Scalable Price Targeting [1]

Jean-Pierre Dubé, *Chicago Booth and NBER*
Sanjog Misra, *Chicago Booth*

June 2017

[1]Dubé: Booth School of Business, University of Chicago, 5807 S Woodlawn Ave, Chicago, IL 60637 (e-mail: jdube@chicagobooth.edu), Misra: Booth School of Business, University of Chicago, 5807 S Woodlawn Ave, Chicago, IL 60637 (e-mail: sanjog.misra@chicagobooth.ed); We are grateful to Ian Siegel and Jeff Zwelling of Ziprecruiter for their support of this project. We would also like to thank the the Ziprecruiter pricing team for their help and work in making the implementation of the field experiments possible. We are also extremely grateful for the extensive feedback and suggestions from Chris Hansen and Matt Taddy. Finally, we benefitted from the comments and suggestions of seminar participants at the Bridge Webinar Series at McGill University, Cornell University, Columbia GSB, Northwestern University, Penn State University, the 2017 Porter Conference at Northwestern University, the University of Chicago Booth School of Business, University of Notre Dame, UNC Chapel Hill, University of Rochester, University of Wisconsin, the 2017 Marketing and Economics Summit and the 2016 Digital Marketing Conference at Stanford GSB. Misra also acknowledges the support of the Kilts Center for Marketing and the Neubauer Faculty Foundation.

Abstract

We study the welfare implications of scalable price targeting, an extreme form of third-degree price discrimination, for a large, digital firm. Targeted prices are computed by solving the firm's Bayesian Decision-Theoretic pricing problem based on a database with a high-dimensional vector of customer features that are observed prior to the price quote. To identify the causal effect of price on demand, we first run a large, randomized price experiment. These data are used to train our demand model. We use lasso regularization to select the set of customer features that moderate the heterogeneous treatment effect of price on demand. We use a weighted likelihood Bayesian bootstrap to quantify the firm's approximate statistical uncertainty in demand and profitability. Theoretically, both firm and customer surplus could rise with scalable price targeting. We test the welfare implications out of sample with a second randomized price experiment with new customers. Optimized uniform pricing improves revenues by 64.9% relative to the control pricing, whereas scalable price targeting improves revenues by 81.5%. Customer surplus declines by less than 1% with price targeting; although nearly 70% of customers are charged less than the uniform price. Our weighted likelihood bootstrap estimator also predicts demand and demand uncertainty out of sample better than several alternative approaches. **Keywords: price discrimination, targeting, scalable price targeting, welfare, lasso regression, weighted likelihood bootstrap, data-mining, field experiment**

1 Introduction

A long literature has studied the theoretical implications of the use of targeted pricing, a form of "price discrimination" that varies the price charged to customers based on differences in their willingess-to-pay (e.g. Pigou (1920); Varian (1980); Stole (2007)). In practice, most targeted pricing practices are limited to third-degree price discrimination, using a coarse segmentation strategy that varies prices across broad groups of customers. Examples include seniors discounts at the movies and geographic or "zone" pricing by retailers across communities in a metropolitan area. Third degree price discrimination has been thought of as a practical way to increase firm profitability:

> "[The monopolist] cannot, except in extraordinary circumstances, introduce either the first or the second degree of discrimination, and that the third degree is of chief practical importance." (Pigou, 1920, Part II, chapter XVI, section 6)

Theorists have long recognized the possibility that with a very granular segmentation scheme, third-degree price discrimination could approximate first-degree, or "perfect," price discrimination:

> "Furthermore, 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 recent years, "...big data and electronic commerce have reduced the costs of targeting and first-degree price discrimination." (CEA, 2015, page 12). In this vein, we study *scalable price targeting* (SPT), a form of price discrimination that leverages the large quantities of observable customer features often available in digital environments to predict individual differences in willingness-to-pay. Like first-degree price discrimination, SPT consists of differential pricing across individual customers. However, SPT typically does not involve differential pricing across each marginal unit sold. In addition, statistical uncertainty typically limits the segmentation to an imperfect form of targetability. Customers with identical, obserable features are targeted the same prices even if they differ along unobservable dimensions. SPT is effectively an extreme form of third-degree price discrimination. To the best of our knowledge, the literature has not yet produced (experimental) field evidence that SPT generates incremental profits in practice.

The extant empirical literature on price targeting has developed econometric methods for devising targeted pricing mechanisms based on a retrospective analysis of detailed customer shopping histories (Rossi, McCulloch, and Allenby, 1996; Chintagunta, Dubé, and Goh, 2005; Shiller, 2015). The implications for targeted pricing are typically studied through model simulations based on demand estimates. These methods have limited applicability beyond markets for fast-moving consumer goods due to the limited availability of customer purchase panels in most markets. Surprisingly, more practical approaches that target based on observable customer features (as opposed to shopping histories) are considered to be of limited value. For example, Rossi, McCulloch, and Allenby (1996) conclude that "...it appears that demographic information is only of limited value" for the targeting 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]."

In addition to the perceived challenges in implementing a practical form of SPT, consumer advocates warn of the potentially harmful effects of price discrimination:

> "[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)

In fact, theoretically the move from uniform to SPT can increase total welfare (Varian, 1989), and can also increase total consumer surplus specifically (Cowan, 2012). The targeting of prices re-allocates value from strong consumers (who are targeted more than the uniform price) to weak consumers (who are charged less than the targeted price). Whether the re-allocation is sufficient to increase total consumer surplus is an empirical question regarding the curvature of demand.

We propose a practical and scalable approach to implement targeted pricing for a firm with access to a large cross-section of customer purchase data and detailed, customer-specific variables. On the demnd side, our approach is structural in the sense that we impose parametric structure on our demand model to ensure the necessary smoothness in prices for implementing price optimization by the firm. 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 characteristics. The firm's empirical goal consists of making statistical inferences about demand from heterogeneous customers, as opposed to making inferences about specific underlying parameters associated with customer characteristics. Thus, we cast our demand analysis as a heterogeneous treatment effects problem using price as a continuous treatment variable.

On the supply side, we accomodate the fact that the firm bases its pricing decisions on statistical estimates of demand. Following Wald (1950) and Savage (1954), we characterize the firm's pricing decision as a Bayesian Decision-Theoretic problem that uses posterior expected profits as the reward function. A simpler "plug-in" approach that optimizes prices based on point estimates of demand would fail to account for this uncertainty correctly and could lead to biased decisions.

To approximate the firm's statistical uncertainty about demand across customers with different observable profiles, we use a weighted likelihood bootstrap (WLB) version of a Lasso logistic regression. WLB provides us with approximate samples from the appropriate posterior density of the parameters of interest. We use these draws to quantify the uncertainty around the firm's demand and profits under different pricing decisions. In principle, we can compute the approximate posterior profitability of any desired optimized pricing structures using the results.

We implement the approach through a collaboration with Ziprecruiter.com, a large, online recruiting firm, to conduct a sequence of experiments with the goal of designing and implementing a real-time, scalable business-to-business pricing algorithm that optimized prices for each potential customer that arrived at their website. The analysis was conducted at the customer acquisition stage, with a focus on the segment of small, "starter" sized firms arriving at Ziprecruiter.com for the first time and seeking a price quote for the website's online job-posting services.

For the analysis, we developed a pricing analytics template consisting of three phases. In practice, price targeting requires estimating a model of demand with differences across customers in their price sensitivities: a heterogeneous treatment effects problem. Therefore, in phase I, we conducted a price experiment for the purposes of measuring the causal effect of price on purchase behavior. The experiment randomly assigned each new starter arriving at the website's paywall throughout the duration of the experimental period to one of ten price buckets ranging from \$19 to \$399, including a control condition of \$99 which was the firm's regular base price at that time. We ran the experiment for an entire month and collected between 750 and 825 starters (subjects) per cell. Descriptive analysis of the data revealed (i) model-free evidence for a downward-sloping demand relationship, (ii) that status quo pricing of \$99 was on the inelastic region of demand, and (iii) model-free evidence of an opportunity to raise prices profitably. Our model-free analysis of demand provides prima facie empirical evidence of the downward-sloping demand relationship in the field. 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 variables on demand (see Einav and Levin 2010). The fact that 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 (see Einav and Levin 2014).

In Phase II, we used the experimental data to estimate a demand model that calibrated price-response as a function of job and starter-firm characteristics. Demand estimation was conducted using the WLB estimator. In our study of "starter firms," we included 266 potential covariates in the model[1]. Our model estimates reveal a considerable degree of heterogeneity in willingness-to-pay across starter firms. The in-sample targeted prices that maximize the posterior expected profit from each starter range from as low as \$142 to as high as \$499; but all the prices exceeded the regular price of \$99[2]. Based on our optimization, we predicted expected gains posterior in profits of 56% and 80% for our uniform and targeted pricing structures, respectively, relative to Ziprecruiter's status quo price of \$99. Moreover, we predicted that our targeting scheme would capture over 40% of the potential profitability from the theoretical benchmark of perfect price discrimination.

[1]These variables were chosen from larger set of over 120,000 covariates available to the firm. This particular subset was relevant to the subset of customers involved in the experiment. 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, e.g. Taddy (2015a) successfully implements the approach in a distributed computing environment for applications with thousands of potential covariates.

[2]Ziprecruiter capped the targeted prices at \$499.

In Phase III, Ziprecruiter implemented a second field experiment with a new sample of prospective starter firms to test the recommended pricing structures against its status quo of $99 per month. The implementation computed a given starter's targeted price using a fast contraction-mapping that converged in less than 20 milliseconds and obviated the need for optimization software. This practical aspect played an important role for implementation since Ziprecruiter does not have optimization software integrated with its customer paywall. We observed profits gains of 68% and 84% for uniform and targeted pricing, respectively, relative to the status quo. Typically, researchers explore the potential counter-factual gains using simulations based on their model estimates. Our second experiment provides a novel opportunity to test the performance of our microeconomic-based counterfactuals out of sample (see also Misra and Nair 2011; Ostrovsky and Schwarz 2016[3]).

We also use Phase III to analyze the impact of SPT on consumer surplus. In the targeting cell of our experiment, 67% of the customers are targeted a lower price than the optimal uniform price. Even though total consumer surplus falls by a small amount (less than 1%), the majority of customers benefit from SPT. Interestingly, the strong customers who are targeted higher prices than the optimal uniform price nevertheless exhibit a higher conversion rate on average than weak customers.

Finally, we use the Phase III experimental data to evaluate the performance of our approximation of the uncertainty in our estimates. We apply the parameter estimates obtained from the Phase I "training data" to the distribution of customers observed in the Phase III experiment. The predicted uncertainty in profits are very close to the empirical distribution of "realized" profits, providing an out-of-sample assessment of our WLB method.

Our empirical analysis of individually-targeted prices builds on a growing empirical literature studying third-degree price discrimination by firms (see the survey by Verboven 2008). To the best of our knowledge, only a few studies have looked specifically at more individualized pricing whereby a firm targets different prices to each customer even though it lacks the complete information required for perfect (or first degree) price discrimination (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. An advantage of our data is that we observe the entire set of customer features that is available to the firm when it sets prices. Unlike this literature, we run field experiments both to estimate demand, to train our SPT structure and to test the performance of our approach out-of-sample, comparing conversion and profits relative to alternative pricing structures. In contrast to past work on SPT, we find that targeting on customer characteristics (as opposed to customer behavior) leads to substantial profit increases. Our work also contributes to the broader marketing literature on targeting prices and other non-price instruments to heterogeneous customers (e.g., Ansari and Mela, 2003; Simester, Sun, and Tsitsiklis, 2006; Dong, Manchanda, and Chintagunta, 2009; Kumar, Sriram, Luo, and Chintagunta, 2011).

[3] 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

Our analysis does not consider the possibility that a strategic customer might alter her behavior to game the firm's targeting strategy (e.g., Fudenberg and Villas-Boas, 2006). In our setting, it is unlikely that a customer would deliberately mis-report its business type at the registration stage since this could interfere with the quality of the service it receives on the platform. Moreover, it is unlikely that a customer would know which of its features to misreport since it would not be able to invert the firm's targeting strategy from the quoted price.

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; Athey and Imbens, 2016a). The extant literature has developed procedures for inference in the context of discrete (typically binary) treatment effects. In contrast, our SPT structure requires 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 2, we set up the prototypical decision-theoretic formulation of monopoly price targeting 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. In section 5, we explore several extensions of our findings and we conclude in section 6.

2 A Model of Decision-Theoretic Monopoly Price Targeting

In this section, we outline the key elements of a data-based approach to monopoly targeted pricing. 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 faces opportunity costs from sub-optimal pricing decisions in terms of missed potential profitability. However, the firm typically faces uncertainty about the sales and profit consequences associated with different prices. We treat the firm's uncertainty as statistical knowledge about customers 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 shortcomings of a simpler approach that "plugs in" point estimates of the uncertain quantities instead of using the full posterior distribution of beliefs. For an early application of Bayesian decision theory to pricing strategy see Green (1963). For a more formal econometric treatment of Bayesian decision-theoretic pricing that integrates consumer demand estimation, see Rossi, McCulloch, and Allenby (1996); Dubé, Fang, Fong, and Luo (2017)[4].

We start by describing the demand setup and defining the sources of statistical uncertainty regarding

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

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 its information (the customer dat), we then define several decision-theoretic (or "data-based") optimal pricing problems for the firm.

2.1 Demand

On the demand side, we start with 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:

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

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 $\varepsilon_i \sim i.id.\ F_\varepsilon(\varepsilon)$ is an i.i.d. random vector of unobserved, random disturbances that are independent of Ψ_i.

2.2 Firm Beliefs and Pricing

Suppose a 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 :

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

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

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

Using the terminology from the literature on price discrimination (e.g. Tirole, 1988; Pigou, 1920), we are technically implementing a form of *third-degree price discrimination*. 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. Unlike most practical implementations of third-degree price discrimination, however, our proposed approach will potentially allow for customer-specific, or "personalized" pricing (e.g. Shiller, 2015). However, in practice the pricing is not exactly personalized since customers with the same posterior expected Ψ_i would be charged the same price even if they differ along unobserved dimensions.

Suppose the firm uses a uniform pricing strategy across all its H customers. The posterior expected profit-maximizing prices if the firm uses uniform prices, p^*, must satisfy the following first-order necessary conditions:

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

The integration of the profit function over the firm's posterior distribution of beliefs adds computational complexity. Consider a simpler "plug-in" approach that instead maximizes the profits evaluated at point estimates of Ψ. For instance, consider the plug-in estimate of profits evaluated at the point estimates $\hat{\Psi}_i = E\left(\Psi|D\right)$:

$$\pi\left(p_i|\hat{\Psi}_i\right) = \left(p_i - c\right)' \bar{q}\left(p;\hat{\Psi}_i,\varepsilon\right). \quad (6)$$

with corresponding optimal *targeted prices* $\tilde{p}_i$, where

$$\tilde{p}_i = c - \left[\nabla_p \bar{q}\left(\tilde{p}_i;\Psi_i,\varepsilon\right)\right]^{-1} \bar{q}\left(\tilde{p}_i;\Psi_i,\varepsilon\right). \quad (7)$$

The price recommendation in equation 7 is computationally simpler to determine than those in 4 because the former avoids the integration of the profit function over the entire posterior distribution. However, by Jensen's inequality, we also know that in general the plug-in approach is biased:

$$\pi\left(p|\hat{\Psi}_i\right) \neq \pi\left(p|D\right).$$

This bias could lead to the manager misestimating the degree of uncertainty and consequently setting suboptimal pricing rules. In our empirical case study below, we will analyze the extent of bias associated the plug-in approach.

2.3 Welfare

For our monopoly SPT, we know by revealed preference that the firm's profits must increase weakly. This is because the optimal targeted prices in 4 could always accommodate charging every customer the uniform price in 5: $p_i^* = p^*, \forall i$.

The impact of targeted prices on consumer surplus is less straightforward. Cowan (2012) proposes a novel approach that interprets the shift from uniform prices, p^*, to a targeted price, p_i^* as an equivalent change in marginal cost. Under targeted prices, the firm sets the price p_i^* to equate the marginal revenue from customer i with the marginal cost, c. Under the uniform price, the firm sets the price p^*, which generates marginal revenue $\bar{MR}_i$from customer i. We can therefore think of the shift from p^* to p_i^* as the equivalent change in price associated with a cost shift from c to $\bar{MR}_i$: $p_i(\bar{MR}_i)$ versus $p_i(c)$. This interpetation of targeting as an equivalent cost shift enables the use of pass-through comparative statics, as in Weyl and Fabinger (2011).

Following Cowan (2012), we define the following pass-through condition whereby the price-elasticity of demand is larger in magnitude than the curvature of the pass-through over the range from c to $\bar{MR}_i$:

$$\left|\eta_{i,p}\right| > p_i(k)\frac{p_i''(k)}{\left(p_i'(k)\right)^2} \tag{8}$$

where $p_i(k)$ is the optimal price charged to consumer i at virtual cost k. When the pass-through condition holds, the consumer surplus function $CS(p_i(k))$ is convex in the virtual cost k. The change in total consumer surplus can therefore be bounded:

$$\sum_i (\bar{MR}_i - c)\, p_i'(c)\, \bar{q}_i(c) \geq \Delta CS \geq \sum_i (\bar{MR}_i - c)\, p_i'(\bar{MR}_i)\, \bar{q}_i(p_i(\bar{MR}_i)). \tag{9}$$

For many demand models, including the one we study in section 4.1 below, the lower bound can be positive, implying that consumer surplus can increase theoretically.

3 Empirical Approach

The execution of the firm's data-based pricing strategies in equations 4 and 5 depends on the ability to construct an estimate of the posterior distribution $F(\Psi_i|D)$. The extant literature on price targeting 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.[5] We consider a scenario with cross-sectional customer

[5]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.

information that includes a detailed set of observable customer characteristics. Our approach consists of using these characteristics to approximate Ψ_i.

3.1 Approximating Individual Types

Suppose we observe data

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

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 $x_i \in \mathbb{R}^K$ 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, hence, targeted prices.

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 for large K and $K \ll N$, 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.

To address these practicality concerns, we use an idea from Taddy, Gardner, Chen, and Draper (2016) to approximate the posterior $F_\Psi(\Psi|D)$ using a variant of the Bayesian Bootstrap (e.g. Rubin (1981); Newton and Raftery (1994); Chamberlain and Imbens (2003); Efron (2012)). The approach generates both a point estimate of Ψ and an approximate sample from the full posterior distribution $F_\Psi(\Psi|\mathbf{D})$. The approach provides an approximation of the posterior distribution required for the decision-theoretic pricing problem. In addition, the approach does not require large-sample approximations. Alternative approaches using asymptotic approximations have been developed but are cumbersome to implement and not easily scalable to the types of scenarios discussed in this paper.

3.2.1 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:

$$F_\Theta(\Theta|D) \propto \ell(D|\Theta) - \sum_{j=1}^{J} \tau_j |\Theta_j| \tag{10}$$

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 (10) can be shown to be equivalent to the Lasso regression:

$$\Theta^{Lasso} = \underset{\Theta \in \mathbb{R}^J}{argmax} \left\{ \ell(D|\Theta) - N\lambda \sum_{j=1}^{J} |\Theta_j| \right\}. \tag{11}$$

In Appendix A, 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)).

3.2.2 Quantifying Uncertainty

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[6]. 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). As discussed in

[6]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.

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 (11). The application of Lasso to each replication will ensure a sparsity structure that facilitates the storage of the draws in memory. This approach has also found its way into econometrics (see e.g. Chamberlain and Imbens (2003)) and is a promising approach to approximating uncertainty in complex models.

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: $\left\{V_i^b\right\}_{i=1}^{N} \sim Exp(\mathbf{1}_N)$
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 (20) in Appendix A
   (a) Construct the regularization path, $\left\{\hat{\Theta}^b|_{\lambda}\right\}_{\lambda=\lambda_1}^{\lambda_T}$
   (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, $\left\{\hat{\Theta}^b\right\}_{b=1}^{B}$, to simulate the posterior of interest, $F_{\Psi}(\Psi_i)$. We construct draws $\left\{\Psi_i^b\right\}_{b=1}^{B}$, where $\Psi_i^b = \Psi\left(x_i; \Theta^b\right)$, 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.

Discussion

One useful interpretation of our proposed model would have us consider the Lasso penalization $\left(\lambda \sum_{j=1}^{J} |\Theta_j|\right)$ as well as the Dirichlet weighting $(f(D))$ as components of our overall prior. Under this interpretation, the proposed sampling algorithm obtains approximate samples, $\left\{\hat{\Theta}^b\right\}_{b=1}^{B}$, from the posterior of interest. While the prior used here is non-standard the framework is coherent from a Bayesian perspective.

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 (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 would use a Lasso to select the relevant model (i.e. the subset of relevant x) and in a second step we would obtain parameter estimates after conditioning on this subset. We term this procedure Post-Lasso-MLE and use it as a benchmark in later sections. The post-Lasso-MLE is somewhat of a strawman 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 Scalable Price Targeting at Ziprecruiter.com

We conduct a case study of targeted pricing at Ziprecruiter.com to illustrate the implementation of the WLB estimator, the application to SPT and to validate our proposed approximation of the posterior of Θ. The case study involves a sequence of two randomized controlled price experiment using a sample of Ziprecruiter's prospective enterprise customers. The data from the first experiment are used to train our demand model and to produce price recommendations. A second experiment is then conducted using a new sample of Ziprecruiter's prospective enterprise customers to validate our recommended pricing structures as well as our inference procedure.

Ziprecruiter.com is an online firm that specializes in matching jobseekers to potential employees. The firm caters to a variety of potential employers across various industries that subscribe to Ziprecruiter.com to gain access to a stream of resumes of matched and qualified candidates from which they might be able to recruit. These firms pay a monthly subscription rate that they can cancel at anytime. Job applicants can use the Ziprecruiter.com platform for free. In a typical month in 2015, Ziprecruiter hosted job postings for over 40,000 registered paying firms.

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 less than 5 employees. Since starters represent nearly 50% of the customer base, we focus our attention on prospective starter firms. At the beginning of this project the base rate for a "starter" firm (a small business with less than 5 employees) 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 firm is asked to input details of the type of jobs they wish to fill as well as characteristics describing the firm itself. During the registration process, the customer reports several characteristics of its business and the specific job posting. Table 2 summarizes the variables we retained for our analysis from the much larger set of registration features. While the set looks small, it generates 133 variables[7]. After completing this registration process, the customer reaches a paywall and receives a price quote. Ziprecruiter currently uses a non-linear price schedule based on the number of months of service for which a new customer is willing to pre-commit to service. The first row of Table 1 reports Ziprecruiter's regular pricing schedule used prior to the experimental period.

4.1 Empirical Model of Demand

In our case study of prospective customers that have reached Ziprecruiter's paywall, demand consists of a binary decision $y_i = 1$ (if purchase) or 0 (if do not purchase). A customer i obtains the following incremental utility from purchasing versus not purchasing

$$\begin{aligned} \Delta U_i &= \alpha_i + \beta_i p_i + \varepsilon_i \\ &= \alpha\left(x_i;\theta_\alpha\right) + \beta\left(x_i;\theta_\beta\right) p_i + \varepsilon_i \end{aligned} \tag{12}$$

where $\alpha\left(x_i;\theta_\alpha\right)$ is an intercept and $\beta\left(x_i;\theta_\beta\right)$ is a slope associated with the price, P_i. To conform with our notation in section 2, we re-write equation 12 as follows

$$\Delta U_i = \tilde{p}_i' \Psi_i + \varepsilon_i \tag{13}$$

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

[7]We used marginal regressions to select these variables for the demand analysis

A customer i has the following probability of buying conditional on x_i

$$\begin{aligned}\mathbb{P}(x_i, p_i; \Psi_i) &= \int 1(\Delta U_i > 0)\, dF_\varepsilon(\varepsilon_i) \\ &= 1 - F_\varepsilon\left(-\tilde{p}_i' \Psi_i\right).\end{aligned}$$

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

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

However, in principle, one could use any arbitrary function of x_i. 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

$$\mathbb{P}(p_i; \Psi_i) = \frac{\exp(\tilde{p}_i' \Psi_i)}{1 + \exp\left(\tilde{p}_i' \Psi_i\right)}. \tag{14}$$

Note that for our demand specification, the treatment effect of price on choice is continuous. In most data-mining applications, variables are treated as categorical. Our structural approach, which will involve optimizing the price on the supply side, motivates our use of a smooth and continuous price effect.

4.2 Pricing

Suppose Ziprecruiter collects a database for a sample of N consumers, $D = (D_1, ..., D_N)$, where $D_i = (y_i, x_i, p_i)$. Suppose also that Ziprecruiter uses the WLB approach described in section 3.2 to estimate the posterior beliefs about the demand parameters, $F_\Psi(\Psi_i|\mathbf{D})$. For SPT, we use the following contraction mapping to enable the real-time calculation of customer-specific prices when a new customer reaches the paywall at ziprecruiter.com. We start with an initial guess p_0 and then iterate the following sequence

$$p_i^{k+1} = c - \frac{\int \mathbb{P}\left(p_i^k; \Psi_i\right) F_\Psi(\Psi_i|\mathbf{D})\, d\Psi_i}{\int \frac{\partial \mathbb{P}\left(p_i^k; \Psi_i\right)}{\partial p} F_\Psi(\Psi_i|\mathbf{D})\, d\Psi_i} \tag{15}$$

until $|p_i^{k+1} - p_i^k| < 1.e-6$. We simulate the integrals over the posterior, $F_\Psi(\Psi_i|D)$ using our WLB draws $\left\{\Psi_i^b\right\}_{b=1}^B$. Using Ziprecruiter's online system, the evaluation of a typical prospective customer's optimal targeted price takes approximately 18.6 microseconds using (15) above. Therefore, the approach is not only fast enough for real-time implementation, it also obviates the need for integrating optimization software with Ziprecruiter's paywall.

4.3 Customer Surplus

We now revisit the lower bound on the change in consumer surplus when the firm switches from uniform to SPT under our logit demand. We assume throughout that marginal cost is 0. First consider the case where there are two customers and the firm has resolved all the uncertainty in the demand parameters so that its posterior is degenerate at $\alpha(x) = (-0.9, 0.3)'$ and $\beta(x) = (-0.01, -0.011)'$. It is straightforward in this case to show that the pass-through condition holds as long as the purchase probability is less than 0.5. The purchase probability is less than 0.3 for both consumers under both uniform pricing and SPT. Moreover, the lower bound on ΔCS is 0.024 and, thus, total consumer surplus and total welfare increase. In fact, consumer surplus increases by \$0.038.

Now suppose the firm faces uncertainty in the demand parameters. For instance, suppose the uncertainty consists of additive Gaussian noise such that $\tilde{\alpha}_i(x) \sim N\left(\begin{bmatrix} \alpha_i(x) \\ \beta_i(x) \end{bmatrix}, \begin{bmatrix} 1.e-4 & 0 \\ 0 & 1.e-6 \end{bmatrix}\right)$ (to ensure that most of the posterior mass over β is negative we set the standard error to about $\frac{1}{10^{th}}$ of the value of the mean). In this case, we verify numerically that the pass-through condition holds over the range of interest. The lower bound on ΔCS is 0.023 and, thus, total consumer surplus and total welfare increase. In fact, consumer surplus increases by \$0.034. These examples illustrate that, theoretically, SPT can increase total consumer surplus for our framework.

4.4 Experiment One

The first phase of the case study consists of a price experiment to generate choice data with exogenous price variation. The 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 pricing schedule, row one of Table 1. To construct our test cells, we changed the monthly rate by some percentage amount relative to the control cell. The corresponding quarterly and annual rates were computed by using the same percentage deviation from 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 1.

4.4.1 Model-free analysis

The results from the first stage experiment appear 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. This means that Ziprecruiter is likely under-pricing its service.

In practice, many firms are reluctant to run field experiments because of the opportunity costs of testing a sub-optimal price (Anderson and Simester (2011)). Figure 2 plots Ziprecruiter's monthly revenues per customer at each tested price level. Interestingly, the experiment itself generated incremental revenues for Ziprecruiter. By running the experiment, Ziprecruiter increased the average monthly revenue per prospective customer by 14% relative to what it would have earned had it charged everyone \$99. The incremental profitability of the higher tested price levels more than offset the high conversion at extremely small test price levels that are well below the control level of \$99.

Figure 2 visualizes Ziprecruiter's pricing incentives. Along our grid of tested price levels, the average monthly revenue per prospective customer is maximized \$399. Although, once we take into account statistical uncertainty, we cannot rule out that the revenue-maximizing price lies somewhere between \$249 and \$399. Ziprecruiter could increase its monthly revenues from new customers by raising its prices by more than 100%. However, the experiment alone may be insufficient to help Ziprecruiter determine the optimal price increase. A model is ultimately needed to design the optimized pricing structures.

4.4.2 Demand estimation

The second phase of the case study consists of using the choice data from the field experiment to estimate the Logit demand model using our WLB estimator discussed in section 4.1[8]. The price experiment avoids the usual concerns about price endogeneity that plague the demand estimation literature. During the registration stage, our prospective customers provided responses on 12 categorical variables. We break the different levels of these variables into 133 dummy variables, 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 slopes, β. For comparison, we also report results for the MLE estimates of a model including all 266 covariates, 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. In section 5.1 below, we apply the WLB to a much higher-dimensional x_i vector using all the main effects and interaction effects. [9]

In Table 3, we report the Bayesian Information Criterion (BIC) associated with the MLE estimator that includes all 266 coefficients and with the Lasso estimator. The BIC includes a penalty for the number of model parameters. We also 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, $F(\Theta)$. As expected, the switch from MLE to Lasso improves the in-sample BIC considerably: 10,018 versus 8,366. This improvement suggests over-fitting with the MLE. Across our 100 bootstrap replications, our WLB

[8]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.

[9]The Lasso algorithm can easily accommodate much larger dimensions in a distributed computing environment. For instance, Taddy (2015a) presents a case study with 11,940 attribute dimensions.

estimator produces a range of BIC values from 7,805 to 8,940.

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 “price”, “job in British Columbia”, “company type: staffing agency,” “employment type: full_time” and “is resume required.” There is no *a priori* “obvious” candidate types of variables that emerge suggesting that the variable selection is important.

As an additional verification, we also examine the out-of-sample fit of each of our estimators. We randomly sample 90% of the firms (with replacement) from the original 7,866 as a training sample. The remaining 10% of firms are held out as a prediction sample. The second column of Table 3 reports the out-of-sample fit for each estimator. Once again, the entire range of BIC values from the WLB is below the BIC of the MLE. These findings are consistent with our concern that MLE may suffer from over-fitting, generating potentially less reliable estimates of the firm’s posterior uncertainty.

4.4.3 Price Optimization

We now use our demand estimates to calibrate Ziprecruiter’s decision-theoretic price optimization problem. 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. During the price optimization procedures, we drop any draws for which $\hat{\beta}(x) \geq 0$ or $\frac{\hat{\alpha}(x)}{\hat{\beta}(x)} \geq 2000$[10]. A summary of the various pricing scenarios analyzed is provided in Table 4.

We begin with an analysis of optimal uniform pricing. At the current price of \$99, the posterior expected own-price elasticity of demand is only -0.36 with a 90% posterior credibility interval of (-0.43,-0.3). Consistent with our model-free analysis above, Ziprecruiter.com is pricing on the inelastic region of demand. Optimal pricing for an information good like Ziprecruiter would be set at the unit-elastic point of demand. Recall from Figure 2 that the revenue-maximizing price appears to lie between \$249 and \$399. We can rule out \$399 as being too high since there is close to a 100% posterior probability that the own-elasticity is well above -1 at that point. At a price of \$249, the posterior expected own-price elasticity is -0.91 and the 90% posterior credibility interval is (-1.09,-0.74). Therefore, we cannot rule out the probability that this price maximizes expected revenues. More formally, the optimized uniform price, as defined in equation 5, is \$280.52. Figure 4 displays Ziprecruiter’s posterior expected revenue function under uniform pricing. The plot visualizes that Ziprecruiter is currently underpricing its service by nearly 63%, when charging \$99 instead of \$280.53.

An important component of the decision-theoretic approach consists of the integration of profits over the firm’s posterior distribution $F_{\Psi}(\Psi|D)$. As explained in section 2.2, a simpler plug-in approach will be biased towards lower profitability, possibly leading to under-pricing. The plug-in approach produces

[10]In total, we only drop 6% of the posterior draws across Ψ_i^b for all $i = 1, ..., N$ and $b = 1, ..., B$..

a recommended price of \$241. If we compute the posterior profits at a price of \$241, the range is not very different from the range in posterior profits at our WLB-based optimized price of \$281. There is nevertheless a 96% posterior probability that \$281 is more profitable than \$241.

In fact, Ziprecruiter subsequently decided to implement a uniform price of \$249 instead of \$281. Taking into account the parameter uncertainty, there is a 96% probability that the \$218 price is more profitable than the \$249 price. But, Table 4 indicates that the two prices produce very similar ranges of posterior profits at the 95% credibility level. Ziprecruiter concluded that \$249 was a more conservative recommendation.

We now explore targeted pricing. Figure 3 summarizes the estimates of heterogeneity. In panel (a), we report the distribution of customers' posterior mean price sensitivities[11]

$$\hat{\beta}_i = \frac{1}{R}\sum_{r=1}^{R} x_i \hat{\beta}^r.$$

The dispersion across customers suggests a potential opportunity for Ziprecruiter to price discriminate. In panel (b), we report the distribution of customers' posterior mean surplus when Ziprecruiter prices its monthly service at \$99:

$$W\hat{T}P_i = \frac{1}{R}\sum_{r=1}^{R} \frac{\log\left(1+exp\left(x_i{}'\hat{\alpha}^r - \$99 \times x_i'\hat{\beta}^r\right)\right)}{\hat{\beta}^r}.$$

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). Panel (b) illustrates the wide dispersion in value consumers derive from the availability of Ziprecruiter when it costs \$99. The 2.5^{th} percentile, median and 97.5^{th} percentile customers derive \$20.49, \$74.84 and \$280.95 in surplus respectively. The magnitudes and degree of dispersion in value indicate an opportunity for Ziprecruiter to target different prices across customers reflecting differences in the value they derive from the service.

Figure 5 summarizes the targeted pricing results. Ziprecruiter wanted to ensure that the targeted prices seemed natural to customers and also did not create a back-lash. Hence, they rounded all the targeted prices down to the closest \$9. For instance, a targeted price of \$251 would be rounded down to \$249. They also capped the prices at \$499. We use our demand estimates to assess the predicted performance of this scheme for our September 2015 customer sample. We observe considerable dispersion in the targeted prices, ranging from as low as \$119 to as high as \$499. The upper bound of \$499 binds for 455 of our customers, or 5% of the sample. All of the targeted prices are strictly larger than the \$99 baseline price. Interestingly, the mean targeted price, \$272.95, is almost identical to the optimized uniform price, \$280.57. Figure 6 plots the relationship between the estimated price sensitivity of each

[11] For each customer, we drop positive draws, i.e. we do not average over draws r where $x_i\hat{\beta}^r < 0$. This trimming is important for the pricing analysis since positive support of the price sensitivity will lead to unbounded pricing. Across our entire sample of customers, we end up dropping only 4.5% of the draws. Without trimming, only 15 of our 7,867 customers would have a positive posterior mean price sensitivity (about 0.19% of our sample).

customer and the corresponding targeted price. As expected, we observe a strong positive correlation between the targeted prices are the price sensitivities. In Table 4, we compare the profits for the Implemented Targeting scheme and the theoretical Targeting scheme without any rounding or capping. The Implemented Targeting scheme While there is a 95% posterior probability that the unrestricted Targeted prices are more profitable than the Implemented Targeted prices, the expected profit difference is only about 4%. Ziprecruiter concluded that this small difference justified the simplicity of the implemented scheme.

Once again, we can compare our decision-theoretic price recommendations to a plug-in approach. Figure 7 displays the density of targeted prices using our decision-theoretic approach and the WLB characterization of uncertainty. The figure also displays the targeted prices using the plug-in approach. As expected, the distribution of prices is shifted to the left using the plug-in estimator, which (by Jensen's Inequality) under-estimates the posterior profitability at any given price. There is a 99% posterior probability that our WLB-based targeted prices generate higher overall profits than the plug-in based targeted prices. In spite of this bias, all of the prices are strictly greater than $99.

We now compare the expected posterior profits per customer from our various pricing structures. The posterior mean profits from the implemented uniform price of $249 and the implemented targeted prices are 56% higher and 71% higher respectively than the profits under the control monthly price level of $99. Taking into account our posterior statistical uncertainty around the parameter estimates, there is a more than 99% posterior probability that baseline profits are lower than uniform and targeted profits, respectively. In the next section, we discuss the follow-up experiment to test the relative profitability of these three pricing structures.

Based on conversations with Ziprecruiter management, we also do not expect any competitive response from other platforms. Since our recommendations involve raising prices, 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 targeting is based on a complex array of customer characteristics, it also seems unlikely that our SPT structure would lead to unintended strategic behavior by Ziprecruiter's customers (e.g., (Fudenberg and Villas-Boas, 2006; Chen, Li, and Sun, 2015)). Moreover, customers need to report their registration characteristics truthfully to ensure that Ziprecruiter's matching algorithm identifies the most appropriate CVs for recruiting purposes.

4.4.4 Degree of Targetability

As explained above, our proposed targeting scheme is imperfect in the sense that we cannot estimate a prospective customer's logistically-distributed idiosyncratic utility shock, ε, as in equation 12. Therefore, our targeted pricing structure, while granular, is a form of third-degree price discrimination. Any set of customers with the same observable traits, x, would all be targeted the same price. We now assess our targeting scheme by comparing it to the theoretical benchmark of perfect price discrimination, or first-degree price discrimination.

Suppose the firm was able to estimate each customer's utility shock, ε. Consumer i's maximum willingness-to-pay (WTP) for Ziprecruiter service is

$$WTP_i = \frac{(\alpha(x_i) + \varepsilon)}{\beta(x_i)}. \tag{16}$$

Under perfect price discrimination, the firm would set the targeted price

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

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

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

$$\begin{aligned} Pr\left(buy|p, = p^{PD}\alpha, \beta\right) &= Pr(WTP \geq 0) \\ &= 1 - \tfrac{1}{1+exp(\alpha)}. \end{aligned} \tag{17}$$

The corresponding expected profit from this consumer is

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

In Appendix B, we show that

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

We can now assess how well our proposed targeting scheme performs relative to the theoretical benchmark of perfect price discrimination. In the final row of Table 4, we report the results if the firm was able to price discriminate. The expected conversion, equation 17, increases considerably, more than double the rate under targeted pricing, since every consumer with a positive WTP would buy. The expected profit per lead, equation 18, also increases considerably to \$98.58. Nevertheless, our proposed targeting scheme is expected to generate 46% of the potential profits under perfect price discrimination. There is a 90% posterior probability that our proposed targeting structure could generate as much as 55% of the profits under perfect price discrimination. These profit differences are visualized in Figure 8 where we plot the posterior CDF of profits in our control, Implemented Uniform and Implemented Targeting pricing structures respectively. In sum, targeting on the observed customer features at the registration stage explain almost half of the customer willingness-to-pay according to our model estimates.

4.5 Experiment Two

The third phase consisted of a second price experiment to validate the price recommendations out of sample and to validate the approximate inference procedure. The experiment was conducted 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 during this period was randomly assigned to one of the three following pricing structures:

1. Control pricing – $99 (25%)
2. Uniform pricing – $249 (25%)
3. Targeted pricing (50%).

We over-sampled the targeted pricing cell to obtain more precision given the dispersion in prices charged across customers. For our optimal uniform pricing cell, all customers were charged a monthly rate of $249. This price was chosen given the fact that (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 targeted pricing cell, customers were targeted a price based on the values of x_i they reported during the registration stage. As explained in the previous section, we then rounded the targeted price down to the nearest $9, discretizing the targeted prices into $10 buckets ranging from $119 to $499. For instance, a customer with a targeted price of $183 would be charged $179.

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 targeted pricing. In the targeting cell, the lowest targeted 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 targeted prices that would have been used had we implemented our targeting method in each cell. Figure 9 reports the density of targeted prices in each cell. 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 targeted prices would look the same across each of the experimental cells.

4.5.1 Model-free analysis

We begin by comparing the realized conversion and subscription revenue across our three pricing structures, control ($99), Optimal Uniform ($249) and SPT. To account for sampling error in our analysis, we bootstrap our sample 1,000 times (sampling with replacement).

Results are summarized in Table 5. As expected, average conversion is higher in the control cell which has the lowest monthly price. Average conversion is almost identical in the uniform and targeted cells, at 15%. However, the average profit per customer is higher in the targeted 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%. Targeted 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 targeting exceed our predictions based on the September sample discussed above in section 4.4.3. Finally, although not reported, our bootstrap generates an 87% probability that targeted profits will exceed uniform profits. These profit differences are visualized in Figure 10 where we plot the posterior CDF of profits in our control, Uniform and Targeted pricing structures respectively. The CDF is computed using our bootstrap draws of the mean profits per customer. Although not reported in the table, a Kolmogorov-Smirnov test easily rejects the hypothesis of identical profit distributions for control and uniform ($p < 0.01$) and of identical profit distributions for uniform and targeted ($p < .01$).

In sum, the November experiment demonstrates the large, permanent increase in profitability achievable by optimized prices and, moreover, by targeting different prices across customers based on their identifiable traits at the registration stage. The targeting scenario performs even better than we had predicted based on our September sample.

4.5.2 Welfare

To analyze the impact of SPT on customer surplus, we focus on the 2,485 customers assigned to the targeting cell. Using the model estimates, we use (9) to compute the bounds on the change in consumer surplus associated with switching from the optimal uniform price ($281) to the optimized targeted prices. Over 98% of the customers satisfy the pass-through condition in (8). We obtain an upper bound of -$3.39 and hence customer surplus is predicted to fall for this sample of customers.

Another advantage of our November experiment is that we can analyze the "actual" behavior of targeted customers. Recall that customers were in fact charged a simplified version of the targeted prices, rounded as explained above. Even though total surplus is predicted to fall, 67% of the prices targeted to these customers are lower than the optimal uniform price. Therefore, SPT strictly benefits the majority of the customers. Furthermore, total predicted conversion increases by close to 1%, meaning that more of the market will likely be covered under SPT. Figure 12 reports the total surplus across all customers assigned to each targeted price cell. As a comparison, we also report the total surplus for those customers had they instead been charged the optimal uniform price. The figure indicates that a small group of customers with very high willingness-to-pay ($499 or above) are subsidizing the majority of customers who are targeted a price less than the uniform rate.

In Figure 13, we look at the realized conversion rate in each cell. In spite of the fact that strong customers subsidize weak customers, the realized conversion rate is actually higher for the strong customers

(16.54%) than for the weak customers (13.9%). Moreover, for the extreme strong customers targeted a price of $499, conversion is higher than in any of the weak customer cells, even though most of the latter are charged prices that are less than half of $499.

Interestingly, there is no *a priori* obvious type of variable that drives the differences in targeted prices. As an exporatory exercise, we correlate the 133 non-price registration features with an indicator for whether each of the 2,485 firms is charged a targeted price lower than the optimal uniform price. The features "Company Type: Small" and "Employment Type: part time" both correlate positively with being targeted prices lower than uniform (correlations of 0.22 and 0.23 respectively). Interestingly, several of the features related to job benefits are strongly negatively correlated with being targeted prices below uniform: "job total benefits," "job medical benefit," "job vision benefit" and "job dental benefit" (with correlations of -0.48, -0.44, -0.40 and -0.44 respectively). These diverse findings suggest an important role for variable selection in determining which of the 133 registration features is best-suited to price targeting.

4.6 Validation of the Proposed Inference Procedure

We now compare the predictions and sampling properties from our WLB estimates and the realized outcomes from the November data. These comparisons allow us to judge how well our proposed WLB approach worked. Since the sample of prospective customers changes in November 2015, we apply the WLB estimates obtained from the September 2015 sample to predict the purchase behavior for the November 2015 sample. Table 6 summarizes our predictions for conversion and profits per customer. The profit predictions are comparable to our predictions from the end of September (see Table 4). The posterior mean conversions do not differ by more than 1 percentage point across cells. The posterior mean profits never differ by more than $1 across cells. Most important, our posterior credibility intervals on profits are very similar, suggesting that the population of prospective customers in November is not too different from the training sample in September.

By comparing Table 6 and Table 5, we can evaluate the properties of our inference approach. The realized mean profits per customer in each of the three cells (Table 5) falls within the predicted 95% credibility intervals for each of the cells (6). The predicted mean conversion rates are also very close to the realized mean conversion rates and fall within the predicted 95% credibility intervals. In sum, the WLB inference approach appears to have produced reliable predictions regarding both conversion and profits in each of the cells.

In Figure 11, we compare the empirical distribution of the realized conversion rates in each of the November pricing structure test cells to the predicted distribution using WLB, post-Lasso MLE and MLE respectively. As described earlier, the post-Lasso MLE follows a two-step approach - the first step implements a Lasso to select the relevant model (i.e. the subset of relevant x) and in a second step obtain parameter estimates after conditioning on this identified subset. The MLE estimator simply uses all available covariates (feasible for the current problem). All confidence intervals for these methods rely on standard Bootstraps. Each panel compares the densities of conversion for each of our compared methods in a given pricing cell. For SPT, we report densities for 6 of the 39 price tiers. The density of realized conversion rates is computed by bootstrapping with replacement from the November data in a given cell.

The figures indicate a relatively good match between our approximate posterior using WLB and the actual observed data. In contrast, the post-Lasso MLE approach predicts considerably less uncertainty than our WLB approach. The post-Lasso MLE would likely lead to managerial over-confidence when compared to the actual conversion rates, which exhibit much more variation. This overconfidence is particularly striking under SPT, where we have much smaller samples for each of the targeted price tiers. Furthermore, comparing to the actual mean conversion, the mean conversion under post-Lasso MLE is systematically less accurate than for the WLB. The figure illustrated additional out-of-sample performance for our WLB procedure.

In each of the three panels, we also report the Kullback-Leibler divergence measure for (1) WLB relative to the true distribution ($D_{KL}(true||WLD)$), and (2) post-Lasso MLE relative to the true distribution ($D_{KL}(true||post - Lasso\, MLE)$). The KLD

$$D_{KL}(A||B) = \int_{\Theta} a(\theta)\, log\left(\frac{a(\Theta)}{b(\Theta)}\right),$$

where a and b denote the densities of A and B respectively, measures the amount of information lost when distribution B is used to approximate distribution A. We find that the KLD is considerably higher for post-Lasso MLE than for WLB, suggesting that WLB is a much better approximation of the true distribution of conversion. Across each of the panels, the percentage difference between the KLD for post-Lasso MLE and for WLB ranges from 50% to 746%. Perhaps not surprisingly, the largest improvements for WLB arise in the control and uniform pricing cells where we have more observations per cell.

The relatively poor performance of post-Lasso MLE reveals the important roles of both variable selection and model uncertainty. Even when we take the best features from the Lasso, the corresponding MLE still performs worse than WLB both on prediction and uncertainty quantification. Although not reported herein, a naive approach that includes all the features in the MLE leads to considerably worse prediction and uncertainty quantification. These findings indicate that price targeting based on registration features is a big data problem for Ziprecruiter.

5 Robustness of Results

5.1 Higher-dimensional Characteristics

Design of second experiment targets based on observables, x_i. So we can evaluate additional pricing structures out of sample. We focus on more granular targeted pricing using higher-dimensional x_i that include all the interaction effects. Our results suggest that the incorporating interactions has limited incremental power. More details to be added.

5.2 Lifetime Value of the Customer

Our analysis of the September 2015 sample above, in section 4.4.3, was based on myopic pricing geared towards instantaneous profits. A potential concern is that raising the price, thereby lowering conversion, might be to lower long-term profitability if there is sufficient repeat business. In this section, we re-investigate the September sample based on four months of renewal behavior for each starter firm up to the end of December 2015.

Figure 14 reports the expected net present value of profits in September over a 4-month horizon. The top panel assumes a discount factor of $\delta = 0$ and corresponds to our static analysis above. The bottom panel assumes a discount factor of $\delta = 0.996$ and assumes a monthly interest rate of 0.4% (or an annual interest rate of 5%). While the net present value of profits is much higher in each cell under $\delta = 0.996$, our ranking of prices is quite similar. To understand this finding, it is helpful to look at both the initial conversion rate along with the retention rates. In Table 7, we report the acquisition and retention rates for each of the experimental price cells. 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. As a result, the optimal Uniform price does not look much different from the myopic (one-month-horizon) case.

6 Conclusions

A long theoretical literature has studied the potential profit improvements associated with monopoly price discrimination. While the extant literature has developed methods to estimate demand and simulate the profitability of price discrimination, we are not aware of a field implementation that demonstrates these profit improvements out of sample. Through a sequence of pricing field experiments, we measure the realized profit improvements from SPT by a large online recruiting company. We find that Ziprecruiter's status quo pricing was on the inelastic region of demand. A substantially higher, uniform optimal price raises the firm's monthly profits by over 60% out of sample. Using machine learning methods, we design a SPT scheme that is fast and scalable. SPT raises monthly profits by about 80% out of sample. Our field experiment also allows us to validate the properties of our proposed method out of sample.

In addition to our substantive evidence, we have also developed a Bayesian Decision Theoretic scalable price targeting method that can accommodate large-dimensional, observable heterogeneity. The approach bridges basic microeconomic principles with machine learning in a manner that is practical and scalable. The approach is potentially generalizable to more complex demand environments with multiple products and non-discrete-choice. An interesting extension would be the application of the method to an oligopolistic setting in which the firm not only faces uncertainty about demand, it also faces uncertainty about its rival's likely actions.

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.

References

ANDERSON, E., AND D. I. SIMESTER (2011): "A Step-by-Step Guide to Smart Business Experiments," *Harvard Business Review*, pp. 1–9.

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.

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

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.

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

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.

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.

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.

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.

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.

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.

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*.

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.

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.

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.

WEYL, E. G., AND M. FABINGER (2011): "A Restatement of the Theory of Monopoly," *SSRN Working Paper*.

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.

	Monthly	*Quarterly*	*Annual*
Control	99	249	590
Test 1	19	49	119
Test 2	39	99	239
Test 3	59	149	359
Test 4	79	199	479
Test 5	159	399	999
Test 6	199	499	1199
Test 7	249	629	1499
Test 8	299	759	1789
Test 9	399	999	2379

Table 1: Experimental Price Cells for Stage One

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

Table 2: Company/Job Variables

Model	**In-Sample BIC**	**Out-of-Sample BIC**
MLE	10018.78	4430.65
Lasso	8366.47	2286.63
WLB range	(7805.11,8940.06)	(3249.34,4071.96)

Table 3: 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.

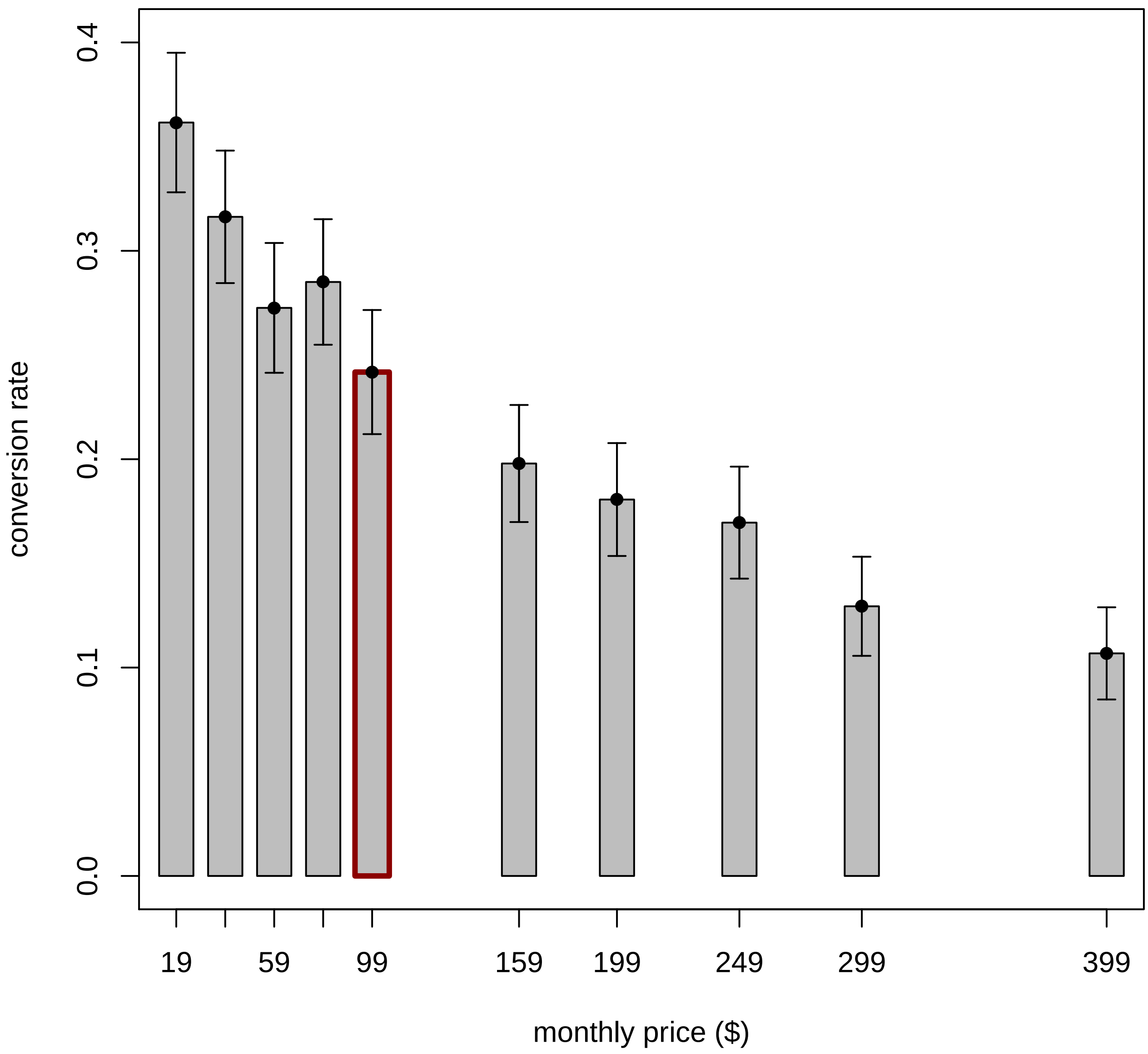


Figure 1: Stage One Experimental Conversion Rates. Each bar corresponds to one of our 10 experimental price cells. The height of the bar corresponds to the average conversion rate within the cell. Error bars indicate the 95% confidence interval for the conversion rate.

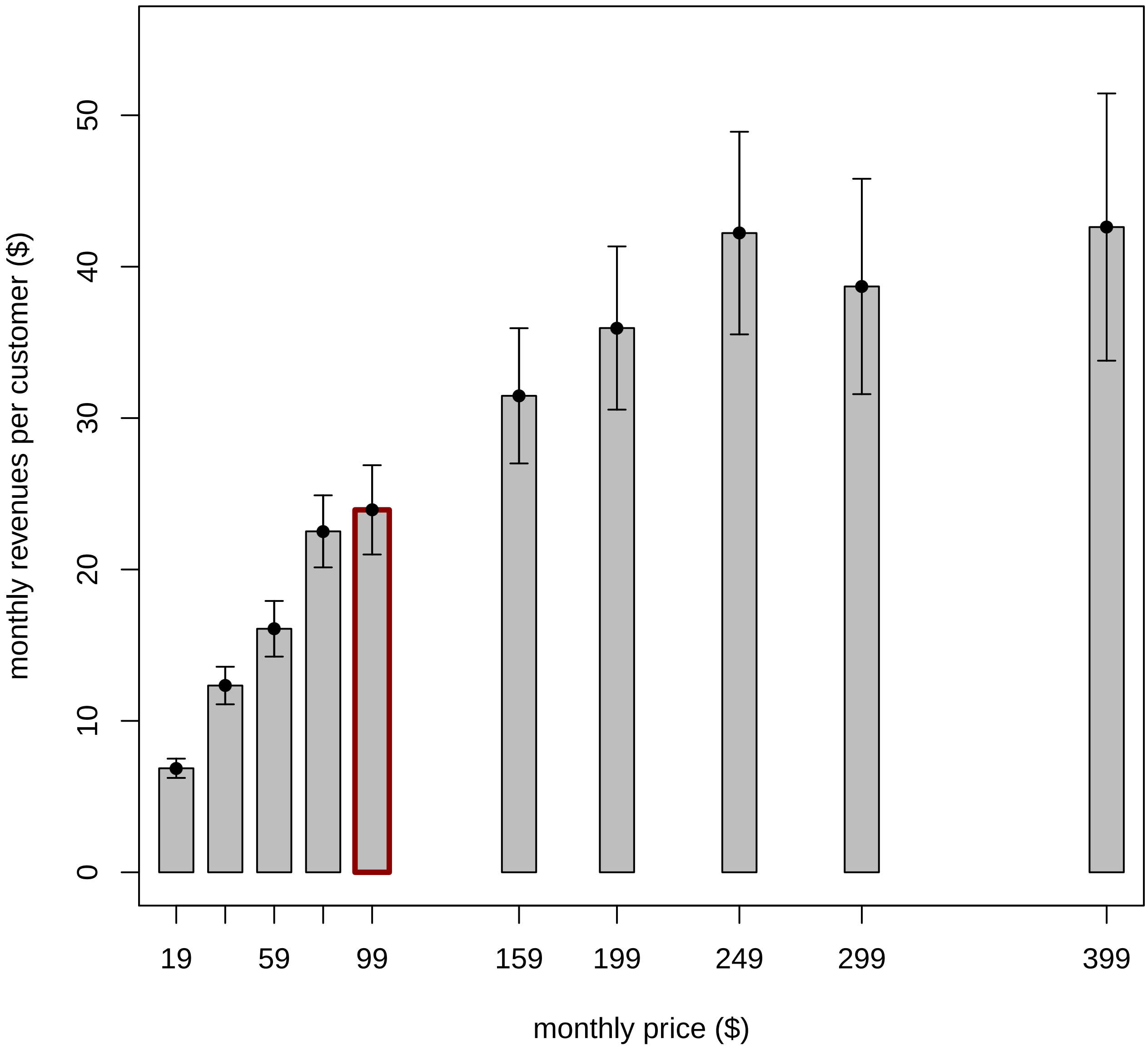


Figure 2: Stage One Experimental Revenues per Customer. Each bar corresponds to one of our 10 experimental price cells. The height of the bar corresponds to the average revenue per prospective customer within the cell. Error bars indicate the 95% confidence interval for the revenues per customer.

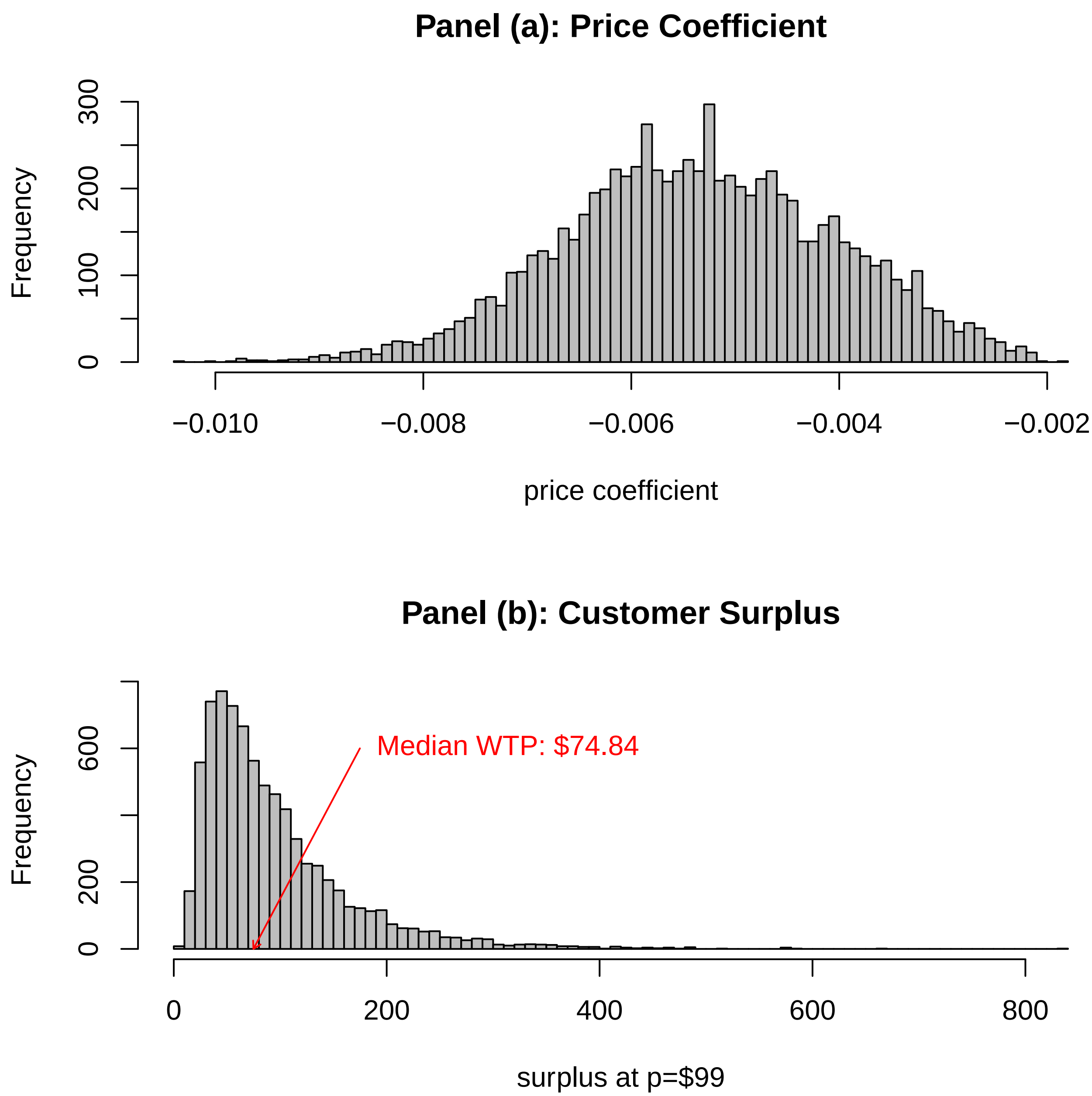


Figure 3: Distribution across customers of posterior mean price sensitivity and posterior surplus from the provision of the service (N=7,867).

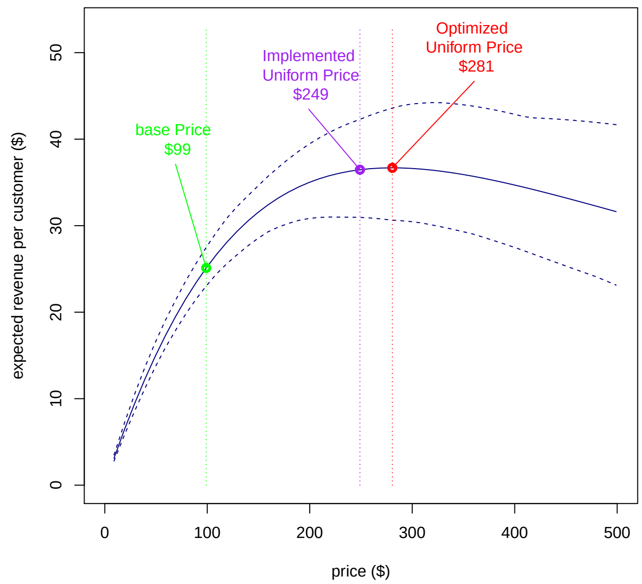


Figure 4: Posterior Monthly Revenues Per Customer (dotted lines represent the 95% posterior credibility interval at each point)

Figure 5: Optimized Prices (N=7,867).

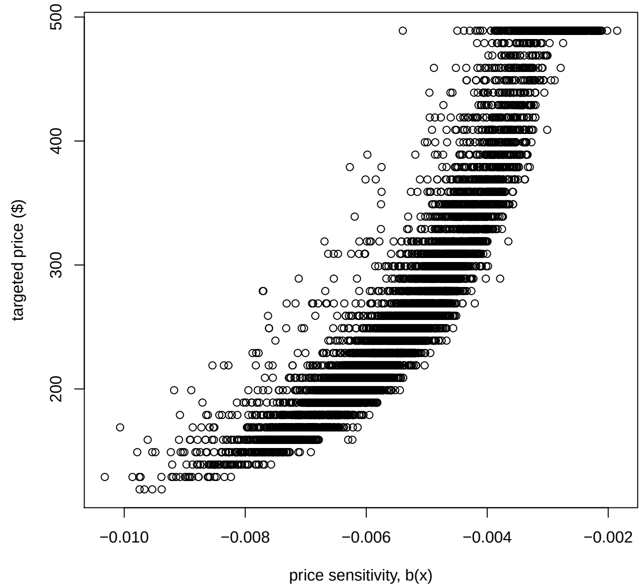


Figure 6: Targeted Prices vs Posterior Mean Price Sensitivities $\left(\hat{\beta}_i\right)$ (N=7,867).

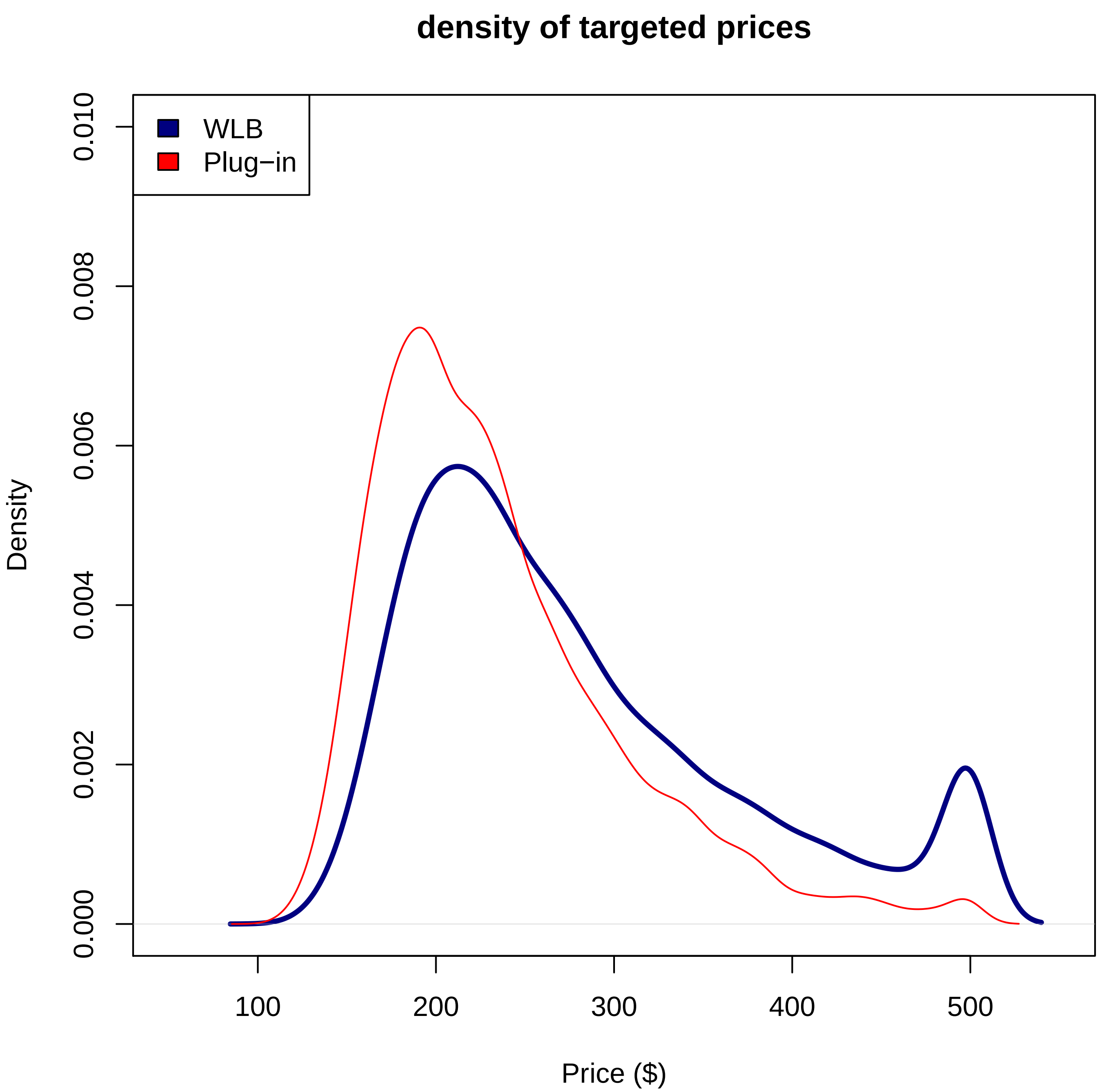
Pricing Structure	**Price Range**	**Conversion Rate**		**Profit per Lead ($)**	
		Mean	**95% Credibility Interval**	**Mean**	**95% Cred. Interval**
Control	$99	0.26	(0.24,0.28)	$25.34	($23.29,$27.68)
Optimized Uniform	$280.57	0.15	(0.12,0.17)	$40.82	($33.98,$48.58)
Implemented Uniform	$249	0.16	(0.14,0.18)	$39.69	($33.93,$45.42)
Targeted	($125.45,$2465.71)	0.15	(0.13,0.18)	$47.61	($36.86,$62.02)
Implemented Targeted	($119,$489)	0.16	(0.13,0.19)	$45.50	($36.47,$55.91)
Perfect	($1.87,$60.5)	0.36	(0.32,0.4)	$98.58	($85.97,$113.5)

Table 4: Stage one posterior profitability by pricing structure (Targeted and Perfect price discrimination cap the prices charged at $499).

Pricing Structure	**# subjects**	**Conversion Rate**		**Profit per Customer ($)**	
		Mean	**95% Conf. Interval**	**Mean**	**95% Conf. Interval**
Control	1360	0.23	(0.21,0.25)	22.55	(20.75,24.39)
Implemented Uniform	1430	0.15	(0.14,0.17)	37.73	(33.78,41.79)
Targeted	2485	0.15	(0.14,0.16)	41.67	(38.34,45.10)

Table 5: Stage two conversion and profitability by pricing structure. (Bootstrapped confidence intervals obtained using 1,000 replications draw with replacement from entire sample in each of the cells).

Figure 7: Distribution of Targeted Prices using WLB and "plug-in." Plug-in estimates are the posterior mean values of Ψ_i.



Pricing Structure	# subjects	Conversion Rate		Profit per Lead ($)	
		Mean	95% Credibility Interval	Mean	95% Cred. Interval
Control	1360	0.26	(0.24,0.29)	25.76	(23.74, 28.5)
Implemented Uniform	1430	0.16	(0.13,0.19)	40.05	(32.97, 47.5)
Targeted	2485	0.15	(0.13,0.18)	44.74	(35.24, 54.09)

Table 6: Stage Two posterior profitability predictions by pricing structure

Figure 8: CDFs of Profit Per Customer in Each Cell (September, 2015)

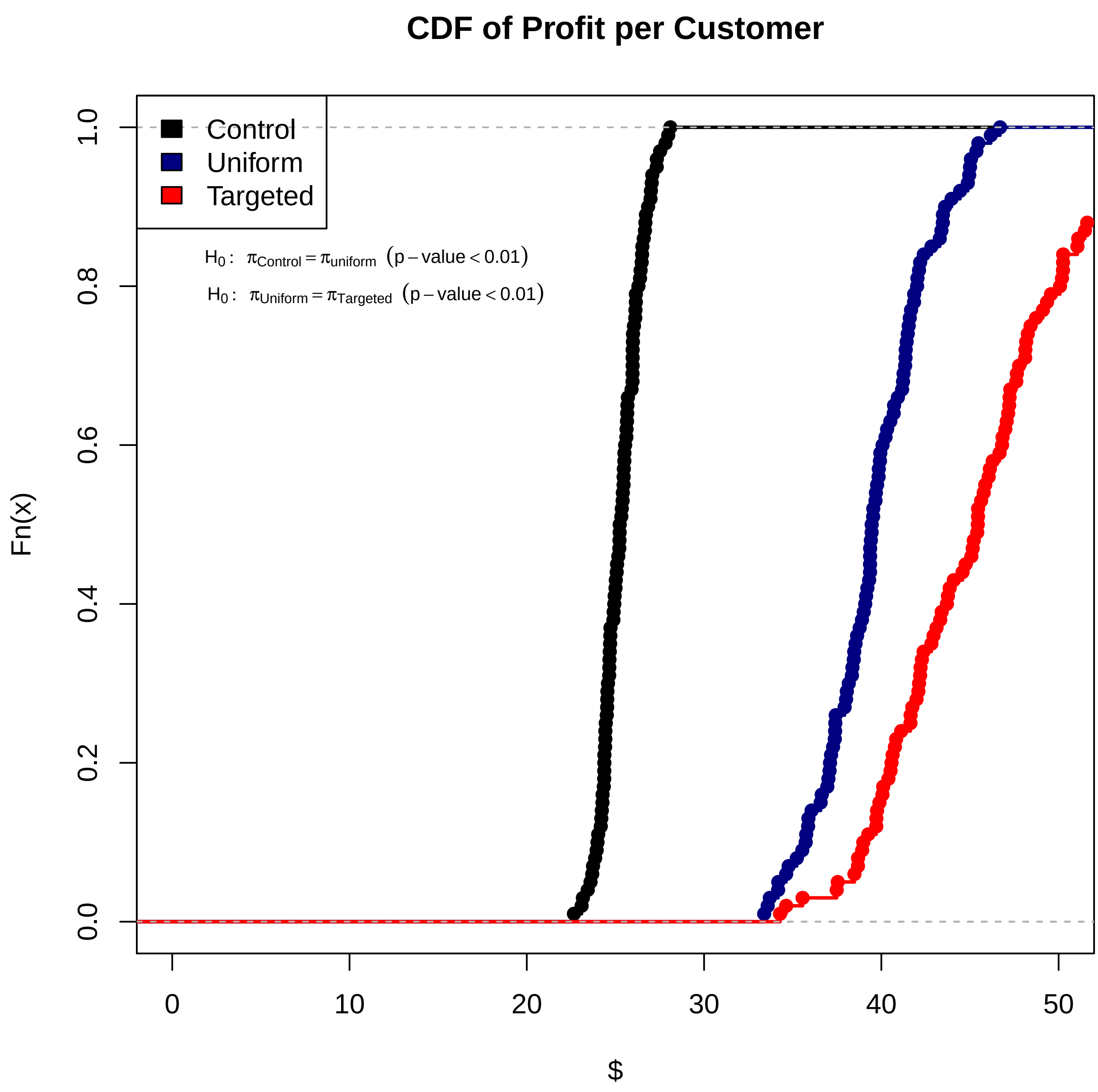


Figure 9: Density of Targeted Prices in Each Cell (November, 2015)

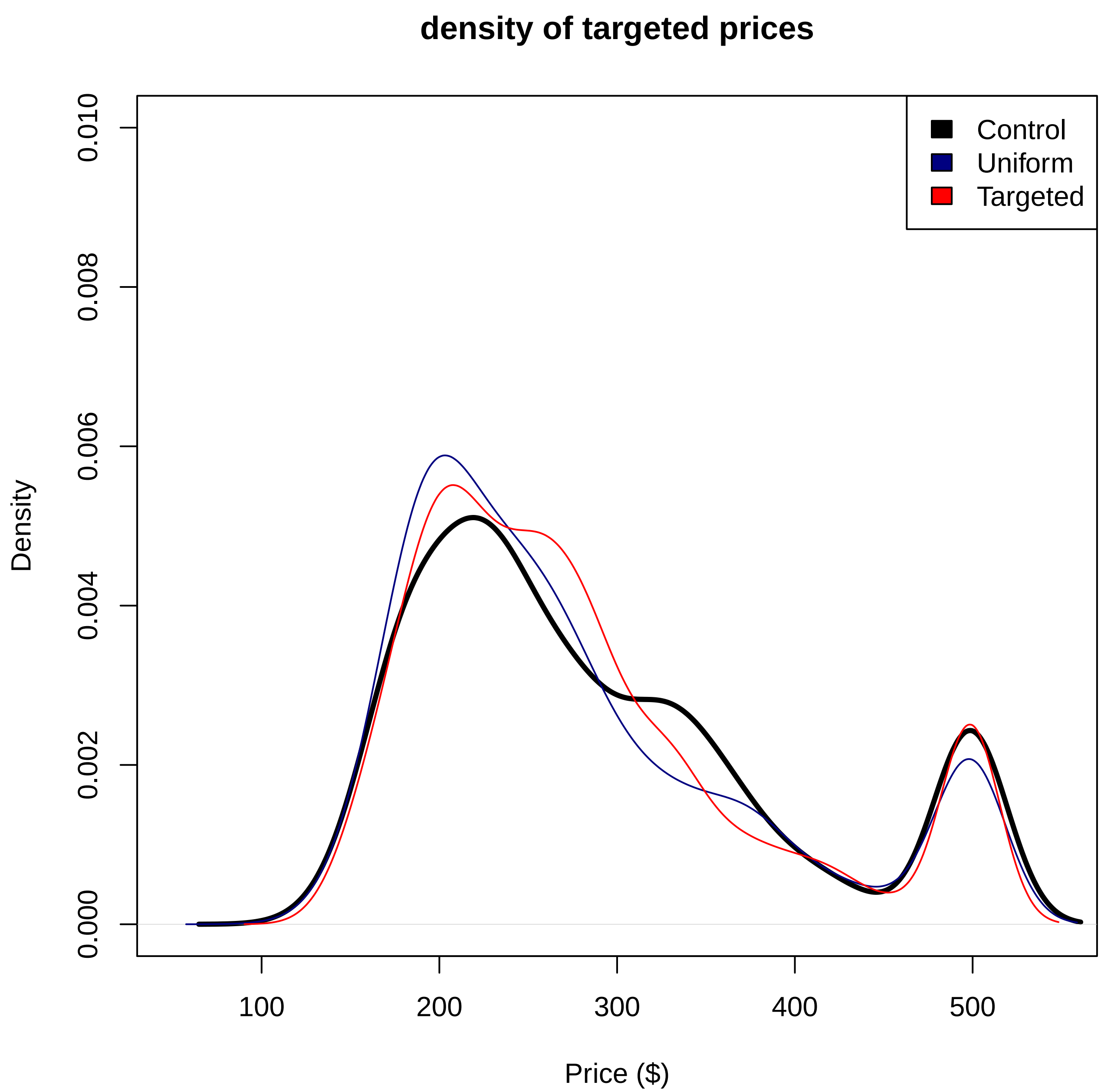


Figure 10: CDFs of Profit Per Customer in Each Cell (November, 2015)

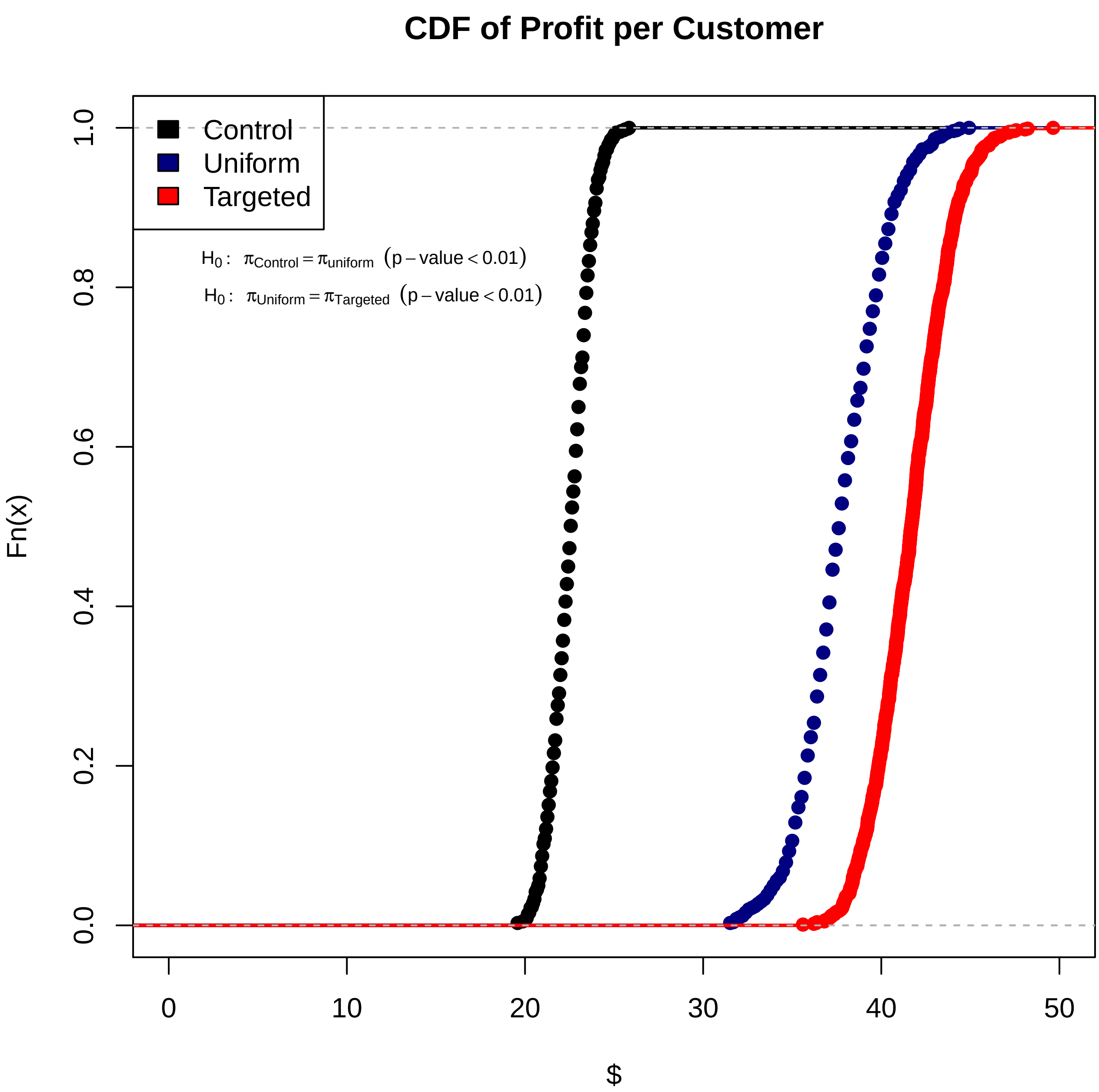
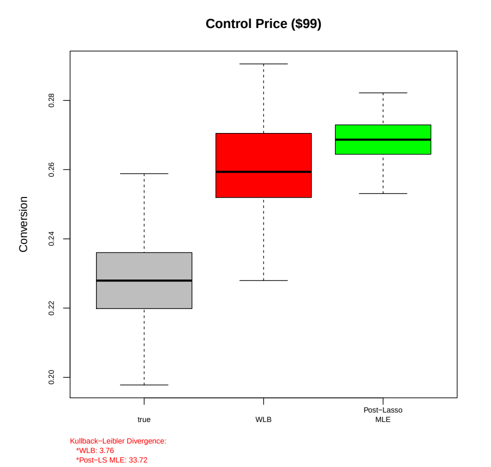
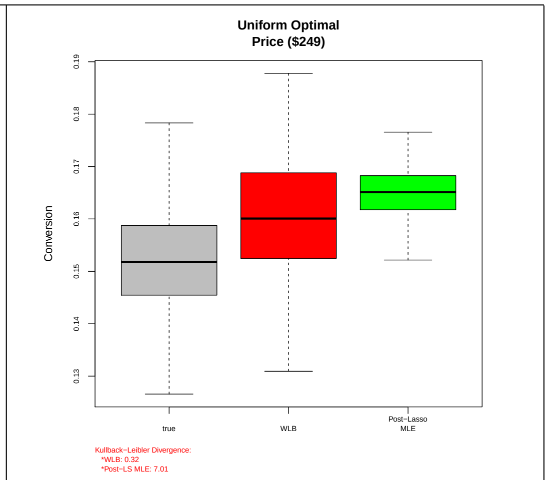


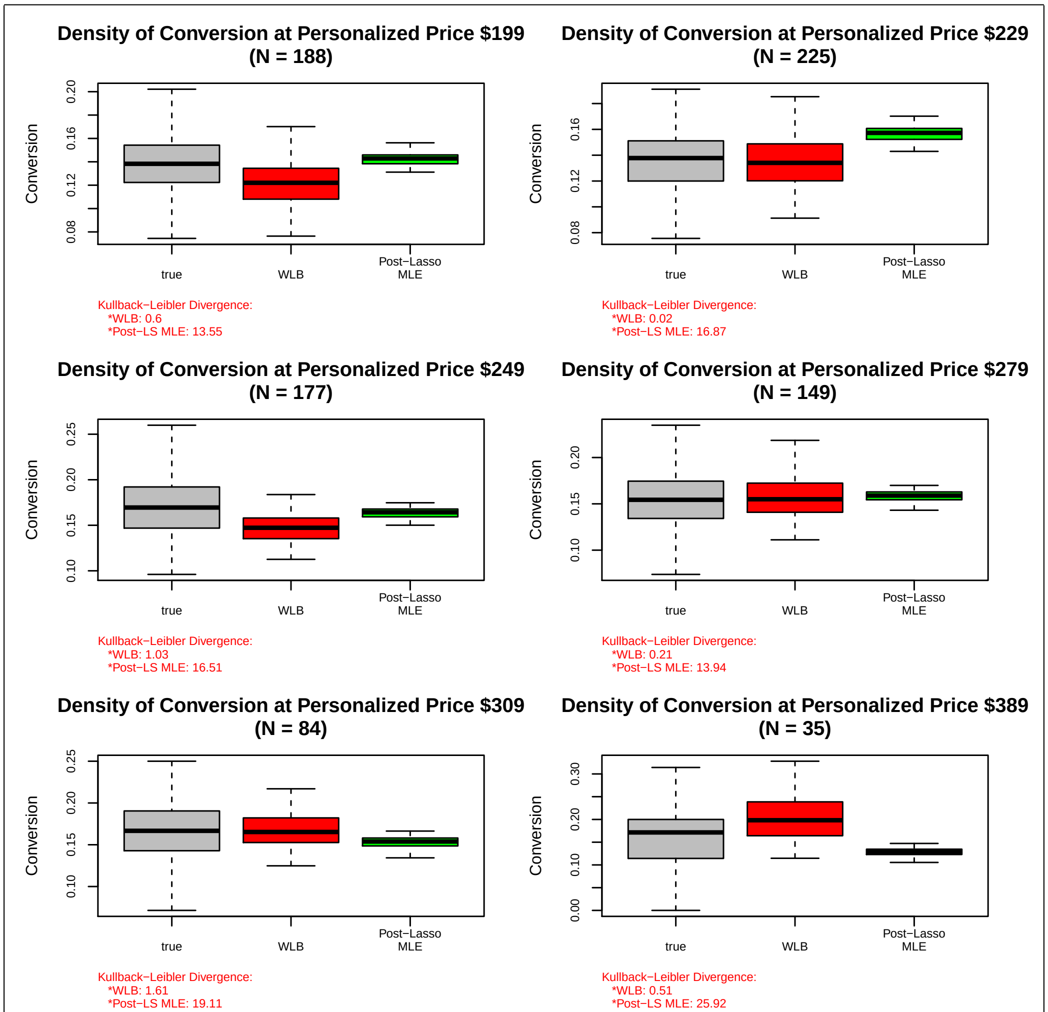
Figure 11: Comparison of Predicted and Realized Conversion



(a) Control Pricing ($99)



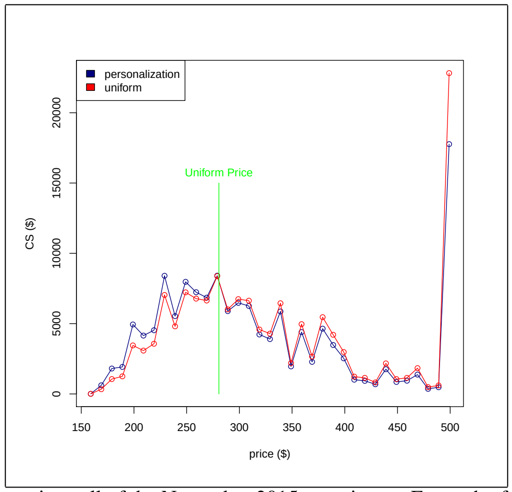
(b) Uniform Optimal Pricing



(c) scalable price targeting

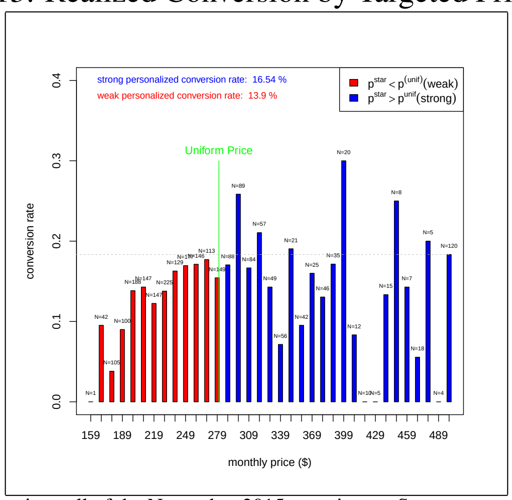
The plots compare the empirical density of realized conversion, for a given pricing structure, to the corresponding predicted densities for WLB, post-Lasso MLE and MLE respectively. The density of realized conversions is computed by bootstrapping (with replacement) from the Nov data.

Figure 12: Comparison of Predicted Total Customer Surplus (by price cell) for Scalable Price Targeting and Uniform Pricing



Results pertain to the 2,485 customers in the targeting cell of the November 2015 experiment. For each of the targeted price cells, we report total surplus across all customers in that cell under SPT (blue). As a comparison, we also report the total surplus had those customers instead been charged the optimal uniform price.

Figure 13: Realized Conversion by Targeted Price Cell

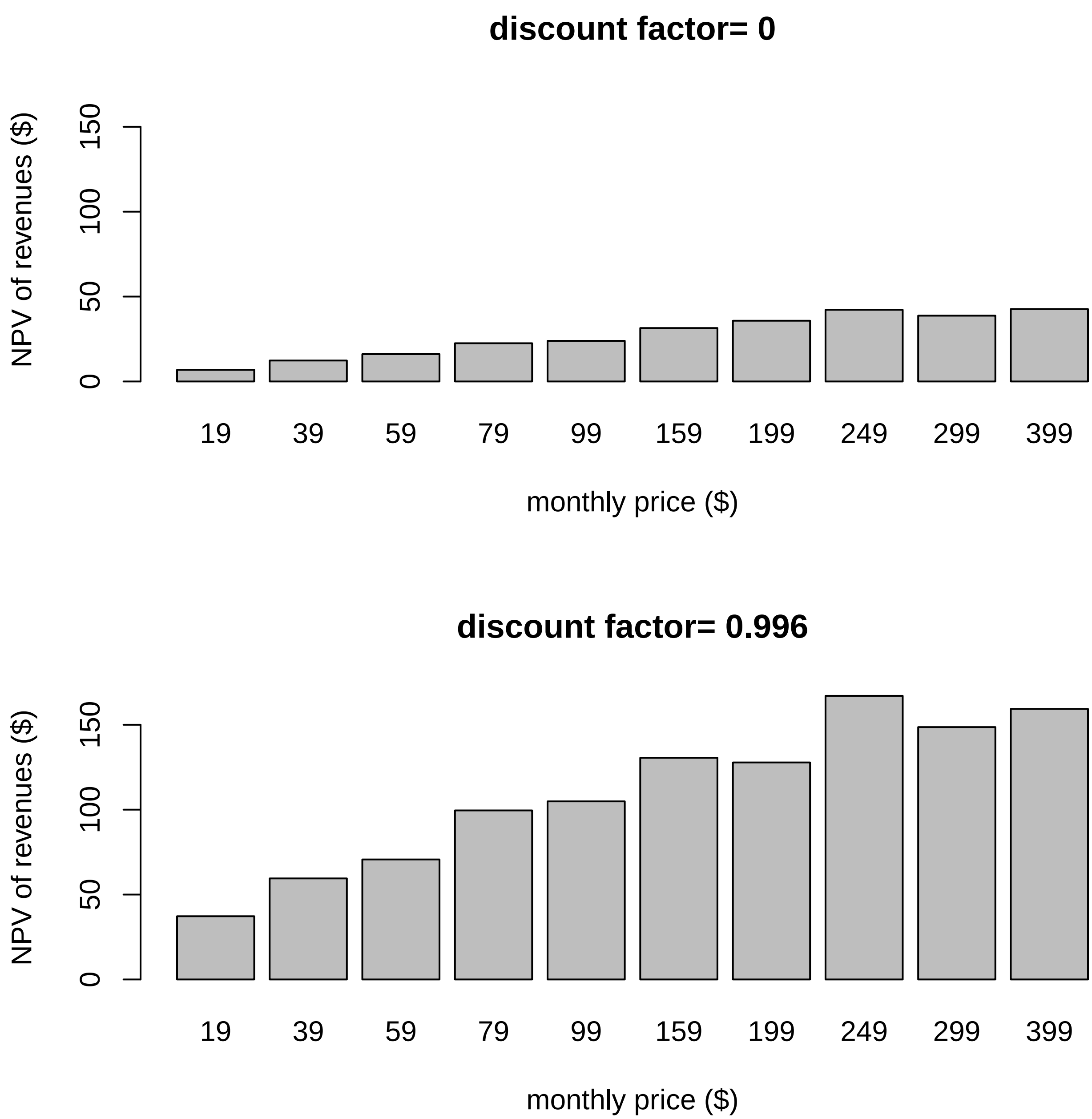


Results pertain to the 2,485 customers in the targeting cell of the November 2015 experiment. Strong customers are targeted a price higher than the uniform price. Weak customers are targeted a price lower than the uniform price.

Table 7: 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

Figure 14: Expected Net Present Value of Monthly Revenues Per Lead over a 4-Month Horizon (September 2015)



A Appendix: Lasso Regression

The penalized Lasso estimator solves for

$$\hat{\Theta}|_{\lambda} = \underset{\Theta \in \mathbb{R}^J}{argmin} \left\{ \ell(\Theta) + N\lambda \sum_{j=1}^{J} |\Theta_j| \right\} \tag{19}$$

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 $t = 2, ..., T$:

$$\begin{aligned} \lambda^t &= \delta\lambda^{t-1} \\ \omega_j^t &= \left(|\Theta_j^{t-1}|\right)^{-1}, \ j \in \hat{S}_t \\ \hat{\Theta}^t &= \underset{\Theta \in \mathbb{R}^J}{argmin} \left\{ \ell(\Theta) + N\textstyle\sum_{j=1}^{J} \lambda^t \omega_j^t |\Theta_j| \right\}. \end{aligned} \tag{20}$$

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 20 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.

B Appendix: Conditional Expectation of Truncated Logistic Random Variable

The random utility component of equation 12 is assumed to be i.i.d. logistic with pdf

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

and CDF

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

The truncated density 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:

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

where

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