Enhancing the Merger Simulation Toolkit with ML/AI

HAROLD D. CHIANG, University of Wisconsin-Madison, USA
JACK COLLISON, University of Wisconsin-Madison, USA
LORENZO MAGNOLFI, University of Wisconsin-Madison, USA
CHRISTOPHER SULLIVAN, University of Calgary, CA

This paper develops a flexible approach to predict the price effects of horizontal mergers using ML/AI methods. While standard merger simulation techniques rely on restrictive assumptions about firm conduct, we propose a data-driven framework that relaxes these constraints when rich market data are available. We develop and identify a flexible nonparametric model of supply that nests a broad range of conduct models and cost functions. To overcome the curse of dimensionality, we adapt the Variational Method of Moments (VMM) [Bennett and Kallus, 2023] to estimate the model, allowing for various forms of strategic interaction. Monte Carlo simulations show that our method significantly outperforms an array of misspecified models and rivals the performance of the true model, both in predictive performance and counterfactual merger simulations. As a way to interpret the economics of the estimated function, we simulate pass-through and reveal that the model learns markup and cost functions that imply approximately correct pass-through behavior. Applied to the American Airlines-US Airways merger, our method produces more accurate post-merger price predictions than traditional approaches. The results demonstrate the potential for machine learning techniques to enhance merger analysis while maintaining economic structure.

Additional Key Words and Phrases: merger simulation, variational method of moments (VMM), neural networks, airline markets

Acknowledgments

<div align="center">Contents</div>

1   Introduction

A central challenge in antitrust policy is assessing the competitive risks of proposed horizontal mergers. As emphasized in the U.S. Horizontal Merger Guidelines, Section 7 of the Clayton Act requires agencies to "arrest anticompetitive tendencies in their incipiency," conducting an assessment of what will likely happen if a merger proceeds as compared to what will likely happen if it does not. While various tools are employed for this analysis, merger simulation based on structural economic models represents a sophisticated quantitative approach that can provide valuable indications of the scale and importance of competition between merging firms. These models typically focus on firms' incentives to change their actions in specific competitive dimensions, such as price, while abstracting from other competitive forces.

The standard merger simulation methodology, developed through seminal contributions [Nevo, 2000, Werden and Froeb, 1994], relies on estimating demand elasticities and modeling firm conduct under specific assumptions about competitive behavior. Typically, firms are assumed to engage in Bertrand-Nash price competition with constant marginal costs. While this framework has proven valuable in many applications, retrospective studies indicate that its predictive performance is mixed. A key limitation identified in the literature is that the restrictive assumptions about firm conduct may not adequately capture the complexity of real-world competitive interactions [see e.g., Peters, 2006]. While one could alternatively pursue a purely data-driven approach using machine learning methods, such an approach would ignore the fundamental simultaneity in market equilibrium and likely produce poor predictions.

This paper proposes a novel approach that maintains the economic structure of merger simulation while relaxing assumptions about firm conduct when sufficiently rich market data are available. Our model maintains the core economic insight that market prices emerge from the simultaneous interaction of demand and supply conditions while relaxing parametric assumptions about both markups and cost functions. The framework respects the structure of oligopoly equilibrium – where prices and quantities jointly solve a system of demand and supply equations. At the same time, we allow for a flexible, nonparametric specification of how firms set markups as a function of market conditions and the marginal cost they face, nesting both the standard Bertrand-Nash model and alternative models of competition. Moreover, economic theory guides our choice of instrumental variables for handling the endogeneity of prices and market shares. However, the high dimensionality of this problem poses significant estimation challenges that cannot be addressed with classical nonparametric methods.

We overcome these challenges by adapting recent advances in machine learning, specifically the Variational Method of Moments (VMM) developed by [Bennett and Kallus, 2023]. VMM combines deep learning with instrumental variables estimation through a novel min-max formulation of 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 avoid 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. NPIV estimators offer flexibility in capturing complex relationships and possess well-established theoretical properties. However, they suffer from the curse of dimensionality, limiting their practicality in settings with multiple endogenous variables. Like many new machine learning-based methods, VMM mitigates this issue by using parametric models (e.g., neural networks) that adaptively grow in complexity with sample size. This approach preserves NPIV's strengths while mitigating dimensionality challenges. 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 post-merger 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.

Monte Carlo simulations demonstrate the strong predictive performance of our approach. When the data-generating process follows a standard model of competition (e.g., Bertrand), VMM recovers predictions close to the true structural model. More importantly, when the true model of competition differs from standard assumptions, our method substantially outperforms misspecified structural approaches. The inclusion of demand derivatives as inputs improves prediction accuracy, particularly with larger neural network architectures in data-rich environments.

Beyond prediction, we address a common criticism of machine learning approaches: their "black box" nature. We devise a method to interpret the supply model implied by our VMM estimates by computing numerical approximations of cost pass-through matrices. We find that the pass-through implied by the flexible model closely aligns with that of the true model, demonstrating that our approach not only performs well in prediction but also captures the underlying economic structure of firm conduct.

The airline industry provides an ideal setting to evaluate our methodology. The industry features substantial variation in market structure both across routes and over time, allowing us to learn about competitive conduct under different market conditions. Additionally, several major mergers in recent years provide opportunities for retrospective analysis. We focus on predicting the price effects of the 2013 American Airlines-US Airways merger, comparing our flexible approach to standard merger simulation techniques.

Our empirical analysis yields several key findings. First, the flexible supply-side specification achieves better in-sample fit than the standard Bertrand-Nash model, reducing the passenger-weighted mean squared error in predicted prices by approximately 40%. More importantly, when predicting post-merger prices in markets affected by the American-US Airways merger, our method produces estimates closer to observed outcomes compared to traditional merger simulation. The passenger-weighted mean squared error in post-merger prices for the Bertrand-Nash model is 365.71, while the passenger-weighted mean squared error for the flexible model is 66.93, a reduction of more than 80%. The median predicted price increase from our flexible approach is 2.05%, closely matching the difference-in-differences estimate of 2.92%, while the standard Bertrand model predicts a median increase of 1.45%.

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 standard market-level price and quantity data. We provide detailed guidance and code for implementation.

As for any method, there are trade-offs to consider. In particular, the method's applicability depends on the empirical setting. In cases where a proposed merger would create market structures radically different from those observed historically, theory-driven approaches may be more appropriate. However, our simulations show that the method performs well even with moderate sample sizes and when the exact post-merger market structure is not observed in the training data, provided there is sufficient variation in market structure to inform prediction. This makes the approach particularly valuable for industries with rich historical data on market structure changes.

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].

Similarly to [Compiani, 2022], who develops a method to estimate demand nonparametrically, our paper proposes a method for nonparametric estimation of the supply side. With similar 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 et al., 2020] provide a theoretical foundation for using deep learning in two-stage econometric procedures, allowing for rich heterogeneity while maintaining model interpretability. [Kaji et al., 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 (NNE) approach trains neural networks to directly estimate structural parameters from data moments. Similar 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].

There has been increasing interest in employing general function approximation techniques - such as deep neural networks and random forests - in a unified approach to IV problems. Several studies, including those by [Dikkala et al., 2020, Lewis and Syrgkanis, 2018, Liao et al., 2020], and [Zhang et al., 2023] have advanced this literature. 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.

This paper also contributes to the literature that seeks to enhance the standard merger simulation toolkit with more flexible and data-driven procedures. [Benkard et al., 2010] and [Bruegge et al., 2024] develop methods based on policy functions. In comparison, our method considers a more restrictive static structure and does not predict post-merger entry or exit. The ML/AI tools we adopt, however, allow us to be nonparametric, and the explicit equilibrium model we develop allows us to quantify welfare and firm profit post-merger.

The remainder of the paper is organized as follows. Section 2 presents the standard merger simulation 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   Background: The Standard Merger Simulation Toolkit

The standard merger simulation toolkit, developed building on the seminal contributions of [Berry and Pakes, 1993, Hausman et al., 1994, Werden and Froeb, 1994], is based on a demand system and an assumed model of conduct. In this section, we introduce the market environment we will consider in the paper and summarize the basic elements of the merger simulation toolkit.[1]

Data-generating Process: We observe pre-merger 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

---

[1]For a more detailed description of the toolkit, see [Miller and Sheu, 2021].

characteristics x_{jt} that enter demand, a vector of cost shifters w_{jt}, and ownership structure $\mathcal{H}_t$ 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:

$$s_t = \jmath(p_t, x_t, \xi_t, \theta_0^D). \tag{1}$$

The term ξ_t is a vector of unobservable product characteristics and θ_0^D is the vector of true demand parameters. 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:

$$p_t = \Delta_0(p_t, s_t, \mathcal{H}_t, \theta_0^D) + c_{0t}. \tag{2}$$

We denote $\Delta_0(p_t, s_t, \mathcal{H}_t, \theta_0^D)$ as the vector of true markups and c_{0t} as the vector of true marginal costs in market t. Therefore, an assumption on conduct, which determines the form of the markup function, and an assumption on how costs are determined are the key ingredients of the supply-side model.

A standard assumption on conduct is Bertrand-Nash, whereby firms are playing a complete information pricing game, and the markup function is given by $\Delta_0(p_t, s_t, \mathcal{H}_t, \theta_0^D) = \left(\mathcal{H}_t \odot \frac{\partial s_t}{\partial p_t}\right)^{-1} s_t$. Other assumptions are possible, including Cournot-Nash, Nash Bargaining, or profit weight models.

Firms' costs are generated by some cost function c, or $c_{0t} = c(q_{jt}, w_{jt}, \omega_{0jt})$, where w_{jt} and ω_{0jt} are, respectively, observed and unobserved cost-shifter variables. The canonical case in the literature is costs that are constant in quantities.

The Merger Simulation Problem: Suppose the researcher wants to predict prices following a horizontal merger that results in a post-merger ownership matrix $\tilde{\mathcal{H}}_t$ in market t. In the standard merger simulation toolkit, the researcher uses a combination of data and assumptions on the data-generating process to assess the effects of the merger.

This analysis typically proceeds in three steps. First, the researcher formulates a parametric demand model and estimates demand primitives $\hat{\theta}^D$.[2] Second, under an assumption on conduct, whereby the researcher specifies a model m with corresponding markup function Δ_m, and the researcher inverts marginal costs $c_{mt} = p_t - \Delta_m(p_t, s_t, \mathcal{H}_t, \hat{\theta}^D)$. Third, the researcher computes post-merger prices $\tilde{p}_t$ under the post-merger ownership structure $\tilde{\mathcal{H}}_t$. Post-merger prices solve a fixed point:

$$\tilde{p}_t = c_{mt} + \Delta_m(\tilde{p}_t, \jmath(\tilde{p}_t, \hat{\theta}^D, \cdot), \tilde{\mathcal{H}}_t, \hat{\theta}^D). \tag{3}$$

For example, under Bertrand-Nash, the markup function is $\left(\tilde{\mathcal{H}}_t \odot \frac{\partial \jmath(\tilde{p}_t, \hat{\theta}^D, \cdot)}{\partial p_t}\right)^{-1} \jmath(\tilde{p}_t, \hat{\theta}^D, \cdot)$. The same procedure applies to other models and is limited by the assumptions on conduct. Assumptions on cost efficiencies generated by the merger, quantified in post-merger cost vectors $\tilde{c}_{mt}$, can be incorporated by using this vector instead of c_{mt} in Equation (3).

Discussion: The framework relies on assumptions on demand and supply. First, the researcher makes assumptions on the demand side to capture how consumers substitute across products. Second, she maintains restrictions on the supply side by specifying firms' cost functions and a model of firm conduct, constant across the pre- and post-merger periods. We discuss these in turn.

In the context of merger review, demand elasticities may be recovered or calibrated from data on diversion, or estimated from simple demand systems [e.g., logit as in Werden and Froeb, 1994]. In the IO literature, a lot of effort has been devoted to estimating flexible models of demand [e.g.,

---

[2]We abstract here from the possibility of demand misspecification and assume the researcher correctly specifies demand.

Berry et al., 1995] that can then be used as input in merger simulation. In what follows, we assume the researcher can either calibrate or estimate substitution.

On the supply side, we first remark on the scope of the exercise. By assuming that conduct is constant before and after the merger, and it is described by oligopoly equilibrium, the model is only capturing unilateral effects of mergers.[3] Within this framework, reliance on the strong assumptions on conduct and cost, such as Bertrand-Nash pricing and constant marginal cost, may have important consequences for prediction. For instance, [Peters, 2006] suggests that alternative assumptions on conduct may lead to more reliable conclusions in his setting of airline mergers. This motivates a more flexible approach.

3 Flexible Model of Supply

We propose a flexible model of supply that maintains the core economic structure of market equilibrium while relaxing standard assumptions about firm conduct. Our approach recognizes that merger simulation is fundamentally a complex prediction problem with simultaneity – prices are equilibrium objects that jointly solve a system of demand and supply equations.

3.1 Model Setup

In general, we can express prices in market t as a function of markups and costs:

$$p_t = \Delta(s_t, p_t, D(s_t, p_t, x_t); \mathcal{H}_t) + c(s_t, w_t, \omega_t), \tag{4}$$

where $\Delta(\cdot)$ is a markup function and $c(\cdot)$ is a cost function. 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 $(w_t, x_t, \omega_t, \xi_t)$ 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:*

$$c(s_t, w_t, \omega_t) = \tilde{c}(s_t, w_t) + \omega_t. \tag{5}$$

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.*

---

[3]Merger may have important coordinated effects [Porter, 2020]. See, e.g., [Miller and Weinberg, 2017] for a study that allows for and measures coordinated effects.

Under these assumptions, we can write the price equation as:

$$p_t = h(s_t, D_t, w_t; \mathcal{H}_t) + \omega_t, \tag{6}$$

where $h(\cdot)$ is a flexible function that captures both markup and cost components. Notably, this formulation does not impose separability between markups and costs. We further discuss this assumption below.

3.2 Discussion

Several aspects of our framework deserve emphasis. First, it significantly generalizes the standard merger simulation toolkit while maintaining its essential economic structure. The formulation in Equation (6) nests the Bertrand-Nash model as a special case but allows for richer forms of strategic interaction. 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]

Second, while our framework is flexible, it is not completely unstructured. The economic content enters through the maintained assumption that prices emerge from equilibrium behavior and through the specification of which variables can enter the pricing function. These restrictions are essential for the identification and interpretation of the results.

There may be more additional economic or statistical restrictions on the h function that a researcher may want to impose. Beyond separability, 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, and we discuss in Appendix B how these can be incorporated within our modeling and estimation framework.

Finally, for merger simulation and other counterfactuals, we can use the estimated function $\hat{h}$ to solve for counterfactual prices that satisfy:

$$\tilde{p}_t = \hat{h}(\jmath(\tilde{p}_t), D(\tilde{p}_t), w_t; \tilde{\mathcal{H}}_t) + \hat{\omega}_t, \tag{7}$$

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 flexible conduct.

3.3 Identification

Equation (6) 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:

$$\mathbb{E}[\omega_{jt} \mid z_{jt}, w_{jt}] = 0, \tag{8}$$

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. Importantly, the simultaneity of the environment, where the demand equation $s_t = \jmath(p_t)$ 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

---

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

relation h_t. Given that Equation (6) 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*

$$\mathbb{E}[\omega_{jt} \mid z_{jt}, w_{jt}] = 0$$

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 $h = \jmath^{-1}$ 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 $B(s_t, D_t, w_t; \mathcal{H}_t)$ with finite expectation, if $\mathbb{E}[B(s_t, D_t, w_t; \mathcal{H}_t) \mid z_{jt}, w_{jt}] = 0$ almost surely, then $B(s_t, D_t, w_t) = 0$ almost surely.*

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

THEOREM 3.1. *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}, w_{jt}):

$$\mathbb{E}[\omega_{jt} \mid z_{jt}, w_{jt}] = \mathbb{E}[h_j(s_t, D_t, w_t; \mathcal{H}_t) \mid z_{jt}, w_{jt}] - \mathbb{E}[p_{jt} \mid z_{jt}, w_{jt}] \tag{9}$$

$$\mathbb{E}[p_{jt} \mid z_{jt}, w_{jt}] = \mathbb{E}[h_j(s_t, D_t, w_t; \mathcal{H}_t) \mid z_{jt}, w_{jt}], \tag{10}$$

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

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

so that

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

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 3.1 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 (4), 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 3.1 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. The next section describes how we overcome these challenges using recent advances in machine learning.

4 Estimation and Inference

We now turn to the estimation of the flexible supply function via VMM. To contextualize VMM as our estimator of choice, we first consider 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, machine learning-based IV estimation methods, such as VMM and the two neural network-based average derivative estimators presented by [Chen et al., 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. Notably, 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.

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

$$\mathbb{E}[p_{jt} - h_j(s_t, D_t, w_t; \theta, \mathcal{H}_t) \mid z_t, w_t] = 0 \tag{11}$$

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

$$\hat{\theta}_N = \operatorname{argmin}_{\theta \in \Theta} \sup_{f \in \mathcal{F}_N} \frac{1}{TJ} \sum_{j,t} f(z_{jt})' \omega_{jt}(\theta) - \frac{1}{4TJ} \sum_{j,t} (f(z_{jt})' \omega_{jt}(\tilde{\theta}_N))^2 - R_N(f, h) \tag{12}$$

$$\text{where} \quad \omega_{jt}(\theta) = p_{jt} - h_j(s_t, D_t, w_t; \theta, \mathcal{H}_t)$$

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 be used to adhere to economic properties of the supply 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 (12) is consistent for the true parameter values θ_0.

4.1 Inference

We now turn to the quantification of uncertainty in the flexible supply function. Importantly, while VMM offers valid element-wise inference procedures for the underlying nuisance parameters, our focus is on predicting post-merger 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 post-merger price predictions through standard errors and confidence intervals,

---

[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.

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 supply function $h : \mathbb{R}^b \supset \Theta \to \mathbb{R}^d$ at a set of d observations, the numerical delta method in [Hong et al., 2015] implies:

$$\sqrt{N}(h(\hat{\theta}_N, \mathcal{H}) - h(\theta_0, \mathcal{H})) \xrightarrow{d} N(0, \nabla_{\theta'} h(\theta_0, \mathcal{H}) \Omega_0^{-1} \nabla_{\theta'} h(\theta_0, \mathcal{H})') \tag{13}$$

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:

$$\beta^T \Omega_0^{-1} \beta = -\frac{1}{4} \inf_{\gamma \in \mathbb{R}^b} \sup_{f \in \mathcal{F}} \mathbb{E}[f(z)' \nabla_\theta \omega(\theta_0) \gamma] - \frac{1}{4} \mathbb{E}[(f(z)' \omega(\theta_0))^2] - 4\gamma'\beta - R_N(f) \tag{14}$$

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 (13). 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]

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

---

for *each index $j \in \{1, ..., d\} \equiv J$* **do**

    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 14

end

Fix critical values $T_\alpha = \{T_{\alpha_k} : k = 1, ..., d\}$ where $\alpha_k = \frac{\alpha}{d+1-k}$

for *each permutation $\tilde{J}$ of indices J* **do**

    Arrange values $\tilde{x}$ and $\hat{\sigma}_{\tilde{x}}$ using permuted indices $\tilde{J}$

    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

end

Collect the simultaneous confidence interval as the union of bounds

---

This novel procedure allows us to quantify uncertainty simultaneously across many predicted prices. Importantly, we can construct confidence intervals on counterfactuals using our method, meaning we can put bounds on predictions for post-merger prices.

---

[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 note that we are still subject to the regularity condition from [Hong et al., 2015].

5   Monte Carlo Simulations

We illustrate the predictive performance of our flexible model with simulations. We evaluate the performance of VMM against the standard merger simulation toolkit, where the researcher imposes a potentially misspecified model. To develop this comparison, we first create an artificial merger simulation exercise, where different models are trained/estimated on pre-merger data, and deployed to predict counterfactual post-merger price changes. To dig deeper into where the differences in performance come from, we also (i) evaluate out-of-sample fit of the different models on a subsample of the pre-merger data, and (ii) compute implied pass-through matrices for each model. Both exercises reveal that VMM outperforms other methods because it is able to learn the true structural relationship on the supply side.

Data Generation: 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 either $J_t = 2$ or $J_t = 3$ for duopolies and triopolies, respectively.

We adopt a simple logit framework for demand with market size normalized to one. Consumer i receives indirect utility from consuming product j in market t, given by:

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

We include a vector of a constant and two observed product characteristics in x_{jt} and the price of the product p_{jt}. 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 $u_{i0t} = \varepsilon_{i0t}$. We draw observed product characteristics independently from normal distributions $N(1, 0.25)$, while ε_{ijt} is assumed to be distributed Type I Extreme Value. The mean taste parameters β are set to $\beta = [-4, 3, 6]$ and the price coefficient is set to $\alpha = -0.25$.

On the supply side, we separately generate data under two conduct models: Bertrand-Nash and profit-weight model. Stacking first-order conditions, the true markups under either model can be represented as:

$$p_t - c_t = \left(-\mathcal{H}_t \odot D'_t(p_t)\right)^{-1} s_t(p_t) \tag{15}$$

Under Bertrand-Nash, the ownership matrix is an identity matrix while under profit-weight conduct, we set $\mathcal{H}_{jkt} = 0.75$ for $j \neq k$ and $\mathcal{H}_{jkt} = 1$ for $j = k$. We assume that the marginal cost of producing product j in market t is given by:

$$c_{jt} = w_{jt}\gamma + \omega_{jt}$$

We include a constant and two observable cost shifters (excluded from demand) in the vector w_{jt}. Marginal cost also depends on a true unobservable cost shock ω_{jt}. We draw observed cost shifters independently from normal distributions $N(1, 0.25)$. The cost parameters are set to $\gamma = [3, 6, 4]$. We adopt the default in `pyblp` by drawing unobserved demand and cost shocks ξ_{jt} and ω_{jt} from a mean-zero bivariate normal distribution with variances of one and correlation of 0.9. Under the given model of conduct, prices solve the fixed point of Equation (15). The Bertrand-Nash and profit-weight simulations yield mean own-price elasticities -3.29 and -4.32, and mean diversions to the outside option of 0.34 and 0.50, respectively.

In the estimation routine, the sample is split into a training set and a hold-out test set, stratified to keep whole markets intact, prevent overfitting, and evaluate out-of-sample performance. The model is trained on duopoly and triopoly markets. We include both market structures to give the estimator variation in market structure to learn the supply function. The simulation environment mimics a merger simulation exercise in that the estimated model has seen enough variation to extrapolate well to post-merger behavior upon merging two firms in a triopoly market.

Merger Simulation Models: We describe now the set of models that we use to generate predictions, and how each is estimated/fit from the pre-merger data. First, within the standard merger simulation toolkit, we impose three different canonical models indexed by m: Bertrand ($m = B$), monopoly ($m = M$), and perfect competition (marginal cost pricing, $m = P$). 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 $p_{jt} - \Delta_{jt}^m = w_{jt}\gamma + \omega_{jt}^m$, which we estimate by OLS. For the VMM model, we estimate the function h:

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

with the VMM formulation, obtaining the implied shocks $\hat{\omega}$.

The neural networks are implemented using the `torch` package in Python. The objective function for the VMM supply function h_j is given by Equation (12). We do not include regularization in our baseline simulations. The variance estimates use the objective function from Equation (14).

When estimating VMM, we consider two aspects of implementation. To illustrate how knowing the demand side can improve performance of the model, we perform estimation with and without the demand derivatives. 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 can improve performance, especially in small samples. We also 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.

5.1 Merger Simulation Results

Now that we have constructed estimators, the ultimate goal of the structural object recovered in estimation is to conduct a merger simulation and compare predictions with the synthetic data.

To obtain post-merger predictions for model m, we find a vector of prices $\tilde{p}_t^m$ in all markets t that simultaneously satisfy the demand system, and the first order conditions given the post-merger ownership matrix $\tilde{\mathcal{H}}_t$. For the standard merger simulation toolkit, we rely on a contraction mapping from [Morrow and Skerlos, 2011]. Using our estimated flexible supply function, we modify this method into a root-finding problem:[7]

$$\tilde{p}_t - \hat{h}(s_t(\tilde{p}_t), D(p_t), w_t; \tilde{\mathcal{H}}_t) - \hat{\omega}_t = 0 \tag{16}$$

We rely on the structure of an estimated demand system $s(\cdot)$ and its derivatives $D(\cdot)$ to solve for equilibrium shares and, subsequently, counterfactual prices under post-merger ownership $\tilde{\mathcal{H}}$. We hold all the parameters in demand and cost at their true or estimated values, and hold the cost shocks (for both the true and imposed model) in the post-merger period fixed at their pre-merger values.

In Table 1, we compare the distributions of percentage discrepancies $(\tilde{p}_{jt}^m - \tilde{p}_{jt})/\tilde{p}_{jt}$ between predicted and actual post-merger prices. We also compute prediction MSEs for each model, obtained as $\sum_{j,t}(\tilde{p}_{jt}^m - \tilde{p}_{jt})^2$ for each model m.

We compare post-merger prices under the true model to prices under different assumptions of conduct to evaluate performance. The monopolist performs the worst with prices far higher than the true prices. Perfect competition significantly underpredicts price changes. Our VMM estimator greatly outperforms all misspecified models and is close to the true prices. It is important to note that the merger of triopolies results in duopolies which the training process is able to see in-sample.

---

[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.

Table 1. Prediction Error for Bertrand DGP Merger Simulation (Duopolies and Triopolies)
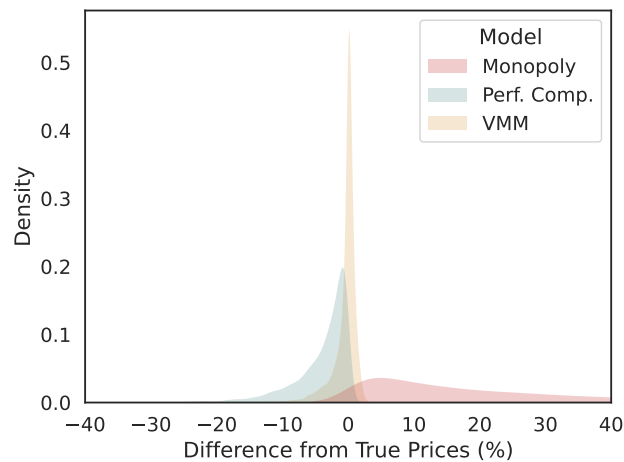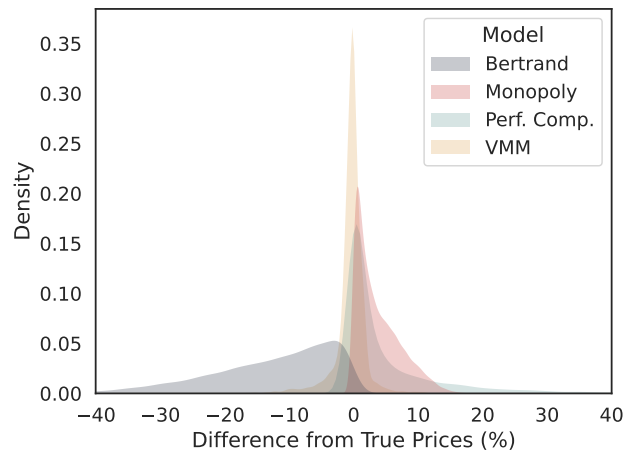

Model	MSE
Monopoly	26.26
Perf. Comp.	1.80
VMM	0.27

Notes: The figure and table show the distribution and mean squared error (MSE) in prices for the true model (Bertrand), a set of standard models, and the VMM model in the merger simulation exercise. 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. The neural network used in estimation is small with a 3×3 hidden layer.

In Table A3, we train on only triopolies so the merger resulting in duopolies is strictly out of sample. The results are very similar: our estimator greatly outperforms all misspecified models.

Table 2. Prediction Error for Profit Weight DGP Merger Simulation (Duopolies and Triopolies)

Model	MSE
Bertrand	15.19
Monopoly	1.55
Perf. Comp.	3.57
VMM	0.24

Notes: The figure and table show the distribution and mean squared error (MSE) in prices for the true model (profit-weight with $\kappa = 0.75$), a set of standard models, and the VMM model in the merger simulation exercise. 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. The neural network used in estimation is small with a 3×3 hidden layer.

The results in Table 2 show similar performance in the profit-weight simulations. Bertrand is now misspecified and performs poorly in the merger simulation exercise. Equilibrium prices generated

from perfect competition and a monopolist also do not perform well. Our method performs again greatly outperforms all misspecified models and is close to the true post-merger prices. We evaluate performance with a model trained solely on tripolies in Table A4 and find that our model again outperforms all others.

5.2   Simulation Results

We compute MSE on a hold-out sample to evaluate performance outside the training sample. The results in Table 3 and Table 4 vary sample sizes, assumptions on conduct, and the architecture of the estimator.[8] The first column describes the sample size in terms of the number of markets and the second column denotes whether demand derivatives were included in training. The remaining columns refer to the MSE of the relevant model, i.e., the true model, Bertrand, monopoly, perfect competition, and VMM, respectively.

Table 3. MSE Across Models (Bertrand DGP, Small Network)

T	True Model	Standard Models			VMM	D_t included
		B	M	P		
100	0.895	0.895	578.773	9.411	1.910	No
					1.708	Yes
1,000	0.889	0.889	1022.812	8.170	2.211	No
					1.065	Yes
10,000	0.962	0.962	1788.816	8.487	2.355	No
					1.146	Yes

Notes: The table reports the mean squared error (MSE) in prices for the true model (Bertrand), 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. The neural network used is small with a 3×3 hidden layer.

Table 3 describes the Bertrand simulations with a small neural network for the flexible models of supply. Generally, the shocks implied by perfect competition ω^P and especially monopoly ω^M are far away from the true shocks ω^B and ω. The shocks implied by our VMM estimator $\hat{\omega}$ outperform all misspecified models. The performance of our estimator without demand derivatives remains stable across sample sizes. In 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. On the whole, our estimator outperforms all misspecified models, sometimes by an order of magnitude.

The results in Table 4 are generated with the profit-weight simulations and a large neural network for the flexible models of supply. The collusive profit-weight model is more difficult for a researcher to estimate because there are free parameters in the ownership matrix. A similar pattern emerges in the results compared to the Bertrand data-generating process. Monopoly- and perfect competition-implied shocks, ω^M and ω^P, respectively, are far from the true shocks, and now Bertrand-implied shocks ω^B are far from the truth as well. Our VMM estimator outperforms all misspecified models with implied shocks $\hat{\omega}$ that are close to the ground truth. The model performs better as the sample size increases, especially with the inclusion of demand derivatives. In Table A2, 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.

---

[8]Additional results can be found in Table A1 and Table A2.

Table 4. MSE Across Models (Profit Weight DGP, Large Network)

T	True Model	Standard Models			VMM	D_t included
		B	M	P		
100	0.895	8.351	5.973	13.372	1.845	No
					2.816	Yes
1,000	0.889	7.877	6.650	9.962	1.673	No
					1.297	Yes
10,000	0.962	8.719	7.049	10.853	1.175	No
					1.177	Yes

Notes: The table reports the mean squared error (MSE) in prices for the true model (profit weight with $\kappa = 0.75$), 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. The neural network used in estimation is large with a 100×100 hidden layer.

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 structural supply function $\hat{h}$. In an attempt to "peek inside the black box," we simulate the pass-through of the resulting supply function and compare it to pass-through under the true model. The pass-through matrix shows us a rough shape of the supply function as we change own and rival costs.

To compute pass-through for the flexible supply function, we increase costs by 10 percent, loading increases on the residuals $\hat{\omega}$. We then solve for equilibrium prices under different models of conduct and compare the prices before and after the cost change. Below, we report pass-through matrices for markets with median post-merger inside shares from our different simulation environments.

Table 5. Pass-through Comparison: Bertrand vs Profit Weight DGP

Panel A: Bertrand DGP				**Panel B: Profit Weight DGP**			
True Model		VMM		True Model ($\kappa = 0.75$)		VMM	
0.60	0.10	0.65	0.12	0.98	-0.00	0.96	-0.01
0.16	0.78	0.12	0.72	-0.43	0.41	-0.44	0.46

Notes: The table reports the pass-through behavior of the true models (Bertrand and profit weight with $\kappa = 0.75$) and VMM. 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. Market characteristics for Panel A: $c_1 = 15.08$, $c_2 = 15.33$, $s_1 = 0.61$, $s_2 = 0.04$. Market characteristics for Panel B: $c_1 = 15.72$, $c_2 = 10.20$, $s_1 = 0.03$, $s_2 = 0.60$.

Table 5 shows pass-through matrices for both Bertrand and profit-weight simulations. Panel A compares the true pass-through matrix in the Bertrand environment with the pass-through implied by VMM, while Panel B presents the same comparison for the profit-weight environment. In both cases, the pass-through implied by VMM is remarkably close to the true pass-through matrix. The flexible model learns markup and cost functions that imply approximately correct pass-throughs. 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.

5.4   Inference Performance

After evaluating predictive performance in sample and in counterfactuals, and the interpretability of the flexible markup and functions, we quantify the uncertainty of the predictions. For ease of exposition, we pick a single observation at which to evaluate our inference procedure.

Table 10.   Inference Comparison by Sample Size (Small Network)

Model	Sample Size	ψ	$\hat{\psi}$	$\hat{\sigma}/\sqrt{N}$	Interval
Bertrand	N = 253	21.014	21.092	0.817	[18.673, 23.512]
Bertrand	N = 2,579	20.341	20.474	0.057	[20.305, 20.642]
Profit Weight	N = 253	17.321	12.907	0.309	[11.991, 13.822]
Profit Weight	N = 2,579	17.375	15.554	0.099	[15.261, 15.847]

Notes: The table reports the VMM variance estimates for a single observation in the data under different models and sample sizes. The model is trained on duopolies and triopolies. The true price ψ and the model-implied price $\hat{\psi}$. The standard error $\hat{\sigma}/\sqrt{N}$ is reported with the implied confidence interval. The neural network used in estimation is small with a 3×3 hidden layer.

The results in Table 10 present the estimated standard errors and confidence intervals for our predictions under different simulation environments. The first two rows show that our predictions $\hat{\psi}$ for the Bertrand simulation are close to the true prices ψ, even in small samples, and the resulting standard errors and confidence intervals are small. Similarly, the profit-weight environment has quantifiable and tight confidence intervals that are shrinking with increasing sample sizes.

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 et al., 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.

However, we emphasize the illustrative nature of our application. Recent papers have highlighted the dynamic nature of pricing and demand in this market [e.g., Aryal et al., 2024, Betancourt et al., 2022, Hortaçsu et al., 2024, Williams, 2022], and the role of endogenous network structure [Bontemps et al., 2023, Ciliberto et al., 2021, Li et al., 2022, Yuan and Barwick, 2024]. We abstract away from these important elements to keep our application tied to the standard merger simulation toolkit.

6.1   Background and Data

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 C1. The earliest merger in our analysis 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 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.

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 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 (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.

We evaluate the retrospective unilateral price effects of major airline mergers following previous studies of this industry (e.g., [Peters, 2006]) and other reduced-form investigations of the price effect of mergers (e.g., [Ashenfelter et al., 2013]) by creating control groups of markets not affected by the merger. The rich cross-section of separate markets where airlines compete allows to distinguish between overlap markets – markets where merging firms both have a sizable presence, which are thus likely to be affected by the mergers – and markets where there is no overlap between merging airlines. As in [Peters, 2006], we define overlap markets as those in which both merging airlines operate flights for the four quarters before the merger. We use the four quarters surrounding the merger approval as our pre- and post-merger samples. The sample is restricted to markets with at most five firms in the pre-merger period. The results, presented in Table C2, show limited effects of the first two mergers while the American-US merger resulted in higher prices.

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 C3. 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 technique described in Section 4.[9] 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. As in the difference-in-differences, 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 C2.

Fig. 1. Merger Simulation Results
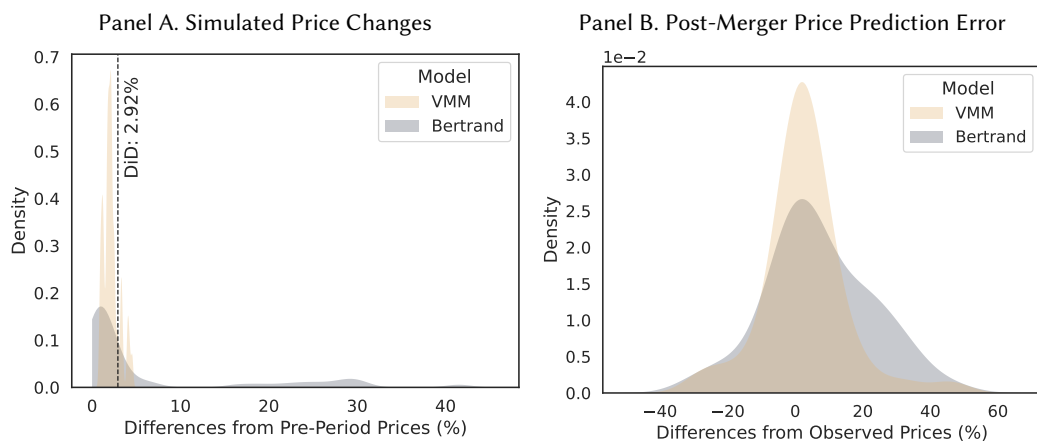
Notes: The figure reports merger simulation results for the flexible model estimated with VMM (in yellow) and the standard merger simulation model (in blue). Panel A reports the distribution of percent differences between pre-merger and post-merger predicted prices. Panel B reports the distribution of percent differences between observed and post-merger predicted prices. Figure C3 breaks down Panel B by merger status.

Figure 1 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 2.16% (median 2.05%). This is in line with

---

[9]In our baseline specification, we do not include demand derivatives, although the model should fit them given the data at hand. As a robustness check, we fit a model that includes demand derivatives but found negligible differences in performance.

the descriptive DiD result, which found a retrospective price change in the data of 2.92% in the markets we examine. 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). The mass of large price increases are predicted 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.

To analyze the source of these different predictions, we now compare whether the flexible model or the standard merger simulation model more closely matches realized post-merger prices. Panel B of Figure 1 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 within 20% of realized prices, and an MSE of 67. Instead, the standard merger simulation method over-predicts changes and has an overall MSE of 366. In sum, our flexible model substantially outperforms the standard toolkit when predicting post-merger prices for the American-US Airways merger.

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.

The results are presented in Appendix Figure C4, where we show 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 (mostly between 5-10 for predicted prices in the range 100-300) in which we can make precise point estimates even with a high-dimensional supply function.

7   Conclusion

This paper demonstrates how machine learning methods can enhance merger analysis while maintaining economic structure. Our approach relaxes standard assumptions about firm conduct when rich market data are available, adapting recent advances in variational inference to estimate a flexible model of supply that nests both Bertrand-Nash and alternative models of competition. Applied to the American Airlines-US Airways merger, the method produces more accurate predictions of post-merger prices compared to traditional merger simulation approaches, while maintaining computational tractability. The framework also provides valid statistical inference, allowing antitrust practitioners to quantify uncertainty in their predictions. Beyond mergers, our methodology opens new possibilities for empirical work in IO by showing how machine learning tools can be integrated into structural economic models without sacrificing their economic content.

The method involves important trade-offs and limitations. To mention one, while our method performs well when the post-merger market structure has some precedent in the training data, it may be less reliable for mergers that would create entirely novel market structures, or for which the researcher lacks relatively rich pre-merger data on market outcomes. An important direction for future research is developing hybrid approaches that combine data-driven flexibility with economic restrictions in settings where data are limited. Despite these limitations, our results suggest that careful application of modern machine learning methods can significantly improve our ability to predict merger effects while maintaining the conceptual foundations that make merger simulation such a valuable tool for antitrust analysis.

References

Gaurab Aryal, Charles Murry, and Jonathan Williams. 2024. Price discrimination in international airline markets. *Review of Economic Studies* 91, 2 (2024), 641–689.

Orley C Ashenfelter, Daniel S Hosken, and Matthew C Weinberg. 2013. The price effects of a large merger of manufacturers: A case study of Maytag-Whirlpool. *American Economic Journal: Economic Policy* 5, 1 (2013), 239–61.

José Azar, Martin C Schmalz, and Isabel Tecu. 2018. Anticompetitive effects of common ownership. *The Journal of Finance* 73, 4 (2018), 1513–1565.

C Lanier Benkard, Aaron Bodoh-Creed, and John Lazarev. 2010. Simulating the dynamic effects of horizontal mergers: Us airlines. *Working Paper* (2010).

Andrew Bennett and Nathan Kallus. 2023. The Variational Method of Moments. *Journal of the Royal Statistical Society Series B: Statistical Methodology* 85, 3 (2023), 810–841.

Steven Berry and Philip Haile. 2014. Identification in Differentiated Products Markets Using Market Level Data. *Econometrica* 82, 5 (2014), 1749–1797.

Steven Berry and Panle Jia. 2010. Tracing the woes: An empirical analysis of the airline industry. *American Economic Journal: Microeconomics* 2, 3 (2010), 1–43.

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

Steven Berry and Ariel Pakes. 1993. Some Applications and Limitations of Recent Advances in Empirical Industrial Organization: Merger Analysis. *American Economic Review* 83, 2 (1993), 247–252.

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

Steven T Berry. 1994. Estimating discrete-choice models of product differentiation. *The RAND Journal of Economics* 25, 2 (1994), 242–262.

Steven T. Berry, Michael Carnall, and Pablo T. Spiller. 2006. Airline hubs: costs, markups and the implications of customer heterogeneity. *Competition policy and antitrust* (2006).

Jose Betancourt, Ali Hortaçsu, Aniko Oery, and Kevin Williams. 2022. *Dynamic price competition: Theory and evidence from airline markets*. Technical Report. NBER working paper #30347.

Christian Bontemps, Cristina Gualdani, and Kevin Remmy. 2023. *Price competition and endogenous product choice in networks: Evidence from the us airline industry*. Technical Report. Working Paper.

Chris Bruegge, Gautam Gowrisankaran, and Alex Gross. 2024. Using Policy Functions to Estimate Merger Impacts: an Application to Jet-Blue Spirit. *Working Paper* (2024).

Marine Carrasco, Jean-Pierre Florens, and Eric Renault. 2007. Linear inverse problems in structural econometrics estimation based on spectral decomposition and regularization. *Handbook of econometrics* 6 (2007), 5633–5751.

Jiafeng Chen, Xiaohong Chen, and Elie Tamer. 2023. Efficient estimation of average derivatives in NPIV models: Simulation comparisons of neural network estimators. *Journal of Econometrics* 235, 2 (2023), 1848–1875.

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

Federico Ciliberto, Charles Murry, and Elie Tamer. 2021. Market Structure and Competition in Airline Markets. *Journal of Political Economy* 129, 11 (2021), 2995–3038.

Federico Ciliberto and Jonathan W Williams. 2014. Does multimarket contact facilitate tacit collusion? Inference on conduct parameters in the airline industry. *The RAND Journal of Economics* 45, 4 (2014), 764–791.

Giovanni Compiani. 2022. Market Counterfactuals and the Specification of Multiproduct Demand: A Nonparametric Approach. *Quantitative Economics* 13, 2 (2022), 545–591.

Christopher Conlon and Jeff Gortmaker. 2020. Best Practices for Differentiated Products Demand Estimation with pyblp. *RAND Journal of Economics* 51, 4 (2020), 1108–1161.

Adam Dearing, Lorenzo Magnolfi, Daniel Quint, Christopher Sullivan, and Sarah Waldfogel. 2024. Learning Firm Conduct: Pass-through as a Foundation for Instrument Relevance. (2024). working paper.

Nishanth Dikkala, Greg Lewis, Lester Mackey, and Vasilis Syrgkanis. 2020. Minimax estimation of conditional moment models. *Advances in Neural Information Processing Systems* 33 (2020), 12248–12262.

Max H Farrell, Tengyuan Liang, and Sanjog Misra. 2020. Deep learning for individual heterogeneity: An automatic inference framework. *arXiv preprint arXiv:2010.14694* (2020).

Amit Gandhi and Jean-François Houde. 2020. *Measuring Firm Conduct in Differentiated Products Industries*. Technical Report. Working Paper.

Jerry Hausman, Gregory Leonard, and J Douglas Zona. 1994. Competitive analysis with differenciated products. *Annales d'Economie et de Statistique* (1994), 159–180.

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

Han Hong, Aprajit Mahajan, and Denis Nekipelov. 2015. Extremum estimation and numerical derivatives. *Journal of Econometrics* (2015), 250–263.

Ali Hortaçsu, Olivia Natan, Hayden Parsley, Timothy Schwieg, and Kevin Williams. 2024. Organizational Structure and Pricing: Evidence from a Large US Airline. *Quarterly Journal of Economics* 139, 2 (2024), 1149–1199.

Tetsuya Kaji, Elena Manresa, and Guillaume Pouliot. 2023. An adversarial approach to structural estimation. *Econometrica* 91, 6 (2023), 2041–2063.

Pauline Kennedy, Daniel P O'Brien, Minjae Song, and Keith Waehrer. 2017. The competitive effects of common ownership: Economic foundations and empirical evidence. *Available at SSRN 3008331* (2017).

Greg Lewis and Vasilis Syrgkanis. 2018. Adversarial generalized method of moments. *arXiv preprint arXiv:1803.07164* (2018).

Sophia Li, Joe Mazur, Yongjoon Park, James Roberts, Andrew Sweeting, and Jun Zhang. 2022. Repositioning and Market Power after Airline Mergers. *RAND Journal of Economics* 53, 1 (2022), 166–199.

Luofeng Liao, You-Lin Chen, Zhuoran Yang, Bo Dai, Mladen Kolar, and Zhaoran Wang. 2020. Provably efficient neural estimation of structural equation models: An adversarial approach. *Advances in Neural Information Processing Systems* 33 (2020), 8947–8958.

Nathan Miller and Gloria Sheu. 2021. Quantitative Methods for Evaluating the Unilateral Effects of Mergers. *Review of Industrial Organization* 58 (2021), 143–177.

Nathan Miller and Matthew Weinberg. 2017. Understanding the Price Effects of the MillerCoors Joint Venture. *Econometrica* 85, 6 (2017), 1763–1791.

Ross Morrow and Steven Skerlos. 2011. Fixed-Point Approaches to Computing Bertrand-Nash Equilibrium Prices Under Mixed-Logit Demand. *Operations Research* 59, 2 (2011), 328–345.

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

Taisuke Otsu and Martin Pesendorfer. 2024. *Conduct in the Soft Drink Market: A Mechanism Design Approach.* Technical Report. Working Paper.

Craig Peters. 2006. Evaluating the Performance of Merger Simulation: Evidence from the US Airline Industry. *Journal of Law and Economics* 49, 2 (2006), 627–649.

Robert Porter. 2020. Mergers and Coordinated Effects. *International Journal of Industrial Organization* 73 (2020), 102583.

Joseph Sill. 1997. Monotonic networks. *Advances in neural information processing systems* 10 (1997).

Antoine Wehenkel and Gilles Louppe. 2019. Unconstrained monotonic neural networks. *Advances in neural information processing systems* 32 (2019).

Yanhao Wei and Zhenling Jiang. 2024. Estimating parameters of structural models using neural networks. *Marketing Science* (2024).

Gregory J Werden and Luke M Froeb. 1994. The effects of mergers in differentiated products industries: Logit demand and merger policy. *JL Econ. & Org.* 10 (1994), 407.

Kevin Williams. 2022. The Welfare Effects of Dynamic Pricing: Evidence from Airline Markets. *Econometrica* 90, 2 (2022), 831–858.

Seungil You, David Ding, Kevin Canini, Jan Pfeifer, and Maya Gupta. 2017. Deep lattice networks and partial monotonic functions. *Advances in neural information processing systems* 30 (2017).

Zhe Yuan and Panle Jia Barwick. 2024. *Network Competition in the Airline Industry: An Empirical Framework.* Technical Report. National Bureau of Economic Research.

Rui Zhang, Masaaki Imaizumi, Bernhard Schölkopf, and Krikamol Muandet. 2023. Instrumental variable regression via kernel maximum moment loss. *Journal of Causal Inference* 11, 1 (2023), 20220073.

A   Additional Simulation Results

Table A1. MSE Across Models (Bertrand DGP, Large Network)

| T	True Model	Standard Models			VMM	D_t included
		B	M	P		
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

Notes: The table reports the mean squared error (MSE) in prices for the true model (Bertrand), 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. The neural network used in estimation is large with a 100×100 hidden layer.

Table A2. MSE Across Models (Profit Weight DGP, Small Network)

T	True Model	Standard Models			VMM	D_t included
		B	M	P		
100	0.895	8.351	5.973	13.372	3.355	No
					3.694	Yes
1,000	0.889	7.877	6.650	9.962	3.120	No
					2.006	Yes
10,000	0.962	8.719	7.049	10.853	2.993	No
					2.170	Yes

Notes: The table reports the mean squared error (MSE) in prices for the true model (profit weight with $\kappa = 0.75$), 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. The neural network used in estimation is small with a 3×3 hidden layer.

Table A3. Prediction Error for Bertrand DGP
Merger Simulation (Triopolies)
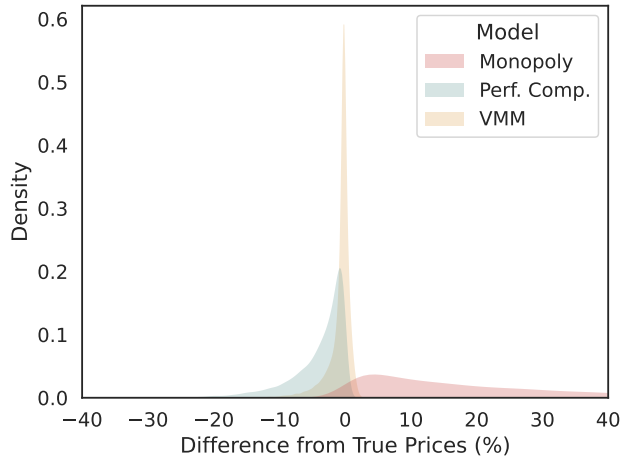
Model	MSE
Bertrand	0.00
Monopoly	25.94
Perf. Comp.	1.81
VMM	0.28

Notes: The figure and table show the distribution and mean squared error (MSE) in prices for the true model (Bertrand), a set of standard models, and the VMM model in the merger simulation exercise. The model is trained on triopolies. The MSE is computed on a hold-out test sample restricted to markets where merging firms are present. The neural network used in estimation is small with a 3×3 hidden layer.
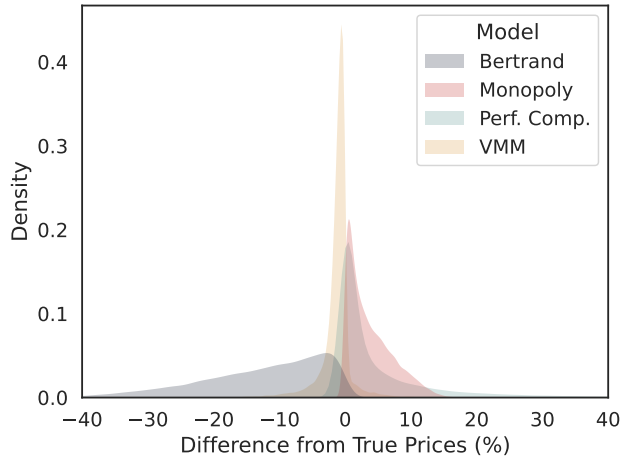
Table A4. Prediction Error for Profit Weight DGP
Merger Simulation (Triopolies)

Model	MSE
Bertrand	15.02
Monopoly	1.53
Perf. Comp.	3.54
VMM	0.28

Notes: The figure and table show the distribution and mean squared error (MSE) in prices for the true model (Bertrand), a set of standard models, and the VMM model in the merger simulation exercise. The model is trained on triopolies. The MSE is computed on a hold-out test sample restricted to markets where merging firms are present. The neural network used in estimation is small with a 3×3 hidden layer.

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 (12). 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 $\mathcal{H}_t$ in the supply function h, suppressing them for notational simplicity. If $h(D_{(i)}) \leq h(D_{(i-1)})$, there is no additional penalty on the supply 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 et al., 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 supply 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.

Fig. C1. Concentration in the Airline Industry
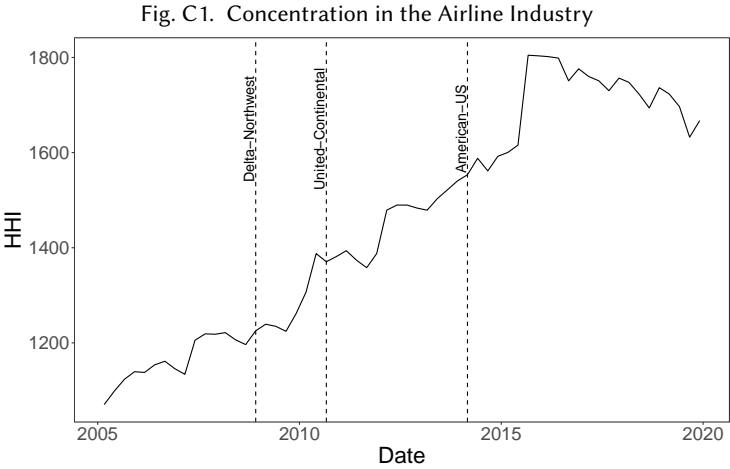
Notes: The figure plots the evolution of the national HHI of the airline industry during the period 2005-2019. We use passenger counts to construct market shares used in the calculation of HHI.

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:

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

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 et al., 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 C3. The results and median own-price elasticities are in line with the literature.

C.3   Supply Estimation

In Section 6.2, we briefly introduced the supply side and focused on the flexible supply 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 C4. We find that distance is positively associated with marginal costs inferred under the Bertrand assumption of conduct.

C.4 Baseline Empirical Results

Table C2. Difference-in-Differences Estimates

| | log(Fare) | | |
| | DL-NW | UA-CO | AA-US |
	(1)	(2)	(3)
Treated × Post	-0.0322	0.0103	0.0292***
	(0.0229)	(0.0165)	(0.0091)
Share Nonstop	-0.2753***	-0.2620***	-0.1160***
	(0.0416)	(0.0513)	(0.0351)
Average Distance	5.66×10^{-5}	7.8×10^{-5}	0.0002***
	(5.83×10^{-5})	(6.88×10^{-5})	(3.18×10^{-5})
R^2	0.5817	0.4897	0.4001
Observations	3,938	5,178	9,965
Treated (%)	19.880	22.210	52.230
Origin-destination fixed effects	✓	✓	✓
Year-quarter fixed effects	✓	✓	✓

Notes: The table reports estimates from the DID regression of log fares on treated × post and market characteristics including the share of nonstop flights and the average distance. We consider three mergers: Delta-Northwestern in Column (1), United-Continental in Column (2), and American-US in Column (3).

Table C3. 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.

Table C4. 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.5 Counterfactuals

Fig. C2. Price Change Distribution
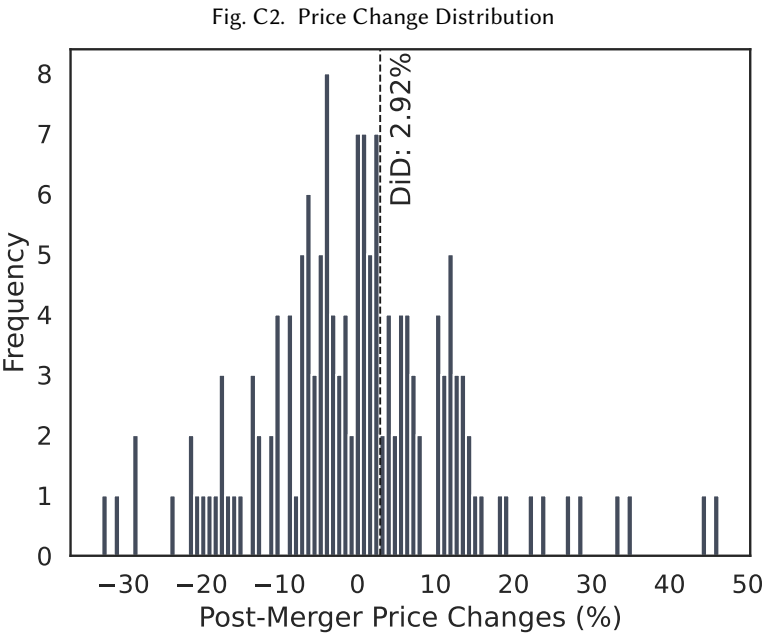
Notes: The figure plots the observed post-merger price changes after the US-AA merger.
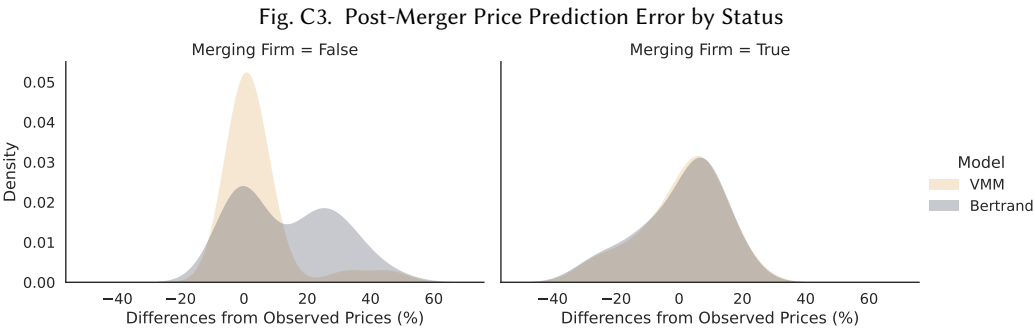
Fig. C3. Post-Merger Price Prediction Error by Status

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.

Fig. C4. Width of Confidence Intervals
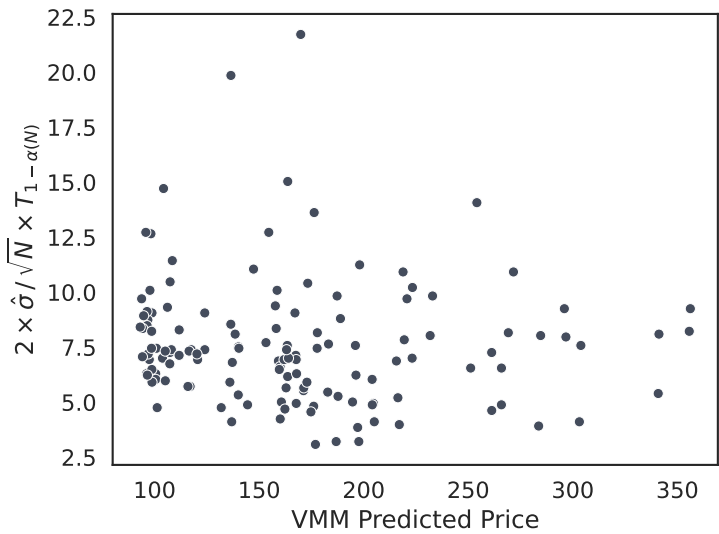
Notes: The figure represents the scatter plot of VMM predicted prices (on the horizontal axis) against the width of the confidence interval for the predictions. Variance is estimated for all firms in markets affected by the US Airways-American merger.