Nonparametric Demand Estimation with Many Products: Microfoundations and Simulations

James Brand*

May 15, 2022

Abstract

In this note I argue for the use of simple machine learning tools to learn key features of consumer demand prior to nonparametric estimation. In particular, I propose the "hierarchical lasso" (Bien, Taylor and Tibshirani, 2013) to select relevant substitutes for each product. Implementing this approach solves three problems. First, it reduces one researcher degree of freedom in demand estimation (namely the choice of "nests"). Second, relative to the basic lasso/elastic net, this approach imposes a natural constraint on the non-zero coefficients which makes the resulting selected model more easily interpretable when demand functions are nonlinear. Finally, this solves a curse of dimensionality in nonparametric demand estimation (Compiani, 2021), thereby permitting the use of nonparametric estimation methods in settings with potentially many products. The argument here is implemented in the in-progress Julia package `NPDemand.jl`, and a number of simulations are presented using this package to demonstrate that the proposed approach performs well in complicated settings and under some forms of misspecification.

*Office of the Chief Economist, Microsoft. email:: jamesbrand@microsoft.com.

1 Introduction

The estimation of structural models of consumer demand has been a central focus of the IO literature for decades. Of particular interest are own- and cross price elasticities, as common counterfactuals involving firm pricing or profitability depend directly on these objects. An important practical question is how best to estimate the Jacobian of demand with respect to price. With J goods, the full Jacobian is a $J \times J$ matrix of derivatives, each of which is a function of all J prices, making this a complicated object to flexibly estimate in practical samples. The fundamental tension is between parsimony and flexibility. Popular models like nested and mixed logits make strong behavioral assumptions in order to reduce the problem to the estimation of an often small number of structural utility parameters (Berry, 1994; Berry, Levinsohn and Pakes, 1995). More flexible approaches, for example the nonparametric estimator introduced by Compiani (2021), relax these behavioral assumptions but introduce a curse of dimensionality which is often insurmountable in settings with many products.

One natural middle ground between these extremes is to allow for preferences for some sets of products to be correlated, even after conditioning on all product characteristics observed by the econometrician. This is commonly incorporated into empirical models by grouping products into disjoint "nests" prior to estimation, in a way which often reflects the available data and the researcher's priors. Because the predicted demand for each product, as well as the substitution patterns between products, depend directly on the structure of these nests,[1] these often-ignored choices can have substantive effects on the model's estimates. Moreover, many counterfactuals of interest depend on estimates of first and second derivatives of the aggregate demand function, which themselves depend on the specified model of demand (Mrázová and Neary, 2017).

In this note I outline what I call a *fully nested* demand model, which is a nested utility model in which groups of products are so segmented that all inter-group cross-price elasticities are zero. One example of such a model would be one in which, at each purchase event, consumers have strict preferences for types of liquor which imply zero substitution between whiskey and tequila for a wide range of prices. I then propose a two-stage estimation procedure in which I first select the relevant substitutes for each product and then estimate the implied demand system. This selection procedure is applied to each product separately, meaning that the substitution patterns it selects need not be symmetric (as is the case, for example, in some models of inattentive consumers), though symmetry can be easily enforced.

This note contributes to the literature concerning flexible demand estimation on a number of fronts. Researchers face many difficult practical choices in modeling and estimating demand. The foremost goal of this paper is to develop a procedure for these steps which is (1) data-driven, (2) applicable to a large number of settings, and (3)

[1] See e.g. equation (5) in Grigolon and Verboven (2014)

easy to implement. On (1), in most existing methods, researchers who estimate nested (mixed) logit models must provide the algorithm with predetermined non-overlapping nests. Although there is new work being done in this direction, this paper is among the first to introduce data-driven methods of determining the appropriate nests for a nested logit model or for verifying whether nests are indeed non-overlapping. As I show in simple simulations, the proposed procedure appears to identify the correct nests with high probability.

On (2), as discussed in Compiani (2021), the parametric restrictions researchers make regarding consumer preference heterogeneity while estimating mixed logit models can have substantive impacts on the resulting model estimates. Examples include the distribution of preferences for price, which are often assumed to be normal, and the common assumption that idiosyncratic preferences follow a Type 1 Extreme Value distribution (for which the variance is identical across products). By focusing on nonparametric estimation of the selected model of demand, we avoid many of these issues. Finally, I have written the Julia package NPDemand.jl to accomplish goal (3). This package can be called from Julia, R, or Python (the latter two through existing tools for cross-language communication), and the methods for model selection and demand estimation are straightforward and simple for users to implement. Because the model selection approach I take is a constrained lasso regression, this step is relatively quick even when cross-validating the relevant regularization parameters. Similarly, because the nonparametric estimator in Compiani (2021) is, in its simplest form, a two-stage least squares estimate, estimation is quick and computationally cheap as long as the selected nests are not too large. This package is thus a quick and easy way to generate baseline or supplemental estimates of consumer demand in research or industry settings.

There is already a growing body of work which combines machine learning tools with traditional demand estimation. Gillen, Shum and Moon (2014) estimate a model much like that of Berry, Levinsohn and Pakes (1995), and focus on settings in which the number of product characteristics is large. Extending this work, Gillen et al. (2018) make use of the linearity of the log of market shares in a logit model and include model selection in the first and second stages of an IV model, followed by a post-selection estimation of the selected model in the spirit of Belloni et al. (2012). Contemporary work has also begun to focus on identifying relevant close substitutes for products in markets with many options by applying machine learning tools to nested logit models or generalizations thereof. The approach here is complimentary to this existing work. Conducting model selection separately for each product's nonparametric demand function, as is done here, is appealing for its ease of implementation and because it can draw on a large body of work using regularized regressions for model selection. However, methods which make use of the mixed and nested logit structure during model selection may be more efficient as well as more easily interpreted. These approaches also provide estimates of utility parameters, which may be important for some empirical questions or counterfactuals.

I run a series of simulations to evaluate the performance of, the proposed approach. I consider cases in which the demand for two groups of products is independent, and the econometrician has data on all products but does not know which are relevant substitutes for others. First, I show that the selection procedure I propose consistently selects only the relevant substitutes for each product in a simple case for which the proposed model is correctly specified. I then conduct a series of more complicated simulations demonstrating that this procedure is robust to some forms of misspecification. In these simulations, I show that the combination of model selection, a violation of an assumed index structure, and purposefully mis-selected nests appears to generate little to no bias in estimates of own-price elasticities.

2 Fully Nested Discrete Choice

2.1 Model

In this section I discuss a simple random utility model in which consumers have strong preferences for small groups of products, meaning they ignore the attributes of all products outside of their preferred group. I call this model a *fully nested* demand model, because it can be thought of as a typical nested logit model with two modifications: preference heterogeneity for inside options have bounded support, and group-level shocks are large relative to differences in characteristics across products. This structure "fully" nests each product, in the sense that *all* substitution will be within-nest and to the outside option.

In each market t, let there be J products, each of which falls into one of G non-overlapping groups.[2] Products are assumed to be differentiated by prices p_t and an unobservable characteristic ξ_t, each a vector with J entries. Let $\chi_t = (p_t, \xi_t)$ be an element of the characteristic space $\mathcal{X}$.[3] Let the indirect utility received by consumer i from product j in group g be

$$u_{ijt} = f(\delta_{jt}(p_{jt}, \xi_{jt})) + \lambda_{ig} + \epsilon_{ijt} \tag{1}$$

and the utility of the outside good be $u_{i0t} = \epsilon_{i0t}$. I emphasize in Equation 1 that each unobservable component of utility can depend on some observables, but suppress that notation for most of the remainder of the discussion (though these functional dependences are permitted throughout). Although I do not include any observable product characteristics other than price, other characteristics can be incorporated easily, and any characteristics which are fixed across markets (e.g. flavor, size, etc) are already permitted in Equation 1 via either product-specific constants in $\delta(\cdot)$ or group-level effects in λ_{ig}.

[2] The latter is simply for exposition; in a nonparametric setting, groups of products can overlap. This may be relevant, for example, if a diet sodas are substitutes for the non-diet version of the same soda and for other diet sodas but not for the non-diet version of competing sodas.

[3] I assume throughout that the characteristic space $\mathcal{X}$ is a Cartesian product.

The first term of Equation 1 (δ_{jt}) is the mean utility of product j in market t, which is shared by all consumers in market t. The second term λ_{ig} is a group-level shock, the distribution of which can depend on group characteristics (fixed across markets), and which shifts all products in a group homogeneously within consumer. The residual term ϵ_{ijt} represents all remaining preference heterogeneity (which can depend on prices). In a traditional mixed logit model, δ_{jt} represents the mean preferences for all product characteristics, and ϵ includes the logit error as well as all preference heterogeneity which is correlated with product characteristics. Equation 1 is far more general, encompassing the nested logit, probit, and most other common random utility models. I now make three restrictions on the unobservables in this model

Assumption 1. *(i) For all i, for some g_i, for all $g_i' \neq g_i$,*

$$\lambda_{ig_i} - \lambda_{ig_i'} > \max_{j \in g_i, j' \in g_i'} \max_{\chi_t \in \mathcal{X}_t} [f(\delta_{j't}) - f(\delta_{jt}) + \epsilon_{ij't} - \epsilon_{ijt}]$$

(ii) $supp(\epsilon_{i0}) = \mathbb{R}$ for all i

(iii) The index δ_{jt} is linear, i.e. $\delta_{jt} \equiv p_{jt} + \xi_{jt}$, and $f(\cdot)$ is strictly monotonically increasing.

These modeling assumptions are, at face value, high-level distributional restrictions on the unobservables λ, δ and ϵ, but they capture a simple intuitive structure which is of empirical interest. If Assumption 1(i) holds then each consumer i will always choose a product in the corresponding group g_i of products under consideration (which varies by consumer). Further, because this inequality holds for all $\chi_t \in \mathcal{X}$, changes in the characteristics of goods not in group g have no effect on the demand for goods in group g, because they do not induce *any* consumers to change their purchasing behavior.[4] This generates what is a the core feature of the model: under Assumption 1, each good only substitutes for relatively few others on the margin.[5] Note that I do not assume that the number or composition of groups G are known, so this is quite general. The biggest limitation of this assumption is that it requires that the groups do not differ by market.

A model which satisfies Assumption 1(i) and in which the distribution of preferences ϵ_{i0} for the outside option is very small (e.g. a point mass at zero) is one in which each group of products does not strictly substitute to any other good in the model. In such a restricted model, it is unclear why products in different groups should be

[4]This inequality does not need to hold for all $\chi_t \in \mathcal{X}$. Rather, it needs to hold within the subset of $\mathcal{X}$ which contains the researcher's data and any counterfactuals she wishes to run. Thus, one does not need to assume (counterintuitively) that consumers would be unresponsive to extreme changes in the prices of products outside group g (e.g. zero prices).

[5]One could generate this pattern alternatively by assuming that there exist a limited number of non-overlapping consideration sets and that each consumer's consideration set is constant over time. Without additional restrictions, such a model would generate identical predictions to those herein.

included in the same model at all. In order to exclude this unintuitive setting, and for other reasons which will be elaborated below, Assumption 1(ii) requires that the utility of the outside option have full support on the real line. Thus, any marginal change in a good's characteristics can induce switching to or from the outside option, regardless of λ_{ig} or any product characteristics. Finally, Assumption 1(iii) is an index restriction common in this literature dating back to Berry (1994). In a traditional logit model in which the variance of utility shocks is left unspecified, researchers often normalize the coefficient on one product characteristic to 1. This assumptions serves a similar purpose here, effectively normalizing the scale of the unobservable ξ to that of prices p_{jt}. I assume for ease of notation that ξ enters δ linearly, though all that is necessary for identification is that δ_{jt} is monotonic in ξ_{jt}.

I have assumed so far that groups partition the set of products, meaning that all groups are non-overlapping. This is an expositional tool to demonstrate a simple case which is sufficient to generate the aggregate model of interest, and is not necessary for identification or in estimation. In other words, substitution patterns may be asymmetric, such as in a model with a leading brand which affects demand for smaller competitors (e.g. because its price is salient to consumers) but which is unaffected by each small firm's prices.[6] All that is required in principle is that the demand for any product depends on the characteristics of relatively few (i.e. less than five or ten) other products and that consumers agree on these groups. What the previous subsection highlights is that this condition can be thought of as requiring that across-group preference heterogeneity is sufficiently larger than product- or consumer-specific heterogeneity.

2.2 Realism of Fully Nested Demand

Although many empirical papers have applied the nested *logit* model, the reader may be concerned as to the realism of the previous model of *fully* nested demand. Though the models are similar in spirit, the nested logit model permits substitution between all products in a demand system, whereas fully nested demand explicitly does not. There are three reasons I view this as a reasonable restriction. First, empirically, estimates of many cross-price elasticities in large demand systems, and especially those for products in different intuitive subsets, tend to be very small. Berry, Levinsohn and Pakes (1995), who study the market for automobiles in the U.S., estimate that a $1000 increase in the price of a Nissan Sentra increases the market share of the Honda Accord by 2.5% and of the Ford Escort by 8.2%, but has a negligible effect on the demand for the BMW or Lincoln models considered. In the market for cereal, Nevo (2001) estimates cross-price elasticities that range from 0.02 to more than 0.3. Thus, even before considering nested logit models explicitly, there is evidence that other

[6] Abaluck and Adams (2018) have shown that asymmetries in the Jacobian of demand are informative of consumers' consideration sets. This intuition applies here as well under additional assumptions on the similarity of the distributions of ϵ_{ij} across different products.

components of product differentiation make some cross-price elasticities very small.

More recently, Miravete, Seim and Thurk (2018) estimate demand for liquor in Pennsylvania in a random coefficient nested logit model and find that within-nest cross-price elasticities are between 3 and 30 times larger than across-nest elasticities. Miller and Seo (2018), in studying the substitution between alcohol and marijuana, estimate that the cross-price elasticities between beer, wine, and liquor are between 0.001 and 0.11, and cannot reject that any of these elasticities are zero at the 95% level. In a study of automobile taxes in Brazil, Chaves (2019) presents estimates of some cross-price elasticities which are as large as 0.8 and others which are smaller than 10^{-3}. Thus, some products are far more relevant determinants of the demand for others. Even if these small cross-elasticities are not exactly zero, which is difficult or impossible to test, the loss from ignoring them may be quite small.

Second, though the nested logit model can be quite flexible, there are well-known restrictions on the substitution patterns it can rationalize. For example, one could say that the nested logit "permits" substitution between all products, but it also *requires* it via the full support of the logit distribution. Even the least desirable high-price, low-quality products will be chosen by each consumer with positive probability. The nonparametric approach is appealing in this sense, as Compiani (2021) provides examples of demand systems which do not fit the standard multinomial logit (nor nested logit) framework but which do satisfy his assumptions. If these small estimated cross-price elasticities represent true zeros, then logit and nested logit models may be too restrictive by assuming substitutability between products that are in fact fully segmented.

Third, there are a variety of existing models which have a similar flavor as the simple choice model presented here. Tversky (1972)'s elimintation by aspects, which models choice as a sequential process of elimination of irrelevant products, is particularly similar. In fact, Train (2009) mentions that Tversky's model is the limit of the nested logit model as the nesting parameter approaches 1 (which eliminates across-group substitution). Similarly, Train and Sonnier (2005) consider a mixed logit model in which the mixing distribution has bounded support. There is also a large marketing literature which studies consideration sets, i.e. the small subset of products among many over which the consumer makes an active comparison (e.g. Mehta, Rajiv and Srinivasan (2003)). Even if consumers' consideration sets are heterogeneous, as long as no two possible consideration sets overlap with one another, demand may be fully nested.

2.3 Aggregation and Inversion

Now we can aggregate the preceding model of individual choice to a model of market-level demand. Let s_{jt} denote the market share of j in t, and s_t the vector of

market shares for market t. Now define the market-level demand function

$$\sigma : \mathcal{X} \to \Delta^J \tag{2}$$

from the characteristic space to the unit simplex in J dimensions. This will be the object the researcher aims to identify from the data. Because we will not attempt to identify the distributions of δ, λ, and ϵ separately, I write σ as a nonparametric function of product characteristics, i.e. $\sigma(\chi_t) = \sigma(p_t, \xi_t) = \sigma(\delta_t)$, where the second equality comes from the index assumption above. To understand the properties of market demand in this setting, consider the probability that consumer i chooses product k:

$$\begin{aligned} P(i \text{ choose } k) &= P(u_{ikt} \geq \max_{j \neq k} u_{ijt}) \\ &= P(f(\delta_{kt}) + \lambda_{ig} + \epsilon_{ikt} > f(\delta_{j^*t}) + \lambda_{ig^*} + \epsilon_{ij^*t}) \\ &= P(\lambda_{ig} - \lambda_{ig^*} > f(\delta_{j^*t}) - f(\delta_{kt}) + \epsilon_{ij^*t} - \epsilon_{ikt}) \end{aligned}$$

where j^* denotes the best alternative to k and g^* is the group which contains j^*. Note that this inequality, by Assumption 1, is never satisfied when $g \neq g^*$, and is unaffected by changes in the product characteristics. Thus, the derivative of these choice probabilities, for any two products in different groups, is zero. Because these inequalities hold for all consumers, market shares clearly inherit this same property. When j^* and k are in the same group, the same choice probability becomes

$$\begin{aligned} P(u_{ikt} \geq \max_{j \neq k} u_{ijt}) &= P(0 > f(\delta_{j^*t}) - f(\delta_{kt}) + \epsilon_{ij^*t} - \epsilon_{ikt}) \\ &= P(f(\delta_{kt}) > f(\delta_{j^*t}) + \epsilon_{ij^*t} - \epsilon_{ikt}) \end{aligned}$$

Clearly, under some minor regularity conditions, consumer demand for product k will be increasing in δ_k and declining in δ_{j^*t}. This brings us to the assumptions we will require the aggregate demand to satisfy.

Assumption 2. *(i)* $\frac{\partial}{\partial \delta_{jt}} \sigma_k(\delta_t) \leq 0$ *for all* $j > 0$, $k \neq j$ *and all* δ_t

(ii) For every $\mathcal{K} \subset \mathcal{J}$ *and every* δ_t, *there exist a* $k \in \mathcal{K}$ *and a* $j \notin \mathcal{K}$ *such that* $\frac{\partial}{\partial \delta_{kt}} \sigma_j(\delta_t) < 0$.

As shown in Berry, Gandhi and Haile (2013), Assumption 2 implies that the demand function σ (and each component j) is invertible in δ_t.[7] Given the invertibility of $\sigma(\cdot)$ in δ_t, we can write, for each j,

$$\begin{aligned} \delta_{jt} &= \sigma_j^{-1}(s_t) \\ p_{jt} &= \sigma_j^{-1}(s_t) - \xi_{jt} \end{aligned}$$

[7] We also require that the characteristic space $\mathcal{X}$ is a Cartesian product, which is assumed at the beginning of the previous subsection

These functions σ_j^{-1}, and functionals thereof, will be our objects of interest for the remainder of the paper, and our task will be to estimate each using only market-level data.

3 Implementation in Two Stages

In this section I describe the implementation of the two-step estimation procedure. In the first step, I select the appropriate model of demand for each product using some form of penalized regression. Second, I take the model selected in the first stage as given and apply the method introduced by Compiani (2021) to estimate each demand function with reasonably few arguments. I discuss these stages in reverse order, beginning with the estimation of demand conditional on a selected model, and then discussing the model selection procedure.

3.1 Second Stage: Demand Estimation

First, suppose that we have already selected a model (specifically, the relevant substitutes for each product). Then, under the selected model, I estimate the demand function nonparametrically following Compiani (2021) directly. As a result, this section borrows heavily from that paper's structure and notation. The researcher is assumed to have data on prices p and other (exogenous) product characteristics x, as well as market shares, for $J^{(j)}$ selected products (different for each j) in T markets. Letting uppercase letters denote the random variables corresponded to lowercase realizations, I begin with the following assumptions from Compiani (2021)

Assumption 3. $\mathbb{E}[\xi|Z] = 0$ *almost surely in* Z

Assumption 4. *For all functions* $B(\cdot)$ *with finite expectation, if* $\mathbb{E}[B(S,P)|Z] = 0$ *almost surely in* Z *then* $B(S,P) = 0$ *almost surely in* (S,P).

Assumption 3 assumes the availability of an instrument for price, and 4 is a standard completeness condition in the NPIV literature. See e.g. Newey and Powell (2003) for details on this assumption. As shown in Compiani (2021), Assumptions 2, 3, and 4 guarantee that the demand function σ is invertible and point-identified. See that paper for a proof and further discussion. I apply the Compiani (2021) estimation sieve-based estimation procedure, which comes originally from Chen and Christensen (2018). Our estimating equation is the product-specific inverse demand function

$$p_{jt} = \sigma_j^{-1}(s_t) - \xi_{jt} \tag{3}$$

Let $\sigma^{-1} \in \Sigma$ (σ used to denote all demand functions σ_j jointly) for some function space Σ, $\tilde{\sigma}(\cdot)$ denote the sieve analog of σ for a given set of parameters, and let Σ_T denote the sieve space for a given sample size T. As in Compiani (2021), I approximate Σ_T as the product of univariate Bernstein polynomial basis functions, the order of which can

vary with T and for each product j. Similarly, for a given sample size I approximate the instrument space by a fully interacted set of univariate Bernstein polynomials in exogenous variables (X, Z). Let A denote this matrix, and let a_t denote the row of A corresponding to market t.[8] For each guess of $\tilde{\sigma}$ I construct instrument-weighted residuals

$$r_{jt}(s_t, p_t, z_t, \tilde{\sigma}^{-1}) \equiv (p_{jt} - \tilde{\sigma}_j^{-1}(s_t)) \times a_t(z_t) \tag{4}$$

and solve the quadratic minimization problem[9]

$$\min \sum_{j=1}^{J} \Big[\sum_{t=1}^{T} r_{jt}(s_t, p_t, z_t; \tilde{\sigma}_j^{-1}) \Big]' (A'A)^{-} \Big[\sum_{t=1}^{T} r_{jt}(s_t, p_t, z_t; \tilde{\sigma}_j^{-1}) \Big] \tag{5}$$

Chen and Christensen (2018) show consistency and inference results for the general case of this estimator, and the appendix of Compiani (2021) contains those results as applied directly to demand estimation, as well as a discussion of the convenient properties of Bernstein polynomials.

3.2 Structured Sparsity

As discussed above, the issue with implementing this estimator directly in a setting with many products is that the number of terms in the sieve $\tilde{\sigma}$ grow exponentially quickly. In markets with 10 products, estimating the full demand system can require estimating more than one million parameters. One route to aiding the search procedure as the number of parameters increases is to introduce constraints which require some parameters to be equal (e.g. exchangeability). This is the approach taken by Compiani (2021), and he demonstrates how to impose these constraints in estimation to reduce the number of estimated parameters drastically.

An alternative approach to to reducing the dimension of the parameters space is to impose the following prior, guided by the fully nested demand model above: some goods are unlikely to ever serve as substitutes for one another. The approach I take to imposing this prior on the model of demand begins with the inverse demand function:

$$p_{jt} = \sigma_j^{-1}(s_t) - \xi_{jt} \tag{6}$$

Now, our task is to determine which products are the most relevant substitutes for product j (for each j). Given my preceding arguments that products in many markets

[8] Note that this matrix, including the order of Bernstein polynomial used to construct it, can differ for each product. I suppress this in my notation here.

[9] In the simulations herein, I do not sum over products J. Rather, I solve each problem independently. I do this because this saves a significant amount of time (as the solution to this problem at the product level has a closed form) and because it does not significantly affect the performance of the estimator at the median in these tests.

are fully nested, our prior should take the form of a function σ^{-1} which is a function of only a subset of competing market shares. Importantly, we want to do this while still permitting σ^{-1} to be nonlinear in its arguments.

For this I turn to a recent literature on "structured sparsity." To understand this approach, let us begin by considering a simple machine learning procedure like the lasso (Tibshirani, 1996). Suppose that we are considering a market with two products, and to apply the lasso to our problem we approximate σ_1^{-1} (the inverse demand function for product 1) with a quadratic function:

$$p_{1t} = \beta_1 s_1 + \beta_2 s_2 + \beta_{12} s_1 s_2 + \beta_{11} s_1^2 + \beta_{22} s_2^2 - \xi_{jt} \tag{7}$$

It is possible that this lasso regression results in estimates indicating that β_1, β_{11}, and β_{12} are nonzero but β_2 and β_{22} are set to zero. In order to make use of this selection pattern, we are forced to make a choice as to whether the inclusion of β_{12} implies that product 2 should be included in the demand function for product 1. We would face a similar problem if β_{12} were zero and β_{22} were not. The safer option in each case is clearly to include product 2, as we can still estimate zero cross-price elasticities in the second step of estimation. However, especially when we are attempting to select among many products, this choice can largely or entirely undo the model selection performed by lasso.

This problem arises in part because the basic lasso penalty does not impose any knowledge on the relationships of different coefficients.[10] In many situations calling for shrinkage this is reasonable, but in our context this excludes some important information. If products 1 and 2 are not substitutes, then β_{12}, β_2, and β_{22} should *all* be zero. In other words, we are interested in imposing *structure* on the sparsity of the coefficients in Equation 7. Namely, we want to impose one of the following constraints:

- β_{12} and β_{22} are nonzero only if β_2 is also nonzero.
- β_{12} is nonzero only if β_1 *and* β_2 are nonzero

For this I employ the hierarchical lasso, developed recently by Bien, Taylor and Tibshirani (2013) and implemented in the R package hiernet. This approach solves a lasso-type problem with additional constraints in order to enforce exactly the type of sparsity we are interested in here. In their framework, Bien, Taylor and Tibshirani (2013) define strong and weak hierarchy as:

$$\begin{aligned} \text{Strong Hierarchy} : \hat{\Theta}_{jk} \neq 0 \rightarrow \quad & \hat{\beta}_j \neq 0 \text{ and } \hat{\beta}_k \neq 0 \\ \text{Weak Hierarchy} : \hat{\Theta}_{jk} \neq 0 \rightarrow \quad & \hat{\beta}_j \neq 0 \text{ or } \hat{\beta}_k \neq 0 \end{aligned}$$

[10] From a Bayesian perspective, lasso corresponds to a prior that all coefficient are independently distributed Laplace random variables (Park and Casella, 2008).

The authors then demonstrate that these constraints can be enforced by solving the following lasso-type problem (with additional constraints) for each product j:

$$\min_{\beta^+,\beta^-,\Theta} \sum (p_{jt} - s_t^T(\beta_j^+ + \beta_j^-) - \frac{1}{2} s_t^T \Theta s_t)^2 + \lambda 1^T(\beta_j^+ + \beta_j^-) + \frac{\lambda}{2}||\Theta||_1 \tag{8}$$
$$s.t. \quad ||\Theta_j||_1 \leq \beta_j^+ + \beta_j^-$$
$$\beta_j^+ \geq 0, \beta_j^- \geq 0$$
$$\Theta^T = \Theta$$

where $||\Theta||_1 = \sum_{j \neq k} |\Theta_{jk}|$, and where the final constraint imposing symmetry of Θ is only imposed under strong hierarchy. This structure incorporates our prior on the demand system in interesting an intuitive ways. When all inequality constraints are slack, the solution to this problem is identical to the solution to a lasso regression in which all pairs of interactions are included. Otherwise, it shrinks interaction terms relative to linear terms in order to satisfy our desired constraints. This is shown precisely in Bien, Taylor and Tibshirani (2013).

One complication in our setting here is that market shares s_t are endogenous. To minimize the impact of endogeneity on model selection, we can predict s_t using a vector of instruments and apply the hiernet procedure to the predicted values $\hat{s}_t$ from this first stage regression. If the purpose of this step were to obtain consistent estimates of a target parameter, this approach would be similar to the "forbidden regression," as it amounts to using fitted values nonlinearly in a second-stage regression. Given that our intent is instead to determine which products' demand predicts attributes of competing goods, the issue is less clear. As I show in simulation exercises, this model selection step appears to perform well even given the use of the forbidden regression.

I make three more modifications to this procedure in NPDemand.jl. First, I run `hierNet` repeatedly on multiple bootstrapped samples for each product, and calculate the selected set of substitutes by taking the intersection of selected products across bootstraps. As demonstrated by Bach (2008) for the standard lasso, this procedure can help to exclude irrelevant substitutes with minimal risk of excluding relevant substitutes. The user can control the number of bootstrapped samples, and I show in Section 4 that even a few bootstrapped iterations can significantly improve model selection. Second, I always require that each product's market share is selected into its own inverse demand system. It is rare that this restriction is necessary in simulations and in limited empirical testing, but it is an important catch. Third, I allow for users to require that if product j is selected as a substitute for product k, then product k must be a substitute for product j. Note that this restriction is strong in some settings. In a model with inattentive consumers in which one product is considered by all consumers (e.g. an industry leader), the correct demand model may be one in which this product enters the demand for all others but where many competing prices do not affect demand for this leader. However, symmetry of this sort is a natural constraint

which may be relevant in some applications.

3.3 Inference

Inference on causal estimates after a model selection step are has been explored by many authors in recent years (e.g. Chernozhukov et al. (2017)). Many of these studies consider generating a second stage (causal) estimate of a single treatment effect which is unaffected by the first stage selection. I am not aware of any approaches which provide uniform or point-wise inference results for an infinite dimensional parameter after model selection with many instruments and regressors. However, it is intuitive that if the researcher had a second large sample of data available she could select her model using the first sample and estimate the selected model using the second. In such a setting, variation in the data which contributes to incorrect model selection cannot affect estimates of the selected model. This intuition has been operationalized by many researchers in the form of sample splitting, in which the researcher splits their data into two subsets and does exactly this. Athey and Imbens (2016), for example, use one subsample to construct regression trees and another to estimate treatment effects in each leaf. This is the perhaps the easiest way to ensure that model selection errors are independent of post-selection estimation errors, and is the recommended approach here. To close this section I summarize the approach described herein in Algorithm 1.

4 Simulation Evidence

To study the performance of the proposed approach, I begin with a simple data-generating process and aim to demonstrate that this approach consistently selects the relevant substitutes for each product and estimates own-price elasticities without significant bias. I simulate data from many markets in which, for each product j in each market t, consumer i receives utility

$$u_{ijt} = \alpha p_{jt} + \xi_{jt} + \epsilon_{ijt} \tag{9}$$

where $\alpha = -0.4$, $\xi_{jt} \sim N(0, 0.15^2)$. Prices are generated by $p_{jt} = 2(z_{jt} + \eta_{jt}) + \xi_{jt}$ for $\eta \sim U(0, 0.1)$ and $z_{jt} \sim U(0, 1)$. Thus, by construction, z_{jt} are strong instruments for p_{jt} and prices are correlated with ξ_{jt}. This is quite similar to the model Compiani (2021) simulates, though I restrict α to be fixed across consumers. I simulate data for 2000 independent markets in this example.

To test the model selection procedure, I generate two groups of products. In the simplest case, I simulate two groups of products entirely independently, both according to Equation 9, which is equivalent to the fully nested demand model discussed above. The model selection procedure here is product-specific, and the results are all with respect to the first simulated product. This simulation procedure has the simplifying but unrealistic quality that all irrelevant characteristics/products are fully independent of

Algorithm 1: Nonparametric Model Selection and Demand Estimation

I. Split data (s, p) into training and testing subsamples, $(s^{train}, p^{train}, z^{train})$, and $(s^{test}, p^{test}, z^{test})$

II. Select the model using the training subsample. For each j,

   (a) Predict each endogenous regressor (s_t^{train}) using all instruments (z^{train})

   (b) Substitute all endogenous variables for predicted values from preceding step, $(\hat{s}^{train})$

   (c) Estimate a hierarchical lasso regression of p^{train} on a polynomial in $(\hat{s}^{train}, \hat{p}^{train})$ to select model

   (d) Apply (c) iteratively on B_1 bootstrapped samples and take the intersection of selected substitutes for each product across bootstraps.

IV. Estimate the selected model using *testing* data, following Compiani (2021) (superscripts omitted for readability):

   - Searching over $\tilde{\sigma}$, construct $r_{jt}(s_t, p_t, z_t, \tilde{\sigma}^{-1}) \equiv (p_{jt} - \tilde{\sigma}_j^{-1}(s_t) \times a_t(z_t)$ to minimize $\sum_{j=1}^{J} \left[\sum_{t=1}^{T} r_{jt}(s_t, p_t, z_t; \tilde{\sigma}_j^{-1}) \right]' (A'A)^{-} \left[\sum_{t=1}^{T} r_{jt}(s_t, p_t, z_t; \tilde{\sigma}_j^{-1}) \right]$

V. (Inference) Draw B_2 bootstrapped samples from the testing data and re-estimate σ_j on each bootstrap sample to construct confidence intervals, taking the selected model as given.

all relevant ξ, which is unlikely to be the case in oligopolistic settings. I have also considered cases in which some irrelevant characteristics are correlated with unobservables (ξ) of relevant products. These will be included in future drafts.

To demonstrate the value of regularization, I simulate a model with two nests of four products each. To nonparametrically estimate demand for eight products would require more than 6000 parameters for each demand function, which is much larger than the sample size and can be quite difficult to handle computationally. To the contrary, each correctly estimated demand function here (which only depends on four products' market shares) requires only 81 parameters.[11] To see how the proposed procedure performs, in Table 1 I present the frequency with which each product is selected as a substitute over 100 simulations. I show these frequencies first using only a single bootstrapped sample to select the model and then using five samples for this purpose. Perfect selection in this table would be two 4×4 matrices of ones stacked diagonally. Clearly, even a single bootstrap often reduces the dimension of the model significantly. The model selection procedure correctly identifies that the first four products are in one group and the latter four are in another. However, many simulations include one or more irrelevant products (e.g. product 5-8 in σ_1). In the bottom half of the table I increase the number of bootstrap iterations to 5. This drastically reduces the number of irrelevant products included, while continuing to include all relevant products at a very high rate.

I show in Figure 1 the resulting estimates of own-price elasticities from this procedure. In black I plot the true own-price elasticities for one of the products along a fixed grid of prices, and in grey I plot the 90% interval of simulation estimates of own-price elasticities along that grid. I find that the 90% interval covers the truth, as expected, and although I do not report the median here it is also close to the truth. Though this example is simple, what is important to note here is that the procedure to create this graph requires no customization to deal with potentially far more products, nor in response to changes to the underlying model of consumer choice (e.g. the distribution of ϵ). The relevant substitution patterns and all price elasticities can be calculated easily and fully nonparametrically using built-in commands in NPDemand.jl.

4.1 Likely Forms of Misspecification

Although the model herein avoids many common distributional assumptions made in parametric models of demand, there are still many ways in which the model can be misspecified. In this section we can explore the costs of mis-specifying demand as a fully nested model, as well as the performance of the method under a more complicated data-generating process. I now simulate markets of consumers with utility functions

[11] Both counts of the number of parameters necessary assume that second order Bernstein polynomials are used to approximate each demand function.

Table 1: Effect of Increasing Number of Bootstraps on Model Selection

$B=1$	1	2	3	4	5	6	7	8
σ_1	1.0	0.95	0.99	1.0	0.08	0.04	0.02	0.07
σ_2	0.96	1.0	0.99	1.0	0.04	0.06	0.07	0.06
σ_3	0.99	0.98	1.0	0.99	0.04	0.02	0.07	0.01
σ_4	0.98	0.98	1.0	1.0	0.04	0.04	0.01	0.05
σ_5	0.11	0.03	0.05	0.03	1.0	0.97	0.98	0.98
σ_6	0.03	0.02	0.06	0.03	0.99	1.0	0.98	0.98
σ_7	0.06	0.09	0.03	0.05	0.99	1.0	1.0	0.97
σ_8	0.02	0.03	0.06	0.04	0.98	1.0	0.99	1.0
$B=5$								
σ_1	1.0	0.99	0.94	0.94	0.0	0.0	0.01	0.0
σ_2	0.97	1.0	0.93	0.97	0.0	0.0	0.0	0.01
σ_3	0.98	0.99	1.0	0.95	0.0	0.0	0.0	0.0
σ_4	0.98	0.95	0.96	1.0	0.0	0.0	0.0	0.0
σ_5	0.0	0.0	0.0	0.01	1.0	0.97	0.99	0.96
σ_6	0.0	0.0	0.01	0.0	0.95	1.0	0.95	0.96
σ_7	0.0	0.0	0.0	0.0	0.94	0.97	1.0	0.97
σ_8	0.0	0.0	0.0	0.0	0.94	0.95	0.92	1.0

Note: Fraction of simulation runs (out of 100) in which the product (columns) were selected into each product-specific demand function (rows). Shown first using only a single bootstrap sample (top) and then with five samples (bottom).

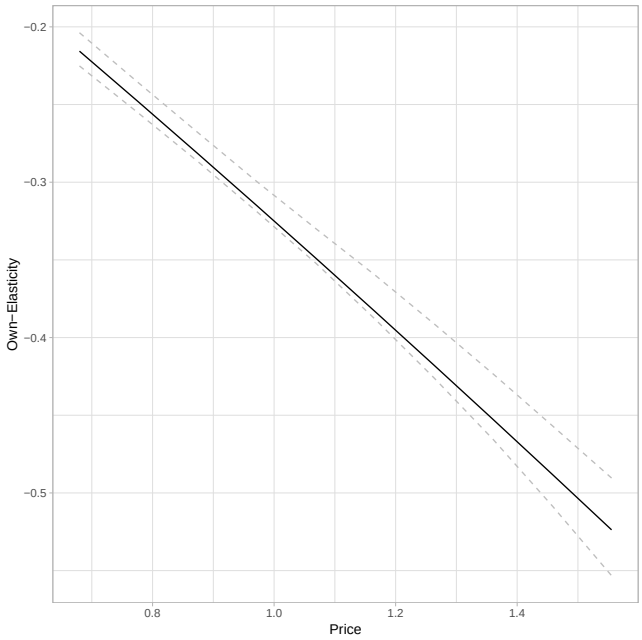
of the form:

$$
\begin{aligned}
u_{ijt} &= \theta_{ij} - \alpha_i p_{jt} + \xi_{jt} + \epsilon_{ijt} \\
\theta_{ij} &\sim N(0, \frac{0.5j}{J}) \\
\alpha_i &\sim N(-1, 0.4)
\end{aligned} \tag{10}
$$

Prices (p_{jt}), market-level shocks (ξ_{jt}), and instruments for prices continue to be simulated as in the preceding section. We have made the utility function significantly more complicated now, as it includes product- and consumer-specific match values which are distributed differently across products as well as heterogenous preferences for prices. This model could be difficult to estimate even parametrically, as it would involve integrating over $J+1$ dimensions of random coefficients. It is also important to note that the fully nested model is mis-specified now, as θ_{ij} violates the required index structure (i.e. because p and ξ have different sets of random coefficients).

The simulations in the preceding section focused on comparing estimated and true price elasticities evaluated at a fixed grid of prices. Another important measure of the

Figure 1: Nonparametric Price Elasticities, Truth and 90% Simulation Range



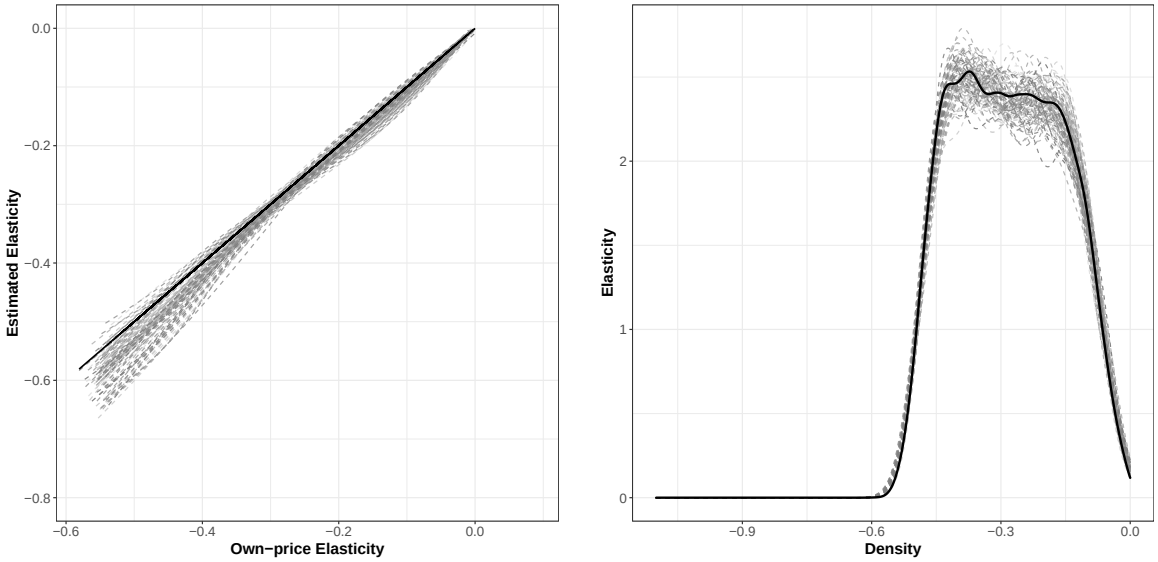
Note: True own-price elasticities, evaluated on a fixed grid of prices, are in black, and dotted lines indicate the 90% interval of estimated own-price elasticities on that grid across simulations.

performance of this method is its ability to match the realized price elasticities in the simulated data. This is the metric on which I focus in this section. For each simulation sample, I calculate $\hat{\mathbb{E}}[\hat{\varepsilon}_{jt}|\varepsilon_{jt}]$, where ε_{jt} denotes the true own-price elasticities of the chosen product, $\hat{\varepsilon}_{jt}$ denotes the nonparametric estimate of that elasticity, and $\hat{\mathbb{E}}[\cdot|\cdot]$ denotes the (conditional) expectation, estimated via a simple local regression (i.e. a kernel smoother).

For my first set of simulations, I generate data for $J = 4$ products in 2000 markets. In Figure 2(a) I plot an estimate of $\hat{\mathbb{E}}[\hat{\varepsilon}_{jt}|\varepsilon_{jt}]$ for 100 simulated samples in grey, and the 45-degree line in black. The closer the grey lines are to the 45-degree line, the better the proposed procedure is at matching the price elasticities in the data. Similarly, in Figure 2(b) I plot kernel density estimates of estimated own-price elasticities (in grey) and the true density function in black.[12] In both figures, the nonparametric procedure appears to work well. There appears to be a small amount of bias when own-price elasticities are relatively large, but Figure 2(b) indicates that these represent a small enough fraction of the data that the estimated distribution of price elasticities closely approximates the truth. At most other points along the horizontal axis, $\hat{\mathbb{E}}[\hat{\varepsilon}_{jt}|\varepsilon_{jt}]$ lies quite close to the 45-degree line. Together, these figures indicate that violation of the index assumption appears to have little impact on the accuracy of nonparametric

[12]I calculate the latter by simulating a single large sample of 20,000 markets and estimating the density of true own-price elasticities within that sample.

Figure 2: Simulation Results: 4 Products



(a) (b)

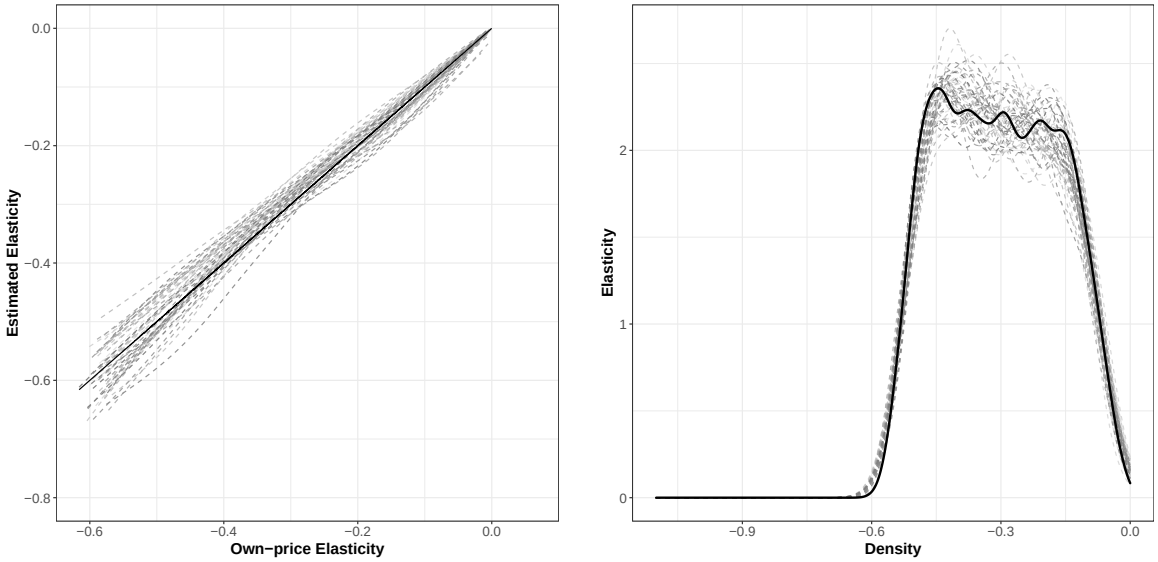
Note: Each figure presents the results of nonparametric demand estimation on 100 simulated samples of 4 products which are simulated according to Equation 10. Figure (a) plots kernel estimates of $\hat{\mathbb{E}}[\hat{\varepsilon}_{jt}|\varepsilon_{jt}]$ and Figure (b) plots kernel estimates of the probability density function of $\hat{\varepsilon}_{jt}$.

estimates.

For the next set of results, for each simulated sample (of 50 total) I generate demand for six nests, each containing three products, following the structure above for 2000 markets. That is, I simulate demand for three products six times and then pool all products together as if they are sold together as competitors. I then apply the hierNet model selection procedure using half of the data set with 5 bootstrapped samples, and then estimate the selected model nonparametrically and calculate price elasticities using the other half of the data. Results of this exercise are shown in Figures 3(a) and (b), which are constructed similarly to Figures 2(a) and (b). The results of this exercise are encouraging. The model selection step, which reduces the demand estimation problem from one which is infeasibly large (one with 18 products) to one which is almost trivial to solve, appears to have made little to no difference on the accuracy of price elasticity estimates on average. It does appear that, relative to Figures 2(a) and (b), that there is more variation in estimates across simulation samples. This is to be expected, as the model selection step often does not select exactly the correct model, meaning that superfluous parameters must be estimated which thereby decreases the precision of the parameters of interest.

Finally, I simulate demand for 20 products in the form of two nests of 10 products each. Because the Bernstein polynomials I use to approximate demand would have an unreasonable number of parameters if 10 products were included, I modify the model

Figure 3: Simulation Results: 6 Nests of 3 Products Each



(a) (b)

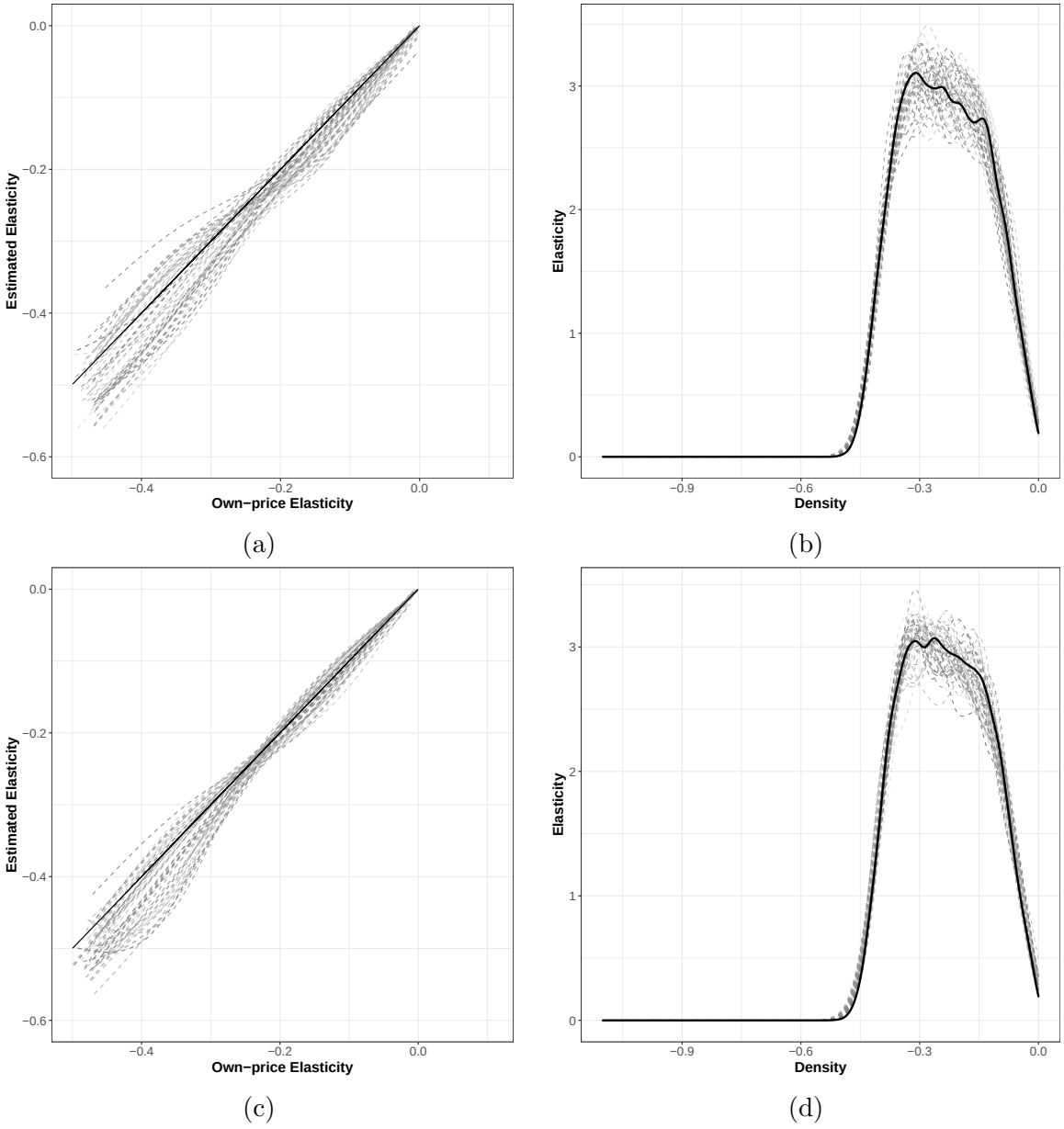
Note: Each figure presents the results of nonparametric demand estimation on 50 simulated samples of 18 products which are simulated according to Equation 10 in 6 nests, each containing 3 products. Figure (a) plots kernel estimates of $\hat{\mathbb{E}}[\hat{\varepsilon}_{jt}|\varepsilon_{jt}]$ and Figure (b) plots kernel estimates of the probability density function of $\hat{\varepsilon}_{jt}$.

selection step in order to constrain the maximum size of estimated nests to be small enough to be quickly estimable. I begin by setting the regularization parameter to a small value, then run the procedure iteratively, increasing the parameter's value by a fixed factor each time. This process is repeated until the number of selected products is no larger than the specified maximum nest size, which in this case I set to either four or two. We now face two forms of misspecification: failure of the index restriction and misspecified nest sizes. The latter may be particularly important in practice, as researchers are rarely able to know the nesting structure of a set of products ex-ante. In Figures 4(a)-(d) I show that this form of misspecification also makes little difference in the accuracy of our estimates. In subfigures (a) and (b), I restrict nests to contain at most four products, and in subfigures (c) and (d) I reduce the maximum nest size to two. Neither of these constraints bias estimates of price elasticities significantly. In fact, because nests of two products require much fewer parameters to estimate, there appears to be noticeably less variation in estimates in Figure 4(c) than in 4(a), even though the former is more misspecified than the latter.

4.2 Timing

One important advantage of nonparametric demand estimation relative to parametric models is the time required to estimate the model. The workhorse BLP model

Figure 4: Simulation Results: 2 Nests of 10 Products Each



Note: Each figure presents the results of nonparametric demand estimation on 50 simulated samples of 20 products which are simulated according to Equation 10 in two nests, each containing 10 products. Figures (a) and (b) present estimates which constrain nests to contain at most four products, and Figures (c) and (d) constrain nests to have two products at most. Figures (a) and (c) plots kernel estimates of $\hat{\mathbb{E}}[\hat{\varepsilon}_{jt}|\varepsilon_{jt}]$ and Figures (b) and (d) plots kernel estimates of the probability density function of $\hat{\epsilon}_{jt}$.

is notoriously difficult to implement in practice, and can take many minutes or hours to estimate once. As researchers may wish to try a number of specifications and/or instruments for endogenous variables, estimating a BLP-style model of demand can be very time consuming. This is even true of state-of-the-art software packages designed to estimate these models, such as *pyblp* (Conlon and Gortmaker, 2019). In contrast, nonparametric demand can be done at the relatively low cost of a series of two-stage least squares regressions. To see the potential benefits on this dimension, in Table 2 I present the estimated time required for model selection and estimation, varying both the number of products and the number of markets of available data. I simulate data in the form of some number of nests N, where demand in each nest is generated according to Equation 10. I show estimates of the time required for N varying from 2 to 10 and with sample sizes ranging from 100 to 1000 markets, running the procedure on 10 bootstrapped samples using 5-fold cross validation to select the tuning parameter. Columns 1 and 2 present the time required for model selection under strong and weak hierarchy, respectively (only the first three rows of Column 1 are included due to the length of time required under strong hierarchy). Even when the data contains 1000 markets with 30 products each, model selection under weak hierarchy takes little more than 10 minutes. Although this is much longer than the times required for fewer products, it is still much quicker than the time necessary for many parametric estimation procedures. Moreover, because estimation remains extremely quick (Column 3), researchers willing to impose a predetermined substitution matrix can estimate demand flexibly at extremely small cost.

Table 2: Time Required for Model Selection and Nonparametric Estimation

Selection Time (s) Strong Hier.	Selection Time (s) Weak Hier.	Estimation Time (s)	N	J	T
166.62	2.7	< 0.01	2	6	100
253.55	5.57	0.02	2	6	500
366.38	11.86	0.01	2	6	1000
	15.85	0.02	5	15	100
	45.55	0.02	5	15	500
	101.83	0.03	5	15	1000
	112.72	0.08	10	30	100
	303.86	0.05	10	30	500
	771.57	0.07	10	30	1000

Note: Approximate number of seconds required for model selection (with 10 bootstrapped samples and 5-fold cross validation) under strong and weak hierarchy and nonparametric demand estimation for varying N, T. N denotes the number of nests, each of which contains three products. T denotes the number of simulated markets. Times are calculated by taking an average over as many times the procedure can be run in 30 seconds. For procedures taking longer than 30 seconds for a single run, a single run-time is reported.

5 Concluding Remarks

In this paper I address the application of straightforward machine learning methods to estimate demand nonparametrically in settings with many products. I show that a simple extension of the standard nested logit model implies that only a small number of products will substitute for each other. As a result, the inverse demand function of interest in Compiani (2021) is a function of the market shares and characteristics of only a small number of products, even when the market includes many options. Though this model is not appropriate for all settings, it greatly reduces the computational burden of estimating the market-level demand function nonparametrically when correctly applied and may be a reasonable approximation even when substantially misspecified. While treating a market with eight products fully nonparametrically can require estimating tens of thousands of parameters, estimating fully nested demand in these settings may require only a few hundred parameters. Because, when the inverse demand function is unconstrained, estimation amounts to a two-stage least squares problem, these few hundred parameters can be estimated extremely quickly.

I have shown that selecting the relevant demand model via a high-dimensional two-stage least squares problem performs well in initial simulations. When products are grouped such that all characteristics of products in one group are independent of another, the procedure correctly identifies relevant substitutes at a very high rate. Nonparametric estimates of price elasticities derived from the selected model closely match true elasticities on average, even when there are many products and the nonparametric model is misspecified in multiple ways. Although there are certainly settings in which the model selection applied here may be misspecified to the point of biasing estimates of elasticities, the speed with which this approach can be applied is encouraging. The time costs shown in Table 2, which scale well with larger samples, are small relative to most workhorse demand models, and the existence of a closed-form solution avoids many issues with non-convergence in more complicated models. Thus, researchers may wish to apply this approach as an early step in their project in order to explore potential substitution patterns in the data (i.e. the substitution patterns estimated by the model selection step) and/or to test the performance of different instruments quickly.

References

Abaluck, Jason, and Abi Adams. 2018. "What Do Consumers Consider Before They Choose? Identification from Asymmetric Demand Responses."

Athey, Susan, and Guido Imbens. 2016. "Recursive partitioning for heterogeneous causal effects." *Proceedings of the National Academy of Sciences*, 113(27): 7353–7360.

Bach, Francis R. 2008. "Bolasso: model consistent lasso estimation through the bootstrap." 33–40, ACM.

Belloni, Alexandre, Daniel Chen, Victor Chernozhukov, and Christian Hansen. 2012. "Sparse models and methods for optimal instruments with an application to eminent domain." *Econometrica*, 80(6): 2369–2429.

Berry, Steven, Amit Gandhi, and Philip Haile. 2013. "Connected substitutes and invertibility of demand." *Econometrica*, 81(5): 2087–2111.

Berry, Steven, James Levinsohn, and Ariel Pakes. 1995. "Automobile Prices in Market Equilibrium." *Econometrica*, 63(4): 841–890.

Berry, Steven T. 1994. "Estimating discrete-choice models of product differentiation." *The RAND Journal of Economics*, 242–262.

Bien, Jacob, Jonathan Taylor, and Robert Tibshirani. 2013. "A lasso for hierarchical interactions." *Annals of statistics*, 41(3): 1111.

Chaves, Daniel. 2019. "Taxation and Product Variety: Evidence from the Brazilian Automobile Industry." Working Paper.

Chen, Xiaohong, and Timothy M Christensen. 2018. "Optimal sup-norm rates and uniform inference on nonlinear functionals of nonparametric IV regression." *Quantitative Economics*, 9(1): 39–84.

Chernozhukov, Victor, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney Newey, and James Robins. 2017. "Double/debiased machine learning for treatment and structural parameters." *The Econometrics Journal*.

Compiani, Giovanni. 2021. "Nonparametric Demand Estimation in Differentiated Products Markets."

Conlon, Christopher, and Jeff Gortmaker. 2019. "Best practices for differentiated products demand estimation with pyblp." Working paper.

Gillen, Benjamin J, Matthew Shum, and Hyungsik Roger Moon. 2014. "Demand estimation with high-dimensional product characteristics." In *Bayesian Model Comparison*. 301–323. Emerald Group Publishing Limited.

Gillen, Benjamin J, Sergio Montero, Hyungsik Roger Moon, and Matthew Shum. 2018. "BLP-LASSO for Aggregate Discrete Choice Models with Rich Covariates."

Grigolon, Laura, and Frank Verboven. 2014. "Nested logit or random coefficients logit? A comparison of alternative discrete choice models of product differentiation." *Review of Economics and Statistics*, 96(5): 916–935.

Mehta, Nitin, Surendra Rajiv, and Kannan Srinivasan. 2003. "Price uncertainty and consumer search: A structural model of consideration set formation." *Marketing science*, 22(1): 58–84.

Miller, Keaton, and Boyoung Seo. 2018. "Tax revenues when substances substitute: Marijuana, alcohol, and tobacco." *Kelley School of Business Research Paper,(18-27)*.

Miravete, Eugenio J, Katja Seim, and Jeff Thurk. 2018. "Market power and the Laffer curve." *Econometrica*, 86(5): 1651–1687.

Mrázová, Monika, and J Peter Neary. 2017. "Not so demanding: Demand structure and firm behavior." *American Economic Review*, 107(12): 3835–74.

Nevo, Aviv. 2001. "Measuring market power in the ready-to-eat cereal industry." *Econometrica*, 69(2): 307–342.

Newey, Whitney K, and James L Powell. 2003. "Instrumental variable estimation of nonparametric models." *Econometrica*, 71(5): 1565–1578.

Park, Trevor, and George Casella. 2008. "The bayesian lasso." *Journal of the American Statistical Association*, 103(482): 681–686.

Tibshirani, Robert. 1996. "Regression shrinkage and selection via the lasso." *Journal of the Royal Statistical Society. Series B (Methodological)*, 267–288.

Train, Kenneth, and Garrett Sonnier. 2005. "Mixed logit with bounded distributions of correlated partworths." In *Applications of simulation methods in environmental and resource economics*. 117–134. Springer.

Train, Kenneth E. 2009. *Discrete choice methods with simulation*. Cambridge university press.

Tversky, Amos. 1972. "Elimination by aspects: A theory of choice." *Psychological review*, 79(4): 281.

A Code Examples

A.1 Minimal Example in Julia

```
using Statistics, NPDemand
using RCall, DataFrames
@rlibrary ggplot2

J = 2; # number of products
T =2000; # number of markets
beta = -0.4; # price coefficient
sdxi = 0.15; # standard deviation of xi

# Returns market shares, prices, instruments,
# and the market demand shock, respectively
s, p, z  = simulate_logit(J, T, beta, sdxi);
df = toDataFrame(s,p,z);

# Estimate demand nonparametrically
inv_sigma, designs = inverse_demand(df);

# Calculate price elasticities at realized prices and market shares
elast, jacobians = price_elasticity(inv_sigma, df, p);
# equation for own-price elasticities in logit model
true_elast = beta.*p.*(1 .- s[:,1])

# Plot kernel densities of estimated and true own-price elasticities
df2 = DataFrame(Estimate = elast, True = true_elast[:,1])
ggplot(df2, aes(x=:True))+
 geom_density(aes(x=:Estimate), color = "gray", linetype = "dashed") +
 geom_density(aes(x=:True), color = "black") + xlab("Elasticity") +
 ylab("Density")
```

A.2 Minimal Example in R

Below is an example of how to use `JuliaConnectR` to call `NPDemand.jl` from R. In this example, I assume that a data.frame called r_data exists in R containing columns as required by `NPDemand` package.

```
install.packages("JuliaConnectoR")
library("JuliaConnectoR")
juliaEval('using Pkg')
juliaEval('Pkg.add(PackageSpec(url=
                "https://github.com/jamesbrandecon/NPDemand.jl"))')
juliaEval('Pkg.add("DataFrames")')

NPDemand <- juliaImport("NPDemand")
juliaEval("using DataFrames")
jl_data <- juliaLet('DataFrames.DataFrame(data);', data = r_data)
results <- NPDemand$inverse_demand(jl_data)
```