Market Counterfactuals and the Specification of Multi-Product Demand: A Nonparametric Approach*

Giovanni Compiani †

October 1, 2019

Abstract

Demand estimates are essential for addressing a wide range of positive and normative questions in economics that are known to depend on the shape—and notably the curvature—of the true demand functions. The existing frontier approaches, while allowing flexible substitution patterns, typically require the researcher to commit to a parametric specification. An open question is whether these a priori restrictions are likely to significantly affect the results. To address this, I develop a nonparametric approach to estimation of demand for differentiated products, which I then apply to California supermarket data. While the approach subsumes workhorse models such as mixed logit, it allows consumer behaviors and preferences beyond standard discrete choice, including continuous choices, complementarities across goods, and consumer inattention. When considering a tax on one good, the nonparametric approach predicts a much lower pass-through than a standard mixed logit model. However, when assessing the market power of a multi-product firm relative to that of a single-product firm, the models give similar results. I also illustrate how the nonparametric approach may be used to guide the choice among parametric specifications.

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

JEL codes: L1, L66

*I am very grateful to Steve Berry, Phil Haile, Penny Goldberg and Xiaohong Chen for their guidance and support, as well as Daniel Ackerberg, Nikhil Agarwal, Matteo Benetton, Judith Chevalier, Tim Christensen, Jan De Loecker, Ying Fan, Sharat Ganapati, Amit Gandhi, Mitsuru Igami, Adam Kapor, Yuichi Kitamura, Lorenzo Magnolfi, Camilla Roncoroni, Marcelo Sant'Anna, Paul Scott, and participants of EARIE 2017, ASSA 2019, and seminars at MIT, UCL, EIEF, NYU, Stanford, LSE, Mannheim, Tilburg, Berkeley 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. This paper was previously circulated under the title "Nonparametric Demand Estimation in Differentiated Products Markets." All errors and omissions are my own.

†University of California, Berkeley. 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, evaluating the effect of a tax or subsidy, merger simulation, assessing the impact of a new product being introduced into the market, understanding the drivers of the well-documented incomplete pass-through of cost and exchange-rate shocks to downstream prices, and determining whether firms play a game with strategic complements or substitutes.[1] 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 flexible 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. Addressing this concern requires an approach that relaxes the parametric assumptions, thus providing results that may be used as a benchmark.

To this end, I propose the first nonparametric approach to estimate demand in differentiated products markets based on aggregate data.[2] Specifically, I focus on markets in which consumers face a range of options that are differentiated in ways that are both observed and unobserved to the researcher. Importantly, the presence of unobserved heterogeneity at the product or market level implies that all the variables that are chosen by firms after observing consumer preferences—e.g. prices in many models—are econometrically endogenous. A vast literature in industrial organization and other fields has focused on the empirical analysis of this type of markets. The current frontier approach is to posit a random coefficients discrete choice logit model[3] and estimate it using the methodology developed by Berry et al. (1995)

[1]References include Berry et al. (1995) and Nevo (2001) for the study of market power, Bulow and Pfleiderer (1983) and Weyl and Fabinger (2013) for the effect of taxes and subsidies, Nevo (2000a) and Capps et al. (2003) for merger simulation, Petrin (2002) for the analysis of new products, Nakamura and Zerom (2010) and Goldberg and Hellerstein (2013) for incomplete pass-through, and Bulow et al. (1985) for strategic complementarity and substitution.

[2]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.

[3]Throughout the paper, I use the terms "random coefficients logit model" and "mixed logit model" interchangeably.

(henceforth BLP).[4] While the methodology in BLP accomplishes the crucial goals of generating reasonable substitution patterns and allowing for price endogeneity, it relies on a number of parametric assumptions which may affect the results of counterfactual exercises. For example, while it is well known that the pass-through of a tax depends on the curvature—i.e. the second derivatives—of the demand functions, it is not a priori clear whether BLP is flexible enough to capture these features of the true demand system. In contrast, the approach proposed in this paper does not rely on any distributional assumptions and imposes only limited functional form restrictions. For instance, one does not need to assume that the idiosyncratic taste shocks or the random coefficients on product characteristics in the utility function 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 consumer theory.

In addition, by directly targeting the demand functions as opposed to the 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,[5] that consumers are aware

[4]Another influential approach to demand estimation is the Almost Ideal Demand System (AIDS) pioneered by Deaton and Muellbauer (1980). I choose to compare my approach to BLP-type models and not AIDS-type models, because the latter restrict the role of the unobserved heterogeneity in a way that is at odds with the differentiated products markets literature from the last twenty years. Specifically, Deaton and Muellbauer (1980) use their model of consumer behavior to obtain a demand equation only involving observables and add a separable error term to carry out estimation (equation (15) in their paper). This implies that the unobservables do not have an immediate "structural" interpretation (such as product quality not captured by the data). One consequence is that the standard arguments used to motivate the issue of (price) endogeneity, as well as justify the instrumental variables solution to it, do not typically apply in the AIDS framework.

[5]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 I view this as a promising avenue for future research.

of all products and their characteristics,[6,7] and that each consumer buys at most one unit of a single product.[8] In contrast to this, my approach allows for a broader range of consumer behaviors and preferences, including complementarities across goods, consumer inattention, and multiple discrete or continuous choices.

In practice, I propose approximating the (inverse of the) demand system using Bernstein polynomials, which make it easy to enforce a number of economic constraints in the estimation routine. Computationally, the objective function to be minimized is convex in the parameters; thus, if the constraints are also convex, standard algorithms are guaranteed to converge to the global optimum. In order to show validity of the standard errors, I adapt proofs from recent work in econometrics on nonparametric instrumental variables regression and I provide primitive conditions for the case where the objects of interest are price elasticities and (counterfactual) equilibrium prices.

As with many nonparametric estimators, one limitation of my approach is that the number of parameters tends to increase quickly with the number of goods and/or covariates. However, I show that one can partially mitigate this curse of dimensionality by imposing restrictions on the demand functions while preserving most of the flexibility of the nonparametric approach. Specifically, I consider (i) an exchangeability restriction; and (ii) constraints on the way covariates and prices enter the demand system. Both (i) and (ii) substantially reduce the number of parameters relative to the most general model. In particular, (ii) highlights that there is a trade-off between functional form restrictions and severity of the curse of dimensionality. In practice, this means that a researcher can—to a certain extent—tailor the model to the specifics of her setting by choosing how many assumptions to impose. For example, if the sample size is moderate, a researcher might choose to assume more in terms of functional form to contain the number of parameters, while still avoiding

[6]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 assumptions in Section 2 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.

[7]One could conceivably use a BLP-type model to estimate consumer preferences on data generated by inattentive consumers. Whether the BLP functional form is flexible enough in such contexts is an open question that depends on the object of interest. The simulation evidence presented in this paper suggests that a BLP-type model tends to under-estimate own-price elasticities and over-estimate cross-price elasticities for one pattern of inattention.

[8]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).

several assumptions on the distribution of the unobservables and consumer behavior relative to a standard discrete choice model. On the other hand, with larger samples, the researcher might be able to relax some of these functional form restrictions.

Besides requiring a nonparametric approach, the assessment of how counterfactual outcomes are affected by parametric restrictions necessitates an amount of data sufficient to obtain informative results in the more flexible model. To this end, I leverage a large sample of store/week-level quantities and prices from Nielsen. Specifically, I focus on strawberry sales in California, which allows me to keep the number of goods low and thus avoid the curse of dimensionality. In addition, given the perishability of the product, I am able to reasonably abstract from dynamic considerations and perform a clean comparison between static demand models. Of course, this is a small product category, but the increasing availability of large data sets suggests that it might be possible to apply nonparametric approaches such as that proposed here to a much broader class of settings.

I consider two counterfactual exercises. The first is to quantify the pass-through of a tax into retail prices. Comparing the results to those given by a standard mixed logit model, I find that the nonparametrically estimated tax pass-through is significantly lower than that delivered by mixed logit for organic strawberries. I relate this to the fact that the nonparametric own-price elasticity for that good increases in absolute value much faster with own-price, which provides an incentive for the retailer to contain the price increase after the tax, all else equal. The second counterfactual concerns the role played by the multi-product nature of retailers in driving up markups (the "portfolio effect" in the terminology of, e.g., Nevo (2001)). In this case, a mixed logit model with product-specific fixed effects matches the nonparametric results very closely. This is not the case for mixed logit models with fewer fixed effects, suggesting that the proposed approach may be used to guide the choice among competing parametric specifications.

Related literature. This paper contributes to the vast literature on models of demand in differentiated products markets pioneered by BLP. In particular, a recent paper by Berry and Haile (2014) (henceforth BH) shows that most of the parametric assumptions imposed by BLP 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 I build on the *identification* result in BH, I focus on a distinct set of issues pertaining nonparametric *estimation*. Other papers developing flexible estimation approaches to demand estimation include Bajari et al. (2007), Fox et al. (2011), Fox et al. (2012), Fox et al. (2016) and Fox and Gandhi (2016). The goal in these papers is to recover the distribution of random coefficients in discrete choice settings, whereas I directly target the structural demand function. On the one hand, this allows for a broader range of consumer behaviors; on the other, as discussed above, it faces a curse of dimensionality. A recent paper by Tebaldi et al. (2019) proposes a method to obtain nonparametric bounds on demand counterfactuals, but does not develop inference procedures.

It should be emphasized that the present paper focuses 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 et al. (2004),[9] and, for more recent nonparametric approaches, Hausman and Newey (2016), Blundell et al. (2017) and Chen and Christensen (2018).

Second, the paper is related to the large literature on incomplete pass-through[10] 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 how much of the incomplete pass-through is explained by sellers adjusting their markups.[11] The present paper contributes to this literature by providing a method to evaluate markups that relaxes a number of restrictions on consumer behavior and preferences. In my empirical setting, I estimate a significantly larger reduction in markups after the tax—and thus a more incomplete pass-through—for the organic product relative to what is predicted by a more restrictive parametric model.[12]

Third, the paper relates to the literature investigating the sources of market power based on demand estimates, notably Nevo (2001).[13] Once again, I offer a more flexible method to disentangle and quantify the different components of market power. In my empirical setting,

[9]Of course, any method based on market-level data may be immediately applied to consumer-level data by simply aggregating the latter at the market level. However, recent work by Berry and Haile (2010) shows that within-market variation makes it possible to identify demand under weaker conditions relative to the case where only market-level data is available. This opens an interesting avenue for future research on nonparametric estimation of demand based on individual-level data.

[10]The literature on estimating pass-through is large and I do not attempt to provide an exhaustive list of references. Here, I 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.

[11]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. Competing explanations for incomplete pass-through considered in these papers are nominal rigidities and the presence of costs not affected by the shocks.

[12]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.

[13]Another approach to studying market power is based on estimates of the firm production function (de Loecker (2011), de Loecker and Warzynski (2012)).

I find that a mixed logit model with product fixed effects matches the nonparametric results very closely, suggesting that standard parametric models might be sufficient to address this type of questions.

Overview. 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. Section 4 presents the results of several Monte Carlo simulations. Section 5 contains the empirical application. Section 6 concludes. All proofs, additional simulations, and more details on the empirical application are presented in the appendices.

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. 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).

Next, I define the structural demand system

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

where $\mathcal{X}$ denotes the support of χ_t and 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. I 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. For example, this case arises when the goods are complements to each other and the interpretation of market shares as fraction of consumers preferring one good over all others does not apply.[14] I also define

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

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

Next, following BH, I restrict the way in which some of the variables in X enter demand. Specifically, I partition x_t as $\left(x_t^{(1)}, x_t^{(2)}\right)$, 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*

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

Then, for every market t, I assume that

$$\sigma\left(\chi_t\right) = \sigma\left(\delta_t, p_t, x_t^{(2)}\right) \tag{1}$$

where $\delta_t \equiv \left(\delta_{1t}, ..., \delta_{Jt}\right)$.[15] Equation (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. In a standard BLP-type discrete choice setting, a simple sufficient condition is that $x_j^{(1)}$ enters good j's indirect utility with a non-random coefficient. On the other hand, $x_t^{(2)}$ is allowed to enter the share function in an unrestricted fashion.[16] Two remarks about the restriction in (1) are in order. First, while (1) requires that the dimension of $x^{(1)}$ be equal to J, one could include more than one covariate in each linear index. In fact, this is one of the strategies for dimension reduction suggested in Section 3.2. Second, $x_j^{(1)}$ could be a characteristic of good j, but it need not be. The model allows for the case where $x^{(1)}$ includes market-level demand shifters that are not necessarily product-specific, as long as the dimension of $x^{(1)}$ is at least J. This is illustrated in the application of Section 5,

[14] See Example 1 in Berry et al. (2013) and the simulation in Section 4.3.

[15] As shown in Appendix B of BH, what is critical for identification is the strict monotonicity of δ_{jt} in ξ_{jt}. Both its linearity in $x_{jt}^{(1)}$ and its separability in ξ_{jt} can be relaxed. However, the assumption in (1) simplifies the estimation procedure in that it leads to an additively separable nonparametric regression model. Given that this is the first attempt at estimating demand nonparametrically for this class of models, maintaining (1) appears to be a reasonable compromise. 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)).

[16] Indeed, the case where $x_t^{(2)}$ does not enter the model at all is allowed.

where $x^{(1)}$ consists of variables that shift consumer preferences for strawberries but do not represent product characteristics.

Throughout the paper, I assume that the structural demand system σ is point-identified, which I record in the next assumption.

Assumption 1. *The structural demand system σ is point-identified.*

BH provide sufficient conditions for Assumption 1.[17] As one would expect, these conditions include the existence of instruments $Z = (Z_1, ..., Z_J)$ that shift prices but do not enter demand. In addition, Z, along with X, is assumed to be exogenous, i.e. $\mathbb{E}(\xi_j|X, Z) = 0$ a.s.-(X, Z) for $j \in \mathcal{J}$. I refer the reader to BH for a more detailed discussion of identification.

3 Nonparametric Estimation

3.1 Setup and asymptotic results

Under the sufficient conditions for identification, BH show that one can invert the demand system in (1) as follows:

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

This inverted system is the starting point of my estimation strategy. Specifically, I rewrite (2) as

$$x_{jt}^{(1)} = \sigma_j^{-1}\left(s_t, p_t, x_t^{(2)}\right) - \xi_{jt} \quad j = 1, ..., J, \tag{3}$$

where I use the normalization $\beta_j = 1$.[18] Equation (3), coupled with the IV exogeneity restriction, $\mathbb{E}(\xi_j|X, Z) = 0$, suggests estimating σ_j^{-1} using nonparametric instrumental variables (NPIV) methods.[19] In particular, I approximate the functions σ_j^{-1} via the method of sieves, i.e. using a sequence of models whose dimension grows with the sample size. For instance, in the case of polynomial approximations, the degree of the polynomials increases with the sample size. Therefore, the approach does not require one to assume any functional form

[17] BH show identification under two alternative approaches. The first is based on a completeness condition, while the second does not rely on this type of assumption and instead leverages joint restrictions from both the demand and the supply side.

[18] As noted in BH, we are free to normalize β_j to 1 since the unobservables ξ_{jt} have no natural scale.

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

asymptotically, which guards against misspecification bias. Implementing the procedure is straightforward in that, in practice, it amounts to estimating a (large) parametric model. On the other hand, proving theoretical properties of the estimator—e.g. establishing the validity of the standard errors for price elasticities—is more complicated due to the fact that the unknown parameter is an entire function as opposed to a finite-dimensional object. Specifically, I cannot rely on standard results from parametric models and I need to adapt recent results from the econometrics literature on NPIV.

Some additional notation is needed to formalize the approach. I denote by T the sample size, i.e. the number of markets in the data. While T grows to infinity asymptotically, the number of goods J is fixed. Let Σ be the space of functions to which σ^{-1} belongs[20] and let $\psi^{(j)}_{M_j}(\cdot) \equiv \left(\psi^{(j)}_{1,M_j}(\cdot), ..., \psi^{(j)}_{M_j,M_j}(\cdot)\right)'$ be the basis functions used to approximate σ_j^{-1} for $j \in \mathcal{J}$.[21] Note that, although I suppress it in the notation, M_j grows with T for all j. Let Σ_T be the resulting sieve space for Σ. Next, I denote by $a^{(j)}_{K_j}(\cdot) \equiv \left(a^{(j)}_{1,K_j}(\cdot), ..., a^{(j)}_{K_j,K_j}(\cdot)\right)'$ the basis functions used to approximate the instrument space for good j's equation, and I let $A_{(j)} = \left(a^{(j)}_{K_j}(x_1, z_1), ..., a^{(j)}_{K_j}(x_T, z_T)\right)'$ for $j \in \mathcal{J}$. Again, I suppress the dependence of $\{K_j\}_{j\in\mathcal{J}}$ on the sample size. I 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, I let $r_{jt}\left(s_t, p_t, x_t, z_t; \tilde{\sigma}_j^{-1}\right) \equiv \left(x^{(1)}_{jt} - \tilde{\sigma}_j^{-1}\left(s_t, p_t, x^{(2)}_t\right)\right) \times a^{(j)}_{K_j}(x_t, z_t)$. Then, the estimator solves the following GMM program[22]

$$\min_{\tilde{\sigma}^{-1}\in\Sigma_T} \sum_{j=1}^{J} \left[\sum_{t=1}^{T} r_{jt}\left(s_t, p_t, x_t, z_t; \tilde{\sigma}_j^{-1}\right)\right]' \left(A'_{(j)}A_{(j)}\right)^{-} \left[\sum_{t=1}^{T} r_{jt}\left(s_t, p_t, x_t, z_t; \tilde{\sigma}_j^{-1}\right)\right] \tag{4}$$

The solution $\hat{\sigma}^{-1}$ to (4) 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. When Σ_T is chosen to be a linear sieve (e.g. polynomials, splines, wavelets), the approximation to σ_j^{-1} will be of the form $\tilde{\sigma}_j^{-1} = \theta_j' \psi^{(j)}_{M_j}(\cdot)$ for $j \in \mathcal{J}$. This, in turn, implies that (4) will be a convex program in the coefficients θ, for which readily available algorithms are guaranteed to converge to the global minimizer. In contrast, BLP-type models typically require minimizing non-convex functions and thus may run into numerical challenges (Knittel and Metaxoglou (2014)). One caveat to the above is that, if one wants to impose non-convex constraints on θ, the optimization

[20] This class of functions will be formally defined in Appendix B.

[21] 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 1 below does not depend on this choice, hence the general notation used in the first part of this section.

[22] In equation (4), $\tilde{A}^{-}$ denotes the Moore-Penrose inverse of a matrix $\tilde{A}$.

problem will become harder. One such constraint is the symmetry of the Jacobian of demand with respect to prices (see Appendix D.1). On the other hand, several other constraints, including monotonicity and the exchangeability constraint considered in Section 3.2, are linear and thus can be handled with off-the-shelf convex optimization methods. For the case with linear constraints, I recommend using the Matlab package CVX,[23] whereas in the presence of nonlinear (and possibly nonconvex) constraints I found Knitro to perform well.[24]

I now state a result that yields asymptotically valid standard errors for functionals of the demand system. In turn, this may be used to obtain confidence intervals for quantities of interest (e.g. price elasticities) and test hypotheses on consumer behavior. The result adapts Theorem D.1 in Chen and Christensen (2017) (henceforth, CC).[25] Note that CC consider a model with only one equation and one unknown function, whereas the setting here involves J equations, each with a distinct unknown function and error term ξ_j. This requires imposing additional (mild) restrictions on the covariance matrix of the errors and modifying the proof accordingly. For conciseness, the formal assumptions needed for the result and the definition of the estimator for the variance of the functional are postponed to Appendix B. In words, the assumptions restrict: (i) the distribution of the error terms by way of standard bounded moment conditions; (ii) the rate at which the dimension of the approximation to the unknown functions grows with the sample size; (iii) the rate of convergence of the nonparametric estimator for the demand functions and their derivatives. The restriction in (iii), which is formalized in Assumption 7 in Appendix B, is high-level and I provide more primitive sufficient conditions for two special cases of interest in Theorems 2 and 3 below.

Theorem 1. *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 B. In addition, let Assumptions 1, 2, 3, 4, 5, 6 and 7 in Appendix B hold. Then,*

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

[23]See Grant and Boyd (2008),Grant and Boyd (2014).

[24]See R. H. Byrd and Waltz (2006).

[25]Note that CC consider inference on functionals of an *unconstrained* sieve estimator of the unknown function, whereas our model features a range of equality and inequality constraints. However, under the assumption that the true demand functions satisfy the inequality constraints strictly, the constrained and unconstrained estimators will coincide asymptotically and thus the unconstrained standard errors will be valid in large samples. Recent papers by Chernozhukov et al. (2015) and Freyberger and Reeves (2018) develop inference procedures for constrained estimators that could be applied to our model. I leave comparing different nonparametric methods for future research.

Proof. See online Appendix C. □

Importantly, the variance term $\hat{v}_T(f)$ in the statement of the theorem is allowed to grow to infinity with the sample size, implying that the result covers the scenario in which the functional f is estimable at a rate slower than the parametric rate $\sqrt{T}$. This will typically be the case when, as in the empirical analysis of Section 5, the functionals of interest are defined for a fixed market, as opposed to being averages across markets.

I now specialize Theorem 1 to two functionals: price elasticities and equilibrium prices. These quantities are key inputs in addressing many (counterfactual) questions in industrial organization. Consistent with the empirical application of Section 5, I also assume that $J = 2$ and that the unknown functions are approximated via Bernstein polynomials. I state the results here and again postpone the full presentation of the assumptions, as well as the proofs, to Appendix B. In words, Theorems 2 and 3 replace the high-level Assumption 7 in Theorem 1 with sufficient conditions on the smoothness of the unknown functions, the support of the endogenous variables, and the growth rate of the sieve approximation. These are standard assumptions in the NPIV literature. Lemmas 3 and 4 in Appendix B provide even more concrete restrictions for the "mildly ill-posed case," i.e. the scenario where a measure of the degree of endogeneity in the nonparametric problem grows polynomially with the dimension of the sieve space.[26]

Theorem 2. *Let f_ϵ be the own-price elasticity functional defined in* (14) *in Appendix B, 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 1, 2, 3, 4(iii), 5, 6 and 8 from Appendix B hold. Then,*

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

Proof. See Appendix B. □

Next, I state a result establishing the asymptotic distribution of equilibrium prices.

Theorem 3. *Let f_{p_1} be the equilibrium price functional defined in* (18) *in Appendix B, let $\hat{v}_T(f_{p_1})$ denote the estimator of the standard deviation of $f_{p_1}(\hat{\sigma}^{-1})$ based on* (13)*, and let Assumptions 1, 2, 3, 4(iii), 5, 6 and 9 from Appendix B hold. Then,*

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

[26]See Blundell et al. (2007) for a formal definition of the measure of ill-posedness and CC for a discussion of its estimation.

Proof. See Appendix B. □

In the empirical application of Section 5, I apply Theorem 2 to obtain confidence intervals for own-and cross-price elasticities, and Theorem 3 to obtain confidence intervals for equilibrium prices under two counterfactual scenarios.

3.2 Constraints

I conclude this section with a discussion of the curse of dimensionality that is inherent in nonparametric estimation, as well as of ways to partially mitigate 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.

One way to mitigate this problem is to impose constraints on the estimated demand functions. I consider a number of restrictions, including exchangeability of the demand functions,[27] monotonicity and lack of income effects. Some constraints—such as 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 an important role in disciplining the estimation routine and obtaining reasonable estimates (e.g. negative own-price elasticities across all markets).[28] I emphasize that this is not an exhaustive list, and one may wish to impose additional constraints in a given application. Conversely, not all constraints discussed in this paper need to be enforced simultaneously in order to make the approach feasible. For example, in the empirical application in Section 5, I do not assume lack of income effects nor exchangeability.

Imposing constraints in model (3) 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, I propose to estimate the functions σ_j^{-1} in (3) using Bernstein polynomials, which are convenient for imposing economic restrictions due to their approximation

[27] Similar exchangeability restrictions are discussed in Pakes (1994), Berry et al. (1995) and Gandhi and Houde (2017) in relation to optimal instruments.

[28] Blundell et al. (2017) make a similar point regarding Slutsky negative semidefiniteness constraints in a model with one good plus a numeraire.

properties. Here I provide an informal discussion to convey the main insight and postpone a more formal presentation to Appendix A. Consider the problem of approximating a bounded function $g(t_1, t_2)$, where $(t_1, t_2) \in [0,1] \times [0,1]$, using the tensor product of the univariate Bernstein basis of degree 2 in each of the arguments t_1, t_2. Since there are three terms for each argument, the approximation is a linear combination of nine terms in total. Let θ be the vector of coefficients on these nine terms. By the properties of Bernstein polynomials, it is possible to choose the coefficients θ such that they are close to the true value of g at a grid of points over the $[0,1] \times [0,1]$ square. This means that one may arrange the coefficients in a matrix such that the following holds[29]

$$\begin{bmatrix} \theta_{11} & \theta_{12} & \theta_{13} \\ \theta_{21} & \theta_{22} & \theta_{23} \\ \theta_{31} & \theta_{32} & \theta_{33} \end{bmatrix} \approx \begin{bmatrix} g(0,0) & g(0,0.5) & g(0,1) \\ g(0.5,0) & g(0.5,0.5) & g(0.5,1) \\ g(1,0) & g(1,0.5) & g(1,1) \end{bmatrix} \tag{5}$$

where the $\approx$ sign indicates that each element on the left matrix will be close to the corresponding element of the right matrix. As the degree of the polynomial grows, the approximation becomes arbitrarily good. This is helpful because it allows us to immediately translate restrictions on g into restrictions on θ. For instance, the constraint that g be weakly increasing in its first argument implies the simple inequalities $\theta_{1i} \leq \theta_{2i} \leq \theta_{3i}$ for $i = 1, 2, 3$. Appendix D discusses monotonicity and other restrictions that economic theory places on the functions σ_j^{-1} and shows how to enforce them using the approximation properties of Bernstein polynomials.

The remainder of this section focuses on two types of constraints that are especially helpful in alleviating the curse of dimensionality: exchangeability and index restrictions In order to define exchangeability, let $\pi : \{1, ..., J\} \mapsto \{1, ..., J\}$ be a permutation with inverse π^{-1} and, for simplicity, let $x^{(2)} = \left(x_1^{(2)}, ..., x_J^{(2)}\right)$, i.e. I assume that $x^{(2)}$ is a vector of product-specific characteristics.[30] Also, let $\tilde{n}_{x^{(2)}}$ be the dimension of each $x_j^{(2)}$, so that $n_{x^{(2)}} = J\tilde{n}_{x^{(2)}}$. Then, the structural demand system σ is exchangeable if

$$\sigma_j\left(\delta, p, x^{(2)}\right) = \sigma_{\pi(j)}\left(\delta_{\pi^{-1}(1)}, ..., \delta_{\pi^{-1}(J)}, p_{\pi^{-1}(1)}, ..., p_{\pi^{-1}(J)}, x^{(2)}_{\pi^{-1}(1)}, ..., x^{(2)}_{\pi^{-1}(J)}\right), \tag{6}$$

for $j = 1, ..., J$. In words, this means that the demand functions do not depend on the

[29]Lemma 2 in Appendix A clarifies which coefficient belongs to each element of the matrix.

[30] This need not be the case in the general model from Section 2. For instance, $x^{(2)}$ could be a vector of market-level variables. In such settings, I say the demand system is exchangeable if $\sigma_j\left(\delta, p, x^{(2)}\right) = \sigma_{\pi(j)}\left(\delta_{\pi^{-1}(1)}, ..., \delta_{\pi^{-1}(J)}, p_{\pi^{-1}(1)}, ..., p_{\pi^{-1}(J)}, x^{(2)}\right)$, 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.

identity of the products, but only on their attributes $\left(\delta, p, x^{(2)}\right)$.[31] For instance, for $J = 3$, exchangeability implies 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 $\left(\delta_1, \overline{\delta}, \underline{\delta}, p_1, \overline{p}, \underline{p}, x_1^{(2)}, \overline{x}^{(2)}, \underline{x}^{(2)}\right)$, i.e. the demand for good 1 is the same if we switch the labels for goods 2 and 3. 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.[32]

Moreover, one may allow for additional flexibility by letting the intercepts of the δ indices vary across goods. This preserves the advantages of exchangeability in terms of dimension reduction, which I discuss below, while simultaneously allowing each unobservable to have a different mean. Relative to existing methods, this is no more restrictive than standard mixed 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 D.3. □

Lemma 1 implies that one can directly impose exchangeability on the target functions σ^{-1}. This can be achieved by simply using the same set of Bernstein coefficients to approximate all demand functions and imposing that the value of each function be invariant to certain permutations of its arguments. By the approximation properties of Bernstein polynomials, these restrictions can be conveniently enforced through linear constraints on the Bernstein coefficients. To see this, note that this invariance property amounts to setting appropriately chosen pairs of coefficients equal to each other. For instance, in the simple illustration in (5), the restriction that g be invariant to permutations of its two arguments translates (asymptotically) into the constraints $\theta_{21} = \theta_{12}$, $\theta_{31} = \theta_{13}$ and $\theta_{32} = \theta_{23}$. One can impose

[31] 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 firm. The arguments in this section would then apply to the subset of products on which the restriction is imposed.

[32] This also uses the fact that the idiosyncratic taste shocks are typically assumed to be iid—and thus exchangeable—across goods.

these as equality restrictions or directly embed them in the estimation routine by minimizing the criterion function over the lower-dimensional space of free parameters (there are six free parameters out of nine in the example in (5)). The latter may be preferable when the number of parameters is large and memory or speed considerations take center stage in computation. A more formal discussion of how to impose exchangeability is provided in Appendix A.

Second, I consider index restrictions. Specifically, suppose we are willing to assume that $x^{(2)}$ enters the demand functions through the indices δ. Then, each demand function goes from having $2J+n_{x^{(2)}}$ to $2J$ arguments, which reduces the number of parameters.[33] Similarly, if we are willing to assume that prices enter the demand functions through the indices δ, each demand function goes from having $2J+n_{x^{(2)}}$ to $J+n_{x^{(2)}}$ arguments. Thus, to a certain extent, it is possible to tailor the approach based on the setting and sample size at hand by choosing how much to assume in terms of functional form. Further, note that, while the index restriction does have bite, including variables in the linear index does not mean that they are restricted to enter the demand functions linearly. As discussed in Section 2, the content of this assumption is that the variables in the index and the unobservables ξ must be perfect substitutes in the "production" of utility. For instance, in a discrete choice model, a sufficient condition is that the variables have nonrandom coefficients, but they are allowed to enter the demand functions in highly nonlinear ways. Further, index restrictions do not impose any constraints on the distribution of the unobservables and are thus consistent with the goal of relaxing the arbitrary distributional assumptions often made in estimating demand parametrically.

To illustrate the role played these constraints in alleviating the curse of dimensionality, I show in Table 1 how the number of parameters for each demand function grows with J depending on whether I do or do not impose exchangeability and the index restriction on p. While the dimension of the model grows large with J in both cases, the curse of dimensionality is much more severe when exchangeability or the index restriction are not imposed—indeed to the point where estimation becomes computationally intractable. Thus, 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.

[33]In particular, the number of Bernstein coefficients for each demand function goes from $(m+1)^{2J+n_{x^{(2)}}}$ to $(m+1)^{2J}$, where for simplicity I assume the degree m of the polynomials is the same for all arguments.

Table 1: Number of parameters with and without exchangeability and index restriction on price

	p in Index		p not in Index	
J	Exchangeability	No Exchangeability	Exchangeability	No Exchangeability
3	10	27	405	729
5	45	243	4,455	59,049
7	84	2,187	27,027	4.78×10^6
10	165	59,049	218,790	3.49×10^9

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

4 Monte Carlo Simulations

This section presents the results of Monte Carlo simulations. There are three goals. First, I illustrate that the estimation procedure works well with moderate sample sizes—indeed much smaller than the sample size used in the empirical application and other readily available supermarket scanner datasets. Second, I 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. Finally, I investigate the performance of the estimator as the number of goods increases.

I compare the performance of the nonparametric demand approach (NPD for short) to that of standard methods. Specifically, I take as a benchmark a random coefficient logit model with normal random coefficients. I refer to this model as BLP. In order to summarize the results, I plot the own- and cross-price elasticities as a function of the own price, since these functions are key inputs to many counterfactuals of interest. For instance, the shape of the own-price elasticity function will turn out to play an important role in determining the pass-through rate of a tax in the application of Section 5. In each plot, all market-level variables different from the own-price are fixed at their median values. All simulations are for the case with $J = 2$ number of goods (except for Section 4.4), $T = 3,000$ number of markets, and 200 Monte Carlo repetitions. Appendix E presents additional simulation designs in which the sample size is lower ($T = 500$), the number of goods is larger than two, and the index restriction is violated.

4.1 Correctly specified BLP model

First, I generate data from a mixed logit model with normal random coefficients. This means that the BLP procedure is correctly specified and therefore performs well. On the other hand, one would 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 sheds 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

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

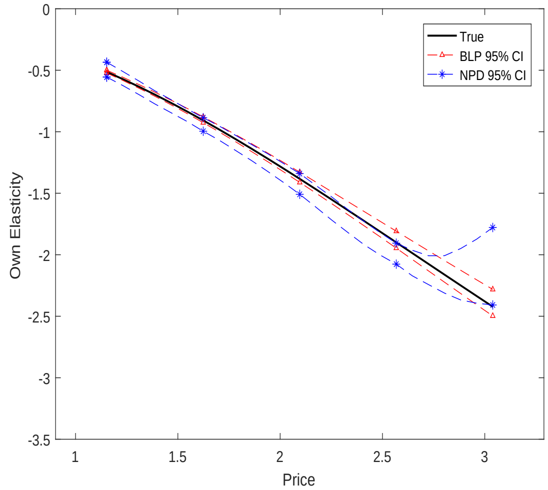
where ϵ_{ij} is independently and identically distributed (iid) extreme value across goods and consumers, α_i is distributed $N(-1, 0.15^2)$ iid across consumers, and I set $\beta = 1$. The exogenous shifters x_j are drawn from a uniform $[0, 2]$ distribution,[34] whereas the unobserved quality indices ξ_j are distributed $N(1, 0.15^2)$. Excluded instruments z_j are drawn from a uniform [0,1] distribution and I generate prices according to $p_j = 2(z_j + \eta_j) + \xi_j$, where η_j is uniform [0,0.1].[35]

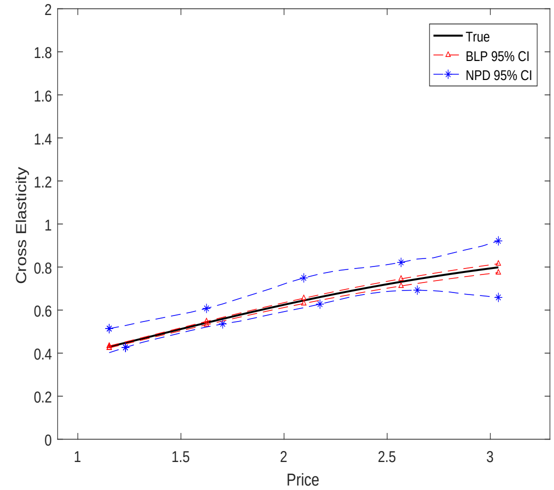
When estimating demand nonparametrically, I impose the following constraints from Section 3 and Appendix D: exchangeability, diagonal dominance of the Jacobian of σ, $\mathbb{J}_\sigma^\delta$, and monotonicity of σ. Figure 1 shows the own- and cross-price elasticity functions for good 1, respectively. 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, I take this as suggestive that the penalty one pays when ignoring correct parametric assumptions in finite samples may not be substantial.

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

[35] Note that, while I do not specify a supply model, 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 (left) and cross-price (right) elasticity functions





One may wonder how robust the nonparametric estimates in Figure 1 are to the choice of the tuning parameter, i.e. the polynomial degree for the Bernstein approximation. Table 2 shows how the estimator for the median own- and cross-price elasticities performs as the tuning parameter changes. While, as expected, the bias tends to decrease and the standard deviation to increase with the polynomial degree, the own- and cross-price elasticities are pinned down reasonably well for a range of tuning parameters. Appendix E.4 provides more results suggesting that this does not just hold for the median levels, but also for the entire elasticity functions.

Table 2: Performance of nonparametric estimator for median own- and cross-price elasticities as the polynomial degree varies

	True	Degree	Bias	S.E.	MSE
Own	-1.339	6	0.079	0.024	0.007
		8	0.035	0.023	0.002
		12	0.003	0.028	0.001
Cross	0.569	6	-0.055	0.009	0.003
		8	-0.059	0.011	0.004
		12	-0.038	0.014	0.002

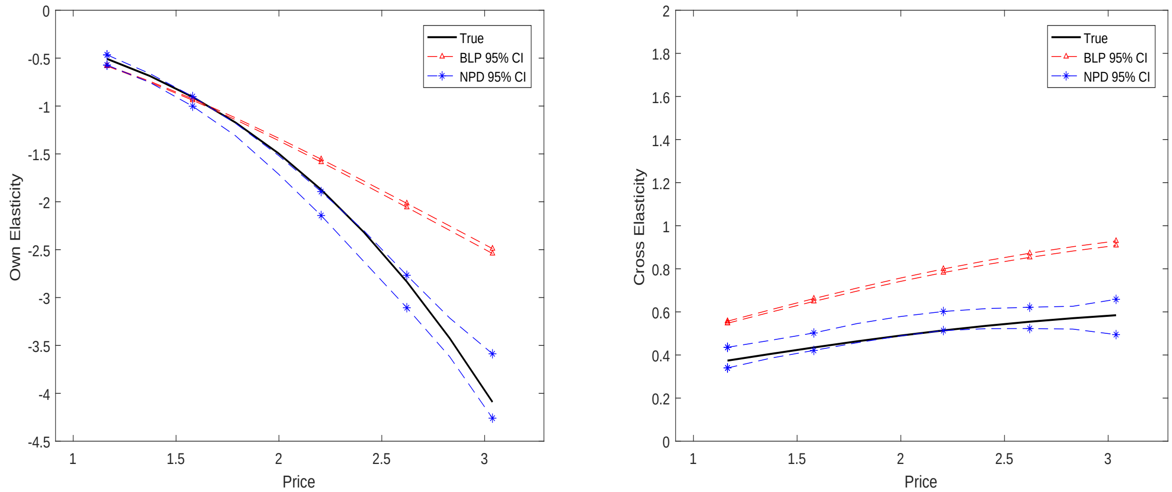
4.2 Inattention

Next, I consider a discrete choice setting with inattention. In any given market, I assume a fraction of consumers ignore good 1 and therefore maximize their utility over good 2 and the outside option only. The remaining consumers consider all goods. I 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. items that are on sale might have a special display in supermarkets or options might be filtered from cheapest to most expensive on a e-commerce platform). Except for the presence of inattentive consumers, the simulation design is the same as in Section 4.1. In nonparametric estimation, I impose the following constraints: diagonal dominance of $\mathbb{J}_{\sigma}^{\delta}$ and monotonicity of σ.[36] Note that I 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, I allow different constants for the two goods.

Figure 2 shows the results for good 1. The nonparametric method captures the shape of both the own- and the cross-price elasticity functions, whereas BLP tends to underestimate the own-price elasticity and overestimate the cross-price elasticity. Intuitively, BLP does not capture the fact that, as the price of good 1 increases, more and more consumers ignore good 1. This results in a BLP own-price elasticity that is too low in absolute value. Similarly, the BLP model does not capture the fact that, as the price of good 2 increases, a fraction of customers will not switch to good 1 because they ignore it. This leads to a BLP cross-price elasticity that is too high.

[36] See Appendix D for a discussion of these constraints.

Figure 2: Inattention: Own-price (left) and cross-price (right) elasticity functions



4.3 Complementary goods

I now consider a setting where good 1 and 2 are not substitutes, but complements. I generate the exogenous covariates and prices as in the previous two simulations,[37] but I 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.[38] In the NPD estimation, I impose the following constraints: monotonicity of σ, diagonal dominance of $\mathbb{J}_\sigma^\delta$ and exchangeability.[39]

Figure 3 shows 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

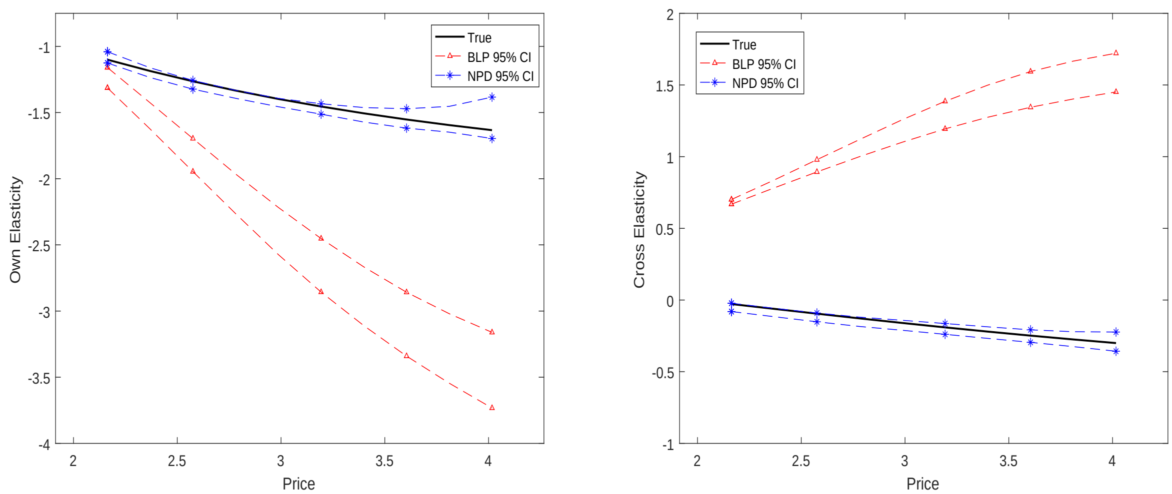
[37] One difference is that I now take the mean of ξ_1 and ξ_2 to be 2 instead of 1 in order to obtain shares that are not too close to zero.

[38] See also Example 1 in Berry et al. (2013).

[39] See Section 3.2 and Appendix D for a discussion of these constraints.

are mostly off target, consistent with the fact that a discrete choice model is not well-suited to estimate demand for (divisible) complements. In particular, the discrete choice framework implies that the goods are substitutes and thus forces the cross-price elasticity to be positive.

Figure 3: Complements: Own-price (left) and cross-price (right) elasticity functions



4.4 $J > 2$ Goods

The simulation designs considered so far featured $J = 2$ goods, which corresponds to the setting of the empirical application. However, researchers are often interested in modeling demand for a larger number of goods. To this end, I now investigate the performance of the estimator as the number of products increases. To tackle the curse of dimensionality that arises as J grows, I both impose exchangeability and restrict prices (as well as the x attributes) to enter the indices δ in estimation. As discussed in Section 3.2, both these constraints substantially reduce the number of parameters to estimate.

The data is generated from the discrete choice dgp from Section 4.1 with one difference: the coefficients (α, β) on the product attributes (p_j, x_j) are now drawn from a discrete distribution and are correlated.[40] Because the product attributes have random coefficients, the index restriction is not satisfied (ξ_j does not enter the demand functions in the same way as x_j or p_j) and thus the nonparametric model I estimate is misspecified. In addition, the

[40]Specifically, I draw α from the distribution that places equal probabilities on the values -3 and -0.5 and set $\beta = -\alpha$, so that there are two types of consumers, one that places low weights and one that places high weights on the observable product attributes.

Table 3: Performance of estimators for median own- and cross-price elasticities as J varies

	J	True	NPD Bias	NPD S.E.	NPD MSE	BLP Bias	BLP S.E.	BLP MSE
Own	3	-1.322	-0.017	0.049	0.003	-0.980	0.052	0.963
	5	-1.458	-0.065	0.078	0.010	-1.479	0.089	2.195
	10	-1.559	0.429	0.088	0.191	-0.857	0.137	0.752
Cross	3	0.277	-0.088	0.022	0.008	0.247	0.217	0.108
	5	0.173	-0.015	0.019	0.001	0.050	0.035	0.004
	10	0.091	-0.048	0.012	0.003	0.017	0.050	0.003

Note: Mixed logit dgp with correlated discrete random coefficients. Both the NPD and the BLP model are misspecified.

BLP model I estimate is also misspecified in that it incorrectly assumes that the random coefficients are normally distributed and independent of each other. Comparing the performance of the two estimators then illustrates the relative impact of two different types of misspecification: (i) that arising from incorrect distributional assumptions in a parametric model, and (ii) that stemming from incorrectly imposing the index restriction in the proposed nonparametric approach. Table 3 shows that, as the number of goods ranges from 3 to 10, the nonparametric approach consistently outperforms the parametric one in pinning down the cross-price and especially the own-price elasticities. This suggests that even the more restrictive version of the nonparametric estimator that imposes both exchangeability and the index restriction in all the product attributes might be preferable to a parametric model that makes incorrect distributional assumptions on the unobservables. Appendix E.3 further explores the robustness of the nonparametric approach to increasing violations of the index restriction, while Appendix E.5 contains estimates for the entire own- and -cross elasticity functions for the $J > 2$ case.

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

In this section, I use the proposed nonparametric procedure to investigate the robustness of two counterfactual exercises to the parametric specification of demand. First, I 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.[41] Therefore, flexibly capturing the shape of the demand function is crucial to obtaining accurate results.

The second counterfactual concerns the role played by the multi-product nature of retailers

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

in driving up markups. Specifically, a firm simultaneously pricing multiple (substitute) goods is able to internalize the competition that would occur if those goods were sold by different firms, thus pushing prices upwards.[42] Quantifying the magnitude of this effect is ultimately an empirical question which again depends on the shape of the demand functions.

5.1 Data

I use data on sales of fresh fruit at stores in California. Specifically, I focus on strawberries, and look at how consumers choose between organic strawberries, non-organic strawberries and other fresh fruit, which I 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, one 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,[43] which means that they cannot be artificially ripened using ethylene.[44] 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 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.

I instrument for prices using Hausman-type IVs.[45] In addition, for the inside goods, I also use shipping-point spot prices, as a proxy for the wholesale prices faced by retailers. Besides prices, I 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 G provides further details on the construction of the dataset, as well as some summary statistics and results for the first-stage regressions.

---

[42]This is one of the determinants of markups considered by Nevo (2001) in his analysis of the ready-to-eat cereal industry.

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

[44]Unlike climacteric fruits, such as bananas.

[45]See Hausman (1996).

5.2 Model

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

$$\begin{aligned}
s_1 &= \sigma_1\left(\delta_{str}, \delta_{org}, p_0, p_1, p_2, x^{(2)}\right) \\
s_2 &= \sigma_2\left(\delta_{str}, \delta_{org}, p_0, p_1, p_2, x^{(2)}\right) \\
\delta_{str} &= \beta_{0,str} - \beta_{1,str} x^{(1)}_{str} + \xi_{str} \\
\delta_{org} &= \beta_{0,org} + \beta_{1,org} x^{(1)}_{org} + \xi_{org}
\end{aligned} \tag{7}$$

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 a measure of taste for organic products,[46] $x^{(1)}_{str}$ 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, 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, I 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, I view this 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, Appendix F shows that many objects of interest, including the counterfactuals in Section 5.4, are robust to certain forms of endogeneity arising through this channel. Note that in model (7), the exogenous shifters $x^{(1)}_{str}$ and $x^{(1)}_{org}$ are not product-

[46] Specifically, I take the percentage of organic sales over total yearly sales in the lettuce category at the store.

[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 sold in California supermarkets systematically varied with the harvesting patterns of other fresh fruits. This does not seem to be a first-order concern given that (i) strawberries are harvested in California essentially year-round; and (ii) more than 90% of all strawberries produced in the US are grown in California (United States Department of Agriculture (2017)).

specific characteristics, but rather market-level variables. As highlighted in Section 2, the framework of the paper accommodates this.

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

$$\begin{aligned} u_{i,1} &= \beta_1 + \left(\beta_{p,i} + \beta_{x^{(2)}} x^{(2)}\right) p_1 + \beta_{p,0} p_0 + \beta^{par}_{str} x^{(1)}_{str} + \xi_1 + \epsilon_{i,1} \\ u_{i,2} &= \beta_2 + \left(\beta_{p,i} + \beta_{x^{(2)}} x^{(2)}\right) p_2 + \beta_{p,0} p_0 + \beta^{par}_{str} x^{(1)}_{str} + \beta^{par}_{org} x^{(2)}_{org} + \xi_2 + \epsilon_{i,2} \end{aligned} \tag{8}$$

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 $\beta_{p,i}$ can take one of two values across consumers.[49]

Comparing model (7) to model (8) illustrates the flexibility of the approach proposed in this paper. The latter model 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 (7) allows for a broader range of consumer behaviors, including continuous choice, as I show in Appendix H.2. This is one of the advantages of targeting the structural demand function directly as opposed to the underlying utility parameters.

In order to perform the counterfactual exercises in Section 5.4, I need to take a stand on the supply side. Following the retail literature, I make the assumption that each store acts as a monopolist when choosing strawberry prices. This model of supply is justified if consumers do not compare prices across stores when making their strawberry purchase decisions, which seems to be a reasonable assumption.

5.3 Estimation Results

First, I present the results from the mixed logit model in Table 4.[50] All coefficients have the expected signs and are significantly different from zero (except for the coefficient on taste for organic products, which is not significant).

[49] Following the original BLP paper, I 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, I present the two-point distribution because it is slightly more flexible than the one with normal coefficients. Specifically, the former has three parameters for the distribution of the random coefficient (the two values plus the probability of, say, the first value), while the latter has two parameters (the mean and the variance of the normal distribution).

[50] Estimation of the mixed logit model follows the recommendations in Nevo (2000b) and Dubé et al. (2012).

Table 4: Mixed logit estimation results

Variable	Type I	Type II
Price	-7.58 (0.07)	-89.85 (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 (0.00)	0.18 (0.00)

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

Turning to nonparametric estimation, I impose the constraints on the Jacobian of demand discussed in Section D.2, but do not impose exchangeability. Thus, I allow the organic and non-organic category to have different demand functions. Further, I choose the degree of the polynomials for the Bernstein approximation based on a two-fold cross-validation procedure.[51] To summarize the nonparametric estimation results, I show in Table 5 the median estimated own- and cross-price elasticities for the two inside goods.

Table 5: Nonparametric estimation results

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

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

In order to compare the fit of the nonparametric model relative to the mixed logit model, I follow the same two-fold cross-validation approach used to choose the degree for the Bernstein polynomial approximation. As shown in Table 6, the greater flexibility of the NPD model translates into a lower average MSE.

[51] See, e.g., Chetverikov and Wilhelm (2017). Specifically, I partition the sample into two subsamples of equal size. Then, I estimate the model using the first subsample and compute the mean squared error (MSE) for the second subsample. I 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. I 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 6: Two-Fold Cross-Validation Results

	NPD	Mixed Logit
MSE	0.93	2.38

5.4 Counterfactuals

I use the estimates to address two counterfactual questions. First, I consider the effects of a per-unit tax on prices.[52] In each market, I compute the equilibrium prices when a tax is levied on each of the inside goods individually. I set the tax equal to 25% of the price for the product in that market. As shown in Table 7, 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 4 I plot uniform confidence bands for the own-price elasticity of the organic product as a function of its price.[53,54] The own-price elasticity estimated nonparametrically is much steeper than the parametric one. This is consistent with the pass-through results. All else equal, 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.

Table 7: 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)	($5 \cdot 10^{-4}$)

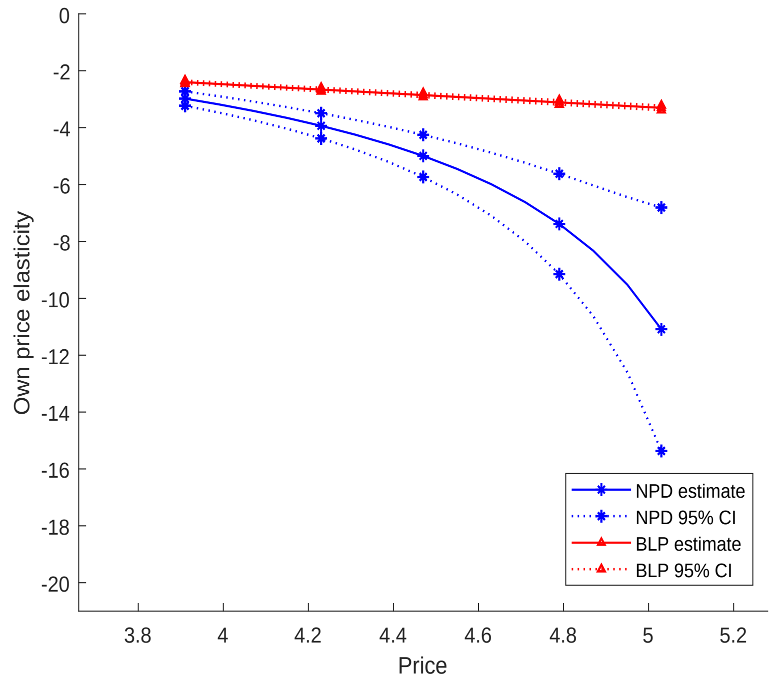
Note: Median changes in prices as a percentage of the tax. 95% confidence intervals in parentheses.

[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, one 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. I set all other variables at their median levels, except for δ_2 which I 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.

[54]The uniform confidence bands are obtained by applying the score bootstrap procedure described in CC.

Figure 4: Organic strawberries: Own-price elasticity function



As a second counterfactual experiment, I quantify the "portfolio effect." Specifically, I 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, instead of a two-product monopolist. I assume the two retailers compete on prices, compute the resulting equilibrium and compare it to the observed (monopoly) prices.[55] This type of exercise is instrumental to assessing the impact of large retailers on consumer prices. Specifically, it provides a measure of the upwards pressure on prices given by the fact that a retailer selling multiple products is able to partially internalize price competition. On the other hand, large retailers might tend to charge lower prices due to, among other things, economies of scale or loss-leader behavior (see , e.g., Lal and Matutes (1994), Lal and Villas-Boas (1998) and Chevalier et al. (2003)). Quantifying these different effects on prices is ultimately an empirical question that requires reliable estimates of demand.

Table 8 reports 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 (8)—labeled Mixed Logit (I)—and the nonparametric approach yield very similar results. In the median market, both models attribute around 10% and just above 40% of markups

[55]Since this counterfactual exercise amounts to splitting a monopoly into a duopoly, it is related to the literature on merger analysis. See, e.g., Nevo (2000a) and Jaffe and Weyl (2013) and references therein.

Table 8: Effect of multi-product pricing

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

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 (8). Mixed Logit (II) refers to the model in (8) with $\beta_1 = \beta_2$; Mixed Logit (III) refers to the model in (8) with $\beta_1 = \beta_2 = 0$.

to the portfolio effect for non-organic and organic strawberries, respectively. In other words, markups would be 10% to 40% lower in the scenario with two competing single-product retailers. One may wonder how robust this result is to modifications of the parametric specification. To this end, I estimate two additional models—labeled Mixed Logit (II) and (III)—that restrict the constants in model (8) to be the same and to be zero, respectively. Thus, while 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 a larger share of markups to the portfolio effect relative to the more flexible parametric specification or the nonparametric approach. 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, I develop and apply a nonparametric approach to estimate demand in differentiated products markets. The methodology relaxes several arguably arbitrary restrictions on consumer behavior and preferences that are embedded in standard discrete choice models. I achieve this by estimating the demand functions nonparametrically and leveraging a number of constraints from consumer theory. Further, I provide primitive conditions sufficient to obtain valid standard errors for quantities of interest.

I then use the approach as a benchmark to test the robustness of counterfactual outcomes given by standard parametric methods. While I find that a standard model yields a higher tax pass-through for one product relative to the nonparametric approach, an exercise designed to quantify the upward pressure on prices given by the multi-product nature of sellers suggests that a flexible enough parametric model captures the patterns in the data well.

This paper opens several avenues for future research. First, it would be interesting to explore additional ways to tackle the curse of dimensionality and thus enhance the applicability of the approach. For example, in markets with dozens of goods the current methodology would typically be unfeasible. However, if good j is effectively only competing with a handful of other products, then the remaining products' prices and characteristics do not enter good j's demand function, which would substantially reduce the dimensionality of the model. Therefore, developing a data-driven way of selecting the relevant set of competitors for a given product appears to be a promising line of research. Second, while the counterfactual analysis in this paper suggest that the nonparametric approach may be used to guide the choice among parametric specifications, additional work is required to make this argument formal. In this respect, the statistics literature on focused model selection might provide valuable insights. Finally, it would be interesting to apply the methodology proposed here to a broader range of empirical settings. For instance, a recent paper by Adao et al. (2017) shows that many questions of interest in international trade may be addressed by considering an economy where countries directly exchange factors of production instead of goods. While their identification argument is nonparametric, they estimate a parametric model in practice. Given that production factors are low dimensional, pursuing a more flexible approach seems feasible in their setting. One could then assess how robust the results are to the maintained parametric assumptions.

References

Abaluck, J. and A. Adams (2017). What do consumers consider before they choose? identification from asymmetric demand responses. IFS Working Paper W17/09.

Adao, R., A. Costinot, and D. Donaldson (2017). Nonparametric counterfactual predictions in neoclassical models of international trade. *American Economic Review 107*, 633–689.

Atkin, D. and D. Donaldson (2015). Who's getting globalized? the size and implications of intra-national trade costs. NBER Working Paper No. 21439.

Bajari, P., J. Fox, and S. Ryan (2007). Linear regression estimation of discrete choice models with nonparametric distributions of random coefficients. *AER Papers and Proceedings*, 459–463.

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).

Berry, S. and P. Haile (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.

Berry, S., J. Levinsohn, and A. Pakes (2004). Differentiated products demand systems from a combination of micro and macro data: The new car market. *Journal of Political Economy 112*, 68–105.

Blundell, R., X. Chen, and D. Kristensen (2007). Semi-nonparametric iv estimation of shape-invariant engel curves. *Econometrica 75*(6), 1613–1669.

Blundell, R., J. Horowitz, and M. Parey (2017). Nonparametric estimation of a nonseparable demand function under the slutsky inequality restriction. *The Review of Economics and Statistics 99*(2), 291–304.

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.

Chen, X. and T. Christensen (2017). Optimal sup-norm rates and uniform inference on nonlinear functionals of nonparametric iv regression. arXiv:1508.03365.

Chen, X. and T. Christensen (2018). Optimal sup-norm rates and uniform inference on nonlinear functionals of nonparametric iv regression. *Quantitative Economics 9*, 39–84.

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

Chen, X. and D. Pouzo (2012). Estimation of nonparametric conditional moment models with possibly nonsmooth generalized residuals. *Econometrica 80*, 277–321.

Chen, X. and D. Pouzo (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.

Chernozhukov, V., W. Newey, and A. Santos (2015). Constrained conditional moment restriction models. Working Paper (arXiv:1509.06311).

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.

de Loecker, J. (2011). Recovering markups from production data. *International Journal of Industrial Organization 29*, 350—-355.

de Loecker, J. and F. Warzynski (2012). Markups and firm-level export status. *American Economic Review 102*, 2437—-2471.

Deaton, A. and J. Muellbauer (1980, June). An almost ideal demand system. *AER 70*(3), 312–326.

Dubé, J., J. Fox, and C.-L. Su (2012). Improving the numerical performance of blp static and dynamic aggregate discrete choice random coefficients demand estimation. *Econometrica 80*, 2231–2267.

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.

Fox, J. and A. Gandhi (2016). Nonparametric identification and estimation of random coefficients in multinomial choice models. *RAND Journal of Economics 47*(1), 118–139.

Fox, J., K. il Kim, S. Ryan, and P. Bajari (2011). A simple estimator for the distribution of random coefficients. *Quantitative Economics 2*, 381–418.

Fox, J., K. il Kim, S. Ryan, and P. Bajari (2012). The random coefficients logit model is identified. *Journal of Econometrics 166*, 204–212.

Fox, J., K. il Kim, and C. Yang (2016). A simple nonparametric approach to estimating the distribution of random coefficients in structural models. *Journal of Econometrics 195*, 236–254.

Freyberger, J. and B. Reeves (2018). Inference under shape restrictions. Working Paper.

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

Gandhi, A. and J. F. Houde (2017). Measuring substitution patterns in differentiated products industries. Working Paper.

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.

Goldberg, P. K. and R. Hellerstein (2013). A structural approach to identifying the sources of local currency price stability. *Review of Economic Studies 80*, 175—210.

Grant, M. and S. Boyd (2008). Graph implementations for nonsmooth convex programs. In V. Blondel, S. Boyd, and H. Kimura (Eds.), *Recent Advances in Learning and Control*, Lecture Notes in Control and Information Sciences, pp. 95–110. Springer-Verlag Limited. `http://stanford.edu/~boyd/graph_dcp.html`.

Grant, M. and S. Boyd (2014, March). CVX: Matlab software for disciplined convex programming, version 2.1. `http://cvxr.com/cvx`.

Hausman, J. and W. Newey (2016). Individual heterogeneity and average welfare. *Econometrica 84*(3), 1225–1248.

Hausman, J. A. (1996). Valuation of new goods under perfect and imperfect competition. In T. F. Bresnahan and R. J. Gordon (Eds.), *The Economics of New Goods*, Chapter 5, pp. 209–248. Chicago: University of Chicago Press.

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.

Jaffe, S. and G. Weyl (2013). The first-order approach to merger analysis. *American Economic Journal: Microeconomics 5*, 188–218.

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.

Knittel, C. R. and K. Metaxoglou (2014). Estimation of random coefficient demand models: Two empiricists' perspectives. *The Review of Economic and Statistics 96*(1), 34–59.

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. (2000a). Mergers with differentiated products: The case of the ready-to-eat cereal industry. *The RAND Journal of Economics 31*, 395–421.

Nevo, A. (2000b). A practitioner's guide to estimation of random coefficients logit models of demand. *Journal of Economics & Management Strategy 9*(4), 513–548.

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

Newey, W. (1997). Convergence rates and asymptotic normality for series estimators. *Journal of Econometrics 79*, 147–168.

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

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.

R. H. Byrd, J. N. and R. Waltz (2006). Knitro: An integrated package for nonlinear optimization. In G. di Pillo and M. Roma (Eds.), *Large-Scale Nonlinear Optimization*, pp. 35–59. Springer. `http://stanford.edu/~boyd/graph_dcp.html`.

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

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

Souza-Rodrigues, E. A. (2014). Nonparametric estimation of a generalized regression model with group effects. working paper, University of Toronto, Department of Economics.

Tebaldi, P., A. Torgovitzky, and H. Yang (2019). Nonparametric estimates of demand in the california health insurance exchange. NBER Working Paper No. 25827.

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.

Appendix A provides some details on Bernstein polynomials. Appendix B presents the assumptions for Theorems 1, 2, 3, and the proofs for Theorems 2 and 3. Online Appendix C contains the proof for Theorem 1 and supplementary results for inference. Online Appendix D discusses additional economic constraints and shows how to enforce them in estimation. Online Appendix E presents the results of additional Monte Carlo simulations. Online Appendix F discusses violations of the exogeneity restriction maintained throughout the paper. Online Appendix G discusses the construction of the data and contains descriptive statistics. Finally, Online Appendix H provides two possible micro-foundations for the demand model estimated in the empirical application.

Appendix A: Bernstein Polynomials

A.1 Approximation Result

For a positive integer m, the Bernstein basis functions are 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 tensor-product 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 I assume that the degree m is the same for each dimension $n = 1, ..., N$. This is not needed, but I 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.[56]

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 $m \to \infty$.

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

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.

A.2 Imposing Exchangeability

I now provide more details on how to impose the exchangeability restriction discussed in Section 3.2 (see equation (6)). As in the previous section, for simplicity I consider an approximation via the tensor-product of univariate Bernstein polynomials all of the same degree m. The case with tensor-product of univariate polynomials of differing degrees makes the notation heavier, but is handled in the same way. First, as in the main text, I consider the case where $x^{(2)}$ is a vector of product-specific characteristics each with dimension $\tilde{n}_{x^{(2)}}$. With J goods, the overall degree of the approximation is then $(2J+\tilde{n}_{x^{(2)}})m$. Let $v^s\equiv(v_1^s,\cdots,v_J^s)$ be a $J-$vector of integers between 0 and m, and define $v^p\equiv(v_1^p,\cdots,v_J^p)$, and $v^x\equiv(v_1^x,\cdots,v_J^x)$ similarly. Next, let $\theta_j\left(v_1^s,\cdots,v_J^s,v_1^p,\cdots,v_J^p,v_1^x,\cdots,v_J^x;m\right)$ denote the coefficient on the term $\Pi_{k=1}^J b_{v_k^s,m}(s_k)\,b_{v_k^p,m}(p_k)\,b_{v_k^x,m}\left(x_k^{(2)}\right)$ in the Bernstein approximation for σ_j^{-1}. Note that, if the inverse demand system σ^{-1} is exchangeable, Lemma 2 implies that the Bernstein coefficients should satisfy an analogous property. To define this precisely, let $\pi:\{1,...,J\}\mapsto\{1,...,J\}$ be any permutation, π^{-1} be its inverse, and $\tilde{\pi}$ 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.

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

$\tilde{\pi}^{-1}$ is defined similarly for π^{-1}. Then, the following property holds for large m

$$\theta_j\left(v^s,v^p,v^x;m\right)=\theta_{\pi(j)}\left(\tilde{\pi}^{-1}(v^s),\tilde{\pi}^{-1}(v^p),\tilde{\pi}^{-1}(v^x);m\right) \tag{9}$$

for all $v_k^s,v_k^p,v_k^x\in\{0,1,\cdots,m\}$. This is a set of linear constraints on the Bernstein coefficients that can be easily be enforced. In fact, one could embed the constraint into the very definition of the vector of Bernstein coefficient, which would effectively reduce the dimension of the program to be solved in estimation (equation (4)).

Without exchangeability, the number of coefficients to estimate for each demand function is equal to $(m+1)^{J\left(2+\tilde{n}_{x^{(2)}}\right)}$. In contrast, when exchangeability is imposed, that number is $\left[\frac{(m+J-1)!}{(J-1)!(m)!}(m+1)\right]^{2+\tilde{n}_{x^{(2)}}}$. To see this, note that θ_j in (9) has $J\left(2+\tilde{n}_{x^{((2))}}\right)$ arguments, of which $2+\tilde{n}_{x^{(2)}}$ are "own" argument (i.e. j's share, price and $x^{(2)}$ attributes) and $(J-1)\left(2+\tilde{n}_{x^{(2)}}\right)$ are other goods' arguments. Exchangeability of σ_j^{-1} means that the function is invariant to rearranging the rival goods's arguments, for any given value of the own arguments. Now, the number of ways $(J-1)$ numbers can be drawn with replacement from a set of size $m+1$ is $\frac{(m+J-1)!}{(J-1)!(m)!}$.[57] Repeating this for $m+1$ possible values of each own argument and for $2+\tilde{n}_{x^{(2)}}$ arguments per good (share, price and $x^{(2)}$ attributes), one obtains the total number of coefficients under

[57]This is sometimes called a multicombination.

exchangeability.

Finally, I consider the case where $x^{(2)}$ is a vector of market-level variables that are not product-specific (e.g. income). The corresponding definition of exchangeability was given in footnote 30. In this case, the analog of equation (9) is

$$\theta_j\left(v^s, v^p, v^x; m\right) = \theta_{\pi(j)}\left(\tilde{\pi}^{-1}\left(v^s\right), \tilde{\pi}^{-1}\left(v^p\right), v^x; m\right) \tag{10}$$

for all $v_k^s, v_k^p, v_k^x \in \{0, 1, \cdots, m\}$. Again, this is a set of linear constraints on the Bernstein coefficients that can be easily be enforced.

Appendix B: Inference

This appendix contains all the notation and assumptions for the inference results in Section 3 of the paper, as well as the proofs for Theorem 2 and 3.

B.1 Setup and Notation

For simplicity, we focus on the case where there are no additional exogenous covariates $x^{(2)}$ in the demand system. Accordingly, we drop $x^{(2)}$ and use x to denote what was denoted by $x^{(1)}$ in the main text. As pointed out by CC (Section 3.3), allowing for $x^{(2)}$ is straightforward and does not change anything in the implementation of the estimator.

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 $Y \equiv (S, P)$ denote the arguments of the unknown functions and $\mathcal{Y}$ denote its support. For every $y \in \mathcal{Y}$, let $h_0\left(y\right) \equiv \left[h_{0,1}\left(y\right), ..., h_{0,J}\left(y\right)\right]' \equiv \left[\sigma_1^{-1}\left(y\right), ...\sigma_J^{-1}\left(y\right)\right]'$, so that the estimating equations become

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

We assume that, for $j \in \mathcal{J}$, $h_{0,j} \in \mathcal{H}$, where $\mathcal{H}$ is the Hölder ball of smoothness r, and we endow it with the norm $||\cdot||_\infty$ defined by $||h||_\infty \equiv \max_{j\in\mathcal{J}} ||h_j||_{1,\infty}$ for a function $h = [h_1, ..., h_J]$, where $||h_j||_{1,\infty}$ denotes the sup-norm for a scalar-valued function h_j. We also let $||v||$ denote the Euclidean norm of a vector v, $||\mathbf{M}||$ denote the norm of an m_1−by−m_2 matrix $\mathbf{M}$ defined as $||\mathbf{M}|| \equiv \sup\left\{||\mathbf{M}v|| : v \in \mathbb{R}^{m_2}, ||v|| = 1\right\}$, and $\left(\mathbf{M}\right)_l^- \equiv \left(\mathbf{M}'\mathbf{M}\right)^{-1}\mathbf{M}'$ be the left-inverse of a matrix $\mathbf{M}$.
Further, we let $\left\{\psi_{1,M_i}^{(i)}, ..., \psi_{M_i,M_i}^{(i)}\right\}$ be the collection of basis functions used to approximate $h_{0,i}$ for $i \in \mathcal{J}$, and let $M = \sum_{j=1}^J M_j$ be the dimension of the overall sieve space for h. Similarly, we let $\left\{a_{1,K_i}^{(i)}, ..., a_{K_i,K_i}^{(i)}\right\}$ 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, letting $\text{diag}\left(mat_1, ..., mat_J\right) \equiv \begin{bmatrix} mat_1 & 0_{d_{1,r}\times d_{2,c}} & \cdots & 0_{d_{1,r}\times d_{J,c}} \\ 0_{d_{2,r}\times d_{1,c}} & mat_2 & \cdots & 0_{d_{2,r}\times d_{J,c}} \\ \vdots & \vdots & \ddots & \vdots \\ 0_{d_{J,r}\times d_{1,c}} & 0_{d_{J,r}\times d_{2,c}} & \cdots & mat_J \end{bmatrix}$ for matrices $mat_j \in \mathbb{R}^{d_{j,r}\times d_{j,c}}$

with $j \in \mathcal{J}$, we define, for $i \in \mathcal{J}$

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

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

Also, we let, for $j, k \in \mathcal{J}$,

$$\begin{aligned}
\Omega_{jk} = \Omega_{kj}' &= \mathbb{E}\left(\xi_{jt}\xi_{kt} a_{K_j}^{(j)}(W_t)\, a_{K_k}^{(k)}(W_t)'\right) && K_j - \text{by} - K_k \\
\Omega &= \begin{bmatrix} \Omega_{11} & \Omega_{12} & \cdots & \Omega_{1J} \\ \Omega_{21} & \Omega_{22} & \cdots & \Omega_{2J} \\ \vdots & \vdots & \ddots & \cdots \\ \Omega_{J1} & \Omega_{J2} & \cdots & \Omega_{JJ} \end{bmatrix} && K - \text{by} - K
\end{aligned}$$

and, similarly,

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

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

where $\hat{\xi}_{jt} = x_{jt} - \hat{h}_j(y_t)$.

For $i \in \mathcal{J}$, 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 $\zeta \equiv \max_{j\in\mathcal{J}}\zeta_j$. The rate at which ζ diverges to infinity with the sample size will play a role in the proofs. When splines (including Bernstein polynomials) are used for a_K and ψ_M, we have $\zeta = O(\sqrt{M})$ (see, e.g., Newey (1997).)

As in CC, we use the following sieve measure of ill-posedness, for $i \in \mathcal{J}$,

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

where $\tilde{\Psi}_{M_i}$ is the closed linear span of $\left\{\psi_{M_i}^{(i)}\right\}$ and we let $\tau_M \equiv \max_{j\in\mathcal{J}}\tau_{M_j}^{(j)}$. The rate at which τ_M diverges to infinity may be viewed as a measure of how difficult the estimation problem is. In order to formalize this, we will need appropriate notation. Specifically, letting a_T and b_T be sequences of positive numbers, the notation $a_T \lesssim b_T$ means $\limsup_{T\to\infty} a_T/b_T < \infty$, and the notation $a_T \asymp b_T$ means $a_T \lesssim b_T$ and $b_T \lesssim a_T$. Next, 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

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

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 $H_{0,j} \equiv (h_{0,j}(y_1), ..., h_{0,j}(y_T))'$ 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

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

For any functional $f : \mathcal{H} \mapsto \mathbb{R}$ and any $(h, v) \in \mathcal{H} \times \mathcal{H}$, we let $Df(h)[v] \equiv \left.\frac{\partial f(h+\tau v)}{\partial \tau}\right|_{\tau=0}$ denote the pathwise derivative of f at h in the direction v (if it exists). Next, letting $vec_{g,J,j}$ be the column $J-$vector valued function that returns all zeros except for the $j-$th element, where it returns the function g, we define

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

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

Finally, we let

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

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

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

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

B.2 Assumptions for Theorem 1

This section collects the assumptions for Theorem 1. The proof can be found in online Appendix C.

Assumption 2. *The variables (X_t, Z_t, P_t, ξ_t) are independent and identically distributed across markets.*

Assumption 3. *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) $\sup_{w\in\mathcal{W}} \mathbb{E}\left[\xi_j^2\mathbb{I}\left\{\sum_{i=1}^J |\xi_i| > \ell(T)\right\}|w\right] = o(1)$ *for any positive sequence $\ell(T) \nearrow \infty$;*
(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 4. *(i)* $\tau_M \zeta \sqrt{M(\log M)/T} = o(1)$;
(ii) $\zeta^{\left(2+\gamma^{(1)}\right)/\gamma^{(1)}} \sqrt{(\log K)/T} = o(1)$ *and* $\zeta^{\left(1+\gamma^{(2)}\right)/\gamma^{(2)}} \sqrt{(\log K)/T} = o(1)$, *where* $\gamma^{(1)}, \gamma^{(2)} > 0$ *are defined in Assumption 3(v)-3(vi);*
(iii) $T \geq K \geq M$, $K \asymp M$, *and* $\zeta = O\left(\sqrt{M}\right)$.

Assumption 5. *The basis used for the instrument space 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 6. $||\hat{h} - h_0||_\infty = o_p(1)$.

Assumption 7. *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 that the sieve variance* $v_T(f)$ *for the functional* f *is strictly positive 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) $\frac{\sqrt{T}}{v_T(f)}\left(||\partial^{\alpha_1}\hat{h} - \partial^{\alpha_1} h_0||_\infty ||\partial^{\alpha_2}\hat{h} - \partial^{\alpha_2} h_0||_\infty + ||\partial^{\alpha}\tilde{h} - \partial^{\alpha} h_0||_\infty\right) = O_p(\eta_T)$ *for a nonnegative sequence* η_T *such that* $\eta_T = o(1)$;

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

Discussion of assumptions. Assumption 3 corresponds to Assumption 2 in CC, modified to account for the fact that my model has J equations and J error terms. Assumption $4(i)$ corresponds to the condition imposed by CC in Theorem D.1, whereas $4(ii)$ is similar to Assumption $3(iii)$ in CC. Assumption $4(iii)$ restricts the growth rates of the sieve spaces for the endogenous variables and the instruments. The requirement that $T \geq K \geq M$ simply says that one needs more instruments than endogenous variables and that both should be smaller than the sample size, analogously to parametric models. On the other hand, the requirement that $\zeta = O\left(\sqrt{M}\right)$ has more bite. It holds, for instance, when splines are used to approximate the unknown functions (see, e.g., Newey (1997)). I impose it since in practice I advocate using Bernstein polynomials, which are a special case of splines. Assumption 5 is not necessary but I impose it for simplicity. Assumption 6 requires $\hat{h}$ to be a consistent estimator. CC provide sufficient conditions for it and characterize the rate of convergence. Assumption 7 corresponds to the sufficient conditions in Remark 4.1 of CC.

B.3 Theorem 2: 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 $(\overline{s}, \overline{p}) \equiv (\overline{s}_1, \overline{s}_2, \overline{p}_1, \overline{p}_2)$ and Bernstein polynomials are used for both the endogenous variables and the instruments. The goal is to provide sufficient, lower-level conditions for Theorem 1. Analogous arguments hold for the own-price elasticity of good 2 and for the cross-price elasticities.
The functional of interest takes the form

$$f_\epsilon(h_0) = -\frac{p_1}{s_1} \frac{\frac{\partial h_{0,2}(\overline{s},\overline{p})}{\partial s_2}\frac{\partial h_{0,1}(\overline{s},\overline{p})}{\partial p_1} - \frac{\partial h_{0,1}(\overline{s},\overline{p})}{\partial s_2}\frac{\partial h_{0,2}(\overline{s},\overline{p})}{\partial p_1}}{\frac{\partial h_{0,1}(\overline{s},\overline{p})}{\partial s_1}\frac{\partial h_{0,2}(\overline{s},\overline{p})}{\partial s_2} - \frac{\partial h_{0,1}(\overline{s},\overline{p})}{\partial s_2}\frac{\partial h_{0,2}(\overline{s},\overline{p})}{\partial s_1}} \equiv -\frac{p_1}{s_1}\frac{N_1 - N_2}{D_1 - D_2} \tag{14}$$

Theorem 2 maintains the following assumption.

Assumption 8. *(i) P has bounded support and (P,S) have densities bounded away from 0 and ∞;*
(ii) The basis used for both the endogenous variables and the instruments is tensor-product Bernstein polynomials. Further, the univariate Bernstein polynomials for the endogenous variables all have the same degree $M^{1/4}$;
(iii) The unknown functions $h_0 = [h_{0,1}, h_{0,2}]'$ belong to the Hölder ball of smoothness $r \geq 8$ and finite radius;
(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 3(v)-3(vi);
(v) $\frac{\sqrt{T}}{v_T(f_\epsilon)} \times \left(M^{\frac{3-r}{4}} + \frac{\tau_M M^{\frac{9-r}{4}}}{\sqrt{T}} + \tau_M^2 M^3 \frac{\log M}{T}\right) = o(1)$.

Discussion of Assumption 8. Assumptions 8(i), 8(iii) and 8(iv) are regularity conditions needed to apply the sup-norm rate results in CC.[58] 8(ii) is assumed for simplicity but it is not necessary. 8(v) corresponds to the second part of Assumption CS(v) in CC and is used to verify Assumption 7. More concrete sufficient conditions for Assumptions 8(iv) and 8(v) may be provided in specific settings. For example, Lemma 3 below gives sufficient conditions for the mildly ill-posed case.[59]

We now provide a proof of Theorem 2.

Proof of Theorem 2. We prove the statement by showing that the assumptions of Theorem 1 hold. Assumptions 2, 3, 4(iii), 5 and 6 are maintained. Assumption 4(i) is implied by Assumptions 4(iii) and 8(v), and Lemma 8. Similarly, Assumption 4(ii) is implied by Assumptions 4(iii) and 8(iv).
We now verify Assumption 7. 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.

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

$$Df_\epsilon(h_0)[v] \equiv \left.\frac{\partial f_\epsilon(h_0 + \tau v)}{\partial \tau}\right|_{\tau=0} = \frac{p_1}{s_1}\left(C_1\frac{\partial v_2}{\partial s_2} + C_2\frac{\partial v_1}{\partial s_2} + C_3\frac{\partial v_1}{\partial p_1} + C_4\frac{\partial v_2}{\partial p_1} + C_5\frac{\partial v_2}{\partial s_1} + C_6\frac{\partial v_1}{\partial s_1}\right) \tag{15}$$

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.

[58] CC establish sup-norm rate results for the case where the unknown function is approximated using B-splines, among others. Since Bernstein polynomials are a special case of splines (see, e.g., Schumaker (2007)), their results apply to the setting considered here.

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

Next, note that, for any $h = [h_1, h_2] \in \mathcal{H}_T$,

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

where the first inequality follows from the triangle inequality and the fundamental theorem of calculus, and the second inequality follows from assumption 8(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 (15) and write

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

which shows that Assumption 7(i) holds with $\alpha = [1, 1, 1, 0]$.

7(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

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

where $(F_i)_{i=1}^6$ are linear combinations 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 $|\alpha_1| = |\alpha_2| = 1$.

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

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

Thus, Assumption 7(*iii*) is implied by Assumption 8(*v*).

7(*iv*) By Remark 4.1 in CC, a sufficient condition for Assumption 7(*iv*) is

$$T_{iv,\epsilon} \equiv \frac{\tau_M \sqrt{\sum_{m=1}^M \left(Df_\epsilon\left(\hat{h}\right)\left[\left(G_\psi^{-1/2}\psi_M\right)_m\right] - Df_\epsilon(h_0)\left[\left(G_\psi^{-1/2}\psi_M\right)_m\right]\right)^2}}{v_T(f_\epsilon)} = o_p(1) \tag{16}$$

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, using Corollary 3.1 in CC and the well-known rate results for splines and their derivatives in, e.g., Newey (1997),

$$T_{iv,\epsilon} \quad = \quad O_p\left(\frac{\sqrt{T}}{v_T(f_\epsilon)} \times \left[\frac{\tau_M M^{(9-r)/4}}{\sqrt{T}} + \frac{\tau_M^2 M^{11/4}\sqrt{\log M}}{T}\right]\right) \tag{17}$$

The conclusion in (16) then follows from Assumption 8(*v*).

□

The next lemma provides more primitive sufficient conditions for Assumptions 8(*iv*) and 8(*v*).

Lemma 3. *Let Assumptions 8(i) and 8(iii) hold. Further, let* $(v_T(f_\epsilon))^2 \asymp M^{a+\varsigma+1}$ *and* $\tau_M \asymp M^{\varsigma/2}$ *for* $a \leq 0, \varsigma \geq 0, a + \varsigma + 1 > 0$, $r + 2a - 4 > 0$.[60] *Then, Assumptions 8(iv) and 8(v) are satisfied if* $M \asymp T^\rho$ *with*

$$\rho \in \left(\frac{2}{r - 3 + 2(a + \varsigma + 1)}, \min\left\{\frac{1}{\varsigma - a + 5}, \frac{\gamma^{(1)}}{2 + \gamma^{(1)}}, \frac{\gamma^{(2)}}{1 + \gamma^{(2)}}\right\}\right)$$

Further, M *may be chosen to satisfy the latter condition if* $r + 4a - 11 > 0$ *and* $\gamma^{(i)}(r + 2a + 2\varsigma - 3) - 4 > 0$ *for* $i \in \{1, 2\}$.

Proof. The result follows by inspection.

□

[60] 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 $v_T(f)$.

B.4 Theorem 3: Equilibrium price functionals

We now specialize Theorem 1 to the case where the functional f is the equilibrium price of good 1 in a market with two goods 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. Again, 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

$$\begin{bmatrix} g_1\left(\overline{\delta}, \overline{p}, \overline{mc}, h_0\right) \\ g_2\left(\overline{\delta}, \overline{p}, \overline{mc}, h_0\right) \end{bmatrix} \equiv -\left[\left(\mathbb{J}_{h_0}^s\right)^{-1} \mathbb{J}_{h_0}^p\right]' \begin{bmatrix} \overline{p}_1 - \overline{mc}_1 \\ \overline{p}_2 - \overline{mc}_2 \end{bmatrix} + \begin{bmatrix} h_{0,1}^{-1}\left(\overline{\delta}, \overline{p}\right) \\ h_{0,2}^{-1}\left(\overline{\delta}, \overline{p}\right) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{18}$$

where

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

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 endogenous variables and the instruments 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 $r \geq 9$ and finite radius;
(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 3(v)-3(vi);
(v) $\frac{\sqrt{T}}{v_T(f_{p_1})} \times \left(M^{\frac{4-r}{4}} + \frac{\tau_M M^{\frac{10-r}{4}}}{\sqrt{T}} + \tau_M^2 M^3 \frac{\log M}{T}\right) = o(1)$.

Discussion of Assumptions. Assumptions 9(i), 9(iii) and 9(iv) are regularity conditions needed 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 7. More concrete sufficient conditions for Assumptions 9(iv) and 9(v) may be provided in specific settings. For example, Lemma 4 below gives sufficient conditions for the mildly ill-posed case.

We now provide a proof of Theorem 3.

Proof of Theorem 3. We prove the statement by showing that the assumptions of Theorem 1 hold. Assumptions 2, 3, 4(iii), 5 and 6 are maintained. Assumption 4(i) is implied by Assumptions 4(iii) and 9(v), and Lemma 8. Similarly, Assumption 4(ii) is implied by Assumptions 4(iii) and 9(iv).
We now verify Assumption 7.

7(i) Applying the implicit function theorem to (18),

$$Df_p(h)[v] = -\begin{bmatrix} \frac{\partial g_1\left(\overline{\delta},\overline{p},\overline{mc},h+\tau v\right)}{\partial p_1} & \frac{\partial g_1\left(\overline{\delta},\overline{p},\overline{mc},h+\tau v\right)}{\partial p_2} \\ \frac{\partial g_2\left(\overline{\delta},\overline{p},\overline{mc},h+\tau v\right)}{\partial p_1} & \frac{\partial g_2\left(\overline{\delta},\overline{p},\overline{mc},h+\tau v\right)}{\partial p_2} \end{bmatrix}^{-1} \begin{bmatrix} \frac{\partial g_1\left(\overline{\delta},\overline{p},\overline{mc},h+\tau v\right)}{\partial \tau} \\ \frac{\partial g_2\left(\overline{\delta},\overline{p},\overline{mc},h+\tau v\right)}{\partial \tau} \end{bmatrix}\Bigg|_{\tau=0} \equiv -\left(\mathbb{J}_g^p\right)^{-1} \mathbb{J}_g^\tau\Big|_{\tau=0} \tag{19}$$

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

$$D f_{p_1}(h)[v] = \sum_{\tilde{\alpha}:|\tilde{\alpha}|\leq 1}\sum_{j=1}^{2} C_{\tilde{\alpha},j}\left(\overline{\delta},\overline{mc},\left\{\partial^{\beta}h : |\beta|\leq 2\right\}\right)\times\partial^{\tilde{\alpha}}v_j\left(h^{-1}\left(\overline{\delta},\overline{p}\right),\overline{p}\right) \tag{20}$$

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 the proof of Theorem 2, 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 7(i) holds with $\alpha=[1,1,1,1]$.

7(ii) As in the proof of Theorem 2, by the mean value theorem, we obtain

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

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 7$(ii)$ holds with $|\alpha_1|=1, |\alpha_2|=2$.

7(iii) Given the choice of α,α_1,α_2 above and by Corollary 3.1 in CC, we have

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

Thus, Assumption 7(iii) is implied by Assumption 9(v).

7(iv) By Remark 4.1 in CC, a sufficient condition for Assumption 7(iv) is

$$T_{iv,p}\equiv\frac{\tau_M\sqrt{\sum_{m=1}^{M}\left(Df_{p_1}\left(\hat{h}\right)\left[\left(G_{\psi}^{-1/2}\psi_M\right)_m\right]-Df_{p_1}(h_0)\left[\left(G_{\psi}^{-1/2}\psi_M\right)_m\right]\right)^2}}{v_T\left(f_{p_1}\right)}=o_p(1) \tag{21}$$

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, using Corollary 3.1 in CC and the well-known rate results for splines and their derivatives in, e.g., Newey (1997), we can write

$$T_{iv,p} = O_p\left(\frac{\sqrt{T}}{v_T\left(f_{p_1}\right)}\times\left[\frac{\tau_M M^{(10-r)/4}}{\sqrt{T}}+\frac{\tau_M^2 M^3\sqrt{\log M}}{T}\right]\right)$$

The conclusion in (21) then follows from Assumption $9(v)$.

□

Finally, the following lemma provides more primitive sufficient conditions for Assumptions $9(iv)$ and $9(v)$.

Lemma 4. *Let Assumptions 9(i) and 9(iii) hold. Further, let* $(\upsilon_T(f_\epsilon))^2 \asymp M^{a+\varsigma+1}$ *and* $\tau_M \asymp M^{\varsigma/2}$ *for* $a \leq 0, \varsigma \geq 0, a+\varsigma+1 > 0$, $r + 2a - 5 > 0$.[61] *Then, Assumptions 9(iv) and 9(v) are satisfied if* $M \asymp T^\rho$ *with*

$$\rho \in \left(\frac{2}{r-4+2(a+\varsigma+1)}, \min\left\{\frac{1}{\varsigma-a+5}, \frac{\gamma^{(1)}}{2+\gamma^{(1)}}, \frac{\gamma^{(2)}}{1+\gamma^{(2)}}\right\}\right)$$

Further, M *may be chosen to satisfy the latter condition if* $r + 4a - 12 > 0$ *and* $\gamma^{(i)}(r+2a+2\varsigma-4) - 4 > 0$ *for* $i \in \{1,2\}$.

Proof. The result follows by inspection.

□

[61] Again, this corresponds to the "mildly ill-posed" case discussed by CC in Corollary 5.1.

Appendices for Online Publication

Appendix C: Supplementary Results for Inference

Proof of Theorem 1.

We prove that

$$\sqrt{T}\frac{\left(f\left(\hat{h}\right)-f\left(h_0\right)\right)}{v_T\left(f\right)}\xrightarrow{d}N\left(0,1\right) \tag{22}$$

The result then follows from Lemma 5 below. By Assumption 7(ii),

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

By Assumption 7(iii), $c_T=o_p\left(1\right)$ and therefore,

$$\sqrt{T}\frac{\left(f\left(\hat{h}\right)-f\left(h_0\right)\right)}{v_T\left(f\right)} = \sqrt{T}\frac{Df\left(h_0\right)[\hat{h}-h_0]}{v_T\left(f\right)}+o_p\left(1\right) \tag{23}$$

Further, by Assumption 7(i)

$$Df\left(h_0\right)[\hat{h}-h_0]=Df\left(h_0\right)[\hat{h}-\tilde{h}]+Df\left(h_0\right)[\tilde{h}-h_0] \tag{24}$$

and

$$Df\left(h_0\right)[\tilde{h}-h_0]\lesssim||\partial^{\alpha}\tilde{h}-\partial^{\alpha}h_0||_\infty \tag{25}$$

By (25) and Assumption 7(iii),

$$\sqrt{T}\frac{Df\left(h_0\right)[\tilde{h}-h_0]}{v_T\left(f\right)}=o_p\left(1\right) \tag{26}$$

Combining (23), (24) and (26), we obtain

$$\sqrt{T}\frac{\left(f\left(\hat{h}\right)-f\left(h_0\right)\right)}{v_T\left(f\right)} = \sqrt{T}\frac{Df\left(h_0\right)[\hat{h}-\tilde{h}]}{v_T\left(f\right)}+o_p\left(1\right) \tag{27}$$

We define

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

and note that $\mathbb{E}\left[\left(R_T(W)\cdot[\xi_1,\cdots,\xi_J]'\right)^2\right]=1$. Then,

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

where $\xi\equiv[\xi_{11},\cdots,\xi_{1T},\xi_{21},\cdots,\xi_{2T},\cdots,\xi_{J1},\cdots,\xi_{JT}]'$.
First, we show that $T_1\xrightarrow{d}N(0,1)$ by the Lindeberg-Feller theorem. The Lindeberg condition requires that, for every $\epsilon>0$,

$$C_{0,T}\equiv\mathbb{E}\left[\left(R_T(W)\cdot[\xi_1,\cdots,\xi_J]'\right)^2\underbrace{\mathbb{I}\left\{\left|R_T(W)\cdot[\xi_1\cdots\xi_J]'\right|>\epsilon\sqrt{T}\right\}}_{Q_T(W,\xi)}\right]=o(1) \tag{28}$$

To show that this condition holds, note that

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

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

$$\begin{aligned}\left|R_T^{(i)}(w_t)\right| &\leq \frac{\left|\left|Df(h_0)\left[\psi_{M_i}^{(i)}\right]'\left(L_i'G_{A,i}^{-1}L_i\right)^{-1}L_i'G_{A,i}^{-1/2}\right|\right|}{v_T(f)}\times\sup_{w\in\mathcal{W}}\left|\left|G_{A,i}^{-1/2}a_{K_i}^{(i)}(w)\right|\right| \\ &\equiv \lambda_i(T)\times\zeta_{A,i}\end{aligned} \tag{29}$$

by the Cauchy-Schwarz inequality and thus

$$\left|\sum_{j=1}^{J}R_T^{(j)}(w_t)\,\xi_{jt}\right|\leq\sum_{j=1}^{J}|\xi_{jt}|\;\times\max_i\left[\lambda_i(T)\times\zeta_{A,i}\right] \tag{30}$$

Equation (30) implies that, for all $w\in\mathcal{W}$ and all $\xi\in\Xi$,

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

where $Q_T(w,\xi)$ was defined in (28). Therefore, using Cauchy-Schwarz and the law of iterated expectations,

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

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

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

where the inequality follows from Lemma 6 below. Further, $\sup_{w\in\mathcal{W}}\mathbb{E}\left[\xi_i^2\overline{Q}_T(\xi)\,|w\right] = o(1)$ by Assumption 3(iv) and the fact that, by Assumption 4(i) and Lemma 6, $\frac{\sqrt{T}}{\max_i[\lambda_i(T)\zeta_{A,i}]} \nearrow \infty$. 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

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

The first inequality follows from some algebra and the Cauchy-Schwarz inequality, the first equality is by the definition in (29), the second inequality holds by the definition of matrix norm, and the second equality is by Lemmas A.1, F.8 and F.10(c) in CC, Lemma 6 below and Assumption 4(iii). Therefore, by Assumption 4(i), we obtain $|T_2| = o_p(1)$. This completes the proof of (22).

□

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 endogenous variables and those used for the instruments form a Riesz basis for the conditional expectation operator. I follow CC in assuming that this is the case.*

Lemma 5. *Let* $||\hat{h}-h_0||_\infty = o_p(1)$ *and let Assumptions 3(i), 3(ii), 3(iii), 3(v), 3(vi), 4, 5, 7(iv) hold. Then*

$$\left|\frac{\hat{v}_T(f)}{v_T(f)} - 1\right| = o_p(1). \tag{31}$$

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

$$\frac{\hat{v}_T^2(f)}{v_T^2(f)} - 1 = \frac{(\hat{\gamma}_T - \gamma_T)' \Omega^o (\hat{\gamma}_T + \gamma_T)}{v_T^2(f)} + \frac{\hat{\gamma}_T' \left(\hat{\Omega}^o - \Omega^o\right) \hat{\gamma}_T}{v_T^2(f)} \equiv T_1 + T_2 \tag{32}$$

where

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

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

and note that $\frac{||\gamma_T||^2}{v_T^2(f)} = \sum_{j=1}^J \lambda_j(T)^2$ by the definition in (29).
We consider T_1 and T_2 in equation (32) in turn. Note that

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

Now,

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

where the last step follows from Assumption 7(iv) and Lemma 6. Further, $T_1^{(2)} = o_p(1)$ by Lemmas F.10(c) and A.1 in CC and Assumption 4(i),and $T_1^{(3)} = o_p(1)$ by Assumption 7(iv). This implies that

$$\frac{||\hat{\gamma}_T - \gamma_T||}{v_T(f)} = o_p(1). \tag{33}$$

Therefore, by Cauchy-Schwarz,

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

where in the last step we also use Lemma 6 and the fact that $||\Omega^o|| < \infty$ by Assumptions 3(i), 3(ii), 3(iii).
Turning to $|T_2|$, note that

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

where the last step follows again from Lemma 6 and (33). We complete the proof by showing that $||\hat{\Omega}^o - \Omega^o|| =$

$o_p(1)$. Note that

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

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

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

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

$$\left|\left| \left(\hat{\Omega}^o - \Omega^o\right) v \right|\right| \;=\; \sum_{j=1}^{J} v_j' \left(\hat{\Omega}^o_{jj} - \Omega^o_{jj}\right) v_j + 2 \sum_{j=1}^{J} \sum_{k=1}^{j-1} v_j' \left(\hat{\Omega}^o_{jk} - \Omega^o_{jk}\right) v_k$$

and thus, by definition of matrix norm and Cauchy-Schwarz,

$$||\hat{\Omega}^o - \Omega^o|| \;\leq\; J \max_{j \in \mathcal{J}} ||\hat{\Omega}^o_{jj} - \Omega^o_{jj}|| + 2J^2 \max_{j,k \in \mathcal{J}, j \neq k} ||\hat{\Omega}^o_{jk} - \Omega^o_{jk}|| \equiv \tilde{T}_1 + \tilde{T}_2$$

Now, $\tilde{T}_1 = o_p(1)$ by Lemma G.3 in CC.For $\tilde{T}_2$, note that, by the triangle inequality, for all $j,k \in \mathcal{J}, j \neq k$,

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

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 5. Using Lemma 7 below, we obtain $||T_{\Omega,1}|| = o_p(1)$. Further, $||T_{\Omega,2}|| = o_p(1)$ by $\left(\hat{\xi}_{jt} - \xi_{jt}\right) \xi_{kt} \leq ||\hat{h}_j - h_{0,j}||_{1,\infty} \left(1 + \xi_{kt}^2\right)$ and Lemma F.7 in CC. Similarly, $||T_{\Omega,4}|| = o_p(1)$. Finally, $||T_{\Omega,3}|| = o_p(1)$ by $\left(\hat{\xi}_{jt} - \xi_{jt}\right)\left(\hat{\xi}_{kt} - \xi_{kt}\right) \leq ||\hat{h} - h_0||_\infty^2$ and Lemma F.7 in CC.

□

Lemma 6. *For $i \in \mathcal{J}$, let* $\lambda_i(T) \equiv \frac{\left|\left| Df(h_0)\left[\psi^{(i)}_{M_i}\right]' \left(L_i' G_{A,i}^{-1} L_i\right)^{-1} L_i' G_{A,i}^{-1/2} \right|\right|}{v_T(f)}$ *and let Assumption 3(ii) hold. Then,* $\limsup_{T \to \infty} \lambda_i(T) < \infty$.

Proof. Note that

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

Further, by Assumption 3(ii)

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

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

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

Since we focus on the case in which the functional f is slower than $\sqrt{T}-$estimable, the denominator in the display above goes to infinity. Since the numerator is at most of the same order as the denominator, the result follows.

□

Lemma 7. *Let Assumptions 3(iii), 3(vi), 4(ii) and 5 hold. Then* $||T_{\Omega,1}|| = O_p(1)$*, where* $T_{\Omega,1}$ *is defined in the proof of Lemma 5.*

Proof. The proof adapts that of Lemma 3.1 in Chen and Christensen (2015). Let $C_T \asymp \zeta^{\left(1+\gamma^{(2)}\right)/\gamma^{(2)}}$ be a sequence of positive numbers with $\gamma^{(2)}$ defined in Assumption 3(vi), and let

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

where

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

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. By definition, $||\Xi_{1,t}|| \leq C_T^2$ and thus, by the triangle inequality and Jensen's inequality ($||\cdot||$ is convex), we have $||\Xi_{1,t} - \mathbb{E}(\Xi_{1,t})|| \leq 2C_T^2$. Further, dropping the t subscripts,

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

where the inequalities are in the sense of positive semi-definite matrices. Then, Corollary 4.1 in Chen and Christensen (2015) yields $||T_{\Omega,1}^{(1)}|| = O_p\left(C_T\sqrt{(\log K)/T}\right)$ and thus $||T_{\Omega,1}^{(1)}|| = o_p(1)$ by Assumption 4(ii). Turning to $||T_{\Omega,1}^{(2)}||$, since $||\Xi_{2,t}|| \leq \zeta^2|\xi_{jt}\xi_{kt}|\mathbb{I}\left\{|\xi_{jt}\xi_{kt}| \geq C_T^2/\zeta^2\right\}$, 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 3(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 8. *Let Assumptions 3 and 8(i)-8(iii) hold. Then, for* $f \in \{f_\epsilon, f_{p_1}\}$,

$$[v_T(f)]^2 \lesssim \tau_M^2 M^4$$

Proof. We prove this for $f = f_\epsilon$. The proof for $f = f_{p_1}$ is identical. As shown in CC,[62] the maintained assumptions imply

$$[v_T(f_\epsilon)]^2 \asymp \tau_M^2 \sum_{m=1}^{M}\left(Df_\epsilon(h_0)\left[\left(G_\psi^{-1/2}\psi_M\right)_m\right]\right)^2 \tag{34}$$

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,

$$\left|Df_\epsilon(h_0)\left[\left(G_\psi^{-1/2}\psi_M\right)_m\right]\right| \lesssim \max_{\tilde{\alpha}:|\tilde{\alpha}|=1}||\partial^{\tilde{\alpha}}\left(G_\psi^{-1/2}\psi_M\right)_m||_\infty \asymp M^{3/2}$$

where the first step follows from (15) and the second step follows from well-known properties of splines (see, e.g., Newey (1997)). Combining this and (34) completes the proof. □

Appendix D: Additional Constraints

In this appendix, I consider several constraints that one might be willing to impose besides those discussed in Section 3.2, and I show how to enforce them in estimation in a computationally tractable way. Because these

[62] See pp.22-23. See also Chen and Pouzo (2015).

constraints are defined conditional on any given value of $x^{(2)}$, I drop this for notational convenience.

D.1 Symmetry of the Jacobians

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. Under the assumption 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 $s=\sigma(\delta, p)$,

$$\mathbb{J}_{\sigma^{-1}}^{s}(s, p)=\left[\mathbb{J}_{\sigma}^{\delta}(\delta, p)\right]^{-1} \tag{35}$$

Because the inverse of a symmetric matrix is symmetric, symmetry of $\mathbb{J}_{\sigma}^{\delta}(\delta, p)$ implies symmetry of $\mathbb{J}_{\sigma^{-1}}^{s}(s, p)$. This, in turn, imposes *linear* constraints on the Bernstein coefficients as the degree of the approximation goes to infinity.[63]

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

$$\mathbb{J}_{\sigma}^{p}(\delta, p)=-\left[\mathbb{J}_{\sigma^{-1}}^{s}(s, p)\right]^{-1}\mathbb{J}_{\sigma^{-1}}^{p}(s, p) \tag{36}$$

which shows that $\mathbb{J}_{\sigma}^{p}$ is a nonlinear function of the derivatives of σ^{-1} and therefore of the Bernstein coefficients. In the implementation, it might be convenient to rewrite (36) as

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

Expressing $\mathbb{J}_{\sigma^{-1}}^{s}$ and $\mathbb{J}_{\sigma^{-1}}^{p}$ as linear combinations of the Bernstein polynomials and introducing extra param-

[63]More precisely, by the approximation result discussed in Appendix A, the difference between two (appropriately chosen) Bernstein coefficients approximates the change in the function σ_j^{-1} given by a change in s_k over a grid of equidistant points. Therefore, in the limit, equality of derivatives may be expressed as equality of differences between Bernstein coefficients.

eters (call them γ) for the entries of $\mathbb{J}_{\sigma}^{p}$, one then obtains a set of constraints that are linear in the Bernstein coefficients, given γ, and linear in γ, given the Bernstein coefficients.[64]

D.2 Additional Properties of the Jacobian of Demand

The matrix $\mathbb{J}_{\sigma}^{\delta}(\delta, p)$ has a number of additional features that one might want to impose in estimation. First, one of the sufficient conditions for identification in BH is a substitutability condition (their Assumption 2) that requires the off-diagonal elements of $\mathbb{J}_{\sigma}^{\delta}(\delta, p)$ to be non-positive. Further, it follows from Remark 2 of Berry et al. (2013) that the diagonal elements must be positive.[65]

Moreover, $\mathbb{J}_{\sigma}^{\delta}(\delta, p)$ belongs to the class of M-matrices, which are the object of a vast literature in linear algebra.[66] One of the most common definitions of this class is as follows.

Definition 1. A square real matrix A is called an M-matrix if (i) it is 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.

I now formalize the aforementioned result, which is a corollary of Theorem 2 in Berry et al. (2013).

Lemma 9. *Let Assumption 2 in BH hold. Then $\mathbb{J}_{\sigma}^{\delta}(\delta, p)$ is an M-matrix for all (δ, p).*

Proof. See Section D.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 I estimate σ^{-1} rather than σ itself. The Jacobian of the function to 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.[67] Thus, we may borrow results from that literature to impose necessary conditions on the Bernstein coefficients for σ^{-1}.

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.

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

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

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

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

Second, under the substitutability condition required for identification in BH, $\mathbb{J}_\sigma^\delta$ 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.

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

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

for $i = 1, ..., m$.

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

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

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

Column diagonal dominance and column entry diagonal dominance are defined analogously. By Theorem 3.2 of McDonald et al. (1995), if an M-matrix A is weakly diagonally dominant matrix of its columns, then $(A)^{-1}$ is weakly diagonally dominant of its row entries. This immediately implies the following result.

Lemma 10. *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 10 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.

D.3 Proofs for Results on Constraints

Proof of Lemma 1. Let $\pi : \{1, ..., J\} \to \{1, ..., J\}$ be any permutation with inverse π^{-1}. Further, 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.

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

and define $\tilde{\pi}^{-1}$ similarly for π^{-1}. Then, we can rewrite the definition of exchangeability for a generic $J-$valued function $g(y_1, y_2, y_3)$ of $3J$ arguments as

$$\tilde{\pi}^{-1}(g(y_1, y_2, y_3)) = g\left(\tilde{\pi}^{-1}(y_1), \tilde{\pi}^{-1}(y_2), \tilde{\pi}^{-1}(y_3)\right).$$

Now take any $\left(\delta, p, x^{(2)}\right)$ and let $s = \sigma\left(\delta, p, x^{(2)}\right)$. We can invert the demand system to obtain

$$\delta = \sigma^{-1}\left(s, p, x^{(2)}\right) \tag{37}$$

By exchangeability of σ,

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

Inverting this demand system, we obtain

$$\tilde{\pi}^{-1}(\delta) = \sigma^{-1}\left(\tilde{\pi}^{-1}(s), \tilde{\pi}^{-1}(p), \tilde{\pi}^{-1}\left(x^{(2)}\right)\right) \tag{38}$$

Combining (37) and (38),

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

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

□

Proof of Lemma 9. Under the maintained assumptions, Theorem 2 in Berry et al. (2013) implies that $\mathbb{J}_{\sigma}^{\delta}\left(\delta,p\right)$ is a P-matrix for every (δ,p), i.e. a square matrix such that all of its principal minors are strictly positive. Next, by the weak substitutability, $\mathbb{J}_{\sigma}^{\delta}\left(\delta,p\right)$ is also a Z-matrix, i.e. a matrix with non-positive off-diagonal entries. Finally, since a Z-matrix which is also a P-matrix is an M-matrix,[68] the result follows.

□

Appendix E: Additional Monte Carlo Simulations

E.1 Reference prices

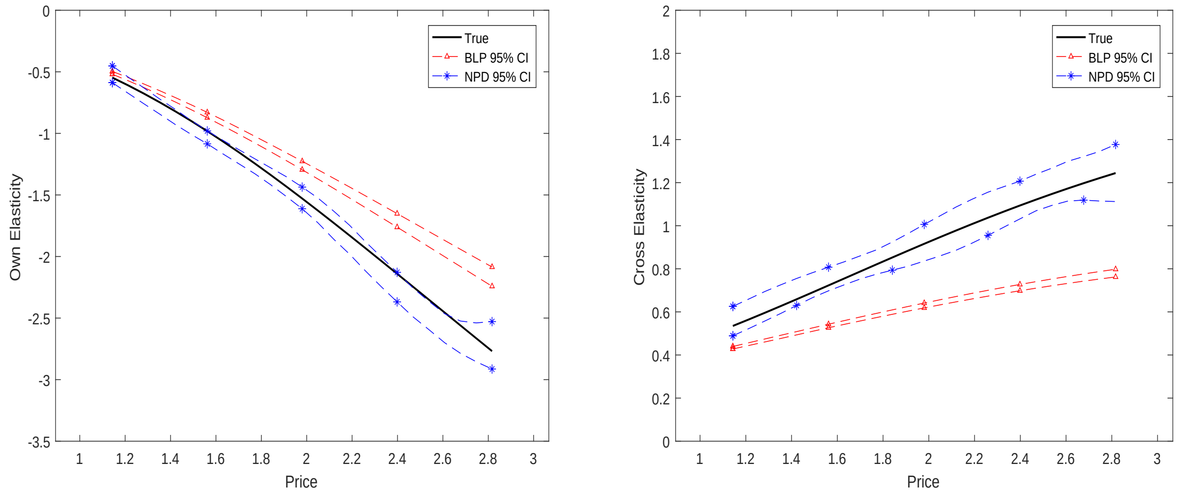
Another type of behavior allowed by the NPD model is one where consumers like (dislike) a product more if its price is lower (higher) than its competitor's, all else equal. The idea is that consumers might enjoy the feeling of getting a bargain and, conversely, might be turned off if they perceive a good is over-priced. I formalize this by letting the utility for good j be a function not only of the price of j but also a (decreasing) function of the difference between the price of j and that of its competitor. 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, I compare the performance of the nonparametric approach 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 nonparametric estimation, I impose the following constraints: monotonicity of σ^{-1}, diagonal dominance of $\mathbb{J}_{\sigma}^{\delta}$ and exchangeability.[69]

Figure 5 shows the own- and cross-price elasticity functions, respectively. While the nonparametric approach is on target, BLP tends to underestimate the magnitude of both due to the fact that it does not capture the reference pricing patterns in the data.

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

[69] See Section 3.2 and Appendix D for a discussion of these constraints.

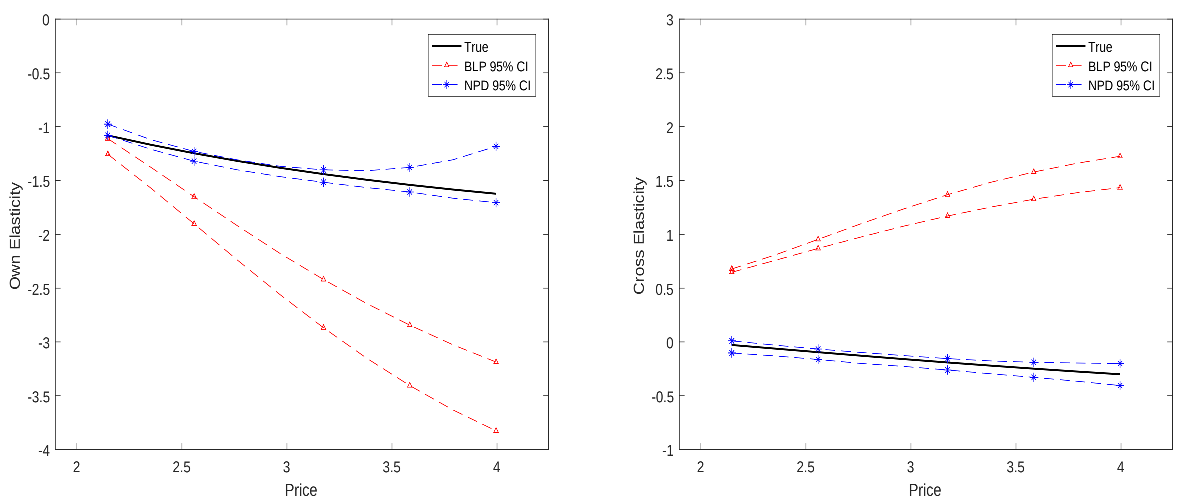
Figure 5: Reference Prices: Own-price (left) and cross-price (right) elasticity functions



E.2 Smaller Sample Size

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

Figure 6: Complements, $T = 500$: Own-price (left) and cross-price (right) elasticity functions



E.3 Violation of the Index Restriction

The NPD estimator is based on the index restriction embedded in Equation (1). Here, I explore how robust the estimator is to violations of this assumption. Specifically, I generate the data from the mixed logit dgp as in Section 4.1, except that I 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 7 to 9 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. These results complement those on the median elasticities (Table 3 in the main text) by showing robustness of the entire own- and cross-elasticity functions.

Figure 7: Violation of Index Restriction, $\sigma_{viol} = 0.10$: Own-price (left) and cross-price (right) elasticity functions

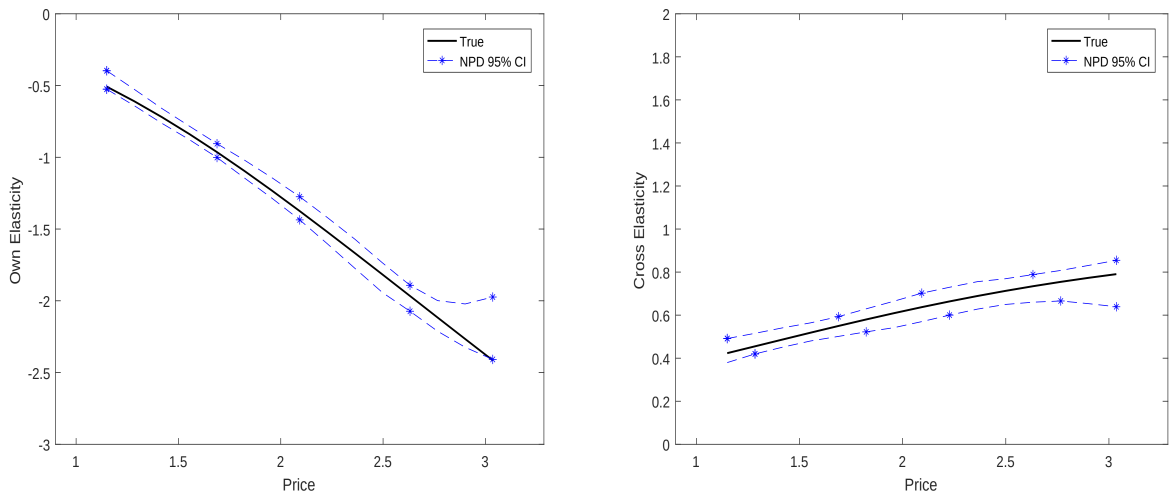


Figure 8: Violation of Index Restriction, $\sigma_{viol} = 0.50$: Own-price (left) and cross-price (right) elasticity functions

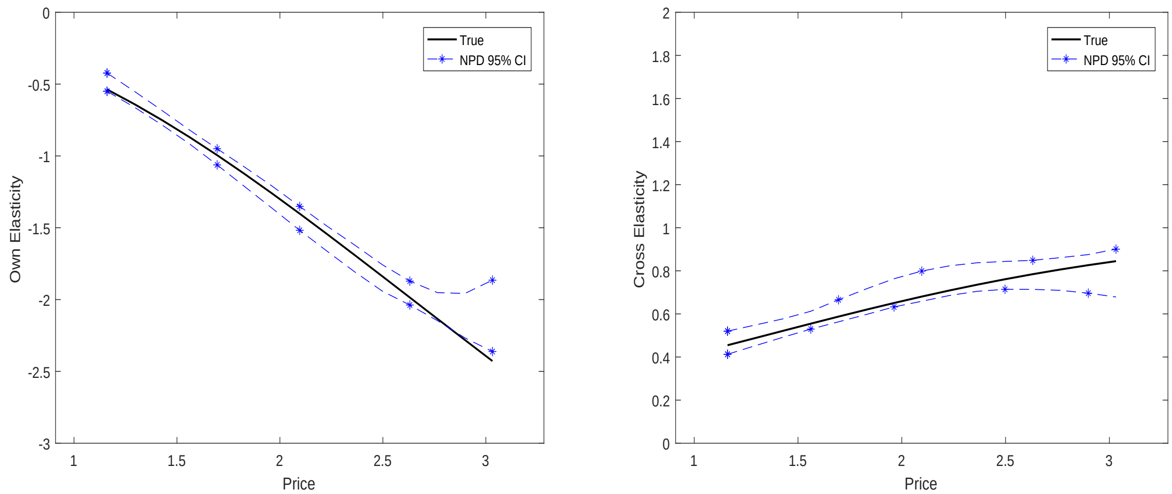
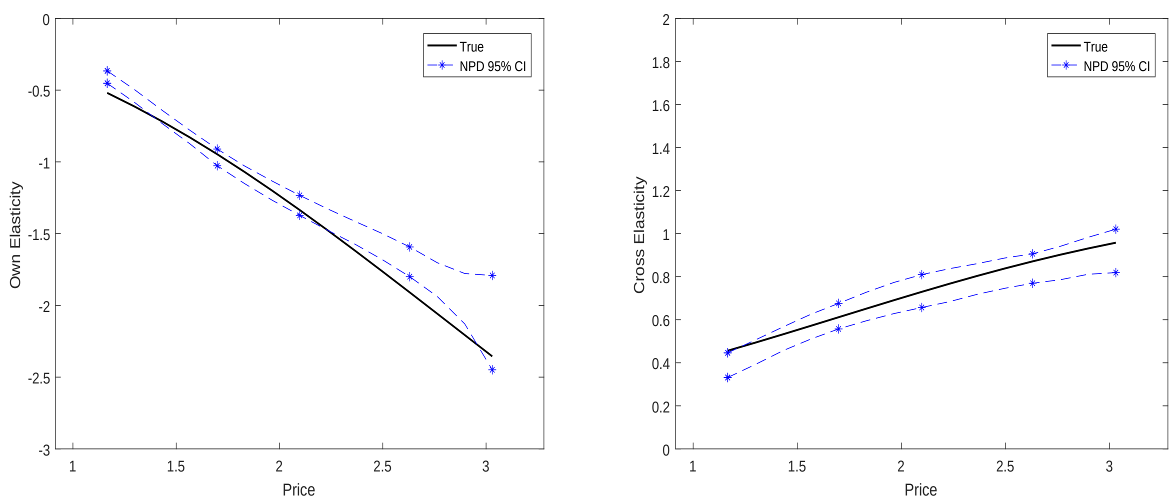


Figure 9: Violation of Index Restriction, $\sigma_{viol} = 1.50$: Own-price (left) and cross-price (right) elasticity functions



E.4 Sensitivity to the Choice of Polynomial Degree

To complement the results in Table 2 in the main text, I consider how the entire own- and cross-elasticity functions estimates vary as the degree for the polynomial approximation changes. I focus on the mixed logit dgp from Section 4.1 and the complements dgp from Section 4.3.

E.4.1 Mixed logit dgp

Figure 10: Mixed Logit Data, degree = 16: Own-price (left) and cross-price (right) elasticity functions

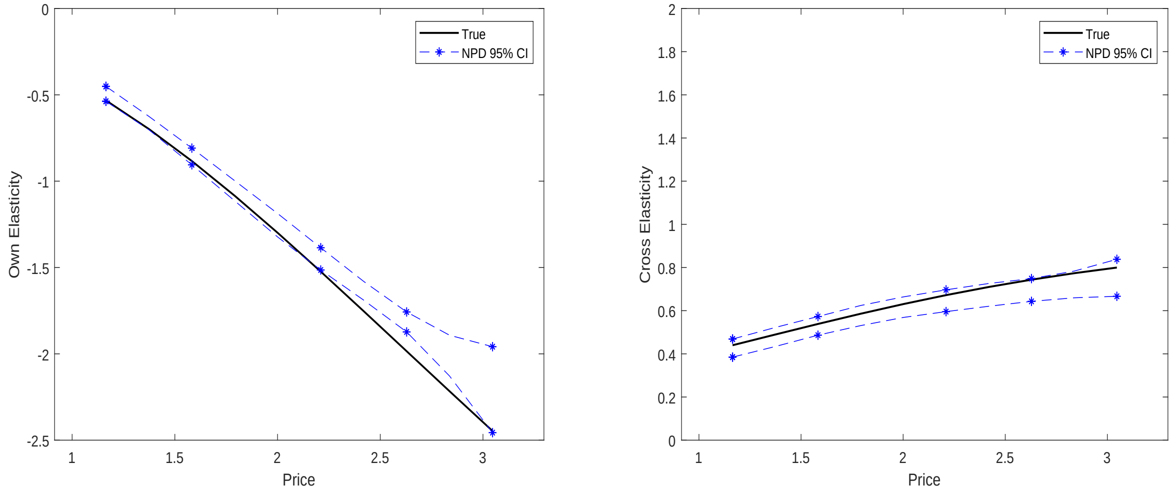


Figure 11: Mixed Logit Data, degree = 12: Own-price (left) and cross-price (right) elasticity functions

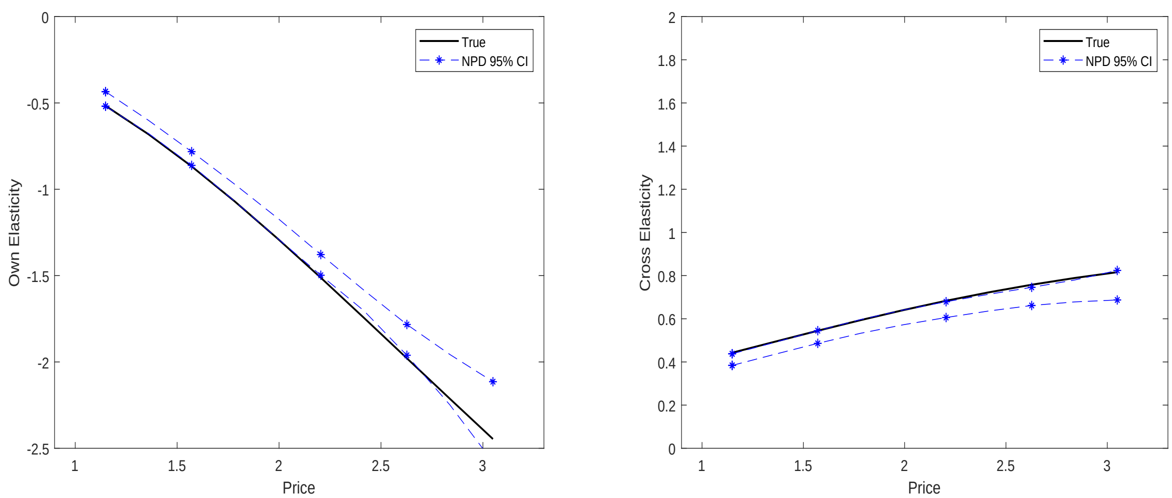


Figure 12: Mixed Logit Data, degree = 8: Own-price (left) and cross-price (right) elasticity functions

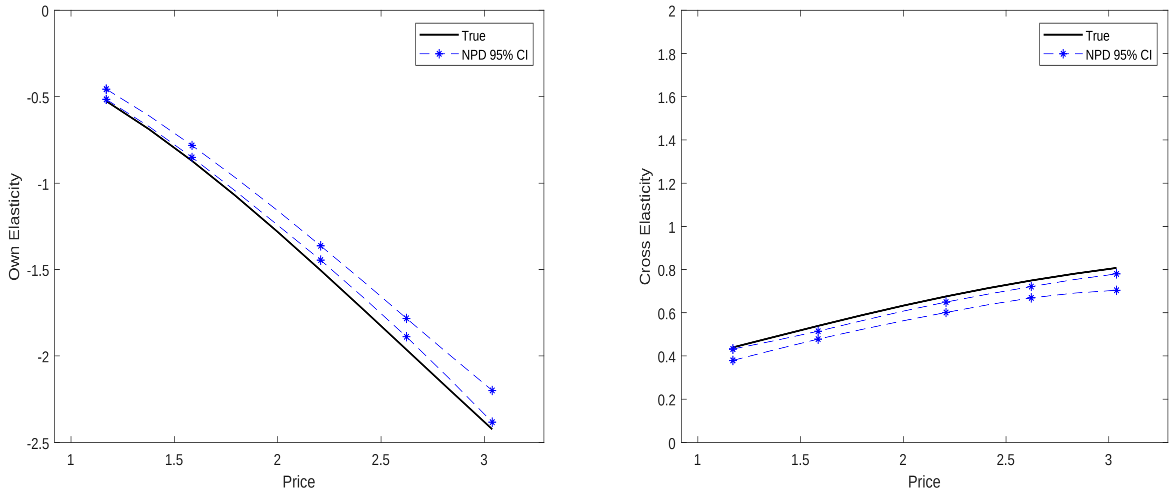


Figure 13: Mixed Logit Data, degree = 6: Own-price (left) and cross-price (right) elasticity functions

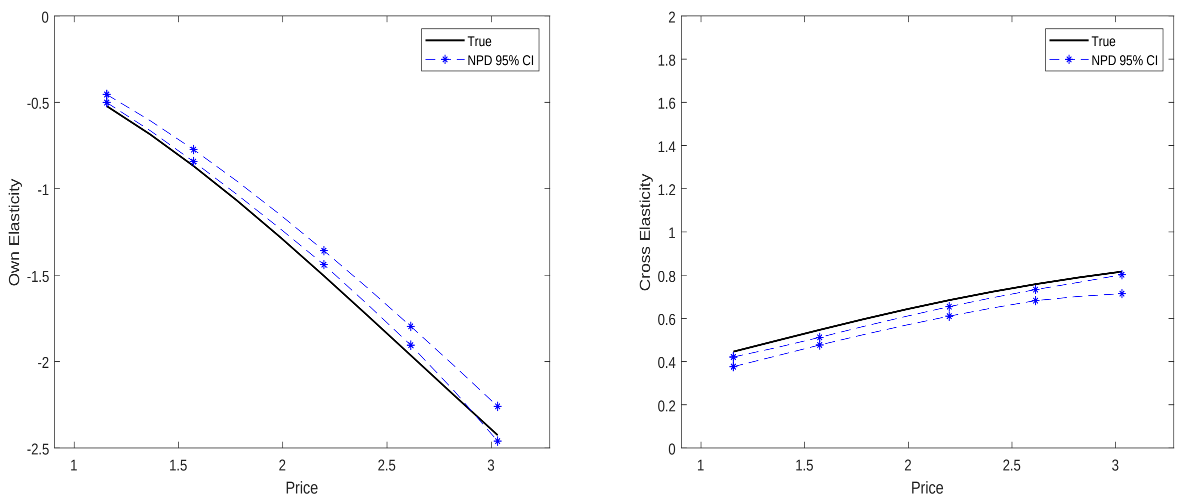
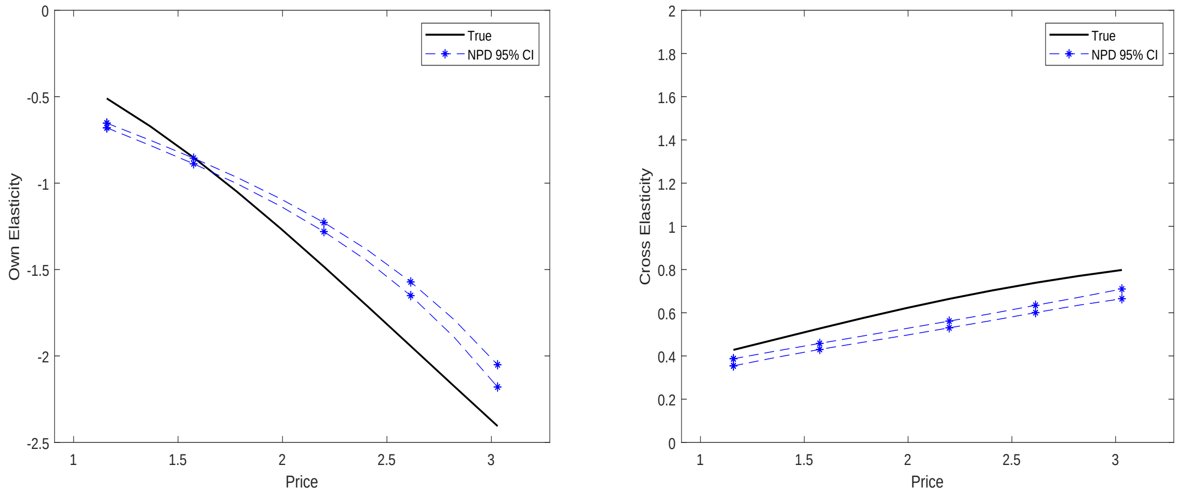


Figure 14: Mixed Logit Data, degree = 4: Own-price (left) and cross-price (right) elasticity functions



E.4.2 Complements dgp

Figure 15: Complements, degree = 16: Own-price (left) and cross-price (right) elasticity functions

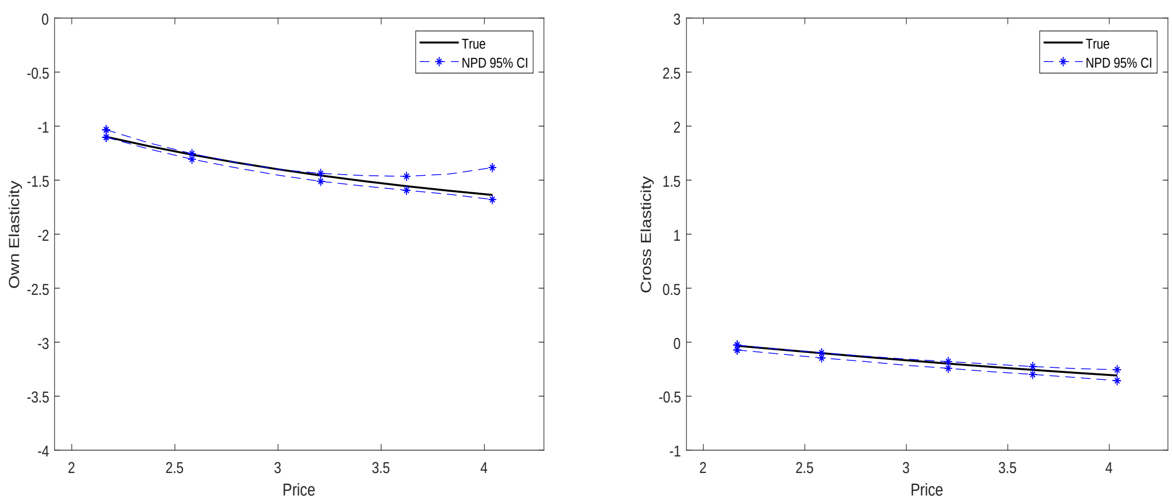
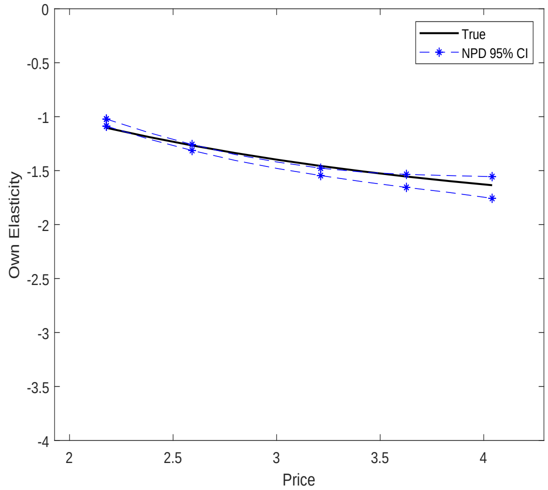


Figure 16: Complements, degree = 12: Own-price (left) and cross-price (right) elasticity functions



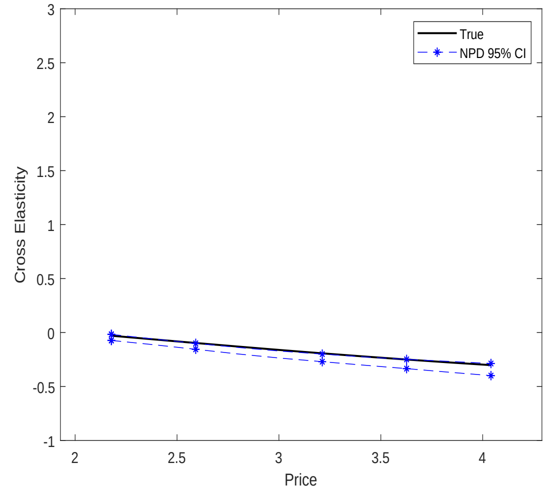
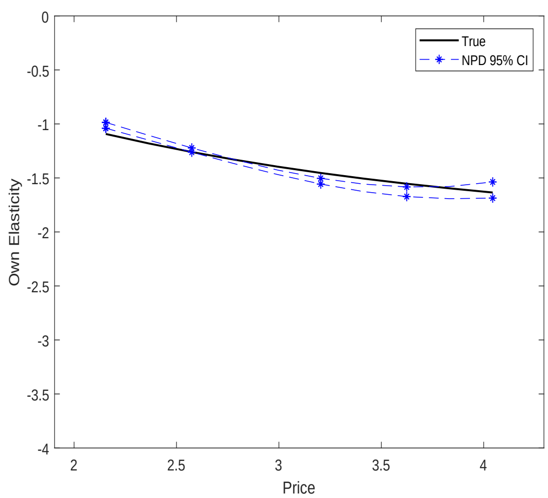


Figure 17: Complements, degree = 8: Own-price (left) and cross-price (right) elasticity functions



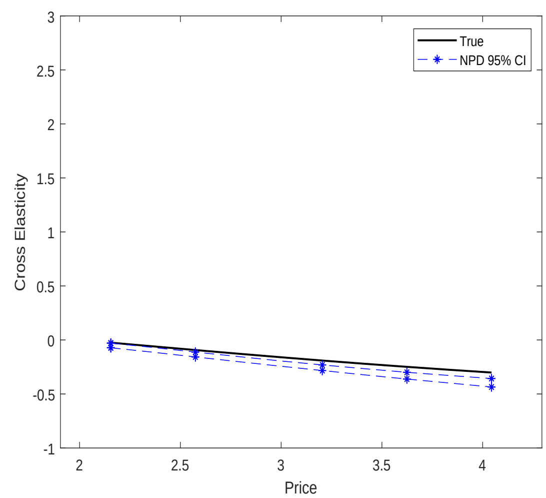


Figure 18: Complements, degree = 6: Own-price (left) and cross-price (right) elasticity functions

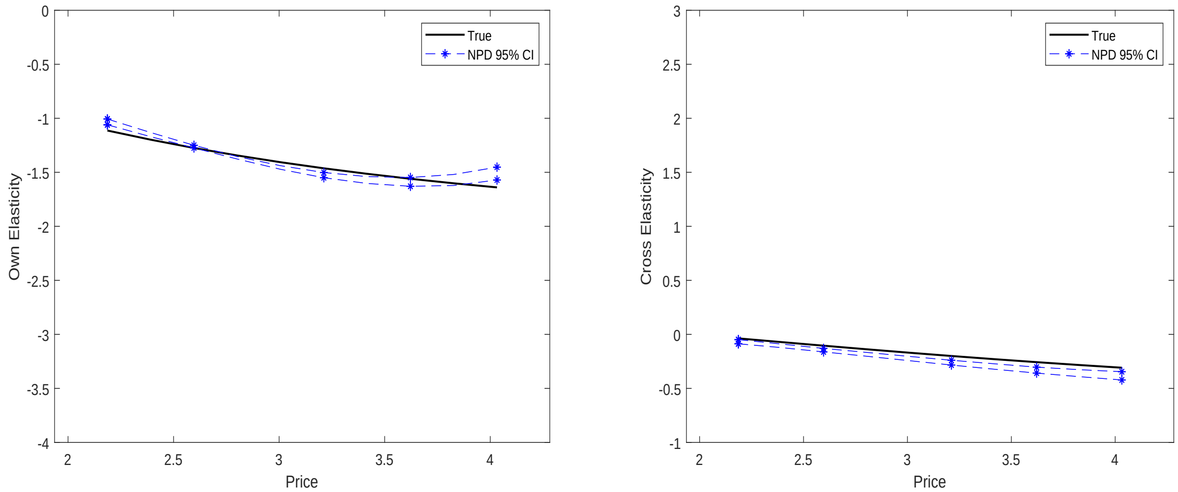
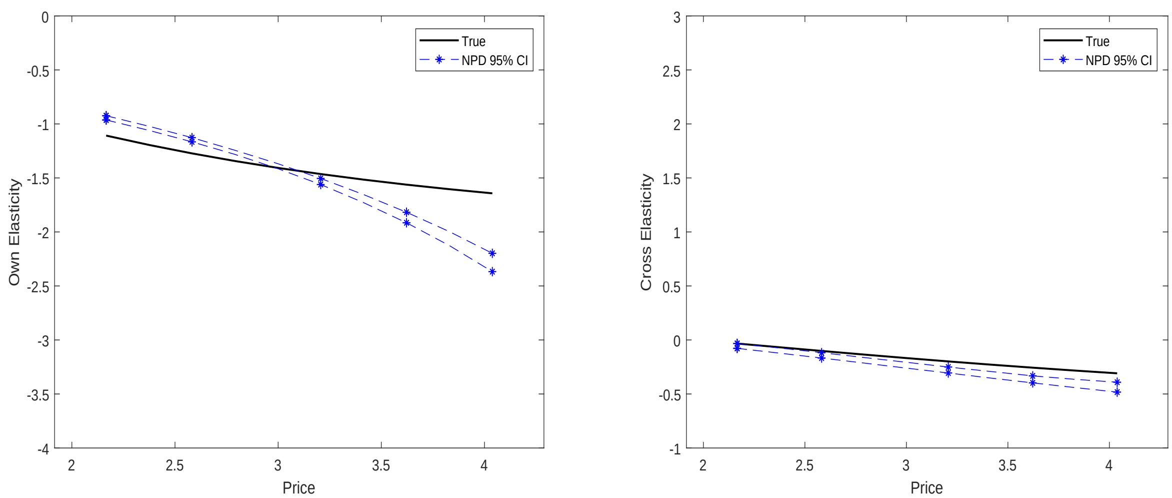


Figure 19: Complements, degree = 4: Own-price (left) and cross-price (right) elasticity functions



E.5 $J > 2$ goods

To complement the results in Table 3 in the main text, here I report estimates for the entire own- and cross-elasticity functions for the $J > 2$ goods case. I generate data from the logit model

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

I choose this simple model as it means that I can put p_j into the linear index δ_j, which reduces the number of parameters to estimate. I 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.[70]

Figure 20: Logit Data, $J = 3$: Own-price (left) and cross-price (right) elasticity functions

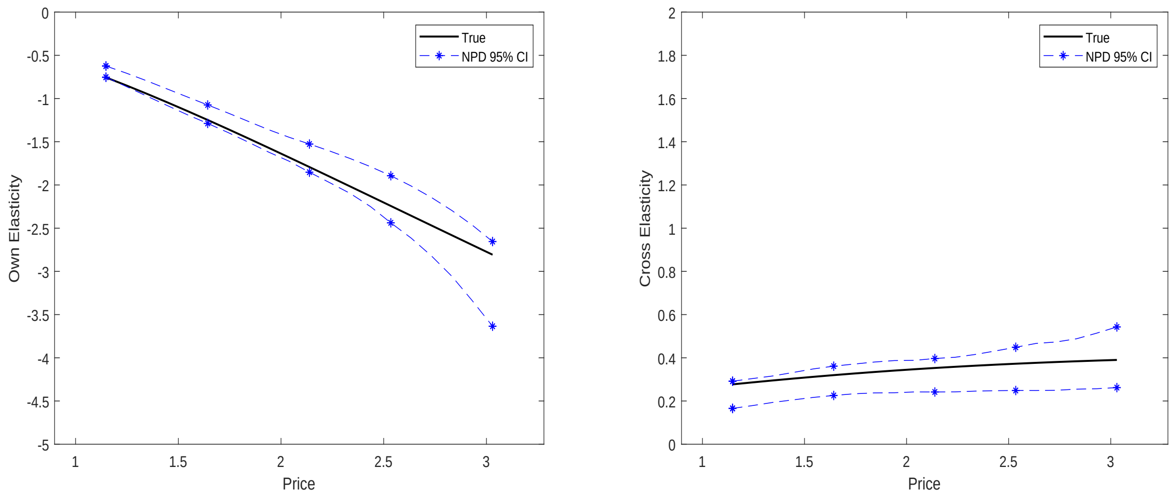
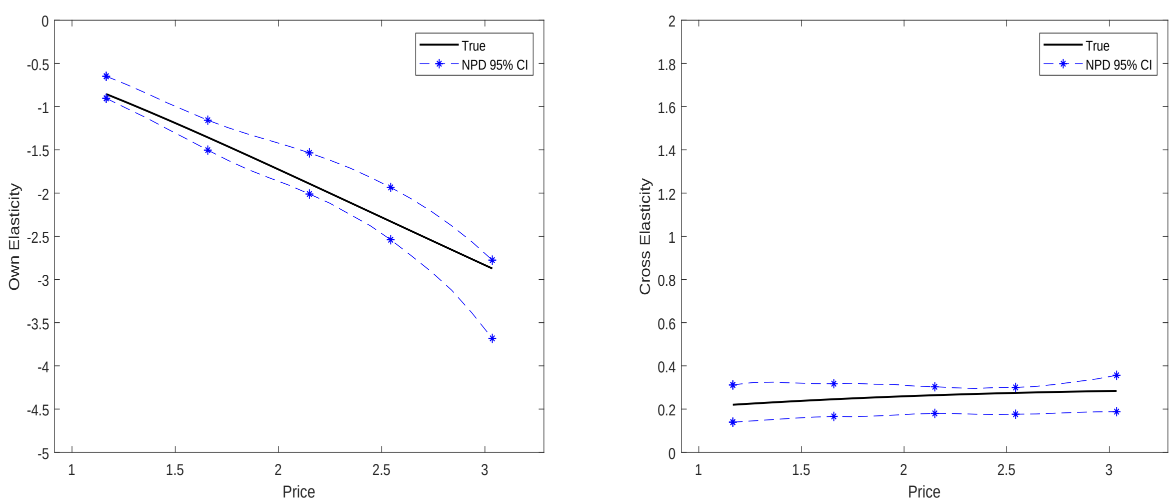
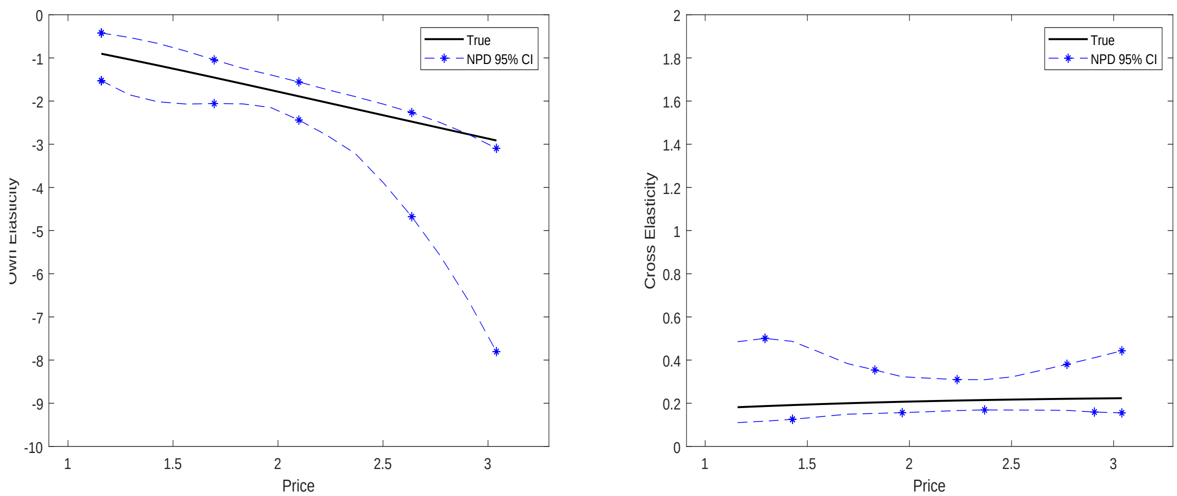


Figure 21: Logit Data, $J = 5$: Own-price (left) and cross-price (right) elasticity functions



[70]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 22: Logit Data, $J = 7$: Own-price (left) and cross-price (right) elasticity functions



Appendix F: Extension to Endogenous Demand Shifters

In this appendix, I consider violations of the exogeneity assumption that take the form $\mathbb{E}(\xi_j|x,z) = \gamma_j x_j$ for all j.[71] By Equation (2), for all j,

$$x_{jt} = \mathbb{E}\left[\frac{1}{\beta_j + \gamma_j}\sigma_j^{-1}\left(s_t, p_t, x_t^{(2)}\right)\bigg|x,z\right] \equiv \mathbb{E}\left[\mu_j\left(s_t, p_t, x_t^{(2)}\right)\bigg|x,z\right] \tag{39}$$

where I let $\mu \equiv [\mu_1, ..., \mu_J]' \equiv M_\mu \sigma^{-1}$ and M_μ is the diagonal matrix with (j,j) entry $\frac{1}{\beta_j+\gamma_j}$. Then, we can identify μ under completeness conditions as in BH. Let $\mathbb{J}_\mu^s$ denote the Jacobian of μ wrt s, and similarly for $\mathbb{J}_\mu^p$, $\mathbb{J}_\mu^{x^{(1)}}$, and $\mathbb{J}_\mu^{x^{(2)}}$. Note that $\mathbb{J}_\sigma^p = -\left(\mathbb{J}_\mu^s\right)^{-1}\mathbb{J}_\mu^p$, so that $\mathbb{J}_\sigma^p$ is identified. An analogous argument applies to $\mathbb{J}_\sigma^{x^{(2)}}$. On the other hand, since $\mathbb{J}_\sigma^{x^{(1)}} = \left(\mathbb{J}_\mu^s\right)^{-1}\tilde{M}_\mu$, where $\tilde{M}_\mu$ is the diagonal matrix with (j,j) entry $\frac{\beta_j}{\beta_j+\gamma_j}$, identifying μ 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.

Appendix G: Data

I take a market to be a week/store combination.[72] Data on prices and quantities come from the 2014 Nielsen scanner data set For each market, the most granular unit of observation in the Nielsen data is a UPC (i.e. a specific bar code). I aggregate UPCs according to whether they bear or do not bear the USDA Organic

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

[72]I use the terms "store" and "retailer" interchangeably.

Seal. When this information is missing, I assume the UPC is non-organic. The aggregate quantities are obtained by simply summing the quantities for the individual UPCs, whereas for prices I take a weighted average where the weights are determined by the yearly share of sales that a given UPC has in that store. Similarly, I aggregate across UPCs for selected non-strawberry fruits.[73] Specifically, I 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, I compute a price index (across UPCs) following the same procedure I 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, I 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[74] 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.[75] The data reports spot prices for the following shipping points: California, Texas, Florida, North Carolina, and Mexico. In absence of information on where supermarkets source their strawberries from, I take a simple average of the prices at the various shipping points in any given week.

I 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 data set in that week. To proxy for consumer tastes for organic produce at a given store, I 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.[76]

The resulting data set has 38,800 markets. Table 9 reports descriptive statistics for each variable and Figure 23 shows the price pattern for a typical store over time. 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 marked than for other items, such as packaged goods. Therefore, the model does not explicitly account for sales.[77]

---

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

[74] Here I follow the Nielsen partition of the United States into Designated Marketing Areas.

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

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

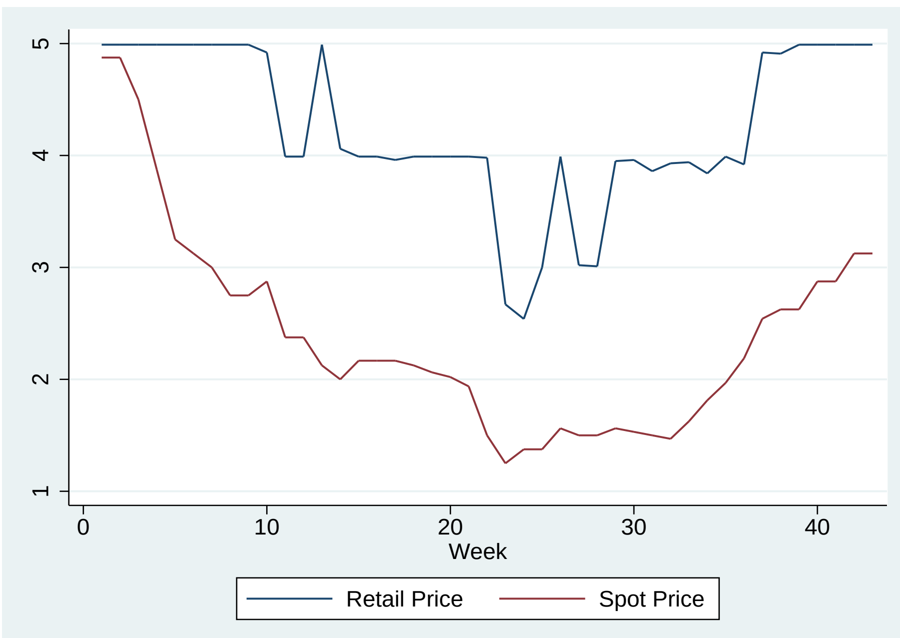
[77] Inventory is often invoked as a justification for sales in models of retail. However, because strawberries are so perishable, it is unlikely that inventory plays a first-order role in driving the retailer's pricing behavior.

Table 9: 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.

Figure 23: Price patterns



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

Next, I present the results of the first-stage regressions in Table 10. 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.

Table 10: 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.

Appendix H: Microfoundation of the Empirical Model

This appendix shows how to map the model estimated on the Nielsen data in Section 5 into the general framework outlined in Section 2. Specifically, I outline two models of consumer behavior that yield the demand system in equation (7) 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 (7) and invertibility, and that the estimation procedure does not rely on any of the parametric restrictions embedded in either model.[78]

H.1 Model 1

I 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 H.2. Moreover, given the prevalence of discrete choice models in the literature, it provides a connection between the demand system in (7) and a more familiar setup.

I 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

$$\begin{aligned} u_{i1} &= \theta_{str}\delta^*_{str} + \alpha_i p_1 + \epsilon_{i1} \\ u_{i2} &= \theta_{str}\delta^*_{str} + \theta_{org}\delta^*_{org} + \alpha_i p_2 + \epsilon_{i2} \\ u_{i0} &= \theta_{0,str}x^{(1)}_{str} + \theta_{0,org}\delta^*_{org} + \alpha_i p_0 + \epsilon_{i0} \end{aligned} \tag{40}$$

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

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. I 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), I let δ^*_{org} enter u_{i0}. In addition, u_{i0} also depends on the richness of the non-strawberry fruits offering, as captured by $x^{(1)}_{str}$. I 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, I 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 I show that the demand system generated by the model above is identified under the following assumption (as well as the standard exogeneity and completeness assumptions discussed in Section 2).

Assumption 10. *The coefficients* $\theta_{str}, \theta_{org}$,$\theta_{0,str}, \theta_{0,org}$ *and* $\theta_{1,org}$ *are non-zero.*

Note that Assumption 10 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 $x^{(1)}_{org}$ is indeed a proxy for taste for organic produce ($\theta_{1,org} > 0$).

Lemma 11. *Under Assumption 10, the demand functions* σ_1 *and* σ_2 *generated by the model in* (40) *are point-identified under the same set of conditions used to obtain identification in BH.*

Proof. Since utility is ordinal, I can subtract $\theta_{0,str}x^{(1)}_{str} + \theta_{0,org}\delta^*_{org} + \alpha_i p_0$ from each equation in (40) and write

$$\begin{aligned}u_{i1} &= \tilde{\delta}_1 - \theta_{0,str}x^{(1)}_{str} + \alpha_i\left(p_1 - p_0\right) + \epsilon_{i1}\\ u_{i2} &= \tilde{\delta}_2 - \theta_{0,str}x^{(1)}_{str} + \alpha_i\left(p_2 - p_0\right) + \epsilon_{i2}\\ u_{i0} &= \epsilon_{i0},\end{aligned} \tag{41}$$

where

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

Using (41) and the fact that the distribution of α_i is allowed to depend on $x^{(2)}$, we can write the demand system as

$$s = \tilde{\sigma}\left(\tilde{\delta}_1 - \theta_{0,str}x^{(1)}_{str}, \tilde{\delta}_2 - \theta_{0,str}x^{(1)}_{str}, p, x^{(2)}\right), \tag{42}$$

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 et al. (2013), we can invert the system in (42) for the mean utility levels as follows

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

where $\tilde{\sigma}_k^{-1}$ denotes the $k-$th element of the inverse, $\tilde{\sigma}^{-1}$, of $\tilde{\sigma}$. I now show that there is a one-to-one mapping between $\left(\delta^*_{str},\delta^*_{org}\right)$ and $\left(\tilde{\delta}_1,\tilde{\delta}_2\right)$. 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 10, we can rewrite (43) as

$$\delta^*=A^{-1}\tilde{\sigma}^{-1}\left(s,p,x^{(2)}\right)+A^{-1}\cdot[1 \quad 1]'\times\theta_{0,str}x_{str}^{(1)} \tag{44}$$

or, equivalently,

$$\begin{aligned}\delta^*_{str} &= \sigma_1^{-1}\left(s,p,x^{(2)}\right)+\theta_1 x_{str}^{(1)} \\ \delta^*_{org} &= \sigma_2^{-1}\left(s,p,x^{(2)}\right)+\theta_2 x_{str}^{(1)},\end{aligned} \tag{45}$$

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 I 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 (45) and using the definitions of δ^*_{str} and δ^*_{org}, we obtain

$$\begin{aligned}\xi_{str} &= \sigma_1^{-1}\left(s,p,x^{(2)}\right)+\frac{\theta_{0,str}}{\theta_{str}}x_{str}^{(1)} \\ \theta_{1,org}x_{org}^{(1)}+\xi_{org} &= \sigma_2^{-1}\left(s,p,x^{(2)}\right)\end{aligned} \tag{46}$$

The final step is to show that we can identify the system in (46), 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

(46) by $\frac{\theta_{0,str}}{\theta_{str}}$ and the second equation by $\theta_{1,org}$ without loss,[79] and rearrange terms as follows

$$-x_{str}^{(1)} = \sigma_1^{-1}\left(s, p, x^{(2)}\right) - \xi_{str} \tag{47}$$

$$x_{org}^{(1)} = \sigma_2^{-1}\left(s, p, x^{(2)}\right) - \xi_{org}, \tag{48}$$

Equations (47) and (48) 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 (47) and (48) yields the demand system in equation (7) that was estimated on the Nielsen data (after normalizations).

□

H.2 Model 2

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

$$\begin{aligned} &\max_{q_0,q_1,q_2} U_i\left(q_0, q_1, q_2\right) \\ &\text{s.t.} \;\; p_0 q_0 + p_1 q_1 + p_2 q_2 \leq y_i^{inc} \end{aligned} \tag{49}$$

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

$$q_j^*\left(d, p, y_i^{inc}, \epsilon_i\right) = \frac{y_i^{inc}}{p_j} \cdot \frac{d_j \epsilon_{i,j}}{\sum_{k=0}^{2} d_k \epsilon_{i,k}} \qquad j = 0, 1, 2 \tag{50}$$

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

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

[79] These divisions are well-defined operations as $\frac{\theta_{0,out}}{\theta_{str}}$ and $\theta_{1,org}$ are nonzero by Assumption 10.

and $\delta^*_{str}, \delta^*_{org}$ are defined as in Section H.1. I can then re-write (50) as

$$q^*_j\left(\tilde{d}, p, y^{inc}_i, \epsilon_i\right) = \frac{y^{inc}_i}{p_j} \cdot \frac{\tilde{d}_j \epsilon_{i,j}}{\sum_{k=0}^{2} \tilde{d}_k \epsilon_{i,k}} \qquad j = 0,1,2 \tag{51}$$

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^{inc}_i and ϵ_i in the market, and define[80]

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

$Q^*_j\left(\tilde{d}, p, x^{(2)}\right)$ 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 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

$$s = \tilde{\tilde{\sigma}}\left(\tilde{d}, p, x^{(2)}\right) \tag{52}$$

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 et al. (2013), we can invert (52) as follows

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

and, taking logs, we can write

$$\begin{aligned}
\theta_{str}\delta^*_{str} - \theta_{0,org}\delta^*_{org} &= \tilde{\sigma}_1^{-1}\left(s, p, x^{(2)}\right) + \theta_{0,str} x^{(1)}_{str} \\
\theta_{str}\delta^*_{str} + \left(\theta_{org} - \theta_{0,org}\right)\delta^*_{org} &= \tilde{\sigma}_2^{-1}\left(s, p, x^{(2)}\right) + \theta_{0,str} x^{(1)}_{str}
\end{aligned} \tag{53}$$

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

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 (53) has the exact same form as (43). Therefore, we can use the argument in Section H.1 to show that the demand system is identified.