Personalized Pricing and Customer Welfare [1]

Jean-Pierre Dubé, *Chicago Booth and NBER*

Sanjog Misra, *Chicago Booth*

May 10, 2019

Abstract

We study the welfare implications of personalized pricing, an extreme form of third-degree price discrimination implemented with machine learning for a large, digital firm. We conduct a randomized controlled pricing field experiment to train a demand model and to conduct inferences about the effects of personalized pricing on firm and customer surplus. In a second experiment, we validate our predictions out of sample. Personalized pricing improves the firm's expected posterior profits by 19%, relative to optimized uniform pricing, and by 86%, relative to the firm's status quo pricing. On the demand side, customer surplus declines slightly under personalized pricing relative to a uniform pricing structure. However, over 60% of customers benefit from personalized prices that are lower than the optimal uniform price. Based on simulations with our demand estimates, we find several cases where customer surplus increases when the firm is allowed to condition on more customer features and classify customers into more granular segments. These findings indicate a need for caution in the current public policy debate regarding data privacy and personalized pricing. Some data restrictions could harm consumers and even reduce total consumer welfare.

Keywords: price discrimination, welfare, field experiment, personalized pricing, targeted marketing, Lasso regression, weighted likelihood bootstrap

1 Introduction

The vast quantities of personal data accessible online have enormous potential economic value. These data represent formidable business assets when firms use them to target decisions, like advertising and pricing, differentially across individuals. Recent events, such as the controversy over Cambridge Analytica's alleged misuse of user data on Facebook (Granville, 2018), the adoption of General Data Protection Regulation (hereafter GDPR) in the EU 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 customer-specific prices using a large number of observable customer features[1]. Prices are set differentially across each combination of observed customer features to capture surplus. The application of 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, p. 3)[2]. 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, page 12), the report mostly drew dire conclusions about the potential harm to customers:

> "[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, page 6).

Similar concerns about the harmful effects of differential pricing have been echoed in the recent mainstream business media (e.g., Useem, 2017; Mohammed, October 20, 2017), leading experts to question the fairness and even legality of these practices (e.g., Krugman, October 4, 2000; Ramasastry, June 20, 2005; Turow, Feldman, and Meltzer, 2005). While the CEA report does not specifically recommend new legislation to regulate differential pricing, privacy legislation, such as the GDPR, will require firms to disclose their usage of customer data "in a concise, transparent, intelligible and easily accessible form,

---

[1]In practice, third-degree price discrimination has typically been based on very coarse segmentation structures that vary prices across broad groups of customers 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).

[2]Even large, digitally-enabled firms like Amazon have committed to an explicit, non-discriminatory pricing policy (Wolverton, 2010).

using clear and plain language.[3]" The GDPR may also require customers 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 are also found in the recent California Consumer Privacy Act.

A potential concern is that hasty, over-regulation of data-based price discrimination could in fact have the unintended consequence of reducing social welfare and, more specifically, harming consumers. 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 increases 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 has an element of classification error and can potentially increase consumer surplus (relative to 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 wealth" has been identified as a fruitful direction for future research in the economics of privacy (Acquisti, Taylor, and Wagman, 2016, page 481).

To analyze the welfare implications of personalized pricing, we conduct an empirical case study in cooperation with a large digital firm. The heart of our analysis consists of a sequence of novel, randomized business-to-business price experiments for new customers. In the first experiment, we randomize the quoted monthly price of service to new customers and use the data to train a demand model with heterogeneous price treatment effects. We assume that the heterogeneity in customers' price sensitivities can be characterized by a sparse subset of an observed, high-dimensional vector of observable customer 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 customers, 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[4]).

To the best of our knowledge, this study is the first to document both the feasibility and implications of scalable, personalized pricing with Big Data. In this regard, we add to a small and growing literature using firm-sanctioned field experiments to obtain plausible estimate of the treatment effect of marketing

---

[3]Article 12, EU GDPR, "Transparent information, communication and modalities for the exercise of the rights of the data subject".

[4]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

variables on demand (e.g., Levitt and List, 2009; Einav and Levin, 2010) [5]. 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).

Our demand estimates 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 customer variables (Rossi, McCulloch, and Allenby, 1996; Shiller and Waldfogel, 2011; Shiller, 2015).

On the demand side, we predict that total customer surplus would fall under decision-theoretic personalized pricing. But, for our case study, personalization is still far removed from the purely theoretical case of perfect price discrimination which would transfer all the customer surplus to the firm. In our experiments, a majority of customers benefit from price discrimination policies. In our validation field experiment, nearly 70% of the customers assigned to the personalized pricing cell are targeted a personalized price that is below the optimal uniform price. Although our experiments are not designed to identify the causal effect of specific individual customer 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.

We also use our model estimates to explore the role of the granularity of customer information on surplus. We explore several alternative personalization schemes that restrict the types of customer features on which the firm is allowed to condition to construct segments and set differential prices. Consistent with Bergemann, Brooks, and Morris (2015), we observe interesting non-monotonicities in customer surplus and the quantity of customer data available to the firm for personalization. In several cases, we find that customer surplus can increase when the firm has access to more customer features and can classify customers into more granular segments. However, none of these restricted personalization schemes generates more customer surplus than the optimal uniform pricing structure.

Our findings relate to the concept of fairness in the social choice literature and add to the on-going 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 customers and lower customer surplus, highlighting the potential trade-offs between fairness and customer welfare. Moreover, in our case study the

---

[5]See also Cohen, Hahn, Hall, Levitt, and Metcalfe (2016) for a quasi price experiment based on Uber surge.

typical strong customer tends to be a larger company with 20 employees (relative to 10 employees for weak customers), suggesting that our personalization scheme redistributes surplus from larger to smaller customers. 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.

Our findings contribute to the empirical literature on third-degree price discrimination (see the survey by Verboven 2008). By running a price experiment, we avoid the typical price endogeneity concerns associated with demand estimation. In the domain of digital marketing, Bauner (2015) and Einav, Farronato, Levin, and Sundaresan (2017) argue that the co-existence of auctions and posted price formats on eBay may be price discriminating between customer segments. Einav, Farronato, Levin, and Sundaresan (2017) 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." Surprisingly, in a large-scale randomized price experiment for an online gaming company that uses almost uniform pricing, Levitt, List, Neckermann, and Nelson (2016) find almost no effect on revenues from various alternative second-degree "non-linear" price discrimination policies. However, they document substantial heterogeneity across consumers which suggests potential gains from the type of third-degree "personalized pricing" studied herein.

Our work also contributes to the broad empirical literature on the targeting of marketing actions across customers (e.g., Ansari and Mela, 2003; Simester, Sun, and Tsitsiklis, 2006; Dong, Manchanda, and Chintagunta, 2009; Kumar, Sriram, Luo, and Chintagunta, 2011). A small subset of this literature has analyzed personalized pricing with different prices charged to each customer (e.g. Rossi, McCulloch, and Allenby 1996; Chintagunta, Dubé, and Goh 2005; Zhang, Netzer, and Ansari 2014; Waldfogel 2015; Shiller 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 customer purchase histories to determine personalized prices. These studies report large predicted profit improvements for firms. 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 our model predictions for the impact on customers and the firm. The extant work's findings and methods also have limited applicability beyond markets for fast-moving consumer goods due to the limited availability of customer purchase panels in most markets. In contrast, we devise a more broadly practical targeting scheme based on observable customer features and cross-sectional data.

The extant literature suggests that basing personalized prices on observable customer 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 membership as a proxy 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 customer features could improve firm profits substantially, supporting the view that such practices could become more commonplace.

Finally, our work is also related to the recent literature conducting inference when machine learning algorithms are used to analyze heterogeneous treatment effects (e.g., Wager and Athey, 2015; Chernozhukov, Chetverikov, Demirer, Duflo, Hansen, Newey, and Robins, 2016; Taddy, Gardner, Chen, and Draper, 2016; Athey and Imbens, 2016a). The extant literature has developed procedures for inference in the context of discrete (typically binary) treatment effects. In contrast, we conducting inference over the heterogeneous effects of price, a continuous treatment, on customer demands.

The remainder of the paper is organized as follows. In section **??**, we set up the prototypical decision-theoretic formulation of monopoly price personalization based on demand estimation. In section 3, we derive our empirical approach for estimating the demand parameters and quantifying uncertainty. We summarize our empirical case study of targeted pricing at Ziprecruiter.com in section 4. We conclude in section 6.

2   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 and also see Hirano 2008 for a short overview along with Green and Frank 1966 and Bradlow, Lenk, Allenby, and Rossi 2004 for a discussion of Bayesian decision theory for marketing problems). The firm trades off the opportunity costs from sub-optimal 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 customer's preferences and demand. Bayes theorem provides the most appropriate manner for the firm to use available data to update its beliefs about customers 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 short-comings 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); Dubé, Fang, Fong,

and Luo (2017)[6].

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

2.1 Demand

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

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

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

2.2 Firm Beliefs and Pricing

In this section, we 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 Ψ_i described by the density $f_\Psi(\Psi_i)$. Let D denote the customer database collected by the firm. We assume the firm uses Bayes Rule to construct the data-based posterior belief about the customer's type:

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

where $\ell(D|\Psi_i)$ is the log-likelihood induced by the demand model, 1 and the uncertainty in the random disturbances, ε_i. Let $F_\Psi(\Psi_i|D)$ denote the corresponding CDF of the posterior beliefs. Note that we assume the firm does not update its beliefs $F_\varepsilon(\varepsilon)$ about the random disturbances, ε_i.

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

---

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

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

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

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

where $\nabla_p \bar{q}(p_i^*; \Psi_i, \varepsilon)$ is the matrix of derivatives of consumer $i's$ demand with respect to prices.

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[7]:

> "... 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, Part II, chapter XVI, section 14)

In fact, the personalized pricing in (4) technically constitutes a form of *third-degree price discrimination* (e.g. Tirole, 1988; Pigou, 1920). In our model, the firm can never learn ε_i even with repeated observations on the same customer (i.e. panel data). Therefore it will never be possible for the firm to extract all of the customer surplus even when all the uncertainty in Ψ_i is resolved. In practice, the prices are not fully personalized since customers with the same posterior expected Ψ_i would always be charged the same price even if they differ along unobserved dimensions.

2.3 Welfare

At the heart of the public policy debate is a wide-spread belief that machine learning and databased 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 customer the same uniform price in **??**: $p_i^* = p^*$, $\forall i$.[8] The predicted impact of personalized prices on consumer surplus is less straightforward. The extant literature has relied on local conditions regarding the curvature of demand and other regularity conditions to determine whether a third-degree price discrimination strategy would increase social surplus (e.g., Varian, 1989) and consumer surplus specifically (e.g., Cowan,

---

[7]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 infra-marginal unit purchased by a customer.

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

7

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." 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 tie-breakers between profit-equivalent choices. In Table 1, we report the results under several information scenarios. Under perfect price discrimination, the firm charges each consumer her valuation, generating \$21 in profits and \$0 consumer welfare. 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.[9] Total welfare however is only \$15 and there is a deadweight loss of \$6.

Now, suppose the firm has a database, **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 it 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 welfare however has increased toy \$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 it 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 allowg for more personalized pricing, consumer surplus increases to \$4. Moreover, total welfare is now \$21, which is equal to total welfare under perfect price discrimination except that some of the value accrues to consumers. Interestingly, there is no deadweight loss in this case.

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 large-scale segmentation schemes.

---

[9]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.

3 Empirical Approach

The execution of the firm's data-based pricing strategies in equations 4 and **??** 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 non-linear panel data methods to estimate $F(\Psi_i|D)$ using repeated purchase observations for each customer panelist (e.g. Rossi, McCulloch, and Allenby 1996; Chintagunta, Dubé, and Goh 2005). In practice, many firms may not have access to panel databases. In many business-to-business and e-commerce settings, for instance, firms are more likely to have access to data for a broad cross-section of customers, but not with repeated observations.[10] We consider a scenario with cross-sectional customer information that includes a detailed set of observable customer features. Our approach consists of using these features to approximate Ψ_i.

3.1 Approximating Individual Types

Suppose we observe data

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

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

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

where Θ_0 is a vector of parameters. Note that for our pricing problem in section 2.2 above, 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:

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

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

We focus on applications where, potentially, $K \gg N$ and Θ_{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 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 customer types and, potentially, personalized prices.

---

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

3.2 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 potentially produce biased estimates due to over-fitting. The literature on regularized regression provides numerous algorithms for parameter selection with a high-dimensional parameter vector, Θ (e.g. Hastie, Tibshirani, and Friedman, 2009). Most of this literature is geared towards 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 customer 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 G.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

$$f(\Theta|\mathbf{D}) \propto \exp(-l(\Theta, \mathbf{D})) h(\Theta) \tag{5}$$

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

$$l(\Theta, \mathbf{D}) = - \left[\sum_{i=1}^{N} \ell(\mathbf{D}_i|\Theta) - \lambda \sum_{j=1}^{J} |\Theta_j| \right] \tag{6}$$

where $\sum_{i=1}^{N} \ell(\mathbf{D}_i|\Theta)$ is the sample log-likelihood induced by the demand model in section 1, and λ 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 the posterior distribution of the model parameters using a weighted likelihood bootstrap algorithm (WLB) as outlined in Newton and Raftery (1994).[11] The approach 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:

---

[11]For a detailed description of our procedure see Appendix B.

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

We show in Appendix B that weights $V_i \sim$ i.i.d. $\text{Exp}(1)$ 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 non-standard implementation. We refer the reader to Bissiri, Holmes, and G.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 model. At the same time, the sampling procedure also accounts for usual parameter uncertainty.

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 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 post-regularization 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.

4   Personalized Pricing at Ziprecruiter.com

Given the dearth of empirical case studies of personalized pricing in practice, we conduct a novel case study. We analyze personalized pricing empirically through a sequence of experiments in collaboration with Ziprecruiter.com. The first experiment uses a sample of prospective, new Ziprecruiter customers to train a demand model with heterogeneous price responses. The second experiment uses a new sample of prospective customers 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 databased price discrimination strategy. Moreover, we want to analyze the implications for customer welfare.

Ziprecruiter.com is an online firm that specializes in matching jobseekers to potential employers. We focus Ziprecruiter's business-to-business decision since they offer their jobseeker services for free

and only charge prospective employers. Hereafter, we refer to prospective employers who could use Ziprecruiter as customers. The firm caters to a variety of potential customers across various industries that can use Ziprecruiter.com to access a stream of resumes 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 customers.

Our analysis focuses on prospective customers who have reached the paywall at Ziprecruiter.com for the first time. Amongst all prospective customers, Ziprecruiter's largest segment consists of the "starters," small firms with typically less than 50 employees, looking to fill between 1 and 3 jobs. Since starters represent nearly 50% of the customer base, we focus our attention on prospective starter firms. Another advantage of focusing on small customers 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 resumes. At the beginning of this project the base rate for a "starter" firm looking for candidates was $99/month.

Each prospective new firm that registers for Ziprecruiter's services navigates a series of pages on the Ziprecruiter.com 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 customer 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 customer reports several characteristics of its business and the specific job posting. Table 3 summarizes the variables we retained for our analysis from the much larger set of registration features[12]. While the set looks small, it generates 133 variables[13]. After completing this registration process, the customer reaches a paywall and receives a price quote. The registration process is used to ensure that Ziprecruiter's matching algorithm connects customers with the most relevant CV's 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 customers strategically mis-report.

4.1 Empirical Model of Demand

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

$$\Delta U_i = \alpha_i + \beta_i p_i + \varepsilon_i$$
$$= \alpha(x_i; \theta_\alpha) + \beta(x_i; \theta_\beta) p_i + \varepsilon_i \tag{7}$$

---

[12]In our personalized pricing application below, we only analyze segmentation schemes based on these features which are voluntarily and knowingly self-reported by customers. We do not use any involuntary information tracked, for instance, through cookies.

[13]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.

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 2, we re-write equation 7 as follows

$$\Delta U_i = \tilde{p}_i' \Psi_i + \varepsilon_i \tag{8}$$

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

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

$$\mathbb{P}(y_i = 1 | p_i; \Psi_i) = \int 1(\Delta U_i > 0) dF_\varepsilon(\varepsilon_i)$$
$$= 1 - F_\varepsilon\left(-\tilde{p}_i' \Psi_i\right)$$

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

For our analysis below, we use a linear specification of the functions α and β

$$\alpha(x_i; \theta_\alpha) = x_i' \theta_\alpha$$
$$\beta(x_i; \theta_\beta) = x_i' \theta_\beta.$$

We also assume that the random utility disturbance ε_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

$$\mathbb{P}(y_i = 1 | p_i; \Psi_i) = \frac{\exp\left(\tilde{p}_i' \Psi_i\right)}{1 + \exp\left(\tilde{p}_i' \Psi_i\right)}. \tag{9}$$

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.

4.2 Experiment One: Demand, Pricing and Customer Welfare

The first experiment was conducted between August 28, 2015 and September 29, 2015. During this period, 7,867 unique prospective customers reached Ziprecruiter's paywall. Each prospective customer was randomly assigned to one of ten experimental pricing cells. The control cell consisted of Ziprecruiter's standard $99 per month price, row one of Table 2. 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 two to ten of Table 2.

4.2.1 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 customer base by an arbitrary amount, the marginal cost per customer is close to $0. Therefore Ziprecruiter is likely under-pricing its service, at least under myopic pricing that optimizes current monthly profits.

Figure 2 plots Ziprecruiter's expected monthly revenue per customer at each of the tested prices. Using the myopic, monthly profit objective, Ziprecruiter is indeed under-pricing considerably. Along our grid of tested price levels, the average monthly revenue per prospective customer is maximized at $399. Although, 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, it may also lower longer-term retention in ways that impact long-term profitability. Figure 3 reports the expected net present value of revenues per customer 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 understand this finding, table 5 reports both the acquisition rate (from September) and the retention rate (for October to 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 customers. In sum, our relative ranking of prices does not change much if we consider a longer-term planning horizon.

The data confirm that Ziprecruiter should raise its prices by more than 100%, which would generate substantial incremental revenues per customer. However, the added structure of the proposed demand model will be required to determine the exact optimal uniform price to charge and the structure of a personalized pricing strategy. In the next section, we discuss demand estimation.

4.2.2 Demand estimation

We now use the data from the field experiment to estimate the Logit demand model using our WLB estimator discussed in section 4.1[14]. Since the experiment randomized the prices charged to each customer,

---

[14]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.

we do 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 customers 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, α, and the 133 interaction effects with price in the slope, β.[15]

In addition to our WLB estimates, we also report results from other approaches. 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 over-fitting. 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, 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 4 reports in-sample and out-of-sample fit measures. For MLE, we report the Bayesian Information Criterion (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, 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, the switch from MLE to Lasso improves the in-sample BIC considerably: 10,018 versus 8,366. This improvement is consistent with our concern that the MLE will over-fit the data. The WLB provides comparable fit to the Lasso in-sample, with an average BIC (across bootstrap replications) very similar to the Lasso BIC.

To see the important role of both variable selection and model uncertainty, note that across the 100 bootstrap replications, we retain as few as 58 to as many as 188 variables in the active set. 172 of the parameters have more than a 50% posterior probability of being non-zero. If we look at the 6 parameters with a more than 90% posterior probability of being non-zero, 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 an additional verification, we also examine the out-of-sample fit of each of our estimators in the second column of Table 4. We first split the sample into training and prediction sub-samples, randomly assigning 90% of the customers to the former and the remaining 10% to the latter. We run each specification using the training sample. We then use the parameter estimates from the training sample to compute

---

[15]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.

the BICs in the prediction sample. Once again, the entire range of BIC values from the WLB is below the BIC of the MLE, again confirming our concerns about over-fitting and highlighting the importance of feature selection and regularization. The WLB performs slightly worse than basic Lasso out-of-sample. As an additional validation, in section 4.4 below, we also test our pricing recommendations out of sample in a second experiment with a new set of test customers.

4.3   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 $\mathbb{E}\left[\beta\left(x_i\right)|D,x_i\right] \geq 0$ at the highest negative value of $\mathbb{E}\left[\beta\left(x_i\right)|\mathbf{D}\right]$ [16].

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 customers that pay for a month of service), posterior expected revenue per customer and posterior expected customer surplus. 95% posterior credibility intervals are also reported for each of these predicted outcomes.

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.com was pricing on the inelastic region of demand prior to 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 **??**, 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 customer 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 now explore decision-theoretic personalized pricing. Figure 4 summarizes the degree of estimated heterogeneity across customers. In panel (a), we report the distribution of customers' posterior mean price sensitivities

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

The dispersion across customers suggests a potential opportunity for Ziprecruiter to price discriminate.

---

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

In panel (b), we report the distribution of customers' posterior mean surplus when Ziprecruiter prices its monthly service at $99:

$$\mathbb{E}\left[WTP\left(x_i\right)|\mathbf{D},x_i\right] = -\frac{1}{B}\sum_{b=1}^{B}\frac{\log\left(1+\exp\left(\alpha^b\left(x_i\right)-\$99\times\beta^b\left(x_i\right)\right)\right)}{\beta^b\left(x_i\right)}. \tag{10}$$

The measure of surplus measures the dollar value created to a customer by the availability of Ziprecruiter's service (versus only the no-purchase option). So we can think of $\mathbb{E}\left[WTP\left(x_i\right)|\mathbf{D},x_i\right]$ as the expected willingness-to-pay of a customer with features x_i to keep Ziprecruiter service available. Panel (b) illustrates the wide dispersion in dollar value customers derive from the availability of Ziprecruiter when it costs $99. The 2.5^{th} percentile, median and 97.5^{th} 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.

The distribution of decision-theoretic personalized prices are plotted in Figure 5. We observe considerable dispersion in the prices, ranging from as low as $126 to as high as $6,292. Across our $N = 7,866$ customers, 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 customers 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 customer by 19% relative to uniform pricing, from $39.01 to $46.57. Moreover, compared to Ziprecruiter's base price of $99, decision-theoretic personalized pricing increases posterior expected revenue per customer 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[17]. In the fourth row of Table 6, we re-compute 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 customer is still 8% higher than under uniform pricing. The expected posterior revenue per customer 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 raising prices above the baseline of $99, mitigating any concerns about triggering a price war. Furthermore, pricing is not transparent in this industry since prices are not posted in a public manner. At Ziprecruiter, for instance, a firm must complete the registration process to obtain a price quote. Since our price discrimination is based on a complex array of customer features as opposed to past behavior, it also seems unlikely that our personalized pricing structure would lead to unintended strategic behavior by Ziprecruiter's customers (e.g., (Fudenberg

---

[17]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 customers.

and Villas-Boas, 2006; Chen, Li, and Sun, 2015; Bonatti and Cisternas, 2018)). For instance, we expect the self-reporting of customer features to be incentive compatible since the responses influence the performance of the matching algorithm to locate prospective recruits.

4.4 Experiment Two: 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 customers that arrived to the Ziprecruiter.com paywall during this period and had not previously paid for the firm's services. Each prospective customer 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 over-sampled the personalized pricing cell to obtain more precision given the dispersion in prices charged across customers.

The tested pricing structures were formulated in part based on Ziprecruiter's own needs. For instance, 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, customers were charged a price based on the values of x_i they reported during the registration stage. As we indicated in the previous section, Ziprecruiter capped the personalized prices at $499. In addition, we rounded the personalized price down to the nearest $9, discretizing the prices into $10 buckets ranging from $119 to $499. For instance, a customer with a targeted price of $183 would be charged $179. Ziprecruiter used this rounding because they believed customers 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 customers reached Ziprecruiter's paywall. Of these prospectives, 5,315 were starters and the remainder were larger firms. Amongst 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 customer 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 6

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.

4.4.1 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 structural model and the corresponding inferences regarding profits under different pricing structures discussed in section 4.3. The box plots in Figure 7 compare the realized sampling distribution for conversion across several of the tested price cells to the corresponding inferences for conversion using our WLB approach versus the the post-Lasso MLE and classical MLE approaches (as discussed at the end of section 3.2). To 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 box plots 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 under-stated 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 over-fitting. 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 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 customer 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 customer in each of the three cells based on our estimates from the September 2015 training sample. Specifically, we use the posterior distribution of the parameter estimates, $F\left(\Theta|\mathbf{D}^{Sept}\right)$ and the observed features from our November subjects, X^{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 which has the lowest monthly price, as expected. Average conversion is almost identical in the uniform and personal-

ized pricing cells, at 15%. However, the average profit per customer is higher in the personalized pricing cell, as one would theoretically expect. Overall, the uniform pricing increases expected profits per customer 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 above in section 4.3. 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 customer 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.

5 Customer Surplus and the Role of Information at Ziprecruiter

5.1 The Information Content of Customer Features

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 business-to-business market. We report the posterior expected customer surplus, $E\left(WTP\left(x_i\right)|\mathbf{D}, x_i\right)$ in equation (10), for each decision-theoretic pricing structure in Table 6. Personalization reduces customer surplus considerably from $94.78 to $71.41, 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 customer). Interestingly, this decline comes from a small minority of customers. In fact, 63% of the customers' personalized prices are lower than the uniform optimal price of $327, indicating that over half our customers benefit from personalization. When we cap the personalized prices above at $499, customer surplus falls by only 1.3% (just over $1) relative to uniform pricing, while still allowing the firm to generate a more than 8% gain in profits. Therefore, the majority of customers benefit from the proposed personalized pricing algorithm.

We now explore the types of customers that benefit from personalized pricing. While our experiment was not designed to recover the causal effect of specific individual firm features, 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. Using the capped personalized prices, these correlations increase to 0.52 and 0.51 respectively. "Company size" is also strongly correlated with prices (small companies have a correlation of -0.15 with unrestricted prices and -0.29 with the capped prices). However, the correlational value of information can be clouded by the fact that certain features, such as state and company type, com-

prise 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 $g = 1,...,6$ 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 $\mathscr{X}$ represent the complete feature set and let $f(p^*|\mathscr{X})$ denote the density of personalized prices based on information set $\mathscr{X}$. 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 $f(p^*|\mathscr{X}_{-g})$ where $-g$ denotes the exclusion of feature group g. We then compute the Kullback-Leibler divergence in the distribution of personalized prices when we exlude feature group g:

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

We effectively treat $f(p^*|\mathscr{X}_g)$ as our target distribution so that $KLD(\mathscr{X}||\mathscr{X}_{-g})$ measures the entropy associated with approximating $f(p^*|\mathscr{X})$ using $f(p^*|\mathscr{X}_{-g})$, 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 ($KLD\left(\mathscr{X}||\mathscr{X}_{-\{state\}}\right) = 0.032$), followed by job category ($KLD\left(\mathscr{X}||\mathscr{X}_{-\{job\ category\}}\right) = 0.029$), benefits ($KLD\left(\mathscr{X}||\mathscr{X}_{-\{benefits\}}\right) = 0.018$), employment type ($KLD\left(\mathscr{X}||\mathscr{X}_{-\{employment\ type\}}\right) = 0.0078$), company type ($KLD\left(D||D_{-\{company\ type\}}\right) = 0.004$) and declared number of job slots ($KLD\left(D||D_{-\{job\ slots\}}\right) = 0.002$). 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 customers. In sum, individual features like 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.

5.2   Data Policies and Customer Surplus

Policies like GDPR and CCPA have been enacted to protect consumer's privacy broadly, but also to prevent firms from surplus extraction. Theoretically, however, it is possible that restricting the types of data firms are permitted to use for personalized pricing could harm consumer surplus (e.g., Bergemann, Brooks, and Morris, 2015). We will now use our Ziprecruiter case study to explore how restrictions over the set of features available to a firm for pricing purposes affects customer surplus.

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 only permitted to use a subset of the $g = 1,...,6$ sets of customer features for the personalization of prices. Let $\mathscr{X}$ represent the complete feature

set, let $\mathscr{X}^o \subset X$ denote the subset of features the firm can use for segmenting customers and setting personalized prices, and let $\mathscr{X}^u \subset X$ represent the features the firm cannot use for segmentation. The firm can partition demand for a customer i with features X_i into the targetable and non-targetable components as follows:

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

where

$$\alpha\left(X_i^o, X_i^u, \Theta\right) = \alpha + X_i^o \alpha_o + X_i^u \alpha_u$$

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

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

$$p_i^* = \arg\max_p \left\{ (p-c)' \int \int \mathbb{P}(p; X_i^o, \Theta) \, dF_{X^u}\left(X^u | X^o\right) dF_\Theta\left(\Theta | \mathbf{D}\right) \right\} \tag{11}$$

where $F_{X^u}\left(X^u | X^o\right)$ represents the firm's beliefs about a customer's unobserved traits, X^u, conditional on her observed traits, X^o. We use an empirical estimate of $F_{X^u}\left(X^u | X^o\right)$ 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 features. We solve the personalized prices 11 corresponding to each of the 62 possible combinations of the $g = 1, ..., 6$ feature groups, which includes the case using all the feature variables[18]4.3

We report the range of feasible personalized pricing outcomes in the surplus triangle in Figure 8, 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, $F_\Theta\left(\Theta | \mathbf{D}\right)$. Point A represents the case where the firm has conducted demand estimation, but does not use any of the customer-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 not only observes all of the customers' features, it also observes their utility shocks, $\{\varepsilon_i\}_{i=1}^N$. In this case, the firm conducts perfect price discrimination. See Appendix D 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 customer 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 customer surplus is zero. Bergemann, Brooks, and Morris (2015) show that every point in this surplus triangle represents a potentially feasible

---

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

segmentation with third-degree price discrimination.

The top panel of Figure 8 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 only generates 30% of the expected posterior profits under perfect price discrimination.

Turning to the demand side, each of our 62 feasible segmentation schemes retains over 50% of the expected customer surplus at point C and over 70% of the expected surplus at point A (uniform pricing). Even though it is theoretically possible for a segmentation scheme to exist that would increase the expected posterior customer 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 customer surplus under uniform pricing. Recall from above that when we cap prices at $499, personalization based on the full feature set achieves over 98% of the customer surplus under uniform pricing while improving posterior expected profits by over 8%. Therefore, while personalized pricing reduces customer surplus somewhat, it does not come close to extracting all the surplus as would have been the case under a purely theoretical perfect price discrimination scheme.

The bottom three panels zoom in on the surplus triangle to examine how different data policies influence customer surplus. The left-most panel indicates that when we only allow the firm to target prices based on "benefits," total customer surplus is almost $1 lower than when we also allow the firm to target on "company type" and "declared number of job slots." The middle panel shows a similar result. Only targeting prices on "job category" generates more than $1 less customer surplus than when the firm is also permitted to target on "employment type," "company type" and "declared number of job slots." However, the right-hand panel indicates that some customer features stricly harm customer surplus. In particular, allowing the firm to target on "job category" and/or "state" reduces customer surplus. These results indicate that allowing the firm to target on more granular data can be good for customer surplus and that granularity per se does not harm customers.

In spite of the decline in total customer surplus, the percentage of customers that benefit from personalization ranges from 59.4% to 62.2% across our 62 segmentation scenarios. Therefore, a minority of customer is baring the cost of personalization. To see this point more clearly, Figure 9 plots density estimates of the change in posterior expected surplus across customers for each of the 62 segmentation scenarios versus uniform pricing. In each case, we see a large mass of customers just to the right of $0, representing the majority who benefit from personalized prices. We then see a long tail to the left of $0 representing the minority of customer who are harmed. If we correlate the incidence that a customer benefits from personalization ($p_i^* < p^{unif}$) with the customer features, we find that two most highly correlated features are "Small Company Type" ($corr = 0.38$) and "Part-Time Employment" ($corr = 0.31$).

At face value, these results suggests 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, e.g., "Total Job Benefits" ($corr = -0.81$), "Full-Time Employment" ($corr = -0.37$) and "Medium Company Type" ($\rho = -0.25$).[19] Therefore, larger companies with full-time employment and high benefits are the most likely to be harmed from personalized pricing. Conceptually, this reallocation of customer surplus from personalized pricing could be rationalized as "fair" under a Pareto-weight scheme that assigns higher social value to smaller, disadvantaged firms.

A striking finding from our analysis is that we do not observe a monotonic relationship between the number of features used for segmentation and total customer welfare or total number of customers who benefit from personalization. Thus, granting the firm more access to customer data does not per se lead to more customer harm. In Figure 8, we can see that the full segmentation using all 6 groups of customer features generates more customer surplus than several of the restricted scenarios. For instance, allowing the firm to condition its prices on all 6 feature groups increases customer 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 customers benefit from personalized prices conditioned on all feature variables, whereas only 59.4% benefit when the firm is only allowed to condition its prices on "benefits." In Figure 9 we see that the density of the change in expected posterior surplus across customers 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 customer as it can lead to segmentation schemes that allocate more surplus to the customer.

6 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 customer information and its potential impact on customer 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 would reduce total customer surplus. However, we also find that certain data policies that would restrict the use of specific customer variables for targeting purposes could in fact reduce customer welfare.We also find that the majority of customer would benefit from being charged lower prices than the uniform rate.

In sum, the re-distributive aspects of personalized pricing cause the majority of customers to benefit. Current public debate surrounding the fairness of differential pricing needs to consider these re distributive aspects of personalized pricing. Most importantly, over-regulation of the types of data firms can use

---

[19]The Large Company Type feature was excluded due to redundancy.

for personalized pricing purposes could potentially harm customers. In a setting where one can assign heterogeneous Pareto weights, this redistributive effect could be socially desirable if the beneficiaries of personalized pricing are from under-represented or disadvantaged customer segments.

The results presented herein are based on a single case study of a large digital human resources platform. 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 customers 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 customer 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 consumers goods market where personalized pricing may be more transparent and less accepted[20]. Finally, our findings focus on the monopoly price discrimination problem for Ziprecruiter.com. 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[21].

In this paper, we approximate the posterior distribution of demand using a weighted likelihood bootstrap of the lasso estimator. 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 with discrete treatments (Athey and Imbens, 2016b,a) and, more recently, with continuous treatments (Hansen, Kozbur, and Misra, 2017). We believe this to be a fertile area for future work on both the theoretical and applied fronts.

---

[20]Negotiated price deals are quite common in B2B pricing, especially with sales agents.

[21]See for instance the empirical analysis of competitive geographic price discrimination in Dubé, Fang, Fong, and Luo (2017), the theoretical work by Corts (1998) and literature survey in Stole (2007)

References

ACQUISTI, A., C. TAYLOR, AND L. WAGMAN (2016): "The Economics of Privacy," *Journal of Economic Literature*, 54(2), 442–492.

ACQUISTI, A., AND H. R. VARIAN (2005): "Conditioning Prices on Purchase History," *Marketing Science*, 24(3), 367–381.

ANSARI, A., AND C. F. MELA (2003): "E-Customization," *Journal of Marketing Research*, 40(2), 131–145.

ATHEY, S., AND G. W. IMBENS (2016a): "Estimation and Inference of Heterogeneous Treatment Effects using Random Forests," *forthcoming Journal of the American Statistical Association*.

——— (2016b): "Recursive Partitioning for Heterogeneous Causal Effects," *PNAS*, 113(27), 7353–7360.

BAUNER, C. (2015): "Mechanism Choice and the Buy-It-Now Auction: A Structural Model of Competing Buyers and Sellers," *International Journal of Industrial Organization*, 38, 19–31.

BERGEMANN, D., B. BROOKS, AND S. MORRIS (2015): "The Limits of Price Discrimination," *American Economic Review*, 105(3), 921–957.

BERGER, J. (1985): *Statistical Decision Theory and Bayesian Analysis*. Springer-Verlag.

BERRY, S. T. (1994): "Estimating Discrete-Choice Models of Product Differentiation," *Rand Journal of Economics*, 25(2), 242–262.

BISSIRI, P. G., C. C. HOLMES, AND S. G.WALKER (2016): "A general framework for updating belief distributions," *Journal of the Royal Statistical Society*, 78(5), 1103–1130.

BONATTI, A., AND G. CISTERNAS (2018): "Consumer Scores and Price Discrimination," *NYU Working Paper*.

BRADLOW, E., P. LENK, G. ALLENBY, AND P. E. ROSSI (2004): *Marketing Research and modeling: Progress and Prospects, A Tribute to Paul Green*chap. When BDT in Marketing Meant Bayesian Decision Theory: The Influence of Paul Green's Research. Kluwer Academic Publishers.

CEA (2015): "Big data and differential pricing / Executive Office of the President of the United States, Council of Economic Advisors.," *Washington, D.C.] : Executive Office of the President of the United States, Council of Economic Advisors, 2015.*, pp. 1–21.

CHAMBERLAIN, G., AND G. W. IMBENS (2003): "Nonparametric Applications of Bayesian Inference," *Journal of Business and Economic Statistics*, 21(1), 12–18.

CHEN, Y., X. LI, AND M. SUN (2015): "Competitive Mobile Targeting," *Working Paper*.

CHERNOZHUKOV, V., D. CHETVERIKOV, M. DEMIRER, E. DUFLO, C. HANSEN, W. NEWEY, AND J. ROBINS (2016): "Double/Debiased Machine Learning for Treatment and Causal Parameters," *arXiv e-prints*.

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 Science*, 51, 832–849.

COHEN, P., R. HAHN, J. HALL, S. LEVITT, AND R. METCALFE (2016): "Using Big Data to Estimate Consumer Surplus: the case of Uber," *NBER Working Paper 22627*.

CORTS, K. S. (1998): "Third-degree price discrimination in oligopoly: all-out competition and strategic commitment," *The RAND Journal of Economics*, 29, 306–323.

COWAN, I. A. S., AND J. VICKERS (2010): "Monopoly Price Discrimination and Demand Curvature," *The American Economic Review*, 100(4), 1601–1615.

COWAN, S. (2012): "Third-Degree Price Discrimination and Consumer Surplus," *Journal of Industrial Economics*, LX(2), 333–345.

DONG, X., P. MANCHANDA, AND P. K. CHINTAGUNTA (2009): "Quantifying the Benefits of Individual-Level Targeting in the Presence of Firm Strategic Behavior," *Journal of Marketing Research*, XLVI, 207–221.

DUBÉ, J.-P., Z. FANG, N. M. FONG, AND X. LUO (2017): "Competitive Price Targeting with Smartphone Coupons," *forthcoming Marketing Science*.

EFRON, B. (2012): "Bayesian Inference and the Parametric Bootstrap," *The Annals of Applied Statistics*, 6, 1971–1997.

EINAV, L., C. FARRONATO, J. LEVIN, AND N. SUNDARESAN (2017): "Auctions versus Posted Prices in Online Markets," *Journal of Political Economy*, forthcoming.

EINAV, L., AND J. LEVIN (2010): "Empirical Industrial Organization: A Progress Report," *Journal of Economics Perspectives*, 24(2), 145–162.

——— (2014): "Economics in the age of big data," *Science*, 346(6210).

FAN, J., Y. FENG, AND R. SONG (2012): "Nonparametric Independence Screening in Sparse Ultra-High-Dimensional Additive Models," *Journal of the American Statistical Association*, 106(494), 544–557.

FAN, J., AND R. LI (2001): "Variable Selection via Nonconcave Penalized Likelihood and Its Oracle Properties," *Journal of the American Statistical Association*, 96(456), 1348–1360.

FOLEY, D. (1967): "Resource allocation and the public sector," *Yale Economic Essays*.

FUDENBERG, D., AND J. M. VILLAS-BOAS (2006): "Behavior-Based Price Discrimination and Customer Recognition," in *Hanbooks in Information Systems: Economics and Information Systems*, ed. by T. Hendershott. Elsevier.

GRANVILLE, K. (2018): "Facebook and Cambridge Analytica: What You Need to Know as Fallout Widens," *New York Times*.

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

GREEN, P. E., AND R. E. FRANK (1966): "Bayesian Statistics and Marketing Research," *Journal of the Royal Statistical Society*, 15(3), 173–190.

HANSEN, C. B., D. KOZBUR, AND S. MISRA (2017): "Targeted Undersmoothing," *Chicago Booth Working Paper*.

HASTIE, T., R. TIBSHIRANI, AND J. FRIEDMAN (2009): *The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition (Springer Series in Statistics) 2nd ed.* Springer.

HIRANO, K. (2008): *The New Palgrave Dictionary of Economics. Second Edition*chap. Decision Theory in Econometrics. Palgrave Macmillan.

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

KRUGMAN, P. (October 4, 2000): "Reckonings; What Price Fairness?," *New York Times*.

KUMAR, V., S. SRIRAM, A. LUO, AND P. CHINTAGUNTA (2011): "Assessing the Effect of Marketing Investments in a Business Marketing Context," *Marketing Science*, 30(5), 924–940.

LEEB, H., AND B. M. POTSCHER (2008): "Sparse estimators and the oracle property, or the return of Hodges estimator," *Journal of Econometrics*, 142, 201–211.

LEVITT, S. D., AND J. A. LIST (2009): "Field experiments in economics: The past, the present, and the future," *European Economic Review*, 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," *Proceedings of the National Academy of Sciences*, 113(27), 7323–7328.

LIST, J. A. (2004): "The Nature and Extent of Discrimination in the Markeplace: evidence from the field," *Quarterly Journal of Economics*, 119(1), 49–89.

LYDDON, S. P., AND C. C. HOLMES (2019): "General Bayesian updating and the loss-likelihood bootstrap," *forthcoming Biometrika*.

MISRA, S., AND H. S. NAIR (2011): "A Structural Model of Sales-Force Compensation Dynamics: Estimation and Field Implementation," *Quantitative Marketing and Economics*, 9, 211–257.

MOHAMMED, R. (October 20, 2017): "How Retailers Use PersPersonalized Prices to Test What You're Willing to Pay," *Harbard Business Review*.

NEWTON, M. A., AND A. E. RAFTERY (1994): "Approximate Bayesian Inference with the Weighted Likelihood Bootstrap," *Journal of the Royal Statistical Society, Series 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," *Journal of the American Statistical Association*, 103, 681–686.

PIGOU, A. (1920): *The Economics of Welfare*. MacMillan.

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

ROSSI, P. E., R. E. MCCULLOCH, AND G. M. ALLENBY (1996): "The Value of Purchase History Data in Target Marketing," *Marketing Science*, 15(4), 321–340.

RUBIN, D. B. (1981): "The Bayesian Bootstrap," *The Annals of Statistics*, 9, 130–134.

SAVAGE, L. (1954): *Statistical Decision Functions*. Wiley.

SHAPIRO, C., AND H. R. VARIAN (1999): *Information Rules*. Harvard Business School Press.

SHILLER, B., AND J. WALDFOGEL (2011): "Music for a song: An Empirical Look at Uniform Song Pricing and Its Alternatives," *The Journal of Industrial Economics*, 59(4), 630–660.

SHILLER, B. R. (2015): "First-Degree Price Discrimination Using Big Data," *Working Paper*.

SIMESTER, D. I., P. SUN, AND J. N. TSITSIKLIS (2006): "Dynamic Catalog Mailing Policies," *Management Science*.

SMITH, M., J. BAILEY, AND E. BRYNJOLFSSON (2000): *Understanding Digital Markets*. MIT Press, Cambridge MA.

STOLE, L. A. (2007): "Price Discrimination and Imperfect Competition," in *Handbook of Industrial Organization*, vol. 3. Elsevier.

TADDY, M. (2015a): "Distributed Multinomial Regression," *The Annals of Applied Statistics*, 9, 1394–1414.

——— (2015b): "One-step estimator paths for concave regularization," *University of Chicago Booth School of Business Working Paper*.

TADDY, M., M. GARDNER, L. CHEN, AND D. DRAPER (2016): "A nonparametric Bayesian analysis of heterogeneous treatment effects in digital experimentation," *Journal of Business and Economic Statistics*, 34(4), 661–672.

THOMSON, W. (2011): *Handbook of Social Choice and Welfare* vol. 2, chap. Fair Allocation Rules, pp. 393–506. Elsevier.

TIBSHIRANI, R. (1996): "Regression Shrinkage and Selection via the Lasso," *Journal of the Royal Statistical Society, Ser. B*, 58, 267–288.

TIROLE, J. (1988): *The Theory of Industrial Organization*. MIT Press.

TUROW, J., L. FELDMAN, AND K. MELTZER (2005): "Open to Exploitation: America 's Shoppers Online and Offline," *Working Paper, Annenberg School for Communication*.

USEEM, J. (2017): "How Online Shopping Makes Suckers of Us All," *The Atlantic*.

VARIAN, H. R. (1980): "A Model of Sales," *American Economic Review*, 70(4), 651–659.

——— (1989): *Price Discrimination* chap. 10, pp. 597–654. Elsevier Science Publishers BV.

VERBOVEN, F. (2008): *The New Palgrave Dictionary of Economics* chap. Price Discrimination (Empirical Studies). Palgrave Macmillan.

WAGER, S., AND S. ATHEY (2015): "Estimation and Inference of Heterogeneous Treatment Effects using Random Forests," *ArXiv e-prints*.

WALD, A. (1950): *Statistical Decision Functions*. John Wiley & Sons.

WALDFOGEL, J. (2015): "First Degree Price Discrimination Goes to School," *Journal of Industrial Economics*, 63(4), 569–597.

WOLVERTON, T. (2010): "MP3 player sale exposes Amazon's flexible prices," *https://www.cnet.com/news/mp3-player-sale-exposes-amazons-flexible-prices/ accessed on 4-9-2019*.

ZHANG, J., O. NETZER, AND A. ANSARI (2014): "Dynamic Targeted Pricing in B2B Relationships," *Marketing Science*, 33(3), 317–337.

ZOU, H. (2006): "The Adaptive Lasso and Its Oracle Properties," *Journal of the American Statistical Association*, 101(476), 1419–1429.

ZOU, H., T. HASTIE, AND R. TIBSHIRANI (2007): "On the Degrees of Freedom of the Lasso," *The Annals of Statistics*, 35(5), 2173–2192.

	Uniform	Perfect PD	Personalized 1 $\mathbf{D} = (\{1\}$ $\{2,3,4,5,6\})$	Personalized 2 $\mathbf{D} = (\{1\},$ $\{2,3\},$ $\{4,5,6\})$
prices	$p_i^U = \$4, \forall i$	$p_i^{PD} = i, \forall i$	$p_{\{1\}}^{PP1} = \$1$ $p_{\{2,3,4,5,6\}}^{PP1} = \$4$	$p_{\{1\}}^{PP2} = \$1$ $p_{\{2,3\}}^{PP2} = \$2$ $p_{\{4,5,6\}}^{PP2} = \$4$
profits	$12	$21	$13	$17
CS	$3	0	$3	$4

Table 1: Data and Welfare

	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

Table 2: Experimental Price Cells for Stage One

Feature Name
job state
company type
hascomm
company declared job slots needed
job total benefits
employment type
is resume required
job medical benefit
job dental benefit
job vision benefit
job life insurance benefit
job category

Table 3: Company/Job Variables
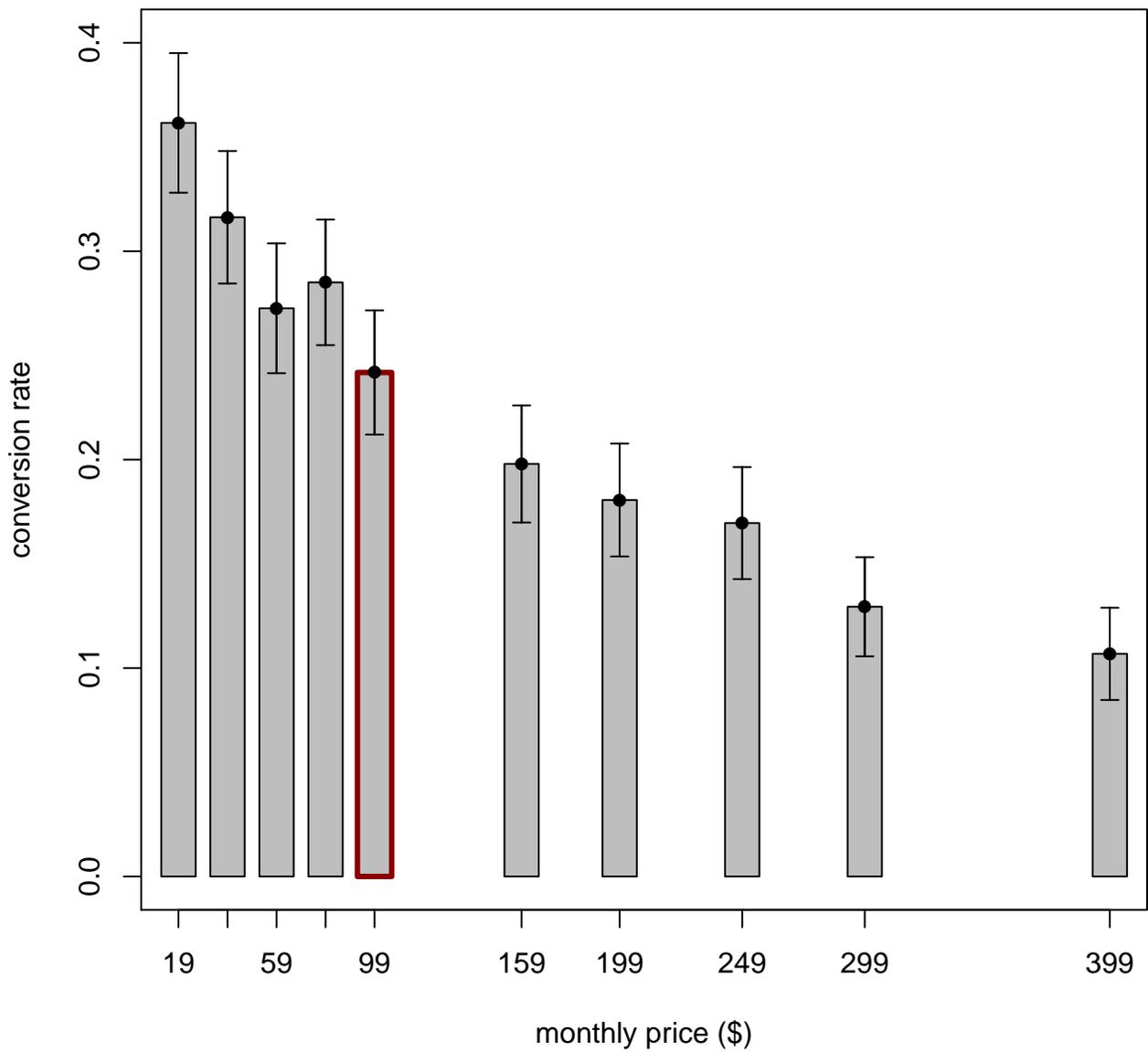
Figure 1: Stage One Experimental Conversion Rates. Each bar corresponds to one of our 10 experimental price cells. The height of the bar corresponds to the average conversion rate within the cell. Error bars indicate the 95% confidence interval for the conversion rate.
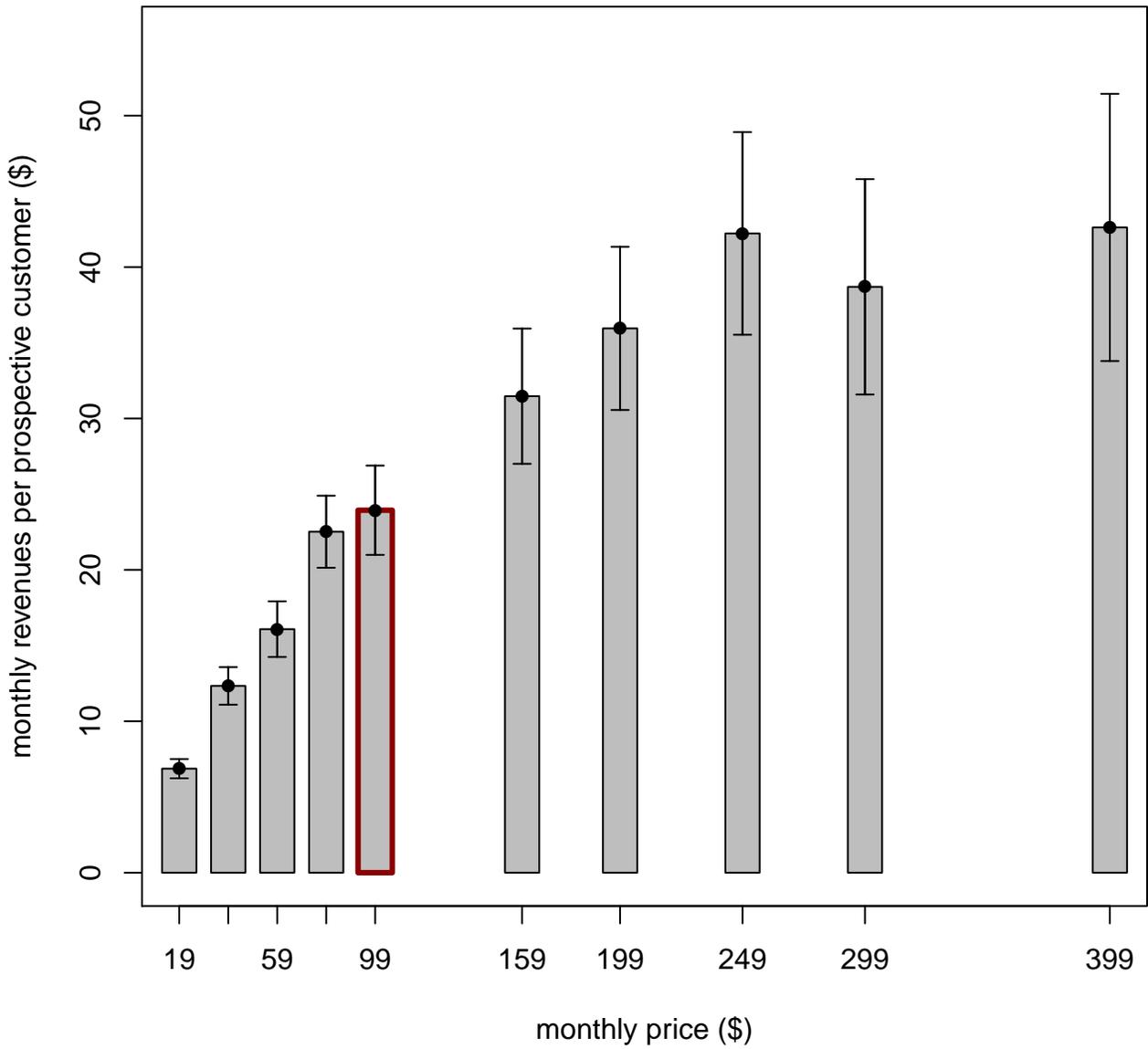
Figure 2: Stage One Experimental Revenues per Customer. Each bar corresponds to one of our 10 experimental price cells. The height of the bar corresponds to the average revenue per prospective customer within the cell. Error bars indicate the 95% confidence interval for the revenues per customer.

Model	In-Sample BIC	Out-of-Sample BIC
MLE	10,018.78	44,30.65
Lasso	8,366.47	2,286.63
WLB range	(7,805.11 , 8,940.06)	(3,249.34 , 4,071.96)

Table 4: Predictive Fit from MLE, Lasso and Weighted Likelihood Bootstrap estimation (WLB) (for WLB we report the range across all 100 bootstrap replications). In-Sample results are based on 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 hold-out sample with the remaining 10% of the firms.

Table 5: 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	0.36	0.8	0.77	0.61	0.56
39	0.32	0.75	0.73	0.52	0.47
59	0.27	0.65	0.63	0.49	0.4
79	0.29	0.69	0.64	0.5	0.39
99	0.24	0.69	0.66	0.48	0.38
159	0.2	0.63	0.61	0.43	0.34
199	0.18	0.56	0.5	0.31	0.19
249	0.17	0.63	0.59	0.39	0.27
299	0.13	0.58	0.53	0.35	0.29
399	0.11	0.54	0.52	0.37	0.25

Decision-Theoretic Pricing Structures	Price	Expected Conversion Rate		Expected Revenue ($ per Customer)		Expected Surplus ($ per Customer)	
		Mean	95% Cred. Int.	Mean	95% Cred. Int.	Mean	95% Cred. Int.
Control	$99	0.25	(0.23,0.28)	25.09	(23.02,27.48)	135.13	(80.35,217.23)
Uniform	$327	0.12	(0.1,0.14)	39.01	(31.9,46.58)	94.78	(42.04,174.98)
Personalized	($126,6292)	0.12	(0.1,0.14)	46.57	(32.89,60.8)	71.41	(32.35,136.84)
Personalized ($499 price cap)	($126,$499)	0.13	(0.11,0.16)	42.21	(33.41,50.53)	93.56	(42.1,172.87)

Table 6: Posterior expected conversion, revenue per customer and surplus per customer by pricing structure for September 2015 experiment.
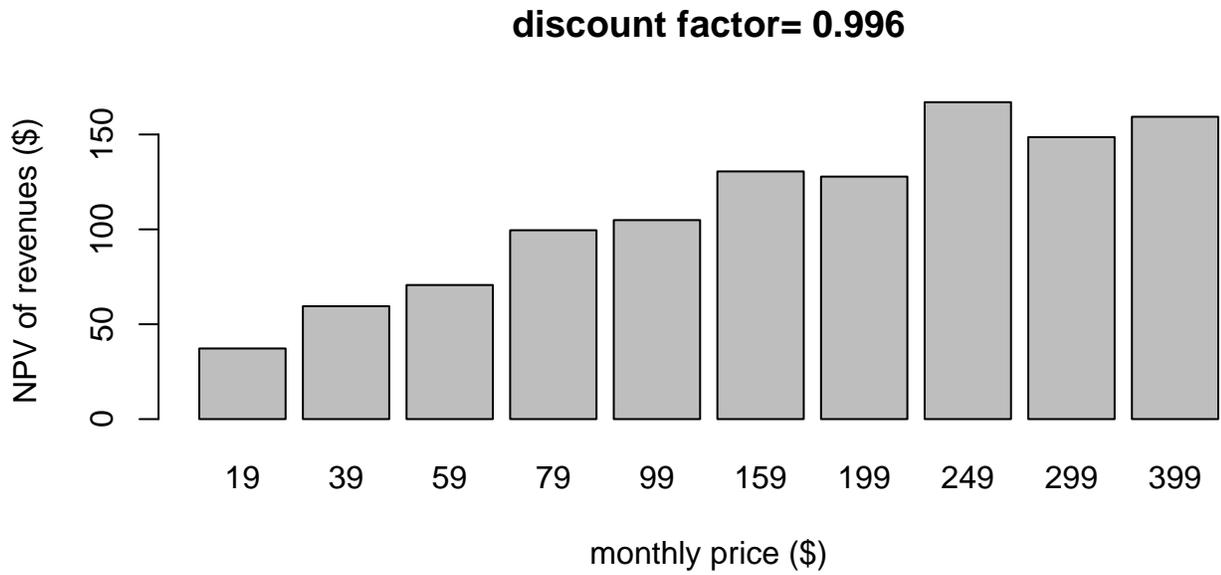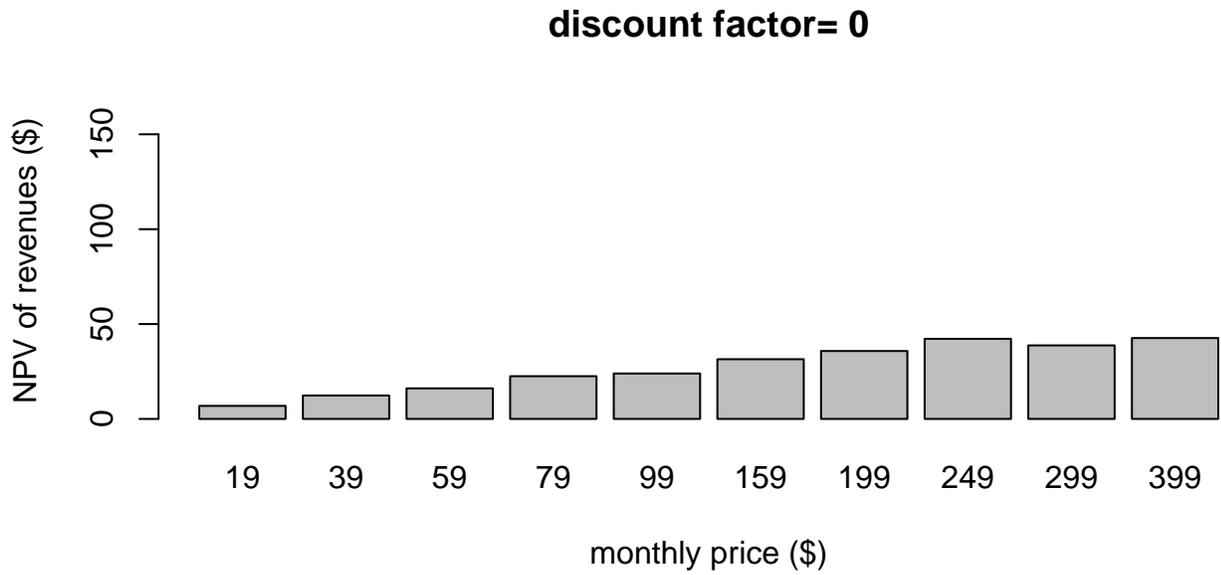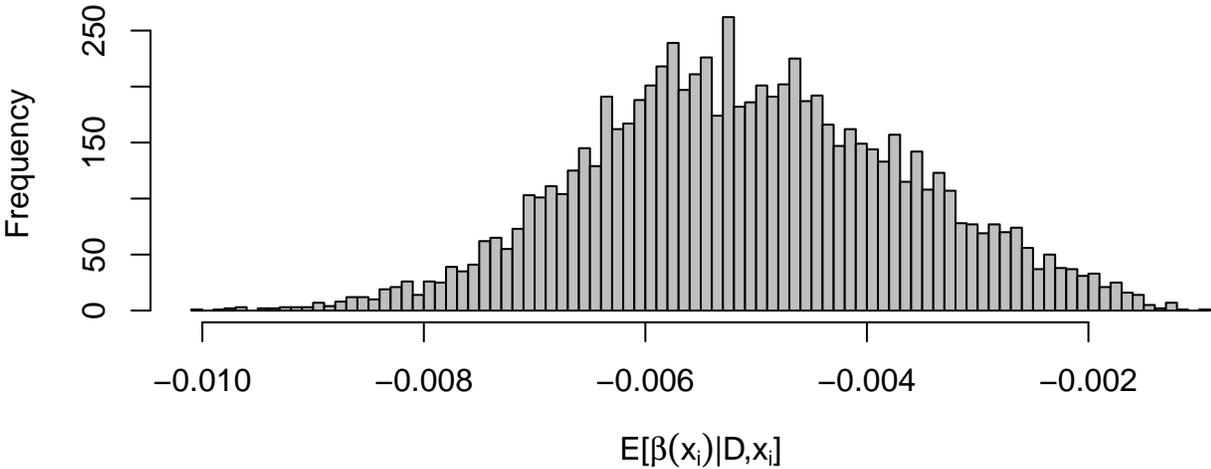
discount factor= 0

discount factor= 0.996

Figure 3: Expected Net Present Value of Monthly Revenues Per Lead over a 4-Month Horizon (September 2015)

Panel (a): Price Coefficient

Panel (b): Customer Surplus when p=$99
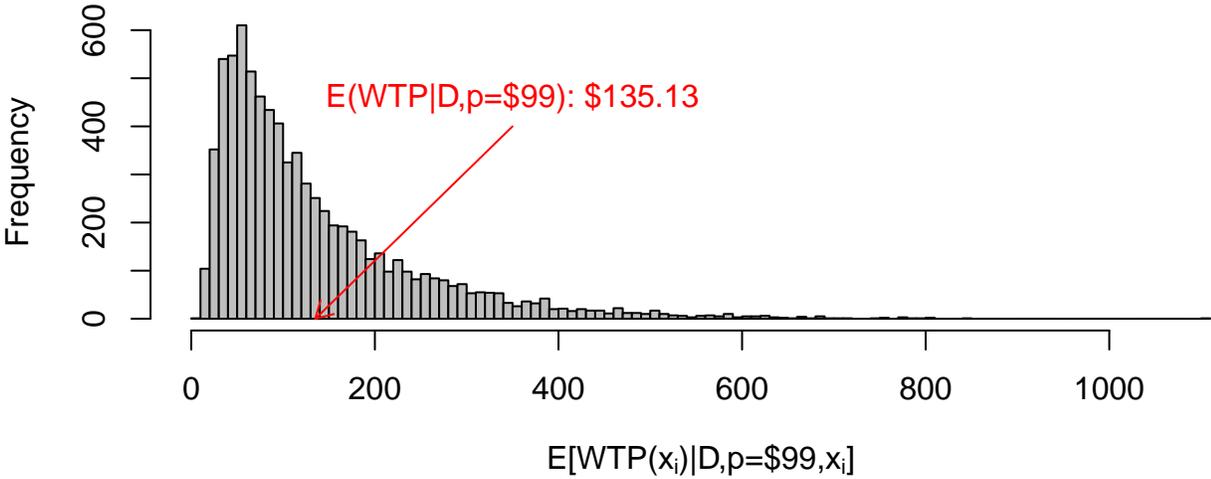

E(WTP|D,p=$99): $135.13

Figure 4: Distribution across customers of posterior mean price sensitivity and posterior surplus from the provision of the service (N=7,867).
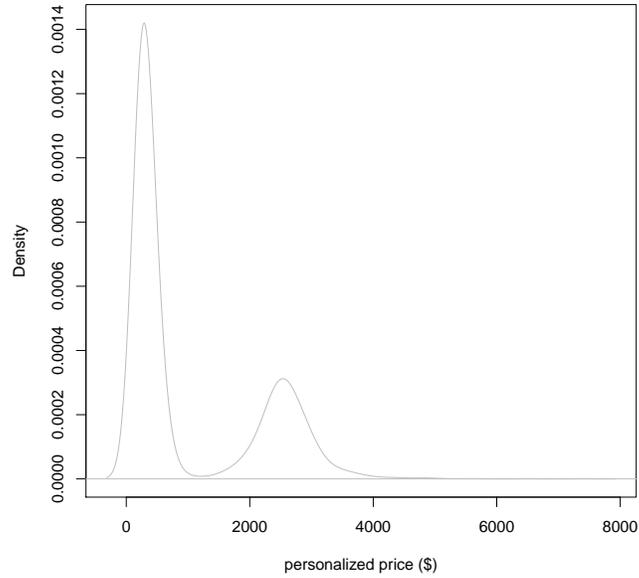
Figure 5: Distribution of Decision-Theoretic Personalized Prices (N=7,867).

	control ($99)	test ($249)	test (personalized pricing)
Sample Size	1,360	1,430	2,485
mean conversion	0.23	0.15	0.15
	(0.21,0.25)	(0.13,0.17)	(0.13,0.16)
mean revenue per customer	$22.57	$37.82	$41.67
	(20.46,24.68)	(33.08,42.66)	(37.87,45.76)
posterior mean conversion	0.26	0.15	0.14
	(0.23,0.29)	(0.13,0.18)	(0.12,0.17)
posterior mean revenue per customer	$25.5	$38.37	$41.05
	(23.26,28.31)	(32.04,44.9)	(33.78,48.78)

Table 7: Predicted versus Realized Outcomes in November 2015 Experiment (Below each realized outcome, we report in brackets the 95% confidence intervals. Below each posterior predicted outcome, we report in brackets the 95% credibility interval.)

density of targeted prices
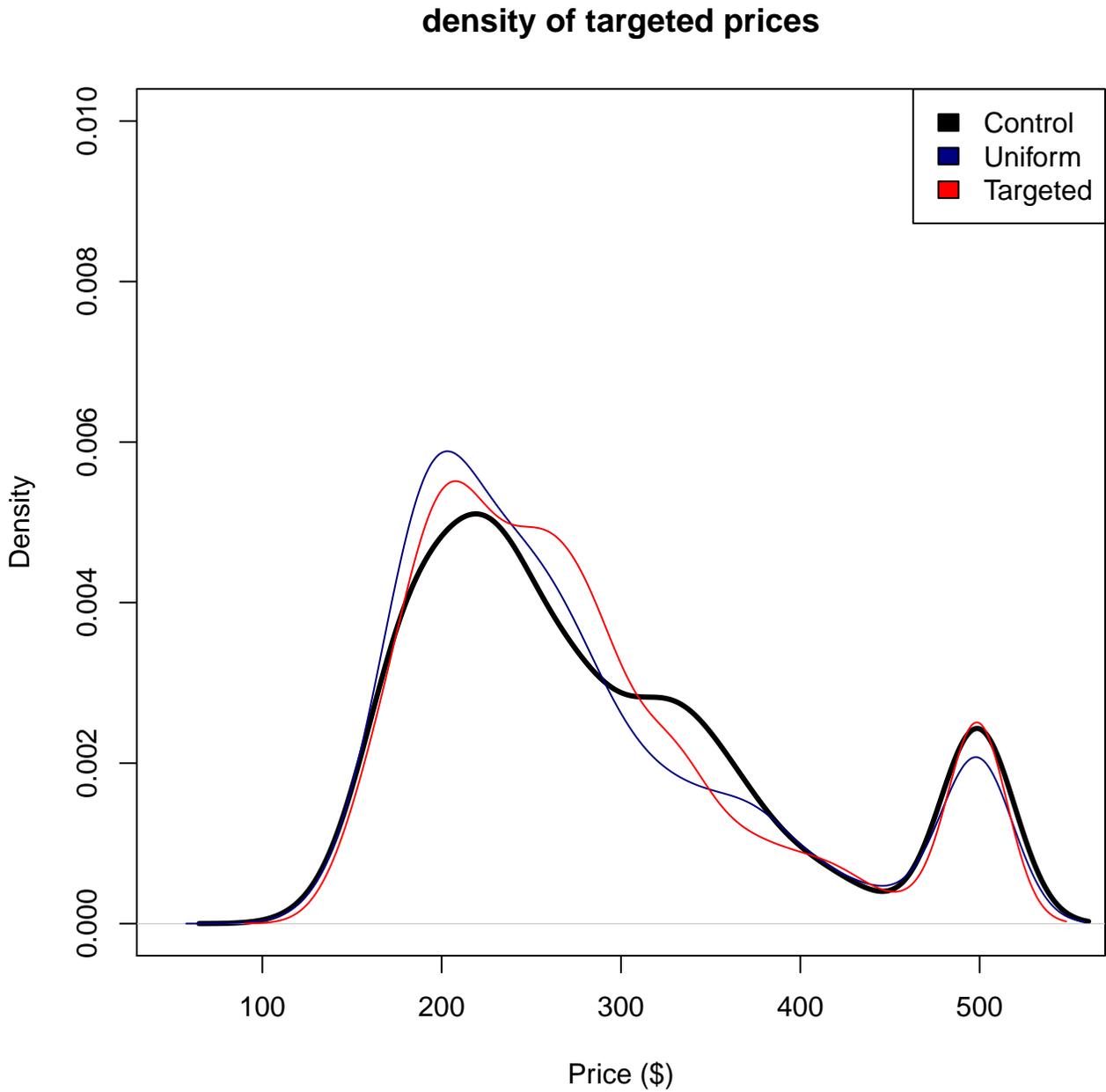
Figure 7: Comparison of Predicted and Realized Conversion

(a) Control Pricing ($99)

(b) Uniform Optimal Pricing

(c) scalable price targeting
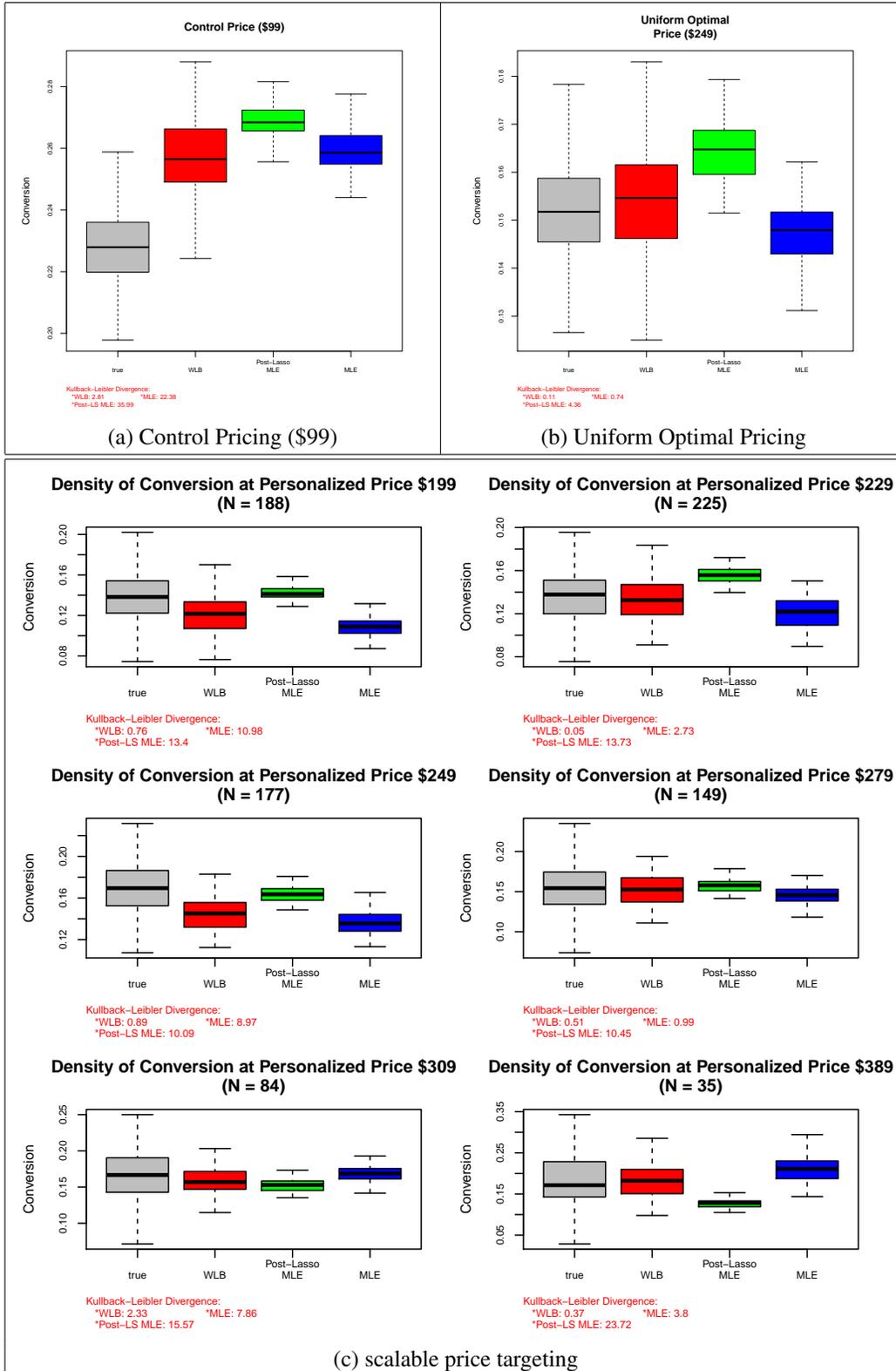
The plots compare the empirical density of realized conversion, for a given pricing structure, to the corresponding predicted densities for WLB, post-Lasso MLE and MLE respectively. The density of realized conversions is computed by bootstrapping (with replacement) from the Nov data.
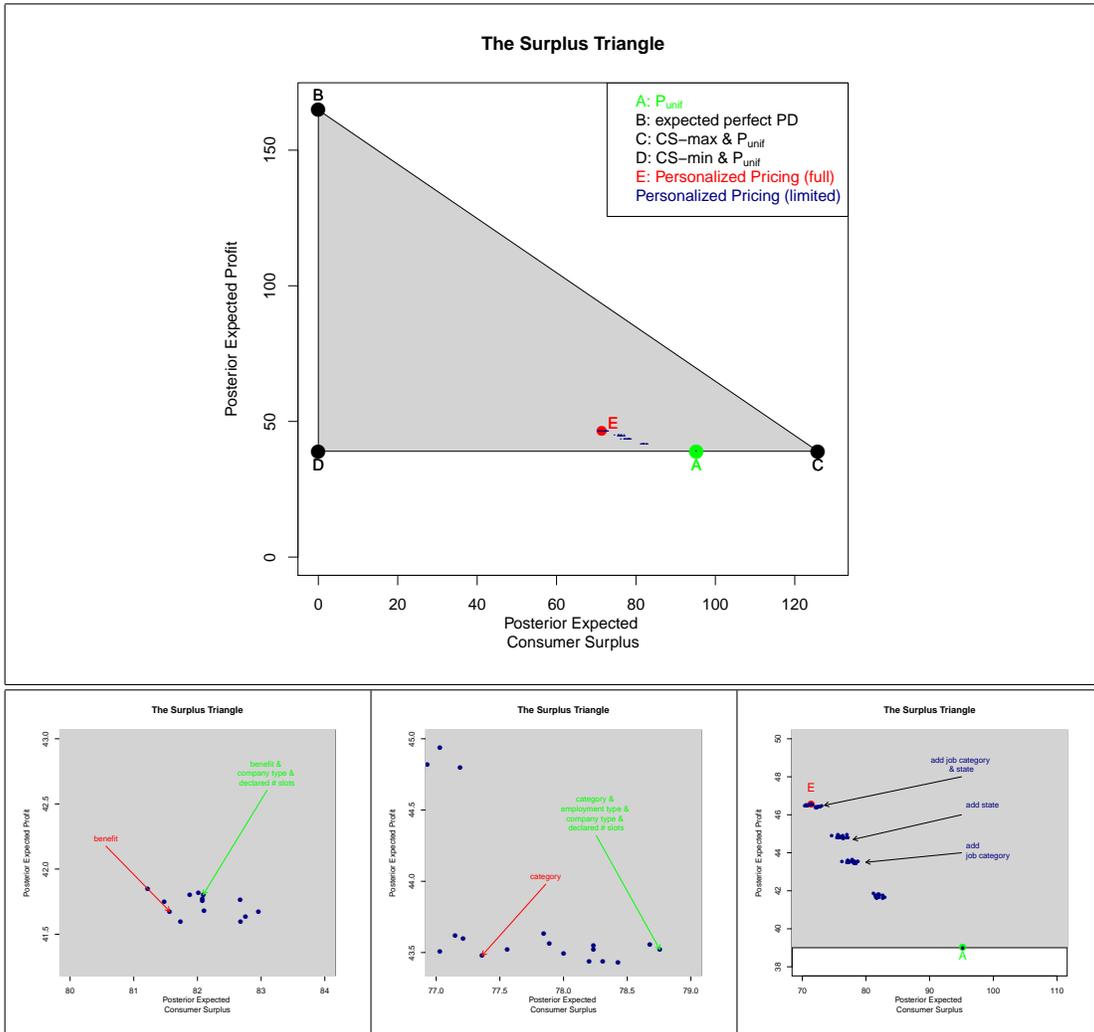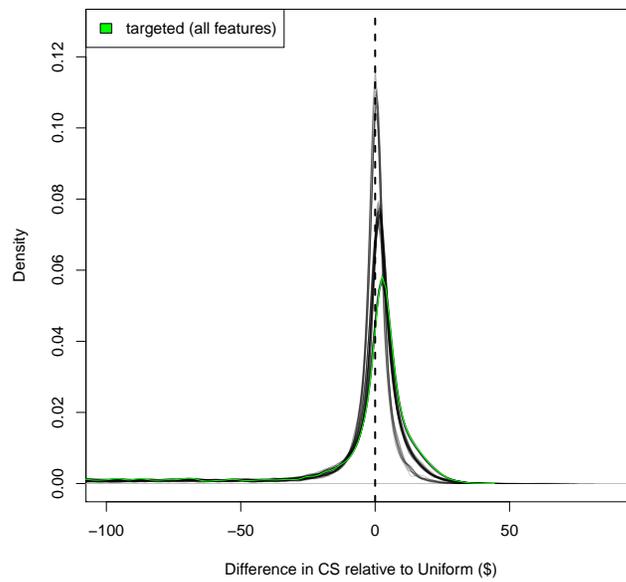
Figure 8: Surplus Triangle

Figure 9: Densities of the Change in Expected Posterior Surplus Across Customers Under Personalized Pricing versus Uniform Pricing

A   The Bayesian Lasso

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

$$F_\Theta(\Theta|D) \propto \ell(D|\Theta) - \sum_{j=1}^{J} \tau_j |\Theta_j| \tag{12}$$

where $\ell(D|\Theta)$ is 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 MAP (maximum a posteriori) estimator that optimizes (12) can be shown to be equivalent to the Lasso regression:

$$\Theta^{Lasso} = \underset{\Theta \in \mathbb{R}^J}{argmax} \left\{ \ell(D|\Theta) - N\lambda \sum_{j=1}^{J} |\Theta_j| \right\}. \tag{13}$$

In Appendix C, we describe the path-of-one-step estimators procedure used to select λ and generate estimates of Θ and its sparsity structure (see also Taddy (2015b)).

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[22]. Instead, we simulate the approximate posterior using a Weighted Likelihood Bootstrap (WLB) of the Lasso problem. The Weighted Likelihood Bootstrap (Newton and Raftery (1994)) is an extension of the Bayesian Bootstrap originally proposed by Rubin (1981).[23] As discussed in Efron (2012), the BB 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 customer 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 (13). 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

---

[22]Challenges include drawing from a large-dimensional distribution, assessing convergence of the MCMC sampler, tuning the algorithm and storing a non-sparse simulated chain in memory.

[23]To be clear, our implementation only uses the first stage of the WLB procedure described in Newton and Raftery (1994) and does not implement the Sampling-Importance-Resampling (SIR) 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.

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 $\mathbf{D} = (D_1, ..., D_N)$. We assume that the data-generating process for D is discrete with support points $(\zeta_1, ..., \zeta_L)$ and corresponding probabilities $\phi = (\phi_1, ..., \phi_L) : Pr(D_i = \zeta_l) = \phi_l$. We can allow L to be arbitrarily large to allow for flexibility in this representation. We assume the following Dirichlet prior on the probabilities

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

Following the convention in the literature, we use the improper prior distribution with $a_l \to 0$. This assumption implies that any support points, ζ_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 $V_l = \sum_{k=1}^{L} \phi_k \phi_l$.

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

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

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

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

2. Run the Lasso
$$\hat{\Theta}^b|_{\lambda} = \underset{\Theta \in \mathbb{R}^J}{argmin} \left\{ \ell^b(\Theta) + N\lambda \sum_{j=1}^{J} |\Theta_j| \right\}$$

   where $\ell^b(D|\Theta) = \sum_{i=1}^{N} V_i^b \ell(D_i|\Theta)$, using the algorithm (15) in Appendix C

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

   (b) Use k-fold-cross validation to determine the optimal penalty, λ^*

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 Ψ_i such as profits and demand elasticities.

C Appendix: Lasso Regression

The penalized Lasso estimator solves for

$$\hat{\Theta}|_\lambda = \underset{\Theta \in \mathbb{R}^J}{argmin} \left\{ \ell(\Theta) + N\lambda \sum_{j=1}^{J} |\Theta_j| \right\} \tag{14}$$

where $\lambda > 0$ controls the overall penalty and $|\Theta_j|$ is the L_1 coefficient cost function. Note that as $\lambda \to 0$, 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 λ is unknown a priori, we derive a regularization path, $\left\{ \hat{\Theta}|_\lambda \right\}_{\lambda=\lambda_1}^{\lambda_T}$, consisting of a sequence of estimates of Θ corresponding to successively lower degrees of penalization. Following Taddy (2015b), we use the following algorithm to construct the path:

1. $\lambda_1 = inf\left\{ \lambda : \hat{\Theta}|_{\lambda_1} = 0 \right\}$

2. set step size of $\delta \in (0,1)$

3. for $t = 2, ..., T$:
$$\begin{aligned} \lambda^t &= \delta \lambda^{t-1} \tag{15} \\ \omega_j^t &= \left(|\Theta_j^{t-1}| \right)^{-1}, \ j \in \hat{S}_t \\ \hat{\Theta}^t &= \underset{\Theta \in \mathbb{R}^J}{argmin} \left\{ \ell(\Theta) + N\sum_{j=1}^{J} \lambda^t \omega_j^t |\Theta_j| \right\}. \end{aligned}$$

The algorithm produces a weighted-L_1 regularization, with weights ω_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 towards zero later in the path. This bias diminishes faster with larger values of γ.

The algorithm in 15 above generates a path of estimates corresponding to different levels of penalization, λ. We use K-fold cross-validation to select the "optimal" penalty, λ^*. We implement the approach using the *cv.gamlr* function from the *gamlr* package in **R**.

D   Appendix: Perfect Price Discrimination

Suppose the firm observed not only the full feature set for a customer i, X_i, but also the random utility shock, ε_i. Under perfect price discrimination, the firm would set the personalized price

$$p_i^{PD} = \max(WTP_i, 0)$$

where WTP_i is customer i's maximum willingness-to-pay (WTP)

$$WTP_i = \frac{(\alpha(X_i) + \varepsilon_i)}{\beta(X_i)}. \tag{16}$$

Customer i would deterministically buy as long as $WTP_i \geq 0$.

Accounting for the fact that the researcher (unlike the firm in this case) does not observe ε, the expected probability that a customer with preferences (α, β) would purchase at the perfect price discrimination price is

$$\begin{aligned}
\mathbb{P}\left(p^{PD}; X_i, \Theta\right) &= Pr(WTP \geq 0) \\
&= 1 - \frac{1}{1 + \exp(\alpha)}.
\end{aligned} \tag{17}$$

The corresponding expected profit from this customer is

$$\pi\left(p^{PD} | \alpha, \beta\right) = E(WTP | WTP \geq 0, \alpha, \beta) Pr\left(buy | p = p^{PD} \alpha, \beta\right). \tag{18}$$

where

$$E(WTP | WTP > 0, \alpha, \beta) = \frac{\alpha}{\beta} + \frac{1}{\beta}\left(-\alpha + \frac{[1 + \exp(\alpha)] ln[1 + \exp(\alpha)]}{\exp(\alpha)}\right). \tag{19}$$

We now derive the result in 19. Recall the random utility shock is assumed to be i.i.d. logistic with PDF

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

and CDF

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

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

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

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

follows:

$$E\left(\Delta\varepsilon|\Delta\varepsilon\geq k\right) = [Pr\left(\Delta\varepsilon\geq k\right)]^{-1}\int_{k}^{-\infty}\Delta\varepsilon f\left(\Delta\varepsilon\right)d\Delta\varepsilon$$

$$= \left[\frac{exp(-k)}{1+exp(-k)}\right]^{-1}\int_{k}^{-\infty}\Delta\varepsilon\frac{exp(-\Delta\varepsilon)}{[1+exp(-\Delta\varepsilon)]^{2}}d\Delta\varepsilon$$

$$= \left[\frac{1+exp(-k)}{exp(-k)}\right]\left[\frac{kexp(-k)+[1+exp(-k)]ln[1+exp(-k)]}{1+exp(-k)}\right]$$

$$= k+\frac{[1+exp(-k)]ln[1+exp(-k)]}{exp(-k)}$$

where

$$\Delta\varepsilon\frac{exp\left(-\Delta\varepsilon\right)}{[1+exp\left(-\Delta\varepsilon\right)]^{2}}=\frac{d\left(-\frac{\Delta\varepsilon e(-\Delta\varepsilon)+[1+e(-\Delta\varepsilon)]ln[1+e(-\Delta\varepsilon)]}{[1+e(-\Delta\varepsilon)]}\right)}{d\Delta\varepsilon}.$$