{\delta},\overline{p}),\overline{p}\right)}{\partial p_1} & \frac{\partial h_{0,2}\left(h_0^{-1}(\overline{\delta},\overline{p}),\overline{p}\right)}{\partial p_2} \end{bmatrix}$$

[64] This corresponds to the "mildly ill-posed" case discussed by CC in Corollary 5.1. CC also provide sufficient conditions for the maintained assumption on the rate of divergence of $\upsilon_T(f)$.

We make the following assumptions.

Assumption 10. *(i) P has bounded support and (P,S) have densities bounded away from 0 and ∞;*
(ii) The basis used for both the sieve space and the instrument space is tensor-product Bernstein polynomials. Further, for the sieve space, the univariate Bernstein polynomials all have the same degree $M^{1/4}$;
(iii) $h_0 = [h_{0,1}, h_{0,2}]$ where $h_{0,1}$ and $h_{0,2}$ belong to the Hölder ball of smoothness $r \geq 5$ and finite radius L, and the order of the tensor-product Bernstein polynomials used for the sieve space is greater than r;
(iv) $M^{\frac{2+\gamma^{(1)}}{2\gamma^{(1)}}}\sqrt{\frac{\log T}{T}} = o(1)$ and $M^{\frac{1+\gamma^{(2)}}{2\gamma^{(2)}}}\sqrt{\frac{\log T}{T}} = o(1)$, where $\gamma^{(1)}, \gamma^{(2)} > 0$ are defined in Assumption 5(v)-5(vi);
(v) $\frac{\sqrt{T}}{v_T(f_{p_1})} \times \left(M^{\frac{4-r}{4}} + \tau_M^2 M^{9/4}\frac{\log M}{T}\right) = o(1)$.

Discussion of Assumptions. Assumptions 10(i), 10(iii) and 10(iv) are regularity conditions sufficient to apply the sup-norm rate results in CC. 10(ii) is made for simplicity but it is not necessary. 10(v) corresponds to the second part of Assumption CS(v) in CC and is used to verify Assumption 8. More concrete sufficient conditions for Assumptions 10(iv) and 10(v) may be provided in specific settings. For example, Lemma 7 below gives sufficient conditions for the mildly ill-posed case.

We now provide a proof of Theorem 4.

Proof of Theorem 4. We prove the statement by showing that the assumptions of Theorem 2 hold. Assumptions 5, 6(iii) and 7 are maintained. Assumption 6(i) is implied by Assumption 10(v) , Lemma 10 and the fact that $\zeta = O_p\left(\sqrt{M}\right)$ for Bernstein polynomials.[65] Similarly, Assumption 6(ii) is implied by Assumption 10(iv) and $\zeta = O_p\left(\sqrt{M}\right)$.
We now verify Assumption 8.

8(*i*) Applying the implicit function theorem to (26),

$$Df_p(h)[v] = -\begin{bmatrix} \frac{\partial g_1\left(\overline{\delta},\overline{p},\overline{mc},h+\tau v\right)}{\partial p_1} & \frac{\partial g_1\left(\overline{\delta},\overline{p},\overline{mc},h+\tau v\right)}{\partial p_2} \\ \frac{\partial g_2\left(\overline{\delta},\overline{p},\overline{mc},h+\tau v\right)}{\partial p_1} & \frac{\partial g_2\left(\overline{\delta},\overline{p},\overline{mc},h+\tau v\right)}{\partial p_2} \end{bmatrix}^{-1} \begin{bmatrix} \frac{\partial g_1\left(\overline{\delta},\overline{p},\overline{mc},h+\tau v\right)}{\partial \tau} \\ \frac{\partial g_2\left(\overline{\delta},\overline{p},\overline{mc},h+\tau v\right)}{\partial \tau} \end{bmatrix}\Bigg|_{\tau=0} \equiv -\left(\mathbb{J}_g^p\right)^{-1}\mathbb{J}_g^\tau\Big|_{\tau=0} \qquad (27)$$

for all $h, v \in \mathcal{H}$. Now, note that $\mathbb{J}_g^p\big|_{\tau=0}$ does not depend on v, and that $\mathbb{J}_g^\tau\big|_{\tau=0}$ is a linear function of $v\left(h^{-1}\left(\overline{\delta},\overline{p}\right),\overline{p}\right)$ and its first derivatives, with coefficients that depend on derivatives of h of order 2 or lower, i.e. we can write

$$Df_{p_1}(h)[v] = \sum_{\tilde{\alpha}:|\tilde{\alpha}|\leq 1}\sum_{j=1}^{2} C_{\tilde{\alpha},j}\left(\overline{\delta},\overline{mc},\left\{\partial^\beta h : |\beta| \leq 2\right\}\right) \times \partial^{\tilde{\alpha}} v_j\left(h^{-1}\left(\overline{\delta},\overline{p}\right),\overline{p}\right) \qquad (28)$$

for real-valued functionals $C_{\tilde{\alpha},j}$. This shows that $Df_p(h_0)[v]$ is linear. Further, by the fundamental theorem of calculus, following an argument analogous to that in part 4(i) of the proof of Theorem 3, we obtain

$$|Df_{p_1}(h_0)[h-h_0]| \leq constant\left\|\frac{\partial^4 h}{\partial s_1\partial s_2\partial p_1\partial p_2} - \frac{\partial^4 h_0}{\partial s_1\partial s_2\partial p_1\partial p_2}\right\|_\infty$$

for all $h \in \mathcal{H}$. Therefore, Assumption 8(i) holds with $\alpha = [1,1,1,1]$.

[65] This follows from the fact that Bernstein polynomials are a special case of B-splines.

8(*ii*) As in part 4(ii) of the proof of Theorem 3, by the mean value theorem, we obtain

$$\left| f_{p_1}\left(\hat{h}\right) - f_{p_1}(h_0) - Df_{p_1}(h_0)[\hat{h}-h_0] \right| \leq$$
$$\sum_{\tilde{\alpha}:|\tilde{\alpha}|\leq 1} \sum_{j=1}^{2} \left[C_{\tilde{\alpha},j}\left(\overline{\delta}, \overline{mc}, \left\{\partial^\beta \hat{h} : |\beta| \leq 2\right\}\right) - C_{\tilde{\alpha},j}\left(\overline{\delta}, \overline{mc}, \left\{\partial^\beta h_0 : |\beta| \leq 2\right\}\right) \right] \times \left|\left| \partial^{\tilde{\alpha}} \hat{h}_j - \partial^{\tilde{\alpha}} h_{0,j} \right|\right|_\infty$$

Since, each of the $C_{\tilde{\alpha},j}\left(\overline{\delta}, \overline{mc}, \left\{\partial^\beta \hat{h} : |\beta| \leq 2\right\}\right) - C_{\tilde{\alpha},j}\left(\overline{\delta}, \overline{mc}, \left\{\partial^\beta h_0 : |\beta| \leq 2\right\}\right)$ terms may be bounded, after some algebra, by a linear combination of $\left\{||\partial^\beta \hat{h} - \partial^\beta h_0||_\infty : |\beta| \leq 2\right\}$, Assumption 8($ii$) holds with $|\alpha_1| = 1, |\alpha_2| = 2$.

8(*iii*) Given the choice of $\alpha, \alpha_1, \alpha_2$ above and by Corollary 3.1 in CC, we have

$$\begin{aligned} \left|\left|\partial^{\alpha_1}\hat{h} - \partial^{\alpha_1} h_0\right|\right|_\infty \left|\left|\partial^{\alpha_2}\hat{h} - \partial^{\alpha_2} h_0\right|\right|_\infty + \left|\left|\partial^{\alpha}\tilde{h} - \partial^{\alpha} h_0\right|\right|_\infty &= O_p\left(M^{\frac{3-2r}{4}} + \tau_M M^{\frac{5-r}{4}} \sqrt{\frac{\log M}{T}} + \tau_M^2 M^{\frac{7}{4}} \frac{\log M}{T}\right) \\ &+ O_p\left(M^{\frac{4-r}{4}}\right) \end{aligned}$$

Thus, Assumption 8(iii) is implied by Assumption 10(v) and Lemma 10.

8(*iv*) By Remark 4.1 in CC, a sufficient condition for Assumption 8(iv) is

$$T_{iv,p} \equiv \frac{\tau_M \sqrt{\sum_{m=1}^M \left(Df_{p_1}\left(\hat{h}\right)\left[\left(G_\psi^{-1/2}\psi_M\right)_m\right] - Df_{p_1}(h_0)\left[\left(G_\psi^{-1/2}\psi_M\right)_m\right]\right)^2}}{v_T(f_{p_1})} = O_p(1)$$

where $\left(G_\psi^{-1/2}\psi_M\right)_m$ denotes the $m-$th row of the matrix $G_\psi^{-1/2}\psi_M$. Note that, after some algebra, we can write $Df_{p_1}\left(\hat{h}\right)\left[\left(G_\psi^{-1/2}\psi_M\right)_m\right] - Df_{p_1}(h_0)\left[\left(G_\psi^{-1/2}\psi_M\right)_m\right]$ for every m as the linear combination of terms, where each term is the difference between a partial derivative of $\hat{h}_i$ of order at most 2 and the same derivative of $h_{0,i}$ for some $i \in \{1,2\}$, and each coefficient is a partial derivative of an element of $\left(G_\psi^{-1/2}\psi_M\right)_m$ of order at most 1. Therefore, we can write

$$\begin{aligned} T_{iv,p} &\lesssim \frac{\tau_M}{v_T(f_{p_1})} \sqrt{M^{3/2}\zeta^2 \times \left[O_p\left(M^{(2-r)/4} + \tau_M M \sqrt{\log M/T}\right)\right]^2} \\ &= O_p\left(\frac{\sqrt{T}}{v_T(f_{p_1})} \times \left[\frac{\tau_M M}{\sqrt{T}} M^{(3-r)/4} + \frac{\tau_M^2 M^{\frac{9}{4}}\sqrt{\log M}}{T}\right]\right) = O_p(1) \end{aligned}$$

where the first step uses Corollary 3.1 in CC and the proof of Lemma 10, the second step uses the fact that, under the maintained assumptions, $\zeta = O\left(\sqrt{M}\right)$, and the last step follows from Assumption 10(v) (which implies $\frac{\tau_M M}{\sqrt{T}} = o(1)$).

□

Lemma 7. *Let Assumptions 10(i) and 10(iii) hold, and let* $(v_T(f_\epsilon))^2 \asymp \tau_M^2 \sum_{m=1}^M m^a$ *for* $a \leq 0$. *Further, assume that* $\tau_M \asymp M^{\varsigma/2}$ *for* $\varsigma \geq 0, a + \varsigma + 1 > 0$.[66] *Then, Assumptions 10(iv) and 10(v) are satisfied if*

[66] This corresponds to the "mildly ill-posed" case discussed by CC in Corollary 5.1. CC also provide sufficient conditions for the maintained assumption on the rate of divergence of $v_T(f)$.

$M \asymp T^{\rho}$ with

$$\rho \in \left(\frac{2}{r-4+2\left(a+\varsigma+1\right)}, \min\left\{ \frac{2}{2\varsigma-2a+7}, \frac{\gamma^{(1)}}{2+\gamma^{(1)}}, \frac{\gamma^{(2)}}{1+\gamma^{(2)}} \right\} \right)$$

Further, M may be chosen to satisfy the latter condition if $r+4a-9>0$ and $\gamma^{(i)}\left(r+2a+2\varsigma-4\right)-4>0$ for $i \in \{1,2\}$.

Proof. As shown in CC, under the maintained assumptions, $\left(v_T\left(f_\epsilon\right)\right)^2 \asymp M^{a+\varsigma+1}$. The result follows by inspection.

□

C.5 Supplementary lemmas and proofs

Lemma 8. *For $i \in \mathcal{J}$, let $\lambda_i\left(T\right) \equiv \frac{\left|\left|Df(h_0)\left[\psi^{(i)}_{M_i}\right]\left(L_i' G_{A,i}^{-1} L_i\right)^{-1} L_i' G_{A,i}^{-1/2}\right|\right|}{v_T(f)}$ and let Assumption 5(ii) hold. Then,* $\limsup_{T\to\infty} \lambda_i\left(T\right) < \infty$.

Proof. Note that

$$\begin{aligned}
v_T^2\left(f\right) &= \sum_{i=1}^{J} Df\left(h_0\right)[\psi^{(i)}_{M_i}]' \left(L_i' G_{A,i}^{-1} L_i\right)^{-1} L_i' G_{A,i}^{-1} \Omega_{ii} G_{A,i}^{-1} L_i \left(L_i' G_{A,i}^{-1} L_i\right)^{-1} Df\left(h_0\right)[\psi^{(i)}_{M_i}] \\
&+ 2\sum_{j=1}^{J}\sum_{k=1}^{j-1} Df\left(h_0\right)[\psi^{(j)}_{M_j}]' \left(L_j' G_{A,j}^{-1} L_j\right)^{-1} L_j' G_{A,j}^{-1} \Omega_{jk} G_{A,k}^{-1} L_k \left(L_k' G_{A,k}^{-1} L_k\right)^{-1} Df\left(h_0\right)[\psi^{(k)}_{M_k}] \\
&\equiv \sum_{i=1}^{J} \sigma_{T,j}^2 + 2\sum_{j=1}^{J}\sum_{k=1}^{j-1} \sigma_{T,j,k}
\end{aligned}$$

Further, by Assumption 5(ii)

$$\left|\left|Df\left(h_0\right)\left[\psi^{(i)}_{M_i}\right]\left(L_i' G_{A,i}^{-1} L_i\right)^{-1} L_i' G_{A,i}^{-1/2}\right|\right|^2 \leq \underline{\sigma}^{-2}\sigma_{T,i}^2$$

for $i \in \mathcal{J}$. Therefore, we can write

$$\left[\lambda_i\left(T\right)\right]^2 \leq \frac{\underline{\sigma}^{-2}\sigma_{T,i}^2}{\sum_{i=1}^{J}\sigma_{T,j}^2 + 2\sum_{j=1}^{J}\sum_{k=1}^{j-1}\sigma_{T,j,k}}$$

from which the result follows.

□

Lemma 9. *Let Assumptions 5(iii), 5(vi) and 6(ii) hold. Then $||T_{\Omega,1}|| = O_p\left(1\right)$.*

Proof. The proof follows that of Lemma 3.1 in Chen and Christensen (2015) with minor changes. Let

$C_T \asymp \zeta^{\left(1+\gamma^{(2)}\right)/\gamma^{(2)}}$ be a sequence of positive numbers with $\gamma^{(2)}$ defined in Assumption 5(vi), and let

$$T_{\Omega,1}^{(1)} \equiv \frac{1}{T}\sum_{t=1}^{T}\left(\Xi_{1,t} - \mathbb{E}\left[\Xi_1\right]\right) \qquad\qquad T_{\Omega,1}^{(2)} \equiv \frac{1}{T}\sum_{t=1}^{T}\left(\Xi_{2,t} - \mathbb{E}\left[\Xi_2\right]\right)$$

where

$$\begin{aligned}
\Xi_{1,t} &\equiv \xi_{j,t}\xi_{k,t}G_{A,j}^{-1/2}a_{K_j}^{(k)}\left(w_t\right)a_{K_j}^{(j)}\left(w_t\right)'G_{A,j}^{-1/2}\mathbb{I}\left\{||\xi_{j,t}\xi_{k,t}G_{A,j}^{-1/2}a_{K_j}^{(j)}\left(w_t\right)a_{K_j}^{(j)}\left(w_t\right)'G_{A,j}^{-1/2}|| \leq C_T^2\right\}\\
\Xi_{2,t} &\equiv \xi_{j,t}\xi_{k,t}G_{A,j}^{-1/2}a_{K_j}^{(j)}\left(w_t\right)a_{K_j}^{(j)}\left(w_t\right)'G_{A,j}^{-1/2}\mathbb{I}\left\{||\xi_{j,t}\xi_{k,t}G_{A,j}^{-1/2}a_{K_j}^{(j)}\left(w_t\right)a_{K_j}^{(j)}\left(w_t\right)'G_{A,j}^{-1/2}|| > C_T^2\right\}
\end{aligned}$$

Note that $T_{\Omega,1} = T_{\Omega,1}^{(1)} + T_{\Omega,1}^{(2)}$, so that $||T_{\Omega,1}^{(1)}|| = O_p(1)$ and $||T_{\Omega,1}^{(2)}|| = O_p(1)$ imply the statement of the lemma.

Control of $||T_{\Omega,1}^{(1)}||$: By definition, $||\Xi_{1,t}|| \leq C_T^2$ and thus, by the triangle inequality and Jensen's inequality ($||\cdot||$ is convex), we have $||\Xi_{1,t} - \mathbb{E}\left(\Xi_1\right)|| \leq 2C_T^2$. Further,

$$\begin{aligned}
&\mathbb{E}\left[\Xi_1 - \mathbb{E}\left(\Xi_1\right)\right]^2 \leq\\
&\mathbb{E}\left[\xi_j^2\xi_k^2||G_{A,j}^{-1/2}a_{K_j}^{(j)}||^2 G_{A,j}^{-1/2}a_{K_j}^{(j)}\left(W\right)a_{K_j}^{(j)}\left(W\right)'G_{A,j}^{-1/2}\mathbb{I}\left\{||\xi_j\xi_k G_{A,j}^{-1/2}a_{K_j}^{(j)}\left(W\right)a_{K_j}^{(j)}\left(W\right)'G_{A,j}^{-1/2}|| \leq C_T^2\right\}\right] \leq\\
&C_T^2\mathbb{E}\left[|\xi_j\xi_k|G_{A,j}^{-1/2}a_{K_j}^{(j)}\left(W\right)a_{K_j}^{(j)}\left(W\right)'G_{A,j}^{-1/2}\mathbb{I}\left\{||\xi_j\xi_k G_{A,j}^{-1/2}a_{K_j}^{(j)}\left(W\right)a_{K_j}^{(j)}\left(W\right)'G_{A,j}^{-1/2}|| \leq C_T^2\right\}\right] \leq\\
&C_T^2\mathbb{E}\left[\mathbb{E}\left(|\xi_j\xi_k||W\right)G_{A,j}^{-1/2}a_{K_j}^{(j)}\left(W\right)a_{K_j}^{(j)}\left(W\right)'G_{A,j}^{-1/2}\right] \lesssim\\
&C_T^2\mathbb{E}\left[G_{A,j}^{-1/2}a_{K_j}^{(j)}\left(W\right)a_{K_j}^{(j)}\left(W\right)'G_{A,j}^{-1/2}\right] = C_T^2 I_{K_j}
\end{aligned}$$

where the inequalities are in the sense of positive semi-definite matrices. The first inequality follows from $\mathbb{E}\left[\Xi_1 - \mathbb{E}\left(\Xi_1\right)\right]^2 \leq \mathbb{E}\left(\Xi_1\right)^2$ (and the fact that matrix multiplication is commutative), the second follows from the fact that $||G_{A,j}^{-1/2}a_{K_j}^{(j)}\left(W\right)a_{K_j}^{(j)}\left(W\right)'G_{A,j}^{-1/2}|| = ||G_{A,j}^{-1/2}a_{K_j}^{(j)}\left(W\right)||^2$, the third from the law of iterated expectations, the fourth follows from Assumption 5(iii), and the last step is because $G_{A,j}^{-1/2}a_{K_j}^{(j)}$ is an orthonormal basis. Then, Corollary 4.1 in Chen and Christensen (2015) yields $||T_{\Omega,1}^{(1)}|| = O_p\left(C_T\sqrt{(\log K)/T}\right) = O_p(1)$, where the last step uses Assumption 6(ii).

Control of $||T_{\Omega,1}^{(2)}||$: Since $||\Xi_{2,t}|| \leq \zeta^2|\xi_{j,t}\xi_{k,t}|\mathbb{I}\left\{|\xi_{j,t}\xi_{k,t}| \geq C_T^2/\zeta^2\right\}$,by the triangle inequality and Jensen's inequality ($||\cdot||$ is convex), we have

$$\mathbb{E}\left[||T_{\Omega,1}^{(2)}||\right] \leq 2\zeta^2\mathbb{E}\left[|\xi_j\xi_k|\mathbb{I}\left\{|\xi_j\xi_k| \geq C_T^2/\zeta^2\right\}\right] \leq \frac{2\zeta^{2(1+\gamma^{(2)})}}{C_T^{2\gamma^{(2)}}}\mathbb{E}\left[|\xi_j\xi_k|^{1+\gamma^{(2)}}\mathbb{I}\left\{|\xi_j\xi_k| \geq C_T^2/\zeta^2\right\}\right] = o(1)$$

where the last step follows from Assumption 5(vi), the fact that $C_T^2/\zeta^2 \asymp \zeta^{2/\gamma^{(2)}} \to \infty$ and that $\zeta^{(1+\gamma^{(2)})}/C_T^{\gamma^{(2)}} \asymp 1$. Thus, $||T_{\Omega,1}^{(2)}|| = O_p(1)$ by Markov's inequality.

□

Lemma 10. *Let Assumption 9(i)-9(iii) hold. Then, for* $f \in \{f_\epsilon, f_{p_1}\}$

$$\left[v_T(f)\right]^2 \lesssim \tau_M^2 M^{5/2}$$

Proof. We prove this for $f = f_\epsilon$. The proof for $f = f_{p_1}$ is identical. As shown in CC,[67] the maintained

[67] See also Chen and Pouzo (2015).

assumption imply $[v_T(f_\epsilon)]^2 \lesssim \tau_M^2 \sum_{m=1}^{M} \left(Df_\epsilon(h_0) \left[\left(G_\psi^{-1/2} \psi_M \right)_m \right] \right)^2$, where $\left(G_\psi^{-1/2} \psi_M \right)_m$ denotes the $m-$th row of the $M-$by$-2-$valued function $G_\psi^{-1/2} \psi_M$. Next, note that

$$\left| Df_\epsilon(h_0) \left[\left(G_\psi^{-1/2} \psi_M \right)_m \right] \right| \lesssim \max_{\tilde{\alpha}:|\tilde{\alpha}|=1} ||\partial^{\tilde{\alpha}} \left(G_\psi^{-1/2} \psi_M \right)_m ||_\infty \leq 4M^{1/4} \sup_{(y)\in\mathcal{S}\times\mathcal{P}} ||G_\psi^{-1/2} \psi_M(y)||_\infty \leq 4M^{1/4}\zeta$$

where we use the fact that the first derivatives of Bernstein polynomials can be written in terms of lower order polynomials, and that the latter can in turn be written as a linear combination of higher order polynomials.Finally, using $\zeta = O\left(\sqrt{M}\right)$, we obtain the result.

□

Appendix D: Additional Monte Carlo Simulations

D.1 Loss aversion

Another type of consumer behavior allowed by the NPD model is one where the consumer disutility from paying a price for good j increases with how much more expensive j is relative to another good. I refer to this as "loss aversion" since it may be viewed as one instance of the pattern studied by Tversky and Kahneman (1991) that goes by this name. Specifically, I let the indirect utility from good 1 depend negatively not only on p_1 but also on $p_1 - p_2$, and similarly for good 2. The underlying idea is that if good j is more expensive than good k, paying p_j will be perceived as a loss and therefore will lead to a greater disutility than if good j were cheaper than good k. I set the coefficient on the price difference to -0.15; the simulation design is otherwise the same as that in Section 4.1. As in the previous simulations, we compare the performance of NPD with that of a mixed logit model. In this case, the latter is misspecified in that it only allows p_1, but not $p_1 - p_2$ to enter the utility of good 1, and similarly for good 2. In the NPD estimation, we impose the following constraints: monotonicity of σ^{-1}, diagonal dominance of $\mathbb{J}_\sigma^\delta$ and exchangeability.[68]

Figures 8 and 9 show the own- and cross-price elasticity functions, respectively. While NPD is on target, BLP tends to underestimate the magnitude of both and the discrepancy grows with price.

---

[68]See Section 3.2 and Appendix B for a discussion of these constraints.

Figure 8: Loss aversion: Own-price elasticity function

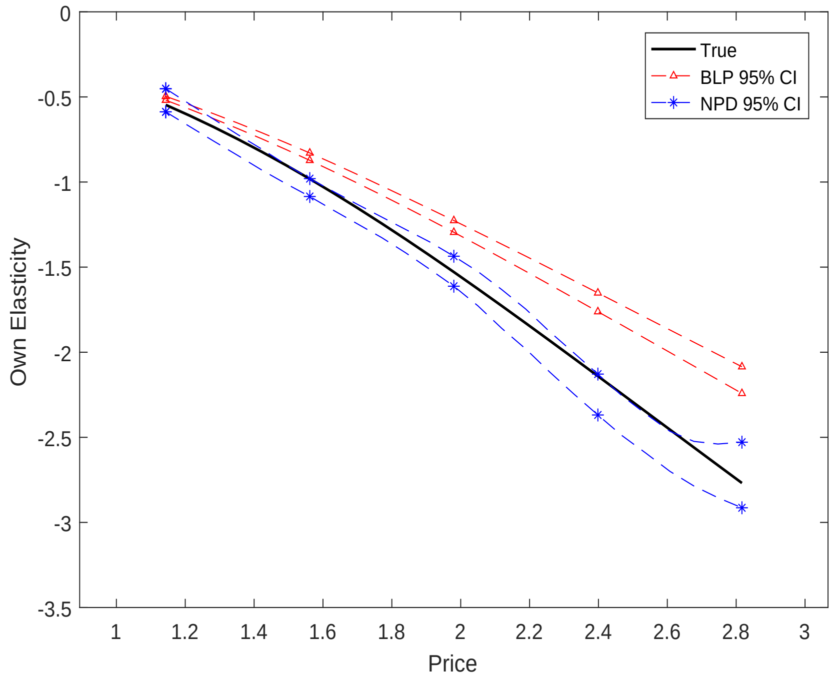
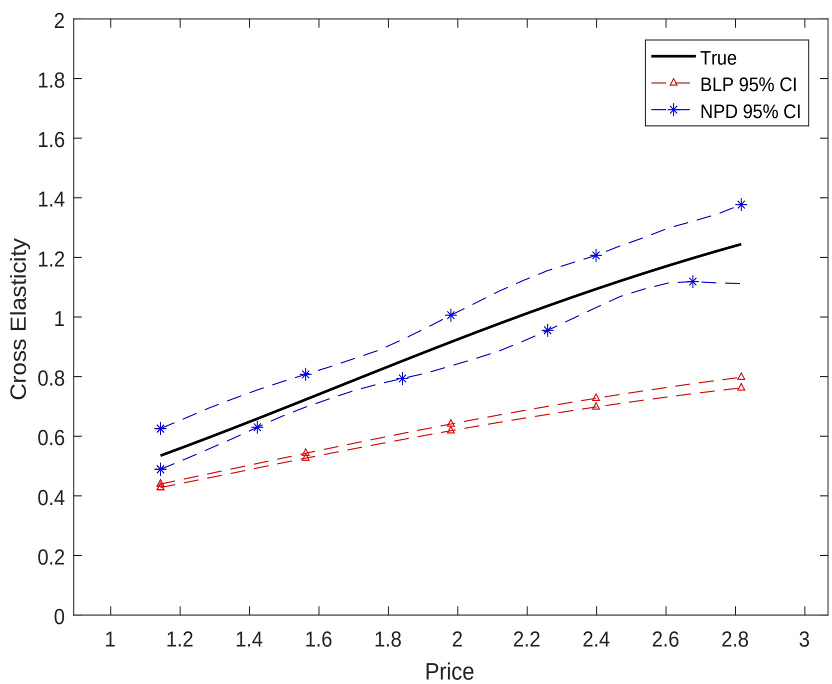


Figure 9: Loss aversion: Cross-price elasticity function



D.2 Smaller Sample Sizes

The simulations in Section 4 were based on sample sizes equal to 3,000. We now investigate how well the NPD estimator performs in a smaller sample size. Specifically, we focus on the complements example from Section 4.3 and repeat the simulation now using a sample of 500 observations.

Figure 10: Complements, $T = 500$: Own-price elasticity function

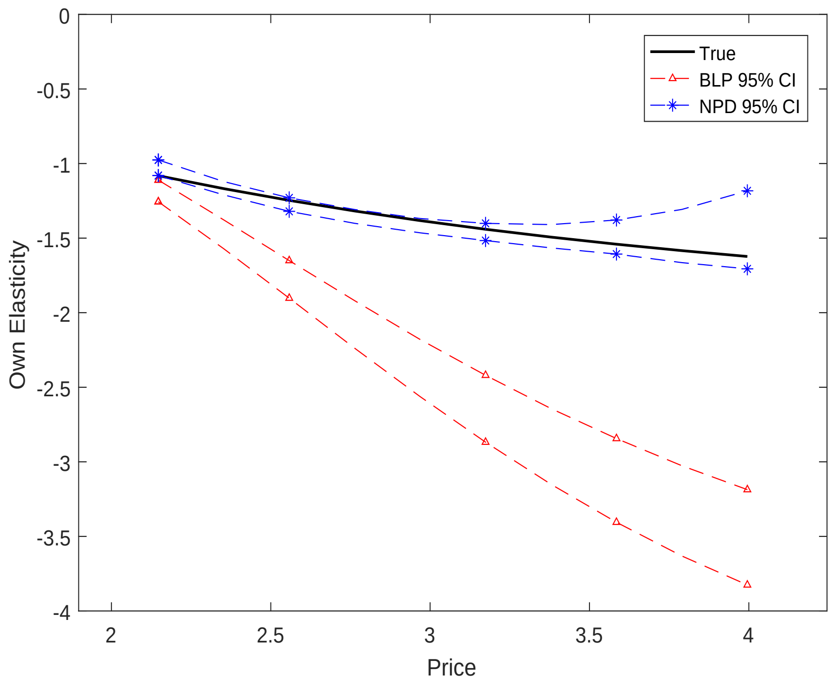
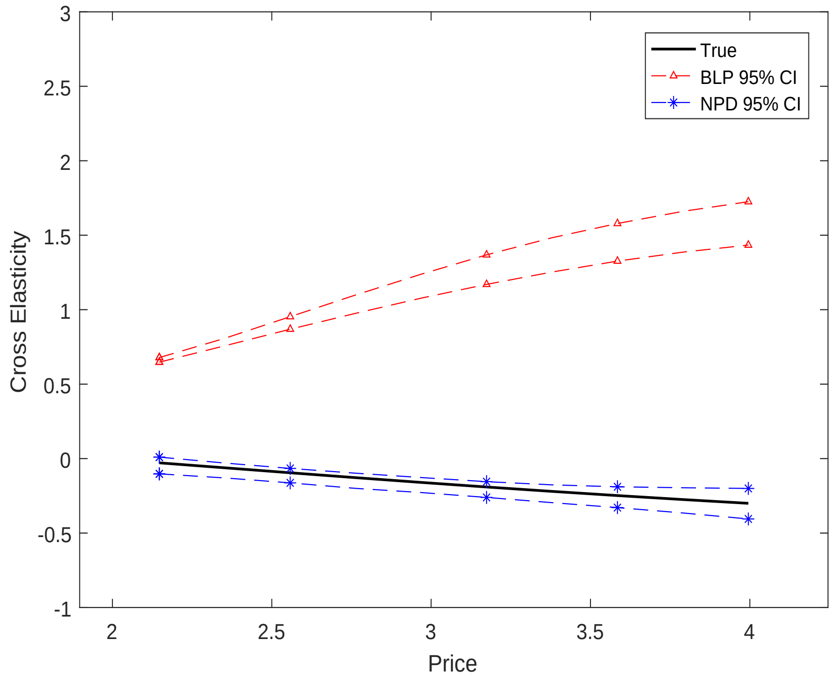


Figure 11: Complements, $T = 500$: Cross-price elasticity function



D.3 Violation of Index Restriction

The NPD estimator is based on the index restriction embedded in Assumption 1. Here, we explore how robust the estimator is to violations of this assumption. Specifically, we generate the data from the mixed logit dgp as in Section 4.1, except that we let the coefficient on the covariate x be random and distributed $N(1, \sigma_{viol})$. Because the coefficient on the unobservable ξ is not random, this induces a violation of the index restriction which becomes more severe as σ_{viol} increases. Figures 12 to 17 show that, except for the own-price elasticity function at large values of own-price, the NPD estimator is quite robust to violations of the index assumption for this dgp.

Figure 12: Violation of Index Restriction, $\sigma_{viol} = 0.10$: Own-price elasticity function

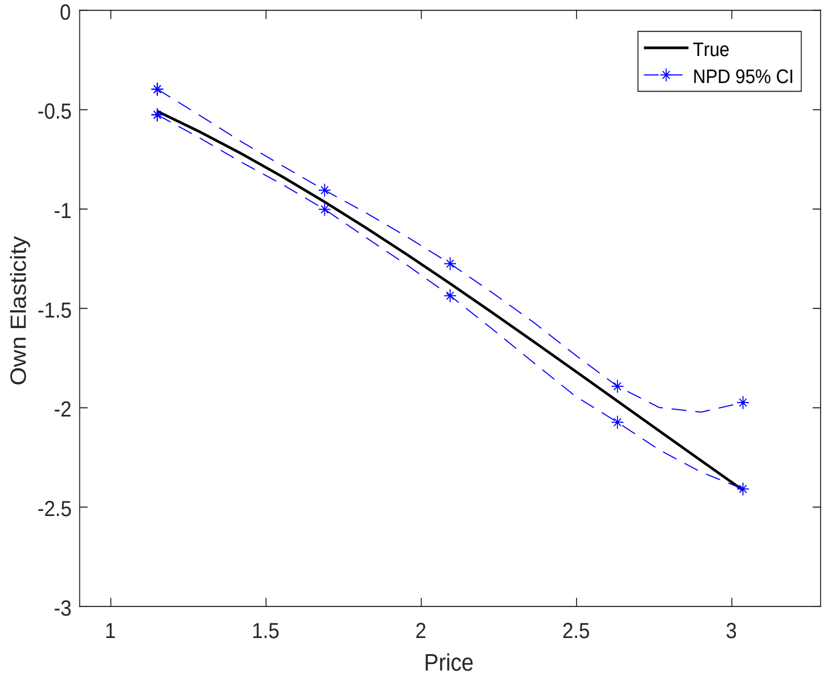


Figure 13: Violation of Index Restriction, $\sigma_{viol} = 0.10$: Cross-price elasticity function

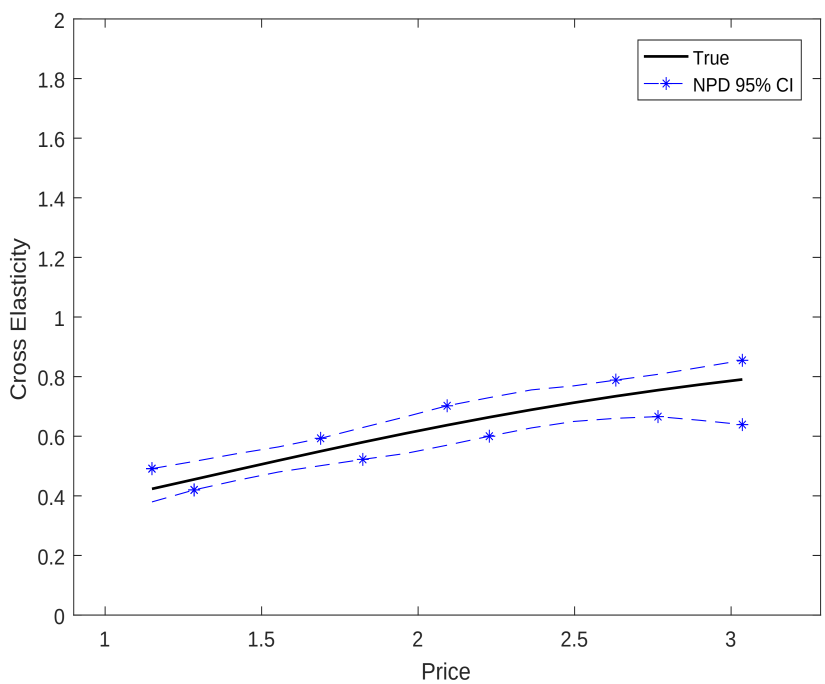


Figure 14: Violation of Index Restriction, $\sigma_{viol} = 0.50$: Own-price elasticity function

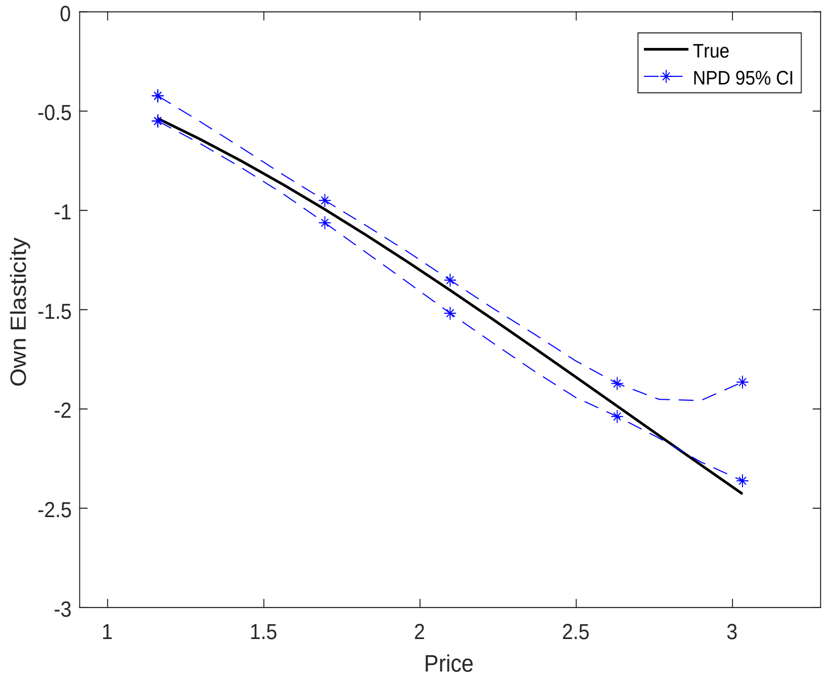


Figure 15: Violation of Index Restriction, $\sigma_{viol} = 0.50$: Cross-price elasticity function

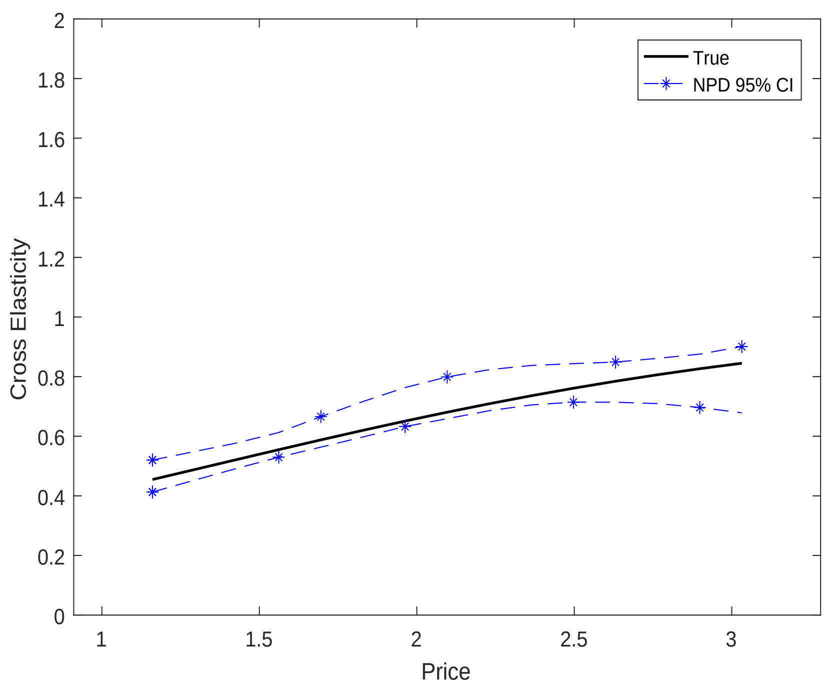


Figure 16: Violation of Index Restriction, $\sigma_{viol} = 1.50$: Own-price elasticity function

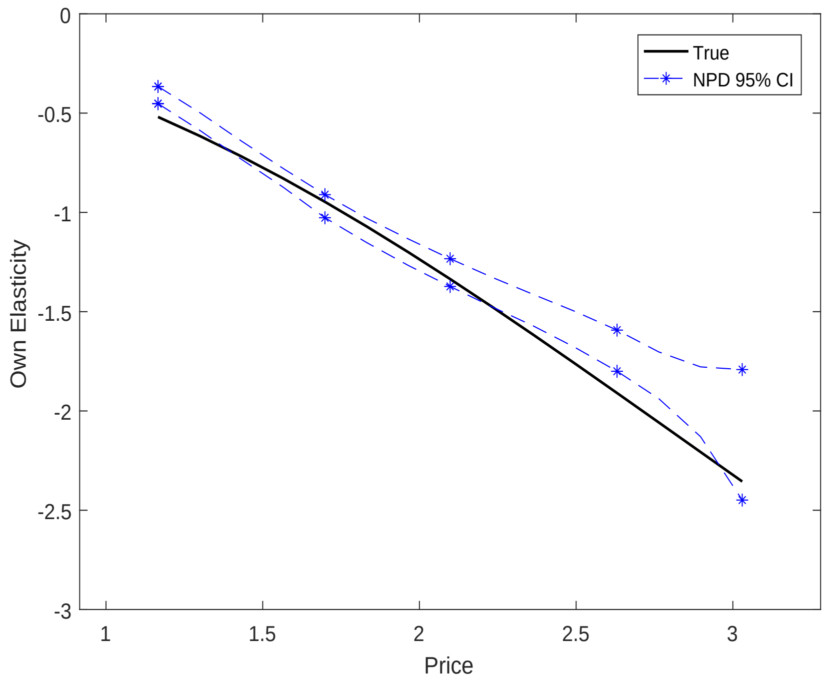
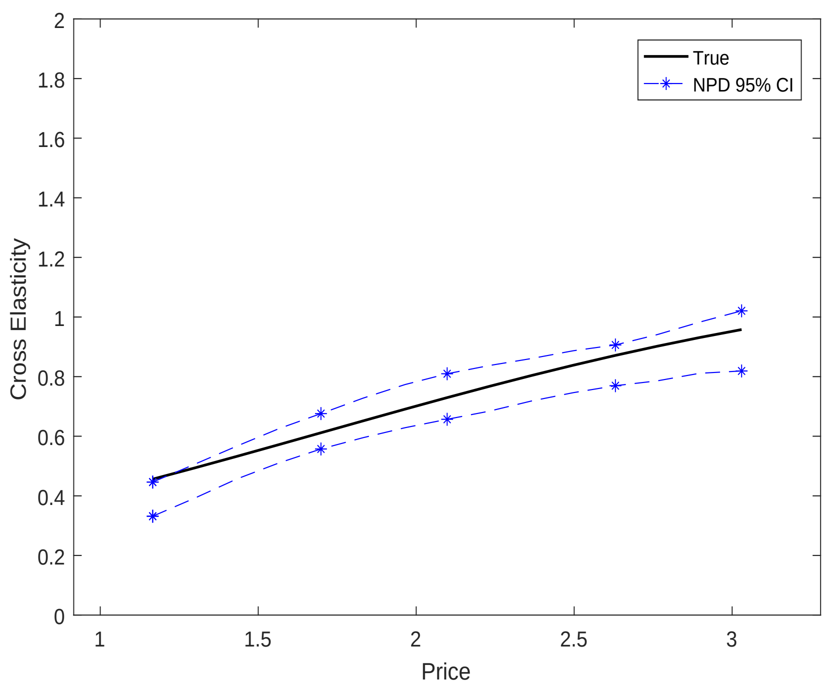


Figure 17: Violation of Index Restriction, $\sigma_{viol} = 1.50$: Cross-price elasticity function



D.4 Sensitivity to the Choice of Polynomial Degree

We now consider how the performance of the NPD estimator varies with the choice of polynomial degree for the Bernstein approximation to the unknown functions. We focus on the mixed logit dgp from Section 4.1 and the complements dgp from Section 4.3. For each dgp, we show how the estimated own- and cross-price elasticity functions vary with the polynomial degree used to approximate each unknown function.

D.4.1 Mixed logit dgp

Figure 18: Mixed Logit, degree = 16: Own-price elasticity function

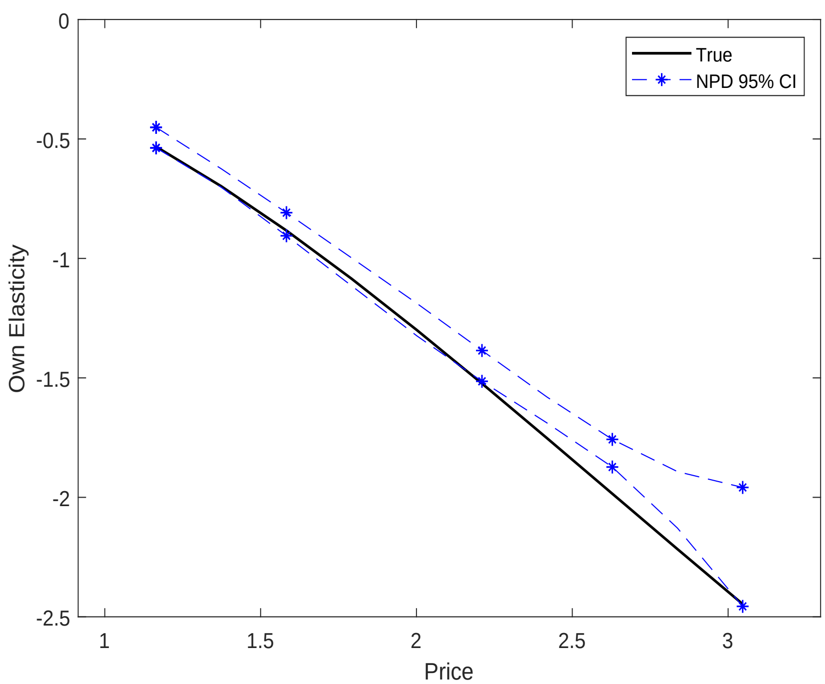


Figure 19: Mixed Logit, degree = 16: Cross-price elasticity function

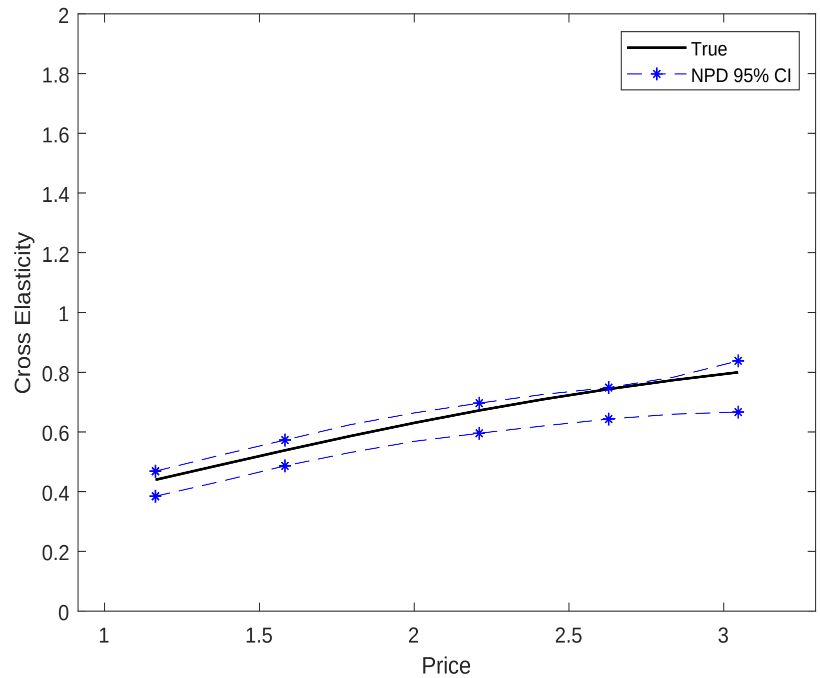


Figure 20: Mixed Logit, degree = 12: Own-price elasticity function

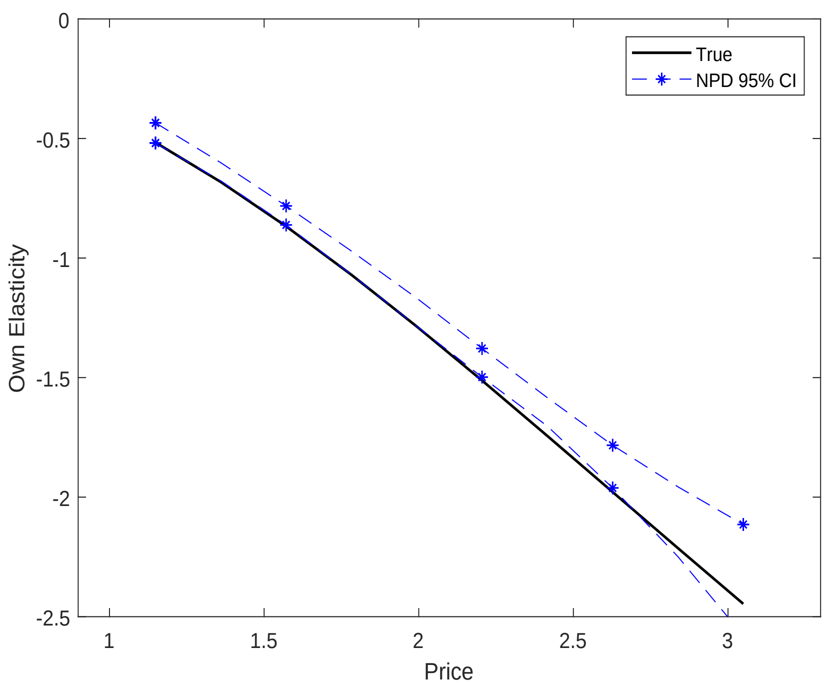


Figure 21: Mixed Logit, degree = 12: Cross-price elasticity function

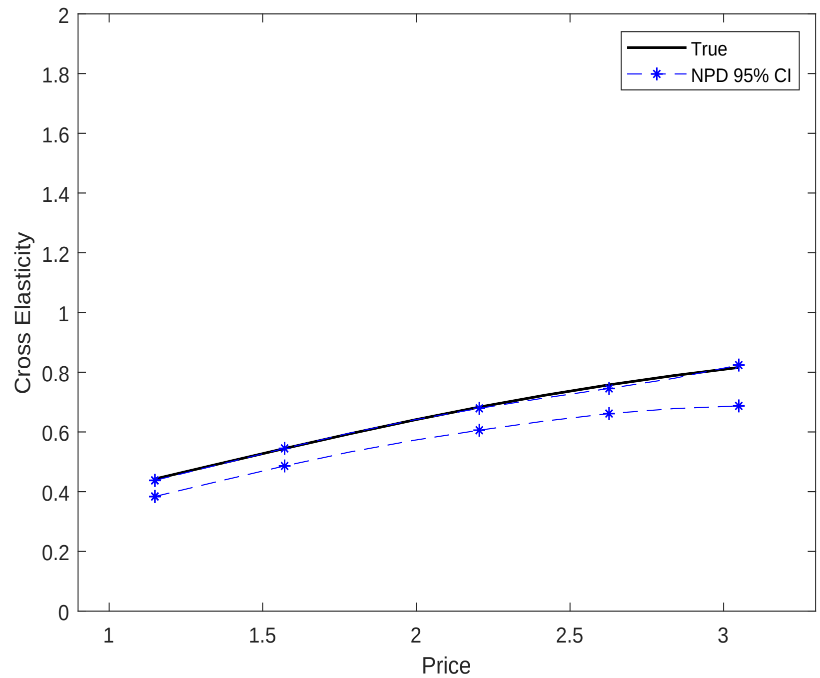


Figure 22: Mixed Logit, degree = 8: Own-price elasticity function

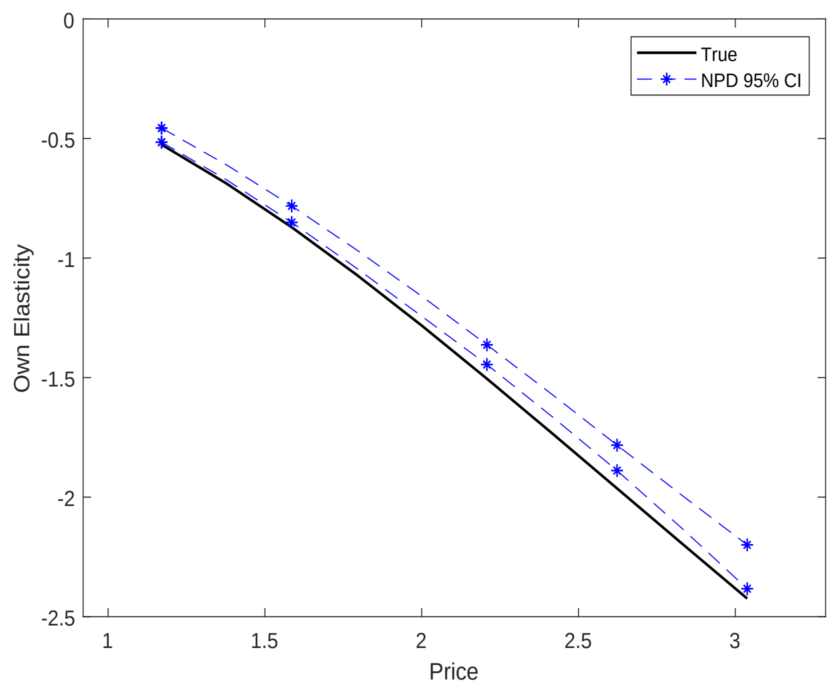


Figure 23: Mixed Logit, degree = 8: Cross-price elasticity function

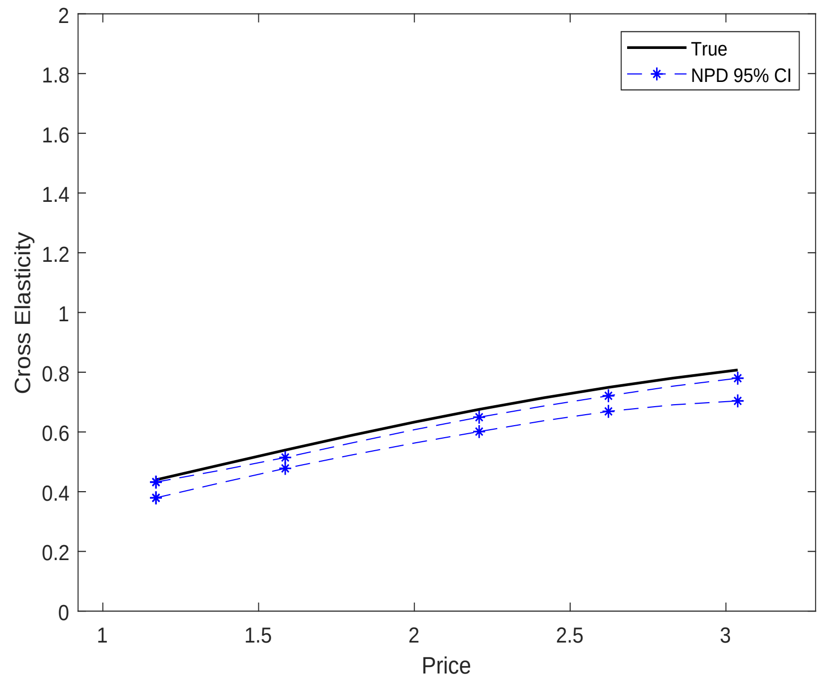


Figure 24: Mixed Logit, degree = 6: Own-price elasticity function

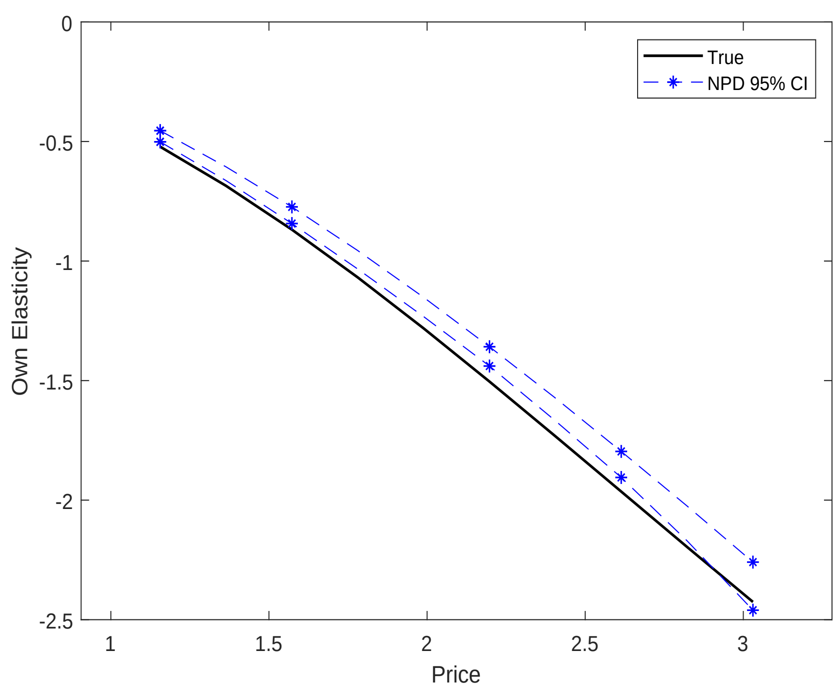


Figure 25: Mixed Logit, degree = 6: Cross-price elasticity function

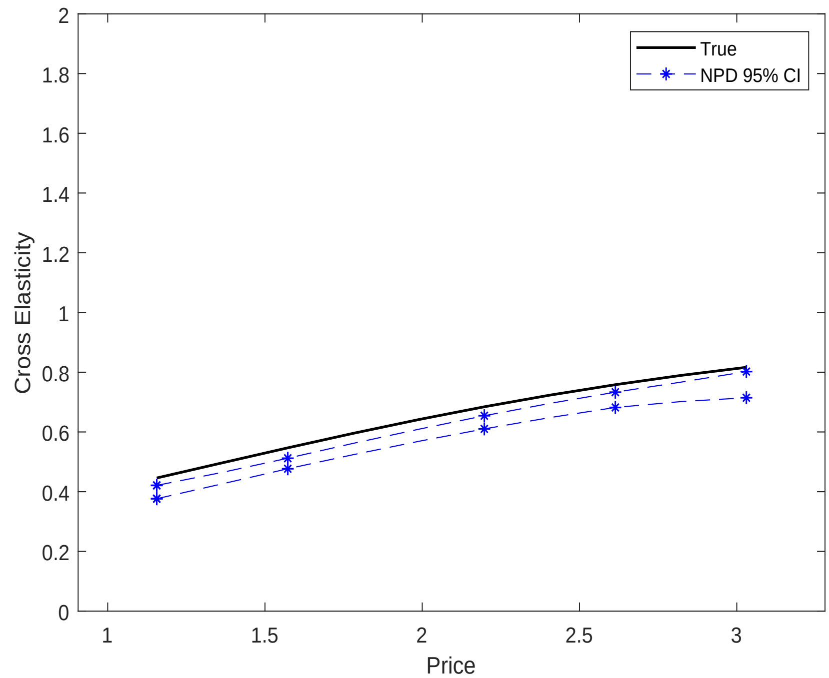


Figure 26: Mixed Logit, degree = 4: Own-price elasticity function

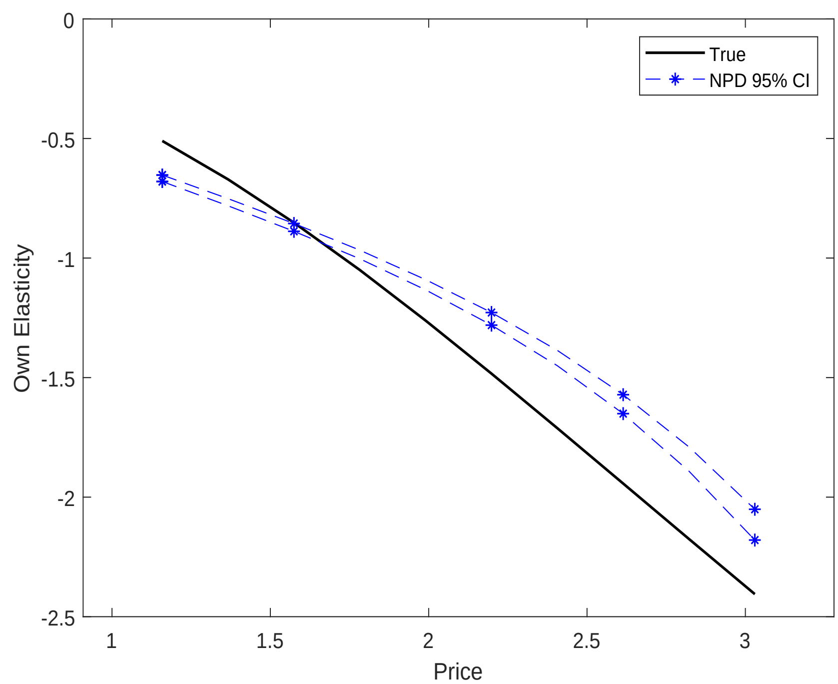
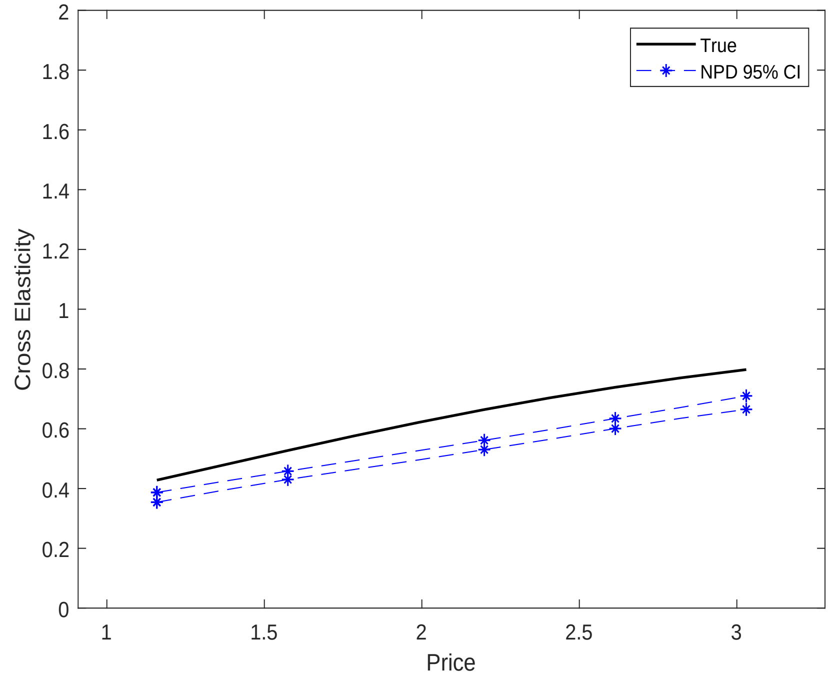


Figure 27: Mixed Logit, degree = 4: Cross-price elasticity function



D.4.2 Complements dgp

Figure 28: Complements, degree = 16: Own-price elasticity function

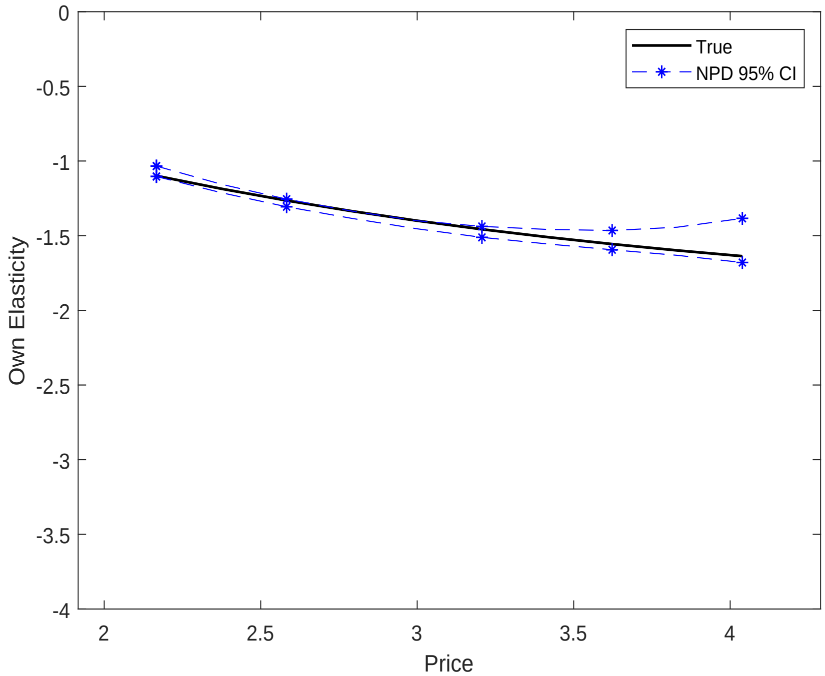


Figure 29: Complements, degree = 16: Cross-price elasticity function

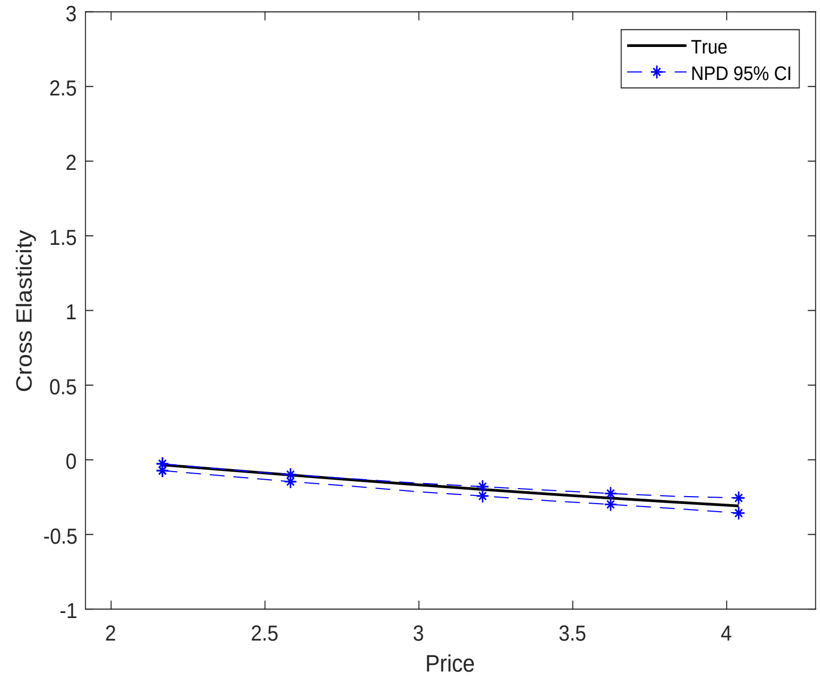


Figure 30: Complements, degree = 12: Own-price elasticity function

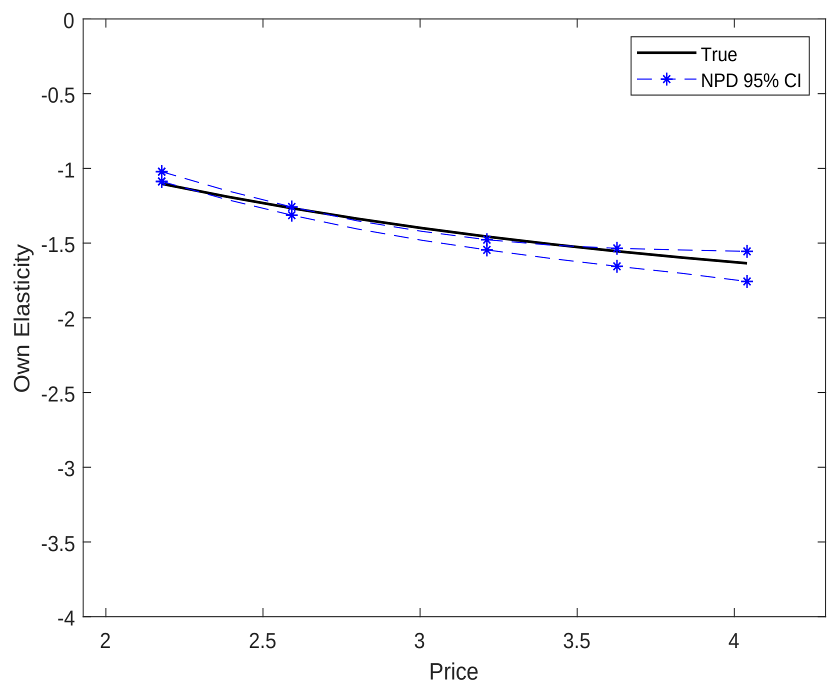


Figure 31: Complements, degree = 12: Cross-price elasticity function

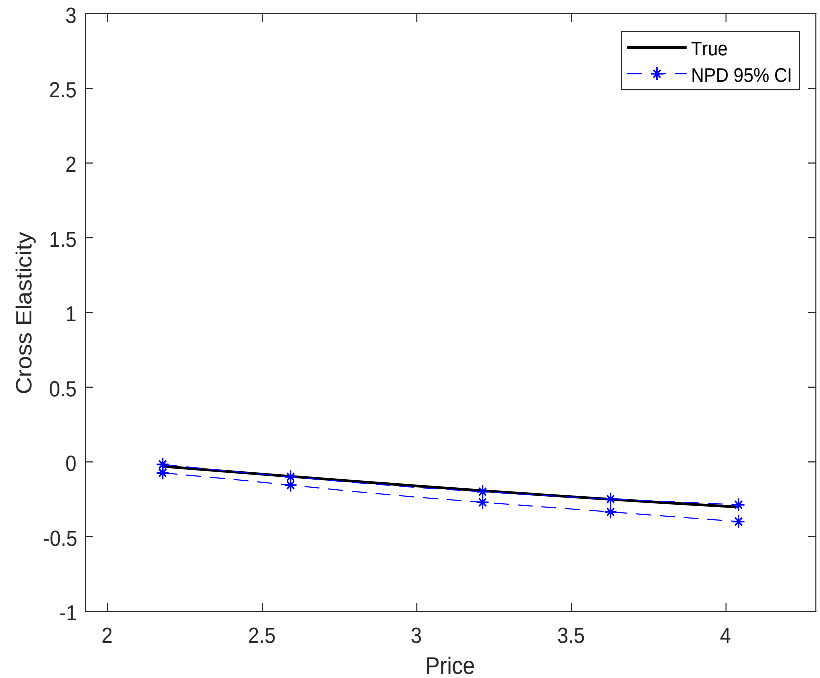


Figure 32: Complements, degree = 8: Own-price elasticity function

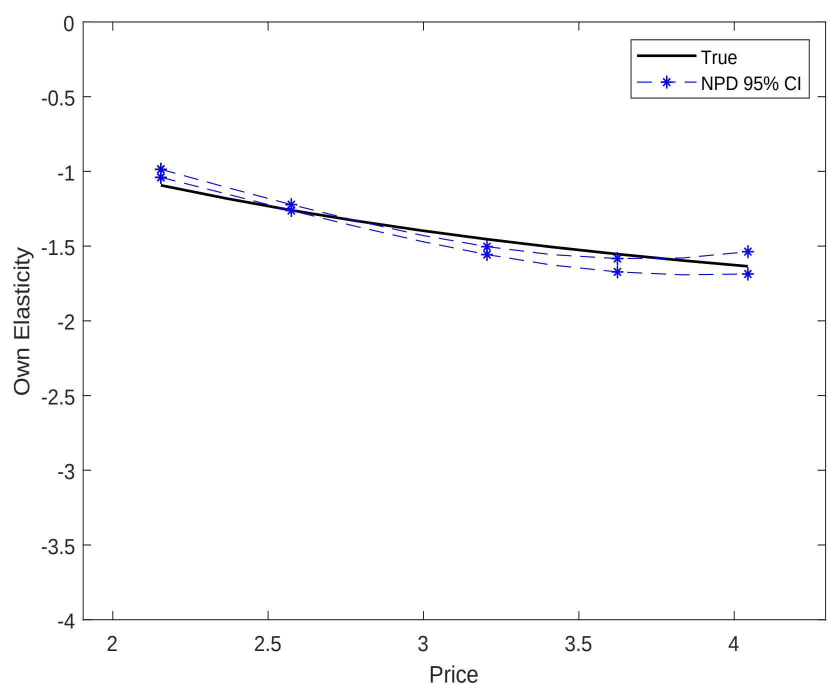


Figure 33: Complements, degree = 8: Cross-price elasticity function

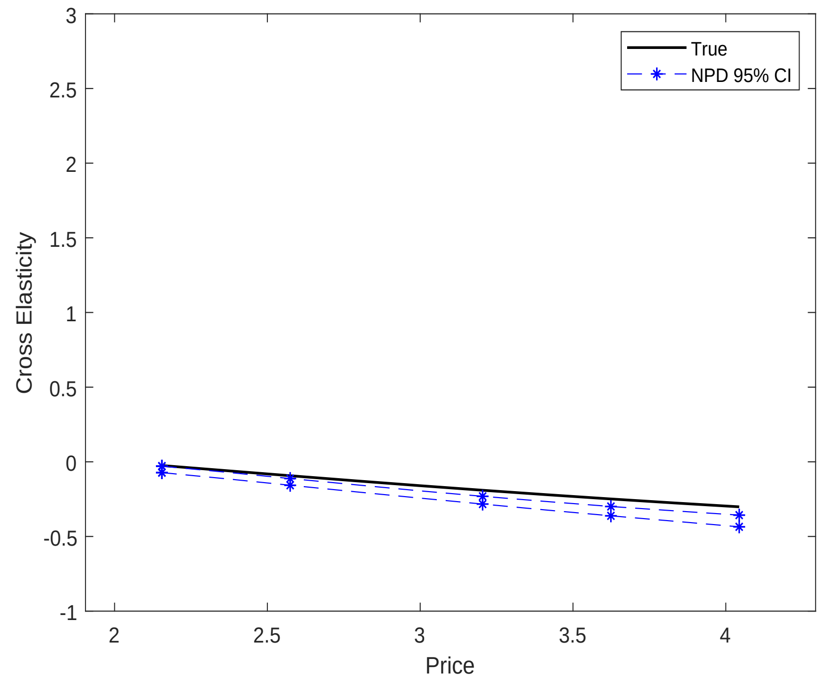


Figure 34: Complements, degree = 6: Own-price elasticity function

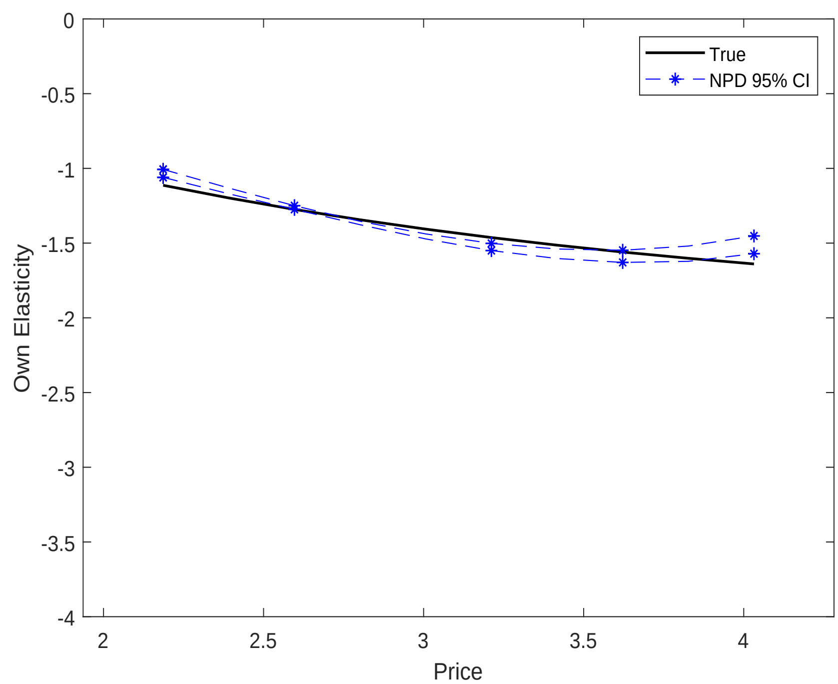


Figure 35: Complements, degree = 6: Cross-price elasticity function

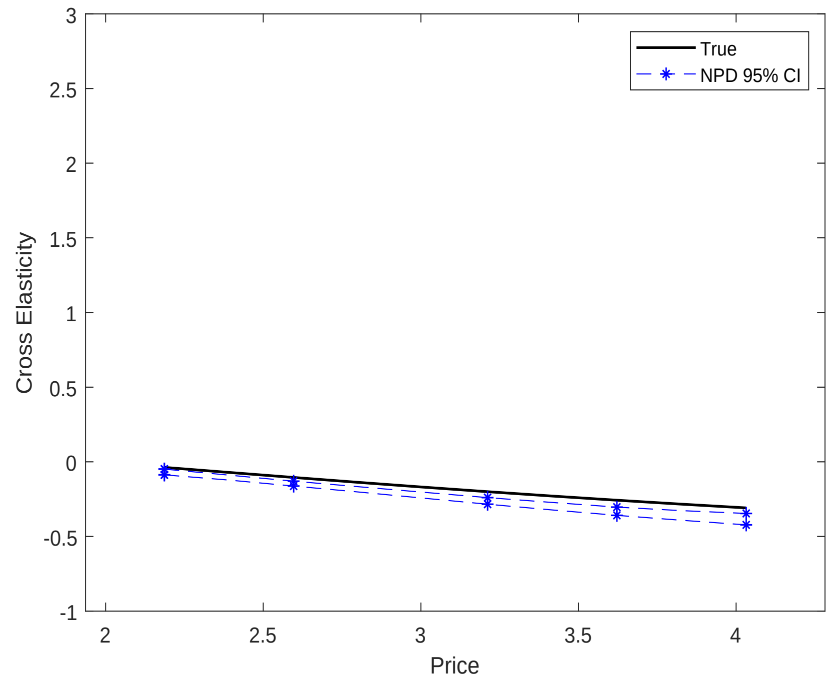


Figure 36: Complements, degree = 4: Own-price elasticity function

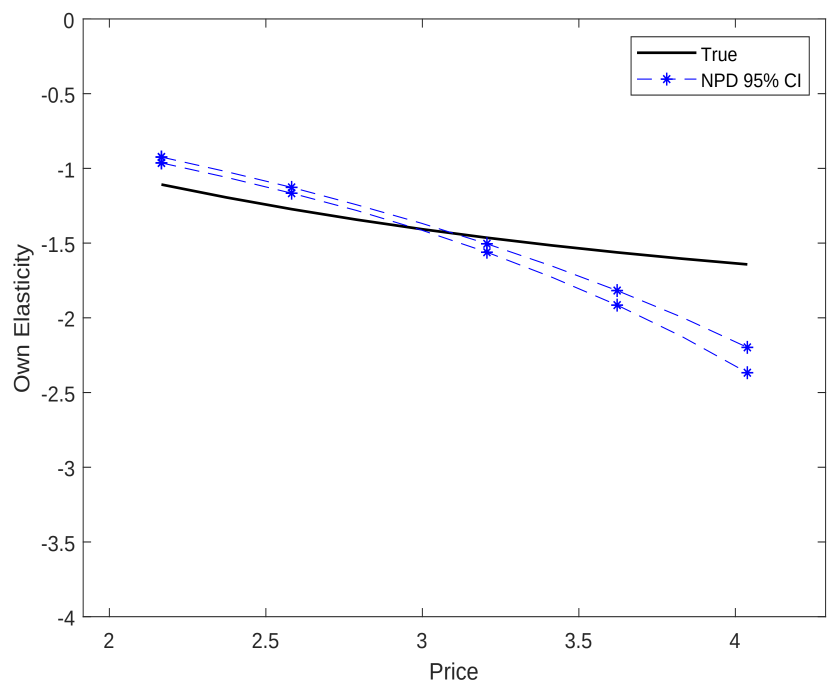
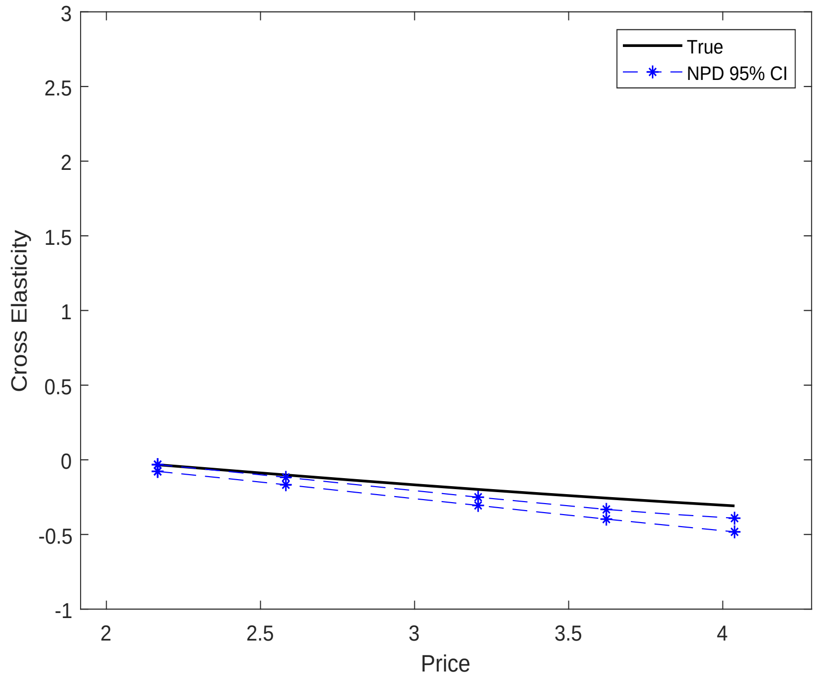


Figure 37: Complements, degree = 4: Cross-price elasticity function



D.5 $J > 2$ goods

The simulations presented so far are all for the case with $J = 2$ goods. While this corresponds to the empirical setting studied in the paper, it is interesting to investigate how the NPD estimator performs with more than two goods. In what follows, we generate data from the logit model

$$u_{ij} = -p_j + x_j + \xi_j + \epsilon_{ij}$$

We choose this simple model as it means that we can put p_j into the linear index δ_j, which reduces the number of parameters to estimate. We report the own-price elasticity of good 1 and the elasticity of good 1 wrt the price of good 2 for $J = 3, J = 5$, and $J = 7$ below.[69]

[69]Since the dgp and the model are symmetric in the different goods, the remaining own- and cross-price elasticities are the same as those reported here.

Figure 38: Logit, $J = 3$: Own-price elasticity function

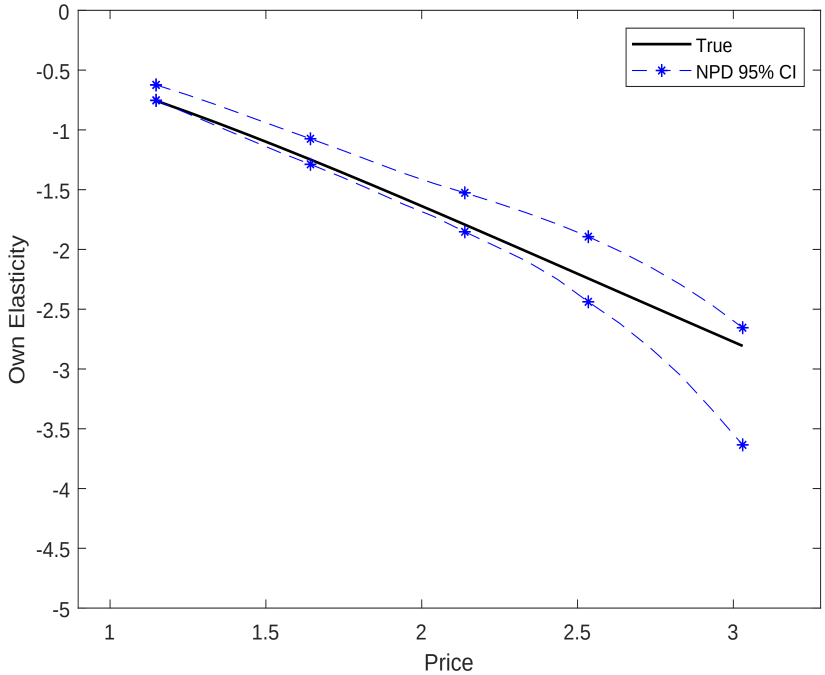


Figure 39: Logit, $J = 3$: Cross-price elasticity function

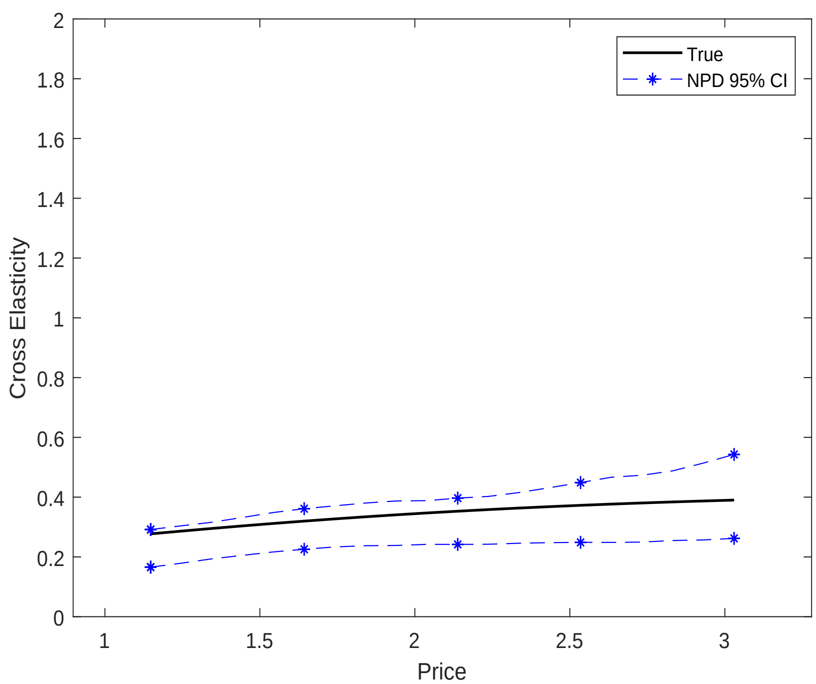


Figure 40: Logit, $J = 5$: Own-price elasticity function

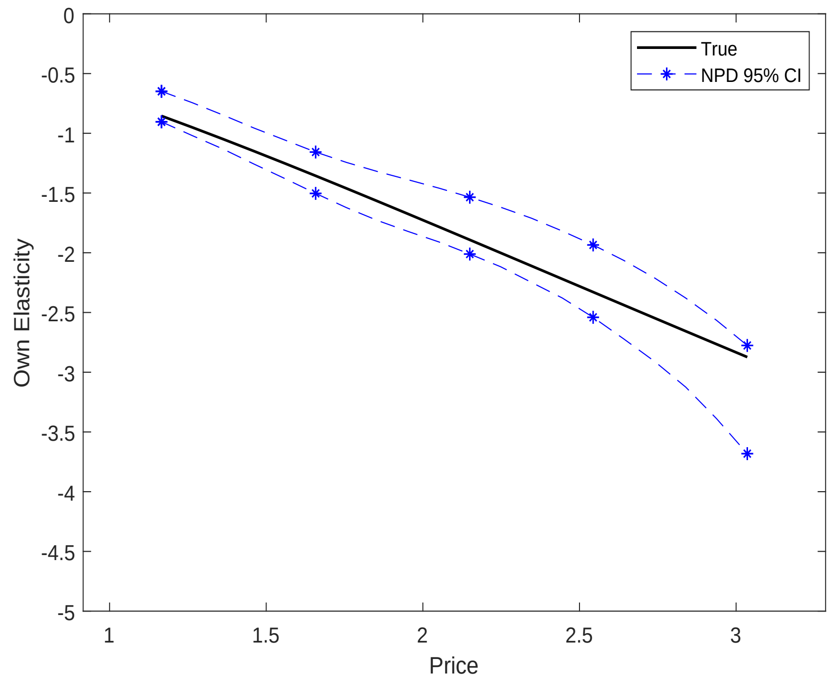
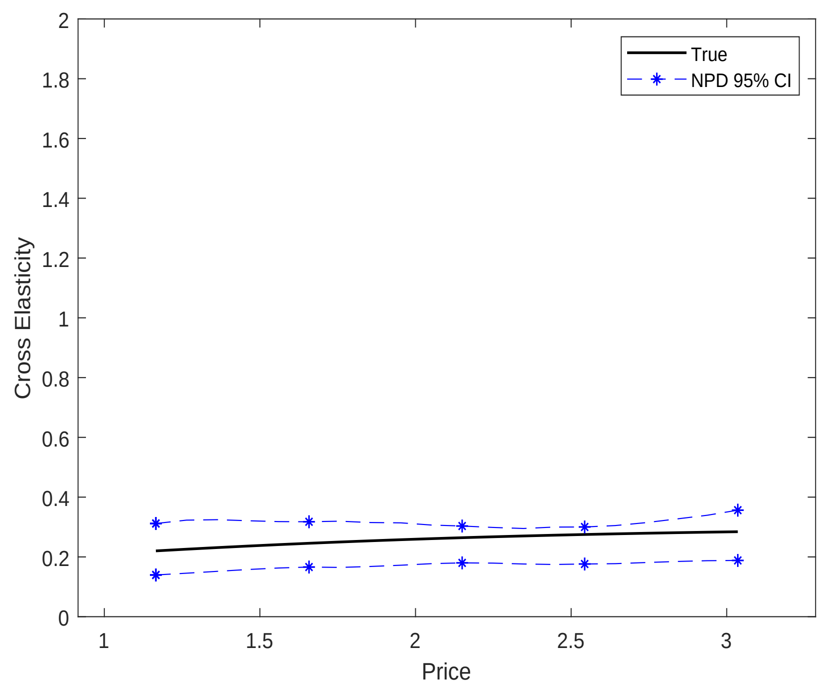


Figure 41: Logit, $J = 5$: Cross-price elasticity function



Appendix E: Data

We take a market to be a week/store combination.[70] Data on prices and quantities come from the 2014 Nielsen scanner dataset. For each market, the most granular unit of observation in the Nielsen data is a UPC (i.e. a specific bar code). Due to the large number of UPCs, we choose to aggregate according to whether they bear or do not bear the USDA Organic Seal. When this information is missing, we assume the UPC is non-organic. The aggregate quantities are obtained by simply summing the quantities for the individual UPCs, whereas for prices we take a weighted average where the weights are determined by the yearly share of sales that a given UPC has in that supermarket. Similarly, we aggregate across UPCs for selected non-strawberry fruits.[71] Specifically, we focus on the top four non-strawberry fruits according to Produce for Better Health Foundation (2015) in terms of per capita consumption nationwide, i.e. bananas, apples, other berries and oranges. For each of these fruits, we compute a price index (across UPCs) following the same procedure we used for strawberries. These fruit-level price indices are then aggregated even further into a single price index using weights that are proportional to the per capita eatings of each fruit and are normalized to sum to one.

Regarding Hausman instruments, we take the mean price of strawberries and the mean price index for the outside option, respectively, across the Californian supermarkets that are not in the same marketing area[72] as a given store. Excluding supermarkets in the same marketing area is meant to alleviate the usual concerns about Hausman instruments, i.e. that likely spatial correlation in the unobserved quality of the products might induce a violation of the exogeneity assumption.

Spot prices for strawberries are obtained from the US Department of Agriculture website.[73] The data reports spot prices for the following shipping points: California, Texas, Florida, North Carolina, and Mexico. In absence of information on where each supermarket sources their strawberries, we take a simple average of the prices at the various shipping points in any given week.

We measure the availability of non-strawberry fresh fruit in any given week at the state level using the total sales of non-strawberry fruits at all stores included in the Nielsen dataset in that week. To proxy for consumer tastes for organic produce at a given store, we compute the percentage of yearly organic lettuce sales over total yearly lettuce sales at the store.

Finally, data on income at the zip-code level is downloaded from the Internal Revenue Service website.[74]

The resulting dataset has 38,800 markets. Table 7 reports descriptive statistics for each variable and Figure 42 shows the price pattern for a typical store over time (the plot refers to organic strawberries). Both the retail price and the spot price exhibit strong seasonality. Moreover, the retail price sometimes displays a pattern in which it drops and then jumps back up to the initial level. This is typical of supermarket prices given the prevalence of periodic sales. However, in the case of strawberries, this pattern is much less

[70] We use the terms "store" and "retailer" interchangeably.

[71] In this case, however, we do not distinguish between organic and non-organic fruits.

[72] Here we follow the Nielsen partition of the United States into Designated Marketing Areas.

[73] http://cat.marketnews.usda.gov/cat/index.html.

[74] https://www.irs.gov/uac/soi-tax-stats-individual-income-tax-statistics-zip-code-data-soi.

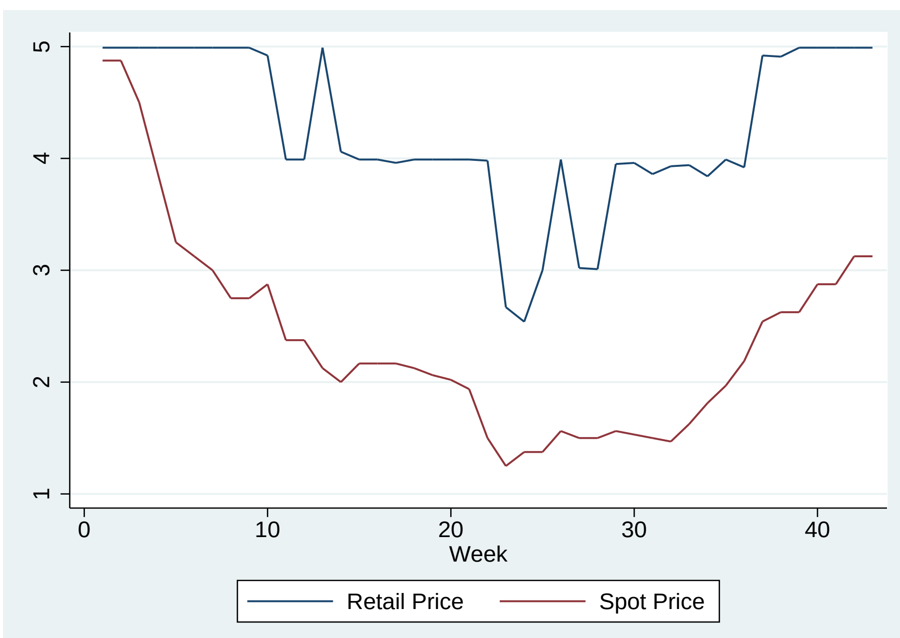
Table 7: Descriptive statistics

	Mean	Median	Min	Max
Quantity non-organic	735.33	581.00	6.00	5,729.00
Quantity organic	128.91	78.00	1.00	2,647.00
Price non-organic	2.97	2.89	0.93	4.99
Price organic	4.26	3.99	1.24	6.99
Price other fruit	3.95	3.80	1.30	13.88
Hausman non-organic	3.00	2.98	2.09	4.05
Hausman organic	4.28	4.07	2.95	5.50
Hausman other fruit	4.50	3.79	1.19	13.33
Spot non-organic	1.46	1.35	0.99	2.32
Spot organic	2.38	2.17	1.25	4.88
Quantity other fruit (per capita)	0.83	0.82	0.60	1.08
Share organic lettuce	0.08	0.06	0.00	0.41
Income	82.54	72.61	33.44	405.09

Note: Prices in dollars per pound. Quantities in pounds. Income in thousands of dollars per household.

marked than for other items, such as packaged goods. Therefore, we do not explicitly model sales in what follows.[75]

Figure 42: Price patterns



Note: Prices in dollars per pound for organic strawberries sold at a representative store.

[75] Inventory is often invoked as a justification for sales in models of retail. However, because strawberries are so perishable, it is unlikely that inventory plays a first-order role in driving the retailer's pricing behavior.

Table 8: First-stage regressions

	Non-organic		Organic	
	Price	Share	Price	Share
Spot price (own)	0.12^{**}	-0.68^{**}	0.35^{**}	-0.26^{**}
Spot price (other)	0.04^{**}	0.10^{**}	-0.21^{**}	0.22^{**}
Hausman (own)	0.70^{**}	-1.30^{**}	0.46^{**}	-0.19^{**}
Hausman (other)	-0.01	0.25^{**}	0.13^{**}	0.22^{**}
Hausman (out)	-0.01^{**}	0.11^{**}	-0.10^{**}	0.04^{**}
Availability other fruit	-0.01^{**}	-0.07^{**}	-0.02^{**}	-0.01^{**}
Share organic lettuce	0.08^{**}	-0.20^{**}	-0.01^{**}	0.10^{**}
Income	-0.02^{**}	0.00^{**}	0.01^{**}	0.04^{**}
R^2	0.46	0.27	0.52	0.16

Note: ** denotes significance at the 95% level. All variables are normalized to belong to the $[0, 1]$ interval.

Next, we present the results of the first-stage regressions in Table 8. As expected, the retail prices significantly increase with the spot prices. Further, the share of organic strawberries increases with the taste for organic products, while the opposite is true of the non-organic share. Finally, the shares of both inside goods decrease with the availability of other fruit.

Appendix F: Extension to Endogenous Demand Shifters

In this appendix, we consider violations of the exogeneity assumption 3 that take the form $\mathbb{E}(\xi_j|x,z) = \gamma_j x_j$ for all j.[76] By Equation (1), for all j,

$$x_{jt} = \mathbb{E}\left[\frac{1}{\beta_j + \gamma_j}\sigma_j^{-1}\left(s_t, p_t, x_t^{(2)}\right)\middle| x, z\right] \equiv \mathbb{E}\left[w_j\left(s_t, p_t, x_t^{(2)}\right)\middle| x, z\right] \tag{29}$$

where we let $w \equiv [w_1, ..., w_J]' \equiv M_w \sigma^{-1}$ and M_w is the diagonal matrix with (j,j) entry $\frac{1}{\beta_j+\gamma_j}$. Under Assumption 4, we can identify w. Let $\mathbb{J}_w^s$ denote the Jacobian of w wrt s, and similarly for $\mathbb{J}_w^p$, $\mathbb{J}_w^{x^{(1)}}$, and $\mathbb{J}_w^{x^{(2)}}$. Note that $\mathbb{J}_\sigma^p = -\left(\mathbb{J}_w^s\right)^{-1}\mathbb{J}_w^p$, so that $\mathbb{J}_\sigma^p$ is identified. An analogous argument applies to $\mathbb{J}_\sigma^{x^{(2)}}$. On the other hand, since $\mathbb{J}_\sigma^{x^{(1)}} = \left(\mathbb{J}_w^s\right)^{-1}\tilde{M}_w$, where $\tilde{M}_w$ is the diagonal matrix with (j,j) entry $\frac{\beta_j}{\beta_j+\gamma_j}$, identifying w is not sufficient to recover $\mathbb{J}_\sigma^{x^{(1)}}$. In other words, the marginal effects of p and $x^{(2)}$ are identified in spite of the endogeneity of $x^{(1)}$, whereas—as one would expect—the marginal effects of $x^{(1)}$ are not. A corollary of this is that counterfactuals that only depend on derivatives wrt prices—such as those considered in Section 5.4—are robust to this type of endogeneity.

[76]For simplicity, here we consider the case where $x_j^{(1)}$ is scalar, since that corresponds to the empirical settings in Section 5.

Appendix G: Microfoundation of the Empirical Model

In this appendix, we show how to map the model used in the empirical application to the general BH framework outlined in Section 2. Specifically, we outline two models of consumer behavior that yield the demand system in equation (5) and prove that the system is indeed invertible. It should be emphasized that these are only two out of many models that are compatible with (5) and invertibility, and that the estimation procedure does not rely on any of the parametric restrictions embedded in either model.[77]

G.1 Model 1

We first consider a standard discrete choice model. While the model is clearly at odds with the fact that consumers buying fresh fruit face an (at least partially) continuous choice, this serves as a building block for the more realistic model discussed in Section G.2. Moreover, given the prevalence of discrete choice models in the literature, it provides a connection between the demand system in (5) and a more familiar setup.

We assume that consumers face a discrete choice between one unit (say, one pound) of non-organic strawberries, one unit of organic strawberries and one unit of other fresh fruit. Consumer i's indirect utilities for each of these goods are, respectively

$$\begin{aligned} u_{i1} &= \theta_{str}\delta^*_{str} + \alpha_i p_1 + \epsilon_{i1} \\ u_{i2} &= \theta_{str}\delta^*_{str} + \theta_{org}\delta^*_{org} + \alpha_i p_2 + \epsilon_{i2} \\ u_{i0} &= \theta_{0,str}x^{(1)}_{str} + \theta_{0,org}\delta^*_{org} + \alpha_i p_0 + \epsilon_{i0} \end{aligned} \tag{30}$$

where

$$\begin{aligned} \delta^*_{str} &= \xi_{str} \\ \delta^*_{org} &= \theta_{1,org}x^{(1)}_{org} + \xi_{org} \end{aligned}$$

and p_1, p_2, p_0 denote the prices of non-organic strawberries, organic strawberries and the price index for other fresh fruit, respectively. We interpret δ^*_{str} as the mean quality of all strawberries in the market and δ^*_{org} as the mean utility for organic products (including—but not limited to—organic strawberries). Because the outside option of buying other fresh fruit includes organic produce (e.g. organic apples), we let δ^*_{org} enter u_{i0}. We use (ξ_{str}, ξ_{org}) to denote the unobserved quality levels for strawberries and organic produce, respectively, and $(\epsilon_{i2}, \epsilon_{i2})$ to denote taste shocks idiosyncratic to consumer i. Unlike BLP, we will not make any parametric assumptions on $(\epsilon_{i2}, \epsilon_{i2})$, nor on the distribution of the price coefficient α_i. In particular, note that the correlation structure of the vector $(\epsilon_{i2}, \epsilon_{i2}, \alpha_i)$ is unrestricted, which allows for patterns such as the fact that wealthier consumers may have a stronger preference for organic produce. Further, the distribution of α_i will be allowed to depend on other covariates such as mean income $x^{(2)}$ in the market.

Now we show that the demand system generated by the model above is identified under the following very mild assumption (as well as the standard exogeneity and completeness assumptions discussed in Section

[77]For instance, while Model 1 below assumes that prices enter linearly in utilities, this restriction is not needed for identification or estimation, given that we do not impose symmetry of the Jacobian of demand with respect to price.

2).

Assumption 11. *The coefficients* $\theta_{str}, \theta_{org}, \theta_{0,str}, \theta_{0,org}$ *and* $\theta_{1,org}$ *are non-zero.*

Note that Assumption 11 is very mild. It is satisfied if (i) consumers care about the quality of strawberries ($\theta_{str} > 0$) and organic produce ($\theta_{org}, \theta_{0,org} > 0$), as well as the availability of non-strawberry fruit $\theta_{0,str} > 0$, when purchasing fresh fruit; and (ii) the variable $x_{org}^{(1)}$ is indeed a proxy for taste for organic produce ($\theta_{1,org} > 0$).

Lemma 11. *Under Assumptions 3, 4 and 11, the functions* σ_1 *and* σ_2 *in* (5) *are point-identified.*

Proof. Since utility is ordinal, we can subtract $\theta_{0,str} x_{str}^{(1)} + \theta_{0,org} \delta_{org}^{*} + \alpha_i p_0$ from each equation in (30) and write

$$\begin{aligned} u_{i1} &= \tilde{\delta}_1 - \theta_{0,str} x_{str}^{(1)} + \alpha_i (p_1 - p_0) + \epsilon_{i1} \\ u_{i2} &= \tilde{\delta}_2 - \theta_{0,str} x_{str}^{(1)} + \alpha_i (p_2 - p_0) + \epsilon_{i2} \\ u_{i0} &= \epsilon_{i0}, \end{aligned} \tag{31}$$

where

$$\begin{aligned} \tilde{\delta}_1 &\equiv \theta_{str} \delta_{str}^{*} - \theta_{0,org} \delta_{org}^{*} \\ \tilde{\delta}_2 &\equiv \theta_{str} \delta_{str}^{*} + (\theta_{org} - \theta_{0,org}) \delta_{org}^{*} \end{aligned}$$

Using (31) and the fact that we allow the distribution of α_i to depend on $x^{(2)}$, we can write the demand system as

$$s = \tilde{\sigma}\left(\tilde{\delta}_1 - \theta_{0,str} x_{str}^{(1)}, \tilde{\delta}_2 - \theta_{0,str} x_{str}^{(1)}, p, x^{(2)}\right), \tag{32}$$

where $p \equiv (p_0, p_1, p_2)$, $s \equiv (s_1, s_2)'$ is the vector of market shares and $\tilde{\sigma}$ is a function from $\mathbb{R}^2 \times \mathbb{R}_+^4$ to the unit 2-simplex. Next, by Theorem 1 of Berry, Gandhi, and Haile (2013), we can invert the system in (32) for the mean utility levels as follows

$$\begin{aligned} \tilde{\delta}_1 &= \tilde{\sigma}_1^{-1}\left(s, p, x^{(2)}\right) + \theta_{0,str} x_{str}^{(1)} \\ \tilde{\delta}_2 &= \tilde{\sigma}_2^{-1}\left(s, p, x^{(2)}\right) + \theta_{0,str} x_{str}^{(1)}, \end{aligned} \tag{33}$$

where $\tilde{\sigma}_k^{-1}$ denotes the $k-$th element of the inverse, $\tilde{\sigma}^{-1}$, of $\tilde{\sigma}$. We now show that there is a one-to-one mapping between $\left(\delta_{str}^{*}, \delta_{org}^{*}\right)$ and $\left(\tilde{\delta}_1, \tilde{\delta}_2\right)$, which means that we can invert the system for the original demand indices. Letting $\delta^{*} \equiv \left(\delta_{str}^{*}, \delta_{org}^{*}\right)'$ and $\tilde{\delta} \equiv \left(\tilde{\delta}_1, \tilde{\delta}_2\right)'$, we have

$$\tilde{\delta} = A\delta^{*},$$

where

$$A \equiv \begin{bmatrix} \theta_{str} & -\theta_{0,org} \\ \theta_{str} & \theta_{org} - \theta_{0,org} \end{bmatrix}$$

Since $\det(A) = \theta_{str}\theta_{org} \neq 0$ under Assumption 11, we can rewrite (33) as

$$\delta^{*} = A^{-1} \tilde{\sigma}^{-1}\left(s, p, x^{(2)}\right) + A^{-1} \cdot [1 \quad 1]' \times \theta_{0,str} x_{str}^{(1)} \tag{34}$$

or, equivalently,

$$\begin{aligned}\delta^*_{str} &= \sigma_1^{-1}\left(s, p, x^{(2)}\right) + \theta_1 x^{(1)}_{str} \\ \delta^*_{org} &= \sigma_2^{-1}\left(s, p, x^{(2)}\right) + \theta_2 x^{(1)}_{str},\end{aligned} \tag{35}$$

for functions $\sigma_i^{-1} : \Delta^2 \times \mathbb{R}^4_+ \to \mathbb{R}^2$, $i = 1, 2$, where Δ^2 denotes the unit 2-simplex. Now we derive expressions for the coefficients θ_1 and θ_2 in terms of the model primitives. Note that

$$A^{-1} = \frac{1}{\theta_{org}} \begin{bmatrix} \frac{\theta_{org} - \theta_{0,org}}{\theta_{str}} & \frac{\theta_{0,org}}{\theta_{str}} \\ -1 & 1 \end{bmatrix}$$

and thus

$$A^{-1} \cdot [1 \quad 1]' = \begin{bmatrix} \frac{1}{\theta_{str}} & 0 \end{bmatrix}',$$

i.e.

$$\begin{aligned}\theta_1 &= \frac{\theta_{0,str}}{\theta_{str}} \\ \theta_2 &= 0\end{aligned}$$

Plugging this into (35) and using the definitions of δ^*_{str} and δ^*_{org}, we obtain

$$\begin{aligned}\xi_{str} &= \sigma_1^{-1}\left(s, p, x^{(2)}\right) + \frac{\theta_{0,str}}{\theta_{str}} x^{(1)}_{str} \\ \theta_{1,org} x^{(1)}_{org} + \xi_{org} &= \sigma_2^{-1}\left(s, p, x^{(2)}\right)\end{aligned} \tag{36}$$

The final step is to show that we can identify the system in (36), given the instruments available. Because we are free to normalize the scale of ξ_{str} and ξ_{org} in the display above, we can divide the first equation of (36) by $\frac{\theta_{0,str}}{\theta_{str}}$ and the second equation by $\theta_{1,org}$ without loss,[78] and rearrange terms as follows

$$-x^{(1)}_{str} = \sigma_1^{-1}\left(s, p, x^{(2)}\right) - \xi_{str} \tag{37}$$

$$x^{(1)}_{org} = \sigma_2^{-1}\left(s, p, x^{(2)}\right) - \xi_{org}, \tag{38}$$

Equations (37) and (38) are in the same form as Equation (6) in BH and thus we can follow their argument to show that σ_1 and σ_2 are identified. Further, note that inverting the system in (37) and (38) yields the demand system in equation (5) (after normalizations).

□

[78] These divisions are well-defined operations as $\frac{\theta_{0,out}}{\theta_{str}}$ and $\theta_{1,org}$ are nonzero by Assumption 11.

G.2 Model 2

We now turn to a model of continuous choice that appears to be a closer approximation to the behavior of consumers buying fresh fruit. Let consumer i face the following maximization problem

$$\begin{aligned} &\max_{q_0,q_1,q_2} U_i\left(q_0,q_1,q_2\right) \\ &\text{s.t. } \; p_0q_0+p_1q_1+p_2q_2 \le y_i^{inc} \end{aligned} \tag{39}$$

where y_i^{inc} denotes the income consumer i allocates to fresh fruit, q_0 is the quantity of non-strawberry fresh fruit, q_1 is the quantity of non-organic strawberries and q_2 is the quantity of organic strawberries, and similarly for prices p_0, p_1, p_2. One could think of y_i^{inc} as being the outcome of a higher-level optimization problem in which the consumer chooses how to allocate total income across different product categories, including fresh fruit. Assume U_i takes the Cobb-Douglas form

$$U_i\left(q_0,q_1,q_2\right) = q_0^{d_0\epsilon_{i,0}} q_1^{d_1\epsilon_{i,1}} q_2^{d_2\epsilon_{i,2}},$$

for positive $d \equiv (d_0, d_1, d_2)$ and $\epsilon_i \equiv (\epsilon_{i,0}, \epsilon_{i,1}, \epsilon_{i,2})$. Then, the optimal quantities chosen by the consumer are

$$q_j^*\left(d,p,y_i^{inc},\epsilon_i\right) = \frac{y_i^{inc}}{p_j}\cdot\frac{d_j\epsilon_{i,j}}{\sum_{k=0}^{2} d_k\epsilon_{i,k}} \qquad j=0,1,2 \tag{40}$$

where $d \equiv (d_0, d_1, d_2)$ and $p \equiv (p_0, p_1, p_2)$. Now assume that

$$\begin{aligned} d_0 &= \gamma_{org}^{\theta_{0,org}}\tilde{x}_{str}^{\theta_{0,str}} \\ d_1 &= \gamma_{str}^{\theta_{str}} \\ d_2 &= \gamma_{str}^{\theta_{str}}\gamma_{org}^{\theta_{org}} \end{aligned}$$

where

$$\begin{aligned} \gamma_{str} &\equiv \exp\left\{\delta_{str}^*\right\} \\ \gamma_{org} &\equiv \exp\left\{\delta_{org}^*\right\} \\ \tilde{x}_{str} &\equiv \exp\left\{x_{str}^{(1)}\right\} \end{aligned}$$

and $\delta_{str}^*, \delta_{org}^*$ are defined as in Section G.1. We can then re-write (40) as

$$q_j^*\left(\tilde{d},p,y_i^{inc},\epsilon_i\right) = \frac{y_i^{inc}}{p_j}\cdot\frac{\tilde{d}_j\epsilon_{i,j}}{\sum_{k=0}^{2} \tilde{d}_k\epsilon_{i,k}} \qquad j=0,1,2 \tag{41}$$

where

$$\begin{aligned} \tilde{d}_0 &\equiv 1 \\ \tilde{d}_1 &\equiv \gamma_{str}^{\theta_{str}}\gamma_{org}^{-\theta_{0,org}}\tilde{x}_{str}^{-\theta_{0,str}} \\ \tilde{d}_2 &\equiv \gamma_{str}^{\theta_{str}}\gamma_{org}^{\theta_{org}-\theta_{0,org}}\tilde{x}_{str}^{-\theta_{0,str}} \end{aligned}$$

and $\tilde{d} \equiv \left(\tilde{d}_0, \tilde{d}_1, \tilde{d}_2\right)$.
Next, let $F_{Y,\epsilon}$ denote the joint distribution of y_i^{inc} and ϵ_i in the market, and define[79]

$$Q_j^*\left(\tilde{d}, p, x^{(2)}\right) = \int q_j^*\left(\tilde{d}, p, y, \epsilon\right) dF_{Y,\epsilon}\left(y, \epsilon; x^{(2)}\right) \qquad j = 0, 1, 2$$

$Q_j^*\left(\tilde{d}, p, x^{(2)}\right)$ is the model counterpart to the market-level quantity Q_j observed in the data.
The last step is to show that there exists a mapping of quantities into (artificial) market shares such that the resulting demand system is invertible. For $j = 0, 1, 2$, define

$$\tilde{\tilde{\sigma}}_j\left(\tilde{d}, p, x^{(2)}\right) = \frac{Q_j^*\left(\tilde{d}, p, x^{(2)}\right)}{\sum_{k=0}^{2} Q_k^*\left(\tilde{d}, p, x^{(2)}\right)}$$

and

$$s_j = \frac{Q_j}{\sum_{k=0}^{2} Q_k}$$

Then, equating observed shares to their model counterparts, we obtain the system

$$s = \tilde{\tilde{\sigma}}\left(\tilde{d}, p, x^{(2)}\right) \tag{42}$$

where $s \equiv (s_0, s_1, s_2)'$ and $\tilde{\tilde{\sigma}}\left(\tilde{d}, p, x^{(2)}\right) \equiv \left(\tilde{\tilde{\sigma}}_0\left(\tilde{d}, p, x^{(2)}\right), \tilde{\tilde{\sigma}}_1\left(\tilde{d}, p, x^{(2)}\right), \tilde{\tilde{\sigma}}_2\left(\tilde{d}, p, x^{(2)}\right)\right)'$.
Because $\tilde{\tilde{\sigma}}_j$ is strictly decreasing in $\tilde{d}_k$ for all j and all $k > 0, k \neq j$, by Theorem 1 in Berry, Gandhi, and Haile (2013), we can invert (42) as follows

$$\tilde{d} = \tilde{\tilde{\sigma}}^{-1}\left(s, p, x^{(2)}\right)$$

and, taking logs, we can write

$$\begin{aligned} \theta_{str}\delta_{str}^* - \theta_{0,org}\delta_{org}^* &= \tilde{\sigma}_1^{-1}\left(s, p, x^{(2)}\right) + \theta_{0,str}x_{str}^{(1)} \\ \theta_{str}\delta_{str}^* + \left(\theta_{org} - \theta_{0,org}\right)\delta_{org}^* &= \tilde{\sigma}_2^{-1}\left(s, p, x^{(2)}\right) + \theta_{0,str}x_{str}^{(1)} \end{aligned} \tag{43}$$

where $\tilde{\sigma}_j^{-1}\left(s, p, x^{(2)}\right) \equiv \log\left(\tilde{\tilde{\sigma}}_j^{-1}\left(s, p, x^{(2)}\right)\right)$ for $j = 1, 2$.
Note that (43) has the exact same form as (33). Therefore, we can use the argument in Section G.1 to show that the demand system is identified.

[79]Note that we let $F_{Y,\epsilon}$ be a function of mean income $x^{(2)}$, consistently with the information available in the data.

References

Abaluck, J., and A. Adams (2017): "What Do Consumers Consider Before They Choose? Identification from Asymmetric Demand Responses," IFS Working Paper W17/09.

Atkin, D., and D. Donaldson (2015): "Who's Getting Globalized? The Size and Implications of Intra-national Trade Costs," NBER Working Paper No. 21439.

Berry, S., A. Gandhi, and P. Haile (2013): "Connected substitutes and invertibility of demand," *Econometrica*, 81, 2087–2111.

Berry, S., and P. Haile (2010): "Nonparametric Identification of Multinomial Choice Demand Models with Heterogeneous Consumers," Working Paper (Yale University).

——— (2014): "Identification in differentiated products markets using market level data," *Econometrica*, 82, 1749–1797.

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

——— (2004): "Differentiated Products Demand Systems from a Combination of Micro and Macro Data: The New Car Market," *Journal of Political Economy*, 112, 68–105.

Bulow, J. I., J. D. Geanakoplos, and P. D. Klemperer (1985): "Multimarket Oligopoly: Strategic Substitutes and Complements," *The Journal of Political Economy*, 93, 488–511.

Bulow, J. I., and P. Pfleiderer (1983): "A Note on the Effect of Cost Changes on Prices," *The Journal of Political Economy*, 91, 182–185.

Capps, C., D. Dranove, and M. Satterthwaite (2003): "Competition and Market Power in Option Demand Markets," *RAND Journal of Economics*, 34, 737–763.

Chen, X., and T. Christensen (2015): "Optimal uniform convergence rates and asymptotic normality for series estimators under weak dependence and weak conditions," *Journal of Econometrics*, 188, 447–465.

——— (forthcoming): "Optimal Sup-norm Rates and Uniform Inference on Nonlinear Functionals of Nonparametric IV Regression," *Quantitative Economics*.

Chen, X., and D. Pouzo (2009): "Efficient estimation of semiparametric conditional moment models with possibly nonsmooth residuals," *Journal of Econometrics*, 152, 46—-60.

——— (2012): "Estimation of nonparametric conditional moment models with possibly nonsmooth generalized residuals," *Econometrica*, 80, 277–321.

——— (2015): "Sieve Wald and QLR Inferences on Semi/Nonparametric Conditional Moment Models," *Econometrica*, 83, 1013—-1079.

Chen, X., and Y. J. J. Qiu (2016): "Methods for Nonparametric and Semiparametric Regressions with Endogeneity: A Gentle Guide," *Annual Review of Economics*, 8, 259–290.

Chernozhukov, V., and C. Hansen (2006): "Instrumental quantile regression inference for structural and treatment effect models," *Journal of Econometrics*, 132, 491–525.

Chetverikov, D., and D. Wilhelm (2017): "Nonparametric Instrumental Variable estimation under monotonicity," *Econometrica*, 85, 1303–1320.

Chevalier, J. A., A. K. Kashyap, and P. E. Rossi (2003): "Why Don't Prices Rise During Periods of Peak Demand? Evidence from Scanner Data," *American Economic Review*, 93, 15–37.

Dubé, J.-P. (2004): "Multiple Discreteness and Product Differentiation: Demand for Carbonated Soft Drinks," *Marketing Science*, 23, 66–81.

Dubin, J. A., and D. L. McFadden (1984): "An Econometric Analysis of Residential Electric Appliance Holdings and Consumptions," *Econometrica*, 52, 345–362.

Gal, S. (2008): *Shape-Preserving Approximation by Real and Complex Polynomials*. Birkhauser, Boston.

Gentzkow, M. (2007): "Valuing New Goods in a Model with Complementarity: Online Newspapers," *American Economic Review*, 97, 713.

Goeree, M. S. (2008): "Limited Information and advertising in the U.S. personal computer industry," *Econometrica*, 76, 1017–1074.

Goldberg, P. K. (1995): "Product Differentiation and Oligopoly in International Markets: The Case of the U.S. Automobile Industry," *Econometrica*, 63, 891–951.

Goldberg, P. K., and R. Hellerstein (2008): "A Structural Approach to Explaining Incomplete Exchange-Rate Pass-Through and Pricing-to-Market," *American Economic Rewview: Papers and Proceedings*, 98, 423—-429.

——— (2013): "A Structural Approach to Identifying the Sources of Local Currency Price Stability," *Review of Economic Studies*, 80, 175—210.

Gordon, B., and W. Hartmann (2016): "Advertising competition in presidential elections," *Quantitative Marketing and Economics*, 14, 1–40.

Hendel, I. (1999): "Estimating Multiple-Discrete Choice Models: An Application to Computerization Returns," *The Review of Economic Studies*, 66, 423–446.

Horowitz, J. (2011): "Applied nonparametric instrumental variable estimation," *Econometrica*, 79, 347–394.

Johnson, C. R., and R. L. Smith (2011): "Inverse M-matrices, II," *Linear Algebra and its Applications*, 435, 953–983.

Knee, M. (2002): *Fruit Quality and Its Biological Basis*. CRC Press.

Lal, R., and C. Matutes (1994): "Retail Pricing and Advertising Strategies," *Journal of Business*, 67, 345—370.

Lal, R., and J. M. Villas-Boas (1998): "Price Promotions and Trade Deals with Multiproduct Retailers," *Management Science*, 44, 935—-949.

McDonald, J., M. Neumann, H. Schneider, and M. Tsatsomeros (1995): "Inverse M-Matrix Inequalities and Generalized Ultrametric Matrices," *Linear Algebra and Its Applications*, 22, 321–341.

Nakamura, E., and D. Zerom (2010): "Accounting for Incomplete Pass-Through," *Review of Economic Studies*, 77, 1192–1230.

Nevo, A. (2000): "Mergers with Differentiated Products: The Case of the Ready-to-Eat Cereal Industry," *The RAND Journal of Economics*, 31, 395–421.

——— (2001): "Measuring market power in the ready-to-eat cereal industry," *Econometrica*, 69, 307–642.

Pakes, A. (1994): "Dynamic structural models, problems and prospects: mixed continuous discrete controls and market interactions," in *Advances in Econometrics: The Sixth World Congress of the Econometric Society*, ed. by J.-J. Laffont, and C. Sims, vol. 2, pp. 171–260. Cambridge University Press, New York.

Petrin, A. (2002): "Quantifying the Benefits of New Products: The Case of the Minivan," *Journal of Political Economy*, 110(4), 705–729.

Plemmons, R. (1977): "M-matrix characterizations," *Linear Algebra and its Applications*, 18, 175–188.

Produce for Better Health Foundation (2015): *State of the plate: 2015 study on America's consumption of fruit and vegetables*. Produce for Better Health Foundation, Web http://www.PBHFoundation.org.

Schumaker, L. (2007): *Spline Functions: Basic Theory*. Cambridge Mathematical Library, Cambridge.

Seber, G. A. (2007): *A Matrix Handbook for Statisticians*. Wiley, New York.

Souza-Rodrigues, E. A. (2014): "Nonparametric Estimation of a Generalized Regression Model with Group Effects," working paper, University of Toronto, Department of Economics.

Tversky, A., and D. Kahneman (1991): "Loss Aversion in Riskless Choice: A Reference-Dependent Model," *The Quarterly Journal of Economics*, 106, 1039—-1061.

United States Department of Agriculture (2017): *Noncitrus Fruits and Nuts 2016 Summary*.

Weyl, E. G., and M. Fabinger (2013): "Pass-Through as an Economic Tool: Principles of Incidence under Imperfect Competition," *Journal of Political Economy*, 121, 528–583.