האוניברסיטה העברית בירושלים
THE HEBREW UNIVERSITY OF JERUSALEM

Estimating Flexible Specifications of the Nested Logit Model

Presented by

Yotam Nir

Supervised by

Prof. Alon Eizenberg

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Arts in Economics

in the

Hebrew University of Jerusalem

Faculty of Social Sciences

Bogen Family Department of Economics

December 2022

Estimating Flexible Specifications of the Nested Logit Model

Yotam Nir*

December 2022

Abstract. This thesis is interested in the estimation of a nested logit demand system in which the functional form of consumers' indirect utility is either not assumed a priori by the researcher, or is of a high degree of complexity, such that standard parametric specifications or estimation methods are not appropriate. I summarize the restrictions on functional form that are imposed by economic theory and by the nested logit model, and propose a two-step approach to estimating the class of functional forms that is consistent with these restrictions. A Monte Carlo analysis demonstrates that despite its slower convergence rate, this approach can potentially provide significant improvements on the bias of a misspecified linear model, as well as providing guidance in the selection of a parametric functional form, even given relatively small datasets. An application to the domestic US airline market returns own price elasticity estimates that are consistent with the literature and cross price elasticity estimates that are around the upper tail of the results obtained in the literature, and reveals certain informative relationships between product characteristics in the functional form of their effects on consumer utility. The empirical application also reveals a possible omitted variable bias that is not apparent in the linear specification.

1. Introduction

Empirical discrete choice models of product differentiation uncover substitution patterns between products by representing them as bundles of characteristics that influence consumer utility. However, clear theoretical guidance on the manner in which these characteristics affect utility is often lacking, which may result in misspecification bias. Moreover, as both products and marketing techniques grow in sophistication, the decision processes of consumers are growing in dimensionality and in complexity. At the same time, datasets are growing both larger and more diverse in structure over time, spurring advancements in statistical analyses of large datasets. In recent years, these advancements have begun to be harnessed by the causal inference literature. This paper leverages these contributions to the literature to explore the possibilities and limitations of relaxing functional form assumptions in the empirical estimation of nested

* I would like to express my sincere gratitude to Professor Alon Eizenberg, the advisor of this thesis, for countless hours of advice, insights, and availability, and for a genuine interest in my learning process. The dataset used for the empirical application was obtained from an ongoing research project led by Professor Alon Eizenberg and funded by the Israeli Science Foundation (ISF) grant No. 2512/21 – "Mergers and market power: expanding the econometrician's toolbox".

logit demand models, especially when the researcher is interested in the effects of product features beyond the price.

I begin by characterizing a general set of functional forms of consumers' indirect utility that are consistent with both consumer theory in general and the nested logit model in particular. The main limitations are on the sensitivity to price, which must be invariant across products and be either linear or logarithmic (note that these limitations are not unique to the present framework, but rather apply to nested logit models in general). Next, I suggest a two-step estimation approach to uncovering the functional form of the utility. In the first step, the linear parameters of the model (the price sensitivity and the nesting parameter) are estimated in a partially linear instrumental variables (IV) regression model, flexibly controlling for the rest of the covariates as in (1.1),

$$Y = \theta_0 T + g_0(X) + U, \tag{1.1}$$

where Y is the outcome, T is a vector of variables that enter the regression linearly, X is a vector of covariates, and U is an unobserved residual. In particular, the outcome here is the usual nested logit dependent variable: the difference between the log of the product's market share and the log of the outside option share. In the model, this quantity is shown to be equal to the mean utility from purchasing the good. Thus, θ_0 and the gradients of g_0 with respect to the covariates reflect the sensitivity of mean consumer utility to changes in T and in X, respectively. The estimated linear components are then subtracted from the left-hand side of the equation, and in the second step, the functional form of the remaining product features with respect to this modified dependent variable is estimated (i.e., we estimate the functional form of the first step's nuisance function):

$$Y - \hat{\theta}_0 T = g_0(X) + U. \tag{1.2}$$

Depending on the richness of the available data and on the goals of the researcher, the estimation of this functional form can range from a fully non-parametric estimation of marginal treatment effects (of the covariate on mean utility) to a partially linear estimation of an "average treatment effect" of a product feature of interest.

For each of these steps, I propose appropriate estimation techniques from the causal machine learning literature. The motivation for using these techniques is their efficient scalability to high-dimensional and highly complex functional forms relative to traditional non- and semi-parametric methods, in terms of both accuracy and computational intensity. This advantage is especially pronounced in IV settings, and all of the techniques employed here accommodate the use of IVs.[1] The scalability of these methods can even extend to complexity

[1] In the first step, this is necessary due to the endogeneity of the price and within-nest share of products. In the second step, this allows analyses of additional endogenous covariates.

levels that violate the assumptions of traditional econometric methods. One illustration of the potential usefulness of this is that, given appropriate data, it is in principle possible to include non-standard data forms – such as images, recordings, and natural language text – in the feature space being controlled for. For instance, objects such as social media discussions of a good, which will generally not be structured in a manner that is immediately translatable into a satisfaction index, can be included in the empirical model, and can explain consumer choice in a flexible manner that is unique to the particular market, beyond what might be obtained from a pre-applied one-dimensional scoring of the reviews (for example, some keywords might correlate with demand in particular markets differently or more strongly than in others).

For the first step, I use the double/debiased machine learning (DML) procedure of Chernozhukov et al. (2018). In the second step, given sufficiently rich data and a continuous covariate of interest (for simplicity, let us call it the treatment), a neural network regressor (or, given an endogenous treatment, the Deep IV procedure of Hartford et al., 2017) can be used to approximate the full functional form of the remaining product features. This allows the researcher to learn consumers' willingness to pay as a continuous function of a product characteristic. Such a function can serve various analyses that are of core interest to industrial organization economists (see, for example, Sheshinski, 1976, Spence, 1975, and Spence, 1976), such as comparing actual investment in quality to its socially optimal level, or providing product selection and market entry predictions under counterfactual equilibrium assumptions (for example, given tax changes that affect product prices).

A second option – which poses lower computational and data intensity requirements – is the generalized random forest (GRF) of Athey, Tibshirani and Wager (2019). GRF provides a conditional average treatment effect (CATE) in each region of the covariate space of all features (excluding the feature selected as the treatment). Thus, complex interactions between the treatment and the covariates, and between the covariates themselves, can be accounted for, but the estimate obtained in this case is slightly less general in that the treatment effect in each region of the covariate space is restricted to be linear (i.e., we obtain only average and not marginal treatment effects). A CATE estimate can be useful, for instance, when deciding on which type of product to embellish with a certain feature (for example, would it be more worthwhile to increase the horsepower of a convertible or of a sedan?), or when predicting which market segments are likely to be most affected by a proposed policy. Finally, if one is simply interested in an unconditional average treatment effect (for example, if the researcher is interested only in an average willingness to pay for a product feature, or just in the average sign of this willingness to pay), DML can be employed again in the second stage, flexibly controlling for the rest of the covariates while imposing linearity on the treatment effect. Notice that in this last case, if the true treatment effect is believed to be constant, then there is no need for two steps – linearity can be imposed on the treatment in the first step and a strictly more accurate estimate (in expectation) would be obtained. However, if the treatment effect is not truly constant, then imposing linearity in the first step generates greater misspecification bias and may result in a poorer approximation of the average effect than the two-step estimation.

The main shortcoming of the two-step approach is that it introduces a larger degree of inefficiency to the second step estimators due to the presence of estimated coefficients in the dependent variable. While the first step DML estimators are $\sqrt{N}$ consistent under a broad range of conditions, the second step estimators are not guaranteed to achieve this convergence rate (although they will be consistent). Deriving formal results on the properties of the second step convergence rates, although of interest, is beyond the scope of this thesis. I resort instead to simulation analysis to demonstrate that convergence, at least in the settings considered, is fast enough for the procedure to be of value as an improvement on linear models. A more complete understanding of the stability of this approach in other settings, without formal results on the estimators' statistical properties, would require a more extensive simulation analysis than that provided here.

I conduct a Monte Carlo analysis to examine the performance of the two-step approach. It is demonstrated that a misspecification of the functional form of the mean utility, even when this functional form is quite simple, may result in significant bias. In particular, in the simulation performed here, a linear utility index (as is used in most studies) heavily overestimates the nesting parameter, while heavily biasing the sensitivity to price toward zero, resulting in highly inaccurate own and cross price elasticities both within and across nests. Most importantly, it is shown that despite the slow convergence rate of the two-step approach, it significantly outperforms the misspecified linear model both in estimating the linear parameters (thus providing more accurate elasticity estimates) and in estimating the effects of the rest of the covariates. Moreover, it is shown that even with a small sample size, the second step estimation can provide a good indication as to the general form of the product features' effects on demand. Therefore, even when a linear methodology is ultimately preferred, the two-step approach can serve as a valuable exploratory exercise to identify important interactions between features or curvatures in the features' effects, and thus mitigate misspecification bias.

Following the simulation analysis, I apply the two-step approach to an empirical dataset of the domestic US airline market. The estimation returns own price elasticities that are consistent with those found in the literature, and cross price elasticities (between inside options) that are at the higher end of results found in the literature, due to a relatively low estimated sensitivity to price and a relatively high estimated importance of the nest structure. The second step finds that consumers are less willing to pay for trips involving long detours when the origin and destination airport are closer to each other and when flying with a small low cost carrier. However, a large subset of flights are characterized by a positive willingness to pay for long detours, hinting at the possible existence of important omitted variables, for which a flexible functional form cannot compensate on its own. Finally, despite the relatively large size of the dataset, the neural network regressor fails to pick up on the differing effect sizes across market distances and airline identities (serving as an important reminder that such estimates must be interpreted with care), and the marginal effect estimates are statistically insignificant.

This paper contributes to the empirical demand estimation literature from two primary angles. First, as already discussed, it contributes to the estimation of logit models in its focus on improving the approximation of functional form, whether the flexible methodology is used as the primary estimation technique or as an exploratory step. As an exploratory step, it can even serve the random coefficients logit estimation framework of Berry, Levinsohn, and Pakes (1995; henceforth BLP). Given the sensitivity of the estimates to the choice of functional form, this focus may be of considerable importance to papers that aim to conduct accurate counterfactual equilibrium analyses. One project that shares a slightly similar goal is the BLP-2LASSO model of Gillen et al. (2019), which employs the double-LASSO procedure of Belloni, Chernozhukov and Hansen (2014) for data-driven covariate selection within the BLP model. While the approach of the current framework, too, facilitates data-driven covariate selection (although it does not accommodate demographic data, but rather relies on product features only), it also adds an emphasis on generalizing the functional form of the covariates' effects.

Second, this paper contributes to a newer approach that aims to leverage the predictive abilities of machine learning techniques to estimate demand. The contribution to this approach is the addition of a theoretically sound structure that both reduces the burden on the data and enables more extensive out-of-sample extrapolation. For example, a relatively early work in this strand of literature by Bajari et al. (2015) proposes to circumvent the endogeneity involved in demand estimation by relying on data that is sufficiently extensive to mitigate omitted variable bias. This approach may be successful if the quality of the available data is such that the estimated residuals are of second-order importance. However, this is a very strong requirement that is unlikely to be achieved in most settings.

Another notable example is the motivating example for the Deep IV procedure. This work takes a significant step forward by directly confronting the endogeneity of product prices and accommodating instruments, thus allowing to trace out the market demand curve. However, in the absence of a structural framework, substitution patterns across different products in the market cannot be inferred and, consequently, counterfactual equilibrium analysis cannot be conducted. Therefore, the approach taken here of incorporating machine learning techniques into a well-developed structural model can enrich this strand of demand estimation, to the benefit of both academics and industry practitioners.

The remainder of the paper is structured as follows. Section 2 outlines the theoretical model, the restrictions it must satisfy, and the resulting empirical model. Section 3 presents the estimation procedure and the specific methodologies used in both steps. Section 4 describes the simulation setup considered in this paper and discusses the simulation results. Section 5 covers the empirical application to the domestic US airline market. Finally, section 6 concludes.

2. The Model

The nested logit model is a highly popular model in the demand estimation literature due the simplicity of its estimation relative to the random coefficients logit model and due to the ability of its nest structure to explain some degree of heterogeneity in consumer preferences. In this section, I begin by presenting the nested logit model in a slightly more general form than usual, relaxing functional form assumptions as permitted. I then explain the restrictions derived from consumer and discrete choice theory and from empirical constraints that prevent further generalization of the model, before demonstrating that an empirical model can be derived as done by Berry (1994) for linear models.

2.1. Indirect utility of the consumer

All of the proceeding analysis is independent of the assumed nest structure, but for expositional simplicity, let us assume a one-level nested logit model. Let T denote the number of markets, $t = 1, \dots, T$, each consisting of $\mathcal{J}_t$ differentiated products, $j = 1, \dots, \mathcal{J}_t$. The products are split into mutually exclusive nests, $g = 0, \dots, G$, based on observable features as determined by the researcher. Denote by d_{jgt} an indicator variable that equals one when product j in market t belongs to nest g, and zero otherwise. The indirect utility of consumer i from product j in market t is of the form

$$v_{ijt} = f(x_{jt}) - \alpha h(y_i, p_{jt}) + \xi_{jt} + \sum_g d_{jgt}\zeta_{ig} + (1-\rho)\epsilon_{ijt}, \quad h(y,p) = \begin{cases} y - p \\ \gamma^{-1} \ln y - \ln p \end{cases} \tag{2.1}$$

where f is a function of observable product features x_{jt}, the price p_{jt} and income y_i enter either linearly or logarithmically, ξ_{jt} is a product-specific unobservable utility shifter, and ϵ_{ijt} is an individual-specific taste parameter for the product that is identically and independently distributed Type-I Extreme Value. The random variable ζ_{ig} is an individual-specific taste parameter that is common to all products in nest g, and has the unique distribution such that $\zeta + (1-\rho)\epsilon$ is also distributed Type-I Extreme Value (Cardell, 1997). The nesting parameter, $\rho \in [0,1)$, determines the importance of the nest structure for consumer choice. Each consumer chooses the product that maximizes their utility. For brevity, I henceforth omit the market subscript.

A first important restriction on (2.1) is that f and α do not vary across consumers. While individually unique consumer preferences are of course theoretically sound and desirable (and, in the case of f, even admit a representative agent analysis), an empirical estimation of such heterogeneity would require prior assumptions on the distributions of f and α, and the estimation procedure would need to accommodate parameters of these distributions in a manner that is not consistent with the empirical framework of the nested logit model. Such assumptions are indeed incorporated into the framework of BLP, but this framework relies on a linear specification of the mean and individual utilities. The feasibility of extending the BLP framework to a more general specification of indirect utility is an interesting avenue for future research, but is likely to run into challenges in terms of both statistical efficiency and computational intensity. Therefore, the present work incorporates consumer heterogeneity only through the nest structure, as in the standard nested logit model.

Another restriction relates to the generality with which the mean utility may be specified. In the bulk of the literature, $f(x_j)$ is defined as specific linear combinations of product features. The main contribution of this work is the demonstration of a practical empirical approach that permits the relaxation of this assumption.[2] However, the disposable income and the price in particular are treated separately from the rest of the product features – they are restricted, as usual, to entering the indirect utility either linearly or logarithmically, and the price and income are kept additively separable in both cases. These are requirements imposed by Roy's identity in order to preserve the regular assumptions of the model regarding the form of consumers' Marshallian demand, which determines the relationship between the indirect utilities and market demand. The usual assumption of unit demand for the consumer's product of choice corresponds to the linear specification. The logarithmic specification is a recent innovation introduced by Bjornerstedt and Verboven (2016), which results in consumers spending a constant share of their income on their product of choice. Due to the centrality of the unit demand assumption in the literature, I henceforth consider only the case of $h(y,p) = y - p$, but all derivations in this work can also be applied naturally to the constant expenditures formulation.

A final restriction in equation (2.1) is that the sensitivity to price α is constant not only across consumers but also across products. It is empirically feasible to relax this assumption and estimate a product-specific (or, more precisely, a product feature-specific) sensitivity to price $\alpha(x_j)$. However, such a form of heterogeneity in price sensitivity is theoretically difficult to motivate. Qualitatively, it means that not only do certain product features change consumers' utility from a product, but they also directly affect consumers' sensitivity to the price itself. Therefore, I limit attention to the standard case of a constant α across products. An additional advantage of this is that it keeps this work within the literature tying random utility discrete choice models to a representative consumer framework.[3] This is because product-varying price sensitivities, in the absence of income effects, amount to the representative consumer's Slutzky

[2] As discussed in the introduction, this relaxation enables not only agnosticism with respect to a relatively simple model, but also an extension to functions of a higher complexity than is permitted in traditional models and to nonstandard data structures.
[3] For a discussion of this topic, see, for example, chapter 3 of Anderson et al. (1992) and Dubé et al. (2022).

matrix being asymmetric, as noted by Anderson et al. (1992), thus removing the guarantee that the aggregate demand system can be represented by a rationalizable utility.

2.2. Aggregate demand and the estimation model

We may now proceed exactly as in Berry (1994). I present the unit demand case here, and refer the reader to Bjornerstedt and Verboven (2016) for an analogous derivation of the constant expenditures demand. Define $\delta_j \equiv f(x_j) + \alpha p_j + \xi_j$. Note that the income is allowed to drop out of the equation. Given the above-specified distributions of the utility errors, the probability π_j that a consumer chooses product j of nest g, is equal to that product's market share, s_j, and is given by

$$s_j = \pi_j(\delta_j) = \frac{\exp\left(\frac{\delta_j}{1-\rho}\right)}{D_g} \cdot \frac{D_g^{1-\rho}}{\sum_g\left(D_g^{1-\rho}\right)} = \frac{\exp\left(\frac{\delta_j}{1-\rho}\right)}{D_g^{\rho}\sum_g\left(D_g^{1-\rho}\right)}, \tag{2.2}$$

where $D_g \equiv \sum_{j \in \mathcal{J}_g} \exp\left(\frac{\delta_j}{1-\rho}\right)$, and $\mathcal{J}_g$ is the set of products in group g. Normalizing the mean utility from the outside option to zero (i.e., $\delta_0 \equiv 0$), we get from (2.2) that $s_0 = \frac{1}{\sum_g\left(D_g^{1-\rho}\right)}$ (because $D_0 = 1$), and therefore

$$\ln s_j - \ln s_0 = \frac{\delta_j}{1-\rho} - \rho \ln D_g = \frac{\delta_j}{1-\rho} - \frac{\rho(\ln s_g - \ln s_0)}{1-\rho}, \tag{2.3}$$

where s_0 is the share of the outside option, s_g is the share of nest g, and using $s_g = \frac{D_g^{1-\rho}}{\sum_g\left(D_g^{1-\rho}\right)}$ for the second equality. Rearranging gives us the familiar form

$$\ln s_j - \ln s_0 = \delta_j + \rho \ln s_{j/g} = f(x_j) + \alpha p_j + \rho \ln s_{j/g} + \xi_j, \tag{2.4}$$

where $s_{j/g} = s_j/s_g$ is the within-nest share of product j. Equation (2.4) (and its constant expenditures analogue) is the empirical object of interest in this paper. The only difference between (2.4) and the specification in Berry (1994) is that the linear index $x_j\beta$ is replaced by the more general $f(x_j)$.

3. Estimation Procedure

3.1. First step

This section details the proposed two-step approach to estimating the flexible nested logit model of equation (2.4). The first step returns estimates of the linear parameters α and ρ, while controlling for $f(x)$ as a nuisance function, using DML.[4] Let $(S, S_0, S_G) \in \mathbb{R}^{\sum_{t=1}^{T}|\mathcal{J}_t| \times 3}$ be the vectors across all products $\mathcal{J}_t$ in all T markets of product shares, outside option shares, and within-nest shares, respectively. Similarly, let $\Xi \in \mathbb{R}^{\sum_{t=1}^{T}|\mathcal{J}_t|}$ be a vector of utility errors. Finally, let Z be a vector of instruments that are correlated with the prices and within-nest shares but not with the utility errors. The empirical model is given by (3.1) and (3.2):

$$\ln S - \ln S_0 = \alpha P + \rho \ln S_G + f(X) + \Xi, \quad \mathrm{E}[\Xi|\mathrm{X}, \mathrm{Z}] = 0 \tag{3.1}$$

$$Z = m(X) + V, \quad E[V|X] = 0.^{5} \tag{3.2}$$

To economize in notation, define $Y \equiv \ln S - \ln S_0$, $T \equiv (P, S_G)$, and $\theta \equiv (\alpha, \rho)$. From (3.1) and (3.2), we may derive the moment condition

$$E\big[\big(Y - E[Y|X] - \theta(T - E[T|X])\big)(Z - E[Z|X])\big] = 0. \tag{3.3}$$

Notice that the outcome, the treatments, and the instruments are orthogonalized with respect to the other covariates. In practice, the estimator $\hat{\theta}$ is obtained as the solution to the linear equation

$$Y - \widehat{E[Y|X]} = \theta\big(\widehat{E[T|X,Z]} - \widehat{E[T|X]}\big) + \Xi.^{6} \tag{3.4}$$

Chernozhukov et al. (2018) show that the estimator $\hat{\theta}$ is $\sqrt{N}$ consistent and asymptotically Gaussian under relatively weak assumptions. In particular, a broad range of machine learning algorithms can be used to estimate the conditional expectation functions, including regularized regressions, tree-based ensembles, and neural networks, under the assumption that their predictions converge at a $N^{1/4}$ rate. In both the Monte Carlo and empirical analyses, I orthogonalize all of the variables using extreme gradient boosting regressors.[7]

This first step is of interest in its own right. While DML has been applied empirically in other settings, the bulk of the logit demand estimation literature employs a linear framework,

---

[4] The DML estimations, as well as the second step GRF estimations, are performed here using the implementation in the *EconML* Python package of Battocchi et al. (2019). The same package also provides an implementation of Deep IV.

[5] Note that this presentation involves a slight abuse of notation: in section 2, $f(\cdot)$ was defined over a vector of characteristics of a single product, while here it is being applied to a matrix of all products' characteristics.

[6] In the non-IV case, the DML estimator is simply estimated by the residual on residual OLS regression $Y - \widehat{E[Y|X]} = \theta\big(T - \widehat{E[T|X]}\big) + \epsilon$.

[7] The number of estimators, maximum depth, and learning rate are tuned separately for each particular case.

and to the best of my knowledge no previous paper has proposed to leverage techniques from the causal machine learning literature to this end. Moreover, traditional non-parametric IV estimation requires assumptions on the data generating process (reducing the value of the agnosticism regarding functional form), and in any case rapidly grows in complexity as the number of covariates increases (Hartford et al., 2017), making DML particularly valuable to this project.

Flexibly controlling for covariates may be especially important for logit demand estimation, given the central role of the estimated coefficients $\hat{\alpha}$ and $\hat{\rho}$ in estimating substitution patterns that are often used to perform counterfactual equilibrium analysis. These two coefficients are the most important components in both own and cross price elasticity estimates (the precise formulas are presented in Appendix C, for convenience). The sensitivity of the coefficients $\hat{\alpha}$ and $\hat{\rho}$ to functional form, combined with the absence of theoretical guidance in the choice of a particular functional form, may produce highly inaccurate counterfactual extrapolations. A second advantage, this time of DML in particular, is that it extends the range of estimable nuisance functions. Not only does it allow much higher dimensionality, but it even permits the inclusion of non-standard data structures such as images, recordings, and natural language text in the covariate space.

3.2. Second step

Having estimated $\hat{\alpha}$ and $\hat{\rho}$ in the first step, the second step is interested in the estimation of $f(x)$. This is done by subtracting the first step estimates of the linear components from the outcome:

$$\ln s_j - \ln s_0 - \hat{\alpha} p_j - \hat{\rho} \ln s_{j/g} = f(x_j) + \xi_j. \tag{3.5}$$

Depending on the quality of the available data and on the interests of the researcher, several approaches are possible to estimating (3.4). To fix ideas, I classify these approaches into three groups: non-parametric marginal effects, conditional average treatment effects (CATEs), and unconditional average treatment effects (ATEs). I suggest estimation methods suitable to each approach that are used in the simulations and in the empirical analysis of this work. These methods are appealing in their efficient scalability and in their ability to accommodate IVs (in the event that some of the second step covariates are treated as endogenous). However, they are not at all exclusive options, and other techniques may be applied as appropriate.

Finally, it is important to note that I do not, at this stage, have analytical derivations of confidence intervals for the second step estimates.[8] These are therefore derived here using bootstrapping, with both steps being estimated in each bootstrap replication. Given the already

[8] Some second step estimators may fall under the conditions of Theorem 6.1 in Newey and McFadden (1994), but time constraints prevent me from pursuing a formal characterization. Others, such as neural network regressors, currently lack guarantees of $\sqrt{N}$ consistency even in a single step.

slower convergence rate of the two-step approach, this further increases the importance of a large sample size.

Estimation of $f(x)$

The most general approach, which is naturally also the most demanding in data quality and in computational resources, is a fully non-parametric estimation of $f(x)$. Given the focus on a potentially high-dimensional $f(x)$, I propose to do this using a neural network regressor. Given an endogenous covariate of interest, its effect can also be mapped using Deep IV, which (in a manner analogous to the intuition behind two-stage least squares) first estimates the conditional density of the covariate and then uses this density to estimate the outcome in two separate neural networks.[9] As mentioned in the introduction, tracing out the full form of $f(x)$ may serve various important interests of industrial organization economists. While researchers usually measure a constant monetary willingness to pay β_j/α, this approach allows us to study a continuous function $\frac{\partial f(x_j)}{\partial x_j}/\alpha_j$, and thus obtain more precise answers to economic questions, such as those involving predictions of market entry and exit in counterfactual equilibria. This willingness to pay function may also be of interest to policymakers who wish to incentivize or disincentivize certain product features, such as fuel efficiency in cars.

While more computationally demanding than the methods discussed below, neural network regressors are an improvement in this respect on kernel regressions. In intuitive terms, although neural networks are capable of approximating any functional form (Hornik et al., 1989), they are not truly non-parametric, but rather are constructed from a large but finite number of parameters. However, due to their relative lack of structure, neural network estimates still converge more slowly than GRF or DML. Moreover, little work has so far been done toward obtaining analytical properties of confidence intervals for neural networks estimates, meaning that closed-form two-step standard errors are an even more distant goal in this case. Recent progress has been made on this front by Farrell, Liang, and Misra (2021), who derive analytical properties of high probability bounds for estimates of feedforward neural networks (the class of neural networks in which information flows in one direction only – from the input toward the output).

Regardless of the slower convergence rates, marginal utility effects will often not be the object of interest in the first place. An average utility effect conditional on the covariates, $\tau(x)$, will often be sufficient and easier to summarize. Indeed, when the covariate of interest is binary, the marginal effect on utility is in any case equivalent to the CATE, making a fully non-parametric specification needlessly and inefficiently general. Here, too, we may benefit from

[9] To prevent confusion, it should be clarified that both these stages would be performed within the second step estimation.

obtaining a more granular willingness to pay statistic – for example, if policymakers are considering taxing or subsidizing a particular product feature only in a certain subset of the market, or would like to know on which subsets of products the policy is most likely to be effective.

In this paper, I estimate CATEs using GRF, which has several attractive features. First, it returns estimates that are shown analytically to be consistent and asymptotically normal, and second, the regression trees provide a data-driven partition of the covariate space that aims to maximize treatment effect heterogeneity between partitions. This adaptive partitioning makes the estimates statistically efficient and the estimation computationally efficient. These efficiencies become especially pronounced as the number of covariates grows, in comparison to other approaches such as k-nearest neighbors or kernel regression.[10] Third, GRF can be extended to accommodate IVs in the estimation.

Finally, the two-step approach can be of value to the estimation of an unconditional average effect of the covariate of interest on consumer utility. If one is interested only in the sign of a feature's effect on utility, or in summarizing a variable effect, then this is adequate. It should be noted that if the effect is truly constant, then it is sufficient to include this effect in the set of linear parameters of the first step DML, and thus obtain a strictly more accurate estimate (in expectation). But if the treatment effect is not truly linear, then imposing linearity on it in a single step may result in an inconsistent estimate of the true ATE. In the analyses below, I estimate the ATE using DML again for the second step.

4. Simulation Analysis

4.1. Simulation design

For tractability, and in order to demonstrate the relevance of a flexible functional form even in standard settings, I consider a data generating process in which only three product features, apart from the price, enter utility.[11] These features enter non-linearly and are interacted with each other as described below. The simulation setup loosely relies on that of Grigolon and Verboven (2014) as a benchmark, with appropriate additions and changes made to address the problem at hand.

For each simulation, I generate 1,000 datasets. Each dataset consists of T markets (the exact number is varied across setups in order to examine the effect of the sample size), with $J =$

[10] See Wager and Athey (2018) for a simulation comparison of causal forests to k-nearest neighbors.
[11] The advantages of machine learning techniques in high dimensional settings are discussed at length in the existing literature, including Chernozhukov et al. (2018) and Wager and Athey (2018), and so constructing a more complex setup here would not contribute valuable insights.

25 products per market. Each product j is produced by a different firm, and is defined by the vector $x_j = (1, p_j, d_j, x_j^1, x_j^2, \xi_j, z_j)$: a constant, a price variable, a binary group variable, two continuous characteristics, an unobserved utility error, and a simulated instrument. The price is drawn from a log-normal distribution in order to mimic its distribution in empirical datasets. The two continuous variables, x^1 and x^2, are both drawn from a uniform distribution on $[0.1, 5]$.[12] More precisely, let d^*, $(x^1)^*$, and $(x^2)^*$ be latent variables such that

$$\begin{pmatrix} \ln p \\ d^* \\ (x^1)^* \\ (x^2)^* \\ \xi \\ z \end{pmatrix} \sim N \begin{pmatrix} 0 & 1 & \sigma & \sigma & \sigma & \sigma & \sigma \\ 0 & \sigma & 1 & 0 & 0 & 0 & 0 \\ 0 & \sigma & 0 & 1 & 0 & 0 & 0 \\ 0' & \sigma & 0 & 0 & 1 & 0 & 0 \\ 0 & \sigma & 0 & 0 & 0 & 1 & 0 \\ 0 & \sigma & 0 & 0 & 0 & 0 & 1 \end{pmatrix}. \tag{4.1}$$

Define $d = 1\{d^* > 0\}$ to obtain a binary variable that is correlated with the price. For the continuous variables, a uniform distribution that is correlated with the price is constructed by evenly positioning $(x^i)^*$, for $i = 1,2$, on the interval $[0.1, 5]$ in order of size, thus producing x^i. Notice that the price is correlated with the unobserved component through σ. The price is similarly correlated with the rest of the covariates, in order to penalize misspecification. I set $\sigma = 0.3$. This results in higher prices when $d = 1$ and when x^1 and x^2 are higher.

The nesting parameter on d, ρ, is set to 0.3, and the sensitivity to price, α, is set to -2. The function $f(x)$ is defined in a manner that adds complexity to the regular linear combination of covariates in two ways. First, the continuous variables x^1 and x^2 enter utility logarithmically, rather than linearly – i.e., the marginal utility of consumers from these features is diminishing. Second, they are both interacted with d such that they are slightly more attractive to the consumer when $d = 1$:

$$f(x_j) = \beta_0 + \beta_d d_j + \beta_{x^1} \ln x_j^1 + \beta_{x^2} \ln x_j^2 + \beta_{dx^1} d_j \cdot \ln x_j^1 + \beta_{dx^2} d_j \cdot \ln x_j^2\,, \tag{4.2}$$

where $\beta = (\beta_0, \beta_d, \beta_{x^1}, \beta_{x^2}, \beta_{dx^1}, \beta_{dx^2}) = (-3, -1, 1, 1, 1, 1)$. The parameter values are chosen, as in Grigolon and Verboven (2014), to obtain an outside option share that is generally between 0.1 and 0.7. Also note that this particular structure is suited to demonstrating the second step approaches discussed in the previous section. The CATE of the GRF should identify the

[12] The distribution is bounded away from zero in order to keep the treatment effects bounded, since a log transformation is applied to x^1 and x^2 in the indirect utility, as detailed below. This is necessary because the boundedness of the treatment effect (besides being conceptually more realistic) is an important assumption of the estimation methods.

interactions of x^1 and x^2 with d, while the neural network regressor should identify both the interactions and the curvature of the effects.

The price and the log of the within-nest share are instrumented for by a combination of the instrument generated in (4.1) and the differentiation instruments of Gandhi and Houde (2017). The differentiation instruments, implemented using the PyBLP package of Conlon and Gortmaker (2020), are constructed here as follows: let x_{jtl} be characteristic $l \in \{d, x^1, x^2\}$ of product j in market t. The differentiation instrument z_{jtl}^{Diff} is the sum of the squared differences between x_{jtl} and the values of characteristic l for all rival products in the market, $J_t \backslash \{j\}$:

$$z_{jtl}^{Diff} \equiv \sum_{k \in J_t \backslash \{j\}} \left(x_{jtl} - x_{ktl}\right)^2. \tag{4.3}$$

Since prices in this setup are not determined in equilibrium, the correlations of the three differentiation instruments with the prices and within-nest shares are weak. The correlations are not zero because the distributions of the product characteristics are such that they are weakly correlated with the instruments' values, and these characteristics are at the same time correlated with the price. However, the finite-sample performance of DML is especially sensitive to the strength of instruments used. I therefore include a single additional instrument that is moderately correlated with the price. The generation of such an instrument is reasonable in the sense that instruments of at least a similar richness are often available to researchers (furthermore, when prices are determined in equilibrium, differentiation instruments should also be stronger). It is therefore important to emphasize that the results presented here are heavily dependent on the chosen instrument quality – stronger instruments should generally bring about faster convergence, and weaker instruments should result in slower convergence.

4.2. Simulation results

This section presents estimates of α and ρ, as well as of the own and cross price elasticities that they imply, and of the gradients of f with respect to d and x^1 (x^2 need not be considered separately as its effect is symmetric to that of x^1). These are estimated in three ways. The first – referred to as the linear model – is a simple two-stage least squares (2SLS) regression where the effect of the covariates on utility is modeled as $f(x) = x\beta$, for $x = (1, d, x^1, x^2)$. The second – referred to as the log model – is similarly a 2SLS regression, but this time correctly specifying that x^1 and x^2 enter utility through the log of their values (though still not accounting for their interactions with d); that is, $x = (1, d, \ln x^1, \ln x^2)$. The third approach is the flexible two-step approach discussed in section 3.

As shown below, the convergence rate of the flexible model is relatively slow – a trait that is mainly due to the IV-reliant first step. However, its finite sample bias is in this setting a significant improvement on the misspecification bias of the linear model, and it approximates the true data generating process reasonably well even given modest sample sizes. Moreover, the flexible model does succeed in characterizing the general functional form of $f(x)$ even with a small sample. This can aid the researcher in selecting an appropriate parametric functional form that should be controlled for even when the sample size is too small to allow a reliable non-parametric estimation.

First step results

Beginning with the first step, panels A and B of figure 1 summarize the distributions of the estimated $\hat{\alpha}$ and $\hat{\rho}$, respectively, in each specification and across different sample sizes. Panel C summarizes the estimates of the statistic $\frac{\alpha}{1-\rho}$, which is an important component of the nested logit formulas of own price and (within-nest) cross price elasticities. A first observation is that the linear model is heavily biased by its failure to account for the concavity of $f(x)$ with respect to x^1 and x^2. The sensitivity to price, α, is estimated as being close to zero, rather than -2, while the nesting parameter, ρ, is estimated as close to one, rather than its true value of 0.3. Taken together in $\frac{\hat{\alpha}}{1-\hat{\rho}}$, these biases compensate for each other to some extent, but do not converge to the true value. The log model, on the other hand, while also slightly misspecified in that it lacks the interaction terms, successfully approximates the truth by averaging the utility effects when $d = 0$ and when $d = 1$.

A second observation is that the flexible model, while $\sqrt{N}$ consistent, converges to the truth more slowly than the fully parametric models converge to their respective biased values, with its generality further compounded by the need to instrument for both the price and the within-nest share. However, it does converge to the truth, and it significantly outperforms the standard linear specification in this setup even when the sample size is small. Furthermore, the finite sample biases of $\hat{\alpha}$ and $\hat{\rho}$ cancel out nicely to return relatively accurate elasticity estimates.

Figure 2 displays the distribution of the resulting root mean squared errors (RMSE) of the elasticity estimates across the simulation iterations. In each simulation iteration, these RMSEs are those of the estimated elasticities for each of the simulated products in comparison to their true values. Panels A, B, and C do this for the own price elasticities, the within-nest cross price elasticities, and the across-nest cross price elasticities, respectively. The accuracy of these estimates directly relates to the estimates summarized in figure 1 (reminder: see Appendix C for the elasticity formulas), and therefore the same descriptions of the convergence patterns above apply here. The value of $\frac{\alpha}{1-\rho}$ is the dominant component in the own price elasticity, while the individual biases of α and ρ are more pronounced in the cross price elasticities.

Second step results

The second step analysis summarizes the estimated marginal effects of x^1, and the estimated CATEs and ATEs of x^1 and d, on mean consumer utility. Note that marginal effects and CATEs are not available in the linear and log models as objects that are distinct from the ATEs. Figure 3 presents the estimated marginal effects of x^1 conditional on several values of d and x^2. Recall from (4.2) that x^1 enters the true $f(x)$ through $\beta_{x^1} \ln x_j^1 + \beta_{dx^1} d_j \cdot \ln x_j^1$, where β_{x^1} and β_{dx^1} both equal 1. Taking the derivative of $f(x)$ with respect to x^1, the true marginal effect of x^1 therefore takes the form

$$\frac{\partial f(x)}{\partial x^1} = \begin{cases} \dfrac{1}{x^1} & if \quad d = 0 \\ \dfrac{2}{x^1} & if \quad d = 1 \end{cases}. \tag{4.3}$$

The marginal effect of x^1 is positive and diminishing in x^1, is higher when $d = 1$, and is independent of x^2. This true marginal effect is denoted in figure 3 by the dashed black lines. It is noteworthy that while a large sample size is required to accurately approximate the full functional form, performing this exercise even with a modest sample size can provide an indication that the marginal effects are positive, are decreasing convexly in x^1, are larger when $d = 1$, and are at least not highly dependent on x^2.[13] Thus, researchers without a large sample may still be able to infer that their parametric model can be made more accurate by taking the logarithm of x^1 and by interacting it with d.

Figure 4 presents the CATE estimates returned by the GRF estimation for d over a continuous range of x^1 and at two values of x^2. Recalling that d enters $f(x)$ through $\beta_d d_j + \beta_{dx^1} d_j \cdot \ln x_j^1 + \beta_{dx^2} d_j \cdot \ln x_j^2$, where $\beta_d = -1$ and β_{dx^1} and β_{dx^2} both equal 1, this CATE is given by

$$\tau_d^{CATE}(x) = -1 + \ln x^1 + \ln x^2\,; \tag{4.4}$$

that is, it is more attractive for a product to be characterized by $d = 1$ at higher values of x^1 and x^2 (for low values of x^1 and x^2, this product characteristic is in fact unattractive, with τ_d^{CATE} negative). Again, a large sample size captures this functional form accurately, but a smaller sample size is already able to hint that the CATE is increasing and concave in both x^1 and x^2.

---

[13] As discussed in the empirical application, finite sample non-effects in neural networks, such as the one correctly returned here with respect to x^2, may result from regularization bias and should be interpreted with care.

Figure 5 presents the CATE estimates for x^1. Here, the dashed black lines which serve as the target that is being approximated have a less straightforward interpretation, because of the concavity with which x^1 enters $f(x)$. This concavity means that the effect of x^1 is not constant for a given level of d and x^2. The best that can be demanded of a CATE estimate in this setting is a best linear predictor of the log form with which x^1 truly enters $f(x)$ (at each point in the covariate space of d and x^2). This is therefore how the target is defined.[14] Even with a small sample size, the CATE can be recognized here as dependent on d and not on x^2, although accurate values require a larger sample size.

The ATE estimates for d and x^1 are displayed in panels A and B of figure 6, respectively. For d, this ATE is simply the average CATE across the distribution of products. For x^1, the ATE, like the CATE, is not well-defined. Again, the best that can be expected from such ATE estimates is that they approximate the best linear predictor of the true functional form with which x^1 enters $f(x)$, with the difference this time being that products are not considered separately at different values of d, but rather we estimate a single average value. Here, it is also possible to compare the ATE of the flexible model to the 2SLS coefficients of the linear and log models.[15] The results in terms of biases and convergence rates are similar to those of the first step estimates.[16] The linear model converges away from the truth (although it happens to be quite close for the ATE of d), the log model successfully estimates the *average* treatment effects, and the flexible model converges to the truth, but slowly. In the case of ATEs, the application of the flexible model is uninformative as a descriptive exercise, but a large sample size may justify confidence that its estimate may outperform that of a parametric specification.

To summarize this section, we have seen that while the flexible model converges to the truth slowly, its finite sample performance may still exceed that of a misspecified linear model even at small sample sizes. The second step estimators, which in the setup considered here involve exogenous covariates only and therefore do not require instrumentation (as is the case for the non-price covariates in most studies), all approximate the true values quite closely given a sufficiently large sample, but they also hint at the general functional form of the indirect utility at smaller sample sizes. Thus, even if a linear specification is ultimately preferred, the flexible

---

[14] In greater detail, this best linear predictor is constructed as follows: Let X be a vector of 1,000 scalars that are evenly spaced on $[0.1,5]$, and let $\beta \in \mathbb{R}$. Define $Y \equiv \beta \ln X$, where $\beta = 1$ for $d = 0$ and $\beta = 2$ for $d = 1$, as specified in (4.2). Then the best linear predictor is constructed as the OLS estimate from regressing Y on X, with the addition of a constant term and with X entering the regression linearly.

[15] Since the log model already takes the log of x^1 as an input, its coefficient on x^1 is on a different scale from the estimates of the other two models. It must therefore undergo a transformation similar to that used in defining the best linear predictors above in order to place it on the same scale. Specifically, let $\hat{\beta}^{2SLS}$ be the 2SLS coefficient returned by the log model on x^1, and let X be defined as in footnote 14. Define $Y \equiv \hat{\beta}^{2SLS} \ln X$, which is the value predicted by x^1 in the log model, holding all other variables at 0. Then the best linear approximator of this estimated relationship is constructed as the OLS estimate from regressing Y on X, with the addition of a constant term.

[16] Indeed, although not shown here, the second step error is driven almost entirely by that of the first step.

model can inform the researcher on the functional form that would minimize misspecification bias.

5. Empirical Application

The previous section evaluated the performance of the flexible model on a stylized simulation setup. The present section takes this approach to an empirical dataset that is extensively used in the literature analyzing the domestic US airline market. While, as discussed below, much can be gained from such an application to the airline market, it is important to note that the main emphasis of this section is an evaluation of the estimation approach itself, and the empirical application is designed in accordance with this goal. The first step returns elasticity estimates of a magnitude that is consistent with the literature. These estimates are produced by a combination of a price sensitivity estimate that is around the lower tail (in absolute value) and a nesting parameter that is around the upper tail, respectively, of those usually seen in the literature. The implication is that substitution is estimated here to be stronger between products and weaker with the outside option than usual. A linear 2SLS estimation using the same data returns a higher price sensitivity and a small nesting parameter, suggesting that this result is not simply attributable to the specific sample or variables used.

The presentation of the second step results focuses, as an example, on consumer sensitivity to the distance travelled on a connecting trip relative to the direct flight distance. The ability of airlines to offer itineraries that are efficient in this respect is, to a large extent, a function of the sizes of their networks. For instance, in 2013, American Airlines and US Airways referenced their complementary networks in partial justification of their decision to merge, and their eventual settlement with the antitrust division of the US Department of Justice involved commitments to maintain pre-merger levels of operations at several airports of concern (especially pre-merger hubs of US Airways). Bontemps, Remmy and Wei (2021) already find evidence that the post-merger optimization of the two airlines' networks contributed to consumer welfare. However, a more detailed mapping of consumer utility from the quality (in distance terms) of a connecting flight may contribute to a more complete analysis of the effects of a merger on consumer welfare. Moreover, such a mapping in a pre-merger simulation analysis may guide antitrust authorities in deciding whether to approve a merger, or in selecting particular airports where restrictions should be imposed on the operations of the merging parties as a condition for approval of the merger.

The second step results show that consumers are more sensitive to long detours when the origin and destination are closer together and when flying with a low cost carrier. However, it is also shown that this sensitivity is often positive, counterintuitively implying that a longer detour is at times considered appealing. This result might be explained by an omitted variable bias, and serves as a valuable reminder that while relaxing functional form assumptions can help mitigate misspecification bias, it can only do so to the extent that the relevant information exists in the available data. Additionally, despite the relatively large size of the dataset, meaningful marginal effect estimates fail to be obtained.

5.1. US domestic airline market data

The main dataset used in this section is the airline origin and destination survey (DB1B) of the Bureau of Transportation Statistics for the second quarter of 2012. The DB1B is a 10 percent random sample of US domestic flight tickets that includes prices, itinerary details, and flight characteristics. This dataset is combined with two secondary sources of information. One is the Bureau of Transportation Statistics' Schedule P-12(a), which provides operating carriers' fuel costs and fuel consumption at the monthly level, used in the estimation as supply side instruments. The other is population estimates of metropolitan statistical areas (MSAs) obtained from the US Census Bureau, used to construct market sizes (the average of the origin and destination populations), and as characteristics in the demand estimation.

In order to maximize comparability to the existing literature, I take care to maintain similarity with the methodological choices made in other papers.[17] A market is defined as a unidirectional trip between two airports. As in Park (2020), I consider markets between the 100 largest mainland US domestic airports by number of passenger boardings in 2012. Out of these, markets between airport pairs that are less than 150 miles apart are dropped from the sample. Additionally, to exclude seasonal and negligibly small markets, respectively, both markets which served fewer than 100 passengers in any quarter from 2012 to 2016 and markets which never served more than 200 passengers in any of those quarters are omitted.

Within each market, a product is defined as a combination of origin airport, destination airport, connecting airport (if the flight is not direct), ticketing carrier, and operating carriers. Tickets with more than one connection, multiple ticketing carriers, or involving ground travel in the itinerary are omitted, as are tickets with a fare credibility questioned by the Department of Transportation or which have a bulk fare equal to 1 (hinting that the ticket was purchased by a travel agency, rather than directly by a consumer). Finally, tickets with a price below 25 dollars or above 2,500 dollars (on one-way flights, 12.5 and 1,250, respectively) are also omitted from the sample. Ownership of a product is, as usual, attributed to the ticketing carrier. The tickets are aggregated over the quarter, with a product's price being the passenger-weighted average price of its tickets.[18] Carriers serving fewer than 90 passengers across all markets in the quarter are dropped from the sample.

[17] For a comparison, see Berry and Jia (2010); Ciliberto and Williams (2014); Park (2020); Bontemps, Remmy and Wei (2021); Sweeting, Roberts and Gedge (2020); Ciliberto, Murry and Tamer (2021); and Turner (2021).

[18] For roundtrip tickets, which are split here between two markets (one in each direction), the price is divided proportionally to the distance flown.

Table 1 presents summary statistics for several product and market level variables of interest. Across 180,342 products in 8,035 markets, the average price in the sample is 274.82 dollars, and the average product serves 454 passengers throughout the quarter. Recall that a product can be, and often is, composed of more than one flight. Such combinations of flights explain the low median relative to the capacity of a regular airplane, with travelers combining flight segments into a wide variety of itineraries. The far majority of products are connecting products simply due to the large number of possible combinations of flight segments. However, they serve only one quarter of passengers. The average origin and destination are around 1,500 miles apart, but many of the products are elongated by connections. The variable "extra miles", defined as the ratio of the total miles flown by the product to the nonstop distance, captures how big a detour a particular connecting flight takes (for nonstop flights, this variable is equal to one). This is the covariate analyzed here in the second step. The origin and destination airport presence variables are often used in airline demand estimation to capture the ability of airlines to provide better service at the airport at a lower cost (for instance, through better airport slots or more convenient gate access) and to reflect the value of airlines' loyalty programs. The origin airport presence of an airline is defined as the ratio of the number of destinations that a carrier serves from the origin airport to the total number of destinations that any carrier serves from the origin airport. The destination presence of an airline is defined analogously with respect to the destination airport. Finally, the average market consists of 22.45 products supplied by 4.32 airlines, jointly carrying 10,188 passengers throughout the quarter. The average market size (recall that this is the average population of the origin and destination MSAs) is around 3.45 million.

5.2. Demand estimation

In keeping with the convention in the literature (see footnote 17), airline demand is modeled as a nested logit model consisting of two nests: one which includes all inside option products, and one which includes only the outside option. Formally, this can be modeled as a two level decision process in which consumers first choose whether to fly or not, and then, if they decide to fly, choose a particular flight. This reflects the fact that flights are closer substitutes to one another than they are to other modes of transportation. Several variables are modeled as affecting consumer utility. These variables indicate whether the flight is nonstop or connecting, the origin presence and destination presence of the airline, and dummy variables for each of the legacy carriers – American Airlines (AA), US Airways (US), Delta Airlines (DL), and holding United Airlines (UA) as a reference group – as well as Southwest Airlines (WN). The remaining airlines are grouped into a single dummy variable of low cost carriers (LCC). Additionally, the demand specification includes the (nonstop) distance between the origin and destination, and the "extra miles" variable. Finally, the population estimates at the origin and destination MSAs are included to capture correlations of population sizes with market characteristics.

In order to address the endogeneity of the price and the (log of the) within nest share, I make use of a set of supply side instruments. Both the presence of an LCC in the market and the overall number of rival airlines in the market can generally be expected to exert a downward pressure on prices, while both a higher average distance of rivals' products (which reduces the appeal of those products) and the monopoly status of an airline should generally be expected to relieve this pressure. Additionally, the fuel cost per gallon of the operating carriers should function as a cost shifter (when a product is operated by two different carriers, their distance-weighted average cost is taken). These relationships are largely reflected in the 2SLS first stage analysis presented in table 2, with the only exception being the highly significant but wrongly signed fuel cost. The reason that two pairs of first stage regressions are presented in table 2 – one excluding fuel costs and the other including them – is that fuel cost data is unavailable for around 27 percent of products. In the DML estimation, fuel costs are included in the set of instruments, accounting for products with missing data by imputing a value of zero and including an indicator that equals one for products with missing fuel costs data.

Table 3 presents 2SLS estimates obtained from the two linear specifications (with and without fuel costs in the set of instruments). The two specifications return similar results. The estimated price sensitivity is around the middle of the wide range of estimates seen in the literature, and the estimated nesting parameter is small relative to the literature. Their combination results in own price elasticity estimates that are relatively small and in very weak substitution patterns across products. Additionally, the estimated effect of an increase in the "extra miles" variable is significantly negative, as should be expected.

Next, I present the estimates returned by the flexible two-step approach. The estimates presented are the means obtained from 800 bootstrap replications. Each replication draws two thirds of the sample with replacement and estimates the first and second steps in succession. The first step estimates – those of the price sensitivity and the nesting parameter – are presented in table 4. These estimates, recall, are obtained using DML while essentially controlling for the remaining product features in a nonparametric nuisance function. Note that while the standard errors are obtained using bootstrapping and therefore do not imply a formal significance level, the distributions of both estimates are very narrow. This time, the price sensitivity estimate is quite small in absolute value, while the nesting parameter estimate is large. The resulting own price elasticities are within the range generally reported in the literature, and the estimated cross price elasticities are several orders of magnitude larger than in the linear case (a comparison to the literature is harder in this case, as cross price elasticities are reported less often, but it may be inferred from the values of the estimated parameters that the cross price elasticities are relatively large).

As discussed above, the second step estimation focuses on the "extra miles" variable. Recall that this variable is the ratio between the total trip distance and the nonstop distance, and serves as a measure of the *degree* of indirectness of a connecting flight. Beginning with the simplest second step estimate, table 5 presents an unconditional ATE estimate of "extra miles". Unexpectedly, this estimate of the average utility effect is around zero (0.0026 with a standard error of 0.0132). As will be shown shortly, this reflects a distribution of CATEs some of which are significantly negative while others are, counterintuitively, significantly positive. Since a longer detour can safely be considered a negative trait in all flight types, such a distribution does not appear reasonable. One possible explanation for this result is the existence of important omitted variables that are coincidentally disguised in the 2SLS estimation, but that are biasing the estimate upward once the covariate space is controlled for flexibly. While this is not necessarily the case, the results here at least justify a careful consideration of this possibility. This serves as an important reminder that a flexible estimation methodology cannot replace a critical consideration of functional form, as such a consideration is required in order to determine whether the available dataset is suitable to answer the problem at hand in the first place.

Despite the upward bias, moving from an ATE to CATEs provides an understanding of the particular product types for which longer detours have a relatively bigger or smaller impact on consumer welfare. Table 6 presents an OLS regression of the GRF-estimated CATES of "extra miles" – let us denote them $\hat{\tau}_{EM}(x)$ – on the remaining covariates of the sample observations. All of the covariates in the regression are scaled to have mean zero and a variance of one, such that the magnitude of each coefficient estimate reflects the strength of the correlation between $\hat{\tau}_{EM}(x)$ and the covariate. This regression is estimated only on connecting products, as "extra miles" does not vary for nonstop flights and its interaction with those flights is therefore not meaningful. It can be seen that four covariates interact particularly strongly with $\hat{\tau}_{EM}(x)$: the nonstop distance of the market, the carrier being an LCC that is not Southwest, and the population estimates at the origin and destination MSAs. The correlations with the population estimates are driven solely by airports in three particularly large cities that are important hubs in the American domestic flight network: New York City, Chicago, and Los Angeles. I therefore restrict attention to the relationship of $\hat{\tau}_{EM}(x)$ with the nonstop distance and the identity of the carrier.

Figure 7 presents the distribution of $\hat{\tau}_{EM}(x)$ over a range of nonstop distances separately for each airline, evaluated at the median values of the remaining covariates. The figure implies that, in general, the closer the origin and destination are to each other, the less consumers are willing to tolerate long detours. This relationship is especially pronounced as the market distances become smaller. This is a reasonable result, reflecting the fact that at closer distances, outside options become more viable (a shortcoming of the nested logit model is that the nesting parameter cannot be interacted with other covariates, with this being an example in which such an interaction might have been appropriate). Another observation is that longer detours are especially unattractive when they are conducted by LCCs (excluding Southwest). This may be due to poorer amenities discouraging consumers from choosing LCCs for longer detours, although the relatively higher willingness to pay for longer detours by Southwest raises question marks over this interpretation.

Finally, it would be informative to progress from CATEs to marginal effects of "extra miles". However, as shown in figure 8, the neural network regressor returns constant marginal effects not only at different levels of "extra miles", but also for different nonstop distances and for different airlines (all statistically insignificant), in contradiction to the CATE estimates. The invariance of the marginal effect estimates, as well as their proximity to zero, suggests that these estimates are suffering from regularization bias in the neural network, which biases the network toward a simpler prediction model that neglects to attribute significant weight to some of the covariates. Therefore, the neural network regressor is uninformative with respect to the marginal effects of "extra miles". However, it does provide two lessons that are of value to the current discussion. One is that accurate marginal effect estimates in this context might require a larger sample size than the 180,342 observations available here, suggesting that in many standard settings, second step marginal effect estimates can at best contribute as an exploratory exercise, and not as main results. The other is that even as part of an exploratory exercise, the neural network estimates, and particularly implied non-effects, must be interpreted with caution.

A further comment, the demonstration of which must be left to a future version of this work, is that convergence rates can typically be accelerated by guiding the models with whatever structure is theoretically available. For example, the researcher may reasonably assume that the marginal reduction in utility from "extra miles" is decreasing (that is, a change at higher values matters less than at lower values). Indeed, Park (2020) accounts for this by including the square of "extra miles" in the demand specification. Doing the same in the neural network model could reduce the burden on the data and accelerate convergence of the estimates. This serves as a second reminder that flexible estimation techniques do not obviate a serious consideration of functional form. On the contrary, and especially given large sample sizes, machine learning models benefit from the inclusion of theoretically important functions and interactions of the covariates in the set of inputs.

To summarize the results of this section, the flexible specification of the mean indirect utility returns first step estimates that produce reasonable substitution patterns, but are at the edges of the distributions of price sensitivity and nesting parameter estimates generally found in the literature. A second step analysis of the ATE of only a single covariate – "extra miles" – reveals a possibility that an important product feature is absent from the data and is biasing the results. An analysis of the CATE of the same covariate appears to suggest that a linear utility index can be improved by interacting it with the identity of the carrier and with a concave function of the nonstop market distance. Finally, in this application, little can be learned about the marginal effects of "extra miles".

6. Conclusion

The increasing availability of large datasets and the concurrent development of estimation techniques that are more efficiently able to handle complex models are making it increasingly possible to relax functional form assumptions where economic theory is agnostic, and thus reduce bias resulting from model misspecification. This paper presents an empirical approach that does this within the theoretical framework of a nested logit demand model. It is shown that relaxing linearity can provide benefits both in accurately estimating the linear parameters of the model – the sensitivity to price and the nesting parameter – and in approximating the functional form of the remaining utility effects (although where theoretical insights on functional form do exist, it is beneficial to impose those insights on the model). Moreover, even when a parametric model is ultimately preferred, this functional form approximation can serve to motivate the choice of specification. Thus, more accurate substitution patterns can potentially be estimated and used in various counterfactual equilibrium analyses, and more detailed functions of willingness to pay for product features can be mapped to better guide both policymakers and industry practitioners.

This work leaves several open ends for further research. First, a wider reliance on the two-step approach used here would benefit greatly from a derivation of analytical properties of the second-step estimators. Second, the nested logit model has been applied in various empirical settings apart from airline markets, and flexibly controlling for the nuisance function may produce estimates in these settings that differ from the values usually obtained, and may shed new insights on the effects of the product characteristics. When the first step returns estimates that differ from those normally found in the literature, a critical reevaluation of functional form assumptions might be in order. Third, the incorporation of greater consumer heterogeneity into the estimation would provide more specific substitution patterns, and thus enable more accurate counterfactual equilibrium analysis.

References

Anderson, Simon P., Andre De Palma, and Jacques-Francois Thisse. *Discrete choice theory of product differentiation*. MIT press, 1992.

Athey, Susan, Julie Tibshirani, and Stefan Wager. "Generalized random forests." *The Annals of Statistics* 47.2 (2019): 1148-1178.

Bajari, Patrick, et al. *Demand estimation with machine learning and model combination*. No. w20955. National Bureau of Economic Research, 2015.

Battocchi, Keith, et al. "EconML: A Python Package for ML-Based Heterogeneous Treatment Effects Estimation." *https://github.com/microsoft/EconML*, 2019. Version 0.x.

Belloni, Alexandre, Victor Chernozhukov, and Christian Hansen. "Inference on treatment effects after selection among high-dimensional controls." *The Review of Economic Studies* 81.2 (2014): 608-650.

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

Berry, Steven, James Levinsohn, and Ariel Pakes. "Automobile prices in market equilibrium." *Econometrica: Journal of the Econometric Society* (1995): 841-890.

Berry, Steven, and Panle Jia. "Tracing the woes: An empirical analysis of the airline industry." *American Economic Journal: Microeconomics* 2.3 (2010): 1-43.

Bjornerstedt, Jonas, and Frank Verboven. "Does merger simulation work? Evidence from the Swedish analgesics market." *American Economic Journal: Applied Economics* 8.3 (2016): 125-64.

Bontemps, Christian, Kevin Remmy, and Jiangyu Wei. "Ex-post Evaluation of the American Airlines–US Airways Merger A Structural Approach." *Journal of Transport Economics and Policy (JTEP)* 56.2 (2022): 129-155.

Cardell, N. Scott. "Variance components structures for the extreme-value and logistic distributions with application to models of heterogeneity." *Econometric Theory* 13.2 (1997): 185-213.

Chernozhukov, Victor, et al. "Double/debiased machine learning for treatment and structural parameters." (2018): C1-C68.

Ciliberto, Federico, and Jonathan W. Williams. "Does multimarket contact facilitate tacit collusion? Inference on conduct parameters in the airline industry." *The RAND Journal of Economics* 45.4 (2014): 764-791.

Ciliberto, Federico, Charles Murry, and Elie Tamer. "Market structure and competition in airline markets." *Journal of Political Economy* 129.11 (2021): 2995-3038.

Conlon, Christopher, and Jeff Gortmaker. "Best practices for differentiated products demand estimation with pyblp." *The RAND Journal of Economics* 51.4 (2020): 1108-1161.

Dubé, Jean-Pierre H., Joonhwi Joo, and Kyeongbae Kim. *Discrete-Choice Models and Representative Consumer Theory*. No. w30130. National Bureau of Economic Research, 2022.

Farrell, Max H., Tengyuan Liang, and Sanjog Misra. "Deep neural networks for estimation and inference." *Econometrica* 89.1 (2021): 181-213.

Gandhi, Amit, and Jean-François Houde. "Measuring substitution patterns in differentiated-products industries." *NBER Working Paper* w26375 (2017).

Gillen, Benjamin J., et al. "BLP-2LASSO for aggregate discrete choice models with rich covariates." *The Econometrics Journal* 22.3 (2019): 262-281.

Grigolon, Laura, and Frank Verboven. "Nested logit or random coefficients logit? A comparison of alternative discrete choice models of product differentiation." *Review of Economics and Statistics* 96.5 (2014): 916-935.

Hartford, Jason, et al. "Deep IV: A flexible approach for counterfactual prediction." *International Conference on Machine Learning*. PMLR, 2017.

Hornik, Kurt, Maxwell Stinchcombe, and Halbert White. "Multilayer feedforward networks are universal approximators." *Neural networks* 2.5 (1989): 359-366.

Newey, Whitney K., and Daniel McFadden. "Large sample estimation and hypothesis testing." *Handbook of econometrics* 4 (1994): 2111-2245.

Park, Yongjoon. "Structural Remedies in Network Industries: An Assessment of Slot Divestitures in the American Airlines/US Airways Merger." *US Airways Merger (October 12, 2020)* (2020).

Sheshinski, Eytan. "Price, quality and quantity regulation in monopoly situations." *Economica* 43.170 (1976): 127-137.

Spence, Michael. "Monopoly, quality, and regulation." *The Bell Journal of Economics* (1975): 417-429.

Spence, Michael. "Product differentiation and welfare." *The American Economic Review* 66.2 (1976): 407-414.

Sweeting, Andrew, James W. Roberts, and Chris Gedge. "A model of dynamic limit pricing with an application to the airline industry." *Journal of Political Economy* 128.3 (2020): 1148-1193.

Turner, Douglas. "Coordinated Effects in the American Airlines-US Airways Merger." *Available at SSRN 3917112* (2021).

Wager, Stefan, and Susan Athey. "Estimation and inference of heterogeneous treatment effects using random forests." *Journal of the American Statistical Association* 113.523 (2018): 1228-1242.

Appendix A: Figures

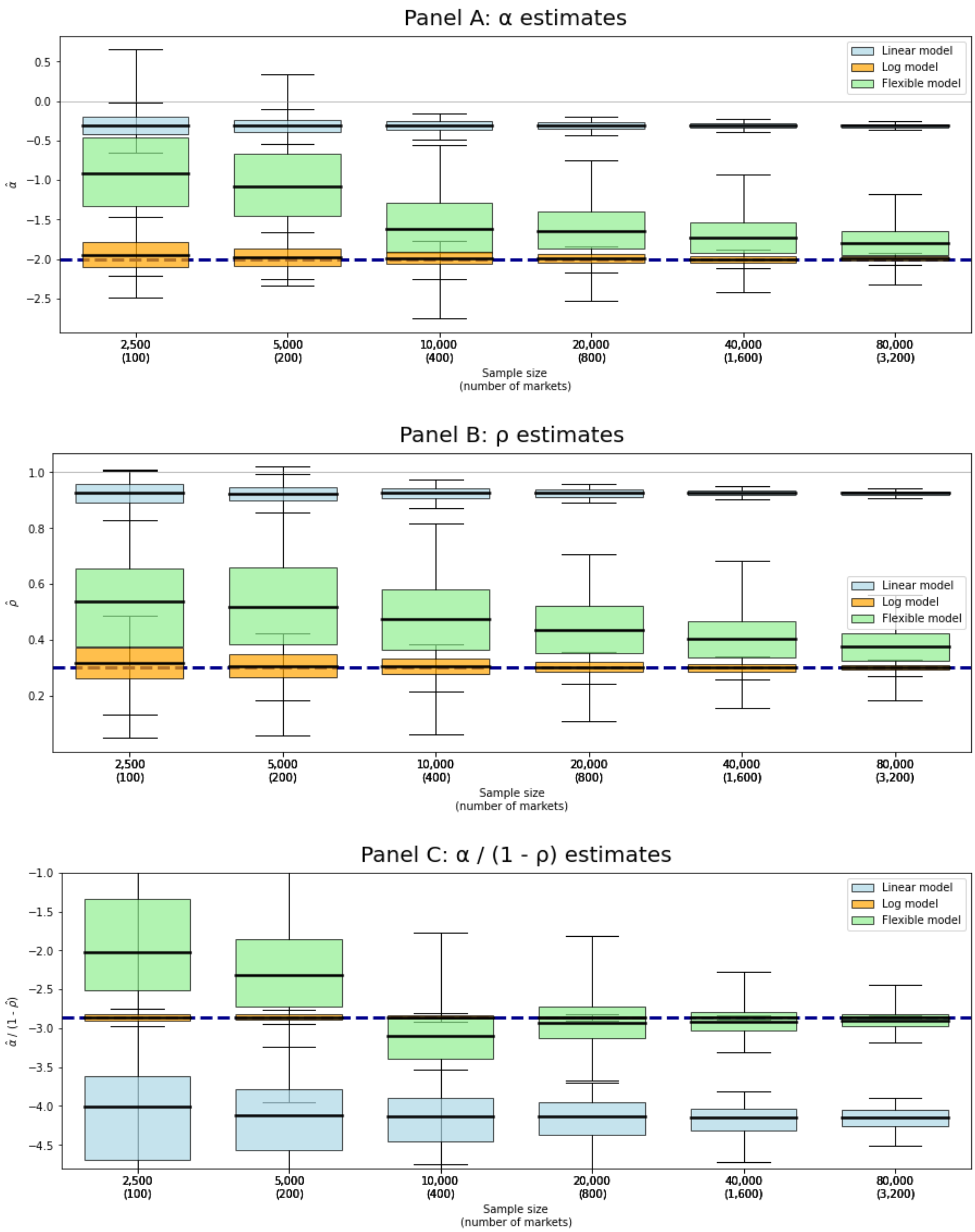

Figure 1. First step parameter estimates. The horizontal dashed blue lines mark the true parameter values. Each boxplot displays the distribution of the estimates across 1,000 simulation iterations. The median of the estimates is marked by the thick central line, the range of the box is the interquartile range of the estimates, and the whiskers extend to the 2.5th and 97.5th percentiles, respectively.

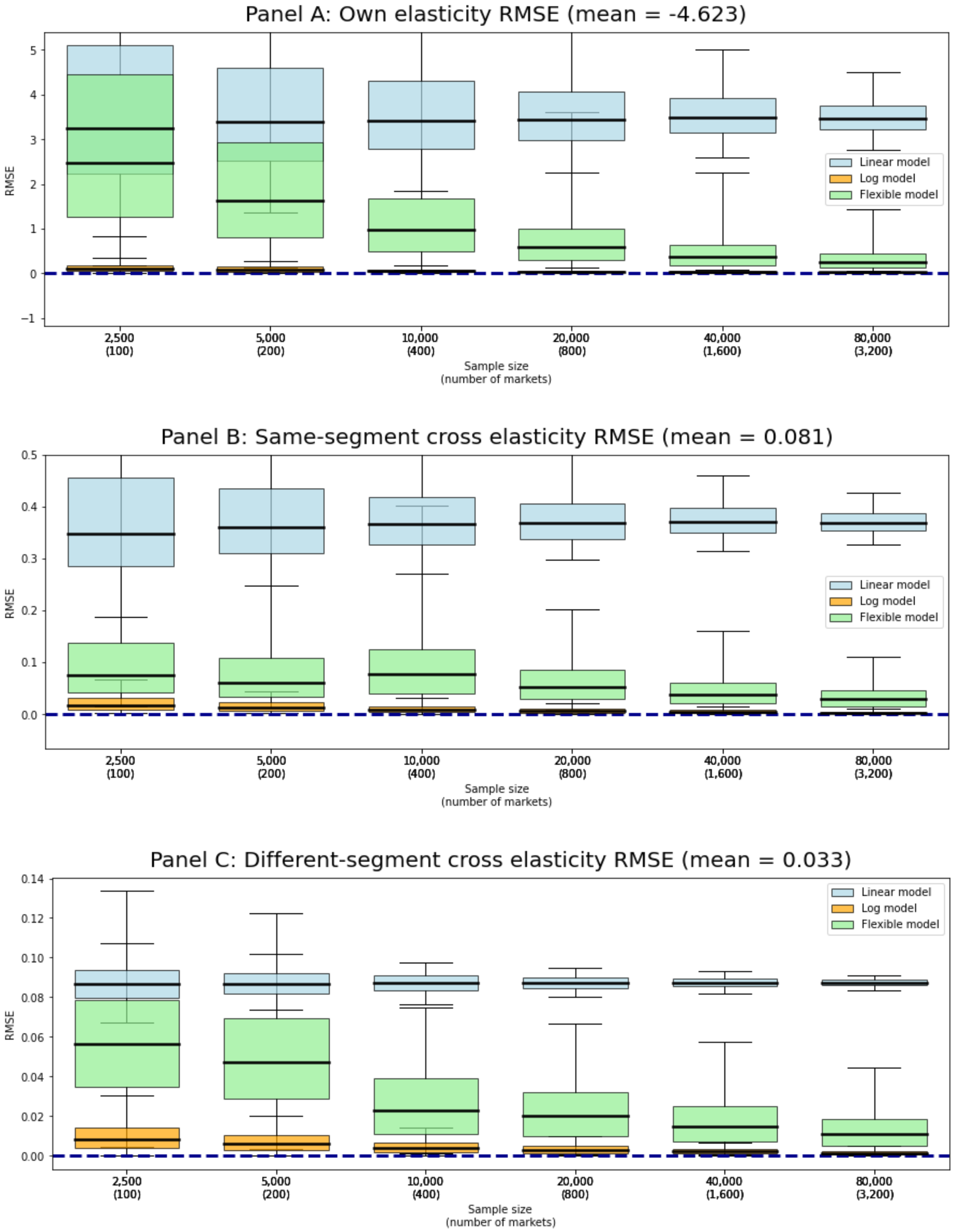

Figure 2. Root mean squared error (RMSE) of the elasticity estimates. Each boxplot displays the distribution of the RMSE across 1,000 simulation iterations. The median of the RMSEs across the iterations is marked by the thick central line, the range of the box is the interquartile range of the RMSEs, and the whiskers extend to the 2.5th and 97.5th percentiles, respectively.

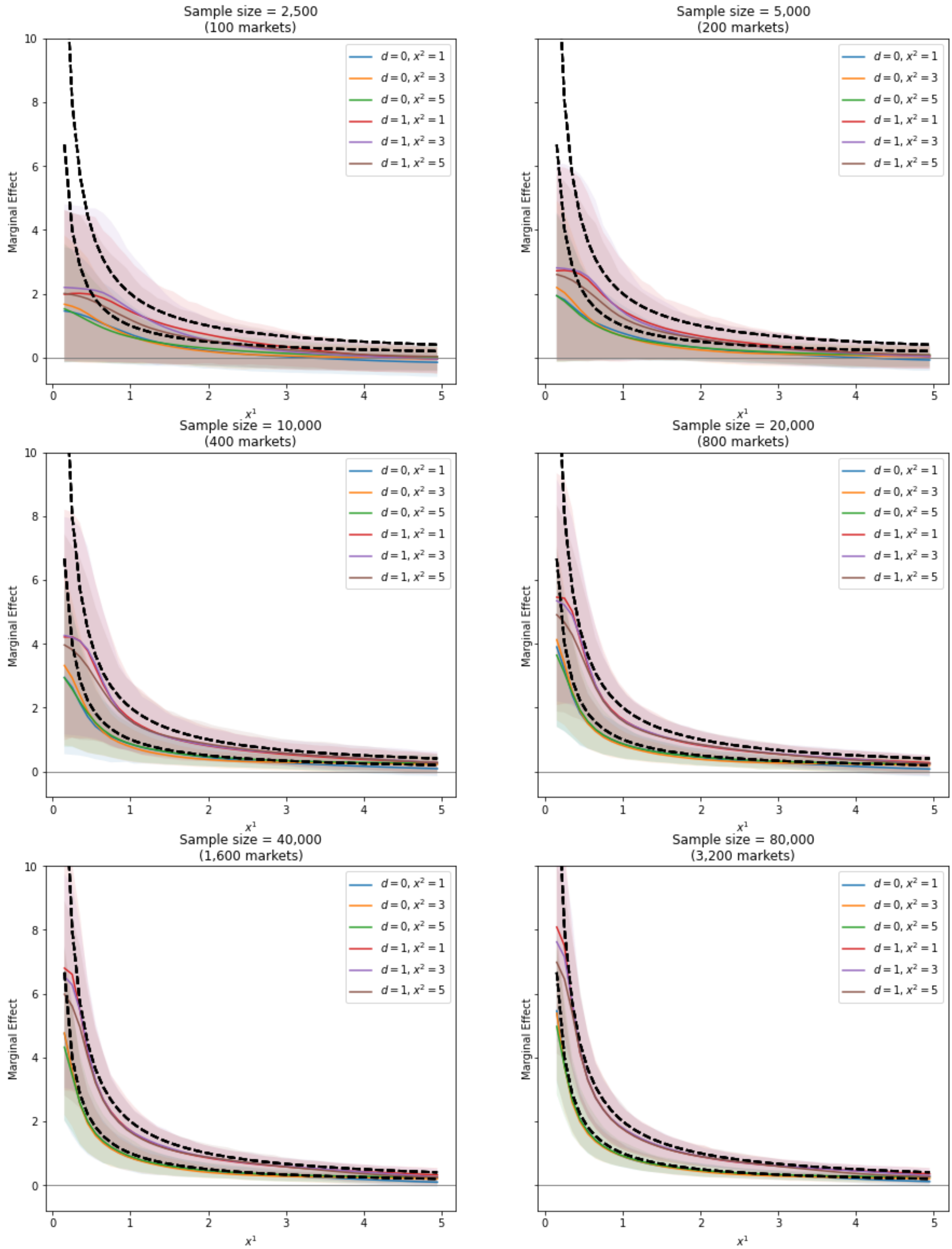

Figure 3. Estimated marginal effects of x^1, conditional on d and x^2. The dashed lines represent the true marginal effects at $d = 0$ (bottom line) and $d = 1$ (top line). The colored lines display the mean marginal effect estimates across 1,000 simulation iterations, estimated with a neural network regressor, and are surrounded by confidence intervals denoting the 2.5th and 97.5th percentiles of estimates, respectively.

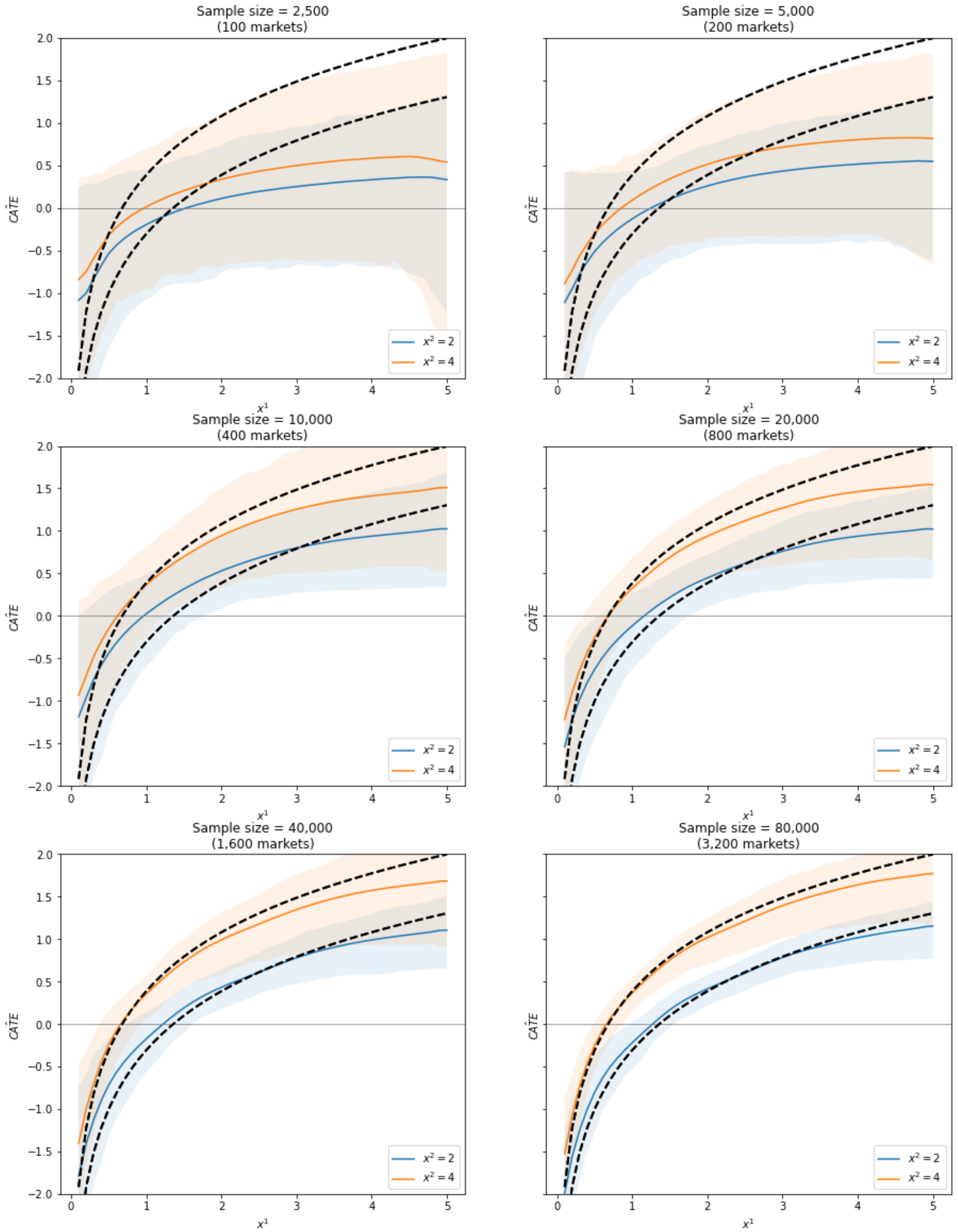

Figure 4. Estimated CATE of d, conditional on x^1 and x^2. The dashed lines represent the true CATE at $x^2 = 2$ (bottom line) and $x^2 = 4$ (top line) as examples. The colored lines display the mean CATE estimate across 1,000 simulation iterations, estimated by GRF, and are surrounded by confidence intervals denoting the 2.5th and 97.5th percentiles of estimates, respectively.

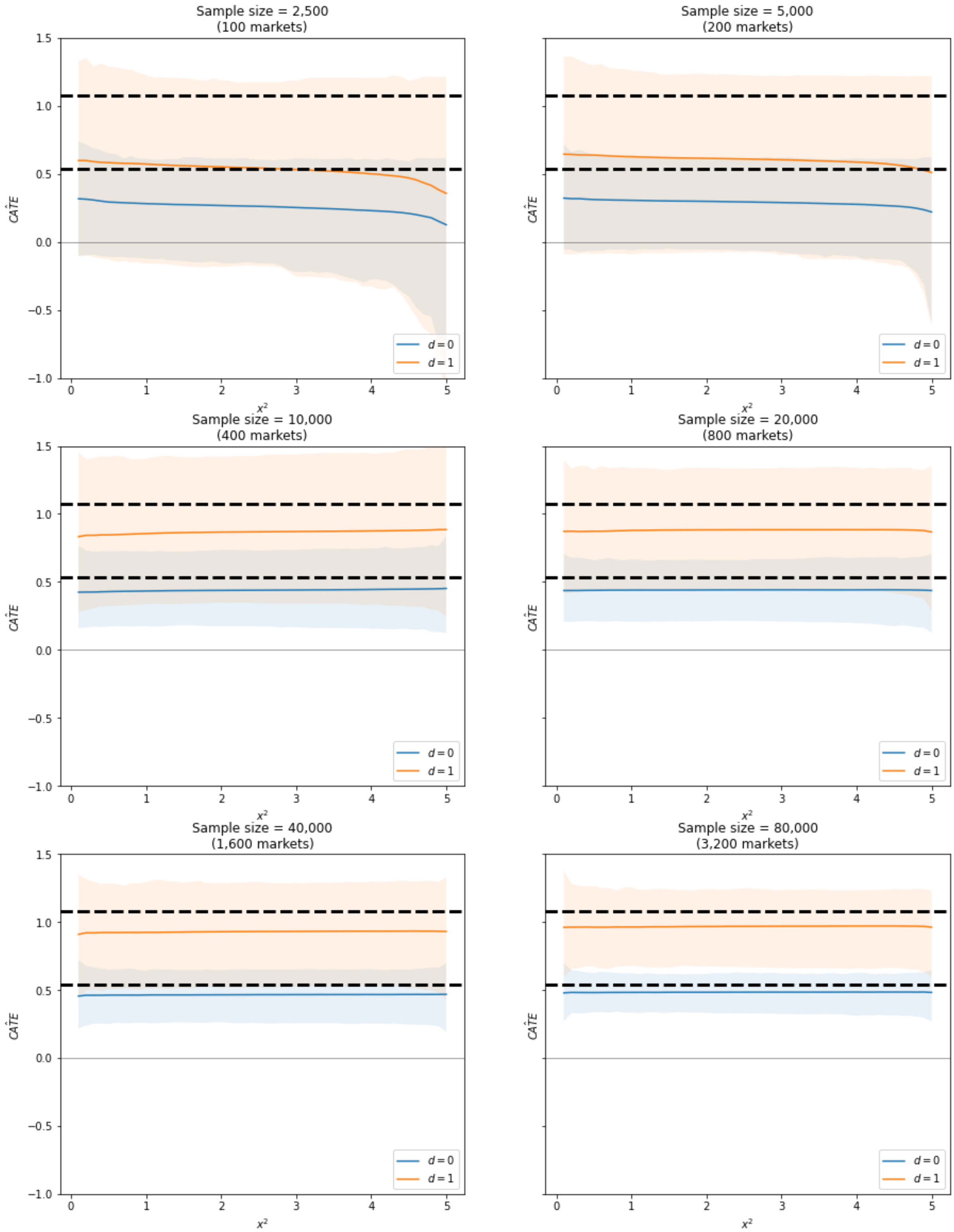

Figure 5. Estimated CATE of x^1, conditional on d and x^2. The dashed horizontal lines represent the true CATE at $d = 0$ (bottom line) and $d = 1$ (top line). The colored lines display the mean CATE estimate across 1,000 simulation iterations, estimated by GRF, and are surrounded by confidence intervals denoting the 2.5th and 97.5th percentiles of estimates, respectively.

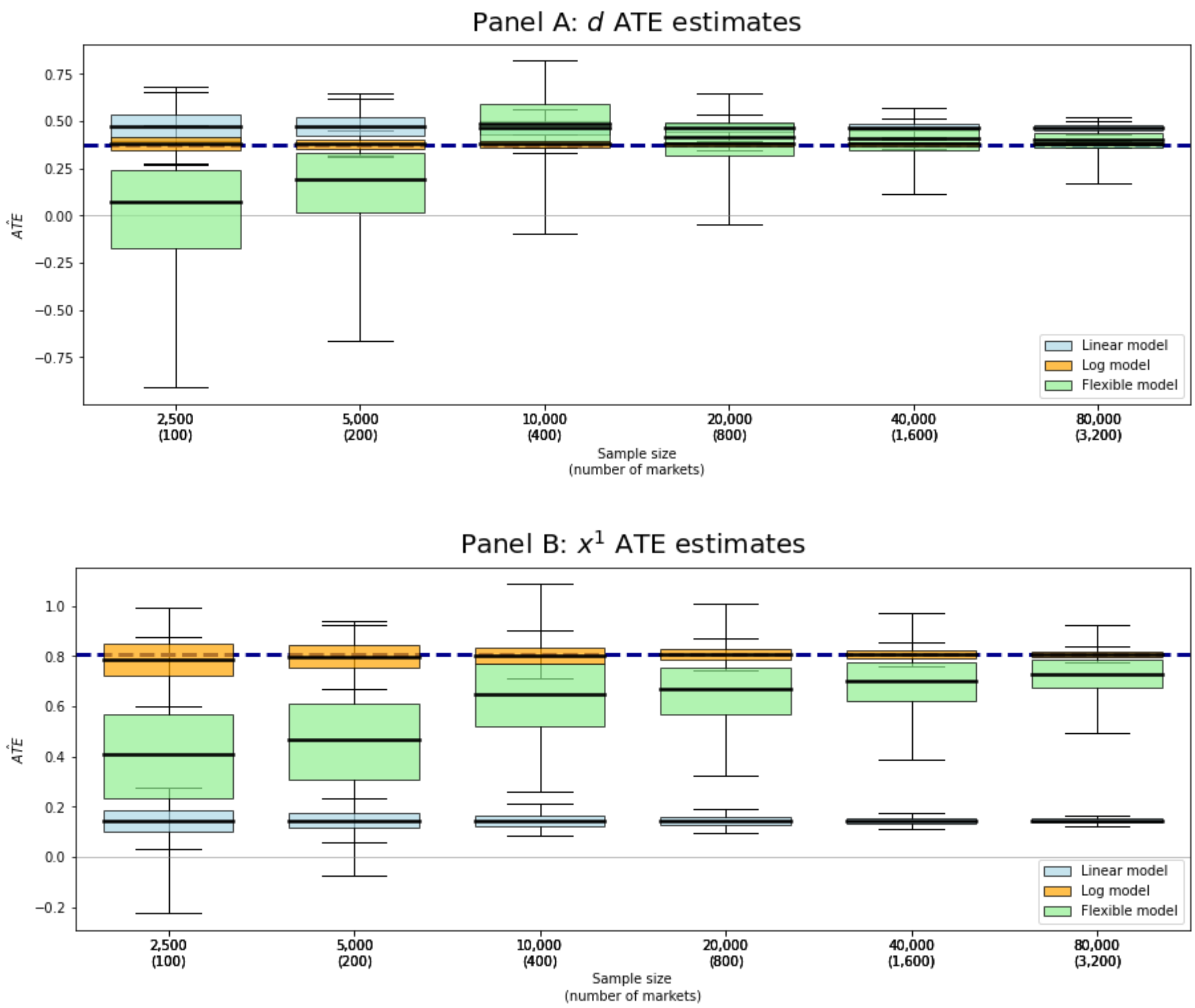

Figure 6. Estimated ATEs of d and x^1. The horizontal dashed blue lines mark the true values. Each boxplot displays the distribution of the estimates across 1,000 simulation iterations. The median of the estimates is marked by the thick central line, the range of the box is the interquartile range of the estimates, and the whiskers extend to the 2.5th and 97.5th percentiles, respectively.

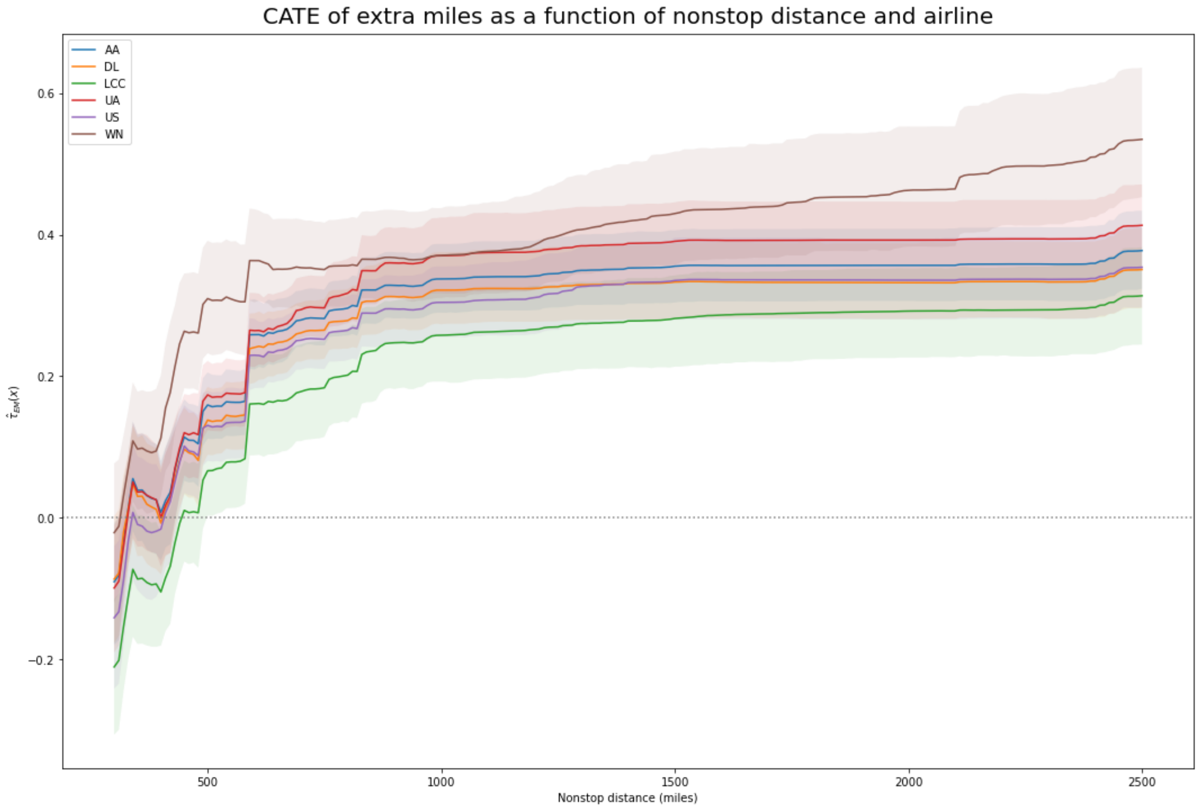

Figure 7. GRF-estimated CATE of "extra miles", $\hat{\tau}_{EM}(x)$, conditional on the nonstop distance of the market and on the identity of the airline, evaluated at the median values of the remaining covariates. The colored lines display the mean CATE estimates across 800 bootstrap replications, and are surrounded by confidence intervals denoting the 2.5th and 97.5th percentiles of estimates, respectively.

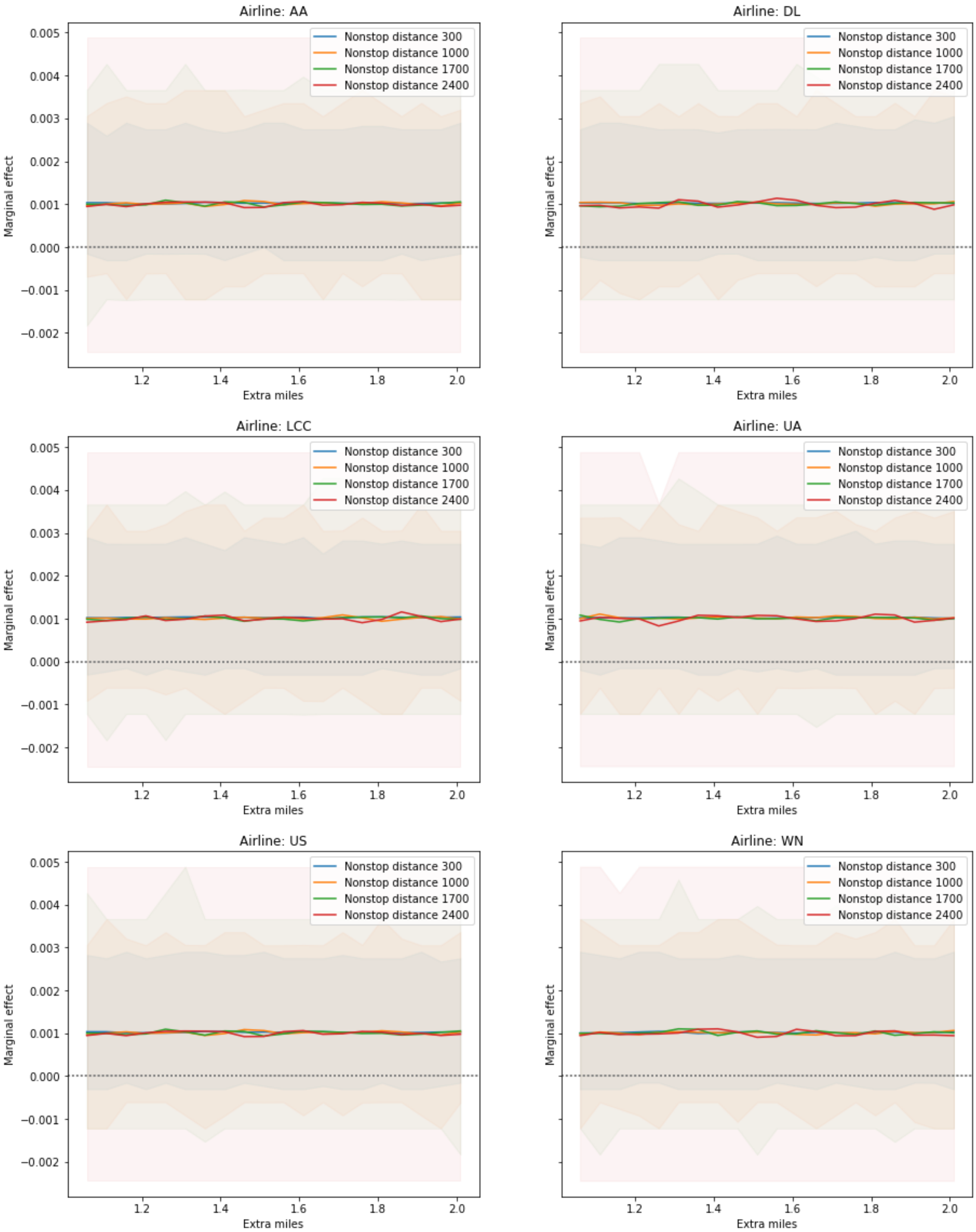

Figure 8. Neural network estimated marginal effects of "extra miles", conditional on different values of the nonstop distance and for different airlines. The colored lines display the mean marginal effect estimates across 800 bootstrap replications, and are surrounded by confidence intervals denoting the 2.5th and 97.5th percentiles of estimates, respectively.

Appendix B: Tables

Panel A: Product level

	Mean	Std. Dev.	Min	Max	Median
Price (USD)	274.82	123.79	14.00	1,250.00	249.72
Passengers	453.91	2,917.72	10.00	114,780.00	40.00
Passengers (nonstop)	4,160.04	9,422.78	10.00	114,780.00	360.00
Nonstop (dummy)	0.08	0.27	0.00	1.00	0.00
Total distance (miles)	1,489.63	808.32	152.00	7,445.00	1,305.00
Extra miles	1.24	0.31	1.00	2.99	1.11
Origin presence	0.84	0.17	0.01	1.00	0.90
Destination presence	0.84	0.17	0.00	1.00	0.90

Panel B: Market level

	Mean	Std. Dev.	Min	Max	Median
Number of products	22.45	15.73	1.00	162.00	18.00
Number of airlines	4.32	1.57	1.00	12.00	4.00
Market passengers	10,187.78	19,091.76	100.00	269,180.00	3,290.00
Market size	3,453,688.28	2,995,483.12	113,573.00	16,451,462.00	2,481,372.50

Table 1. Summary statistics for all 180,342 (14,676 of which are nonstop) products and 8,035 markets in the domestic US airline market in the second quarter of 2012.

	(1)		(2)	
	Excluding fuel costs		Including fuel costs	
Dependent Variable:	Price	$\ln s_{j/g}$	Price	$\ln s_{j/g}$
(Intercept)	-31.80***	-0.6540***	-15.85	-1.097***
	(4.696)	(0.065)	(9.387)	(0.1349)
LCC presence	-35.25***	-0.5907***	-37.31***	-0.5559***
	(0.695)	(0.0096)	(0.8499)	(0.0122)
Rivals' average distance	0.0581***	0.0031***	0.0456***	0.0029***
	(0.0024)	(3.32E-05)	(0.0028)	(3.97E-05)
Number of rivals	-0.4318***	-0.0387***	-0.3803***	-0.0386***
	(0.0194)	(0.0003)	(0.0224)	(0.0003)
Monopoly	47.44***	4.442***	37.30***	4.604***
	(5.137)	(0.0711)	(5.901)	(0.0848)
Nonstop	-4.130***	1.322***	-4.503***	1.196***
	(1.031)	(0.0143)	(1.123)	(0.0161)
Origin presence	77.05***	-1.020***	78.71***	-0.9333***
	(3.233)	(0.0447)	(3.668)	(0.0527)
Destination presence	83.01***	-1.021***	83.85***	-0.9170***
	(3.21)	(0.0444)	(3.642)	(0.0523)
Nonstop miles	0.0101***	-0.0029***	0.0255***	-0.0027***
	(0.0025)	(3.39E-05)	(0.0028)	(4.05E-05)
Extra miles	81.16***	-1.286***	89.99***	-1.393***
	(0.9575)	(0.0132)	(1.127)	(0.0162)
Origin population	3.58e-7***	-5.87e-8***	2.85e-7***	-5.55e-8***
	(6.34E-08)	(8.78E-10)	(7.40E-08)	(1.06E-09)
Destination population	5.53e-7***	-6.02e-8***	5.89e-7***	-5.75e-8***
	(6.38E-08)	(8.82E-10)	(7.44E-08)	(1.07E-09)
AA	25.98***	0.8003***	14.49***	0.9217***
	(1.365)	(0.0189)	(1.72)	(0.0247)
DL	-2.996***	0.4147***	-3.466**	0.6660***
	(0.6978)	(0.0097)	(1.059)	(0.0152)
LCC	49.29***	0.0221	54.71***	-0.1273*
	(3.024)	(0.0418)	(3.551)	(0.051)
US	39.97***	0.1443***	41.03***	0.1428***
	(0.8853)	(0.0122)	(0.9846)	(0.0141)
WN	20.13***	0.5459***	20.12***	0.5778***
	(1.488)	(0.0206)	(1.611)	(0.0232)
Fuel cost per gallon	–	–	-9.631***	0.1286***
	–	–	(2.455)	(0.0353)
S.E. type	Standard	Standard	Standard	Standard
Observations	180,342	180,342	131,015	131,015
R2	0.17744	0.3609	0.20343	0.34767

Table 2. 2SLS first stage regression estimates for each of the two endogenous variables, under two specifications that differ only in that the second specification includes fuel costs in the set of instruments. Note that fuel cost data is available for roughly 72.5% of the sample.

	(1)	(2)
	2SLS excluding fuel costs	2SLS including fuel costs
Dependent Variable:	$\ln s_j - \ln s_0$	$\ln s_j - \ln s_0$
(Intercept)	-12.54***	-12.38***
	(0.0551)	(0.0666)
Price ($100)	-0.6323***	-0.6410***
	(0.0322)	(0.0355)
$\ln s_{j/g}$	0.1545***	0.1423***
	(0.0075)	(0.0081)
Nonstop	1.526***	1.409***
	(0.0166)	(0.0178)
Origin presence	1.006***	0.9350***
	(0.0489)	(0.0563)
Destination presence	1.043***	0.9887***
	(0.0498)	(0.057)
Nonstop miles	0.0004***	0.0004***
	(2.06E-05)	(2.41E-05)
Extra miles	-0.3181***	-0.3591***
	(0.0362)	(0.0436)
Origin population	-9.36e-8***	-9.1e-8***
	(9.55E-10)	(1.10E-09)
Destination population	-9.2e-8***	-8.91e-8***
	(9.97E-10)	(1.16E-09)
AA	1.120***	1.177***
	(0.0164)	(0.0211)
DL	0.2468***	0.5687***
	(0.0089)	(0.0132)
LCC	2.009***	1.809***
	(0.0379)	(0.0466)
US	0.5236***	0.5106***
	(0.0161)	(0.0185)
WN	1.285***	1.253***
	(0.0172)	(0.0197)
Median own elasticity	-1.8457	-1.8495
Median cross elasticity	0.0024	0.0022
S.E. type	Standard	Standard
Observations	180,342	131,015
R2	0.39938	0.38538

Table 3. Regression outputs of two 2SLS specifications – the first omitting fuel costs from the set of instruments, and the latter including them.

	First step DML
Dependent Variable:	$\ln s_j - \ln s_0$
Price ($100)	-0.1762
	(0.0107)
	[-0.1997, -0.1550]
$\ln s_{j/g}$	0.8592
	(0.0073)
	[0.8468, 0.8745]
Median own elasticity	-3.3240
Median cross elasticity	0.1271
S.E. type	Bootstrapped
Observations	180,342

Table 4. First step DML estimates, obtained using 800 bootstrap replications. The estimation controls for all product features included in the 2SLS estimation presented above, and relies on all of the same instruments, including fuel costs. The square brackets present the range between the 2.5th and 97.5th percentiles of estimates.

	Second step DML
Dependent Variable:	$\ln s_j - \ln s_0 - \alpha p_j - \rho \ln s_{j/g}$
Extra miles	0.0026
	(0.0132)
	[-0.0231, 0.0278]
S.E. type	Bootstrapped
Observations	180,342

Table 5. Second step DML estimates, obtained using 800 bootstrap replications. The estimation controls for all product features apart from the price and "extra miles" itself, on which linearity is imposed with the purpose of obtaining an ATE. The square brackets present the range between the 2.5th and 97.5th percentiles of estimates.

	OLS of CATEs on standardized covariates
Dependent Variable:	$\hat{\tau}_{EM}(x)$
(Intercept)	0.3296***
	(0.0003)
Origin presence	-0.0168***
	(0.0006)
Destination presence	-0.0182***
	(0.0006)
Nonstop miles	0.0786***
	(0.0003)
Origin population	0.0684***
	(0.0003)
Destination population	0.0935***
	(0.0003)
DL	-0.0256***
	(0.0008)
LCC	-0.1055***
	(0.0006)
UA	-0.0179***
	(0.0008)
US	-0.0258***
	(0.0006)
WN	-0.0069***
	(0.0005)
S.E. type	Standard
Observations	165,666
R2	0.62385

Table 6. OLS regression of the CATE estimates of "extra miles" for connecting products on the remaining covariates, with the remaining covariates demeaned and scaled to have a variance of one. Thus, the sizes of the coefficients reflect the strength of the relationships between the CATE and these covariates.

Appendix C: Nested Logit Elasticity Formulas

General formula:

$$\textit{Own elasticity}: \quad \eta_{jj} = \alpha \cdot \frac{p_j}{1-\rho}\left(1 - s_j\left[\rho D_g^{\rho-1}\sum_g D_g^{1-\rho} + 1 - \rho\right]\right) \qquad (A.1)$$

$$\textit{Cross elasticity}: \quad \eta_{jk} = \begin{cases} -\alpha \cdot \dfrac{p_k}{1-\rho} \cdot s_k\left[\rho D_g^{\rho-1}\sum_g D_g^{1-\rho} + 1 - \rho\right] & \textit{for} \quad j \neq k \in g \\ -\alpha \cdot p_k \cdot s_k\left(\dfrac{D_g}{D_{g'}}\right)^{\rho} & \textit{for} \quad j \in g, k \in g', g \neq g' \end{cases} \qquad (A.2)$$

where, for a given nest g comprised of products $\mathcal{J}_g$, $D_g \equiv \sum_{j \in \mathcal{J}_g} e^{\frac{\delta_j}{1-\rho}}$.

Formula where (as in the airline application) all inside option products are in a single nest:

$$\textit{Own elasticity}: \quad \eta_{jj} = \alpha \cdot \frac{p_j}{1-\rho}\left(1 - s_j\left[\rho D_g^{\rho-1} + 1\right]\right) \qquad (A.3)$$

$$\textit{Cross elasticity}: \quad \eta_{jk} = -\alpha \cdot \frac{p_k}{1-\rho} \cdot s_k\left[\rho D_g^{\rho-1} + 1\right] \qquad (A.4)$$