# Personalized Pricing and Consumer Welfare

Jean-Pierre Dubé

*University of Chicago and National Bureau of Economic Research*

Sanjog Misra

*University of Chicago*

We study the welfare implications of personalized pricing implemented with machine learning. We use data from a randomized controlled pricing field experiment to construct personalized prices and validate these in the field. We find that unexercised market power increases profit by 55%. Personalization improves expected profits by an additional 19% and by 86% relative to the nonoptimized price. While total consumer surplus declines under personalized pricing, over 60% of consumers benefit from personalization. Under some inequity-averse welfare functions, consumer welfare may even increase. Simulations reveal a nonmonotonic relationship between the granularity of data and consumer surplus under personalization.

## I. Introduction

The vast quantities of personal data available to firms today have enormous economic potential. These data represent valuable business assets when firms use them to target decisions—such as advertising and pricing—differentially

We are grateful to Ian Siegel and Jeff Zwelling of ZipRecruiter for their support of this project. We also thank the ZipRecruiter pricing team for their help and work in making the implementation of the field experiments possible. We are also extremely grateful for the extensive feedback and suggestions from Dirk Bergemann, Chris Hansen, Matt Taddy, Gautam Gowrisankaran, and Ben Shiller. Finally, we benefited from the comments and suggestions of seminar participants at the Bridge Webinar Series at McGill University, Cornell University, Columbia Business School, Institut Européen d'Administration des Affaires,

Electronically published December 9, 2022

*Journal of Political Economy*, volume 131, number 1, January 2023.

https://doi.org/10.1086/720793

across individuals. Recent events, such as the controversy over Cambridge Analytica's alleged misuse of user data on Facebook (Granville 2018), the adoption of the General Data Protection Regulation (GDPR) in the European Union, and the passage of the California Consumer Privacy Act (CCPA) of 2018, have created a surge in public interest and debate over acceptable commercial uses of consumer data. The data policies that have emerged or are currently under debate as a consequence of these events have restricted commercial uses of consumer data, ostensibly to protect consumers and their privacy. However, the overall welfare implications of such privacy and data policies are not completely transparent and could have the unintended consequence of harming consumer surplus.

In this paper, we study the welfare implications of one particular controversial form of data-based decision-making: personalized pricing. Personalized pricing represents an extreme form of third-degree price discrimination that implements consumer-specific prices,[1] using a large number of observable consumer features.[2] Prices are set differentially across each combination of observed consumer features to capture surplus. The application of modern machine-learning tools enables firms to apply such segmented pricing at scale.

The current extent of personalized pricing used in practice is unknown, and "examples remain fairly limited" (CEA 2015, 3).[3] In practice, third-degree price discrimination is still less common than second-degree price discrimination policies involving nonlinear pricing schedules or menus

the 2017 Microsoft Digital Economics Conference, Massachusetts Institute of Technology, Penn State University, the 2017 Porter Conference at Northwestern University, Stanford Graduate School of Business, the University of Chicago Booth School of Business, the University of Notre Dame, the 2019 Triangle Microeconomics Conference at the University of North Carolina at Chapel Hill, the University of Rochester, the University of Wisconsin, the Wharton School, Yale University, the 2017 Marketing and Economics Summit, the 2016 Digital Marketing Conference at Stanford Graduate School of Business, and the 2017 Summer National Bureau of Economic Research meetings in Economics and Digitization. Dubé and Misra acknowledge the support of the Kilts Center for Marketing. Misra also acknowledges the support of the Neubauer Family Foundation. Dubé also acknowledges the support of the Charles E. Merrill faculty research fund. Replication files are available in a zip file. This paper was edited by Chad Syverson.

[1] Even though our business-to-business (B2B) case study involves enterprise customers, we use the term "consumer" herein to refer to the buyers to conform with the terminology typically used in economics literature on demand-side welfare.

[2] In practice, third-degree price discrimination has typically been based on very coarse segmentation structures that vary prices across broad groups of consumers, such as senior citizens' and children's discounts at the movies, and geographic or "zone" retail pricing by chain stores across different neighborhoods within a metropolitan area. Only with the recent rise of the commercial internet and digitization has the potential for more granular, personalized segmentation structures become practical and scalable for marketing purposes (Shapiro and Varian 1999; Smith, Bailey, and Brynjolfsson 2000).

[3] Even large, digitally enabled firms such as Amazon have committed to an explicit, nondiscriminatory pricing policy (Wolverton 2002).

of differentiated substitute products (e.g., Mussa and Rosen 1978). Nevertheless, growing public policy concern over the prospect of differential pricing on scale prompted a 2015 report by the Counsel of Economic Advisors (CEA) devoted entirely to differential pricing with "big data" (CEA 2015). Recognizing how "big data and electronic commerce have reduced the costs of targeting and first-degree price discrimination" (CEA 2015, 12), the report mostly drew dire conclusions about the potential harm to consumers: "[differential pricing] transfers value from consumers to shareholders, which generally leads to an increase in inequality and can therefore be inefficient from a utilitarian standpoint" (CEA 2015, 6). Similar concerns about the harmful effects of differential pricing have been echoed in the recent mainstream business media (e.g., Useem 2017; Mohammed 2017), leading experts to question the fairness and even legality of these practices (e.g., Krugman 2000; Ramasastry 2005; Turow, Feldman, and Meltzer 2005). While the CEA report does not specifically recommend new legislation to regulate personalized pricing, privacy legislation, such as the GDPR, will require firms to disclose their usage of consumer data "in a concise, transparent, intelligible and easily accessible form, using clear and plain language."[4] The GDPR may also require consumers to give consent before receiving personalized prices, which could limit the granularity of price discrimination and the types of variables firms are allowed to use when they set their prices. A similar set of clauses is also found in the recent CCPA.

A potential concern is that overregulation of data-based price discrimination could in fact have the unintended consequence of reducing social welfare and, more specifically, harming consumers in some situations. While it is well understood that in a monopoly setting, price discrimination will typically benefit the firm, there is no general result as far as consumer welfare is concerned. The research in this area has derived, in a variety of settings, sufficient conditions on the shape of demand to determine whether third-degree price discrimination would increase social welfare (e.g., Pigou 1920; Varian 1989; Cowan and Vickers 2010) and consumer welfare specifically (e.g., Cowan 2012). In a more recent theoretical analysis, Bergemann, Brooks, and Morris (2015) show that the consumer welfare implications of third-degree price discrimination depend on the attainable set of consumer segmentation structures using a firm's database. Unlike perfect price discrimination, which transfers all the consumer surplus to the firm, personalized pricing often includes an element of classification error and theoretically could increase consumer surplus relative to optimal uniform pricing. Determining the extent to which "the combination of sophisticated analytics and massive amounts of data will lead to an increase in aggregate welfare" versus "mere changes in the allocation of

4 See article 12 of the GDPR, "Transparent Information, Communication and Modalities for the Exercise of the Rights of the Data Subject."

wealth" has been identified as a fruitful direction for future research in the economics of privacy (Acquisti, Taylor, and Wagman 2016, 481).

To analyze the welfare implications of personalized pricing, we conduct an empirical case study in cooperation with a large digital firm that was in the early stages of reexamining its pricing policy. The heart of our analysis consists of a sequence of novel, randomized B2B price experiments for new consumers. In the first experiment, we randomize the quoted monthly price of service to new consumers and use the data to train a demand model with heterogeneous price treatment effects. We assume that the heterogeneity in consumers' price sensitivities can be characterized by a sparse subset of an observed, high-dimensional vector of observable consumer features. The demand estimates allow us to design an optimized uniform pricing structure and an optimized personalized pricing structure. We use a Bayesian decision-theoretic formulation of the firm's pricing decision problem (Wald 1950; Savage 1954), defining the posterior expected profits as the reward function to account for statistical uncertainty. In a second experiment with a new sample of consumers, we then test our model pricing recommendations and inference procedure out of sample, a novel feature of our analysis (see also Misra and Nair 2011; Ostrovsky and Schwarz 2016).[5]

To the best of our knowledge, this study is the first to document both the feasibility and the implications of scalable personalized pricing. In this regard, we add to a small and growing literature using firm-sanctioned field experiments to obtain plausible estimates of the treatment effect of marketing variables on demand (e.g., Levitt and List 2009; Einav and Levin 2010).[6] The fact that our corporate partner, ZipRecruiter, has authorized us to disclose its identity and the details of the underlying experiment also supports the growing importance of transparency and disclosure when using firm-sponsored experiments for scientific research (Einav and Levin 2014).

While not the main focus of the paper, the field experiment reveals a striking degree of unexercised market power. The data-based, optimal uniform price is 230% higher than the firm's status quo pricing, an opportunity to increase profits by 55%. These large price and profit improvements are robust to the optimization of longer-term discounted profits that also account for future consumer retention rates. The second experiment confirms the profit increases from data-based pricing out of sample. In fact, shortly after the first experiment, ZipRecruiter permanently increased its price to $249, at least until as recently as November 2020.[7]

[5] Misra and Nair (2011) test the performance of a more efficient incentives-based compensation scheme for sales agents in a large firm, and Ostrovsky and Schwarz 2016 test the performance of optimally derived reserve prices for Yahoo's sponsored search auctions.

[6] See also Cohen et al. (2016) for a quasi price experiment based on Uber surge.

[7] More recently, the firm has implemented a menu of prices that includes $249 as the price of the base product.

Our demand estimates also reveal a considerable degree of heterogeneity in willingness to pay. We predict that decision-theoretic personalized pricing would increase the firm's posterior expected profits by 86% relative to its status quo price of $99 and by 19% relative to the decision-theoretic optimal uniform price of $327. These predicted profit improvements are robust to a longer-term time horizon of several months. We validate the predicted profit gains out of sample using our second experiment. Although the gains in profits are not surprising theoretically, the magnitudes are considerably higher than those predicted in past work using observable consumer variables (Rossi, McCulloch, and Allenby 1996; Shiller and Waldfogel 2011; Shiller 2015).

On the demand side, the evaluation of consumer welfare in a setting without a representative-consumer formulation requires the specification of a social welfare function. Under a total consumer surplus standard, we predict that consumer welfare would fall under decision-theoretic personalized pricing relative to optimal uniform pricing. In this regard, our findings confirm some of the concerns about consumer harm in the public policy debate. But for our case study, personalization is still far removed from the purely theoretical case of perfect price discrimination, which transfers all the consumer surplus to the firm. Simulations based on the estimates from the first experiment predict that the majority of consumers benefit from personalization relative to the optimal uniform price, indicating redistributive benefits, albeit at the expense of the highest-willingness-to-pay consumers. In our second validation field experiment, nearly 70% of the consumers assigned to the personalized pricing cell are targeted a personalized price that is below the optimal uniform price. Under alternative inequality-averse consumer welfare functions (Atkinson 1970; Jorgenson 1990; Lewbel and Pendakur 2017), we find that these redistributive benefits could outweigh the losses in total consumer surplus depending on the degree of the social planner's inequality aversion. Although our experiments are not designed to identify the causal effect of specific individual consumer features on demand, in an exploratory exercise, we find that the "firm size" and "benefits offered to employees" features are the most highly correlated with incidence of receiving a personalized price below the uniform rate. Therefore, personalization appears to benefit smaller and more disadvantaged firms, albeit at the cost of an overall decrease in total consumer surplus. Our results do not appear to be an artifact of the use of a standard LASSO (least absolute shrinkage and selection operator) regularization algorithm. Qualitatively, our findings are robust to a recently developed, alternative deep-learning approach developed by Farrell, Liang, and Misra (2021a, 2021b).

The main focus of our analysis is on the use of our model estimates to explore the role of the granularity of consumer information on surplus. We examine several alternative personalization schemes that restrict the types of consumer features on which the firm is allowed to condition to construct segments and set differential prices. Consistent with Bergemann, Brooks, and Morris (2015), we find a nonmonotonic relationship between consumer surplus and the quantity of consumer data available to the firm for personalization. While all of our personalization scenarios generate less consumer surplus than uniform optimal pricing, we find several cases where restricting the firm's information set leads to even less consumer surplus in spite of the coarsening of the segments. This nonmonotonicity is also robust to the use of the deep-learning algorithm. This empirical finding that consumer surplus is nonmonotonic in the degree of consumer information suggests that any regulation of consumer data might need to carefully consider the welfare implications caused by downstream decisions based on such data.

Our findings contribute to the empirical literature on third-degree price discrimination (see the survey by Verboven 2008). The price experiment avoids the typical price endogeneity concerns associated with demand estimation based on observational data and offers a clean study of the impact of third-degree price discrimination on the firm's outcomes. In the domain of digital marketing, Bauner (2015) and Einav et al. (2018) argue that the coexistence of auctions and posted price formats on eBay may be a sign of price discrimination across consumer segments. Einav et al. (2018) conclude that "richer econometric models of e-commerce that incorporate different forms of heterogeneity . . . and might help rationalize different types of price discrimination would be a worthwhile goal for future research." In a large-scale randomized price experiment for an online gaming company that uses almost uniform pricing, Levitt et al. (2016) find almost no effect on revenues from various alternative second-degree nonlinear price discrimination policies. However, they document substantial heterogeneity across consumers, which suggests potential gains from the type of third-degree personalized pricing studied herein. Subsequent to the writing of this paper, Kehoe, Larsen, and Pastorino (2020) also analyzed the potential consumer welfare–increasing effects of personalized pricing in a dynamic durable goods duopoly market.

Our work also contributes to the broader empirical literature on the targeting of marketing actions across consumers (e.g., Ansari and Mela 2003; Simester, Sun, and Tsitsiklis 2006; Dong, Manchanda, and Chintagunta 2009; Kumar et al. 2011). A small subset of this literature has analyzed personalized pricing with different prices charged to each consumer (e.g., Rossi, McCulloch, and Allenby 1996; Chintagunta, Dubé, and Goh 2005; Zhang, Netzer, and Ansari 2014; Shiller 2015; Waldfogel 2015). Our work is closest to Shiller (2015), who also uses machine learning to estimate heterogeneous demand. Most of this research uses a retrospective analysis of detailed consumer purchase histories to determine personalized prices. These studies report large predicted profit improvements for firms when they target consumers' purchase history behavior. However, the implications for targeted pricing are typically studied through model simulations based on demand estimates. In contrast, we run field experiments, not only to estimate demand but also to provide an out-of-sample field validation of the model predictions for the impact on consumers and the firm. The extant work's findings and methods also have limited applicability beyond markets for fast-moving consumer goods owing to the limited availability of consumer purchase panels in most markets. In contrast, we devise a more broadly practical targeting scheme based on observable consumer features and cross-sectional data.

The extant literature suggests that basing personalized prices on observable consumer features, as opposed to purchase histories, generates modest gains for firms, casting doubts on the likelihood that firms would invest in implementing such pricing practices. For example, Rossi, McCulloch, and Allenby (1996) conclude that "it appears that demographic information is only of limited value" for the personalization of prices of branded consumer goods. Similarly, Shiller and Waldfogel (2011) claim that "despite the large revenue enhancing effects of individually customized uniform prices, forms of third degree price discrimination that might more feasibly be implemented produce only negligible revenue improvements." In the internet domain, Shiller (2015) finds "demographics alone to tailor prices raises profits by 0.8% [at Netflix]." These findings may explain the lack of empirical examples of large-scale personalized pricing in practice. One exception is List (2004), who finds that sports card dealers actively use minority status as a proxy for differences in consumer willingness to pay, though he does not explore the profit implications. In contrast, our findings suggest that personalized pricing based on observable consumer features could improve firm profits substantially, supporting the view that such practices could become more commonplace.

Our findings also relate to the concept of fairness in the social choice literature and add to the ongoing public policy debate regarding the fairness aspects of differential pricing. In our discrete-choice demand setting, only a uniform pricing policy would satisfy the "no envy" criterion of the fair allocations studied in the social choice literature (Foley 1967; Thomson 2011). Absent wealth transfers, in our case study this fair outcome could lead to fewer served consumers and lower consumer surplus, highlighting the potential trade-offs between fairness and consumer welfare. Moreover, in our case study the typical strong consumer tends to be a larger company with 20 employees (relative to 10 employees for weak consumers), suggesting that our personalization scheme redistributes surplus from larger to smaller consumers. This type of reallocation could be rationalized as fair under a Pareto-weight scheme that assigns higher social value to smaller, disadvantaged firms. In this regard, our findings also contribute to the emerging literature on the economics of privacy (e.g., Acquisti, Taylor, and Wagman 2016) by documenting potential benefits to consumers from personalization.

The remainder of the paper is organized as follows. In section II, we set up the prototypical decision-theoretic formulation of monopoly price personalization based on demand estimation. In section III, we derive our empirical approach for estimating the demand parameters and quantifying uncertainty. We summarize our empirical case study of targeted pricing at ZipRecruiter in section IV. In section V, we explore the welfare implications of different targeting databases, and in section VI we explore the robustness of our findings to a more sophisticated deep learning algorithm. We conclude in section VII.

## II. A Model of Decision-Theoretic Monopoly Price Personalization

In this section, we outline the key elements of a data-based approach to monopoly price discrimination. We cast the firm's pricing decision as a Bayesian statistical decision theory problem (e.g., Wald 1950; Savage 1954; Berger 1985; for a short overview, see Hirano 2008; for a discussion of Bayesian decision theory for marketing problems, see Green and Frank 1966; Bradlow et al. 2004). The firm trades off the opportunity costs from suboptimal pricing and the statistical uncertainty associated with sales and profits at different prices. We cast the firm's uncertainty as a lack of precise statistical information about an individual consumer's preferences and demand. Bayes's theorem provides the most appropriate manner for the firm to use available data to update its beliefs about consumers and make informed pricing decisions. Failure to incorporate this uncertainty into pricing decisions could lead to bias, as we discuss below. We also discuss herein the potential shortcomings of a simpler approach that plugs in point estimates of the uncertain quantities instead of using the full posterior distribution of beliefs. For an early application of Bayesian decision theory to pricing strategy, see Green (1963). For a more formal econometric treatment of Bayesian decision-theoretic pricing that integrates consumer demand estimation, see Rossi, McCulloch, and Allenby (1996) and Dubé et al. (2017).[8]

We start by describing the demand setup and defining the sources of statistical uncertainty regarding consumers and their demand. The demand model represents the firm's prior beliefs about the consumer. On the supply side, we then define the firm's information set about the consumer. By combining the firm's prior beliefs (the demand model) and available information (the consumer data), we then define several decision-theoretic (or "data-based") optimal pricing problems for the firm.

[8] See Hitsch (2006) for an application of Bayesian decision-theoretic sequential experimentation.

### A. *Demand*

Below we present a relatively agnostic, multiproduct derivation of demand to illustrate the generalizability of our approach across a wide class of empirical demand settings. Consider a population of $i = 1, \dots, H$ consumers. Each consumer $i$ chooses a consumption bundle $q = (q_1, \dots, q_J) \in \mathbb{R}^J_+$ to maximize her utility as follows:

$$\bar{q}(p_i; \Psi_i, \epsilon_i) = \arg\max_q \{U(q; \Psi_i, \epsilon_i) \quad : \quad p_i' q \leqslant I\}, \tag{1}$$

where $U(q; \Psi_i, \epsilon_i)$ is continuously differentiable, strictly quasi-concave, and increasing in $q$; $I$ represents a budget; $p_i = (p_{i1}, \dots, p_{iJ}) \in \mathbb{R}^J_+$ is the vector of prices charged to consumer $i$; $\Psi_i$ represents consumer $I$'s potentially observable "type" (or preferences); and $\epsilon_i \sim$ i.i.d. $F_\epsilon(\epsilon)$ is an independent and identically distributed (i.i.d.) random vector of unobserved, random disturbances that are independent of $\Psi_i$. In our analysis below, we distinguish between the aspects of demand about which a firm can learn, $\Psi_i$, and about which it cannot learn, $\epsilon_i$.

### B. *Firm Beliefs and Pricing*

We now define the personalized pricing problem and its relationship to the price discrimination literature. To capture the marketplace realities of data-based marketing, we model the firm's design of personalized pricing as a statistical decision problem.

Suppose the firm knows the form of demand, (1), and has prior beliefs about $\Psi_i$ described by the density $f_\Psi(\Psi_i)$. Let $\mathbf{D}$ denote the consumer database collected by the firm. We assume that the firm uses Bayes's rule to construct the data-based posterior belief about the consumer's type:

$$f_\Psi(\Psi_i|\mathbf{D}) = \frac{\ell(\mathbf{D}|\Psi_i) f_\Psi(\Psi_i)}{\int \ell(\mathbf{D}|\Psi_i) f_\Psi(\Psi_i)\, d\Psi_i}, \tag{2}$$

where $\ell(\mathbf{D}|\Psi_i)$ represents the log likelihood induced by the demand model, (1), and the uncertainty in the random disturbances, $\epsilon_i$. Let $F_\Psi(\Psi_i|\mathbf{D})$ denote the corresponding cumulative distribution function (CDF) of the posterior beliefs. Note that we assume that the firm does not update its beliefs $F_\epsilon(\epsilon)$ about the random disturbances, $\epsilon_i$.

Given the posterior $F_\Psi(\Psi_i|\mathbf{D})$, the firm makes decision-theoretic, data-based pricing decisions. We assume that the firm is risk neutral and faces unit costs $c = (c_1, \dots, c_J)$ for each of its products. For each consumer $i$, the firm anticipates the following posterior expected profits from charging prices $p_i$:

$$\pi(p_i|\mathbf{D}) = (p_i - c)' \int\int \bar{q}(p; \Psi_i, \epsilon)\, dF_\epsilon(\epsilon)\, dF_\Psi(\Psi_i|\mathbf{D}). \tag{3}$$

The firm's optimal *personalized prices* for consumer $i$, $p_i^*$, must therefore satisfy the following first-order necessary conditions:

$$p_i^* = c - \left[\int\int \nabla_p \bar{q}\left(p_i^*; \Psi_i, \epsilon\right) dF_\epsilon(\epsilon)\, dF_\Psi(\Psi_i|\mathbf{D})\right]^{-1} \int\int \bar{q}\left(p_i^*; \Psi_i, \epsilon\right) dF_\epsilon(\epsilon)\, dF_\Psi(\Psi_i|\mathbf{D}), \tag{4}$$

where $\nabla_p \bar{q}(p_i^*; \Psi_i, \epsilon)$ represents the matrix of derivatives of consumer $i$'s demand with respect to prices. If the firm instead implements a *uniform* pricing strategy across all its $H$ consumers, the posterior expected profit-maximizing uniform prices, $p^*$, must satisfy the following first-order necessary conditions:

$$p^* = c - \left[\sum_i^H \int\int \nabla_p \bar{q}(p^*; \Psi_i, \epsilon)\, dF_\epsilon(\epsilon)\, dF_\Psi(\Psi_i|\mathbf{D})\right]^{-1} \sum_i^H \int\int \bar{q}(p^*; \Psi_i, \epsilon)\, dF_\epsilon(\epsilon)\, dF_\Psi(\Psi_i|\mathbf{D}). \tag{5}$$

The recent public policy debate regarding consumer data and targeted pricing has frequently associated personalized pricing with traditional first-degree price discrimination. While *first-degree* or *perfect price discrimination* has typically been viewed as a polar, theoretical case (e.g., Pigou 1920; Varian 1980; Stole 2007; Bergemann, Brooks, and Morris 2015), theorists have long recognized the possibility that with a very granular segmentation scheme, third-degree price discrimination could approximate first-degree price discrimination:[9] "it is evident that discrimination of the third degree approximates towards discrimination of the first degree as the number of markets into which demands can be divided approximate toward the number of units for which any demand exists" (Pigou 1920, 287). In fact, the personalized pricing in (4) technically constitutes a form of *third-degree price discrimination* (e.g., Pigou 1920; Tirole 1988). In our model, the firm can never learn $\epsilon_i$ even with repeated observations on the same consumer (i.e., panel data). Therefore, it will never be possible for the firm to extract all of the consumer surplus even when all the uncertainty in $\Psi_i$ is resolved. In practice, the prices are not fully

[9] Statistical uncertainty typically limits the segmentation to an imperfect form of targetability. The approximation is also typically closer under unit demand since personalization typically cannot target a different price to each inframarginal unit purchased by a consumer.

personalized since consumers with the same posterior expected $\Psi_i$ would always be charged the same price even if they differ along unobserved dimensions.

## C. *Welfare*

### 1. Welfare Implications of Personalization

At the heart of the public policy debate is a widespread belief that machine learning and data-based marketing will harm consumers per se. Monopoly personalized pricing will always weakly increase the firm's profits since, by revealed preference, the firm can always choose to charge every consumer the same uniform price in (5): $p_i^* = p^*, \forall\ i$.[10] The predicted impact of personalized prices on consumer surplus is less straightforward. Under perfect price discrimination, the monopolist extracts all the consumer surplus. As consumer data converge to the point where a firm can perfectly predict a consumer's willingness to pay for each marginal unit, we would expect additional information to reduce consumer surplus per se. But perfect price discrimination is at best a theoretical polar case. Even in fast-moving consumer goods industries where the firm can track the same consumer's shopping choices repeatedly over time, potentially at different prices, researchers still observe a substantial amount of random (unpredictable) switches in consumer choices (e.g., Rossi, McCulloch, and Allenby 1996). Therefore, for the foreseeable future, personalized pricing will at best achieve an extremely granular form of third-degree (as opposed to first-degree) price discrimination.

The extant literature on monopoly third-degree price discrimination has relied on local conditions regarding the curvature of demand and other regularity conditions to determine the impact on social surplus (e.g., Varian 1989) and consumer surplus specifically (e.g., Cowan 2012). More recently, Bergemann, Brooks, and Morris (2015) show that, theoretically, third-degree price discrimination "can achieve every combination of consumer surplus and producer surplus such that: (i) consumer surplus is nonnegative, (ii) producer surplus is at least as high as profits under the uniform monopoly price, and (iii) total surplus does not exceed the surplus generated by efficient trade" (921). Therefore, the impact of the personalized prices characterized by (4) on consumer surplus is ultimately an empirical question about the segments constructed with the database **D**.

To illustrate this point, consider a market with six consumers $\{i\}_{i=1}^{6}$ with valuations $\Psi_i = \$i$. Assume that costs are negligible (close to zero) and are relevant only as tiebreakers between profit-equivalent choices. In table 1, we report the results under several information scenarios. Under

[10] We make the usual assumption of no arbitrage between consumers.

TABLE 1
Data and Welfare

| | Uniform | Perfect Price Discrimination | Personalized 1: $\mathbf{D}$ = ({1} {2, 3, 4, 5, 6}) | Personalized 2: $\mathbf{D}$ = ({1}, {2, 3}, {4, 5, 6}) |
|---|---|---|---|---|
| Prices | $p_i^U = 4, \forall\ i$ | $p_i^{\mathrm{PD}} = i, \forall\ i$ | $p_{\{1\}}^{\mathrm{PP1}} = 1$<br>$p_{\{2,3,4,5,6\}}^{\mathrm{PP1}} = 4$ | $p_{\{1\}}^{\mathrm{PP2}} = 1$<br>$p_{\{2,3\}}^{\mathrm{PP2}} = 2$<br>$p_{\{4,5,6\}}^{\mathrm{PP2}} = 4$ |
| Profits | 12 | 21 | 13 | 17 |
| Customer surplus | 3 | 0 | 3 | 4 |

Note.—All values are in dollars.

perfect price discrimination, the firm charges each consumer her valuation, generating \$21 in profits and \$0 in consumer surplus. Under a profit-maximizing uniform pricing policy, the firm charges $p_i = \$4\ \forall\ i$, which generates \$12 in profits and \$3 in consumer surplus.[11] Total surplus, however, is only \$15, and there is a deadweight loss of \$6.

Now suppose the firm has a database, $\mathbf{D}$, that signals information about consumers' types, allowing it to distinguish between the following two segments: {1} and {2, 3, 4, 5, 6}. Under third-degree price discrimination, the firm can increase its profits to \$13 by charging the segment prices $p_{\{1\}} = \$1$ and $p_{\{2,3,4,5,6\}} = \$4$. In this case, consumer surplus remains fixed at \$3. Total surplus, however, has increased by \$16, and the deadweight loss is now only \$5.

Now consider the more granular database, $\tilde{\mathbf{D}}$, that allows the firm to classify the consumers into the following three segments: {1}, {2, 3}, and {4, 5, 6}. For instance, suppose that a change in public policy that previously protected the identity of consumers 2 and 3 (e.g., race or gender) is relaxed, allowing the firm to target this segment with differential prices. Under third-degree price discrimination, the firm can now increase its profits to \$17 by charging the segment prices $p_{\{1\}} = \$1$, $p_{\{2,3\}} = \$2$, and $p_{\{4,5,6\}} = \$4$. As we increase the granularity of the database and allow for more personalized pricing, consumer surplus *increases* to \$4. Moreover, total surplus is now \$21, which is equal to total surplus under perfect price discrimination except that some of the value accrues to consumers. Interestingly, there is no deadweight loss in this case.

These findings are robust to the inclusion of classification error (i.e., an untargetable type I extreme value random utility shock): $\Psi_i = i + \epsilon_i$. In this case, both segmentation schemes increase both firm profit and consumer surplus relative to the uniform pricing scenario. However, the granular database increases consumer surplus by less than the coarse database,

[11] The firm does not charge a uniform price equal to \$3 because of our assumption of a small but positive marginal cost to break the tie between \$3 and \$4.

indicating a nonmonotonicity in the relationship between consumer surplus and the degree of granularity of the segmentation scheme.

This example merely illustrates that increasing the granularity of the consumer data available to a firm can increase consumer surplus and even reduce deadweight loss. Obviously, there are other databases that could lead to segmentation schemes that would have different welfare implications. But the example indicates that the consumer welfare implication of personalized pricing is ultimately an empirical question that depends on the databases available to firms for marketing decision-making. In the next section, we discuss how a firm can use large consumer databases and machine learning to construct scalable segmentation schemes.

### 2. Welfare Aggregation

The discussion above assumes that society values only the total consumer surplus, with no weight assigned to the allocation. This perspective is reflected in the commonly used linear aggregation of total consumer surplus as a welfare measure

$$S(\mathbf{p}) = \frac{1}{N}\sum(V(\mathbf{p}, x_i)), \tag{6}$$

where $\mathbf{p} = \{\tilde{p}_i\}_{i=1}^{N}$ is the vector of prices charged to consumers and $V_i(\mathbf{p})$ denotes consumer $i$'s realized surplus in dollars.[12] This measure of surplus fails to account for any distributional effects besides the average. In his classic industrial organization textbook, Tirole (1988) describes the limitations of this approach as follows: "the government has efficiency concerns but no redistribution concerns. Of course, one of the main policy issues in regard to price discrimination is its effect on income distribution" (139).

To account for distributional effects, we examine alternative aggregation metrics and—following Lewbel and Pendakur (2017)—consider a range of welfare functions derived from Atkinson's (1970) mean of order $r$ class:

$$S_r(\mathbf{p}) = \begin{cases} \left[\frac{1}{N}\sum(V_i(\mathbf{p}))^r\right]^{1/r} & \text{for } r \neq 0, \\ \exp\left(\frac{1}{N}\sum \ln V_i(\mathbf{p})\right) & \text{for } r = 0. \end{cases} \tag{7}$$

In equation (7), $r$ determines society's preferences over allocations of surplus. The special case $r = 1$ (arithmetic mean) nests the commonly used

[12] In our empirical case study below, we follow the convention in the empirical literature and approximate $V_i(\mathbf{p})$ using the Hicksian compensating variation.

linear aggregation scheme in (6) above and reflects an inequality-neutral societal preference. As in Lewbel and Pendakur (2017), we focus on $r = -1, 0, 1$. The cases where $r \in \{-1, 0\}$ (harmonic and geometric mean) correspond to inequality-averse welfare functions that may select a personalized pricing policy that reduces total consumer surplus but at the same time disproportionately reduces inequality. This form of the welfare function is closely related to those proposed by Jorgenson (1990) and Jorgenson and Slesnick (2014), who also consider various generalized mean definitions to aggregate consumer surplus and evaluate the allocation.

## III. Empirical Approach

The execution of the firm's data-based pricing strategies in equations (4) and (5) depends on the ability to construct an estimate of the posterior distribution $F(\Psi_i|\mathbf{D})$. The extant literature on price discrimination has developed nonlinear panel data methods to estimate $F(\Psi_i|\mathbf{D})$ using repeated purchase observations for each consumer panelist (e.g., Rossi, McCulloch, and Allenby 1996; Chintagunta, Dubé, and Goh 2005). In practice, many firms do not have access to panel databases. In many B2B and e-commerce settings, for instance, firms are more likely to have access to data for a broad cross section of consumers, but not with repeated observations.[13] We consider a scenario with cross-sectional consumer information that includes a detailed set of observable consumer features. Our approach consists of using these features to approximate $\Psi_i$.

### A. *Approximating Individual Types*

Suppose we observe data

$$\mathbf{D} = \{(q_i, x_i, p_i)\}_{i=1}^{N}$$

for a sample of $N$ consumers, where $q_i \in \mathbb{R}_+^J$ is a vector of purchase quantities, $p_i \in \mathbb{R}_+^J$ represents the prices, and $x_i \in \mathcal{X} \subseteq \mathbb{R}^K$ is a vector of consumer characteristics. We assume that $x_i$ is high dimensional and fully characterizes the preferences, $\Psi_i$. We consider the projection of the individual tastes, $\Psi_i$, onto $x_i$:

$$\Psi_i = \Psi(x_i; \Theta_0),$$

where $\Theta_0$ is a vector of parameters. Note that for our pricing problem in section II.B, we are not interested in the interpretation of the arguments of the function $\Psi(x_i; \Theta)$, so we could be agnostic with our specification. For instance, we could represent the function $\Psi(x_i; \Theta)$ as a series expansion:

[13] Ideal panel data would allow the firm estimate types using fixed effects estimators but there would remain the issue of pricing to new consumers, which is our focus here.

$$\Psi(x_i;\Theta_0) = \sum_{s=1}^{\infty}\theta_{0s}\psi_s(x_i),$$

where $\{\psi_n(x_i)\}_{n\geq 0}$ is a set of orthonormal basis functions and $\Theta_{n0} = (\theta_1, \ldots, \theta_n)$ denotes the parameters for an expansion of degree $n$. We are implicitly assuming that some sparse subset of the vector $x_i$ is informative about $\Psi_i$ and that we possess some methods to identify this sparse subset.

We focus on applications where $K$ is large (potentially, $K \gg N$) and $\Theta_{n0}$ is relatively sparse. Even though our approach consists of a form of third-degree price discrimination, in practice, it can capture very rich patterns of heterogeneity. We assume that the firm has a very high-dimensional direct signal about demand, $x$. For instance, if the dimension of $x_i$ is $K = 30$, our approach would allow for as many as $2^K = 1{,}073{,}741{,}824$ distinct consumer types and, potentially, personalized prices.

### B. *Approximating $F(\Psi_i \,/\, \mathbf{D})$: The Weighted Likelihood Bootstrapped LASSO*

With $K \gg N$, maximum likelihood is infeasible unless one has a theory to guide the choice of coefficients to include or exclude. Even in cases where $K$ is large and $K < N$, maximum likelihood could be problematic and lead to overfitting. The literature on regularized regression provides numerous algorithms for parameter selection with a high-dimensional parameter vector, $\Theta$ (e.g., Hastie, Tibshirani, and Friedman 2009). Most of this literature is geared toward prediction. Our application requires us to quantify the uncertainty around our estimated coefficient vector, $\hat{\Theta}$, and around various economic outcomes, such as price elasticities, firm profits, and consumer value, to implement decision-theoretic optimized pricing structures. In addition, the approach must be fast enough for real-time demand forecasting and price recommendations.

Our framework conducts rational Bayesian updating with the goal of obtaining the posterior distribution of interest using a *loss* function, as opposed to a likelihood function. Bissiri, Holmes, and Walker (2016) show that for a prior, $h(\Theta)$, data, $\mathbf{D}$, and some loss function $l(\Theta, \mathbf{D})$, the object $f(\Theta|\mathbf{D})$ defined by

$$f(\Theta|\mathbf{D}) \propto \exp(-l(\Theta, \mathbf{D}))h(\Theta) \tag{8}$$

represents a coherent update of beliefs under loss function $l(\Theta, \mathbf{D})$. As such, it represents posterior beliefs about the parameter vector $\Theta$ given the data as encoded by the loss function $l(\Theta, \mathbf{D})$. In our setting, we specify the loss function as an $L_1$ penalized (LASSO) negative log likelihood:

$$l(\Theta, \mathbf{D}) = -\left[\sum_{i=1}^{N} \ell(\mathbf{D}_i|\Theta) - \lambda \sum_{j=1}^{J} |\Theta_j|\right], \tag{9}$$

where $\Sigma_{i=1}^{N} \ell(\boldsymbol{D}_i|\Theta)$ represents the sample log likelihood induced by the demand model in section I and $\lambda$ is a penalization parameter.

We then approximate the posterior $F_\Psi(\Psi|\mathbf{D})$ using a variant of the Bayesian bootstrap (e.g., Rubin 1981; Newton and Raftery 1994; Chamberlain and Imbens 2003; Efron 2012). In particular, we simulate draws from the posterior distribution of the model parameters using a weighted likelihood bootstrap (WLB) algorithm as outlined in Newton and Raftery (1994).[14] The approach that we follow is similar to the "loss likelihood bootstrap" outlined in Lyddon and Holmes (2019), who also derive the large-sample properties for these estimators. Broadly speaking, our procedure operates by assigning weights, drawn from a Dirichlet distribution, to each observation and implementing the LASSO estimator that conditions on these weights. Repeating this $B$ times gives us an approximate sample from the full posterior distribution $F_\Psi(\Psi|\mathbf{D})$, which can be used to compute the posterior distribution and other derived quantities required for the decision-theoretic pricing problem.

Formally, our estimator consists of $B$ replications of the following weighted-likelihood LASSO regression, where at step $b$,

$$\hat{\Theta}^b = \arg\max_{\Theta \in \mathbb{R}^J} \left\{ \sum_{i=1}^{N} V_i^b \ell(D_i|\Theta) - N\lambda \sum_{j=1}^{J} |\Theta_j| \right\}.$$

We show in appendix B that weights $V_i \sim$ i.i.d. $\exp(1)$ are equivalent to Dirichlet weights. Our procedure does not provide draws from the exact posterior, and consequently $\{\hat{\Theta}^b\}_{b=1}^{B}$ should be treated as an approximate sample from the posterior of interest. One interpretation of our approach is that it represents the draws from the posterior that minimizes the Kullback-Leibler divergence between the parametric class we adopt and the true data-generating process. This framework is coherent from a Bayesian perspective in spite of the nonstandard implementation. We refer the reader to Bissiri, Holmes, and Walker (2016) and Lyddon and Holmes (2019) for a more thorough discussion.

Our proposed algorithm deals with two sources of uncertainty simultaneously. In particular, by repeatedly constructing weighted LASSO type estimators, we are in effect integrating over the model space spanned by the set of covariates. As such, our draws can also be used to construct posterior probabilities associated with the set of covariates retained in the

[14] For a detailed description of our procedure, see app. B.

model. At the same time, the sampling procedure also accounts for usual parameter uncertainty. An additional advantage of using the loss likelihood approach is that we do not have to make parametric assumptions about our priors over the model space, allowing for additional robustness of our results. Subsequent to our analysis, new research has emerged with formal results on the sampling properties of similar machine-learning estimators applied to settings with high-dimensional observed heterogeneity (Athey and Imbens 2016a, 2016b). In our analysis below, we compare our findings with the WLB to a more sophisticated, nonparametric *deep-learning* algorithm (Farrell, Liang, and Misra 2021a, 2021b). See appendix D for details of the deep-learning algorithm. Owing to the binary nature of most of our consumer feature variables, this deep-learning algorithm produces results qualitatively similar to WLB.

The extant literature has often followed a two-step approach based on the oracle property of the LASSO (e.g., Fan and Li 2001; Zou 2006). When the implementation of the LASSO is an oracle procedure, it will select the correct sparsity structure for the model and will possess the optimal estimation rate. Accordingly, in a first step we could use a LASSO to select the relevant model (i.e., the subset of relevant $x$), and in a second step we could obtain parameter estimates after conditioning on this subset. We term this procedure post-LASSO-MLE (maximum likelihood estimation) and use it as a benchmark in later sections. In practice, the post-LASSO-MLE is a straw man since several authors have already found poor small-sample properties for such postregularization estimators (e.g., Leeb and Potscher 2008) that effectively ignore the model uncertainty by placing a degenerate prior with infinite mass on the model selected by the first-stage LASSO.

## IV. Personalized Pricing at ZipRecruiter

We analyze personalized pricing empirically through a sequence of experiments in collaboration with ZipRecruiter. The first experiment uses a sample of prospective, new ZipRecruiter consumers to train a demand model with heterogeneous price responses. The second experiment uses a new sample of prospective consumers to validate the predictions of the model and performance of the personalized pricing structure out of sample. Of interest is whether a firm like ZipRecruiter could in fact generate sufficient incremental profits to want to pursue a data-based price discrimination strategy. Moreover, we want to analyze the implications for consumer welfare.

ZipRecruiter is an online firm that specializes in matching job seekers to potential employers. We focus on ZipRecruiter's B2B decision since they offer their job-seeker services for free and charge only prospective employers. Hereafter, we refer to prospective employers who could use ZipRecruiter's service as consumers. The firm caters to a variety of potential consumers across various industries who can use ZipRecruiter to access a stream of résumés of matched and qualified candidates for recruiting purposes. Customers pay a monthly subscription rate that they can cancel at any time. In a typical month in 2015, ZipRecruiter hosted job postings for over 40,000 registered paying consumers. During the late spring of 2015, ZipRecruiter was in the process of reevaluating its pricing policy, making them open to our proposal to run randomized field experiments to measure demand and market power.

Our analysis focuses on prospective consumers who have reached the paywall at ZipRecruiter for the first time. Among all prospective consumers, ZipRecruiter's largest segment consists of the "starters," small firms with typically fewer than 50 employees, looking to fill between one and three jobs. Since starters represent nearly 50% of the consumer base, we focus our attention on prospective starter firms. Another advantage of focusing on small consumers is that they are unlikely to create externalities on the two-sided platform that would warrant lower pricing. For instance, ZipRecruiter might want to target low prices to certain very large recruiters in spite of high willingness to pay to create indirect network effects that stimulate demand from the set of applicants submitting their résumés. At the beginning of this project, the base rate for a starter firm looking for candidates was $99 per month.

Each prospective new firm that registers for ZipRecruiter's services navigates a series of pages on the ZipRecruiter website until they reach the paywall. At the paywall, they must use a credit card to pay the subscription fee. Immediately before the request for credit card information, a consumer is required to input details regarding the type of jobs they wish to fill as well as characteristics describing the firm itself. During this registration process, the consumer reports several characteristics of its business and the specific job posting. Table 2 summarizes the variables we retained for our analysis from the much larger set of registration features.[15] While the set looks small, it generates 133 variables.[16] After completing this registration process, the consumer reaches a paywall and receives a price quote. The registration process is used to ensure that ZipRecruiter's matching algorithm connects consumers with the most relevant résumés of potential applicants. In this case, we believe that the self-reported information is incentive compatible and that we do not need to worry whether consumers strategically misreport.

[15] In our personalized pricing application below, we analyze only segmentation schemes based on these features that are voluntarily and knowingly self-reported by consumers. We do not use any involuntary information tracked, e.g., through cookies.

[16] An initial set of marginal regressions were used to select these variables from the broader set of thousands of features for the demand analysis (e.g., Fan, Feng, and Song 2012). For our analysis here, we take these selected variables as given.

TABLE 2
COMPANY/JOB VARIABLES

| Feature Name |
|---|
| Job state |
| Company type |
| Commissions offered |
| Number of job slots needed |
| Total benefits |
| Employment type |
| Resume required |
| Medical benefit |
| Dental benefit |
| Vision benefit |
| Life insurance benefit |
| Job category |

### A. *Empirical Model of Demand*

Assume that a prospective, new consumer $i$ with observable features $x_i$ obtains the following incremental utility from purchasing versus not purchasing:

$$\begin{aligned}\Delta U_i &= \alpha_i + \beta_i p_i + \epsilon_i \\ &= \alpha(x_i; \theta_\alpha) + \beta(x_i; \theta_\beta) p_i + \epsilon_i,\end{aligned} \tag{10}$$

where $\alpha(x_i; \theta_\alpha)$ is an intercept and $\beta(x_i; \theta_\beta)$ is a slope associated with the price, $p_i$. To conform with our notation in section II, we rewrite equation (10) as follows:

$$\Delta U_i = \tilde{p}_i' \Psi_i + \epsilon_i, \tag{11}$$

where $\Psi_i = (\alpha(x_i; \theta_\alpha), \beta(x_i; \theta_\beta))'$ and $\tilde{p}_i = (1 p_i)'$.

The probability that consumer $i$ buys a month of service at price $p_i$ is

$$\begin{aligned}\mathbb{P}(y_i = 1 | p_i; \Psi_i) &= \int 1(\Delta U_i > 0)\, dF_\epsilon(\epsilon_i) \\ &= 1 - F_\epsilon(-\tilde{p}_i' \Psi_i),\end{aligned}$$

where $y_i = 1$ if she purchases or 0 otherwise.

For our analysis below, we use a linear specification of the functions $\alpha$ and $\beta$:

$$\begin{aligned}\alpha(x_i; \theta_\alpha) &= x_i' \theta_\alpha, \\ \beta(x_i; \theta_\beta) &= x_i' \theta_\beta.\end{aligned}$$

We also assume that the random utility disturbance $\epsilon_i$ is distributed i.i.d. logistic with scale parameter 1 and location parameter 0. These assumptions give rise to the standard binary logit choice probability

$$\mathbb{P}(y_i = 1|p_i;\Psi_i) = \frac{\exp(\tilde{p}_i'\Psi_i)}{1 + \exp(\tilde{p}_i'\Psi_i)}. \quad (12)$$

Note that our demand specification assigns a continuous treatment effect to prices since one of our objectives will consist of optimizing prices on the supply side. This smooth and continuous price treatment effect is an important distinction from most applications of machine learning, which involve categorical treatment variables.

### *B. Experiment 1: Demand, Pricing, and Consumer Welfare*

The first experiment was conducted between August 28, 2015, and September 29, 2015. During this period, 7,867 unique prospective consumers reached ZipRecruiter's paywall. Each prospective consumer was randomly assigned to one of 10 experimental pricing cells. The control cell consisted of ZipRecruiter's standard $99 per month price (row 1 of table 3). To construct our test cells, we changed the monthly rate by some percentage amount relative to the control cell. Following ZipRecruiter's practices, we then rounded up each rate to the nearest $9. The nine test cells are summarized in rows 2–10 of table 3.

#### 1. Model-Free Analysis

We report the results from the first experiment in figure 1. As expected, we observe a statistically significant, monotonically downward-sloping pattern of demand. Demand is considerably less price elastic than ZipRecruiter's current pricing would imply. A 100% increase in the price from $99 to $199 generates only a 25% decline in conversions. Given that most of ZipRecruiter's services are automated and it currently has enough capacity to increase its current consumer base by an arbitrary

TABLE 3
Experimental Price Cells for Stage 1

| | Monthly Price |
|---|---|
| Control | 99 |
| Test 1 | 19 |
| Test 2 | 39 |
| Test 3 | 59 |
| Test 4 | 79 |
| Test 5 | 159 |
| Test 6 | 199 |
| Test 7 | 249 |
| Test 8 | 299 |
| Test 9 | 399 |

Note.—All values are in dollars.

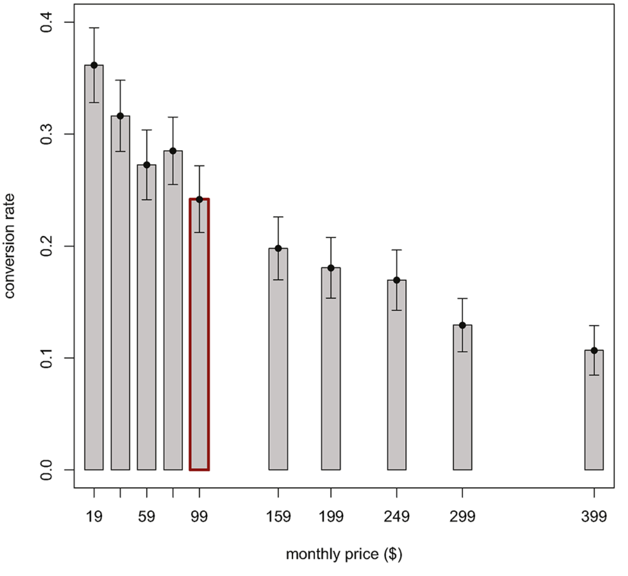


FIG. 1.—Stage 1 experimental conversion rates. Each bar corresponds to one of our 10 experimental price cells. The height of the bar corresponds to the average conversion rate within the cell. Error bars indicate the 95% confidence interval for the conversion rate.

amount, the marginal cost per consumer is close to $0. Therefore, ZipRecruiter is likely underpricing its service, at least under myopic pricing that optimizes current monthly profits.

Figure 2 plots ZipRecruiter's expected monthly revenue per consumer at each of the tested prices. The plot reveals a considerable degree of unexercised market power, suggesting that ZipRecruiter is significantly underpricing. Along our grid of tested price levels, the average monthly revenue per prospective consumer is maximized at $399. However, once we take into account statistical uncertainty, we cannot rule out that the revenue-maximizing price lies somewhere between $249 and $399 or even above $399.

The static profit analysis does not account for the fact that raising the monthly price today not only lowers current conversion but may also lower longer-term retention in ways that impact long-term profitability. Figure 3 reports the expected net present value of revenues per consumer over the 4-month horizon from September to December 2015. The top panel assumes a discount factor of $\delta = 0$ and therefore repeats the static expected revenues discussed above. The bottom panel assumes a discount factor of $\delta = 0.996$, implying a monthly interest rate of 0.4% (or an annual interest rate of 5%). While the net present value of profits is much higher at each of the tested prices, our ranking of prices is quite similar. To

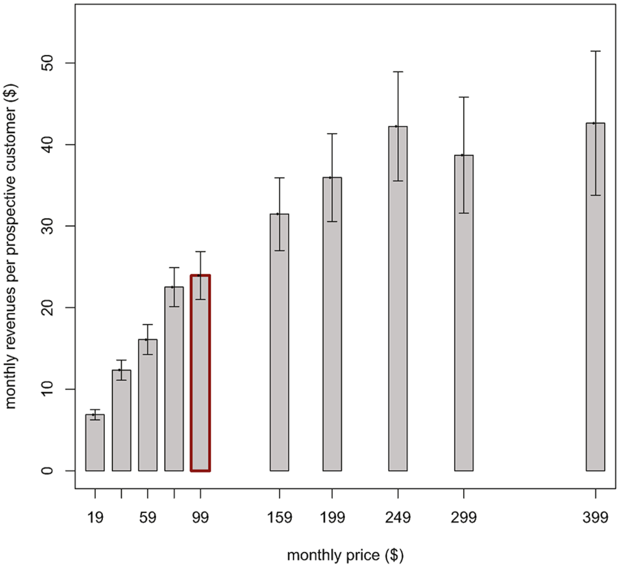


FIG. 2.—Stage 1 experimental revenues per customer. Each bar corresponds to one of our 10 experimental price cells. The height of the bar corresponds to the average revenue per prospective consumer within the cell. Error bars indicate the 95% confidence interval for the revenues per consumer.

understand this finding, table 4 reports both the acquisition rate (from September) and the retention rate (for October–December) for each of the tested price levels. As expected, conversion and retention both fall in the higher-price cells. However, survival rates are still low enough that the profit implications in the first month overwhelm the expected future profits from surviving consumers. In sum, our relative ranking of prices does not change much if we consider a longer-term planning horizon.[17] In fact, 1 month after the experiment, ZipRecruiter increased its price to $249 per month and has retained this base price until at least as recently as May 2021.

Although not the main focus of our studies, even in the absence of consumer information, purchase and price data alone reveal unexercised market power in this case study. ZipRecruiter should raise its prices by more than 100%, which would generate substantial incremental revenues per consumer. A price increase mechanically reduces consumer

[17] Our discussion here assumes that all customers who churn out of ZipRecruiter's business will never return. In practice, consumers may have heterogeneous reasons for churning out, including ranging from the satiation of their current recruiting needs to dissatisfaction with the service.

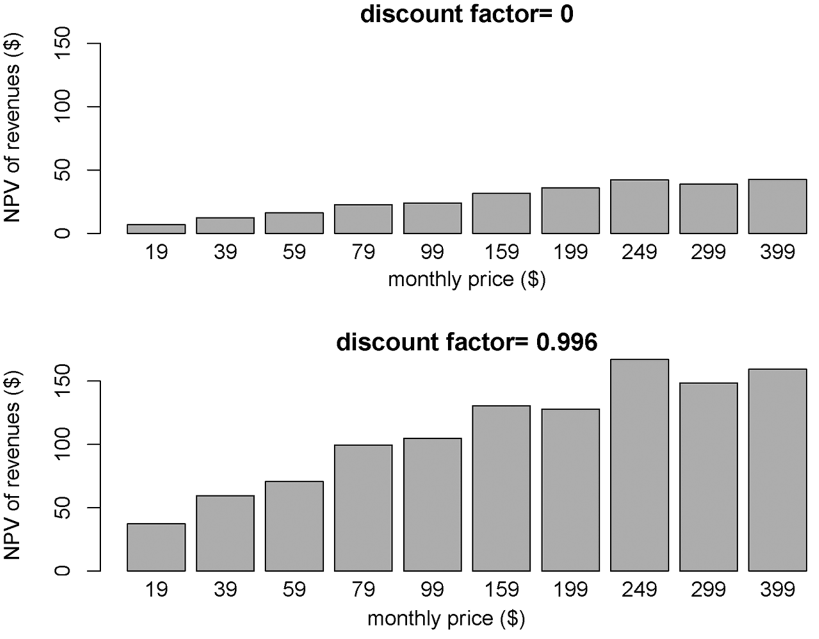


FIG. 3.—Expected net present value (NPV) of monthly revenues per lead over a 4-month horizon (September 2015).

surplus; however, ZipRecruiter would have eventually learned its demand and raised its price as predicted by any standard microeconomics textbook. The determination of the exact optimal uniform price and the personalized pricing structure requires us to estimate the proposed demand model. In the next section, we discuss the demand estimates.

TABLE 4
ACQUISITION AND RETENTION RATES (September 2015)

| Price ($) | Acquisition | At Least 1 Month | At Least 2 Months | At Least 3 Months | At Least 4 Months |
|---|---|---|---|---|---|
| 19 | .36 | .8 | .77 | .61 | .56 |
| 39 | .32 | .75 | .73 | .52 | .47 |
| 59 | .27 | .65 | .63 | .49 | .4 |
| 79 | .29 | .69 | .64 | .5 | .39 |
| 99 | .24 | .69 | .66 | .48 | .38 |
| 159 | .2 | .63 | .61 | .43 | .34 |
| 199 | .18 | .56 | .5 | .31 | .19 |
| 249 | .17 | .63 | .59 | .39 | .27 |
| 299 | .13 | .58 | .53 | .35 | .29 |
| 399 | .11 | .54 | .52 | .37 | .25 |

### 2. Demand Estimation

We now use the data from the field experiment to estimate the logit demand model using our WLB estimator discussed in section III.B.[18] Since the experiment randomized the prices charged to each consumer, we do not face the usual price endogeneity concerns associated with demand estimation using observational databases (e.g., Berry 1994).

Our demand specification allows for a heterogeneous treatment effect of the price on demand. To accommodate heterogeneity, we use 12 categorical feature variables that are self-reported by the prospective consumers during the registration stage. We break the different levels of these variables into 133 dummy variables, summarized in the vector $x_i$. We include the main effects of these 133 dummy variables in the intercepts of our model, $\alpha$, and the 133 interaction effects with price in the slope, $\beta$.[19]

In addition to our WLB estimates, we also report results from other approaches that are easier to implement than WLB. We report the MLE estimates of a model that includes all 266 covariates (main effects and interaction effects with price), which we expect would suffer from overfitting. MLE is much easier to estimate computationally but faces potential overfitting problems. In addition, we report results from the unweighted LASSO penalized regression estimates with optimal penalty selected by cross validation. While the LASSO is easier to implement than WLB, it has the disadvantage of not allowing us to characterize statistical uncertainty and conduct inference. For both the LASSO and the WLB, we always retain the main effect of price. However, even when we do not force price to be retained, the main price effect is always found to be part of the active set.

To compare these specifications, table 5 reports in-sample and out-of-sample fit measures. We assess model fit using the Bayesian information criterion (BIC), the asymptotic approximation of the Bayes factor, which can be used to select between models based on their posterior probabilities (Schwarz 1978). Since the BIC includes a penalty for the number of parameters, it is robust to overfitting concerns. For MLE, we report the BIC. For LASSO, the BIC includes a penalty for the number of model parameters (e.g., Zou, Hastie, and Tibshirani 2007). For our WLB estimator,

[18] We use the gamlr function in the R package gamlr to implement the logistic LASSO at each iteration of our Bayesian bootstrap. We simulate the weighted LASSO procedure as follows. For each iteration, we draw a vector of weights for each observation in our sample. We then draw a subsample by drawing with replacement from the original sample using our weights. The logistic LASSO is then applied to this new subsample.

[19] The methods proposed herein scale well with larger sets—we have implemented a version for the firm with the complete set of covariates. Others have had success with the general approach. For instance, Taddy (2015a) successfully implements the approach in a distributed computing environment for applications with thousands of potential covariates.

TABLE 5
Predictive Fit from MLE, LASSO, and WLB Estimation

| Model | In-Sample BIC (1) | Out-of-Sample RMSE (2) | Out-of-Sample Hit Rate (%) (3) |
|---|---|---|---|
| MLE | 10,018.78 | .412 | 70.3 |
| LASSO | 8,366.47 | .410 | 76.9 |
| WLB range | (7,805.11, 8,940.06) | .405 | 76.9 |

Note.—For WLB, we report the range across all 100 bootstrap replications. In-sample results are based on the entire September 2015 sample with 7,866 firms. Out-of-sample results are based on a randomly selected (without replacement) training sample representing 90% of the firms and a holdout sample with the remaining 10% of the firms.

we report the range of BIC values across the 100 bootstrap replications of the LASSO estimator used for constructing our Bayesian bootstrap estimate of the posterior, $F(\Theta)$.

We evaluate in-sample fit using the entire sample. As expected, table 5 shows that the switch from MLE to LASSO improves the in-sample BIC: 10,018 versus 8,366. This improvement is consistent with our concern that the MLE using all the features will overfit the data. Recall that our objective with the WLB is not prediction but rather inference. The fact that WLB provides comparable fit to the LASSO in-sample, with an average BIC (across bootstrap replications) very similar to the LASSO's BIC, indicates that we have not sacrificed predictive power in the process.

Our results suggest that regularization matters quite a bit, which speaks to the importance of variable selection and model uncertainty. Across the 100 bootstrap replications we conduct, models retain as few as 58 to as many as 188 features in the active set—the variables included in the model. Among the features, 172 have a more than 50% posterior probability of being nonzero (i.e., are retained in over 50% of the bootstrap replications). If we look at the six parameters with more than 90% posterior probability of being nonzero, these include diverse factors such as "job in British Columbia," "company type: staffing agency," "employment type: full-time," and "is resume required." The fact that we do not see a systematic type of variable exhibiting high posterior probability reinforces the importance of using regularization to select model features as opposed to selecting features manually based on managerial judgment.

As an additional verification, we also examine the out-of-sample predictive fit of each of our estimators in column 2 of table 5. We first split the sample into training and prediction subsamples, randomly assigning 90% of the consumers to the training sample and the remaining 10% to the prediction sample. We run each specification using the training sample. We report the out-of-sample RMSE and hit rate to assess model prediction. The hit rate classifies each respondent as choosing the alternative with the highest predicted probability. The WLB slightly outperforms

TABLE 6
POSTERIOR EXPECTED CONVERSION AND REVENUE PER CONSUMER BY PRICING STRUCTURE FOR SEPTEMBER 2015 EXPERIMENT

| | | Expected Conversion Rate | | Expected Revenue ($ per Consumer) | |
|---|---|---|---|---|---|
| Decision-Theoretic Pricing Structure | Price ($) (1) | Mean (2) | 95% Credibility Interval (3) | Mean (4) | 95% Credibility Interval (5) |
| Control | 99 | .25 | (.23, .28) | 25.09 | (23.02, 27.48) |
| Uniform | 327 | .12 | (.1, .14) | 39.01 | (31.9, 46.58) |
| Personalized | (126, 6,292) | .12 | (.1, .14) | 46.57 | (32.89, 60.8) |
| Personalized ($499 price cap) | (126, 499) | .13 | (.11, .16) | 42.21 | (33.41, 50.53) |

both alternative models on RMSE. While it also generates better out-of-sample choice predictions than MLE, it provides identical choice predictions to the basic LASSO. This latter result is not altogether surprising and merely highlights the importance of regularization for our high-dimensional feature set. The key advantage of WLB lies in its ability to generate reliable inferences, as demonstrated in section IV.E, where we use a second field experiment to assess the sampling properties of the three estimators.

### C. *Decision-Theoretic Pricing*

We now use our WLB demand estimates to calibrate ZipRecruiter's decision-theoretic price-optimization problems. Since we do not impose any restrictions on the range of parameter values, we cannot rule out the possibility of positive price coefficients or excessively large willingness to pay, two issues that could interfere with the optimization. For the price optimization procedures, we top-coded any draws for which $\mathbb{E}[\beta(x_i)|\mathbf{D}, x_i] \geq 0$ at the highest negative value of $\mathbb{E}[\beta(x_i)|\mathbf{D}]$.[20] In section VI, we explore the sensitivity of our results to a more sophisticated deep-learning algorithm. All of the price coefficients are found to be negative under the deep-learning algorithm. We also show that our main pricing-related findings based on the LASSO are robust to the deep-learning algorithm.

Table 6 summarizes the predicted economic outcomes associated with the different price structures considered. For each pricing structure, we report the corresponding posterior expected conversion rate (i.e., share of consumers who pay for a month of service), posterior expected revenue per consumer, and posterior expected consumer surplus. Ninety-five percent posterior credibility intervals are also reported for each of these predicted outcomes.

[20] This top-coding affects only 6% of the posterior draws of $\{\beta^b(x_i)\}_{b=1}^B$.

We begin with an analysis of optimal uniform pricing. At ZipRecruiter's base price of \$99, the posterior expected own-price elasticity of demand is only $-0.33$ with a 95% posterior credibility interval of $(-0.41, -0.26)$. Consistent with our model-free analysis above, ZipRecruiter was pricing on the inelastic region of demand before the experiment. Recall from figure 2 that the revenue-maximizing price appeared to lie between \$249 and \$399. The posterior expected own-price elasticity is $-0.82$ for a price of \$249 and $-1.15$ for a price of \$399.

The decision-theoretic optimal uniform price, as defined in equation (5), is \$327. Comparing column 3 of the first and second rows of table 6, we can see that the optimized uniform pricing policy increases ZipRecruiter's posterior expected revenue per consumer by over 55% relative to its \$99 base price, in spite of lowering conversion from 25% to 12%. Not surprisingly, we find an approximately 100% posterior probability that uniform optimal pricing is more profitable than \$99.

We can use our demand estimates to conduct another check that ZipRecruiter would indeed optimally increase its price relative to \$99, even after accounting for the discounted future cash flows from retained consumers. Assume that a consumer's retention probability in any given month is identical to the acquisition probability. The uniform optimal price that maximizes discounted cash flows is then

$$p^{\mathrm{NPV}} = \arg\max_{p} \frac{1}{1 - \delta \mathbb{P}(y_i = 1|p)} p \mathbb{P}(y_i = 1|p), \tag{13}$$

where $\delta$ represents the discount factor. If we assume that $\delta = 0.996$, we obtain $p^{\mathrm{NPV}} = \$261$, which once again confirms the suboptimality of the \$99 price.

We now explore decision-theoretic personalized pricing. Figure 4 summarizes the degree of estimated heterogeneity across consumers. In figure 4*A*, we report the distribution of consumers' posterior mean price sensitivities

$$\mathbb{E}[\beta(x_i)|\mathbf{D}, x_i] = \frac{1}{B}\sum_{b=1}^{B} \beta^b(x_i).$$

The dispersion across consumers suggests a potential opportunity for ZipRecruiter to price discriminate. In figure 4*B*, we report the distribution of posterior mean surplus across consumers when ZipRecruiter prices its monthly service at \$99:

$$\mathbb{E}[V(p, x)|\mathbf{D}, x_i, p = \$99] = -\frac{1}{B}\sum_{b=1}^{B} \frac{\log(1 + \exp(\alpha^b(x_i) - \$99 \times \beta^b(x_i)))}{\beta^b(x_i)}. \tag{14}$$

Figure 4*B* illustrates the wide dispersion in dollar value that consumers derive from the availability of ZipRecruiter when it costs \$99. The

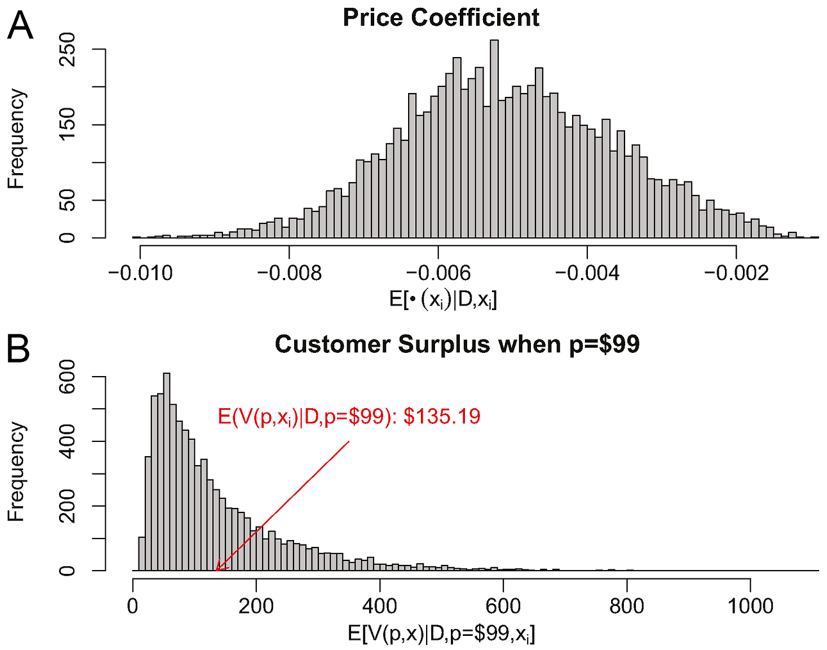


FIG. 4.—Distribution across consumers of posterior mean price sensitivity and posterior surplus from the provision of the service ($N$ = 7,867).

2.5 percentile, median, and 97.5 percentile willingness to pay are $23.55, $99.04, and $443.59, respectively. The magnitudes and degree of dispersion in value indicate an opportunity for ZipRecruiter to price discriminate using the registration features as a segmentation scheme.

We find considerable dispersion in the prices, ranging from as low as $126 to as high as $6,292. Across our $N = 7{,}866$ consumers, all of the personalized prices are strictly larger than ZipRecruiter's $99 baseline price. In spite of the range of prices, some exceeding $1,000, the median price is $277, which is much lower than the optimal uniform price, $327. Therefore, the majority of consumers would benefit from personalized pricing relative to uniform pricing. Comparing column 3 of the second and third rows of table 6, we see that the decision-theoretic personalized pricing increases ZipRecruiter's posterior expected revenue per consumer by 19% relative to uniform pricing, from $39.01 to $46.57. Moreover, compared with ZipRecruiter's base price of $99, decision-theoretic personalized pricing increases posterior expected revenue per consumer by 86%.

A concern with our personalization scenario is that about one-quarter of our recommended prices exceed the highest price in the experiment, $399, with many in excess of $1,000. ZipRecruiter's management team indicated that they would be unlikely to consider prices above $499.[21] In the fourth row of table 6, we recompute the decision-theoretic prices when we impose an upper bound of $499. As expected, this cap increases the posterior expected conversion to 13%. Expected posterior revenue per consumer is still 8% higher than under uniform pricing. The expected posterior revenue per consumer from capped personalized pricing exceeds that of uniform pricing with a posterior probability of 98%.

Based on conversations with ZipRecruiter management, we also do not expect any competitive response from other platforms. Our recommendations involve increasing (not decreasing) prices above the baseline of $99, mitigating any concerns about triggering a price war.

The incremental profitability of personalization in general depends crucially on the "no-arbitrage" condition, which rules out unintended strategic behavior by consumers (e.g., Fudenberg and Villas-Boas 2006; Chen, Li, and Sun 2015; Bonatti and Cisternas 2018). In the ZipRecruiter context, the no-arbitrage condition requires that consumers self-report their company features truthfully during the registration stage. There is no way for us to verify the accuracy of the self-reported features. However, we showed above that company features predict demand, and in section IV.E, we show that personalization generates higher profits out of sample than alternative pricing structures (e.g., uniform optimal pricing) that do not rely on self-reported features. We also believe that truthful self-reporting will remain incentive compatible at ZipRecruiter in the longer term for at least three reasons. First, most consumers would not learn about differential pricing because ZipRecruiter does not post its prices in a public manner. A firm must complete the registration process to obtain a price quote, making it difficult to use software to scrape ZipRecruiter's prices under different registration profile responses. Second, consumers face an arbitrage cost in the sense that misreporting features has an adverse effect on ZipRecruiter's key service: the résumé-matching algorithm uses company features to determine the ideal recruiting prospects. Arbitrage costs are prevalent in other industries that have studied personalized pricing. For instance, in the consumer packaged goods industry, consumer transaction histories are used to determine differential price elasticities (e.g., Rossi, McCulloch, and Allenby 1996; Chintagunta, Dubé, and Goh 2005). A high-willingness-to-pay consumer would need to purchase less preferred brands on a regular basis for her purchase history to generate a high-price-elasticity signal.[22] Third, it would not be possible

[21] This cap reflected both concerns with projecting too far outside the range of the data and, more importantly, charging prices that they felt might create negative goodwill with consumers.

[22] Arbitrage costs also arise in the emerging trend of geographic targeting using mobile coupons. High-willingness-to-pay consumers would need to incur time and travel costs to visit and dwell in locations associated with lower willingness to pay to receive a discount (e.g., Dubé et al. 2017).

for a consumer to determine which combination of features generates low prices purely because of the complexity of the WLB algorithm that uses 133 features. Nevertheless, we cannot rule out challenges with the no-arbitrage condition in the longer term at ZipRecruiter or for other industries with lower transaction costs for price discovery (e.g., without a registration requirement).

### D. *The Information Content of Features*

We now explore the types of consumers who benefit from personalized pricing. While our experiment was not designed to recover the causal effect of specific individual firm features on willingness to pay, it is nevertheless interesting to analyze the role of feature information as an exploratory exercise. We find that the job benefit features are the most highly correlated with the personalized prices. For instance, "job total benefits" and the presence of "medical benefits" have a correlation of 0.31 and 0.27, respectively, with the personalized price levels. However, the correlational value of information can be clouded by the fact that certain features, such as state and company type, comprise many underlying dummy variables (e.g., 62 state/province dummy variables) that may be important drivers of prices collectively.

As an exploratory exercise, we classify each of the feature variables into $g = 1, \ldots, 6$ groups: state, benefits, job category, employment type, company type, and declared number of job slots. We then use entropy to measure the incremental information content associated with a feature group. Let $\mathcal{X}$ represent the complete feature set, and let $f(p^*|\mathcal{X})$ denote the density of personalized prices based on information set $\mathcal{X}$. To assess the targetable information in each group $g$, we drop all of its corresponding features and rerun the WLB algorithm and the personalized pricing calculations to derive $f(p^*|\mathcal{X}_{-g})$, where $-g$ denotes the exclusion of feature group $g$. We then compute the Kullback-Leibler divergence in the distribution of personalized prices when we exclude feature group $g$:

$$\mathrm{KLD}(\mathcal{X}||\mathcal{X}_{-g}) = \int_p f(p|\mathcal{X}) \log\left(\frac{f(p|\mathcal{X})}{f(p|\mathcal{X}_{-g})}\right).$$

We effectively treat $f(p^*|\mathcal{X}_g)$ as our target distribution so that $\mathrm{KLD}(\mathcal{X} \| \mathcal{X}_{-g})$ measures the entropy associated with approximating $f(p^*|\mathcal{X})$ using $f(p^*|\mathcal{X}_{-g})$, the distribution of prices based on the narrower information set that excludes the feature group $g$.

We can now assess the relative incremental information associated with each feature group by ranking them in terms of divergence. State is the most informative group ($\mathrm{KLD}(\mathcal{X}||\mathcal{X}_{-\{\text{state}\}}) = 0.032$), followed by job category ($\mathrm{KLD}(\mathcal{X}||\mathcal{X}_{-\{\text{job category}\}}) = 0.029$), benefits ($\mathrm{KLD}(\mathcal{X}||\mathcal{X}_{-\{\text{benefits}\}}) = 0.018$), employment type ($\text{KLD}(\mathcal{X}||\mathcal{X}_{-\{\text{employment type}\}}) = 0.0078$), company type ($\text{KLD}(\mathbf{D}||\mathbf{D}_{-\{\text{company type}\}}) = 0.004$), and declared number of job slots ($\text{KLD}(\mathbf{D}||\mathbf{D}_{-\{\text{job slots}\}}) = 0.002$). Since company type and state each require only a single categorical question during the registration process on ZipRecruiter's website, these information sources are more efficient to elicit from prospective consumers. In sum, individual features such as company size and benefits are the most correlated with personalized prices. However, aggregating information into groups, the distribution of personalized prices seems most influenced by broad job categories and geographic locations.

### *E. Experiment 2: Validation*

A novel feature of our study is that we conducted a second field experiment to test the policy recommendations based on our empirical analysis of the first experiment. This second experiment allows us to confirm the predictive validity of our structural analysis in the previous section.

We conducted the second field experiment between October 27, 2015, and November 17, 2015, using a new sample of prospective consumers who arrived to the ZipRecruiter paywall during this period and had not previously paid for the firm's services. Each prospective consumer was randomly assigned to one of the three following pricing structures:

1. Control pricing—$99 (25%).
2. Uniform pricing—$249 (25%).
3. Personalized pricing (50%).

We oversampled the personalized pricing cell to obtain more precision given the dispersion in prices charged across consumers.

The tested pricing structures were formulated in part based on ZipRecruiter's own needs. For instance, as we explained earlier, they chose a uniform price of $249 because, based on the earlier experiment, (i) the profit implications relative to the optimum were minimal and (ii) the management believed that $249 would be more palatable on account of similar prices being used elsewhere in the industry. For our personalized pricing cell, consumers were charged a price based on the values of $x_i$ that they reported during the registration stage. As we indicated in the above section, ZipRecruiter capped the personalized prices at $499. In addition, they asked us to round the personalized price down to the nearest $9, discretizing the prices into $10 buckets ranging from $119 to $499. For instance, a consumer with a targeted price of $183 would be charged $179. ZipRecruiter used this rounding because they believed consumers would find the $9 endings on prices more natural. Based on our demand estimates, this rounding has very little impact on the predicted profits of personalization.

During this period, 12,381 prospective consumers reached ZipRecruiter's paywall. Of these prospectives, 5,315 were starters and the remainder were larger firms. Among our starters in the November 2015 study, 26% were assigned to control pricing, 27% to the uniform pricing, and 47% to the personalized pricing. In the personalized pricing cell, the lowest price was $99, and hence neither of our test cells ever charged a prospective consumer less than the baseline price of $99.

To verify that our three experimental cells are balanced, we compare the personalized prices that would have been used had we implemented our personalized pricing method in each cell. Figure 5 reports the density of personalized prices in each cell. For the control cell ($99) and test cell ($249), these are the personalized prices that subjects would have been shown had they been assigned to the personalized pricing test cell instead. The three densities are qualitatively similar, indicating that the nature of heterogeneity and willingness to pay is comparable in each cell. This comparison provides a compelling test for the balance of our randomization, as it indicates that our distribution of personalized prices would look the same across each of the experimental cells.

*Out-of-sample validation of model predictions.*—A novel feature of our case study is the ability to use the November 2015 experiment to validate our proposed WLB inference procedure along with the predictions from our

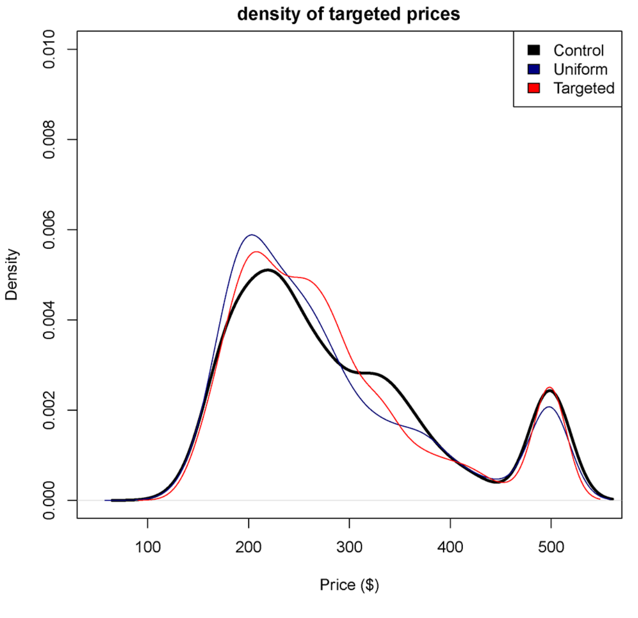


Fig. 5.—Density of targeted prices in each cell (November 2015). For each of the cells, we plot the estimated density using a Gaussian kernel.

structural model and the corresponding inferences regarding profits under different pricing structures discussed in section IV.C. The boxplots in figure 6 compare the realized sampling distribution for conversion across several of the tested price cells with the corresponding inferences for conversion using our WLB approach versus the post-LASSO MLE and classical MLE approaches (as discussed at the end of sec. III.B). To

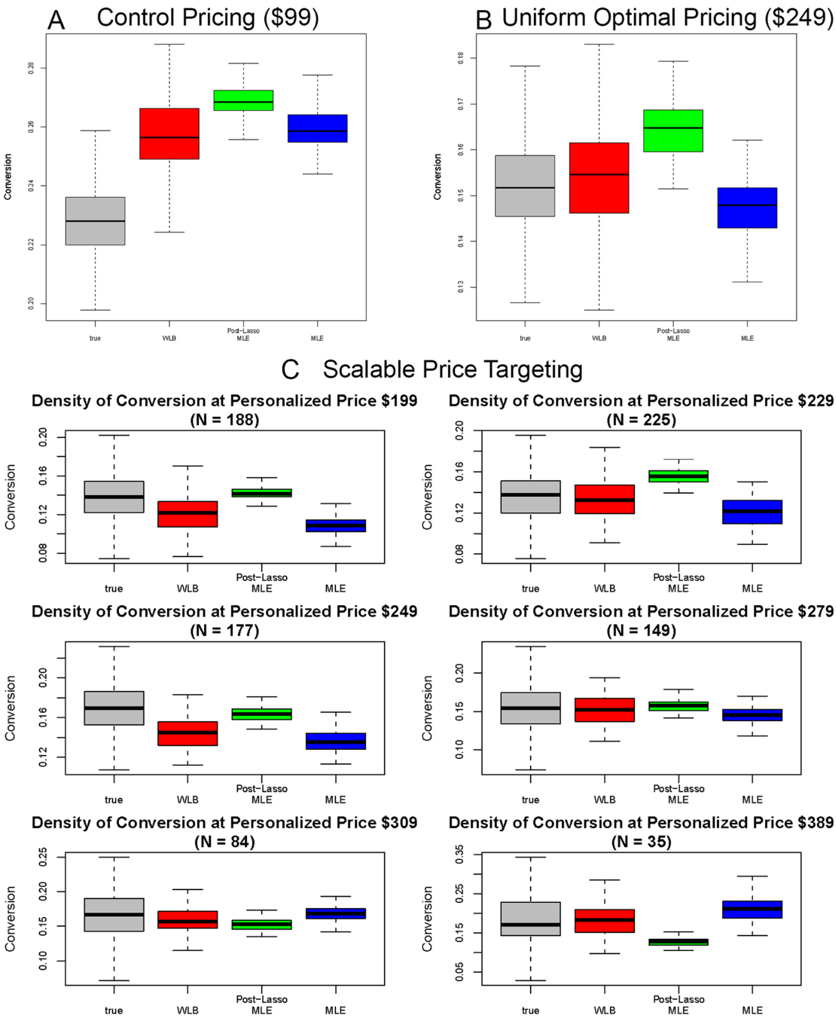


FIG. 6.—Comparison of predicted and realized conversion. The plots compare the empirical density of realized conversion for a given pricing structure to the corresponding predicted densities for WLB, post-LASSO MLE, and MLE, respectively. The density of realized conversions is computed by bootstrapping (with replacement) from the November data.

account for sampling error in our realized outcomes, we bootstrap our sample 1,000 times (sampling with replacement). For WLB, we use the draws from the posterior distribution. For post-LASSO MLE and MLE, we use a parametric bootstrap from the asymptotic covariance matrix. The boxplots indicate that WLB comes much closer to approximating the observed sampling distribution in conversion rates across price cells. Relative to WLB, both post-LASSO MLE and MLE generate what appear to be strikingly understated degrees of statistical uncertainty. This is not surprising since, unlike post-LASSO MLE, WLB accounts for model uncertainty. Unlike MLE, WLB uses regularization to avoid model overfitting. At the bottom of each panel, we report the Kullbach-Leibler divergence for each of our three estimators relative to the true distribution of realized conversions. The divergence of WLB is always considerably smaller than for post-LASSO MLE and MLE, often by orders of magnitude. These findings suggest that WLB is providing a reasonable approximation of the posterior uncertainty over both the model specification and the feature weights. The results also suggest that personalized pricing for a company like ZipRecruiter is a big data problem in the sense that the selection of model features plays an important role in addition to the usual estimation of feature weights.

In table 7, we report the realized conversion rates and revenue per consumer across our three pricing structures: control ($99), test ($249), and test (personalized pricing). For realized outcomes, we report the 95% confidence interval. We also report the posterior expected conversion rate and revenue per consumer in each of the three cells based on our estimates from the September 2015 training sample. Specifically, we

TABLE 7
Predicted versus Realized Outcomes in November 2015 Experiment

| | Control ($99) | Test ($249) | Test (Personalized Pricing) |
|---|---|---|---|
| Sample size | 1,360 | 1,430 | 2,485 |
| Mean conversion | .23 | .15 | .15 |
| | (.21, .25) | (.13, .17) | (.13, .16) |
| Mean revenue per consumer ($) | 22.57 | 37.79 | 41.59 |
| | (20.36, 24.77) | (33.15, 42.42) | (37.49, 45.7) |
| Posterior mean conversion | .26 | .15 | .14 |
| | (.23, .29) | (.13, .18) | (.12, .17) |
| Posterior mean revenue per consumer ($) | 25.5 | 38.37 | 41.05 |
| | (23.26, 28.31) | (32.04, 44.9) | (33.78, 48.78) |

Note.—Below each realized outcome, we report the 95% confidence intervals in parentheses. Below each posterior predicted outcome, we report the 95% credibility interval in parentheses.

use the posterior distribution of the parameter estimates, $F(\Theta|\mathbf{D}^{\text{Sept}})$, and the observed features from our November subjects, $X^{\text{Nov}}$, to form our predictions. For each posterior mean, we also report the corresponding 95% credibility interval.

Starting with the realized outcomes, average conversion is higher in the control cell that has the lowest monthly price, as expected. Average conversion is almost identical in the uniform and personalized pricing cells, at 15%. However, the average profit per consumer is higher in the personalized pricing cell, as one would theoretically expect. Overall, the uniform pricing increases expected profits per consumer by 67.74% relative to control pricing, although our bootstrapped confidence interval admits a change as low as 46%. Personalized pricing increases expected profits by 84.4% relative to control pricing, although our bootstrapped confidence interval admits a change as low as 64%. These improvements from price discrimination are consistent with our predictions based on the September sample discussed in section IV.C. Finally, although not reported, our bootstrap generates an 87% probability that personalized pricing profits will exceed uniform profits.

These realized conversion rates and revenues per consumer are broadly consistent with our model predictions. In particular, the predicted outcomes for the uniform pricing at \$249 and the personalized pricing are almost identical to the realized values. These findings provide out-of-sample validation of the predictive value of our WLB estimator and our structural demand model. The second experiment also allows us to test our pricing policies out of sample. A test of the hypothesis that uniform pricing at \$249 is more profitable than uniform pricing at \$99 is strongly significant ($p < .01$). A test of the hypothesis that personalized pricing is more profitable than uniform pricing at \$249 is less precise ($p = .069$), although the point estimates for both cells closely correspond with our Bayesian predictions.

## V. Personalization, Data Policies, and Consumer Welfare

Having established that personalized pricing (large-scale third-degree price discrimination) generates a substantial increase in producer surplus, we now turn to the demand side of ZipRecruiter's B2B market. As explained above, ZipRecruiter was in the process of exploring ways to collect demand data and improve its pricing when we began the collaboration. Therefore, we use the optimal uniform price as our base case, not \$99, since the former reflects the textbook inverse-elasticity-rule pricing that would be predicted for a maturing company. Our analysis also focuses on the role of conditioning on features $x_i$ to set prices. One could implement uniform pricing with a demand model that does not condition on $X_i$

for estimation, instead using only price and conversion data. Although not reported herein, the optimal uniform price is almost identical in that case (i.e., \$324 as opposed to \$327). So for the remainder of our analysis, each of our pricing structures uses the same demand estimates, $\mathbb{P}(y_i = 1|p; x_i)$.

In what follows, we examine two aspects of consumer welfare: (i) the aggregate welfare differential created by the change in pricing policy and (ii) the impact of data policies on consumer surplus.

### A. *Consumer Welfare*

To analyze the consumer welfare implications of personalized pricing relative to optimal uniform pricing, recall that we use Atkinson's (1970) mean of order $r$ class of consumer welfare functions, which in our empirical setting corresponds to

$$S_r(\mathbf{p}) = \left[\frac{1}{N}\sum\mathbb{E}(V(\mathbf{p}, x_i))^r\right]^{1/r}, \tag{15}$$

where $V(\mathbf{p}, x_i)$ corresponds to the individual-level surplus as in equation (14). As discussed earlier, we follow Lewbel and Pendakur (2017) and restrict our attention to $r \in \{-1, 0, 1\}$, corresponding to the harmonic, geometric, and arithmetic means, the first two of which reflect inequality-averse preferences (on the part of the planner).

Panel A of table 8 reports our consumer welfare results for each decision-theoretic pricing structure. We start with row 3, corresponding to the

TABLE 8
Consumer Welfare and Data-Based Pricing

| Measure | $r$ | $S_r(\mathbf{p}^{\text{pers}})$ | $S_r(\mathbf{p}^{\text{unif}})$ | $\Delta = S_r(\mathbf{p}^{\text{pers}}) - S_r(\mathbf{p}^{\text{unif}})$ | $\%\Delta S_r(\mathbf{p})$ |
|---|---|---|---|---|---|
| | A. Comparing Theoretically Optimal Pricing Policies | | | | |
| Harmonic mean | −1 | 46.8255 | 33.6011 | 13.2244 | 39.36 |
| Geometric mean | 0 | 58.2786 | 57.5773 | .70127 | 1.22 |
| Arithmetic mean | +1 | 71.4094 | 95.2247 | −23.8153 | −25.01 |
| | B. Implemented Personalized versus Optimal Uniform Pricing Policies | | | | |
| Harmonic mean | −1 | 50.1969 | 33.6011 | 16.5958 | 49.39 |
| Geometric mean | 0 | 67.3144 | 57.5773 | .7371 | 16.91 |
| Arithmetic mean | +1 | 93.2841 | 95.2247 | −1.9406 | −2.04 |
| | C. Implemented Personalized versus Implemented Uniform Pricing Policies | | | | |
| Harmonic mean | −1 | 50.1969 | 43.4767 | 6.7202 | 15.46 |
| Geometric mean | 0 | 67.3144 | 68.1889 | −.8745 | −1.28 |
| Arithmetic mean | +1 | 93.2841 | 105.3496 | −12.0656 | −11.45 |

conventional "total consumer surplus" standard, $r = 1$, for which the welfare function (15) is inequality neutral. Personalization reduces linearly aggregated consumer surplus considerably relative to uniform optimal pricing from \$94.78 to \$71.41 (25%) and by more than the increase in profits. Given the decline in conversion under personalized pricing, it is not surprising that we observe a decline in total surplus (firm and consumer). This decline in total surplus comes from less than half of the consumers. In fact, 63% of the consumers' personalized prices are lower than the uniform optimal price of \$327, indicating that over half of our consumers benefit from personalization even though total surplus is lower. Since all consumers are weighted equally, a small number of consumers exert an inordinate amount of influence on the average. We now turn to the inequality-averse consumer welfare functions, $r = -1$ and $r = 0$, respectively, and report the corresponding results in the first and second rows of table 8, panel A. Under both inequality-averse consumer welfare functions, personalization is preferred because the allocative benefits outweigh the decline in total surplus. These findings indicate how the articulation of the aggregate consumer welfare effects of a change in pricing policy depends on the planner's preferences and the choice of surplus aggregation metric. Although not reported in the table, we find that welfare is equal under uniform and personalized pricing at $r = 0.06$, suggesting that some amount of inequality aversion is required for social welfare to improve under personalized pricing.

For completeness, panels B and C of table 8 provide welfare calculations for two other comparisons: "implemented personalized versus optimal uniform" and "implemented personalized versus implemented uniform," respectively. The term "implemented" refers to the \$499 personalization cap and the uniform price of \$249 that were implemented by ZipRecruiter in practice. As one would expect, introducing a price cap at \$499 increases total consumer surplus considerably (under all metrics). It follows then that personalization is viewed more favorably than in panel A. In particular, linearly aggregated consumer surplus falls by 2.04% (in contrast to 25% in panel A) relative to uniform pricing, while still allowing the firm to generate a more than 8% gain in profits. We do not claim that the use of such caps and the results herein would generalize to other firms and/or other industries where personalized pricing could be implemented. Finally, panel C of table 8 shows that when comparing the implemented versions of personalized and uniform pricing, personalization is preferred only under the more extreme inequality-averse welfare function (harmonic mean with $r = -1$). The shift toward uniform pricing reflects the fact that ZipRecruiter implemented a much lower price than optimal (\$249 vs. \$327) by selecting a value off the test grid instead of maximizing its posterior expected profits. In this case, the lower uniform price more than offsets the benefits of a more equitable allocation of surplus unless inequality aversion is strong.

*B. Data Policies and Consumer Surplus*

Policies such as GDPR and CCPA have been enacted to protect consumers' privacy broadly but also to prevent firms from surplus extraction. Theoretically, however, it is possible that restricting the types of data that firms are permitted to use for personalized pricing could harm consumer surplus (e.g., Bergemann, Brooks, and Morris 2015). We now use our ZipRecruiter case study to explore how restrictions over the set of features available to a firm for pricing purposes affects consumer surplus. For most of the analysis that follows, we focus on the usual aggregate surplus metric with linear aggregation (i.e., $r = 1$).[23]

Formally, we need to recast the analysis in Bergemann, Brooks, and Morris (2015) for the context of data-based marketing. Suppose the firm uses all the available data to estimate the demand parameters, $F_\Theta(\Theta|\mathbf{D})$, as before. However, suppose also that the firm is permitted to use only a subset of the $g = 1, \ldots, 6$ sets of consumer features for the personalization of prices. Let $\mathcal{X}$ represent the complete feature set, let $\mathcal{X}^o \subset X$ denote the subset of features the firm can use for segmenting consumers and setting personalized prices, and let $\mathcal{X}^u \subset X$ represent the features the firm cannot use for segmentation. The firm can partition demand for a consumer $i$ with features $X_i$ into the targetable and nontargetable components as follows:

$$\mathbb{P}(p; X_i^o, \Theta) = \frac{1}{1 + \exp(-(\alpha(X_i^o, X_i^u, \Theta) + \beta(X_i^o, X_i^u, \Theta)p))},$$

where

$$\alpha(X_i^o, X_i^u, \Theta) = \alpha + X_i^o\alpha_o + X_i^u\alpha_u,$$
$$\beta(X_i^o, X_i^u, \Theta) = \alpha + X_i^o\alpha_o + X_i^u\alpha_u.$$

For a given segmentation structure, $\mathcal{X}^o$, the personalized pricing problem is

$$p_i^* = \arg\max_p\left\{(p - c)' \int\int \mathbb{P}(p; X_i^o, \Theta)\, dF_{X^u}(X^u|X^o)\, dF_\Theta(\Theta|\mathbf{D})\right\}, \quad (16)$$

where $F_{X^u}(X^u|X^o)$ represents the firm's beliefs about a consumer's unobserved traits, $X^u$, conditional on her observed traits, $X^o$. We use an empirical estimate of $F_{X^u}(X^u|X^o)$ to capture the fact that even though the firm cannot segment on $X^u$ directly, it can nevertheless form an expectation about those unobserved traits from the empirical correlation between

[23] Results for $r \in \{0, -1\}$ are available from the authors upon request. In line with our previous discussion, it is possible for personalized pricing to be surplus positive relative to uniform pricing if one uses inequality-averse aggregation. The nonmonotonicity finding pertaining to data that we discuss below holds even with the alternate aggregation metrics.

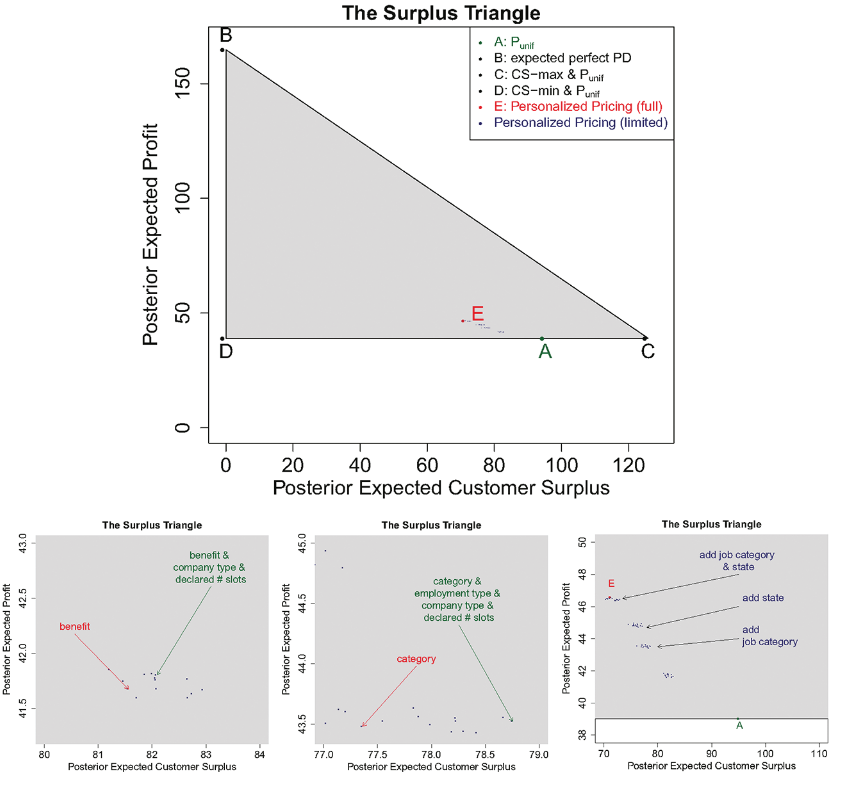


FIG. 7.—Surplus triangle. CS = customer surplus; PD = price discrimination.

features. We solve the personalized prices, 16 corresponding to each of the 62 possible combinations of the $g = 1, \ldots, 6$ feature groups, which includes the case using all the feature variables.[24]

We report the range of feasible personalized pricing outcomes in the surplus triangle in figure 7, the statistical decision-theoretic analog of the feasible surplus allocations examined in Bergemann, Brooks, and Morris (2015). All expectations for posterior surplus are taken over the full posterior distribution, $F_\Theta(\Theta|\mathbf{D})$. Point A represents the case where the firm has conducted demand estimation but does not use any of the consumer-level features for segmentation. In this case, the firm charges the optimal uniform price and earns the standard, uniform monopoly profits. Point B represents the purely theoretical case where the firm

[24] We simulate the integrals by using our posterior WLB draws from $F_\Theta(\Theta|\mathbf{D})$ and 100 independent draws from $F_{X^u}(X^u|X^o)$. We use a $K$-nearest neighbor approach to estimate $F_{X^u}(X^u|X^o)$ using the Hamming distance between each of the observations in our training sample and $K = 200$ as our cutoff.

observes not only all of the consumers' features but also their utility shocks, $\{\epsilon_i\}_{i=1}^{N}$. In this case, the firm conducts perfect price discrimination. See appendix E for details on the calculation of the expected posterior first-degree price discrimination outcomes (i.e., where the uncertainty is for the analyst and not for the firm). Point C represents the case where consumer surplus is maximized subject to the constraint that the firm earns the expected posterior uniform monopoly profits. Finally, point D represents the case where expected posterior social surplus is minimized, with the firm earning the expected posterior uniform monopoly profits and consumer surplus is zero. Bergemann, Brooks, and Morris (2015) show that every point in this surplus triangle represents a potentially feasible segmentation with third-degree price discrimination.

The top panel of figure 7 also indicates in blue all of the 62 possible segmentation schemes based on our observed feature set. Point E corresponds to the personalized pricing scenario already discussed and represents the most granular segmentation using all of the observed features. As expected, each of the 62 feasible segmentation schemes is more profitable than uniform pricing. However, these personalized pricing schemes are not nearly as profitable, in expectation, as perfect price discrimination. Even when all the features are used, personalization generates only 30% of the expected posterior profits under perfect price discrimination.

Turning to the demand side, each of our 62 feasible segmentation schemes reduces consumer surplus relative to points C and A (uniform pricing), sometimes by as much as 30% relative to point A. Even though it is theoretically possible for a segmentation scheme to exist that would increase the expected posterior consumer surplus relative to the case of uniform pricing, none of the 62 scenarios achieves this outcome. The best-case scenario, which conditions prices only on the "employment" and "number of declared job slots" features, generates 87% of the consumer surplus under uniform pricing. Recall from above that when we implement ZipRecruiter's price cap at $499, personalization based on the full feature set reduces consumer surplus by 2% while improving posterior expected profits by over 8%. Our data do not allow us to determine whether firms would implement such price caps in general. We also highlight the point that personalized pricing does come close to the case of true perfect price discrimination, which would extract all the consumer surplus. Even with expanded data collection, it is unlikely that a firm could truly perfectly price discriminate using consumer data. Even in the brand choice literature where pricing could be conditioned on detailed, individual-level transaction histories, there is still a lot of unpredictable, random brand switches (e.g., Dubé, Hitsch, and Rossi 2010).

The bottom three panels zoom in on the surplus triangle to examine how different data policies influence consumer surplus. The leftmost panel indicates that when we allow the firm to target prices only based on "benefits," total expected posterior consumer surplus is almost $1 lower than when we also allow the firm to target on "company type" and "declared number of job slots." The removal of these latter two features and using only "benefits" reduces consumer surplus with 87% posterior probability. The middle panel shows a similar result. Targeting prices only based on "job category" generates more than $1 less consumer surplus than when the firm is also permitted to target based on "employment type," "company type," and "declared number of job slots." The removal of these latter three features and using only "job category" reduces consumer surplus with 98% posterior probability. However, the rightmost panel indicates that some consumer features strictly harm consumer surplus. In particular, allowing the firm to target on "job category" and/or "state" reduces consumer surplus. These results indicate that allowing the firm to target on more granular data can be good for consumer surplus and that granularity per se does not harm consumers.

In spite of the decline in total consumer surplus, the percentage of consumers who benefit from personalization ranges from 59.4% to 62.2% across our 62 segmentation scenarios. Therefore, less than half the consumers bear the cost of personalization. To see this point more clearly, figure 8 plots density estimates of the change in posterior expected

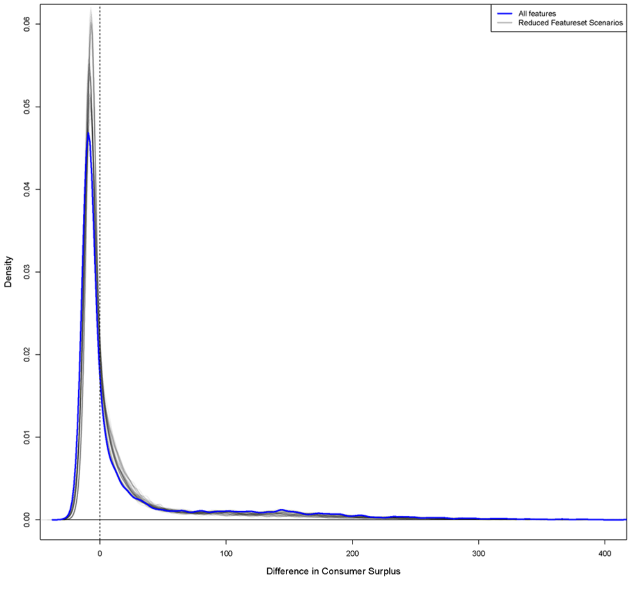


FIG. 8.—Densities of the change in expected posterior surplus across customers under personalized pricing versus uniform pricing. We report densities for all reduced featureset database scenarios, each in gray. We highlight the main case that uses all the features in blue.

surplus across consumers for each of the 62 segmentation scenarios versus uniform pricing. In each case, we see a large mass of consumers just to the right of \$0, representing the majority who benefit from personalized prices. We then see a long tail to the left of \$0 representing the minority of consumers who are harmed. If we correlate the incidence that a consumer benefits from personalization ($p_i^* < p^{\text{unif}}$) with the consumer features, we find that the two most highly correlated features are "small company type" (corr = 0.38) and "part-time employment" (corr = 0.31). At face value, these results suggest that smaller companies with part-time staff are the most likely to benefit from personalization. In contrast, the most negatively correlated features are all related to job benefits—for example, "total job benefits" (corr = −0.81), "full-time employment" (corr = −0.37) and "medium company type" ($\rho = -0.25$).[25] Therefore, larger companies with full-time employment and high benefits are the most likely to be harmed from personalized pricing. Conceptually, this reallocation of consumer surplus from personalized pricing could be rationalized as fair under a Pareto-weight scheme that assigns higher social value to smaller, disadvantaged firms.

A key finding from our analysis is that we do not observe a monotonic relationship between the number of features used for segmentation and total consumer surplus or total number of consumers who benefit from personalization. Thus, granting the firm more access to consumer data does not per se lead to more consumer harm. However, this finding must be balanced against the fact that total consumer surplus falls for each of the segmentation scenarios considered relative to the base case of uniform pricing that does not condition on consumer features. In figure 7, we can see that the full segmentation using all six groups of consumer features generates more consumer surplus than several of the restricted scenarios. For instance, allowing the firm to condition its prices on all six feature groups increases consumer surplus by 1.4% relative to restricting the firm to conditioning on "job state," "benefit," and "company category" (i.e., removing all the features associated with "job benefits," "number of declared job slots," and "employment type"). Similarly, 61% of the consumers benefit from personalized prices conditioned on all feature variables, whereas only 59.4% benefit when the firm is allowed to condition its prices only on "benefits." In figure 8, we see that the density of the change in expected posterior surplus across consumers for full personalized pricing versus uniform pricing is shifted to the right of several of the other restricted segmentation scenarios. In sum, granting the firm access to more information is not per se worse for the consumer, as it can lead to segmentation schemes that allocate more surplus to the consumer.

[25] The "large company type" feature was excluded due to redundancy.

## VI. Robustness

Our results above used a LASSO regularization algorithm to determine the functional parameters $\{\alpha(x), \beta(x)\}$. In this section, we examine the robustness of our results to a more sophisticated machine-learning algorithm to model parameter heterogeneity using the deep-learning framework based on Farrell, Liang, and Misra (2021a, 2021b).

### A. *A Deep-Learning Approach*

Unlike the application of machine learning for prediction purposes, the choice of machine-learning algorithm and the need for model structure are more important in the context of demand estimation and inference. For example, the direct application of a random forest with a standard splitting rule to our demand estimation problem will lead to infinite prices for some subset of consumers. The forest will predict a constant purchase probability for any price at or above $399, the maximum-tested price in the experiment. The corresponding revenues will therefore increase without bound in prices, and no interior solution will exist. The implementation of shape restrictions on demand to obtain a unique, interior optimal price is difficult for most machine-learning tools and beyond the scope of this paper.

As explained in Farrell, Liang, and Misra (2021a), not all machine-learning methods are "structural compatible" in the sense that they can be embedded directly into a structural parametric model. For example, deep neural networks (DNNs) are structurally compatible, while random forests are not. We now examine the robustness of our results to more flexible DNNs that retain the logit structure of the choice model.

#### 1. Deep Learning

As before, a consumer with features $x$ facing prices $\tilde{p}_i = (1\ p_i)'$ derives the incremental utility from buying

$$\begin{aligned} \Delta U_i &= \alpha_i + \beta_i p_i + \epsilon_i \\ &= \alpha(x_i) + \beta(x_i) p_i + \epsilon_i \end{aligned} \tag{17}$$

and has corresponding choice probability

$$\mathbb{P}(y_i = 1 | p_i; \Psi_i) = \frac{\exp(\tilde{p}_i' \Psi_i)}{1 + \exp(\tilde{p}_i' \Psi_i)}.$$

We now model the parameter vector as a DNN:

$$\Psi_i = (\alpha(x_i), \beta(x_i))' = \Psi_{\text{DNN}}(x_i; \theta_{\text{DNN}}).$$

Since the observed consumer features in our data are discrete, the advantage of the DNN is limited to finding (possibly higher-order) interactions that might be relevant in explaining consumer choices. We use two architectures, one with two hidden layers and another with three layers. In each case, the specification allows for zero nodes in each layer. We limit the complexity of the model on account of the limited data ($N < 8{,}000$) in our application. Our results do not change qualitatively when we perturb the architecture while retaining a comparable degree of complexity of the network. We refer the interested reader to Farrell, Liang, and Misra (2021a, 2021b) for a more rigorous discussion of the algorithm and its implementation. As with our LASSO-based framework, we implement the Bayesian bootstrap by optimizing the objective function with randomized Dirichlet weights for $R = 100$ repetitions. This gives us our draws from the approximate posterior ($\hat{\theta}^r_{\mathrm{DNN}}$), and consequently we obtain $\Psi_{\mathrm{DNN}}(x_i; \hat{\theta}^r_{\mathrm{DNN}})$. Our subsequent demand and pricing analysis is analogous to our approach using the WLB above.

### 2. Comparison of Results Using LASSO and Deep Learning

To assess any potential differences between the LASSO and deep-learning algorithms, we compare the following sets of results: (1) individual parameter estimates, (2) uniform and personalized prices, and (3) the differences in consumer welfare across pricing policies.

*1 (Individual parameter estimates).*—Figure 9 compares the distribution of the posterior means across consumers for the three methods: LASSO, two-layer deep learning (DNN-2Layer), and three-layer deep learning (DNN-3Layer). The left panel plots the density of the parameters, while the right panel displays the boxplots and interquartile ranges for each of our algorithms. The dark line in the boxplot indicates the median and not the mean. The three distributions are qualitatively similar and cover very similar ranges of the parameter space. The means of the three distributions are quite close, with mean $\beta(x)$ of $-0.0058$, $-0.0054$, and $-0.006$ for the DNN-2Layer, DNN-3Layer, and LASSO, respectively.

However, we do observe some noteworthy differences. First, the deep-learning-based estimators tend to restrict the range of the price coefficient to be negative in spite of the fact that we have not imposed any sign restrictions. We conjecture that the potential for interaction effects between features may be leading to a better fit of the price effect. Second, the deep-learning parameters imply a higher degree of heterogeneity than those from the LASSO. In particular, the price coefficients exhibit higher variance and skewness than their LASSO counterpart.

*2 (Pricing policies).*—We obtain qualitatively similar optimal uniform prices under each of our three approaches: \$301.92, \$363.81, and \$323.34

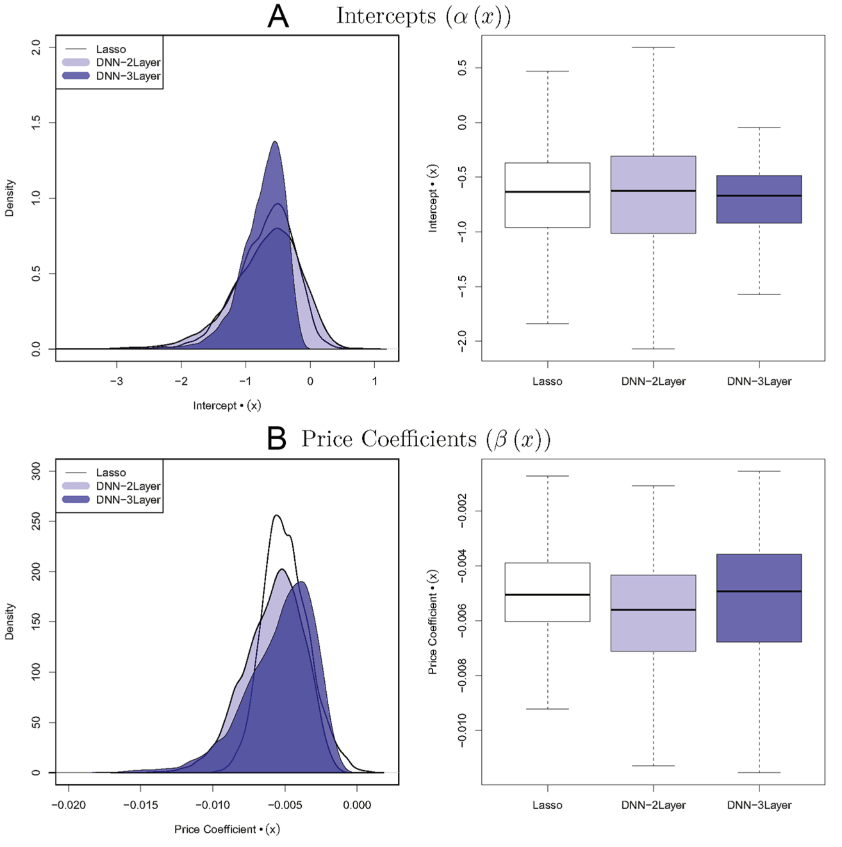


FIG. 9.—Comparison of individual posterior means of parameters.

for the DNN-2Layer, DNN-3Layer, and LASSO, respectively. To compare personalized prices, we plot the three sets of distributions in figure 10. While the median personalized prices (as seen in the boxplot) are close, the heterogeneity in these prices is quite different. In particular, the 3-Layer specification exhibits a higher variance corresponding to the higher variance in the parameter estimates. In spite of these differences, the LASSO specification does not show any systematic bias.

*3 (Welfare).*—For each of our three approaches, we compare welfare under the uniform and personalized pricing policies. As with the parameters and the optimal prices, our three approaches generate similar differences in welfare under the two pricing policies, including comparable medians (see fig. 11). All three methods find a large difference between the median and the mean (dotted line). In all cases, the mean is positive and the median is negative. This difference in sign between the mean

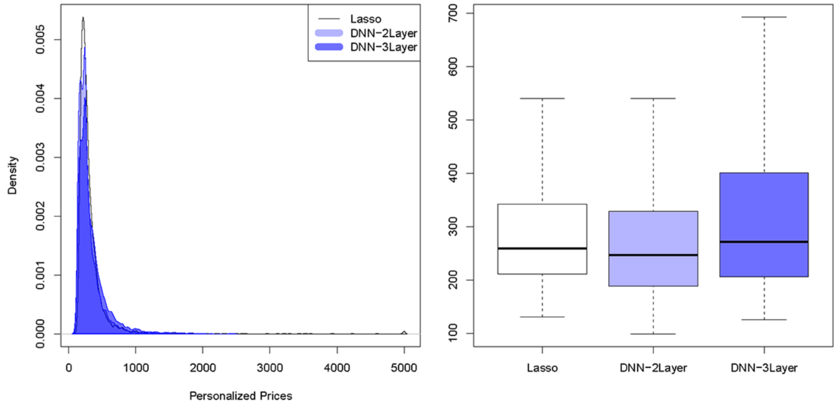


FIG. 10.—Comparison of personalized prices ($p^*(x)$).

and median once again indicates the sensitivity of welfare conclusions to the exact manner in which consumer surplus is aggregated by the social welfare function. Interestingly, the proportion of consumers who are worse off under uniform pricing is higher under the deep-learning framework. The intuition here is straightforward: since the price coefficients are well behaved relative to the LASSO (i.e., fewer values near or greater than zero), the consumer surplus values are less exaggerated under deep learning. Consequently, both the levels of consumer surplus and the differences are less variable in the deep-learning framework.

In summary, our key qualitative findings under the LASSO are robust to a more sophisticated deep-learning algorithm. All three of our estimators

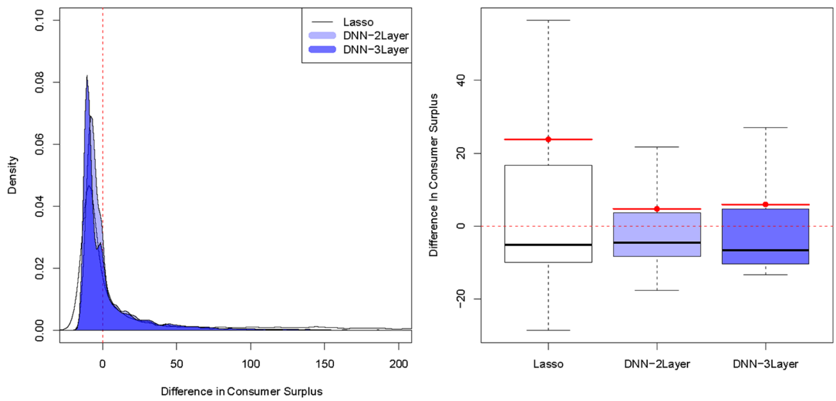


FIG. 11.—Comparison of differences in consumer surplus.

predict that total consumer surplus falls under personalized pricing. However, alternative inequality-averse welfare functions would likely favor personalization over uniform pricing.

### B. *Discussion*

In light of the public scrutiny of data-based marketing, of interest is how the results from the case study herein affect our beliefs about the welfare implications of personalized pricing (Maniadis, Tufano, and List 2014).[26] As discussed earlier, the popular press and public policy debate indicates a strong negative prior belief about the impact of personalized pricing on consumer welfare and a strong positive prior about the impact on firm profitability, in spite of the more neutral prior implied by the extant empirical literature. A formal Bayesian update as in Maniadis, Tufano, and List (2014) is infeasible and beyond the scope of this analysis. Having said that, we can use the ideas therein to articulate what the reader might reasonably conclude from our results. As with any Bayesian econometric analysis, the reader's posterior beliefs depend on the evidence, which in this case is a function of the model specification and the data.

Our analysis finds that, on the supply side, personalization increases profitability and hence a firm would be likely to implement a personalized pricing structure in a setting like ours, in contrast with recent work that fails to detect incremental profits from discriminating based on observed consumer feature variables. These results are robust to perturbations in data and the methodology used. As such, our analysis strengthens prior beliefs that personalized pricing improves profits.

The demand-side implications are more ambiguous. Our analysis demonstrates that posterior beliefs about consumer welfare are potentially affected by at least three factors: the social planner's preferences over the distribution of consumer surplus (i.e., the welfare function), the amount and nature of data conditioned on for personalized pricing, and the methodology used to analyze the data and classify consumers into types.

From a methodological perspective, section VI.A.2 shows that our key findings appear to be robust to different machine-learning algorithms. Hence, we would conjecture that posterior beliefs about the welfare effects of personalization are not dependent on the machine-learning method.

In contrast, the consumer welfare implications are quite sensitive to the specific welfare function used. Society's exact degree of inequality

[26] We thank the editors for suggesting this discussion.

aversion ($r$) should be part of the reader's subjective prior, and any updating of beliefs about personalized pricing will crucially depend on this quantity. For instance, total, linearly aggregated consumer surplus falls. Therefore, under inequality-neutral societal preferences ($r = 1$), our case study supports the a priori concerns expressed in CEA (2015) and would lead to stronger posterior beliefs about the adverse consumer welfare effects of personalized pricing. Under inequality-averse societal preferences ($r \in \{-1, 0\}$) that place some weight on the distribution of surplus across consumers, our case study supports a more favorable posterior belief about personalization due to the allocative effects.

As a concrete example, consider a public policy that might potentially restrict the granularity of data used by the firm for personalized pricing. The posterior belief about the welfare implications of this policy in the context of our case study will be sensitive to the social welfare function adopted. Figure 12 plots the relationship between consumer welfare and the number of features used for personalization for each of the three Atkinson welfare functions, indexed by its respective inequality-aversion

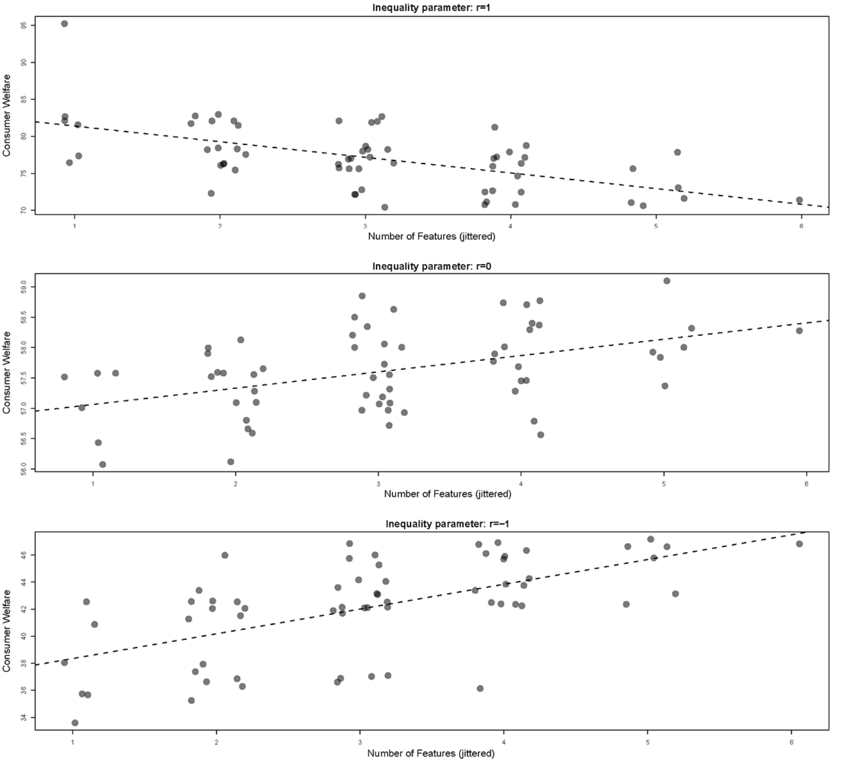


FIG. 12.—Difference in consumer surplus versus number of features used.

parameter $r$. The dotted line in each panel is our regression estimate of the relationship between the level of consumer welfare and the total number of features used for personalization. We can think of the slope of this line as the posterior belief about the welfare effects of data granularity. Once again, we see a stark difference between the inequality-neutral ($r = 1$) and inequality-averse ($r \in \{-1, 0\}$) welfare functions. The former, which considers only total surplus, implies a negative relationship between consumer welfare and the granularity of the targeting data. However, the inequality-averse welfare functions tend to favor personalized pricing and more data granularity. While this analysis does not provide a definitive case for or against the welfare effects of data granularity, it does confirm the need for more academic discourse on how to think about consumer welfare in the context of empirically realistic models of heterogeneous demand when a representative consumer framework is untenable.

In spite of these nuances, we believe our results should challenge the prior that personalized prices are per se bad from a consumer point of a view. To be clear, we are not advocating for personalization as welfare-increasing. Rather, we believe the evidence suggests that data and privacy policies that treat personalized pricing as per se harmful may have unintended consequences and warrant further study.

## VII. Conclusions

A long theoretical literature has studied the welfare implications of monopoly price discrimination. In the digital era, large-scale price discrimination is becoming an empirical reality, raising an important public policy debate about the role of consumer information and its potential impact on consumer well-being. In our case study, we find that personalized pricing using machine learning increases firm profits by over 10% relative to uniform pricing, both in and out of sample, even when we cap the prices at \$499. On the demand side, we find that personalized pricing reduces total consumer surplus. However, we also find that certain data policies that would restrict the use of specific consumer variables for targeting purposes could in fact exacerbate rather than offset the declines in consumer welfare. In our case study, we also find that the majority of consumers would benefit from being charged lower prices than the uniform rate even though total consumer surplus declines. Under standard alternative consumer welfare functions that value the allocation of surplus in addition to the level, we find that the allocative benefits of personalization (through a reduction in inequality) can outweigh the loss in total surplus. These allocative benefits primarily accrue to smaller firms.

The current public policy debate surrounding the fairness of differential pricing might consider the redistributive aspects of personalized pricing in addition to the total surplus implications. In addition, over-regulation of the types of data that firms can use for personalized pricing purposes could exacerbate rather than offset some of the harm to consumers. For instance, we find instances of a nonmonotonic relationship between consumer welfare and the total number of feature variables available for price-targeting purposes.

The results presented herein are based on a single case study of a large digital human resources platform with enterprise consumers. The generalizability of our findings may be limited beyond settings where, like ours, consumers are unlikely to be able to game the personalizing structure. We assume that consumers are unable to misrepresent their types to obtain lower prices (e.g., Acquisti and Varian 2005; Fudenberg and Villas-Boas 2006; Bonatti and Cisternas 2018). Our findings also do not consider the potential role of longer-term consumer backlash based on subjective fairness concerns regarding differential pricing, which could lead to more price-elastic demand in the long run under personalized pricing. This type of backlash might be more problematic in a consumer goods market where personalized pricing may be more transparent and less accepted.[27] Finally, our findings focus on the monopoly price discrimination problem for ZipRecruiter. We do not consider the impact of personalized pricing in a competitive market, where the potential toughening or softening of price competition would also impact the welfare implications.[28]

In addition, our study was conducted in the context of a B2B digital platform selling to enterprise customers. An important direction for future research will be the study of personalized pricing in the context of consumer goods and the welfare implications for consumers with different incomes and socioeconomic status.

## Appendix A

### The Bayesian LASSO

We start with our regularization procedure. Following Tibshirani (1996), suppose that each model parameter, $\Theta_j$, is assigned an i.i.d. Laplace prior with scale $\tau > 0$: $\Theta_j \sim \mathrm{La}(\tau)$, where $\tau = N\lambda$. We can write the posterior distribution of $\Theta$ analytically:

$$F_{\Theta}(\Theta|\mathbf{D}) \propto \ell(\mathbf{D}|\Theta) - \sum_{j=1}^{J} \tau_j |\Theta_j|, \tag{A1}$$

27 Negotiated price deals are quite common in B2B pricing, especially with sales agents.

28 See, e.g., the empirical analysis of competitive geographic price discrimination in Dubé et al. (2017), the theoretical work by Corts (1998), and the literature survey in Stole (2007).

where $\ell(\mathbf{D}|\Theta)$ represents the log likelihood of the demand data as before. This framework is termed the Bayesian LASSO (Park and Casella 2008) on account of the Bayesian interpretation of the LASSO penalized objective function. The maximum a posteriori (MAP) estimator that optimizes (A1) can be shown to be equivalent to the LASSO regression:

$$\Theta^{\text{LASSO}} = \arg\max_{\Theta\in\mathbb{R}^J}\left\{\ell(\mathbf{D}|\Theta) - N\lambda\sum_{j=1}^{J}|\Theta_j|\right\}. \tag{A2}$$

In appendix C, we describe the path-of-one-step estimators procedure used to select $\lambda$ and generate estimates of $\Theta$ and its sparsity structure (see also Taddy 2015b).

## Appendix B

### The Weighted Likelihood Bootstrap

While the MAP estimator generates a point estimate of the posterior mode, it does not offer a simple way to calibrate the uncertainty in these estimates. Park and Casella (2008) propose a Gibbs sampler for a fully Bayesian implementation of the LASSO, but the approach would not scale well to settings with very large-dimensional $x_i$.[29] Instead, we simulate the approximate posterior using a WLB of the LASSO problem. The WLB (Newton and Raftery 1994) is an extension of the Bayesian bootstrap originally proposed by Rubin (1981).[30] As discussed in Efron (2012), the Bayesian bootstrap and the WLB are computationally simple alternatives to MCMC approaches. In our context, the approach is scalable to settings with a large-dimensional parameter space and is relatively fast, making consumer classification and price discrimination practical to implement in real time. Conceptually, the approach consists of drawing weights associated with the observed data sample and solving a weighted version of (A2). The application of LASSO to each replication ensures a sparsity structure that facilitates the storage of the draws in memory. This is a promising approach to approximating uncertainty in complex econometric models (see, e.g., Chamberlain and Imbens 2003).

We construct a novel WLB type procedure to derive the posterior distribution of $\hat{\Theta}|_{\lambda^*}$, $F(\Theta)$. Consider our data sample $\mathbf{D} = (D_1, \ldots, D_N)$. We assume that the data-generating process for $\mathbf{D}$ is discrete with support points $(\zeta_1, \ldots, \zeta_L)$ and corresponding probabilities $\phi = (\phi_1, \ldots, \phi_L):\ \Pr(D_i = \zeta_l) = \phi_l$. We can allow $L$ to

[29] Challenges include drawing from a large-dimensional distribution, assessing convergence of the Markov chain Monte Carlo (MCMC) sampler, tuning the algorithm, and storing a nonsparse simulated chain in memory.

[30] To be clear, our implementation uses only the first stage of the WLB procedure described in Newton and Raftery (1994) and does not implement the sampling-importance-resampling stage. Newton and Raftery (1994) show that the first stage is sufficient to obtain a first-order approximation of the posterior. We could also describe our implementation simply as a variant of the Bayesian bootstrap, but we chose to call it the WLB to acknowledge the contribution of Newton and Raftery (1994), who first outlined the possibility of recasting the Rubin (1981) framework of using the weighted likelihoods.

be arbitrarily large to allow for flexibility in this representation. We assume the following Dirichlet prior on the probabilities

$$\phi \sim \mathrm{Dir}(\mathbf{a}) \propto \prod_{l=1}^{L} \phi_l^{a_l - 1},\ a_l > 0.$$

Following the convention in the literature, we use the improper prior distribution with $a_l \rightarrow 0$. This assumption implies that any support points, $\zeta_l$, not observed in the data will have $\phi_l = 0$ with posterior probability one: $\Pr(\phi_l = 0) = 1, \forall\ \zeta_l \notin \mathbf{D}$. This prior is equivalent to using the following independent exponential prior: $V_l \sim \exp(1)$, where $V_l = \Sigma_{k=1}^{L} \phi_k \phi_l$.

We can now write the posterior distribution of observing a given data point, $D$, as follows:

$$f(D) = \sum_{i=1}^{N} V_i \mathbf{1}_{\{D=\zeta_i\}},\ V_i \sim \text{i.i.d. } \exp(1).$$

The algorithm is implemented as follows. For each of the bootstrap replications $b = 1, \ldots, B$:

1. Draw weights: $\{V_i^b\}_{i=1}^{N} \sim \exp(\mathbf{1}_N)$.
2. Run the LASSO

$$\hat{\Theta}^b|_\lambda = \underset{\Theta \in \mathbb{R}^J}{\arg\min} \left\{ \ell^b(\Theta) + N\lambda \sum_{j=1}^{J} |\Theta_j| \right\},$$

where $\ell^b(\mathbf{D}|\Theta) = \Sigma_{i=1}^{N} V_i^b \ell(D_i|\Theta)$, using the algorithm (C2) in appendix C.

2a. Construct the regularization path, $\{\hat{\Theta}^b|_\lambda\}_{\lambda=\lambda_1}^{\lambda_T}$.

2b. Use $k$-fold cross validation to determine the optimal penalty, $\lambda^*$.

3. Retain $\hat{\Theta}^b \equiv \hat{\Theta}^b|_{\lambda^{b*}}$.

We can then use the bootstrap draws, $\{\hat{\Theta}^b\}_{b=1}^{B}$, to simulate the posterior of interest, $F_\Psi(\Psi_i)$. We construct draws $\{\Psi_i^b\}_{b=1}^{B}$, where $\Psi_i^b = \Psi(x_i; \Theta^b)$, which can be used to simulate the posterior $F_\Psi(\Psi_i)$. We will use this sample to quantify the uncertainty associated with various functions of $\Psi_i$, such as profits and demand elasticities.

## Appendix C

### LASSO Regression

The penalized LASSO estimator solves for

$$\hat{\Theta}|_\lambda = \underset{\Theta \in \mathbb{R}^J}{\arg\min} \left\{ \ell(\Theta) + N\lambda \sum_{j=1}^{J} |\Theta_j| \right\}, \tag{C1}$$

where $\lambda > 0$ controls the overall penalty and $|\Theta_j|$ is the $L_1$ coefficient cost function. Note that as $\lambda \rightarrow 0$, we approach the standard maximum likelihood estimator. For

$\lambda > 0$, we derive simpler "regularized" models with low (or zero) weight assigned to many of the coefficients. Since the ideal $\lambda$ is unknown a priori, we derive a regularization path, $\{\hat{\Theta}|_{\lambda}\}_{\lambda=\lambda_1}^{\lambda_T}$, consisting of a sequence of estimates of $\Theta$ corresponding to successively lower degrees of penalization. Following Taddy (2015b), we use the following algorithm to construct the path:

1. $\lambda_1 = \inf\{\lambda : \hat{\Theta}|_{\lambda_1} = 0\}$;
2. set step size of $\delta \in (0, 1)$;
3. for $t = 2, \dots, T$:

$$
\begin{aligned}
\lambda^t &= \delta\lambda^{t-1}, \\
\omega_j^t &= \left(\left|\Theta_j^{t-1}\right|\right)^{-1}, j \in \hat{S}_t, \\
\widehat{\Theta}^t &= \underset{\Theta \in \mathbb{R}^J}{\arg\min}\left\{\ell(\Theta) + N\textstyle\sum_{j=1}^J \lambda^t \omega_j^t \left|\Theta_j\right|\right\}.
\end{aligned}
\tag{C2}
$$

The algorithm produces a weighted-$L_1$ regularization, with weights $\omega_j$. The concavity ensures that the weight on the penalty on $\hat{\Theta}_j^t$ falls with the magnitude of $|\hat{\Theta}_j^t|$. As a result, coefficients with large values earlier in the path will be less biased toward zero later in the path. This bias diminishes faster with larger values of $\gamma$.

The algorithm in (C2) generates a path of estimates corresponding to different levels of penalization, $\lambda$. We use $k$-fold cross validation to select the "optimal" penalty, $\lambda^*$. We implement the approach using the cv.gamlr function from the gamlr package in R.

## Appendix D

### The Deep-Learning Framework

The deep-learning estimator follows the WLB procedure for the LASSO described in section III.B. The key point of departure is in the definition of the loss function. We use $\ell_{\text{DNN}}^b(\Theta)$ to denote the logit loss function, and we approximate the structural parameters $\{\alpha(x), \beta(x)\}$ with DNNs. The architecture of the DNNs are chosen to match the complexity levels to the data (for a more in-depth discussion, see Farrell, Liang, and Misra 2021a, 2021b). For the analysis presented herein, we used two specifications of the network with two and three hidden layers of 10 nodes each. The total number of parameters in each of these models is 1,440 and 2,340, respectively. The minimization procedure was coded in Tensorflow (Abadi et al. 2015) and used the default specification of the Adam learning algorithm (Kingma and Ba 2015).

The algorithm is implemented as follows. For each of the bootstrap replications $b = 1, \dots, B$:

1. draw weights: $\{V_i^b\}_{i=1}^N \sim \exp(\mathbf{1}_N)$;

2. run the DNN learning algorithm to obtain

$$\hat{\Theta}^b_{\mathrm{DNN}} = \underset{\Theta \in \mathbb{R}^J}{\arg\min}\{\ell^b_{\mathrm{DNN}}(\Theta)\},$$

where $\ell^b_{\mathrm{DNN}}(\boldsymbol{D}|\Theta) = \Sigma^N_{i=1} V^b_i \ell_{\mathrm{DNN}}(D_i|\Theta)$;

3. retain $\hat{\Theta}^b_{\mathrm{DNN}}$.

As before, we use the bootstrap draws, $\{\hat{\Theta}^b_{\mathrm{DNN}}\}^B_{b=1}$, to simulate the posterior of interest, $F_\Psi(\Psi_i)$. We construct draws $\{\Psi^b_i\}^B_{b=1}$, where $\Psi^b_i = \Psi(x_i; \Theta^b)$, which can be used to simulate the posterior $F_\Psi(\Psi_i)$.

## Appendix E

### Perfect Price Discrimination

Suppose that the firm observed not only the full feature set for a consumer $i$, $X_i$, but also the random utility shock, $\epsilon_i$. Under perfect price discrimination, the firm would set the personalized price

$$p^{\mathrm{PD}}_i = \max(\mathrm{WTP}_i, 0)$$

where $\mathrm{WTP}_i$ represents consumer $i$'s maximum willingness to pay (WTP)

$$\mathrm{WTP}_i = \frac{(\alpha(X_i) + \epsilon_i)}{\beta(X_i)}. \tag{E1}$$

Customer $i$ would deterministically buy as long as $\mathrm{WTP}_i \geq 0$.

Accounting for the fact that the researcher (unlike the firm in this case) does not observe $\epsilon$, the expected probability that a consumer with preferences $(\alpha, \beta)$ would purchase at the perfect price discrimination price is

$$\begin{aligned} \mathbb{P}(p^{\mathrm{PD}}; X_i, \Theta) &= \Pr(\mathrm{WTP} \geq 0) \\ &= 1 - \frac{1}{1 + \exp(\alpha)}. \end{aligned} \tag{E2}$$

The corresponding expected profit from this consumer is

$$\pi(p^{\mathrm{PD}}|\alpha, \beta) = E(\mathrm{WTP}|\mathrm{WTP} \geq 0, \alpha, \beta) \Pr(\mathrm{buy}|p = p^{\mathrm{PD}}\alpha, \beta), \tag{E3}$$

where

$$E(\mathrm{WTP}|\mathrm{WTP} > 0, \alpha, \beta) = \frac{\alpha}{\beta} + \frac{1}{\beta}\left(-\alpha + \frac{[1 + \exp(\alpha)] \ln[1 + \exp(\alpha)]}{\exp(\alpha)}\right). \tag{E4}$$

We now derive the result in (E4). Recall that the random utility shock is assumed to be i.i.d. logistic with probability density function (PDF)

$$f(\Delta\epsilon) = \frac{\exp(-\Delta\epsilon)}{[1 + \exp(-\Delta\epsilon)]^2}$$

and CDF

$$F(\Delta\epsilon) = \frac{1}{1 + \exp(-\Delta\epsilon)}.$$

The truncated PDF for $\Delta\epsilon$ when it is known to be strictly greater than $k > 0$ is

$$f(\Delta\epsilon|\Delta\epsilon \geq k) = \frac{f(\Delta\epsilon)}{\Pr(\Delta\epsilon \geq k)} = \left[\frac{\exp(-k)}{1 + \exp(-k)}\right]^{-1} \frac{\exp(-\Delta\epsilon)}{[1 + \exp(-\Delta\epsilon)]^2}.$$

We can then compute the conditional expectation of the truncated random variable $\Delta\epsilon$ when $k > 0$ as follows:

$$\begin{aligned}
E(\Delta\epsilon|\Delta\epsilon \geq k) &= [\Pr(\Delta\epsilon \geq k)]^{-1} \int_k^{-\infty} \Delta\epsilon f(\Delta\epsilon)\, d\Delta\epsilon \\
&= \left[\frac{\exp(-k)}{1 + \exp(-k)}\right]^{-1} \int_k^{-\infty} \Delta\epsilon \frac{\exp(-\Delta\epsilon)}{[1 + \exp(-\Delta\epsilon)]^2}\, d\Delta\epsilon \\
&= \left[\frac{1 + \exp(-k)}{\exp(-k)}\right] \left[\frac{k \exp(-k) + [1 + \exp(-k)] \ln[1 + \exp(-k)]}{1 + \exp(-k)}\right] \\
&= k + \frac{[1 + \exp(-k)] \ln[1 + \exp(-k)]}{\exp(-k)},
\end{aligned}$$

where

$$\begin{aligned}
&\Delta\epsilon \frac{\exp(-\Delta\epsilon)}{[1 + \exp(-\Delta\epsilon)]^2} \\
&\qquad = \frac{d(-((\Delta\epsilon e(-\Delta\epsilon) + [1 + e(-\Delta\epsilon)] \ln[1 + e(-\Delta\epsilon)])/([1 + e(-\Delta\epsilon)])))}{d\Delta\epsilon}.
\end{aligned}$$

## References


Abadi, M., A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, et al. 2015. "TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems." http://tensorflow.org.

Acquisti, A., C. Taylor, and L. Wagman. 2016. "The Economics of Privacy." *J. Econ. Literature* 54 (2): 442–92.

Acquisti, A., and H. R. Varian. 2005. "Conditioning Prices on Purchase History." *Marketing Sci.* 24 (3): 367–81.

Aguirre, I., S. Cowan, and J. Vickers. 2010. "Monopoly Price Discrimination and Demand Curvature." *A.E.R.* 100 (4): 1601–15.

Ansari, A., and C. F. Mela. 2003. "E-Customization." *J. Marketing Res.* 40 (2): 131–45.

Athey, S., and G. W. Imbens. 2016a. "Estimation and Inference of Heterogeneous Treatment Effects using Random Forests." *J. American Statis. Assoc.* 113 (523): 1228–42.

———. 2016b. "Recursive Partitioning for Heterogeneous Causal Effects." *Proc. Nat. Acad. Sci. USA* 113 (27): 7353–60.

Atkinson, A. B. 1970. "On the Measurement of Inequality." *J. Econ. Theory* 2:244–63.

Bauner, C. 2015. "Mechanism Choice and the Buy-It-Now Auction: A Structural Model of Competing Buyers and Sellers." *Internat. J. Indus. Org.* 38:19–31.

Bergemann, D., B. Brooks, and S. Morris. 2015. "The Limits of Price Discrimination." *A.E.R.* 105 (3): 921–57.

Berger, J. 1985. *Statistical Decision Theory and Bayesian Analysis*. New York: Springer.

Berry, S. T. 1994. "Estimating Discrete-Choice Models of Product Differentiation." *RAND J. Econ.* 25 (2): 242–62.

Bissiri, P. G., C. C. Holmes, and S. G. Walker. 2016. "A General Framework for Updating Belief Distributions." *J. Royal Statis. Soc. B* 78 (5): 1103–30.

Bonatti, A., and G. Cisternas. 2018. "Consumer Scores and Price Discrimination." Working paper, New York Univ.

Bradlow, E., P. Lenk, G. Allenby, and P. E. Rossi. 2004. "When BDT in Marketing Meant Bayesian Decision Theory: The Influence of Paul Green's Research." In *Marketing Research and Modeling: Progress and Prospects*, edited by Yoram Wind and Paul E. Green, 17–42. Norwell, MA: Kluwer.

CEA (Council of Economic Advisors). 2015. "Big Data and Differential Pricing." Report, Council Econ. Advisors, Washington, DC.

Chamberlain, G., and G. W. Imbens. 2003. "Nonparametric Applications of Bayesian Inference." *J. Bus. and Econ. Statis.* 21 (1): 12–18.

Chen, Y., X. Li, and M. Sun. 2015. "Competitive Mobile Targeting." Working paper.

Chintagunta, P., J.-P. Dubé, and K.-Y. Goh. 2005. "Beyond the Endogeneity Bias: The Effect of Unmeasured Brand Characteristics on Household-Level Brand Choice Models." *Management Sci.* 51:832–49.

Cohen, P., R. Hahn, J. Hall, S. Levitt, and R. Metcalfe. 2016. "Using Big Data to Estimate Consumer Surplus: The Case of Uber." Working Paper no. 22627, NBER, Cambridge, MA.

Corts, K. S. 1998. "Third-Degree Price Discrimination in Oligopoly: All-Out Competition and Strategic Commitment." *RAND J. Econ.* 29:306–23.

Cowan, S. 2012. "Third-Degree Price Discrimination and Consumer Surplus." *J. Indus. Econ.* LX (2): 333–45.

Dong, X., P. Manchanda, and P. K. Chintagunta. 2009. "Quantifying the Benefits of Individual-Level Targeting in the Presence of Firm Strategic Behavior." *J. Marketing Res.* XLVI:207–21.

Dubé, J.-P., Z. Fang, N. M. Fong, and X. Luo. 2017. "Competitive Price Targeting with Smartphone Coupons." *Marketing Sci.* 36 (6): 944–75.

Dubé, J.-P., G. J. Hitsch, and P. E. Rossi. 2010. "State Dependence and Alternative Explanations for Consumer Inertia." *RAND J. Econ.* 41 (3): 417–45.

Efron, B. 2012. "Bayesian Inference and the Parametric Bootstrap." *Ann. Appl. Statis.* 6:1971–97.

Einav, L., C. Farronato, J. Levin, and N. Sundaresan. 2018. "Auctions versus Posted Prices in Online Markets." *J.P.E.* 126 (1): 178–215.

Einav, L., and J. Levin. 2010. "Empirical Industrial Organization: A Progress Report." *J. Econ. Perspectives* 24 (2): 145–62.

———. 2014. "Economics in the Age of Big Data." *Science* 346 (6210). https://www.science.org/doi/10.1126/science.1243089.

Fan, J., Y. Feng, and R. Song. 2012. "Nonparametric Independence Screening in Sparse Ultra-High-Dimensional Additive Models." *J. American Statis. Assoc.* 106 (494): 544–57.

Fan, J., and R. Li. 2001. "Variable Selection via Nonconcave Penalized Likelihood and Its Oracle Properties." *J. American Statis. Assoc.* 96 (456): 1348–60.

Farrell, M. H., T. Liang, and S. Misra. 2021a. "Deep Learning for Individual Heterogeneity." Working paper.

———. 2021b. "Deep Neural Networks for Estimation and Inference." *Econometrica* 89 (1): 181–213.

Foley, D. 1967. "Resource Allocation and the Public Sector." PhD diss., Yale Univ.

Fudenberg, D., and J. M. Villas-Boas. 2006. "Behavior-Based Price Discrimination and Customer Recognition." In *Handbooks in Information Systems: Economics and Information Systems*, edited by T. Hendershott, 377–436. New York: Elsevier.

Granville, K. 2018. "Facebook and Cambridge Analytica: What You Need to Know as Fallout Widens." *New York Times*, March 19.

Green, P. E. 1963. "Bayesian Decision Theory in Pricing Strategy." *J. Marketing* 27 (1): 5–14.

Green, P. E., and R. E. Frank. 1966. "Bayesian Statistics and Marketing Research." *J. Royal Statis. Soc. B* 15 (3): 173–90.

Hastie, T., R. Tibshirani, and J. Friedman. 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed. New York: Springer.

Hirano, K. 2008. "Decision Theory in Econometrics." In *The New Palgrave Dictionary of Economics*, 2nd ed., edited by S. Durlauf and L. E. Blume, 2636–41. London: Palgrave Macmillan.

Hitsch, G. J. 2006. "An Empirical Model of Optimal Dynamic Product Launch and Exit under Demand Uncertainty." *Marketing Sci.* 25 (1): 25–50.

Jorgenson, D. W. 1990. "Aggregate Consumer Behavior and the Measurement of Social Welfare." *Econometrica* 58 (5): 1007–40.

Jorgenson, D. W., and D. T. Slesnick. 2014. "Measuring Social Welfare in the U.S. National Accounts." In *Measuring Economic Stability and Progress*, edited by D. W. Jorgenson, J. S. Landefeld, and P. Schreyer, 43–88. Chicago: Univ. Chicago Press.

Kehoe, P. J., B. J. Larsen, and E. Pastorino. 2020. "Dynamic Competition in the Era of Big Data." Working paper, Stanford Univ., Stanford, CA.

Kingma, D. P., and J. Ba. 2015. "Adam: A Method for Stochastic Optimization." Paper presented at the 2015 International Conference on Learning Representations (ICLR), San Diego, CA, May 7–9.

Krugman, P. 2000. "Reckonings; What Price Fairness?" *New York Times*, October 4.

Kumar, V., S. Sriram, A. Luo, and P. Chintagunta. 2011. "Assessing the Effect of Marketing Investments in a Business Marketing Context." *Marketing Sci.* 30 (5): 924–40.

Leeb, H., and B. M. Potscher. 2008. "Sparse Estimators and the Oracle Property, or the Return of Hodges Estimator." *J. Econometrics* 142:201–11.

Levitt, S. D., and J. A. List. 2009. "Field Experiments in Economics: The Past, the Present, and the Future." *European Econ. Rev.* 53 (1): 1–18.

Levitt, S. D., J. A. List, S. Neckermann, and D. Nelson. 2016. "Quantity Discounts on a Virtual Good: The Results of a Massive Pricing Experiment at King Digital Entertainment." *Proc. Nat. Acad. Sci. USA* 113 (27): 7323–28.

Lewbel, A., and K. Pendakur. 2017. "Unobserved Preference Heterogeneity in Demand Using Generalized Random Coefficients." *J.P.E.* 125 (4): 1100–148.

List, J. A. 2004. "The Nature and Extent of Discrimination in the Marketplace: Evidence from the Field." *Q.J.E.* 119 (1): 49–89.

Lyddon, S. P., and C. C. Holmes. 2019. "General Bayesian Updating and the Loss-Likelihood Bootstrap." *Biometrika* 106 (2): 465–78.

Maniadis, Z., F. Tufano, and J. A. List. 2014. "One Swallow Doesn't Make a Summer: New Evidence on Anchoring Effects." *A.E.R.* 104 (1): 277–90.

Misra, S., and H. S. Nair. 2011. "A Structural Model of Sales-Force Compensation Dynamics: Estimation and Field Implementation." *Quantitative Marketing and Econ.* 9:211–57.

Mohammed, R. 2017. "How Retailers Use Personalized Prices to Test What You're Willing to Pay." *Harvard Bus. Rev.*, October 20.

Mussa, M., and S. Rosen. 1978. "Monopoly and Product Quality." *J. Econ. Theory* 18:301–17.

Newton, M. A., and A. E. Raftery. 1994. "Approximate Bayesian Inference with the Weighted Likelihood Bootstrap." *J. Royal Statis. Soc. B* 56 (1): 3–48.

Ostrovsky, M., and M. Schwarz. 2016. "Reserve Prices in Internet Advertising Auctions: A Field Experiment." Working paper.

Park, T., and G. Casella. 2008. "The Bayesian Lasso." *J. American Statis. Assoc.* 103:681–86.

Pigou, A. 1920. *The Economics of Welfare*. London: Macmillan.

Ramasastry, A. 2005. "Websites That Charge Different Customers Different Prices: Is Their 'Price Customization' Illegal? Should It Be?" *FindLaw*, June 20.

Rossi, P. E., R. E. McCulloch, and G. M. Allenby. 1996. "The Value of Purchase History Data in Target Marketing." *Marketing Sci.* 15 (4): 321–40.

Rubin, D. B. 1981. "The Bayesian Bootstrap." *Ann. Statis.* 9:130–34.

Savage, L. 1954. *Statistical Decision Functions*. New York: Wiley.

Schwarz, G. E. 1978. "Estimating the Dimension of a Model." *Ann. Statis.* 6 (2): 461–64.

Shapiro, C., and H. R. Varian. 1999. *Information Rules*. Boston: Harvard Bus. School Press.

Shiller, B. 2015. "First-Degree Price Discrimination Using Big Data." Working paper.

Shiller, B., and J. Waldfogel. 2011. "Music for a Song: An Empirical Look at Uniform Song Pricing and Its Alternatives." *J. Indus. Econ.* 59 (4): 630–60.

Simester, D. I., P. Sun, and J. N. Tsitsiklis. 2006. "Dynamic Catalog Mailing Policies." *Management Sci.* 52 (5): 683–96.

Smith, M., J. Bailey, and E. Brynjolfsson. 2000. *Understanding Digital Markets*. Cambridge, MA: MIT Press.

Stole, L. A. 2007. "Price Discrimination and Imperfect Competition." In *Handbook of Industrial Organization*, vol. 3, edited by M. Armstrong and R. H. Porter, 2221–99. Amsterdam: North-Holland.

Taddy, M. 2015a. "Distributed Multinomial Regression." *Ann. Appl. Statis.* 9:1394–414.

———. 2015b. "One-Step Estimator Paths for Concave Regularization." Working paper, Univ. Chicago Booth School Bus.

Thomson, W. 2011. "Fair Allocation Rules." In *Handbook of Social Choice and Welfare*, vol. 2, edited by K. J. Arrow, A. Sen, and K. Suzumura, 393–506. Amsterdam: North-Holland.

Tibshirani, R. 1996. "Regression Shrinkage and Selection via the Lasso." *J. Royal Statis. Soc. B* 58:267–88.

Tirole, J. 1988. *The Theory of Industrial Organization*. Cambridge, MA: MIT Press.

Turow, J., L. Feldman, and K. Meltzer. 2005. "Open to Exploitation: America's Shoppers Online and Offline." Working paper, Annenberg School Communication, Philadelphia.

Useem, J. 2017. "How Online Shopping Makes Suckers of Us All." *Atlantic*, May 15.

Varian, H. R. 1980. "A Model of Sales." *A.E.R.* 70 (4): 651–59.

———. 1989. "Price Discrimination." In *Handbook of Industrial Organization*, vol. 1, edited by R. Schmalensee and R. Willig, 597–654. Amsterdam: North-Holland.
Verboven, F. 2008. "Price Discrimination (Empirical Studies)." In *The New Palgrave Dictionary of Economics*, edited by S. Durlauf and L. E. Blume, 10683–87. London: Palgrave Macmillan.
Wald, A. 1950. *Statistical Decision Functions.* New York: Wiley.
Waldfogel, J. 2015. "First Degree Price Discrimination Goes to School." *J. Indus. Econ.* 63 (4): 569–97.
Wolverton, T. 2002. "MP3 Player Sale Exposes Amazon's Flexible Prices." CNET, January 2. https://www.cnet.com/news/mp3-player-sale-exposes-amazons-flexible-prices.
Zhang, J., O. Netzer, and A. Ansari. 2014. "Dynamic Targeted Pricing in B2B Relationships." *Marketing Sci.* 33 (3): 317–37.
Zou, H. 2006. "The Adaptive Lasso and Its Oracle Properties." *J. American Statis. Assoc.* 101 (476): 1419–29.
Zou, H., T. Hastie, and R. Tibshirani. 2007. "On the Degrees of Freedom of the Lasso." *Ann. Statis.* 35 (5): 2173–92.