Market Counterfactuals with Nonparametric Supply: An ML/AI Approach*

Harold Chiang† Jack Collison‡ Lorenzo Magnolfi§

Christopher Sullivan¶

March 21, 2025

Abstract

This paper develops a flexible approach to perform market counterfactuals using machine learning methods and nonparametric structure from economics. While standard structural methods rely on restrictive assumptions about firm conduct and cost, we propose a data-driven framework that relaxes these constraints when rich market data are available. Building on the identification results of Berry and Haile (2014), we develop a nonparametric model of supply that nests traditional conduct specifications while allowing for more complex competitive interactions. We adapt the Variational Method of Moments (VMM) (Bennett and Kallus, 2023) to estimate this flexible model, addressing the endogeneity of market shares and the high dimensionality of the problem. Our approach enables a wide range of counterfactual exercises including tax policy analysis, product regulation, and merger simulation. Monte Carlo simulations demonstrate that our method substantially outperforms standard approaches; applied to the American Airlines-US Airways merger, our method produces more accurate post-merger price predictions.

KEYWORDS: counterfactual analysis, variational method of moments (VMM), neural networks, merger simulation, airline markets

*We thank the audience at the Midwest IO Fest 2024 for helpful discussions. Any errors are our own. The authors declare no financial relationships or other potential conflicts of interest related to this article.

†Department of Economics, University of Wisconsin-Madison. Email: hdchiang@wisc.edu

‡Department of Economics, University of Wisconsin-Madison. Email: jcollison@wisc.edu

§Department of Economics, University of Wisconsin-Madison. Email: magnolfi@wisc.edu

¶Department of Economics, University of Wisconsin-Madison. Email: cjsullivan@wisc.edu

1 Introduction

Economists and policymakers routinely rely on market counterfactuals to evaluate potential mergers (e.g., Nevo, 2000), taxes (e.g., Miravete, Seim, and Thurk, 2018), product regulation (e.g., Dubois, Griffith, and O'Connell, 2018), and other market interventions. The standard approach to conducting these counterfactuals relies on parametric models of demand and supply to capture consumers' and firms' reactions to changes in policy. While recent advances have introduced more flexibility on the demand side (e.g., Compiani, 2022), supply-side modeling typically remains highly constrained by parametric assumptions about competition, often defaulting to models of Bertrand-Nash and constant marginal cost. These restrictive assumptions may lead to misleading predictions when the true model of firm conduct differs from the imposed structure.

This paper proposes a novel approach that maintains the economic structure of counterfactuals while relaxing assumptions about firm conduct and cost when sufficiently rich market data are available. Our model builds on the nonparametric identification framework of Berry and Haile (2014) but focuses specifically on developing a flexible supply-side specification. We model equilibrium prices as a function of market shares, demand derivatives, cost shifters, and ownership structure, allowing for a nonparametric relationship between these variables. This framework nests standard supply models like Bertrand-Nash competition with constant marginal costs while permitting more complex competitive interactions, including partial collusion, and richer cost functions, including economies of scale. Importantly, we maintain the core economic structure where prices emerge from equilibrium conditions while relaxing specific functional form assumptions about how markups and costs are determined. This approach requires addressing the inherent endogeneity of market shares, which we accomplish through appropriate instrumental variables motivated by economic theory. However, the high dimensionality of this problem poses significant estimation challenges that are hard to address with classical nonparametric methods.

We overcome these challenges by adapting recent advances in machine learning (ML) and artificial intelligence (AI), specifically the Variational Method of Moments (VMM) developed by Bennett and Kallus (2023). VMM combines deep learning methods with instrumental variables estimation through a novel min-max formulation of the moment conditions. This approach, by incorporating instruments, can thus address the simultaneity inherent in market equilibrium models, where prices and quantities are jointly determined. By parameterizing the supply function as a flexible neural network and optimizing over a growing class of functions, we can use VMM to estimate a flexible model that avoids restrictive parametric assumptions while maintaining the core economic structure of oligopoly competition.

To motivate the choice of VMM, we contrast it with conventional Nonparametric Instrumental Variable (NPIV) methods. While NPIV estimators are flexible in capturing complex relationships and possess well-established theoretical properties, they suffer from the curse of dimensionality, which limits their practicality in settings with multiple endogenous variables. Recent advances in neural network-based methods have demonstrated a notable capacity to adapt automatically to various function structures, particularly when the true function can be expressed as a composition of simpler, lower-dimensional components (e.g., Bauer and Kohler, 2019; Schmidt-Hieber, 2020; Kurisu, Fukami, and Koike, 2025), thereby mitigating the curse of dimensionality and achieving faster convergence rates. Moreover, in fixed-dimensional parametric settings, VMM coincides with the optimally weighted Generalized Method of Moments (GMM). A key innovation in our adaptation of VMM addresses a key aspect of our application: while Bennett and Kallus (2023) provides valid element-wise inference for one-dimensional nuisance parameters, our focus is on predicting counterfactual outcomes (e.g., prices), a more complex functional of these parameters. To address this, we develop an inference procedure that leverages the numerical delta method, Holm's step-down procedure, and a permutation-based implementation. This ensures computationally feasible standard errors and confidence intervals.

Our approach can address a wide range of market counterfactuals. Through extensive Monte Carlo simulations, we demonstrate the method's effectiveness in predicting counterfactuals. First, we simulate data under four distinct supply side models (combinations of Nash-Bertrand and profit weight models of conduct and constant and non-constant marginal costs) and show that in each setting, the trained flexible model can accurately predict prices out of sample outperforming predictions generated from models with misspecified conduct or cost. We further elucidate the source of our method's predictive ability, by comparing numerical approximations of the cost pass-through predicted by the flexible model to the true cost-pass through matrices. This exercise allows us to address a common criticism of ML/AI approaches: their "black box" nature. We show that our flexible approach is able to capture the underlying economic structure of firm conduct and cost by closely matching the pass-through matrix of the true model.

Then, motivated by empirical work, we consider four policy interventions: (i) predict price changes in response to regulation that increases marginal cost, (ii) predict consumption changes in response to regulation that alters product characteristics, (iii) predict changes in consumer welfare after product exit, and iv) predict government revenue at counterfactual ad valorem and unit tax levels. In each case, our flexible approach outperforms standard parametric models, especially when the true data-generating process deviates from Bertrand-Nash competition with constant marginal costs. Our simulations show that the method performs

well even with moderate sample sizes and when the exact counterfactual market structure is not observed in the training data, provided there is sufficient variation to inform prediction.

While our approach offers significant advantages in flexibility, it comes with important trade-offs that should guide its application. First, without additional restrictions, the method cannot separately identify markups from marginal costs – though this is an important goal in some applications,[1] it is unnecessary for many counterfactuals of interest. Second, like all nonparametric approaches, the method's reliability depends on the support of the training data; predictions for counterfactual scenarios that represent radical departures from observed market structures will be less reliable. Finally, the method places greater demands on exogenous identifying variation compared to standard parametric approaches (Magnolfi and Sullivan, 2022). In this sense, our method is complementary to testing approaches that use data to select among a menu of parametric conduct (and cost) models (Backus, Conlon, and Sinkinson, 2021; Duarte, Magnolfi, Sølvsten, and Sullivan, 2024; Dearing, Magnolfi, Quint, Sullivan, and Waldfogel, 2024).

To demonstrate the practical value of our approach, we apply it to the U.S. airline industry, focusing on the 2013 American Airlines-US Airways merger. The industry is well suited to illustrate our method given the rich variation in market structure both across routes and over time. Our flexible approach achieves better in-sample fit than the standard Bertrand-Nash model and, more importantly, produces post-merger price predictions that more closely match observed outcomes. Specifically, the flexible model offers a five-fold reduction in prediction MSE compared to the model of Bertrand-Nash pricing with constant marginal cost. These differences have implications for predicted changes in total consumer surplus arising from the merger. While the Nash-Bertrand model predicts a reduction in total consumer surplus of $40 million, our flexible model predicts a more modest reduction of $26 million.

Our method's computational requirements are manageable by modern standards, with estimation typically completed within hours on standard hardware. The approach is highly portable across empirical settings, requiring only market-level price and quantity data.[2]

This paper contributes to the IO literature that proposes nonparametric models of market equilibrium. We build our flexible model of supply on the identification results of Berry and Haile (2014), obtaining new results on the nonparametric identification of supply. Similarly to Compiani (2022), who develops a method to estimate demand nonparametrically,[3] our paper proposes a method for nonparametric estimation of the supply side. With similar

[1] Measuring markups is the primary goal of the "production" approach (see, e.g., De Loecker and Warzynski, 2012), which uses production functions and a cost minimization assumption.

[2] We provide detailed guidance and code for implementation in Appendix D.

[3] Other recent nonparametric approaches to demand estimation in markets for differentiated products include Tebaldi, Torgovitsky, and Yang (2023) and Brand and Smith (2025).

motivation, recent papers by Gandhi and Houde (2020) (proposing a linear approximation of the markup function) and Otsu and Pesendorfer (2024) (using the revelation principle) also develop methods to bring more flexible supply models to data. We complement these approaches by proposing a method that leverages advances in ML/AI coupled with a nonparametric structure.

A growing literature seeks to combine ML/AI methods with structure coming from economic theory. Farrell, Liang, and Misra (2020) provide a theoretical foundation for using deep learning in two-stage econometric procedures, allowing for rich heterogeneity while maintaining model interpretability. Kaji, Manresa, and Pouliot (2023) introduce "adversarial estimation," which formulates the estimator as a minimax problem between a generator and discriminator, achieving parametric efficiency under correct specification. Wei and Jiang (2024)'s Neural Net Estimator approach trains neural networks to directly estimate structural parameters from data moments. Similarly to these methods, we aim to leverage the flexibility and computational efficiency of machine learning while preserving the economic structure and interpretability of traditional models. We differ in the specific implementation and application: we formulate a nonparametric structural model of market equilibrium, and estimate it using the VMM formulation of Bennett and Kallus (2023).

Finally, we note that a growing recent literature focuses on applying general function approximation techniques—including deep neural networks and random forests—to instrumental variables problems within a unified framework (see, e.g., Hartford, Lewis, Leyton-Brown, and Taddy, 2017; Dikkala, Lewis, Mackey, and Syrgkanis, 2020; Lewis and Syrgkanis, 2018; Liao, Chen, Yang, Dai, Kolar, and Wang, 2020; Zhang, Imaizumi, Schölkopf, and Muandet, 2023). Our study is based on VMM for two main reasons. First, Bennett and Kallus (2023) derived its asymptotic distribution, which enables us to construct confidence intervals for the unknown supply-side function of interest. Second, because VMM subsumes optimally-weighted two-step GMM as a special case, it represents a natural evolution from the existing parametric estimation methods in IO.

The remainder of the paper is organized as follows. Section 2 presents the standard framework that serves as our benchmark. Section 3 develops our flexible approach and discusses identification. Section 4 describes the VMM estimation procedure and derives its statistical properties. Section 5 presents Monte Carlo evidence comparing our method to standard approaches. Section 6 applies our methodology to airline mergers and discusses the results. Section 7 concludes with a discussion of limitations and directions for future research.

2 Market Equilibria and Counterfactuals

We develop in this section a general model of equilibrium for a market for differentiated products, which we use as our data generating process. We then proceed to give a general description of market counterfactuals, followed by an example that foreshadows our application: merger simulation.

Data-generating Process: We observe data on a set of products $\mathcal{J}$ offered by firms across a set of $\mathcal{T}$ markets. The researcher observes prices p_{jt}, market shares s_{jt}, a vector of product characteristics x_{jt} that enter demand, and a vector of cost shifters w_{jt} for each product-market pair (j,t). For any variable y_{jt}, we denote y_t as the vector of values in market t. We assume, across all markets, that the true demand system is given by:

$$s_t = \mathcal{s}(p_t, x_t, \xi_t).$$

The term ξ_t is a vector of unobservable product characteristics. Data on endogenous prices and quantities are generated by equilibrium in market t. On the supply side, firm behavior is characterized by a set of first-order conditions for the firms' profit maximization problems:

$$p_t = \Delta_{0t} + c_{0t}, \tag{1}$$

where $\Delta_{0t} = \Delta_0(s_t, p_t, D(s_t, p_t, x_t); \mathcal{H}_t)$ is the vector of true markups, $D_t = \frac{\partial s_t}{\partial p_t}$ is the matrix of demand derivatives, and c_{0t} is the vector of true marginal costs in market t. Firms' costs are generated by some cost function c, or $c_{0jt} = c(q_t, \mathrm{w}_{jt}, \omega_{0jt})$, where w_t and ω_{0t} are, respectively, vectors of observed and unobserved cost-shifter variables. Therefore, a model of conduct, which determines the form of the markup function $\Delta_0(\cdot)$, and a model of how costs are determined are the key ingredients of the supply-side model. The canonical case in the literature is costs that are constant in quantities, but dependence on q_t allows for the presence of economies of scale and scope.

Example 1: A canonical assumption in the literature is that the true model of conduct is Bertrand-Nash, whereby firms are playing a complete information pricing game, with constant marginal cost that is a linear index of cost shifters. Under this assumption, the true markup function is $\Delta_0(p_t, s_t) = (\mathcal{H}_t \odot D_t)^{-1} s_t$, where $\mathcal{H}_t$ is the ownership matrix. Furthermore, the true cost function becomes $c_{0t} = \mathrm{w}_t\gamma + \omega_{0t}$.

To summarize the environment, we have described three functions that pin down the

structure of demand (ϑ) and supply (Δ_0 and c). The endogenous observables (p_t, s_t) are generated by these primitives as a function of exogenous observables (x_t, w_t) and unobservables (ξ_t, ω_t). This formulation nests the standard Bertrand-Nash model but allows for more general forms of strategic interaction. With respect to the framework in Section 4.4 in Berry and Haile (2014), the expression above unpacks marginal revenue as the sum of price and markup, which entails no loss of generality. We further maintain two assumptions also imposed in Berry and Haile (2014):

Assumption 1. *(Equilibrium Selection)* There exists a unique equilibrium, or the equilibrium selection rule is such that the same p_t arises whenever the vector $(\mathrm{w}_t, x_t, \omega_t, \xi_t)$ is the same.

This assumption, similar to Assumption 13 in Berry and Haile (2014), ensures that observed prices reflect stable equilibrium behavior.

Assumption 2. *(Separability of Cost)* The cost function is separable in unobservable shocks:

$$c(s_t, \mathrm{w}_t, \omega_t) = \bar{c}(s_t, \mathrm{w}_t) + \omega_t.$$

This separability assumption follows Berry and Haile (2014) and is essentially without loss of generality, as the unobservable component ω_t can be defined as the residual between total costs and the component explained by observables.

To identify our model of supply, we will assume that demand is identified. This mirrors the two-step procedure in standard merger simulation, where demand is first estimated or calibrated.

Assumption 3. *(Known Demand)* The matrix of demand derivatives is known, so that $D_t = D(s_t, p_t, x_t)$ is observed.

We further restrict the class of models we consider, so that prices enter markups only through demand derivatives:

Assumption 4. *(Markup Dependence)* The markup function Δ depends only on market shares s_t and the matrix of demand derivatives D_t.

Assumption 4 is satisfied by a broad range of conduct models beyond Bertrand-Nash, including Cournot competition, Stackelberg leadership, various forms of partial collusion, and models where firms maximize weighted combinations of profits and consumer surplus.[4]

[4]For a thorough discussion of how the models above (and more) satisfy this assumption, see Appendix C in Dearing et al. (2024).

Market counterfactuals: Most policy changes of interest in applied research can be modeled as *market counterfactuals*. We will predict counterfactuals for a specific market (period) $\tilde{t}$ outside of the sample; for simplicity, we suppress market indexes in what follows and use tildes to denote counterfactual objects.

A market counterfactual must involve a change in either demand or supply functions, or any of their exogenous arguments. For instance, exogenous product characteristics may now take a value $\tilde{x}$, or cost functions be altered to $\tilde{c}$. Under the new primitives, consumers and firms react to exogenous changes and counterfactual market outcomes arise from a new market equilibrium $(\tilde{p}, \tilde{s})$. These outcomes need to satisfy demand and supply equations, and can be found as the fixed point to:

$$\tilde{p} = \tilde{\Delta}(\tilde{p}, \tilde{\jmath}(\tilde{p}, \tilde{x}, \tilde{\xi})) + \tilde{c}(\tilde{q}, \tilde{\mathrm{w}}, \tilde{\omega}).$$

Counterfactuals of interests can then be expressed as a function $F(\tilde{p}, \tilde{s}; \cdot)$, where we suppress the dependence on (counterfactual) structural objects and exogenous variables. To illustrate the general notion of a market counterfactual, we consider next the standard merger simulation problem.

Example 1 (continued): Suppose the researcher wants to predict prices following a horizontal merger. Further suppose that the true model of conduct is Bertrand-Nash; other assumptions are possible, including Cournot-Nash, Nash Bargaining, or profit-weight models. As a baseline case, the merger results only in a deterministic change to the ownership matrix $\tilde{\mathscr{H}}$ in our counterfactual market $\tilde{t}$. If the researcher is only interested in predicting counterfactual $(\tilde{p}, \tilde{s})$, the map F is the identity. Otherwise, the map F could return other objects of interest such as counterfactual consumer surplus and firm profit.

Estimating a Counterfactual: Evaluating F, requires knowledge of the (counterfactual) primitives, exogenous observables, and unobservables. In practice, researchers use a combination of data and assumptions. Typically, the functions $\jmath$, and c are estimated with parametric models, and Δ_0 is assumed. This allows the researcher to obtain estimates of in-sample ξ_t and ω_t, which are used to specify counterfactual $\tilde{\xi}$ and $\tilde{\omega}$. Counterfactual $\tilde{\jmath}, \tilde{c}$ and $\tilde{\Delta}$ are either kept fixed, or changed deterministically.

Example 1 (continued): The standard merger simulation toolkit typically proceeds in three steps. First, the researcher formulates a parametric demand model $\jmath(\cdot; \theta^D)$ and estimates demand primitives $\hat{\theta}^D$, which yield demand derivatives $\hat{D}_t$ and $\hat{\xi}_t$ in each market t. Second,

the researcher assumes a supply model, which encompasses models of cost and markups. In the standard toolkit, these are typically a constant marginal cost function, and Nash-Bertrand conduct. Marginal costs can thus be inverted as $\hat{c}_t = p_t - \left(\mathcal{H}_t \odot \hat{D}_t\right)^{-1} s_t$. Third, the researcher computes post-merger prices $\tilde{p}$ under the post-merger ownership structure $\tilde{\mathcal{H}}$ holding everything else fixed. Post-merger prices solve a fixed point:

$$\tilde{p} = \hat{c} + \left(\tilde{\mathcal{H}}_t \odot D(\tilde{p}, x, \hat{\xi}; \hat{\theta}^D)\right)^{-1} \jmath(\tilde{p}, x, \hat{\xi}; \hat{\theta}^D). \tag{2}$$

The same procedure applies to alternative assumptions on cost and conduct. For instance, assumptions on cost efficiencies generated by the merger, quantified in post-merger cost vectors $\tilde{c}$, can be incorporated by using this vector instead of $\hat{c}$ in Equation (2).

When performing market counterfactuals in applications, researchers trade off practicality and data limitations, which suggest tight parametric specifications, with the dangers of misspecification, leading to misleading conclusions. Compiani (2022), building on the nonparametric demand framework in Berry and Haile (2014), shows how to nonparametrically estimate demand, and what the consequences of restrictions on demand are for counterfactuals. Complementary to these efforts, we propose an approach to flexibly estimating supply to relax the parametric restrictions on costs and markups when performing market counterfactuals.

3 Flexible Supply for Market Counterfactuals

In this section, we develop a flexible model of supply that maintains the core economic structure of market equilibrium while relaxing standard assumptions about firm conduct and cost.

3.1 A Flexible Supply Function

Under Assumptions 1-4, we can write the price equation as:

$$p_t = h(s_t, D_t, \mathrm{w}_t; \mathcal{H}_t) + \omega_t, \tag{3}$$

where $h(\cdot) \equiv \Delta_0(\cdot) + \bar{c}(\cdot)$ is a flexible function that captures both markup and cost components. Notably, this formulation does not impose separability between markups and costs. This is because we want to pursue a nonparametric approach to the supply side, and given the presence of s_t in both Δ_0 and $\bar{c}$, it is not possible to jointly identify these functions

without further restrictions. While we could estimate nonparametrically markups under a parametric specification of costs, or vice versa, our composite supply function allows us to avoid parametric restrictions on either object.

Example 1 (continued): Consider again the case of Bertrand-Nash conduct, but now with (linear) economies of scale in cost. Notice that the markup function $\Delta_0 = (\mathcal{H}_t \odot D_t)^{-1} s_t$ is a function of s_t, D_t and $\mathcal{H}_t$. The marginal cost of product j is given by $\bar{c}_{jt} = \mathrm{w}_{jt}\gamma + M_t s_{jt}\gamma_{ES}$. Thus, their sum can be represented by the flexible supply function $h(s_t, D_t, \mathrm{w}_t; \mathcal{H}_t)$. Suppose the researcher wants to use this function for a merger simulation counterfactual. Then, assuming the researcher knows or estimates the demand function $\hat{\jmath}$, and that the function h is identified, she can estimate h and the unobserved shocks to marginal cost. She can then find counterfactual prices and shares as a fixed point to the function $\hat{h}$:

$$\tilde{p}_t = \hat{h}(\hat{\jmath}(\tilde{p}_t), \hat{D}(\tilde{p}_t), \mathrm{w}_t; \tilde{\mathcal{H}}_t) + \hat{\omega}_t,$$

where $\tilde{\mathcal{H}}_t$ represents the post-merger ownership structure and $\hat{\omega}_t$ are the estimated residuals. This maintains the equilibrium structure of standard merger simulation while allowing for richer firm conduct and cost.

Before discussing identification, we highlight that our approach based on a flexible supply function has consequences on the range of counterfactuals that can be tackled with it. The model accommodates several important classes of counterfactuals. First, the framework naturally handles changes in ownership structure, such as those arising from mergers or product exit, by modifying ownership matrices or setting market shares to zero. Second, the model can evaluate tax or subsidy policy analysis. For example, unit tax changes can be incorporated as cost shifters in the supply relation (we illustrate unit and ad valorem tax counterfactuals in Section 5). Moreover, the model can also assess impacts of regulations that alter product characteristics or observable cost shifters. Examples include fuel economy standards or safety requirements that simultaneously affect both demand and cost structures. We show examples of counterfactuals in all these classes in our simulations in Section 5.

For all these counterfactuals, we can measure equilibrium prices, quantities, and changes in consumer welfare (and, in the case of a tax, government revenue and incidence). However, an important limitation of our approach is that we cannot separately identify the levels of markups and costs, which in turn prevents measurement of firm profits. This also prevents counterfactuals where the cost or markup functions are altered in a known way (e.g., if a merger leads to coordinated effects, or a change in the degree of scale economies). Additional

economic restrictions might enable such measurements (see Appendix B on regularization and economic assumptions).

This discussion highlights how the method is complementary to other tools in the IO literature. When the main goal of the analysis is to measure markups, or learn the true model of conduct, either additional restrictions or parametric methods such as those in the production approach (De Loecker and Warzynski, 2012) or in the testing-demand approach (Backus et al., 2021; Duarte et al., 2024), can be used.

3.2 Identification

Equation (3) expresses equilibrium prices in a market as a function of endogenous market shares. Therefore, the identification of the flexible supply model must rely on instrumental variables. We specify a moment condition:

$$\mathbb{E}[\omega_{jt} \mid z_{jt}, \mathrm{w}_{jt}] \quad = \quad \mathbb{E}[p_{jt} - h_{jt}(\cdot) \mid z_{jt}, \mathrm{w}_{jt}] \quad = \quad 0,$$

where z_{jt} is a vector of instruments. Following standard practice in the literature, candidate instruments include own and rival product characteristics, rival cost shifters, and taxes (Berry and Haile, 2014; Dearing et al., 2024). Importantly, the simultaneity of the environment, where the demand equation $s_t = \mathcal{s}(p_t)$ expresses quantities as a function of prices, makes it necessary to maintain exclusion restrictions to identify h_t.

To see why these are necessary, consider the case of simple logit demand where price and characteristics enter demand through a linear index $\delta_{jt} = \alpha p_{jt} + \beta x_{jt} + \xi_{jt}$. Under the Berry (1994) inversion, $\delta_{jt} = \log s_{jt} - \log s_{0t}$, so that inverse demand is $p_{jt} = \frac{\log s_{jt} - \log s_{0t}}{\alpha} - \frac{\beta x_{jt} + \xi_{jt}}{\alpha}$. A concern then is that our flexible strategy would back out this inverse demand function, instead of the supply relation h_t. Given that Equation (3) is restricted to be a function of s_t, D_t, and w_t, an exclusion restriction that prevents our procedure from identifying inverse demand is as follows:

Assumption 5. *(Instrument Exogeneity and Exclusion)* The vector of instruments z_{jt} that satisfies

$$\mathbb{E}[\omega_{jt} \mid z_{jt}, \mathrm{w}_{jt}] = 0$$

contains demand shifter(s) $x_{jt}^{(e)}$ that are excluded from the vector w_{jt}.

Under this exclusion, in the logit example above, the moment condition would be violated for $h = \mathcal{s}^{-1}$ as the implied ω_{jt} explicitly depends on $x_{jt}^{(e)}$. We also maintain, similar to Berry and Haile (2014), a completeness assumption on the instruments.

Assumption 6. *(Completeness)* For all functions $B(s_t, D_t, \mathrm{w}_t; \mathcal{H}_t)$ with finite expectation, if $\mathbb{E}[B(s_t, D_t, \mathrm{w}_t; \mathcal{H}_t) \mid z_{jt}, \mathrm{w}_{jt}] = 0$ almost surely, then $B(s_t, D_t, \mathrm{w}_t) = 0$ almost surely.

Under these assumptions, we can prove that the function h_j is identified:

Theorem 1. *Under Assumptions 1-6, the function h_j is identified for all $j = 1, ..., J$.*

Proof. The proof follows Theorem 1 in Berry and Haile (2014). For any j, for the true h_j function we have that at all fixed values of $(z_{jt}, \mathrm{w}_{jt})$:

$$\mathbb{E}[\omega_{jt} \mid z_{jt}, \mathrm{w}_{jt}] = \mathbb{E}[h_j(s_t, D_t, \mathrm{w}_t; \mathcal{H}_t) \mid z_{jt}, \mathrm{w}_{jt}] - \mathbb{E}[p_{jt} \mid z_{jt}, \mathrm{w}_{jt}]$$
$$\mathbb{E}[p_{jt} \mid z_{jt}, \mathrm{w}_{jt}] = \mathbb{E}[h_j(s_t, D_t, \mathrm{w}_t; \mathcal{H}_t) \mid z_{jt}, \mathrm{w}_{jt}],$$

where the first equality follows by Assumption 5, and $\mathbb{E}[p_{jt} \mid z_{jt}, \mathrm{w}_{jt}]$ is identified by the data. Suppose that the function $\tilde{h}_j$ is also such that:

$$\mathbb{E}[p_{jt} \mid z_{jt}, \mathrm{w}_{jt}] = \mathbb{E}[\tilde{h}_j(s_t, D_t, \mathrm{w}_t; \mathcal{H}_t) \mid z_{jt}, \mathrm{w}_{jt}],$$

so that

$$\mathbb{E}[h_j(s_t, D_t, \mathrm{w}_t; \mathcal{H}_t) \mid z_{jt}, \mathrm{w}_{jt}] - \mathbb{E}[\tilde{h}_j(s_t, D_t, \mathrm{w}_t; \mathcal{H}_t) \mid z_{jt}, \mathrm{w}_{jt}] = 0 \qquad a.s.$$

By Assumption 6, it follows that the function $B = h_j - \tilde{h}_j = 0$ almost surely, which in turn implies that h_j is identified. Note that Assumption 5 also rules out that the function $\tilde{h}_j$ could be the inverse demand function. □

It is useful to draw a parallel between our result in Theorem 1 and the identification results in Section 4.4 of Berry and Haile (2014). While they establish the identification of the cost shocks without imposing a supply model, we show identification of the full supply function. This is because our assumptions go beyond Assumption 7b in Berry and Haile (2014) in several important ways. First, we impose a functional form on marginal revenue in Equation (1), and unpack marginal revenue as the sum of price and markups. Second, we allow prices to only impact markups through shares and demand derivatives (Assumption 4). Third, we assume that demand derivatives are known (Assumption 3). These assumptions come at a cost in terms of generality but are fit for our purposes and are satisfied by the standard conduct models considered in most of the empirical literature. In particular, the models in Remark 1 of Berry and Haile (2014) that satisfy their assumptions 7a and 7b, also satisfy our assumptions 1-6.

While Theorem 1 establishes identification of our flexible model, estimation poses significant challenges. Standard nonparametric instrumental variables techniques are likely to

perform poorly in finite samples due to the curse of dimensionality. We discuss in section 4 how we overcome these challenges using recent advances in ML/AI.

4 Estimation and Inference

We now turn to the estimation of the flexible markup function via VMM. While alternative NPIV estimators (e.g., Chen 2007; Carrasco, Florens, and Renault 2007) offer flexibility in modeling and robust asymptotic properties, they are limited by the curse of dimensionality, which restricts their practical application to low-dimensional settings. In contrast, neural network-based IV estimation methods, such as VMM, address this issue by using parametric models that grow in complexity with sample size. VMM also coincides with the optimally weighted Generalized Method of Moments (GMM) when the unknown parameter is of fixed, finite dimensionality (Lemma 1 in Bennett and Kallus 2023), making it appealing in our setting where previous approaches to estimating parametric classes of supply models have relied on GMM-based estimators. We further compare VMM to other NPIV approaches at the end of this section.

We now formally explain our VMM-based estimation procedure. Explicitly, we use a moment condition for structural markup:

$$\mathbb{E}[p_{jt} - h_j(s_t, D_t, w_t; \theta, \mathcal{H}_t) \mid z_t, w_t] = 0$$

Given a preliminary consistent estimate $\tilde{\theta}_N$ estimator solves a min-max program:

$$\hat{\theta}_N = \text{argmin}_{\theta\in\Theta}\sup_{f\in\mathcal{F}_N} \frac{1}{TJ}\sum_{j,t} f(z_{jt})'\omega_{jt}(\theta) - \frac{1}{4TJ}\sum_{j,t}(f(z_{jt})'\omega_{jt}(\tilde{\theta}_N))^2 - R_N(f,h) \quad (4)$$
$$\text{where} \quad \omega_{jt}(\theta) = p_{jt} - h_j(s_t, D_t, w_t; \theta, \mathcal{H}_t)$$

Following the identification arguments in Section 3, the estimator includes a vector of observable cost shifters w_{jt} and a vector of instruments z_{jt}.[5] The instruments are necessary to address the endogeneity of market shares s_t and demand derivatives D_t. In general, $f \in \mathcal{F}_N$ and $h \in \mathcal{H}_N$ belong to sequences of function classes. We use classes of neural networks with growing width and depth, allowing flexible controls of model complexity to cope with the curse of dimensionality. R_N is a regularizer $R_N : \mathcal{F}_N \times \mathcal{H}_N \to [0, \infty]$ that penalizes the complexity of the neural network; we note in Appendix B that regularization can also

[5] In practice, we exclude observable cost shifters from the set of instruments. Given that the function f is highly nonlinear, the cost shifters no longer explicitly instrument for themselves. Simulations reveal better estimates upon the exclusion of w from the set of instruments.

be used to adhere to economic properties of the markup function. θ is a potentially large set of parameters that pins down the function h. Under standard regularity assumptions, then Theorem 4 in Bennett and Kallus (2023) shows that the estimator in Equation (4) is consistent for the true parameter values θ_0.

4.1 Inference

We now turn to the quantification of uncertainty in the flexible markup function. Importantly, while VMM offers valid element-wise inference procedures for the underlying nuisance parameters, our focus is on predicting counterfactual prices – a complex functional of these parameters. To address this, we develop an inference method that employs the numerical delta method in conjunction with Holm's step-down procedure and a permutation procedure. This approach enables the quantification of uncertainty in counterfactual price predictions through standard errors and confidence intervals, ensuring computational feasibility. If $\tilde{\theta}_N \xrightarrow{p} \theta_0$, under regularity conditions, Theorems 2-3 in Bennett and Kallus (2023) imply:

$$\sqrt{N}(\hat{\theta}_N - \theta_0) \xrightarrow{d} N(0, \Omega_0^{-1})$$
$$\Omega_0 = \mathbb{E}\left[\mathbb{E}[\nabla_\theta \omega(\theta_0) \mid z, w]' \mathbb{E}[\omega(\theta)\omega(\theta)' \mid z, w]^{-1} \mathbb{E}[\nabla_\theta \omega(\theta_0) \mid z, w]\right]$$

For inference on the markup function $h : \mathbb{R}^b \supset \Theta \to \mathbb{R}^d$ at a set of d observations, the numerical delta method in Hong, Mahajan, and Nekipelov (2015) implies:

$$\sqrt{N}(h(\hat{\theta}_N, \mathcal{H}) - h(\theta_0, \mathcal{H})) \xrightarrow{d} N(0, \nabla_{\theta'} h(\theta_0, \mathcal{H}) \Omega_0^{-1} \nabla_{\theta'} h(\theta_0, \mathcal{H})') \tag{5}$$

Note that $\nabla_\theta h(\theta_0; \mathcal{H})$ has dimension $d \times b$. In the simplest case, suppose that $d = 1$ and the parameter of interest is $h_x(\theta_0, \mathcal{H})$. Lemma (9) in Bennett and Kallus (2023) states that, for any $\beta \in \mathbb{R}^b$, we have:

$$\beta^T \Omega_0^{-1} \beta = -\frac{1}{4} \inf_{\gamma \in \mathbb{R}^b} \sup_{f \in \mathcal{F}} \mathbb{E}[f(z)' \nabla_\theta \omega(\theta_0)\gamma] - \frac{1}{4}\mathbb{E}[(f(z)'\omega(\theta_0))^2] - 4\gamma'\beta - R_N(f) \tag{6}$$

Taking $\beta = \nabla_\theta h_x(\theta_0, \mathcal{H})$ for x, the solution to the optimization problem above yields the asymptotic variance in Equation (5). The gradient $\nabla_\theta h_x(\theta_0, \mathcal{H})$ is difficult to compute analytically but numerical differentiation can be employed, e.g., Hong et al. (2015), and $\hat{\theta}_N$ can be used in place of θ_0. For numerical differentiation to be valid, the ϵ in the difference has to satisfy $\epsilon \to 0$ and $N\epsilon / \log N \to \infty$ from Theorem 1 in Hong et al. (2015).[6]

[6] In practice, we use automatic differentiation in `torch`. Broadly speaking, the module stores a computational graph as the neural network is fit; training requires and stores the gradients in backpropagation. We

However, this approach cannot obtain a covariance matrix when $d \geq 2$. We extend the method to provide a simultaneous confidence interval by adapting Holm's Step-Down procedure (Holm, 1979) with the estimates for $\hat{\sigma}^2_{x_j}(\hat{\theta}, \mathcal{H})$ and $h_{x_j}(\hat{\theta}, \mathcal{H})$ for each $j = 1, ..., d$. The set of critical values T_α is known for significance levels $\frac{\alpha}{d+1-k}$ with $k = 1, ..., d$. For any ordering of x and fixed ordering T_α, we can compute the (vectorized) confidence interval $h_x(\hat{\theta}, \mathcal{H}) \pm N^{-\frac{1}{2}} \hat{\sigma}_x(\hat{\theta}, \mathcal{H}) T_\alpha$. To implement the procedure, we construct the intervals for all permutations of $j = 1, ..., d$, resulting in $d!$ permutations of x. This is because we must consider any possible ordering of the p-values of $x_1, ..., x_d$. The simultaneous confidence interval is subsequently the union of the bounds in each of the $d!$ permutations. Formally, the proposed procedure is as follows:

Algorithm 1 Simultaneous Confidence Interval for VMM

1: **for** each index $j \in \{1, ..., d\} \equiv J$ **do**
2: Estimate $\hat{\sigma}^2_{x_j}(\hat{\theta}, \mathcal{H})$ for $\sigma^2_{x_j}(\theta_0, \mathcal{H})$ at x_j by solving Equation 6
3: **end for**
4: Fix critical values $T_\alpha = \left\{T_{\alpha_k} : k = 1, ..., d\right\}$ where $\alpha_k = \frac{\alpha}{d+1-k}$
5: **for** each permutation $\tilde{J}$ of indices J **do**
6: Arrange values $\tilde{x}$ and $\hat{\sigma}_{\tilde{x}}$ using permuted indices $\tilde{J}$
7: Construct bounds as $h_{\tilde{x}}(\hat{\theta}, \mathcal{H}) \pm N^{-\frac{1}{2}} \hat{\sigma}_{\tilde{x}}(\hat{\theta}, \mathcal{H}) T_\alpha$ for fixed T_α
8: **end for**
9: Collect the simultaneous confidence interval as the union of bounds from Step (5)-(8)

This novel procedure allows us to quantify uncertainty simultaneously across many predicted prices. Importantly, we can construct confidence intervals on counterfactuals using our method.

4.2 Further Discussion of VMM

We now further elaborate on our choice of VMM as our estimator of choice, comparing it to Nonparametric Instrumental Variable (NPIV) methods. NPIV estimators are well-documented (e.g., Chen 2007; Carrasco et al. 2007) and offer flexibility in modeling complex relationships without restrictive parametric assumptions. Their theoretical foundations provide robust asymptotic properties under appropriate conditions, making them appealing for situations where the functional form of structural equations is unknown.

However, conventional NPIV methods are limited by the curse of dimensionality, which restricts their practical application to low-dimensional settings. In contrast, neural network-

note that we are still subject to the regularity condition from Hong et al. (2015).

based IV estimation methods, such as VMM and the average derivative estimators presented by Chen, Chen, and Tamer (2023) address this issue by using parametric models that grow in complexity with sample size. VMM retains conventional NPIV's strengths, such as handling complex variable relationships, while mitigating dimensionality challenges. Neural networks differ from traditional sieve methods in that their architecture enables them to automatically adapt to composite function structures, such as those arising from single-index models. In these cases, the network effectively decomposes a high-dimensional estimation problem into a series of lower-dimensional tasks, thereby mitigating the curse of dimensionality. As a result, deep neural networks can achieve substantially faster convergence rates when the target function possesses a compositional form. This advantage is theoretically substantiated in e.g. Bauer and Kohler (2019); Schmidt-Hieber (2020). In addition, VMM coincides with the optimally weighted Generalized Method of Moments (GMM) when the unknown parameter is of fixed, finite dimensionality (Lemma 1 in Bennett and Kallus 2023). This underscores VMM's role as a direct generalization of traditional parametric estimators, adaptable to high-dimensional contexts.

Beyond our specific method, we note that machine learning methods are increasingly being integrated in economics, offering novel approaches to address challenges in high-dimensional and complex data environments. For example, Kaji et al. (2023) push the frontier by introducing adversarial estimation, a GAN-inspired method that bridges simulation-based estimation techniques such as the Simulated Method of Moments and Maximum Likelihood Estimation. Their framework achieves parametric efficiency while allowing flexible moment selection and robustness under model misspecification. Fernández-Villaverde, Nuño, and Perla (2024) apply deep learning to quantitative economic models, illustrating how deep neural networks can mitigate the curse of dimensionality and enhance the solution of dynamic equilibrium models through efficient state space representations. These advancements highlight the potential of ML in economics research, enabling researchers to tackle more complex models and extract richer insights from data.

5 Monte Carlo Simulations

We illustrate the performance of our flexible model with simulations designed to contrast it to the standard toolkit where the researcher imposes a potentially misspecified model. In each of the data environments we develop, we use both the flexible model estimated with VMM and the standard IO toolkit to predict prices in a hold-out sample: the flexible model performs similarly to the IO toolkit when the assumptions of the toolkit are valid and outperforms the toolkit when either cost or conduct are misspecified. Comparing the pass-

through matrices implied by the flexible model to the true ones, we show that the flexible model is able to learn the true structural relationship on the supply side.

Then, motivated by standard empirical exercises in the economics literature, we illustrate the superior performance of our flexible model in predicting counterfactual outcomes relative to the standard IO toolkit. We consider four experiments: first, we "regulate" the level of an observed cost shifter and predict prices. Second, we "regulate" the level of an observed product characteristic and predict the quantity consumed. Third, we simulate product exit and predict consumer welfare. Finally, we compute government revenue at various levels of both a unit and an ad-valorem tax. For all four experiments, our approach performs similarly to the IO toolkit when it makes the correct assumptions on cost and conduct but outperforms the IO toolkit when cost or conduct are misspecified – this occurs even in relatively small sample sizes, and when the counterfactual is outside of the support of the data.

5.1 Simulation Environment

We consider price-competition between firms across multiple markets, where market size is normalized to one in each market. As our focus is on the supply side, we adopt a simple logit model of demand. Consumer i receives indirect utility from consuming product j in market t, given by:

$$u_{ijt} = \alpha p_{jt} + x_{jt}\beta + \xi_{jt} + \varepsilon_{ijt},$$

where x_{jt} is a vector containing a constant and two observed product characteristics, p_{jt} is the price of the product, and ξ_{jt} and ε_{ijt} are unobservable shocks at the product-market and individual-product-market-level, respectively. The utility of the outside option is normalized to $u_{i0t} = \varepsilon_{i0t}$.

To contrast our flexible approach to the standard prediction method in IO, we simulate data under four different supply models: the canonical Nash-Bertrand pricing with a constant marginal cost model, a model of Nash-Bertrand pricing with economies of scale, a profit weight model with constant marginal costs, and a profit weight model with economies of scale. These four environments allow us to show that the flexible model can perform similarly to the IO toolkit when assumptions on cost and conduct match the truth, but can outperform the toolkit when either conduct, cost, or both, are misspecified.

Stacking first-order conditions, the true markups in each of the supply models in each market t can be represented as:

$$(1 - \nu_t)p_t - c_t - \tau_t = \left(-\mathcal{H}_t \odot D_t'\right)^{-1}(1 - \nu_t)s_t. \tag{7}$$

Here, ν_t and τ_t respectively represent a market-wide ad valorem tax (levied on consumers) and unit tax (levied on firms). While the own-product elements of the ownership matrix $\mathcal{H}_t$ are ones across all four models, the rival-product elements allow us to control the true model of conduct generating the data. We set these elements to zero when the model of conduct is Nash-Bertrand pricing and 0.5 for the profit weight model. We further assume that the marginal cost of producing product j in market t is given by:

$$c_{jt} = \mathrm{w}_{jt}\gamma + g(s_{jt}) + \omega_{jt}$$

where w_{jt} contains a constant and two observable cost shifters that are excluded from demand and ω_{jt} is the true unobserved cost shock. The function $g(s_t)$ allows us to control economies of scale. When marginal cost is constant, we set $g(s_t) = 0$. Instead, to simulate data with economies of scale, we set $g(s_{jt}) = \mu + \lambda_0 s_{jt} + \lambda_1 s_{jt}^2$,

Data Generation: Under each of the four supply models described above, we simulate data using the `pyblp` framework from Conlon and Gortmaker (2020) for $T = 100,\ 1,000,$ and $10,000$ markets. In each market t, the number of single-product firms J_t is randomly assigned with equal probability to either $J_t = 2$ or $J_t = 3$ for duopolies and triopolies, respectively.

We draw the two observed product characteristics and the two observed cost shifters independently from normal distributions $N(1, 0.25)$. For the unobservable shocks we adopt the default in `pyblp` and draw ξ_{jt} and ω_{jt} from a mean-zero bivariate normal distribution with variances of one and correlation of 0.9. As is standard, ε_{ijt} is assumed to be distributed Type I Extreme Value. For the demand system, we set the mean taste parameters as $\beta = [-4, 3, 6]$ and the price coefficient as $\alpha = -0.25$.

The cost parameters on the observed cost shifters are set to $\gamma = [3, 6, 4]$ in all specifications. When we impose economies of scale in the data generating process, we set the parameters in the function g to $\lambda_0 = 20$, $\lambda_1 = -30$, and $\mu = -2$. With scale economies, marginal costs have a parabolic shape in market shares, with costs increasing in shares for smaller market shares and then decreasing after the threshold of $s_{jt} = 0.5$. For most of our simulations, we set the value of the ad valorem and unit taxes to zero. However, for exercises where we want to predict the Laffer curve, we impose taxes on the data generating process. Specifically, we draw the market wide unit taxes from the uniform distribution $U(4, 8)$ and the ad valorem taxes from the uniform distribution $U(0, 0.8)$.

Under each of the four true supply models, prices solve the fixed point of Equation (7). The Bertrand-Nash and profit-weight simulations (without taxes) and with constant

marginal cost yield mean own-price elasticities -3.29 and -4.32, and mean diversions to the outside option of 0.34 and 0.50, respectively. For each of the datasets we generate, we randomly sample 80% of the markets and include all observations from these in a training sample; the observations from the remaining 20% of the markets represent a hold-out sample. Splitting our data in this fashion allows us to prevent overfitting of the flexible model and to evaluate out-of-sample performance in the hold-out sample.

Fitting IO Toolkit/Estimating Flexible Model: As our goal is to make progress on the supply side, we assume the researcher knows the true demand system. Then on each training set, we first follow the standard IO toolkit and fit assumed supply models m to recover implied cost shocks ω_{jt}^m. Specifically, we consider three different canonical models of conduct: Bertrand ($m = B$), monopoly (joint profit maximization, $m = M$), and perfect competition (marginal cost pricing, $m = P$) with and without economies of scale. Under each model m, we can compute the implied markups Δ^m and recover marginal cost shocks ω^m as the residuals to the regression $(1 - \nu_t)(p_{jt} - \Delta_{jt}^m) - \tau_t = w_{jt}\gamma + g(s_t) + \omega_{jt}^m$. For Bertrand-Nash and monopoly, $\Delta_t^m = \left(-\mathcal{H}_t^m \odot D_t'\right)^{-1} s_t$, where $\mathcal{H}_t^B = I_t$ under Bertrand and $\mathcal{H}_t^B$ is a matrix of ones under monopoly. Under perfect competition, we set markups $\Delta_{jt}^p = 0$.

For our flexible model , we estimate the function h:

$$(1 - \nu_t)p_{jt} - \tau_t = h_j(s_t, D_t, w_{jt}, \nu_t; \mathcal{H}_t) + \omega_{jt}$$

using the VMM estimator defined in Equation (4), which yields the implied shocks $\hat{\omega}$. The neural networks are implemented using the `torch` package in Python. We do not include regularization in our baseline simulations. The variance estimates use the objective function from Equation (6). When estimating VMM, we consider two aspects of implementation. First we explore how knowledge of the demand system can improve performance. Given that the demand system is simple logit, neural nets could learn the demand derivatives from market share data alone. However, providing the derivatives directly could improve performance, especially in small samples. Thus, we contrast the performance of our flexible model to one where D_t is not included as an argument. Second, we experiment with different sizes of the neural net: a "small" specification that uses a 3×3 hidden layer, and a "large" specification that uses a 100×100 hidden layer. We provide further details on implementation of the VMM estimator in Appendix D.

Computing Predictions: We obtain counterfactual prices under each model m of the standard IO toolkit as well as our flexible model. In each market t, the vector of prices $\tilde{p}_t^m$ simultaneously satisfies the known true demand system and the model's first order conditions

holding fixed all supply-side parameters and the unobserved cost shocks at their model implied values. For the standard models, we rely on a contraction mapping from Morrow and Skerlos (2011). For the flexible supply function, we find the prices that satisfy

$$(1-\tilde{\nu}_t)\tilde{p}_t - \tilde{\tau}_t - \hat{h}(s_t(\tilde{p}_t), D(\tilde{p}_t), w_t, \tilde{\nu}_t; \mathcal{H}_t) - \tau_t - \hat{\omega}_t = 0 \tag{8}$$

via root-finding methods.[7]

Throughout, we show the distributions of percentage discrepancies $(\tilde{x}^m_{jt} - \tilde{x}_{jt})/\tilde{x}_{jt}$ between predicted and actual counterfactual objects of interest. To summarize the performance of each model, we also compute prediction MSEs for each model m, obtained as:

$$\sum_{j,t} (\tilde{x}^m_{jt} - \tilde{x}_{jt})^2.$$

5.2 Hold-out Sample Performance

We first contrast the performance of our flexible model to the standard IO toolkit when predicting prices for markets in our hold-out samples. For each of the hold-out sample markets, we predict prices under standard IO toolkit models, and under our flexible supply model, by setting cost shocks $\omega^m_t = 0$ and using Equation (8). We obtain MSEs for the prices predicted by each model. Table 1 reports simulation results.[8]

Panel A of Table 1 considers data generated under Bertrand conduct with constant marginal costs. We fit the flexible model with a VMM procedure that relies on a small neural network (3×3 hidden layer). Generally, the prices predicted by the misspecified perfect competition and monopoly models are far away from the truth. The prices predicted by our flexible model estimated with the VMM estimator outperform all misspecified models, and are close to the irreducible error of the correctly specified parametric model (Bertrand with constant cost). Including demand derivatives improves the performance significantly, especially as the sample size (expressed here as the number of markets T) increases.[9] On the whole, our model outperforms all misspecified models, sometimes by several orders of

[7] Alternatively, we set up the problem as a contraction wherein we iterate on prices. In practice, we found that root-finding converged market-wise to a solution to a tolerance of $1e-6$ and the solution matched that of the contraction. We opted for the root-finding technique because it was significantly faster.

[8] In this table, we report MSE results from a single simulation. In Table A2, we report quantiles of the distribution of MSEs across 100 simulations where models are retrained with different weight initializations. The ranges are relatively tight, suggesting results are not particularly sensitive to simulation error and initialization.

[9] In Panel A of Table A1, we increase the size of the neural network and allow more parameters to better fit the nonlinear relationship; we see performance improve with sample sizes even without demand derivatives, fitting the intuition.

magnitude, and gets very close to the performance of the correctly specified parametric model as the sample size increases.

TABLE 1: MSE Across Models

T	True Model	Standard Models			Flexible	D_t included
		B	M	P		
Panel A: Bertrand DGP, Small Network						
100	0.90	0.90	578.77	9.41	1.91	No
					1.71	Yes
1,000	0.89	0.89	1022.81	8.17	2.21	No
					1.07	Yes
10,000	0.962	0.96	1788.82	8.49	2.36	No
					1.15	Yes
Panel B: Profit-Weight DGP, Large Network						
100	0.90	3.71	43.98	10.69	3.04	No
					2.72	Yes
1,000	0.89	3.69	66.42	8.22	1.73	No
					1.74	Yes
10,000	0.96	4.02	77.08	8.79	1.23	No
					1.05	Yes

Notes: The table reports the mean squared error (MSE) in prices for the true model, a set of standard parametric models (B = Bertrand with constant cost, M = joint profit maximization with constant cost, P = marginal cost pricing with constant cost), and the flexible model. In Panel A, the true supply model generating the data is Bertrand with constant marginal cost while in Panel B, it is a profit weight model with economies of scale. In Panel A, the flexible model is trained with a small neural network (3×3 hidden layer) while in Panel B we fit the model on the training sample with a large neural network (100×100 hidden layer). For each imposed supply model under each DGP, the MSE in predicted prices is computed on a hold-out test sample.

The results in Panel B of Table 1 are generated under the profit-weight model of conduct with economies of scale. We use a large (100×100 hidden layer) neural network to fit the flexible models of supply. A similar pattern emerges in the results compared to the Bertrand data-generating process. All of the standard models we consider are now misspecified, and are outperformed by our flexible model, which gets close to match the MSE of the correctly specified model in large sample sizes and when demand derivatives are included. In Table A1, we include results for a small neural network that cannot capture nonlinearities as well as a large neural network. It remains the case that we outperform all misspecified models even with a small neural network.

5.3 Interpretation with Pass-through

A key concern of using a black box method, such as neural networks, is the interpretability of the resulting supply function $\hat{h}$. While the results in Table 1 suggest that the flexible model can correctly predict prices in a hold out sample, the source of its predictive performance is not immediately clear. In an attempt to peek inside the black box, we simulate the pass-through implied by our flexible supply function and compare it to pass-through under the true model. The pass-through matrix allows us to explore whether the flexible model is in fact learning the right relationship on the supply side. To compute pass-through for the flexible supply model, we increase costs by 10 percent, loading increases on the residuals $\hat{\omega}$. We then solve for the equilibrium prices given the increase in cost under the true parametric model and our flexible model, computing pass-through as price changes divided by cost changes.

TABLE 2: Simulated Pass-through Matrices

Panel A: Bertrand DGP				Panel B: Profit-Weight DGP			
True Model		VMM		True Model ($\kappa = 0.5$)		VMM	
0.77	0.17	0.69	0.12	0.35	-0.31	0.44	-0.21
0.10	0.60	0.12	0.63	0.02	0.98	0.01	0.91

Notes: The table reports the simulated pass-through matrices of the true models (Bertrand and profit-weight with $\kappa = 0.5$) and the flexible supply model. Own and rival costs are increased by 10% (separately) and firms respond in prices under each model, loading cost increases on the model-implied ω terms. Reported numbers are price changes divided by the relevant cost changes.

Table 2 shows the true pass-through matrices and those implied by the flexible model in a representative market.[10] In Panel A, the true supply model is Bertrand with constant marginal cost while in Panel B the true supply model is profit-weight with economies of scale. In both these settings, the pass-through implied by the flexible supply model is remarkably close to the true pass-through matrix. The flexible model thus learns markup and cost functions that imply approximately correct pass-throughs, reflecting the economics of different underlying oligopoly models. For instance, in the Bertrand environment, the model learns to pass-through own and rival cost increases and, in the profit-weight environment, the model learns to internalize some of the rival price increases after cost increases.

[10] In Table A3, we compute the element-wise mean-squared error of pass-through tables across many markets, finding that the result is robust to the choice of market.

5.4 Market Counterfactuals

Next, we explore the predictive performance of our flexible approach in counterfactual settings. Motivated by the empirical economics literature, we consider four distinct scenarios: (i) we simulate regulation of a known cost factor and forecast resulting prices; (ii) we model regulation of a product feature and predict consumption; (iv) we predict consumer welfare after one product exits; (iii) we calculate government revenue across various unit and ad valorem tax rates. In each setting, we solve for counterfactual prices under our flexible model and the parametric models of the standard IO toolkit. Then we obtain counterfactual objects by applying the appropriate transformation F to these prices.

5.4.1 Regulating Cost Shifters

Researchers might be interested in how various product regulations affect firms' pricing behavior.[11] For example, environmental regulations could increase the costs of producing a product. We examine these counterfactuals under each of our data generating processes by increasing an observed cost shifter. As w_1 is drawn from a normal distribution $N(1, 0.25)$, scaling it by a factor of two to generates an out-of-sample environment, as the average of this variable falls outside the support of the data used to fit the flexible model.[12]

The results in each data generating process (representing different combinations of Bertrand vs. profit weight conduct and constant marginal cost vs. economies of scale) are presented in Figure 1. Our flexible model outperforms all misspecified parametric models in predicting prices – the MSE across markets is close to zero for the flexible model and considerably lower than those generated by the standard IO toolkit. In all four panels, the distribution of the market-level prediction error in prices for the flexible model is centered tightly around zero.

5.4.2 Regulating Product Characteristics

Governments may also directly regulate product characteristics that influence demand. For instance, bans on menthol in cigarettes or caps on sugar content in packaged foods are often imposed to change consumer behavior, and researchers may want to predict how such regulations affect overall consumption. We examine such counterfactuals by scaling product characteristics. Similar to the example above, we scale one cost shifter x_1 by a factor of

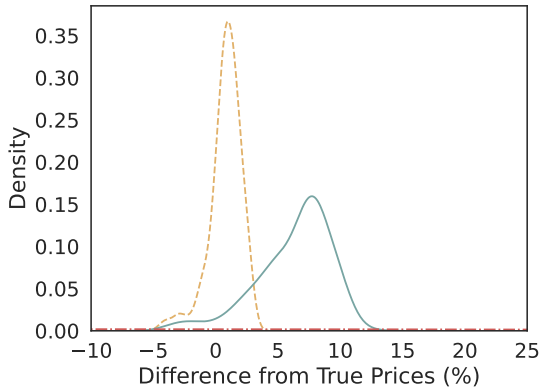
[11] The analysis could extend to consider how other non-price characteristics respond to regulation. While models where firms have more than one choice variable (e.g., price and characteristics) can potentially be estimated flexibly using extensions of the techniques we describe in this paper, we leave this avenue to future research.

[12] In Figure A4, we vary the multiplier on cost shifters to examine the performance of different models as we move more out-of-sample.

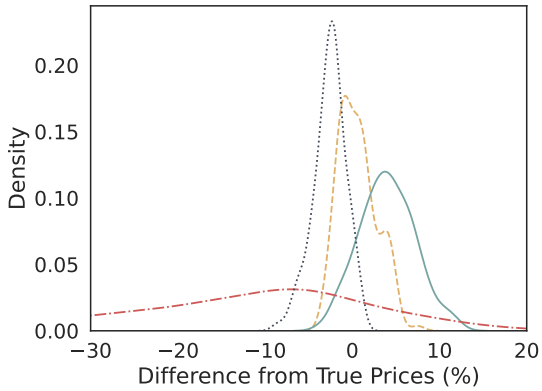
two, making the counterfactual environment out-of-sample.[13] We evaluate the ability of the flexible model to predict changes in consumption under the four supply model data generating processes in Figure 2.

FIGURE 1: Regulation of Cost Shifters

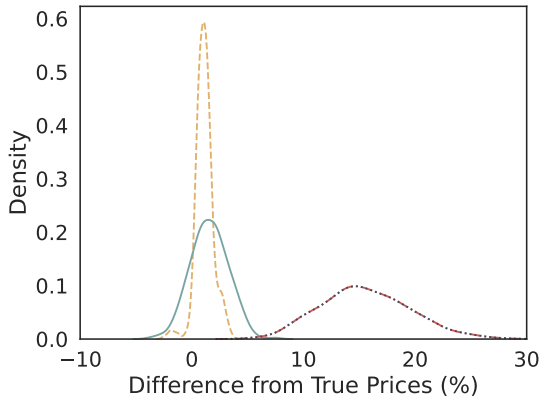
Panel A. Bertrand DGP



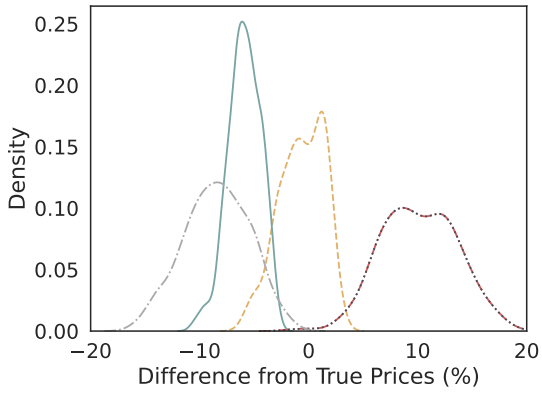
Panel B. Profit-Weight DGP



Panel C. Bertrand DGP, Economies of Scale



Panel D. Profit-Weight DGP, Econ. of Scale



Panel E. Legend and MSEs (in prices) across Panels A-D

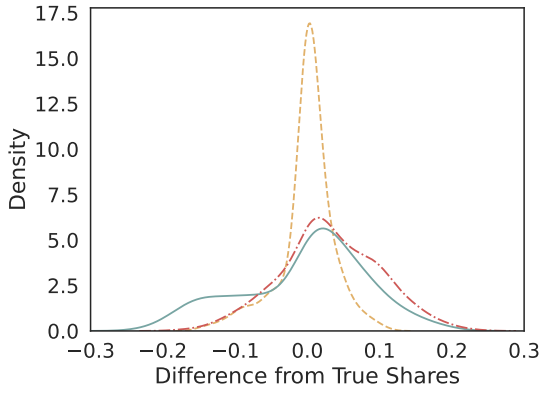
Fitted Model	Panel DGPs			
	A. Bertrand	B. Profit-Weight	C. Bertrand (Scale)	D. Profit-Weight (Scale)
Bertrand (Scale)	--	--	--	4.85
Bertrand (Const.)	--	0.67	13.20	7.04
Monopoly	1758.88	20.58	13.20	7.04
Perf Comp	2.94	1.92	0.25	2.39
Flex Supply	0.15	0.33	0.11	0.32

Figure displays price prediction error for scaling cost shifter w_1 by a factor of two. The DGP in Panels A and B are Bertrand and profit-weight conduct, respectively, with constant marginal costs while Panels C and D incorporate economies of scale. Panel E reports the mean squared error in prices. The neural network used in estimation is small with a 3×3 hidden layer and $T = 1,000$ markets.

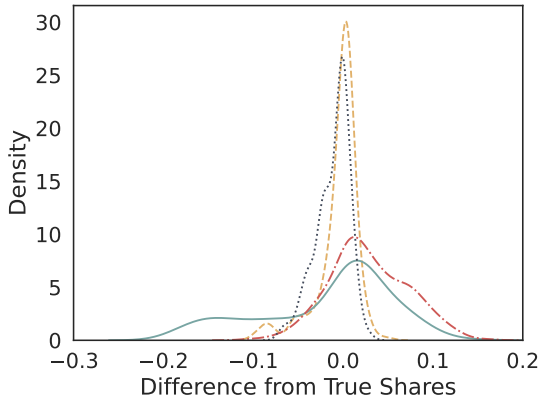
[13] In Figure A3, we vary the multiplier on product characteristics to examine the performance of different models as we move more out-of-sample.

FIGURE 2: Regulation of Product Characteristics

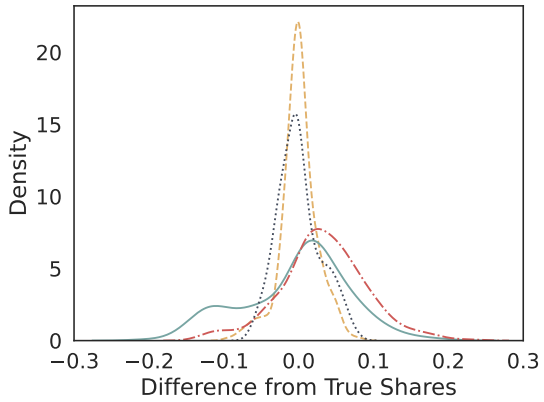
Panel A. Bertrand DGP



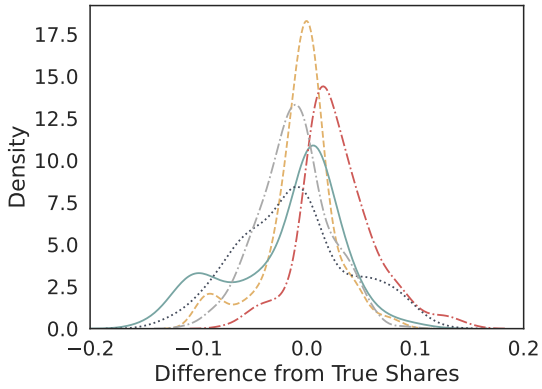
Panel B. Profit-Weight DGP



Panel C. Bertrand DGP, Economies of Scale



Panel D. Profit-Weight DGP, Econ. of Scale



Panel E. Legend and MSEs (in shares) across Panels A-D

Fitted Model	Panel DGPs			
	A. Bertrand	B. Profit-Weight	C. Bertrand (Scale)	D. Profit-Weight (Scale)
Bertrand (Scale)	--	--	--	0.14
Bertrand (Const.)	--	0.05	0.08	0.22
Monopoly	0.56	0.27	0.43	0.69
Perf Comp	0.74	0.57	0.50	0.69
Flex Supply	0.15	0.06	0.06	0.07

Figure displays share prediction error for scaling characteristic x_1 by a factor of two. The DGP in Panels A and B are Bertrand and profit-weight conduct, respectively, with constant marginal costs while Panels C and D incorporate economies of scale. Panel E reports the mean squared error in shares. The neural network used in estimation is small with a 3×3 hidden layer and $T = 1,000$ markets.

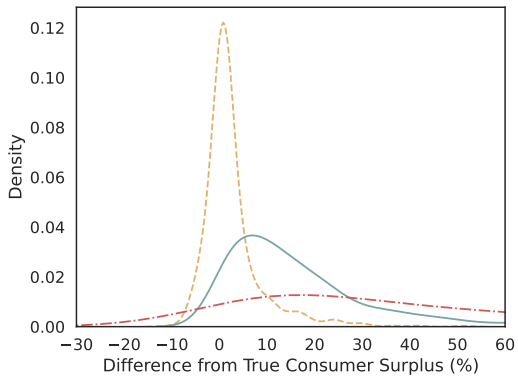
Figure 2 illustrates that the flexible model better predicts changes in market shares. When the true supply model is profit weight with economies of scale, our flexible model predicts a total increase in consumption across markets of 21.83 (recall that market sizes are normalized to one). This is close to the true increase of 19.94, and outperforms the misspecified Bertrand model with constant marginal cost, which predicts an increase of 23.67.

5.4.3 Product Exit

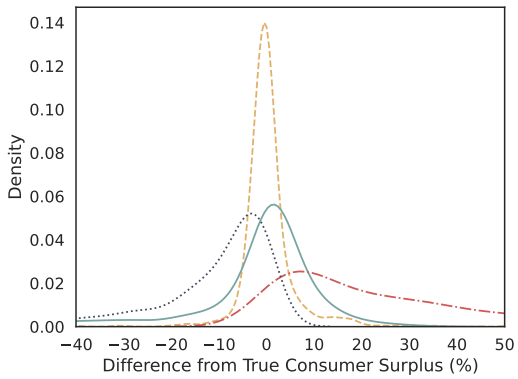
We now consider removing a product from offered set. This counterfactual allows a researcher to measure the benefits from product entry (e.g., Petrin, 2002) or to predict the effects of exit (e.g., if a merged firm drops duplicate products). We use the flexible model to predict changes in consumer welfare, which we measure according to Small and Rosen (1981).

FIGURE 3: Product Exit

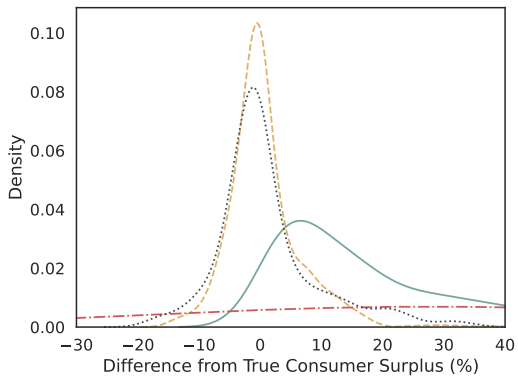
Panel A. Bertrand DGP



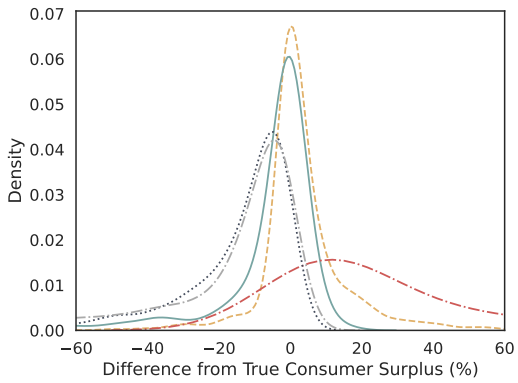
Panel B. Profit-Weight DGP



Panel C. Bertrand DGP, Economies of Scale



Panel D. Profit-Weight DGP, Econ. of Scale



Panel E. Legend and MSEs (in surplus) across Panels A-D

Fitted Model	Panel DGPs			
	A. Bertrand	B. Profit-Weight	C. Bertrand (Scale)	D. Profit-Weight (Scale)
Bertrand (Scale)	--	--	--	0.40
Bertrand (Const.)	--	0.33	0.06	0.43
Monopoly	6.92	1.46	6.93	1.35
Perf Comp	1.57	0.23	1.12	0.15
Flex Supply	0.19	0.03	0.12	0.08

Figure displays consumer surplus prediction error from product exit. The DGP in Panels A and B are Bertrand and profit-weight conduct, respectively, with constant marginal costs while Panels C and D incorporate economies of scale. Panel E reports the mean squared error in consumer surplus. The neural network used in estimation is small with a 3×3 hidden layer. $T = 1,000$ markets.

Figure 3 presents the prediction error in the effect of product exit on consumer surplus across our four data-generating supply models. Across simulations, the flexible model outperforms all misspecified models and is close to predicting the true welfare changes. Under the profit weight DGP with economies of scale, the flexible model predicts that product exit reduces total consumer surplus across all markets by $\$1,316$, virtually identical to the true reduction of $\$1,318$. Instead, the imposed Bertrand with constant marginal cost model predicts a larger reduction of $\$1,796$. While our training data contains both triopolies into duopolies, making the counterfactual somewhat in sample, we show in Figure A1 that the flexible model performs similarly well when the training data only contains triopolies.

5.4.4 Laffer Curves

We next consider a researcher who wants to predict the tax revenue resulting from counterfactual levels of unit or ad-valorem taxes, tracing out the Laffer curve (see, e.g., Miravete et al., 2018). To do so, we return to the simulation environment with single-product duopolies and triopolies but now incorporate market-level taxes when generating data as described in Section 5.1. For each imposed model, including the flexible one, we predict counterfactual prices, shares, and government revenue at different levels of the ad-valorem tax holding fixed the unit tax to construct the ad valorem Laffer curve (and vice versa, vary unit taxes and hold constant the ad valorem tax for the unit tax Laffer curve). We solve in each scenario for counterfactual prices using Equation (8), and compute the corresponding revenues.[14]

Figure 4 presents the Laffer curves under the true model, a host of misspecified models, and the flexible model for unit and ad valorem taxes. Our VMM estimator recovers a Laffer curve very close to the true curve, which is substantially different from the curves recovered under misspecified models. Importantly, our method requires variation in the ad valorem tax rate. Appendix Figure A2 decreases the variation in observed taxes to show the importance of this type of variation in the data.

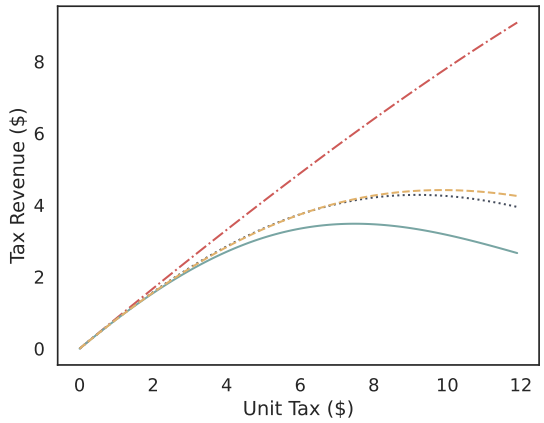
5.5 Inference Performance

After comparing the predictive performance of our flexible supply model to standard models in counterfactuals and a hold-out sample, we quantify uncertainty in the counterfactual predictions. We use a product exit counterfactual to exemplify the inference procedure. The results are presented in Figure 5.

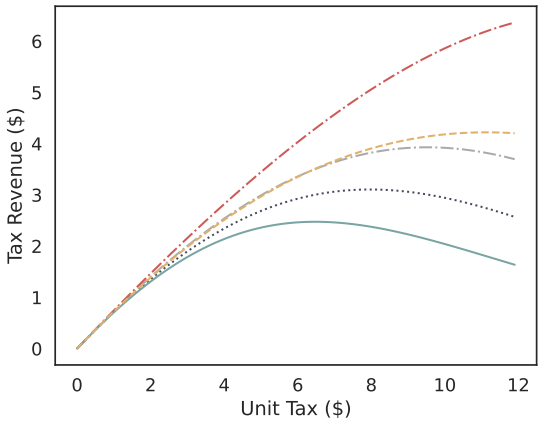
[14] We load taxes $\tilde{\tau}_t$ on the residual $\hat{\omega}_t$ to compute counterfactual prices. Tax revenue is calculated as $R_t(\nu, \tau) = \sum_j \tau s_{jt}(p_{jt}(\nu, \tau)) + \sum_j \nu p_{jt}(\nu, \tau) s_{jt}(p_{jt}(\nu, \tau))$, where ν and τ are the given levels of ad valorem and unit taxes, respectively.

FIGURE 4: Laffer Curves for Unit and Ad Valorem Taxes

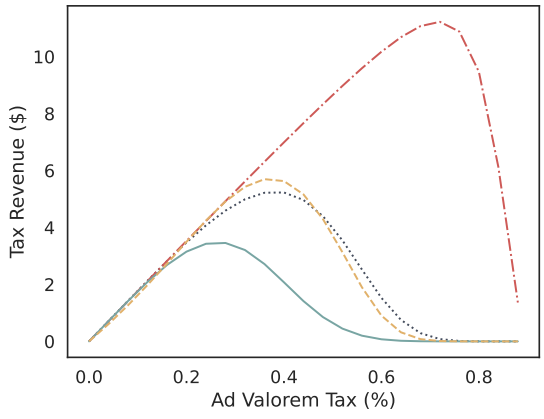
Panel A. Bertrand DGP, Unit Taxes



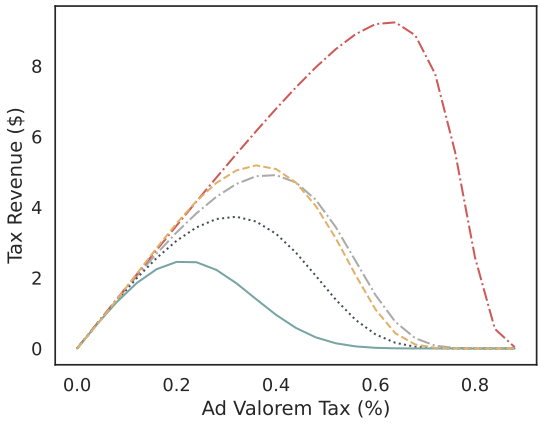
Panel B. Profit-Weight DGP, Unit Taxes



Panel C. Bertrand DGP, Ad Valorem Taxes



Panel D. Profit-Weight DGP, Ad Valorem Taxes



Panel E. Legend and MSEs (in prices) across Panels A-D

Fitted Model	Panel DGPs			
	A. Bertrand (Unit)	B. Profit-Weight (Unit)	C. Bertrand (AV)	D. Profit-Weight (AV)
Bertrand	--	0.38	--	1.01
Monopoly	5.09	1.49	33.38	17.21
Perf Comp	0.45	1.43	2.79	4.11
Flex Supply	0.01	0.03	0.09	0.05

Figure displays predicted Laffer curves. The DGP in Panels A and C are Bertrand with constant cost, while Panels B and D are under profit-weights conduct and economies of scale. Panel E reports the MSE in prices. The neural network used in estimation is small with a 3×3 hidden layer. $T = 100$ triopoly markets.

Panels A and B show inference performance under Bertrand conduct with constant costs, with sample sizes $T = 100$ and $T = 1,000$ markets, respectively. Panels C and D are constructed under profit-weight conduct and economies of scale, with $T = 100$ and $T = 1,000$ markets, respectively. The horizontal axis represents the counterfactual price predicted by the nonparametric model and the vertical axis is the width of the confidence interval. In

FIGURE 5: Inference on Counterfactual Merger Simulation Prices

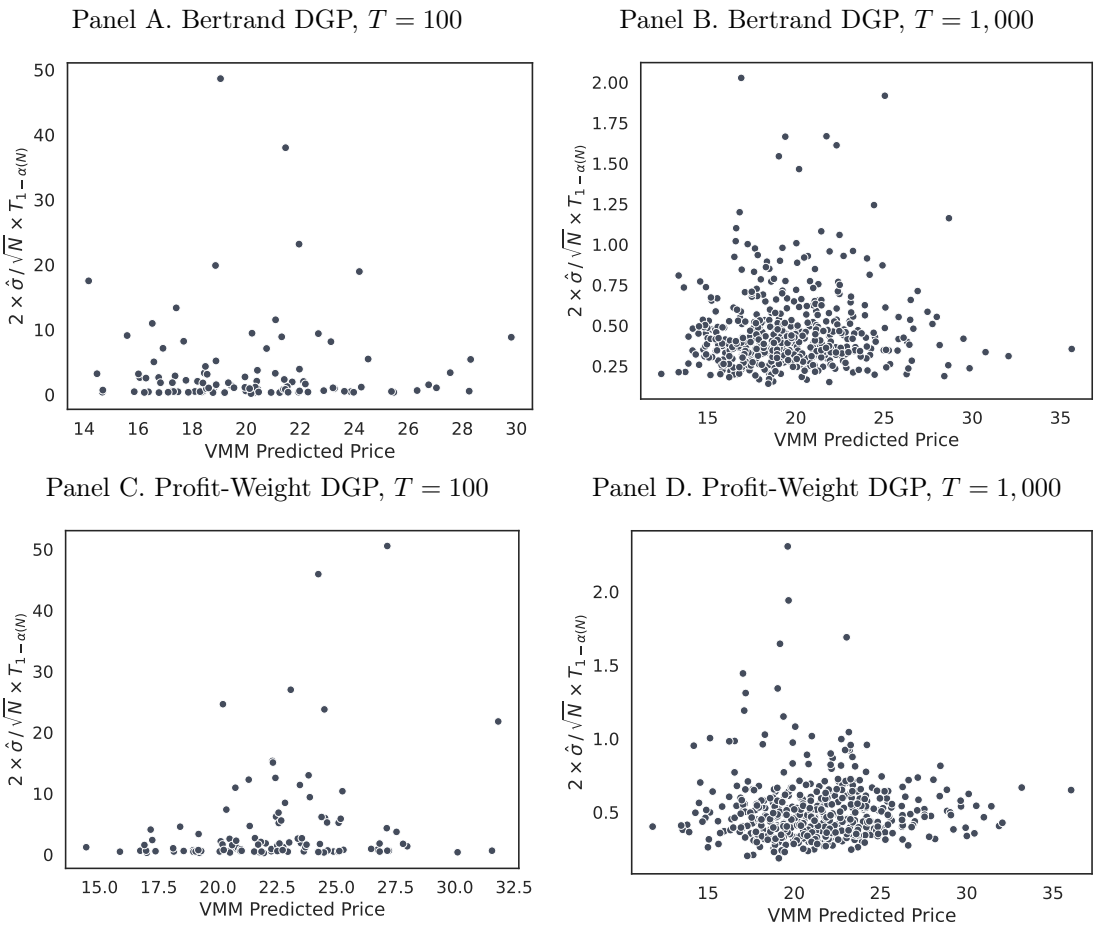


Figure displays the width of confidence intervals for prices resulting from a counterfactual product exit under a Bonferroni correction. Panels A and B show the results under Bertrand conduct and constant costs with $T = 100$ and $T = 1,000$, respectively. Panels C and D show the results under profit-weight conduct and economies of scale with $T = 100$ and $T = 1,000$, respectively. The neural networks used in estimation are small with 3×3 hidden layers.

Panels A and C, the smaller sample size generates some outliers, with large confidence intervals for the prices predicted in a few markets, but most confidence intervals are a small fraction of predicted prices. Moving to Panels B and D, the larger sample size mitigates the outlier issues and brings most intervals to a much tighter range. This exercise suggests that our flexible supply model is not only accurate but produces precise predictions.

6 Empirical Application

As a showcase of our method, we examine the US airline industry. The airline industry has received substantial attention from the IO literature (starting with Berry, 1992; Berry,

Carnall, and Spiller, 2006) given the rich available data and significant consolidation over the last two decades. The retrospective studies of large mergers have had mixed results (Peters, 2006), potentially linked to non-Bertrand conduct (see, e.g., Ciliberto and Williams, 2014, for evidence of non-competitive conduct). Our method applies well here given the large amount of data and variation in market structure.[15]

6.1 Background and Data

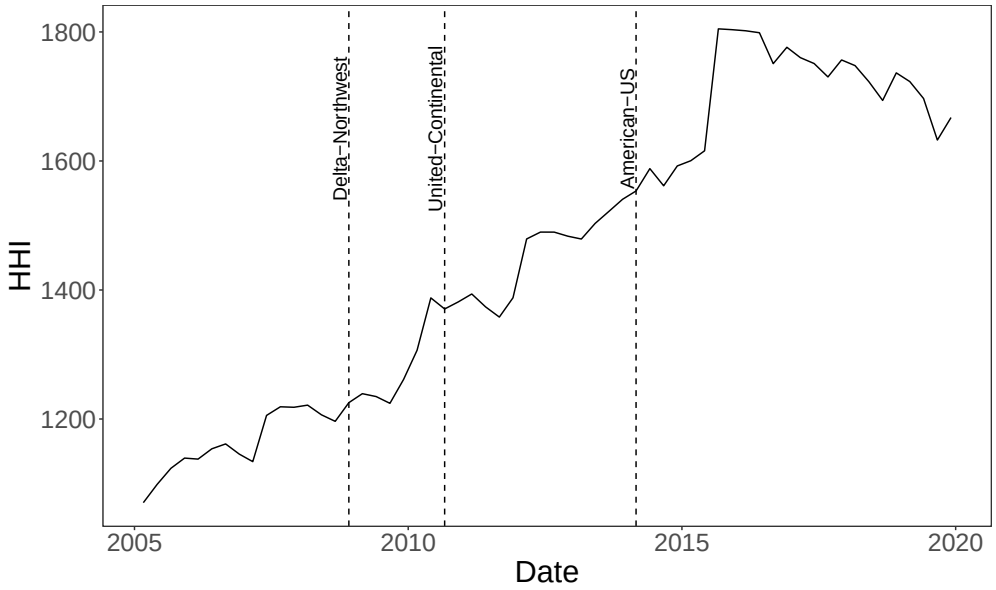
We construct a database of the US airline industry during the period 2005-2019. We use the 10 percent sample of purchased airline tickets from the well-known Airline Origin and Destination Survey (DB1B) database released by the US Department of Transportation. Following Azar, Schmalz, and Tecu (2018) and Kennedy, O'Brien, Song, and Waehrer (2017), a market is defined as a pair of cities, regardless of the flight direction. We match cities to Metropolitan Statistical Areas (MSA) and collect data on the populations of these MSAs from the Bureau of Economic Analysis. The geometric mean of endpoint populations is used as a measure of the market size. A product is a one-way trip that services a particular city-pair and is defined at the carrier-market-quarter level. Additional details on the construction of the data can be found in Appendix C.1.

The U.S. airline industry has experienced substantial consolidation in the last two decades with legacy carriers and low-cost airlines participating in large mergers. We show a descriptive increase in the Herfindahl-Hirschman Index (HHI) in Figure 6. The earliest merger in our data is the Delta-Northwest merger in 2008. The merger was announced on April 14, 2008, and was approved on October 29, 2008, after roughly six months of review by the US Department of Justice (DOJ). Given the limited overlap between the merging airlines' operations, the merger was perceived as having a modest impact on competition. The second merger included is the United-Continental merger in 2010. The DOJ approved the merger after four months of review on August 27, 2010. As a condition of approval, the merged entity was required to lease slots to Southwest at Newark Liberty Airport in New Jersey. Finally, we consider the controversial merger of American Airlines and US Airways. The last of the "mega-mergers" that involved two airlines, the deal was announced on November 12, 2013. According to the settlement terms, the merged entity was required to divest slots at several major airports, most prominently at Ronald Reagan Washington National

[15] We emphasize the illustrative nature of our application. Recent papers have highlighted the dynamic nature of pricing and demand in this market (e.g., Williams, 2022; Betancourt, Hortaçsu, Oery, and Williams, 2022; Aryal, Murry, and Williams, 2024; Hortaçsu, Natan, Parsley, Schwieg, and Williams, 2024), and the role of endogenous network structure (Ciliberto, Murry, and Tamer, 2021; Li, Mazur, Park, Roberts, Sweeting, and Zhang, 2022; Bontemps, Gualdani, and Remmy, 2023; Yuan and Barwick, 2024). We abstract away from these important elements to keep our application tied to the standard merger simulation toolkit.

Airport and New York's LaGuardia Airport. More recently, outside our analysis, Alaska Airlines acquired Virgin America and Hawaiian Airlines, and a federal court blocked JetBlue's attempted acquisition of Spirit.

FIGURE 6: Concentration in the Airline Industry



Notes: The figure plots the evolution of the national HHI of the airline industry during the period 2005-2019. We use passenger counts to construct market shares used in the calculation of HHI.

6.2 Demand and Supply Estimation

Demand Estimation: We follow Berry and Jia (2010) in adopting a nested logit demand model. We briefly summarize the model here and provide additional details in Appendix C.2. Product characteristics include average fares, the share of nonstop flights, the average distance in thousands of miles, and a squared distance term. We restrict our attention to the major carriers, controlling for the number of fringe firms to capture variation in market structure over time across origin-destination pairs. We include origin-destination fixed effects. Our nesting structure includes all inside goods in one nest. We include instruments to handle endogeneity issues for prices and nests. We use BLP instruments as the average rival distance, the average number of markets a rival serves, and the number of rival carriers.

The results for demand estimation are reported in Table C2. In line with the previous literature, we find that consumers prefer a higher share of nonstop flights and incur disutility from more miles traveled. There is strong within-nest substitution. The median own-price elasticity of -5.17 matches the literature well.

Pre-merger Supply Estimation: We consider two models of conduct for our supply specifications: (i) Bertrand pricing and (ii) a flexible supply function as described in Section 3. In this section, we focus on the flexible specification and relegate details of the Bertrand specification to Appendix C.3.

We estimate the flexible supply model with the VMM estimator described in Section 4. We include in the supply function market shares and the average distance in thousands of miles as an observable cost shifter. We also include origin-destination fixed effects. We instrument the endogenous market shares with BLP instruments formed with the following characteristics: average rival distance, average number of markets a rival serves, and number of rival carriers. Additionally, we include own-product characteristics that do not directly impact marginal costs – the share of nonstop flights and squared average distance in thousands of miles – as excluded instruments.

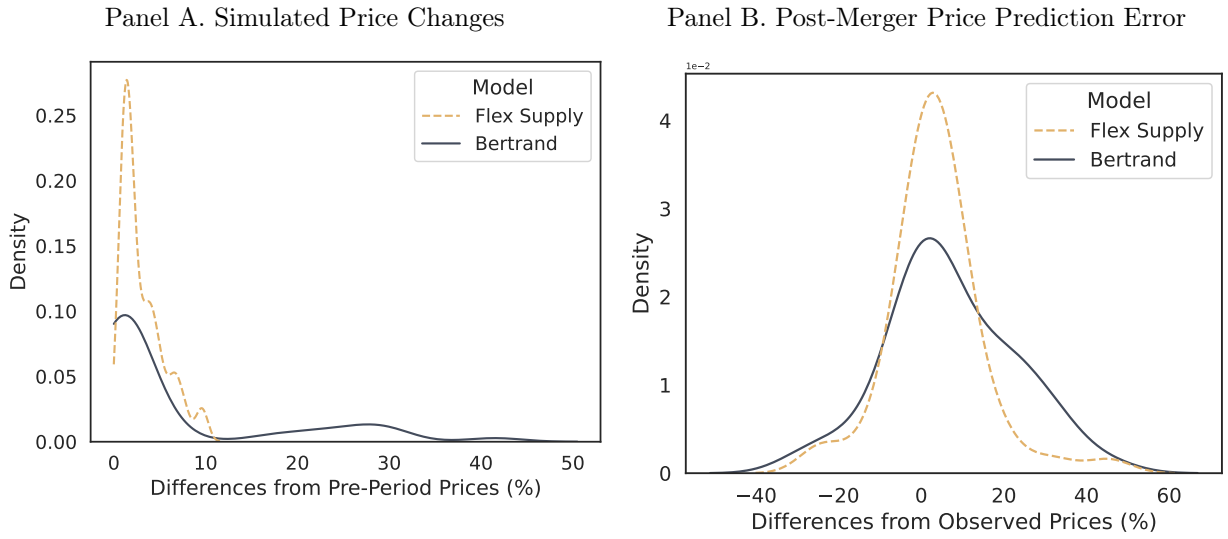
Having estimated the two models, we can evaluate their fit in the pre-merger data. The flexible supply function estimated with VMM significantly outperforms the Bertrand conduct assumption by around 36%; this margin is similar in the training data and a test sample. We stratify the data by market and split it into 80% of markets for training, leaving the remaining 20% of markets as a hold-out test sample to evaluate the fit.

6.3 Merger Simulation Results

We examine the merger of American Airlines and US Airways in our counterfactuals. We focus on markets with three firms in the pre-merger period and two firms in the post-merger period. We compare the predictions of the models to the true post-merger prices. The observed price differences are presented in Figure C1.

Figure 7 presents the results of the merger simulation exercise. Panel A represents the distribution of predicted price increases from the merger under the flexible supply model and under the standard merger simulation with Bertrand. We find that the flexible model predicts a fairly tight distribution of price changes, with an average predicted change of 3.12% (median 2.08%). In contrast, the standard merger simulation procedure with Bertrand conduct predicts an average price change of 6.66% (median 1.45%), which is largely driven by a few large predicted changes (up to 40% increases). Large price increases are predicted by Bertrand for markets that move closer to a monopoly after the merger. More specifically, the market was dominated by US Airways and American Airlines; the third firm was a fringe in the market with a very small market share. After the merger, the market appears close to a monopoly, leading to large price increases under the Bertrand assumption.

FIGURE 7: Merger Simulation Results



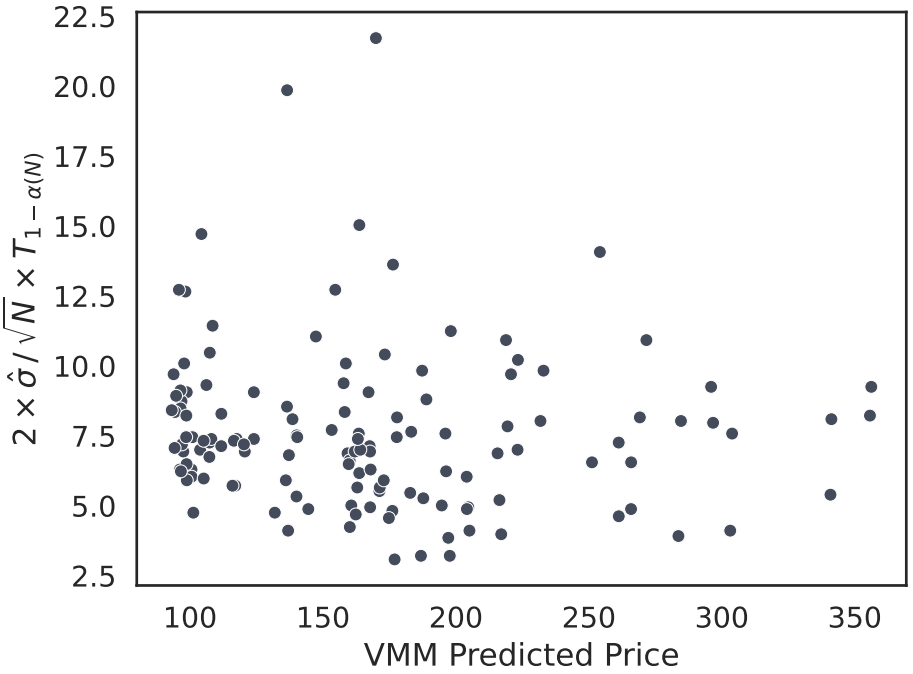
Notes: The figure reports merger simulation results for the flexible model estimated with VMM (in yellow) and the standard merger simulation model (in blue). Panel A reports the distribution of percent differences between pre-merger and post-merger predicted prices. Panel B reports the distribution of percent differences between observed and post-merger predicted prices. Figure C2 breaks down Panel B by merger status.

To analyze the source of these different predictions, we now compare whether the flexible model or the standard merger simulation Bertrand model more closely matches the realized post-merger prices. Panel B of Figure 7 plots the distribution of percentage differences between predicted and observed prices for the two methods in the post-merger period. The flexible model estimated with VMM is centered at zero, with a large fraction of predicted prices falling within 20% of realized prices, and an MSE of 68.31. Instead, the standard merger simulation method systematically over-predicts changes, and has an overall MSE of 365.72. In sum, our flexible model substantially outperforms the standard simulation toolkit when predicting post-merger prices of the American-US Airways merger. Under the model of Bertrand conduct with constant cost, the predicted loss in consumer surplus per consumer as a result of the merger is \$14.17 while our flexible model predicts more modest welfare losses of \$9.59 per consumer. Aggregating across markets, the total predicted loss in consumer surplus under the Bertrand model with constant marginal cost is \$40,143,771 while the flexible model predicts a total loss of consumer surplus of \$26,117,699.

Quantifying Uncertainty: Finally, we quantify the uncertainty of our predictions in the merger simulation exercise. We follow Algorithm 1 to construct the confidence intervals. Notably, we use a more conservative Bonferroni correction for ease of exposition, allowing us to present the results with a single set of bounds for each point. We construct bounds for

all points in the sample selected for merger simulation.

FIGURE 8: Width of Confidence Intervals



Notes: The figure represents the scatter plot of VMM predicted prices (on the horizontal axis) against the width of the confidence interval for the predictions. Variance is estimated for all firms in markets affected by the US Airways-American merger.

The results are presented in Figure 8. We present the total width of the confidence interval as a summary of uncertainty. The predictions of the inference exercise show that uncertainty is roughly constant with prediction levels. Further, and more importantly, the width is at a reasonably small level in which we can make precise point estimates even with a high-dimensional markup function.

7 Conclusion

This paper demonstrates how machine learning methods can enhance market counterfactual analysis while maintaining economic structure. Our approach relaxes standard assumptions about firm conduct and cost when rich market data are available, adapting recent advances in variational inference to estimate a flexible model of supply that nests traditional specifications while allowing for more complex competitive interactions. The method integrates ML/AI tools into structural economic models without sacrificing their economic content, representing a middle ground between theory-driven parametric approaches and purely

data-driven prediction. Applied to the American Airlines-US Airways merger, our flexible model produces more accurate post-merger price predictions, outperforming the standard Bertrand-Nash approach while maintaining computational tractability.

While our approach offers significant advantages in flexibility and predictive performance, several limitations deserve mention. First, our method's reliability depends on the support of the training data; predictions for counterfactual scenarios that represent radical departures from observed market structures will be less reliable. Second, the method requires richer exogenous identifying variation compared to standard parametric approaches. Third, without additional restrictions, the method cannot separately identify markups from marginal costs. Future research could develop hybrid approaches that combine data-driven flexibility with economic restrictions in settings where data are limited, and extend the framework to dynamic settings incorporating entry, exit, and investment. Despite these limitations, our results suggest that careful application of modern machine learning methods can significantly improve counterfactual prediction accuracy while maintaining the conceptual foundations that make structural analysis valuable for policy evaluation.

References

Aryal, G., C. Murry, and J. Williams (2024): "Price discrimination in international airline markets," *Review of Economic Studies*, 91, 641–689.

Azar, J., M. C. Schmalz, and I. Tecu (2018): "Anticompetitive effects of common ownership," *The Journal of Finance*, 73, 1513–1565.

Backus, M., C. Conlon, and M. Sinkinson (2021): "Common Ownership and Competition in the Ready-To-Eat Cereal Industry," NBER working paper #28350.

Bauer, B. and M. Kohler (2019): "On Deep Learning as a Remedy for the Curse of Dimensionality in Nonparametric Regression," *The Annals of Statistics*, 47, 2261–2285.

Bennett, A. and N. Kallus (2023): "The Variational Method of Moments," *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 85, 810–841.

Berry, S. and P. Haile (2014): "Identification in Differentiated Products Markets Using Market Level Data," *Econometrica*, 82, 1749–1797.

Berry, S. and P. Jia (2010): "Tracing the woes: An empirical analysis of the airline industry," *American Economic Journal: Microeconomics*, 2, 1–43.

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

Berry, S. T. (1992): "Estimation of a Model of Entry in the Airline Industry," *Econometrica*, 889–917.

——— (1994): "Estimating discrete-choice models of product differentiation," *The RAND Journal of Economics*, 25, 242–262.

Berry, S. T., M. Carnall, and P. T. Spiller (2006): "Airline hubs: costs, markups and the implications of customer heterogeneity," *Competition policy and antitrust*.

Betancourt, J., A. Hortaçsu, A. Oery, and K. Williams (2022): "Dynamic price competition: Theory and evidence from airline markets," Tech. rep., nBER working paper #30347.

Bontemps, C., C. Gualdani, and K. Remmy (2023): "Price competition and endogenous product choice in networks: Evidence from the us airline industry," Tech. rep., Working Paper.

Brand, J. and A. N. Smith (2025): "A Quasi-Bayes Approach to Nonparametric Demand Estimation with Economic Constraints," *Available at SSRN 5100826*.

Carrasco, M., J.-P. Florens, and E. Renault (2007): "Linear inverse problems in structural econometrics estimation based on spectral decomposition and regularization," *Handbook of econometrics*, 6, 5633–5751.

Chen, J., X. Chen, and E. Tamer (2023): "Efficient estimation of average derivatives in NPIV models: Simulation comparisons of neural network estimators," *Journal of Econometrics*, 235, 1848–1875.

Chen, X. (2007): "Large sample sieve estimation of semi-nonparametric models," *Handbook of econometrics*, 6, 5549–5632.

Ciliberto, F., C. Murry, and E. Tamer (2021): "Market Structure and Competition in Airline Markets," *Journal of Political Economy*, 129, 2995–3038.

Ciliberto, F. and J. W. Williams (2014): "Does multimarket contact facilitate tacit collusion? Inference on conduct parameters in the airline industry," *The RAND Journal of Economics*, 45, 764–791.

Compiani, G. (2022): "Market Counterfactuals and the Specification of Multiproduct Demand: A Nonparametric Approach," *Quantitative Economics*, 13, 545–591.

Conlon, C. and J. Gortmaker (2020): "Best Practices for Differentiated Products Demand Estimation with pyblp," *RAND Journal of Economics*, 51, 1108–1161.

De Loecker, J. and F. Warzynski (2012): "Markups and firm-level export status," *American economic review*, 102, 2437–2471.

Dearing, A., L. Magnolfi, D. Quint, C. Sullivan, and S. Waldfogel (2024): "Learning Firm Conduct: Pass-through as a Foundation for Instrument Relevance," Working paper.

Dikkala, N., G. Lewis, L. Mackey, and V. Syrgkanis (2020): "Minimax estimation of conditional moment models," *Advances in Neural Information Processing Systems*, 33, 12248–12262.

Duarte, M., L. Magnolfi, M. Sølvsten, and C. Sullivan (2024): "Testing Firm Conduct," *Quantitative Economics*, forthcoming.

Dubois, P., R. Griffith, and M. O'Connell (2018): "The Effects of Banning Advertising in Junk Food Markets," *Review of Economic Studies*, 85, 396–436.

Farrell, M. H., T. Liang, and S. Misra (2020): "Deep learning for individual heterogeneity: An automatic inference framework," *arXiv preprint arXiv:2010.14694*.

Fernández-Villaverde, J., G. Nuño, and J. Perla (2024): "Taming the curse of dimensionality: quantitative economics with deep learning," Tech. rep., National Bureau of Economic Research.

Gandhi, A. and J.-F. Houde (2020): "Measuring Firm Conduct in Differentiated Products Industries," Tech. rep., Working Paper.

Hartford, J., G. Lewis, K. Leyton-Brown, and M. Taddy (2017): "Deep IV: A flexible approach for counterfactual prediction," in *International Conference on Machine Learning*, PMLR, 1414–1423.

Holm, S. (1979): "A simple sequentially rejective multiple test procedure," *Scandinavian journal of statistics*, 65–70.

Hong, H., A. Mahajan, and D. Nekipelov (2015): "Extremum estimation and numerical derivatives," *Journal of Econometrics*, 250–263.

Hortaçsu, A., O. Natan, H. Parsley, T. Schwieg, and K. Williams (2024): "Organizational Structure and Pricing: Evidence from a Large US Airline," *Quarterly Journal of Economics*, 139, 1149–1199.

Kaji, T., E. Manresa, and G. Pouliot (2023): "An adversarial approach to structural estimation," *Econometrica*, 91, 2041–2063.

Kennedy, P., D. P. O'Brien, M. Song, and K. Waehrer (2017): "The competitive effects of common ownership: Economic foundations and empirical evidence," *Available at SSRN 3008331*.

Kurisu, D., R. Fukami, and Y. Koike (2025): "Adaptive deep learning for nonlinear time series models," *Bernoulli*, 31, 240–270.

Lewis, G. and V. Syrgkanis (2018): "Adversarial generalized method of moments," *arXiv preprint arXiv:1803.07164*.

Li, S., J. Mazur, Y. Park, J. Roberts, A. Sweeting, and J. Zhang (2022): "Repositioning and Market Power after Airline Mergers," *RAND Journal of Economics*, 53, 166–199.

Liao, L., Y.-L. Chen, Z. Yang, B. Dai, M. Kolar, and Z. Wang (2020): "Provably efficient neural estimation of structural equation models: An adversarial approach," *Advances in Neural Information Processing Systems*, 33, 8947–8958.

Magnolfi, L. and C. Sullivan (2022): "A Comparison of Testing and Estimation of Firm Conduct," *Economics Letters*, 212, 110316.

Miravete, E., K. Seim, and J. Thurk (2018): "Market Power and the Laffer Curve," *Econometrica*, 86, 1651–1687.

Morrow, R. and S. Skerlos (2011): "Fixed-Point Approaches to Computing Bertrand-Nash Equilibrium Prices Under Mixed-Logit Demand," *Operations Research*, 59, 328–345.

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

Otsu, T. and M. Pesendorfer (2024): "Conduct in the Soft Drink Market: A Mechanism Design Approach," Tech. rep., Working Paper.

Peters, C. (2006): "Evaluating the Performance of Merger Simulation: Evidence from the US Airline Industry," *Journal of Law and Economics*, 49, 627–649.

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

Schmidt-Hieber, J. (2020): "Nonparametric regression using deep neural networks with ReLU activation function," *The Annals of Statistics*, 48.

Sill, J. (1997): "Monotonic networks," *Advances in neural information processing systems*, 10.

Small, K. and H. Rosen (1981): "Applied Welfare Economics with Discrete Choice Models," *Econometrica*, 105–130.

Tebaldi, P., A. Torgovitsky, and H. Yang (2023): "Nonparametric estimates of demand in the california health insurance exchange," *Econometrica*, 91, 107–146.

Wehenkel, A. and G. Louppe (2019): "Unconstrained monotonic neural networks," *Advances in neural information processing systems*, 32.

Wei, Y. and Z. Jiang (2024): "Estimating parameters of structural models using neural networks," *Marketing Science*.

Williams, K. (2022): "The Welfare Effects of Dynamic Pricing: Evidence from Airline Markets," *Econometrica*, 90, 831–858.

You, S., D. Ding, K. Canini, J. Pfeifer, and M. Gupta (2017): "Deep lattice networks and partial monotonic functions," *Advances in neural information processing systems*, 30.

Yuan, Z. and P. J. Barwick (2024): "Network Competition in the Airline Industry: An Empirical Framework," Tech. rep., National Bureau of Economic Research.

Zhang, R., M. Imaizumi, B. Schölkopf, and K. Muandet (2023): "Instrumental variable regression via kernel maximum moment loss," *Journal of Causal Inference*, 11, 20220073.

A Additional Simulation Results

TABLE A1: MSE Across Models

T	True Model	Standard Models			VMM	D_t included
		B	M	P		
Panel A: Bertrand DGP, Large Network						
100	0.895	0.895	578.773	9.411	2.083	No
					1.126	Yes
1,000	0.889	0.889	1022.812	8.170	1.091	No
					1.103	Yes
10,000	0.962	0.962	1788.816	8.487	1.226	No
					1.385	Yes
Panel B: Profit-Weight DGP, Small Network						
100	0.895	3.711	43.981	10.694	3.653	No
					2.811	Yes
1,000	0.889	3.689	66.424	8.224	3.037	No
					1.666	Yes
10,000	0.962	4.016	77.079	8.792	2.976	No
					1.736	Yes

Notes: The table reports the mean squared error (MSE) in prices for the true models, a set of standard models, and the VMM model. The model is trained on duopolies and triopolies. The MSE is computed on a hold-out test sample restricted to markets where merging firms are present. Panel A reports the results for a Bertrand DGP with a large neural network (100×100 hidden layer) and Panel B reports results for the profit-weight DGP ($\kappa = 0.5$) with a small neural network (3×3 hidden layer)

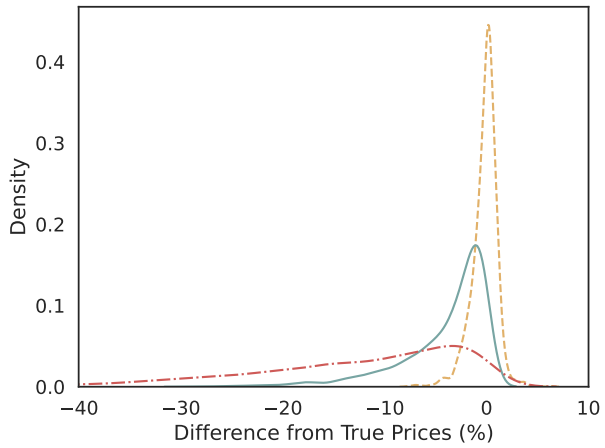
TABLE A2: MSE Ranges Across Models

T	Network Size	D_t Included	P5	P25	P50	P75	P90
Panel A: Bertrand DGP							
100	Small	No	1.881	1.99	2.117	2.182	2.358
		Yes	1.525	1.674	1.801	1.917	2.075
	Large	No	0.963	1.332	2.476	3.58	4.818
		Yes	0.878	1.448	2.258	3.284	4.786
$1,000$	Small	No	1.955	2.03	2.123	2.345	2.502
		Yes	1.014	1.096	1.215	1.336	1.599
	Large	No	0.903	1.022	1.13	1.269	1.625
		Yes	0.854	0.996	1.082	1.294	1.557
Panel B: Profit-Weight DGP							
100	Small	No	3.058	3.269	3.437	3.564	5.550
		Yes	2.135	2.459	2.688	2.872	3.953
	Large	No	1.165	1.766	2.695	3.898	5.041
		Yes	0.980	1.49	2.312	3.642	4.999
$1,000$	Small	No	2.383	2.493	2.608	2.769	3.053
		Yes	1.451	1.506	1.608	1.753	2.017
	Large	No	0.857	0.948	1.08	1.372	1.766
		Yes	0.838	1.004	1.14	1.436	1.864

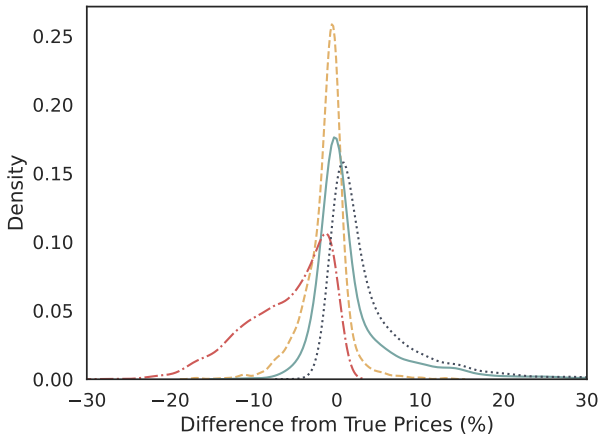
Notes: The table reports the moments of distribution of mean squared errors (MSE) in prices for the VMM model with 100 different neural network weight initializations. The models are trained on duopolies and triopolies. The MSE is computed on a hold-out test sample restricted to markets where merging firms are present. Panel A reports the results for a Bertrand DGP with small neural networks (3×3 hidden layer) and large neural networks (100×100). Panel B reports results for the profit-weight DGP ($\kappa = 0.5$) with small neural networks (3×3) and large neural networks (100×100 hidden layer).

FIGURE A1: Single-Product Merger Simulation

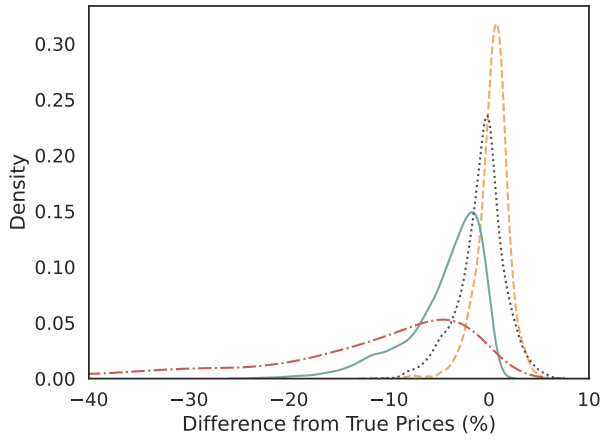
Panel A. Bertrand DGP



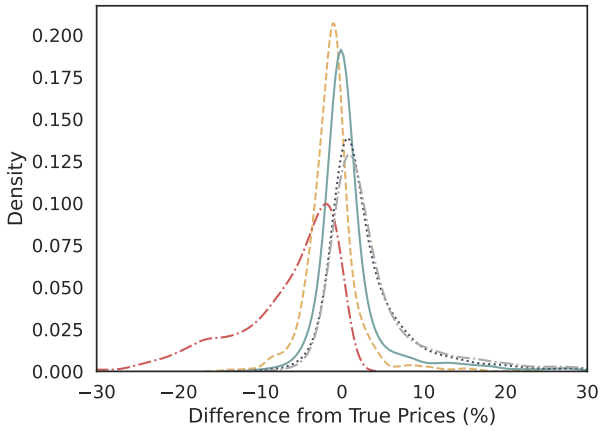
Panel B. Profit-Weight DGP



Panel C. Bertrand DGP, Economies of Scale



Panel D. Profit-Weight DGP, Economies of Scale



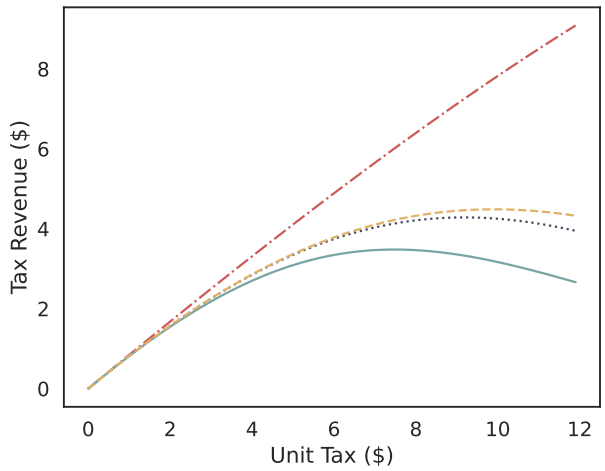
Panel E. Legend and MSEs across Panels A-D

Fitted Model	Panel DGPs			
	A. Bertrand	B. Profit-Weight	C. Bertrand (Scale)	D. Profit-Weight (Scale)
Bertrand (Scale)	--	--	--	2.83
Bertrand (Const.)	--	1.94	0.23	2.11
Monopoly	10.44	2.73	11.81	3.61
Perf Comp	1.76	1.37	1.53	1.37
Flex Supply	0.06	0.27	0.14	0.50

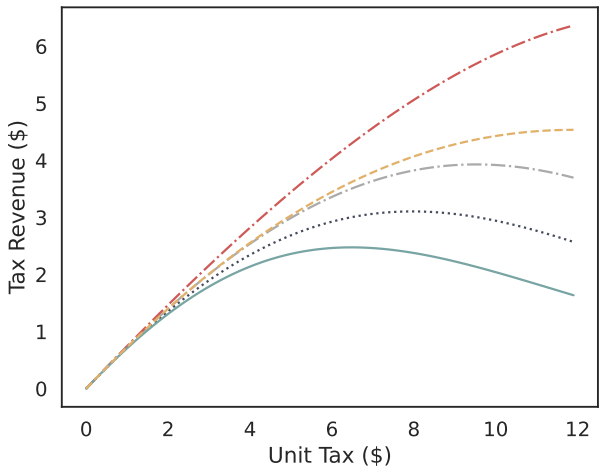
Notes: The figure and table show the distribution and mean squared error (MSE) in prices for the true model (Bertrand and profit-weight with $\kappa = 0.5$), standard models, and the flexible supply model in the single-product merger simulation exercise. Panels A and B correspond to results for Bertrand and profit-weight DGPs without economies of scale, while Panels C and D include economies of scale. The model is trained on triopolies. The MSE is computed on a hold-out test sample restricted to markets where merging firms exist. The neural network used in estimation is small with a 3×3 hidden layer. The models are trained on $T = 1,000$ markets. Profit-weight models include demand derivatives.

FIGURE A2: Laffer Curves for Unit and Ad Valorem Taxes

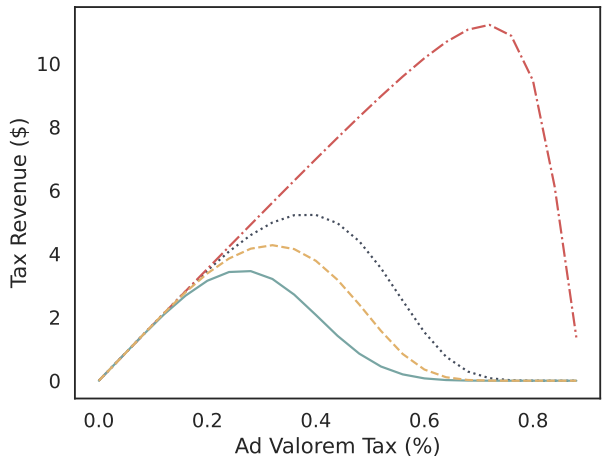
Panel A. Bertrand DGP, Unit Taxes



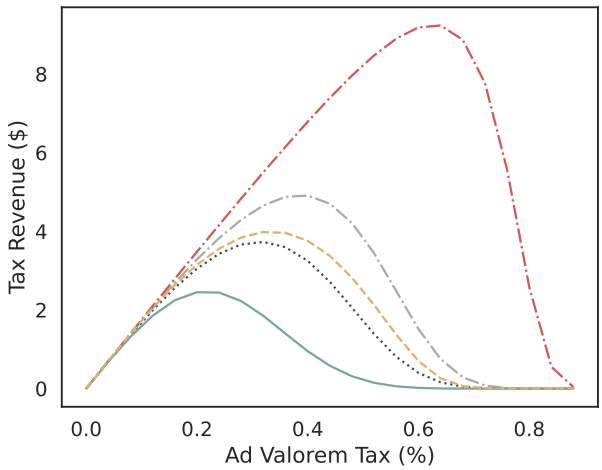
Panel B. Profit-Weight DGP, Unit Taxes



Panel C. Bertrand DGP, Ad Valorem Taxes



Panel D. Profit-Weight DGP, Ad Valorem Taxes

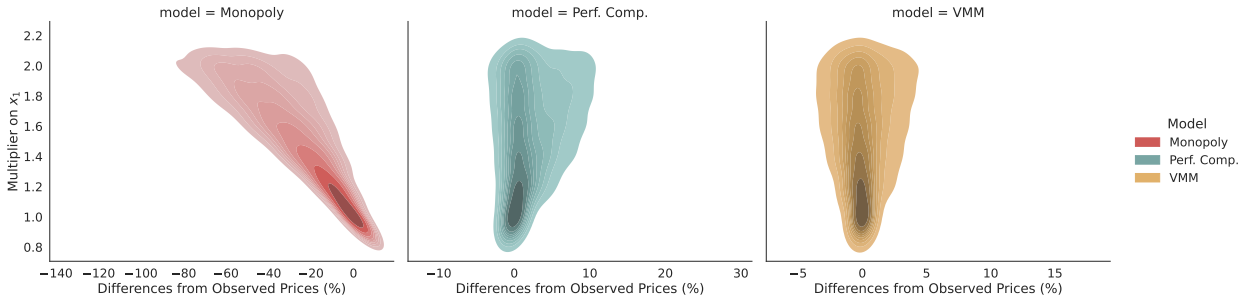


Panel E. Legend and MSEs across Panels A-D

Fitted Model	Panel DGPs			
	A. Bertrand (Unit)	B. Profit-Weight (Unit)	C. Bertrand (AV)	D. Profit-Weight (AV)
Bertrand	--	0.38	--	1.01
Monopoly	5.09	1.49	33.38	17.21
Perf Comp	0.45	1.43	2.79	4.11
Flex Supply	0.02	0.10	0.86	0.46

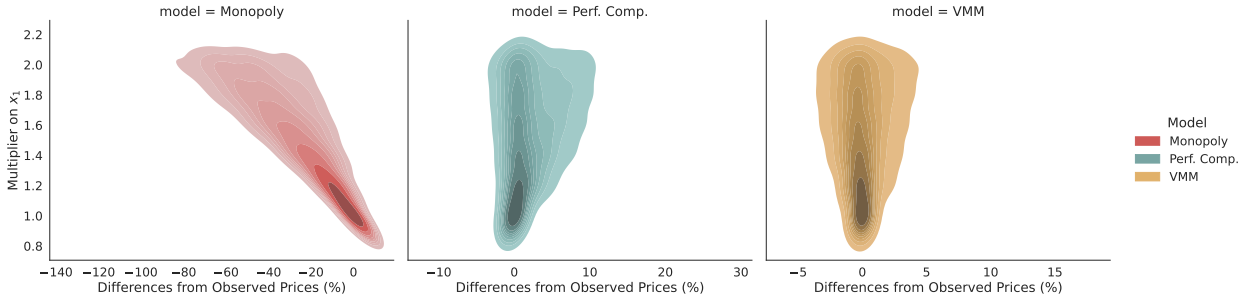
Notes: The figure displays Laffer curves for the flexible model estimated with VMM and a set of standard models under different conduct assumptions. This set of simulations limits the variation available in tax rates. Specifically, unit taxes are drawn from the uniform distribution $U(1,2)$ and ad valorem tax rates are drawn from the uniform distribution $U(0,0.2)$. Panel A shows the Laffer curve for unit taxes under a Bertrand assumption. Panel B shows the Laffer curve for unit taxes under a profit-weight assumption with $\kappa = 0.5$. Panel C shows the Laffer curve for ad valorem taxes under a Bertrand assumption. Panel D shows the Laffer curve for ad valorem taxes under a profit-weight assumption with $\kappa = 0.5$. The neural network used in estimation is small with a 3×3 hidden layer and $T = 100$ markets.

FIGURE A3: Product Characteristics Regulation, Bertrand DGP



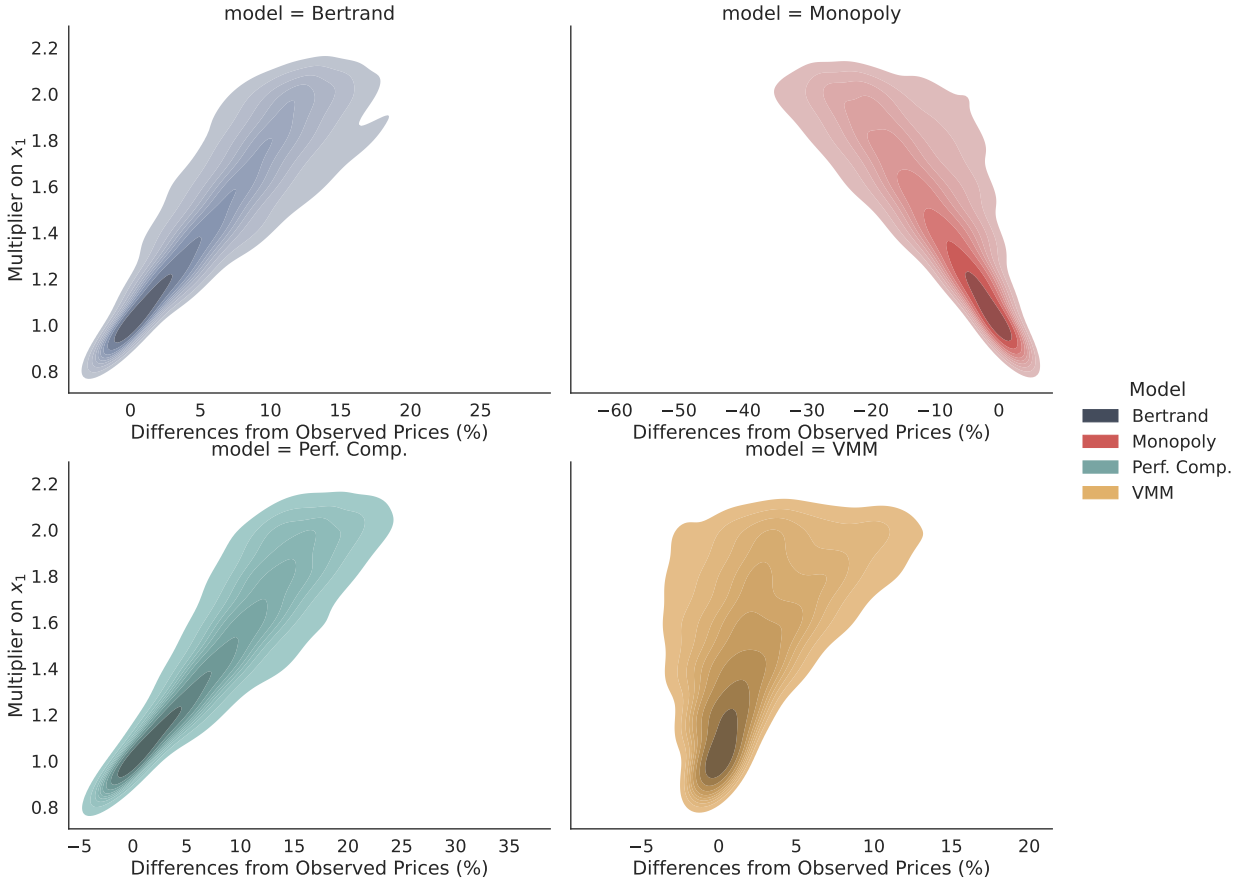
Notes: The figure displays counterfactuals in which product characteristics change for the Bertrand DGP. The regulation increases x_1 by varying factors. The neural network used in estimation is small with a 3×3 hidden layer and $T = 10,000$ markets.

FIGURE A4: Cost Shifter Regulation, Bertrand DGP



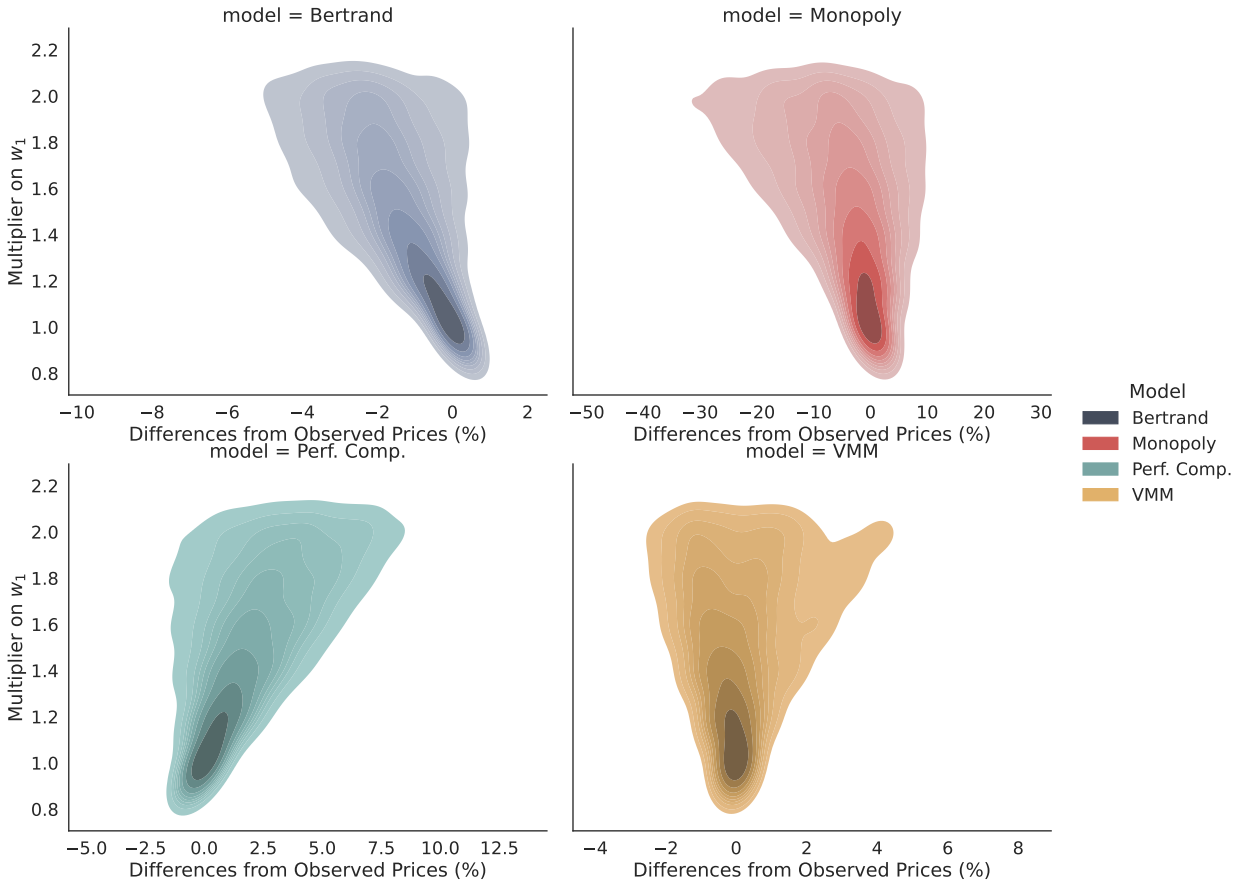
Notes: The figure displays counterfactuals in which product characteristics change for the Bertrand DGP. The regulation increases w_1 by varying factors. The neural network used in estimation is small with a 3×3 hidden layer and $T = 10,000$ markets.

FIGURE A5: Product Characteristics Regulation, Profit-Weight DGP



Notes: The figure displays counterfactuals in which product characteristics change for the profit-weight DGP ($\kappa = 0.5$). The regulation increases x_1 by varying factors. The neural network used in estimation is small with a 3×3 hidden layer and $T = 10,000$ markets.

FIGURE A6: Cost Shifter Regulation, Profit-Weight DGP



Notes: The figure displays counterfactuals in which cost shifters change for the profit-weight DGP ($\kappa = 0.5$). The regulation increases x_1 by varying factors. The neural network used in estimation is small with a 3×3 hidden layer and $T = 10,000$ markets.

TABLE A3: Pass-through MSE

Panel A: Flex Supply in Bertrand DGP		Panel B: Flex Supply Profit-Weight DGP	
0.028	0.004	0.024	0.011
0.006	0.015	0.007	0.026

Notes: The table reports the pass-through behavior of the true models (Bertrand and profit-weight with $\kappa = 0.5$) and the flexible supply model. Own and rival costs are increased by 10% (separately) and firms respond in prices under each model, loading cost increases on the model-implied ω terms. Reported numbers are mean-squared errors of price changes divided by the relevant cost changes for 50 markets in each DGP. Panel A reports the MSE of the flexible supply model in the Bertrand DGP. Panel B reports the same in the profit-weight DGP.

B Economic and Statistical Restrictions

As noted in the main text, one may want to impose additional economic or statistical restrictions on the h function. For instance, one may want h_j to be decreasing in that product's own demand elasticity, as is the case in many standard models. This is in the spirit of the micro-founded economic restrictions that are imposed on nonparametric demand systems in Compiani (2022). It is possible to add these restrictions to our model through regularization, restrictions on weights and activation functions, neural network architecture, or some combination of these. We discuss each of these in turn within the example of monotonicity in own demand elasticities.

The first approach is to incorporate additional components in the regularization term R_N in Equation (4). Define a set of n own-demand elasticities in increasing order as $D = (D_{(1)}, ..., D_{(n)})$. An example of a simple regularization term R^M for monotonicity is the following:

$$R^M(h) = \sum_{i=2}^{n} (\max\{h(D_{(i)}) - h(D_{(i-1)}), 0\})^2$$

We condition on s_t, w_t, θ, and $\mathscr{H}_t$ in the markup function h, suppressing them for notational simplicity. If $h(D_{(i)}) \leq h(D_{(i-1)})$, there is no additional penalty on the markup function, but otherwise, we penalize the squared first difference of own-demand elasticities in the spirit of a ridge regression. We note that the choice of regularization is a degree of freedom for the researcher.

The last two approaches are related to the rich computer science literature on monotonic neural networks, starting with Sill (1997). The first approach enforces constraints on weights and activation functions in particular neural network layers, e.g., You, Ding, Canini, Pfeifer, and Gupta (2017). The user can specify inputs, such as own-demand elasticities, in which the output is monotonic. A second approach, detailed in Wehenkel and Louppe (2019), uses the architecture of the neural network to enforce a constant sign of the derivative of the approximated function without imposing additional constraints. In our specific example, we can restrict the derivative of h with respect to own-demand elasticity to be negative.

The discussion above focuses on a single example of an economic restriction. These approaches can be adapted and combined to impose additional economic or statistical restrictions on the markup function.

C Additional Empirical Results

C.1 Data Construction

We construct a database of the US airline industry from 2005-2019. We obtained a quarterly random 10 percent sample of purchased airline tickets from the well-known Airline Origin and Destination Survey (DB1B) database released by the US Department of Transportation. Following Azar et al. (2018) and Kennedy et al. (2017), a market is defined as a pair of cities, regardless of the flight direction. We match cities to Metropolitan Statistical Areas and collect data on the populations of these MSAs from the Bureau of Economic Analysis. A product is a one-way trip that services a particular city-pair and is defined at the carrier-market-quarter level. Market sizes are measured as the geometric mean of the origin-destination endpoint populations

C.1.1 Sample Selection

We exclude markets with fewer than 20 passengers per day, as airline behavior on these thin, possibly seasonal, routes is unlikely to represent normal competitive behavior in the industry. We also drop itineraries with a ticket carrier change at the connecting airport since these tickets cannot be assigned to a unique ticketing carrier. Finally, we drop every ticket with a fare lower than $25 and higher than $2,500 since these tickets are likely the result of reporting errors.

For each carrier-market-quarter, we begin by calculating the product's average price, total passengers, and average distance. Additionally, we construct each product's extra miles - the difference between the average distance in miles and the nonstop distance in the market - and the fraction of nonstop tickets sold. Averages are weighted by the number of passengers. We remove any products with fewer than 800 quarterly passengers. These products arguably have a weak impact on the competitive behavior of carriers with higher market shares, and although this is standard practice in this literature (e.g. Berry and Jia (2010)) because the dimensionality of the input space is very important in our application, this rule allows us to find more markets with effectively fewer carriers. Summary statistics for the analysis sample are presented in Table C1.

TABLE C1: Summary Statistics

Statistic	Mean	St. Dev.	Min	Pctl(25)	Median	Pctl(75)	Max
Average Fare	216.782	82.819	25.000	169.774	209.072	252.704	2,492.017
Total Passengers	352.395	1,184.007	1	19	69	210	70,909
Average Distance	1,386.534	688.694	67.000	861.000	1,255.081	1,872.045	7,731.500
Average Nonstop Miles	1,170.974	618.988	67.000	678.000	1,035.000	1,600.000	2,783.000
Average Extra Miles	215.560	247.441	−1.000	37.548	136.000	305.862	5,118.000
Share Nonstop	0.205	0.363	0.000	0.000	0.000	0.179	1.000
Origin Hub	0.153	0.360	0	0	0	0	1
Dest. Vacation	0.096	0.294	0	0	0	0	1
LCC	0.207	0.405	0	0	0	0	1
Major	0.920	0.271	0	1	1	1	1
Legacy	0.733	0.443	0	0	1	1	1
Presence	0.148	0.142	0.00000	0.050	0.099	0.201	1.000
Num Markets	50.434	30.156	1	27	47	72	146
Share	0.001	0.003	0.00000	0.0001	0.0004	0.001	0.095
Within Share	0.184	0.219	0.00001	0.022	0.096	0.268	1.000

Notes: The table reports summary statistics for our analysis sample. We include the mean, standard deviation, 25th percentile, median, 75th percentile, and max of each of the variables included in our sample.

C.2 Demand Estimation

We include additional details on demand estimation introduced in Section 6.2. In our demand system, consumer i receives utility from product j in market t with the following indirect utility:

$$u_{ijt} = \alpha p_{jt} + x_{jt}\beta + \xi_t + \xi_{jt} + \zeta_{it} + (1-\rho)\varepsilon_{ijt}$$

The vector x includes the share of nonstop flights, average distance in thousands of miles, the squared term of average distance in thousands of miles, and the logged number of fringe firms (plus one to avoid zero issues). The last term is included to focus on the demand for major carriers while controlling for additional variation over time in market structure across origin-destination pairs. The term ξ_t is a set of origin-destination fixed effects. ξ_{jt} and $\zeta_{it} + \varepsilon_{ijt}$ are unobservable shocks at the product-market and individual-product-market levels, respectively. We assume that ε_{ijt} is distributed Type I Extreme Value and ζ_{it} is distributed according to the conjugate distribution. We close the model by normalizing the

utility of consumer i from the outside option to $u_{i0t} = \varepsilon_{i0t}$. Given the structure of utility and distributional assumptions, market shares s_{jt} are a function of observables, unobservables, and parameters in the standard form from Berry, Levinsohn, and Pakes (1995).

The identifying assumption for demand is that the moment condition $\mathbb{E}[\xi_{jt} z_{jt}^D] = 0$ holds for a vector of demand instruments z_{jt}^D. Following Berry et al. (1995), we include the average rival distance, the average number of markets a rival serves, and the number of rival carriers. The last instrument is especially useful for identifying the nesting parameter. The results of demand estimation are presented in Table C2. The results and median own-price elasticities are in line with the literature.

TABLE C2: Demand Estimates

	$\log(s_{jt})$ - $\log(s_{0t})$
Average Fare	-0.0048***
	(0.0004)
$\log(S_t)$	0.8356***
	(0.0133)
Share Nonstop	0.4030***
	(0.0282)
Average Distance (1,000's)	-0.4881***
	(0.0498)
Average Distance^2 (1,000's)	0.0485***
	(0.0045)
log(1 + Num. Fringe)	-0.2642***
	(0.0057)
R^2	0.94238
Observations	1,283,472
Own-price elasticity	-5.1652
Origin-destination fixed effects	✓

Notes: The table presents results from demand estimation. Prices and log-within-shares $\log(S_t)$ are instrumented with the average rival distance, the average number of markets a rival serves, and the number of rival carriers. We include origin-destination fixed effects and cluster standard errors at the origin-destination level. The median own-price elasticity is -5.1652 which is in line with the literature.

C.3 Supply Estimation

In Section 6.2, we briefly introduced the supply side and focused on the flexible markup function. In the Bertrand specification, we assume that inferred marginal costs are linear in observable cost shifters:

$$p_{jt} - \Delta_{jt}^{B} \equiv c_{jt} = w_{jt}\gamma + \Gamma_{m(j)} + \omega_{jt}$$

We include only the average distance in thousands of miles in w_{jt} and origin-destination fixed effects. The results are presented in Table C3. We find that distance is positively associated with marginal costs inferred under the Bertrand assumption of conduct.

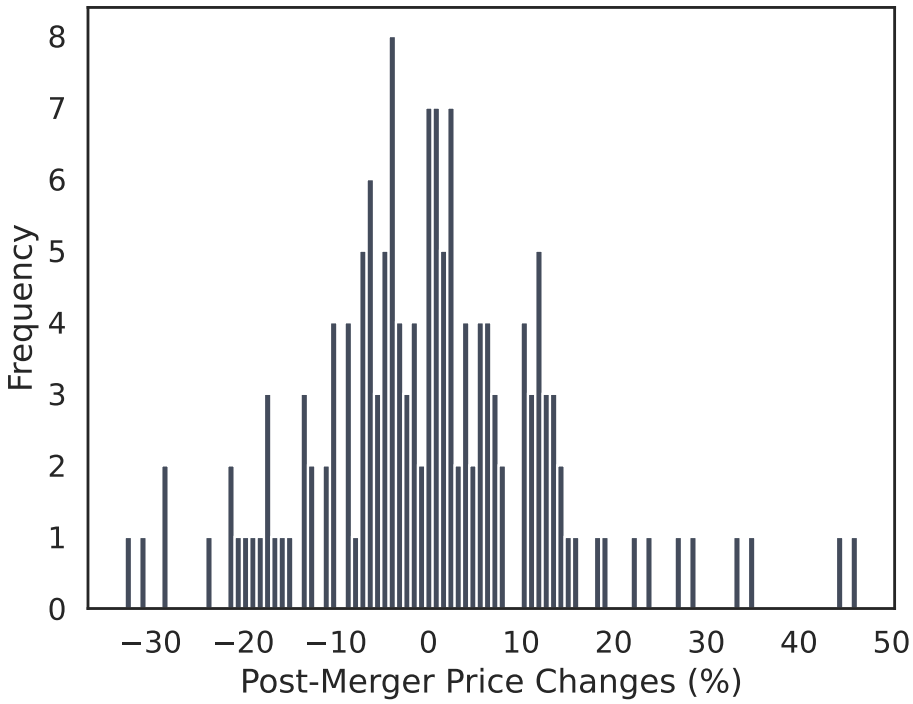
TABLE C3: Bertrand-Implied Marginal Cost Estimates

	Marginal Cost
Average Distance (1,000's)	63.17^{***}
	(0.9502)
R^2	0.42757
Observations	1,283,472
Origin-destination fixed effects	✓

Notes: The table presents results from supply estimation. We include origin-destination fixed effects and cluster standard errors at the origin-destination level.

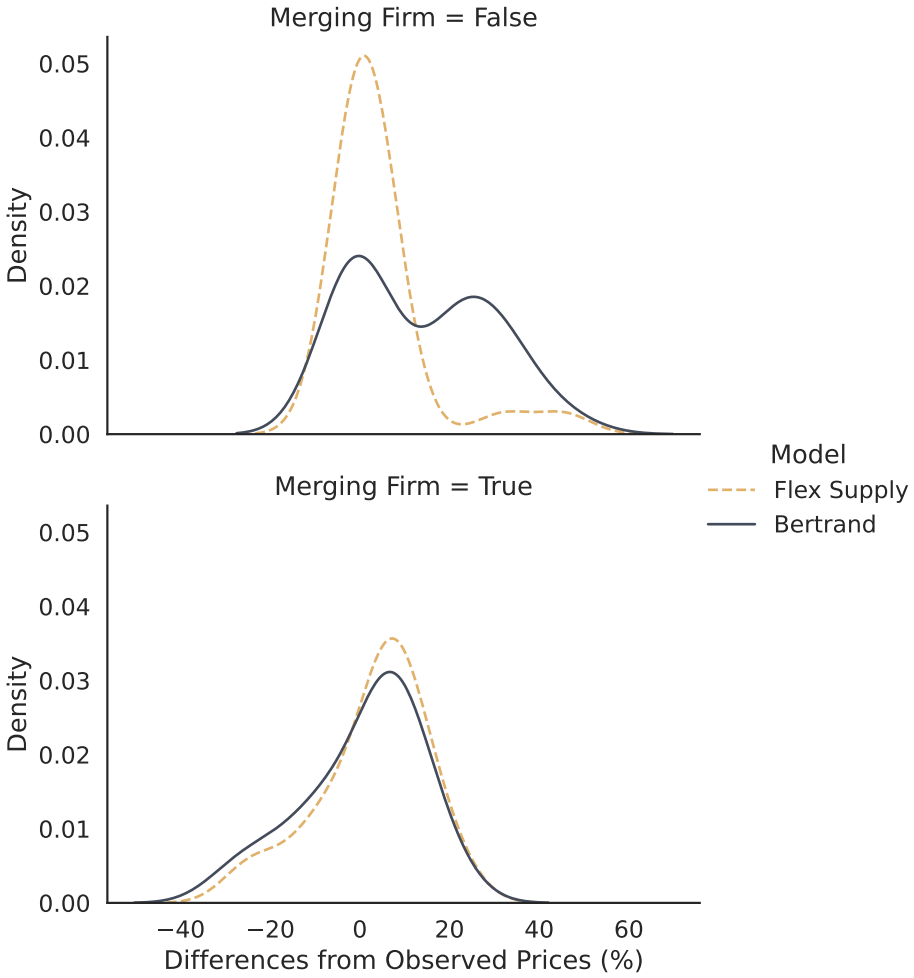
C.4 Counterfactuals

FIGURE C1: Price Change Distribution



Notes: The figure plots the observed post-merger price changes after the US-AA merger.

FIGURE C2: Post-Merger Price Prediction Error by Status



Notes: The figure reports merger simulation results for the flexible model estimated with VMM (in yellow) and the standard merger simulation model (in blue) broken down by firm status. It reports the distribution of percent differences between observed and post-merger predicted prices.

D Using VMM

Researchers can easily use our implementation of VMM in simulations and empirical applications. The models are implemented using the `torch` package in Python. Compute times are manageable by modern standards. Our simulations for $T = 100$ finish within a few minutes; $T = 1,000$ within an hour; and $T = 10,000$ in about a day. The model for our empirical application runs in about 16 hours. We build an easy-to-use wrapper around the machinery of VMM. Researchers specify cost shifters, product characteristics, and instruments they would like to include. We allow the user to optionally include demand derivatives, which we build in the background. We also add an option to look for and use a GPU, if available.

```
# Import packages
from VMM import *
import pandas as pd
import pyblp

# Load data, set variables
data = pd.read_csv('data.csv')
exog, chars, ins = ['w1'], ['x1', 'x2'], ['z1', 'z2', 'z3']
F = np.max(data.groupby('market_ids')['firm_ids'].nunique())

# Estimate demand
results = pyblp.Problem(pyblp.Formulation('0 + prices + x1 + x2'), data).solve()

# Set up scenario
scenario = EmpiricalLogitScenario()
scenario.setup(data, results, exog=exog chars=chars, ins=ins, F=F, add_dsdp=True)
train = scenario.get_dataset('train')
test = scenario.get_dataset('test')
dev = scenario.get_dataset('dev')

# Fit model
markets = data.iloc[train.index].market_ids
n = len(np.unique(markets))
model = get_estimator(scenario, train.z, n=n, use_gpu=True)
model.fit(x=train.x, z=train.z, m=markets, x_dev=dev.x, z_dev=dev.z)
```