Competitive Model Selection in Algorithmic Targeting*

Ganesh Iyer
University of California at Berkeley
giyer@berkeley.edu

T. Tony Ke
Chinese University of Hong Kong
tonyke@cuhk.edu.hk

June 2022

*Very preliminary. Comments welcome.

Competitive Model Selection in Algorithmic Targeting

Abstract

We consider two firms, each appointing a data analyst to estimate demand for multiple consumer segments by running a supervised learning algorithm. Based on the estimates reported by his analyst, each firm devises a targeting policy to maximize estimated profit. The firms face the general trade-off between bias and variance in choosing estimation methods. We show that competition may drive them to choose less flexible algorithms that introduce bias. This implies that more flexible algorithms such as deep learning are more valuable to firms with greater monopoly power.

Keywords: *targeting, algorithmic competition, model simplicity, data analytics, model selection, supervised learning*

1 Introduction

The digital economy has made available unparalleled amounts of consumer data to firms. Over the past decade firms are increasingly delegating the tasks of reaching and targeting consumers to machine learning algorithms which use large amounts of data on consumer characteristics and behavior. One of the defining characteristics of big data environments is the rich and high dimensional information on consumer characteristics, attitudes, opinions and behaviors. Often the number of variables and aspects of consumer behavior that is present can be comparable to the size of the dataset.

Consequently, in designing targeting algorithms firms have to not only be concerned about predictive accuracy, but also about selecting the most relevant variables to include in the model. Specifically, big data environments might confront the firms with what is called the over-fitting problem: The algorithm may use a large number of available consumer predictor variables and more complex functions to map the data onto targeting predictions, but this increases the variance of the estimated predictions. Alternatively, the algorithm can be regularized wherein the more complex functions can be penalized leading to the selection of only the most relevant variables. This would reduce the variance of the estimated predictions but then may introduce bias in the estimates. This paper considers the optimal algorithmic design by firms in competitive markets which trades-off model selection, namely, the selection of the most relevant predictor variables and the predictive accuracy of consumer targeting. Under competition the success of a firm's algorithmic technology depends not only on its predictive ability and its strategic choice of variables, but also on the choices made by its competitors.

In the model we represent the algorithmic design problem of model selection and predictive accuracy using the running example of a basic model like the LASSO which adds a cost function to the standard regression analysis and penalizes non-zero predictor variable coefficients. Thus the procedure accommodates both model selection (selecting which variables will enter the prediction algorithm) as well as the estimation of the selected variables' coefficients.

We consider a market in which firms compete by targeting consumers who are heterogeneous in some characteristic. Firms observe consumer characteristics (in their data) but are uncertain about the profitability of different consumer types. They have access to data which they use to estimate the profit. Given the specialized expertise needed to deploy predictive algorithms firms delegate the task of running them to data analytics groups. However, the firm's management can make strategic design choices

on the complexity of the algorithms and the number of dimensions of consumer behavior that the algorithm should consider. To capture this we construct a two stage simultaneous move game. In the first stage prior to getting the data each firm chooses the tuning or regularization parameter for the algorithm. In the second stage each firm is endowed with a private dataset which is available to its data analyst for its data analyst to run the predictive algorithm to generate the profit estimates. Based on these estimates the firms choose their targeting strategy to maximize profits.

We first analyze the monopoly benchmark, where there is only one firm in the market. We find that it is optimal for the firm to choose zero penalization. In other words, a monopoly firm prefers a more flexible algorithmic design which admits more variance but has lower bias. This allows the firm to target the more profitable consumer segment with greater likelihood. Then, we proceed to analyze the competitive market and find that it can be an equilibrium for both firms to choose positive penalization which introduces bias while reducing variance. In other words, we find that competition favors simpler models for targeting. Under competition firms have two incentives: i) to correctly target the more profitable segment and ii) avoid competition and the overlap in targeting. Allowing for bias helps to avoid competition by reducing the equilibrium overlap in targeting. Overall, the suggestion of our analysis is that more flexible and complex algorithms such as deep learning are likely to be of higher value and be used by firms with greater monopoly power.

2 Related Research

Our paper is broadly related to new and emerging research literature which examines strategic interactions and incentives with algorithms. One strand of research tackles the problem of algorithmic design when it is faced by rational and strategic agents who can manipulate the information that is provided to the algorithm. For example, Björkegren et al. (2020) considers individuals who may observe the rules of the machine learning algorithms and strategically manipulate their behavior to get desired outcomes. The paper derives an equilibrium estimator that is robust to manipulation and models the costs of manipulating different behaviors. Another paper by Eliaz and Spiegler (2019) study a statistical algorithm faced with an agent who strategically self-reports her personal data and they highlights the role of model selection and the incentive issues that it creates. Our paper examines the model selection problem but in a competitive market where firms choose the equilibrium design of their consumer targeting algorithms. Thus in our paper the extent to which firms choose more or less

flexible algorithms and the associated trade-off between model selection and predictive accuracy is governed by the equilibrium consumer targeting incentives of the firms.

There is a recent emerging literature on competitive interactions between multiple algorithms. Salant and Cherry (2020) consider statistical inference in games, where each player obtains a small random sample of other players' actions, uses statistical inference to estimate their actions, and chooses an optimal action based on the estimate. Liang (2020) considers games of incomplete information in which the players have data and use algorithms to derive their beliefs. Olea et al. (2019) study a game between agents competing to predict a common variable, and where agents obtain the same data but differ in the algorithms they utilize for prediction. In all these papers, the algorithms under consideration are fixed exogenously. Our paper, in contrast, focuses on the strategic choice of algorithms in competitive environments.

There is also a growing literature on algorithmic pricing in repeated oligopoly to understand whether algorithms can consistently learn to charge supra-competitive prices. For example, Calvano et al. (2020) examine firms endowed with Q-learning algorithms in repeated interactions to show that they can robustly learn to cooperate with communicating with each other. Lastly, we contribute to the traditional literature on competitive targeting strategies (e.g., Shaffer and Zhang 1995; Chen et al. 2001; Iyer et al. 2005; Bergemann and Bonatti 2011) by introducing the algorithmic design and decisions on model selection to the consumer targeting strategies of firms.

3 Model Setup

Consider a market consisting of consumers who are heterogeneous in a characteristic $x \in \{1, 0\}$. ϕ fraction of consumers have $x = 1$ and the remaining $1 - \phi$ fraction have $x = 0$, where $\phi \in (0, 1)$. For example, x_i may represent consumer i's demographics (1 for men and 0 for women), or past consumer behaviors (1 for those who have visited some website and 0 otherwise), etc. This case of a single characteristic offers the simplest setup for the development of the idea.

There are two firms competing for consumers in the market, indexed by $j = 1, 2$. Firms can observe each consumer i's characteristic x_i and decides which type(s) of consumers to target. Each firm has the ability to reach and target $\theta \in (0, 1)$ fraction of the consumer population in the market. Targeting can therefore be also interpreted as a form of costly informative advertising that informs consumers the existence of the product (Butters 1977). If consumer i is only targeted by firm j, the consumer will only buy from the firm, and the firm earns a monopolistic profit of $\pi_j(x_i)$; on the other hand,

if the consumer is targeted by both firms, she will randomly choose a firm to make a purchase, and thus firm j's expected profit is $\pi_j(x_i)/2$. Lastly, if a consumer is not targeted by either of the two firms, she will not make a purchase from the two firms. To focus the exposition on the effects of algorithmic targeting, we have abstracted away the two firms' decisions on prices.[1]

Given x being binary, it is without loss of generality to write down $\pi_j(x)$ as the following linear function,

$$\pi_j(x) = \alpha_j + \beta_j x. \tag{1}$$

Firm j does not know α_j, β_j a priori. We assume a common prior for α_j, β_j, which follow distribution functions A and B respectively. A is supported in $[\underline{\alpha}, \overline{\alpha}]$, and B is a symmetric distribution around zero, supported in $[-\overline{\beta}, \overline{\beta}]$. α_1, β_1, α_2 and β_2 are independently distributed. The firm is interested in estimating α_j and β_j given the available data. It delegates the task of estimation and prediction to a data analytics department which is equipped with the technology of running prediction and model selection algorithms. Specifically, assume that the analyst uses the technology of running Lasso regressions and that a complete contract between the firm and the data analyst is not possible. Rather, the firm can only specify the tuning parameter of the Lasso regression. Then given the tuning parameter specified by the firm, the analyst runs the Lasso regression on the data to generate an estimate of α_j and β_j.

It is assumed that each firm j and its data analyst have a private access to a dataset with two observations. The l-th observation contains a pair of (x^l, y_j^l) for $l = 0, 1$, where, $x^0 = 0$, $x^1 = 1$ and

$$y_j^l = \pi_j(x^l) + \varepsilon_j^l = \alpha_j + \beta_j x^l + \varepsilon_j^l. \tag{2}$$

The error term, ε_j^l is i.i.d. across j and l and follows distribution function G, which is symmetric around zero and supported in $[-\overline{\varepsilon}, \overline{\varepsilon}]$. We further define $\Delta\varepsilon_j \equiv \varepsilon_j^1 - \varepsilon_j^0$, which follows distribution function $\widetilde{G}$, where $\widetilde{G}(e) = \Pr(\varepsilon_j^1 - \varepsilon_j^0 \leq e) = \int_{-\overline{\varepsilon}}^{\overline{\varepsilon}} G(e' + e) dG(e')$. It is further assumed that $\widetilde{G}$ is single-peaked; that is, $\widetilde{G}'(e)$ weakly decreases (increases) with e for $e > 0$ ($e < 0$).

Based on the data, the analyst runs a Lasso regression, which is represented by the

[1] If price discrimination based on targeting outcomes is allowed, we can endogenize prices in a trivial way. If a consumer is targeted by only one firm, the firm sets the monopoly price and still earns a monopoly profit; on the other hand, if a consumer is targeted by two firms, they engage in a Bertrand competition, which drives the price to be the marginal cost and each firm's profit to be zero. This setting will generate qualitatively the same result as in the model with exogenous prices.

following minimization problem:

$$\left(\hat{\alpha}_j(\lambda_j), \hat{\beta}_j(\lambda_j)\right) = \underset{(a_j,b_j)}{\arg\min} \sum_{l=0}^{1} \left(y_j^l - a_j - b_j x^l\right)^2 + \lambda_j |b_j|, \tag{3}$$

where $\lambda_j \geq 0$ is the tuning parameter specified by firm j that measures the degree of penalization on $\hat{\beta}_j(\lambda_j)$. The choice of λ_j indicates the model selection decision of the firm: At the one extreme when $\lambda_j = 0$ this corresponds to the case of a standard ordinary least square (OLS) regression and in this setup this is equivalent to the firm deciding on the maximum maximum model flexibility and choosing all the available predictor variables. In contrast when λ_j is large and the penalization is large then the model would shrink and have lower flexibility with fewer admitted predictors.

By taking derivatives and the corresponding first- and second-order optimality conditions, it is straightforward to solve the data analyst's estimation problem in equation (3):

$$\hat{\alpha}_j(\lambda_j) = \frac{1}{2}\left(y_j^1 + y_j^0 - \hat{\beta}_j(\lambda_j)\right), \tag{4}$$

$$\hat{\beta}_j(\lambda_j) = \begin{cases} \max\{y_j^1 - y_j^0 - \lambda_j, 0\}, & \text{if } y_j^1 - y_j^0 \geq 0, \\ \min\{y_j^1 - y_j^0 + \lambda_j, 0\}, & \text{otherwise.} \end{cases} . \tag{5}$$

The expression of $\hat{\alpha}_j(\lambda_j)$ in equation (4) is the same as the standard OLS estimator, because there is no penalization on $\hat{\alpha}_j(\lambda_j)$. It is assumed that $\underline{\alpha}$ is large enough so that the realization of $\hat{\alpha}_j(\lambda_j)$ is always positive. This guarantees that firm j will always want to target some consumers in the market. To understand the expression of $\hat{\beta}_j(\lambda_j)$ intuitively, notice that if $\lambda_j = 0$, we have $\hat{\beta}_j(\lambda) = y_j^1 - y_j^0$, which is the OLS estimator; if $0 < \lambda_j < |y_j^1 - y_j^0|$, $\hat{\beta}_j(\lambda_j)$ is in the same sign with $y_j^1 - y_j^0$ but penalized toward zero; finally, if $\lambda_j \geq |y_j^1 - y_j^0|$, the penalization is so severe that $\hat{\beta}_j(\lambda_j) = 0$.

We consider a simultaneous-move game between the two firms in two periods. First, each firm j chooses the tuning parameter λ_j, which remains private for the entire game.Second, each firm j is endowed with a private dataset (x^l, y_j^l) for $l = 0, 1$, based on which firm's analyst generates the estimates $\hat{\alpha}_j(\lambda_j)$ and $\hat{\beta}_j(\lambda_j)$ by running a Lasso regression. Lastly, each firm devises the targeting strategy to maximize the estimated profit. Figure 1 summarizes the game timeline. Before we proceed to analyze the game, we elaborate on the rationale and interpretation of our modeling choices.

First, the reader may wonder that the simple setup above with a data set with just two observations and a binary characteristic ($x \in (1, 0)$) is a far cry from the big data

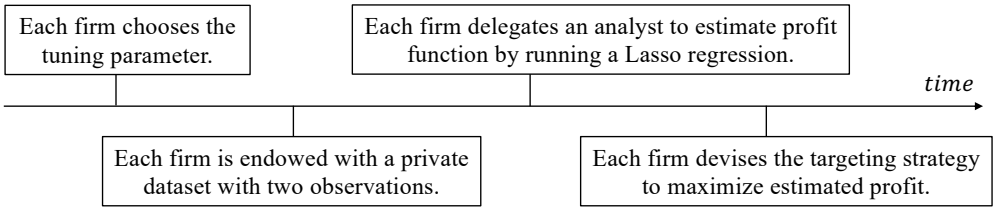


Figure 1: Timeline of a competitive algorithmic targeting game.

situations confronting firms. Machine learning models are typically high dimensional and complex involving numerous dimensions available in big data. Nevertheless, the model is designed to handle the crucial aspects of the so-called "over-fitting" problem encountered in algorithmic decision making by firms, namely, that the potential number of explanatory variables may be large and comparable to the sample size. So unless there is a method for model selection and shrinkage of the number of explanatory variables there is a risk of over-fitting. For example, an unpenalized regression estimator may perfectly fit the data-set but would have high variance and poor predictive performance compared to an estimator with shrinkage. However, a model with shrinkage may potentially be subject to the introduction of bias in the estimated coefficients. The model with the Lasso regression above with the endogenous choice of the tuning parameter λ_j helps to capture the essence of the trade-offs underlying the over-fitting problem, and in doing so endogenize the model selection to the equilibrium incentives of the firms.

Second, in our model, the firms choose the tuning parameters before getting the data. This is consistent with the statistical learning literature which prescribes that the tuning parameter should not be determined based on the training data per se in order to avoid over-fitting.

Third, while we use the Lasso regression as a specific estimation procedure, our results are more general in the sense that λ_j determines the general trade-off between bias and variance in any supervised learning method, where higher values of λ_j is associated with lower the variance but higher bias. Therefore, firm j's choice of λ_j can be interpreted as choosing between different statistical learning models that differ in bias-variance trade-off. Thus the problem can be seen as the strategic choice of the bias-variance trade-off in algorithmic design of the firm's targeted advertising strategy.

Furthermore, different statistical models differ in flexibility, as shown by Figure 2. Those with higher flexibility typically have lower bias but higher variance. Here we will focus on the comparison between Lasso and OLS, where OLS has higher flexibility and lower bias, while Lasso with some level of regularization has lower flexibility

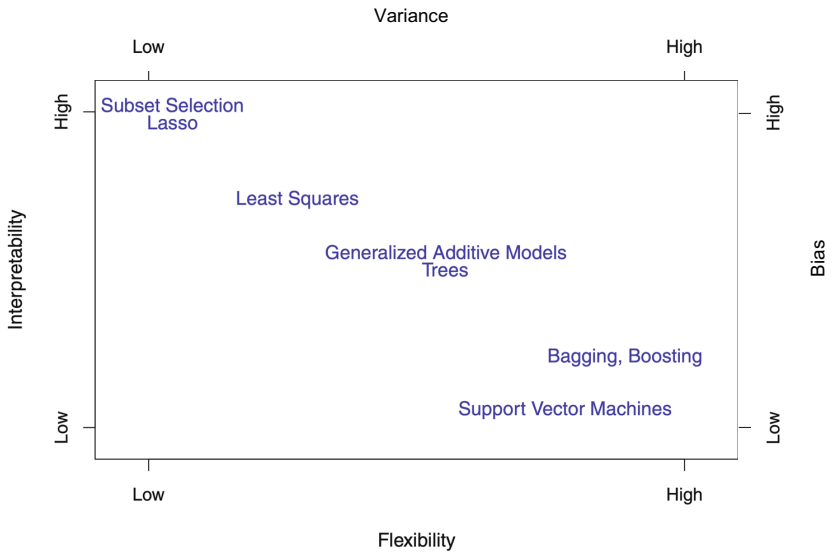


Figure 2: Tradeoff between flexibility and interpretability and tradeoff between bias and variance across different statistical learning methods (excerpted from James et al. (2013) page 25 and adapted).

and higher bias when compared to OLS. By this understanding, the Lasso regression does not necessarily need to represent a "machine-learning" algorithm while OLS a traditional algorithm. In fact, in practice, a firm may decide whether to adopt a very flexible machine-learning algorithm like neural networks compared with a less flexible benchmark algorithm, in which case, the neural networks will correspond to OLS in our framework.

Finally, it has been assumed that the firms do not have the analytical capability themselves and can only rely on data analysts for the estimation procedure; moreover complete contracts are not available between a firm and its analyst. This assumption maps with common practices in companies where managers rely on analysis by data analytics groups to make strategic decisions. This has two important implications:

1. In the last stage of the game, instead of performing a Bayesian update based on the data to calculate the posterior belief of α_j and β_j, each firm relies on the data analyst to run the Lasso regression on the data to get point estimates of $\hat{\alpha}_j(\lambda_j)$ and $\hat{\beta}_j(\lambda_j)$; correspondingly, instead of maximizing the expected profit based on the posterior belief, each firm makes the targeting decision by maximizing the "estimated profit" based on the estimate, $\hat{\alpha}_j(\lambda_j)$ and $\hat{\beta}_j(\lambda_j)$. The standard rational economic model for this problem would involve fully Bayesian decision making with common priors for all agents. However, as argued below the reality of data

based algorithmic decision making in firms does not reconcile with the standard approach as machine learning algorithms like Lasso which are based on the minimization as in (3) are non-bayesian procedures. By separating the estimation problems from the decision-maker (firms), and assigning them to the delegates (analysts), we are able to rationalize the reality of data-driven decision making in firms. Methodologically this is a crucial feature of our framework which underlies the algorithmic decision making in firms.[2]

2. Because it is the data analyst instead of the firm that performs the estimation procedure, this implies the minimization of mean squared error instead of profit maximization as the objective in estimating the parameters in the second stage. This makes our results directly comparable with the standard statistical learning literature. This is also consistent with the industry practice due to several considerations. First, minimization of mean squared error is available and used by companies in standard ready-to-use statistical packages while profit maximization requires customization, which could be costly for the firms. Second, information pertaining to the profit function may be scattered in silos within the organization so that even if the data analyst in charge of the estimation task wants to use profit maximization as the objective, she may fail to gather all relevant information.

Next, we first study the monopoly setting with only one firm in the market as the benchmark, and then proceed to study the main model in duopoly.

4 Monopoly Benchmark

Given only one firm, we will drop the subscript j. Given $\hat{\alpha}(\lambda)$ and $\hat{\beta}(\lambda)$, we have the estimated profit from a targeted consumer, $\hat{\pi}(x) = \hat{\alpha}(\lambda) + \hat{\beta}(\lambda)x$. To maximize the estimated profit, the firm prefers to target consumers with $x = 1$ if $\hat{\beta}(\lambda) > 0$. In this case, the firm will target θ consumers with $x = 1$ if $\theta \leq \phi$; otherwise, the firm will

[2]In an alternative setting in absence of the data analysts, we can assign a Laplace prior distribution to each firm j's prior belief of β_j, with the probability density function $f(\beta_j) = \lambda_j/2 \cdot \exp(-\lambda_j|\beta_j|)$. Then, based on the data (x^l, y_j^l) for $l = 0, 1$ and assuming ε_j^l follows a standard normal distribution, firm j forms a posterior belief of α_j and β_j by Bayes' rule, which can be shown to be equivalent to running the Lasso regression in equation (3) (Tibshirani 1996). However, there are two problems with this Bayesian approach. First, the tuning parameter λ_j is not firm j's choice but rather, a model primitive that is exogenously given. Second, the point estimates generated by the Lasso regression, $\hat{\alpha}(\lambda_j)$ and $\hat{\beta}(\lambda_j)$ in equations (4) and (5) are mode instead of mean of the posterior belief of α_j and β_j (Friedman et al. 2001); meanwhile, to calculate expected profit, we will be mostly concerned with the posterior mean instead of mode. Due to these two problems, we do not adopt the Bayesian approach for Lasso regressions.

target ϕ consumers with $x = 1$ and $\theta - \phi$ consumers with $x = 0$. On the other hand, if $\hat{\beta}(\lambda) < 0$, the firm prefers to target consumers with $x = 0$. Similarly, in this case, the firm will target θ consumers with $x = 0$ if $\theta \leq 1 - \phi$; otherwise, the firm will target $1 - \phi$ consumers with $x = 0$ and $\theta - (1 - \phi)$ consumers with $x = 1$. Finally, if $\hat{\beta}(\lambda) = 0$, the firm is indifferent between the two types of consumers, and it is assumed that the firm will target $k \in [0, \theta]$ consumers with $x = 1$. A priori, the firm chooses λ to maximize the expected profit from all consumers:

$$\begin{aligned}\Pi(\lambda) = \theta \mathrm{E}[\alpha] &+ \min\{\theta, \phi\} \Pr(\hat{\beta}(\lambda) > 0) \mathrm{E}[\beta | \hat{\beta}(\lambda) > 0] \\ &+ \max\{\theta - (1 - \phi), 0\} \Pr(\hat{\beta}(\lambda) < 0) \mathrm{E}[\beta | \hat{\beta}(\lambda) < 0] + k \Pr(\hat{\beta}(\lambda) = 0) \mathrm{E}[\beta | \hat{\beta}(\lambda) = 0],\end{aligned}$$

The expectations inside the expression of $\Pi(\lambda)$ are taken over the prior beliefs of α, β and the realizations of the error terms, ε^0 and ε^1.

First, notice that by equation (5), $\hat{\beta}(\lambda) = 0$ is equivalent to $|\beta + \Delta\varepsilon| \leq \lambda$, which, combining with the fact that B and $\widetilde{G}$ are symmetric distributions around zero, further implies that $\mathrm{E}[\beta | \hat{\beta}(\lambda) = 0] = \mathrm{E}[\beta | |\beta + \Delta\varepsilon| \leq \lambda] = 0$. Therefore, the choice of k has no impact on firm profit and thus the tie-breaking rule has no bite on the result. Based on this and the observations that $\hat{\beta}(\lambda) > 0 \Leftrightarrow \beta + \Delta\varepsilon > \lambda$ and $\hat{\beta}(\lambda) < 0 \Leftrightarrow \beta + \Delta\varepsilon < -\lambda$, we can simplify $\Pi(\lambda)$ as the following,

$$\begin{aligned}\Pi(\lambda) =& \theta \mathrm{E}[\alpha] + \min\{\theta, \phi\} \Pr(\beta + \Delta\varepsilon > \lambda) \mathrm{E}[\beta | \beta + \Delta\varepsilon > \lambda] \\ &+ \max\{\theta - (1 - \phi), 0\} \Pr(\beta + \Delta\varepsilon < -\lambda) \mathrm{E}[\beta | \beta + \Delta\varepsilon < -\lambda] \\ =& \theta \mathrm{E}[\alpha] + \min\{\theta, \phi\} \int_{-2\overline{\varepsilon}}^{2\overline{\varepsilon}} d\widetilde{G}(e) \int_{\lambda - e}^{\overline{\beta}} b dB(b) \\ &+ \max\{\theta - (1 - \phi), 0\} \int_{-2\overline{\varepsilon}}^{2\overline{\varepsilon}} d\widetilde{G}(e) \int_{-\overline{\beta}}^{-\lambda - e} b dB(b) \\ =& \theta \mathrm{E}[\alpha] + \min\{\theta, \phi\} \int_{-2\overline{\varepsilon}}^{2\overline{\varepsilon}} d\widetilde{G}(e) \int_{\lambda - e}^{\overline{\beta}} b dB(b) \\ &+ \max\{\theta - (1 - \phi), 0\} \int_{-2\overline{\varepsilon}}^{2\overline{\varepsilon}} d\widetilde{G}(e) \int_{-\overline{\beta}}^{-\lambda + e} b dB(b) \\ =& \theta \mathrm{E}[\alpha] + (\min\{\theta, \phi\} - \max\{\theta - (1 - \phi), 0\}) \int_{-2\overline{\varepsilon}}^{2\overline{\varepsilon}} d\widetilde{G}(e) \int_{\lambda - e}^{\overline{\beta}} b dB(b) \\ =& \theta \mathrm{E}[\alpha] + \min\{\theta, 1 - \theta, \phi, 1 - \phi\} \int_{-2\overline{\varepsilon}}^{2\overline{\varepsilon}} d\widetilde{G}(e) \int_{\lambda - e}^{\overline{\beta}} b dB(b), \end{aligned} \tag{6}$$

where the third and fourth equations above are due to the symmetry of $\widetilde{G}$ and the sym-

metry of B respectively. Given $\widetilde{G}$ is single-peaked, one can show that $\Pi(\lambda)$ decreases with λ, so we have the following proposition.

Proposition 1. *Under monopoly, the firm chooses the tuning parameter* $\lambda^M = 0$.

Proof.

$$\Pi'(\lambda) = -\min\{\theta, 1-\theta, \phi, 1-\phi\} \int_{-2\overline{\varepsilon}}^{2\overline{\varepsilon}} (\lambda - e) B'(\lambda - e) \widetilde{G}'(e) de.$$

If $\lambda \geq 2\overline{\varepsilon}$, obviously, $\Pi'(\lambda) \leq 0$. Otherwise, if $\lambda < 2\overline{\varepsilon}$, we have

$$\begin{aligned}
\Pi'(\lambda) &\propto -\left(\int_{-2\overline{\varepsilon}}^{2\lambda - 2\overline{\varepsilon}} + \int_{2\lambda - 2\overline{\varepsilon}}^{\lambda} + \int_{\lambda}^{2\overline{\varepsilon}} \right) (\lambda - e) B'(\lambda - e) \widetilde{G}'(e) de \\
&= -\int_{-2\overline{\varepsilon}}^{2\lambda - 2\overline{\varepsilon}} (\lambda - e) B'(\lambda - e) \widetilde{G}'(e) de - \theta \int_{0}^{2\overline{\varepsilon} - \lambda} z B'(z) \left(\widetilde{G}'(\lambda - z) - \widetilde{G}'(\lambda + z) \right) dz \\
&\leq 0,
\end{aligned}$$

where, to get the second equality above, we have changed the variable $e = \lambda - z$ for the second integral from $2\lambda - 2\overline{\varepsilon}$ to λ, and $e = \lambda + z$ for the third integral from λ to $2\overline{\varepsilon}$; moreover, we have utilized $B'(z) = B'(-z)$. To get the last inequality, notice that given $\widetilde{G}$ being single-peaked and symmetric around zero, we have $\widetilde{G}'(\lambda - z) \geq \widetilde{G}'(\lambda + z)$ for any $z \geq 0$ and $\lambda \geq 0$. To summarize, we have shown that $\Pi'(\lambda) \leq 0$, so the optimal λ should be $\lambda^M = 0$. □

Proposition 1 implies that a monopoly firm within the setup of the model prefers the OLS regression to a Lasso. The qualitative implication is that a monopolist optimally prefers a more flexible algorithmic design which accommodates all the variables (in our case one) and which may risk over-fitting the data. In other words, the monopoly prefers low algorithmic bias but at the expense of increased variance. This result serves as benchmark and motivates our analysis below of the competitive incentives for algorithmic targeting.

5 Main Analysis

Now we analyze the main model with two firms. We solve the equilibrium by backward induction.

5.1 Targeting Decision

Given firm j's choice of the tuning parameter as λ_j and his private dataset, his analyst's estimates, $\hat{\alpha}(\lambda_j)$ and $\hat{\beta}(\lambda_j)$ are given by equation (5). Suppose firm j decides to target k_j consumers with $x = 1$ and $\theta - k_j$ consumers with $x = 0$ for $j = 1, 2$. By definition, we have that $0 \leq k_j \leq \phi$ and $0 \leq \theta - k_j \leq 1 - \phi$, which imply that $\max\{0, \theta + \phi - 1\} \leq k_j \leq \min\{\theta, \phi\}$. Firm j does not observe the other firm's choice of the tuning parameter nor his dataset. Denote firm j's expectation of the other firm's choice of the tuning parameter as λ^*_{-j}. Furthermore, from firm j's perspective, the other firm's choice of k_{-j} depends on the realization of his private dataset and thus is a random variable, which is denoted as $\widetilde{k}_{-j}$. Let's calculate firm j's estimated profit:

$$\begin{aligned}
\Pi_j(k_j, \widetilde{k}_{-j}) =& k_j \left(\frac{\widetilde{k}_{-j}}{\phi} \cdot \frac{1}{2} + 1 - \frac{\widetilde{k}_{-j}}{\phi} \right) \left(\hat{\alpha}_j(\lambda_j) + \hat{\beta}_j(\lambda_j) \right) \\
&+ (\theta - k_j) \left(\frac{\theta - \widetilde{k}_{-j}}{1 - \phi} \cdot \frac{1}{2} + 1 - \frac{\theta - \widetilde{k}_{-j}}{1 - \phi} \right) \hat{\alpha}_j(\lambda_j) \\
=& \theta \left(1 - \frac{\theta - \widetilde{k}_{-j}}{2(1 - \phi)} \right) \hat{\alpha}_j(\lambda_j) \\
&+ k_j \left(\frac{\phi\theta - \widetilde{k}_{-j}}{2\phi(1 - \phi)} \hat{\alpha}_j(\lambda_j) + \left(1 - \frac{\widetilde{k}_{-j}}{2\phi} \right) \hat{\beta}_j(\lambda_j) \right). \qquad (7)
\end{aligned}$$

To understand the first equation above, notice that firm j targets k_j consumers with $x = 1$, each of whom is also targeted by the other firm $-j$ with probability $\widetilde{k}_{-j}/\phi$. If this happens, firm j gets an estimated profit of $\left(\hat{\alpha}_j(\lambda_j) + \hat{\beta}_j(\lambda_j) \right)/2$; otherwise, with probability $1 - \widetilde{k}_{-j}/\phi$, this consumer is not targeted by firm $-j$, and firm j's estimated profit is $\left(\hat{\alpha}_j(\lambda_j) + \hat{\beta}_j(\lambda_j) \right)$. Similarly, we can perform the same calculation to get firm j's estimated profit from $\theta - k_j$ consumers with $x = 0$.

Firm j chooses $k_j \in [\max\{0, \theta + \phi - 1\}, \min\{\theta, \phi\}]$ to maximize the expected estimated profit, $\mathrm{E}\big[\Pi_j(k_j, \widetilde{k}_{-j})\big] = \Pi_j(k_j, \mathrm{E}[\widetilde{k}_{-j}])$, where we have utilized the observation that $\Pi_j(k_j, \widetilde{k}_{-j})$ is linear in $\widetilde{k}_{-j}$.[3] Furthermore, notice that $\Pi_j(k_j, \mathrm{E}[\widetilde{k}_{-j}])$ is linear in k_j

[3]Notice that as α_1, β_1, α_2 and β_2 are independently distributed, firm j's private dataset provides information only on α_j and β_j but has no implication for α_{-j} and β_{-j}. Therefore, $\mathrm{E}[\widetilde{k}_{-j}|\text{firm } j\text{'s dataset}] = \mathrm{E}[\widetilde{k}_{-j}]$.

with

$$\frac{\partial \Pi(k_j, \mathrm{E}[\widetilde{k}_{-j}])}{\partial k_j} = \underbrace{\frac{\phi\theta - \mathrm{E}[\widetilde{k}_{-j}]}{2\phi(1-\phi)}\hat{\alpha}_j(\lambda_j)}_{\text{to avoid competition}} + \underbrace{\left(1 - \frac{\mathrm{E}[\widetilde{k}_{-j}]}{2\phi}\right)\hat{\beta}_j(\lambda_j)}_{\substack{\text{to target the more} \\ \text{profitable segment}}}. \tag{8}$$

The expression of $\partial_{k_j}\Pi(k_j, \mathrm{E}[\widetilde{k}_{-j}])$ in equation (8) consists of two terms. The second term plays a similar role with the counterpart under the monopoly benchmark—the firm wants to target consumers with $x = 1$ when $\hat{\beta}_j(\lambda_j) > 0$, and $x = 0$ when $\hat{\beta}_j(\lambda_j) < 0$. The first term introduces incentives for the two firms to coordinate so as to avoid competition. Particularly, firm j wants to target consumers with $x = 1$ when $\mathrm{E}[\widetilde{k}_{-j}]/\theta < \phi$, that is, when the other firm target proportionally more consumers with $x = 0$; similarly, firm j wants to target consumers with $x = 0$ when $\mathrm{E}[\widetilde{k}_{-j}]/\theta > \phi$, that is, when the other firm target proportionally more consumers with $x = 1$.

$\Pi_j(k_j, \mathrm{E}[\widetilde{k}_{-j}])$ being linear in k_j together with equation (8) immediately implies the firm's optimal targeting decision. Specifically, if $\partial_{k_j}\Pi(k_j, \mathrm{E}[\widetilde{k}_{-j}]) > 0$, firm j should set $k_j = \min\{\theta, \phi\}$ to target as many consumers with $x = 1$ as possible; if $\partial_{k_j}\Pi(k_j, \mathrm{E}[\widetilde{k}_{-j}]) < 0$, he should set $k_j = \max\{0, \theta + \phi - 1\}$ to target as many consumers with $x = 0$ as possible. Lastly, from an ex-ante perspective before the realization of firm j's private dataset, $\hat{\alpha}_j(\lambda_j)$ follows a continuous distribution and thus $\partial_{k_j}\Pi(k_j, \mathrm{E}[\widetilde{k}_{-j}]) = 0$ is a knife-edge case that happens with zero probability; consequently, the tie-breaking rule for which consumer to target at $\partial_{k_j}\Pi(k_j, \mathrm{E}[\widetilde{k}_{-j}]) = 0$ has no consequence for the firm's choice of λ_j, which we analyze next.

5.2 Model Selection

Let's first introduce the following notation. From firm, j's perspective, the probability that the other firm $-j$ will set $\widetilde{k}_{-j} = \min\{\theta, \phi\}$ is:

$$\begin{aligned} p_{-j} &\equiv \Pr\left(\widetilde{k}_{-j} = \min\{\theta, \phi\}\right) \\ &= \Pr\left(\frac{\phi\theta - \mathrm{E}[\widetilde{k}_j]}{2\phi(1-\phi)}\hat{\alpha}_{-j}(\lambda^*_{-j}) + \left(1 - \frac{\mathrm{E}[\widetilde{k}_j]}{2\phi}\right)\hat{\beta}_{-j}(\lambda^*_{-j}) > 0\right) \\ &= \Pr\left(\frac{\phi\theta - (\max\{0, \theta + \phi - 1\} + p_j \min\{\theta, 1-\theta, \phi, 1-\phi\})}{2\phi(1-\phi)}\hat{\alpha}_{-j}(\lambda^*_{-j})\right. \end{aligned}$$

$$+\left(1-\frac{\max\{0,\theta+\phi-1\}+p_j\min\{\theta,1-\theta,\phi,1-\phi\}}{2\phi}\right)\hat{\beta}_{-j}(\lambda^*_{-j})>0\Bigg), \quad (9)$$

where, to get the last equality in (9), we have utilized that

$$\begin{aligned}\mathrm{E}[\widetilde{k}_j] &= p_j\min\{\theta,\phi\}+(1-p_j)\max\{0,\theta+\phi-1\}\\ &= \max\{0,\theta+\phi-1\}+p_j\min\{\theta,1-\theta,\phi,1-\phi\}.\end{aligned}$$

By combining equation (9) for $j=1,2$, we should be able to solve p_j for $j=1,2$.

Below, we will restrict our attention to the symmetric equilibrium with $\lambda_1^*=\lambda_2^*=\lambda^*$, $\widetilde{k}_1=\widetilde{k}_2=\widetilde{k}$ and correspondingly, $p_1=p_2=p^*$. In the symmetric equilibrium, we have

$$\begin{aligned}p^* = \Pr\Bigg(&\frac{\phi\theta-(\max\{0,\theta+\phi-1\}+p^*\min\{\theta,1-\theta,\phi,1-\phi\})}{2\phi(1-\phi)}\hat{\alpha}_j(\lambda^*)\\ &+\left(1-\frac{\max\{0,\theta+\phi-1\}+p^*\min\{\theta,1-\theta,\phi,1-\phi\}}{2\phi}\right)\hat{\beta}_j(\lambda^*)>0\Bigg). \quad (10)\end{aligned}$$

Given λ^*, we can solve p^* from equation (10).

Next, we determine λ^* by calculating firm j's expected profit before obtaining the private dataset, which takes the same form as the firm's estimated profit $\Pi_j(k_j,\widetilde{k}_{-j})$ in equation(7) except that we need to replace $\hat{\alpha}_j(\lambda_j)$ and $\hat{\beta}_j(\lambda_j)$ by α_j and β_j respectively and then take expectation.

$$\begin{aligned}&\Pi_j(\lambda_j)\equiv \mathrm{E}\left[\theta\left(1-\frac{\theta-\widetilde{k}_{-j}}{2(1-\phi)}\right)\alpha_j+k_j\left(\frac{\phi\theta-\widetilde{k}_{-j}}{2\phi(1-\phi)}\alpha_j+\left(1-\frac{\widetilde{k}_{-j}}{2\phi}\right)\beta_j\right)\right]\\
&=\theta\left(1-\frac{\theta-\mathrm{E}[\widetilde{k}_{-j}]}{2(1-\phi)}\right)\mathrm{E}\left[\alpha_j\right]\\
&+\Pr\left(\frac{\phi\theta-\mathrm{E}[\widetilde{k}_{-j}]}{2\phi(1-\phi)}\hat{\alpha}_j(\lambda_j)+\left(1-\frac{\mathrm{E}[\widetilde{k}_{-j}]}{2\phi}\right)\hat{\beta}_j(\lambda_j)>0\right)\min\{\theta,\phi\}\\
&\times\mathrm{E}\left[\frac{\phi\theta-\widetilde{k}_{-j}}{2\phi(1-\phi)}\alpha_j+\left(1-\frac{\widetilde{k}_{-j}}{2\phi}\right)\beta_j\middle|\frac{\phi\theta-\mathrm{E}[\widetilde{k}_{-j}]}{2\phi(1-\phi)}\hat{\alpha}_j(\lambda_j)+\left(1-\frac{\mathrm{E}[\widetilde{k}_{-j}]}{2\phi}\right)\hat{\beta}_j(\lambda_j)>0\right]\\
&+\Pr\left(\frac{\phi\theta-\mathrm{E}[\widetilde{k}_{-j}]}{2\phi(1-\phi)}\hat{\alpha}_j(\lambda_j)+\left(1-\frac{\mathrm{E}[\widetilde{k}_{-j}]}{2\phi}\right)\hat{\beta}_j(\lambda_j)<0\right)\max\{0,\theta+\phi-1\}\\
&\times\mathrm{E}\left[\frac{\phi\theta-\widetilde{k}_{-j}}{2\phi(1-\phi)}\alpha_j+\left(1-\frac{\widetilde{k}_{-j}}{2\phi}\right)\beta_j\middle|\frac{\phi\theta-\mathrm{E}[\widetilde{k}_{-j}]}{2\phi(1-\phi)}\hat{\alpha}_j(\lambda_j)+\left(1-\frac{\mathrm{E}[\widetilde{k}_{-j}]}{2\phi}\right)\hat{\beta}_j(\lambda_j)<0\right].\end{aligned}$$

The equilibrium choice of the tuning parameter, $\lambda^* = \arg\max_{\lambda_j} \Pi_j(\lambda_j)$. The following proposition presents the main result of the paper.

Proposition 2. *When $\phi \neq 1/2$ and $\mathrm{E}[\alpha_j]$ is sufficiently high, $\lambda^* > 0$.*

The complete proof to Proposition 2 is provided in Appendix. Below, we sketch the intuition behind the proposition by focusing on the case with $\theta \leq \min\{\phi, 1-\phi\}$, so that there are enough number of consumers in each segment for a firm to target. The alternative case is relegated to Appendix.

5.3 Intuition for the Main Result

First, notice that we can without loss of generality assume that $\phi \in (0, 1/2)$. The other case with $\phi \in (1/2, 1)$ can be obtained by symmetry and thus is omitted. Let's introduce $\bar{\lambda}$, defined implicitly by the following equation:

$$1 - \phi = \Pr\left(\beta_j + \Delta\varepsilon_j \leq \bar{\lambda}\right) = \int_{-2\bar{\varepsilon}}^{2\bar{\varepsilon}} B(\bar{\lambda} - e) d\widetilde{G}(e).$$

By definition, $\phi \in (0, 1/2)$ is equivalent to $\bar{\lambda} \in (0, \overline{\beta} + 2\bar{\varepsilon})$. Let's prove the following lemma.

Lemma 1. *If $0 < \theta \leq \phi < 1/2$ and $\lambda^* < \bar{\lambda}$, then $p^* \in (\phi, 1/2)$.*

Proof. Given $0 < \theta \leq \phi < 1/2$, the right hand side (RHS) of equation (10) can be rewritten as

$$\Pr\left(\frac{1}{1-\phi}\frac{\theta\phi - \theta p^*}{2\phi - \theta p^*}\hat{\alpha}_j(\lambda^*) + \hat{\beta}_j(\lambda^*) > 0\right),$$

which obviously decreases continuously with p^*. Given $p^* = \phi$, RHS$= \Pr(\hat{\beta}_j(\lambda^*) > 0) = \Pr(\beta_j + \Delta\varepsilon_j > \lambda^*) > \Pr(\beta_j + \Delta\varepsilon_j > \bar{\lambda}) = \phi = p^*$; on the other hand, given $p^* = 1/2$, RHS$= \Pr\left(\hat{\beta}_j(\lambda^*) > \frac{1}{1-\phi}\frac{\theta(1/2-\phi)}{2\phi - \theta p^*}\hat{\alpha}_j(\lambda^*)\right) < 1/2 = p^*$. This implies that $p^* \in (\phi, 1/2)$. □

Lemma 1 implies that $p^* < 1/2$. In fact, this holds regardless of whether $\lambda^* < \bar{\lambda}$. To understand, notice that equation (8) shows that when deciding which consumer segment to target, a firm has two considerations: to target more profitable segment and to avoid competition. Because $\hat{\beta}_j(\lambda_j)$ is symmetrically distributed, if the firm only wanted to achieve the first goal of targeting more profitable segment, he would set $p^* = 1/2$. The second goal of avoiding competition drives the firm to reduce the targeting

allocation for consumer segment with $x = 1$, which is more competitive, because $\phi < 1/2$. This leads to $p^* < 1/2$. Furthermore, Lemma 1 shows that $p^* > \phi$ holds when $\lambda^* < \bar{\lambda}$; otherwise, when λ^* is sufficiently high, the firm forgoes the first goal of targeting more profitable segment and will thus set $p^* \to 0$.

Next, under $0 < \theta \leq \phi < 1/2$ and utilizing the expression of $\hat{\alpha}_j(\lambda_j^*)$ and $\hat{\beta}_j(\lambda_j^*)$ in equations (4) and (5), $\Pi_j(\lambda_j)$ can be simplified as:

$$\Pi_j(\lambda_j) = \theta\left(1 - \frac{(1-p^*)\theta}{2(1-\phi)}\right)\mathrm{E}[\alpha_j] - \frac{(p^*-\phi)\theta^2}{2\phi(1-\phi)}\Pr(\mathscr{A})\,\mathrm{E}[\alpha_j|\mathscr{A}] + \frac{\theta(2\phi-\theta p^*)}{2\phi}\Pr(\mathscr{A})\,\mathrm{E}[\beta_j|\mathscr{A}], \quad (11)$$

where

$$\mathscr{A} \equiv \left\{ 2\frac{(1-\phi)(2\phi-\theta p^*)(\beta_j+\varepsilon_j^1) - \theta(p^*-\phi)\alpha_j - \phi\left(2(1-\phi)-\theta(1-p^*)\right)\varepsilon_j^0}{4\phi(1-\phi)-\theta\left(\phi+(1-2\phi)p^*\right)} > \lambda_j \right\}.$$

Let's assume $\lambda^* < \bar{\lambda}$ for the following discussion so that we have $p^* \in (\phi, 1/2)$ by Lemma 1. There are three terms in the expression of $\Pi_j(\lambda_j)$ in equation (11). The first term does not depend on λ_j. For the second term, notice that both $\Pr(\mathscr{A})$ and $\mathrm{E}[\alpha_j|\mathscr{A}]$ decreases with λ_j and thus the second term increases with λ_j. The third term plays a similar role with its counterpart in the monopoly benchmark. Because of the second term, we can have $\Pi_j'(\lambda_j) > 0$ for $\mathrm{E}[\alpha_j]$ sufficiently high. This will lead to an equilibrium with $\lambda^* > 0$. Intuitively, a higher penalization of λ_j reduces the overlapping of the two firms' targeting in the competitive segment of $x = 1$.

Furthermore, notice that Proposition 2 requires that $\phi \neq 1/2$; otherwise, the two segments would be equally competitive so there is no incentive for the firms to choose a positive penalization to reduce overlapping in the more competitive segment. More generally, when B, the distribution of β_j is not symmetric around zero, we can have a different critical value, at which, the two segments are equally competitive. Still, except for the knife-edge case where ϕ is equal to the critical value, the two segments will not be equally competitive, which creates incentives for the firms to choose a positive penalization to reduce overlapping in competition.

6 Conclusion

In this paper, we explore how competitive firms employ algorithms to estimate demand and based on the estimates, make strategic targeting decisions so as to maximize expected profit. Different algorithms essentially mean different model selections, which tradeoff bias and variance differently under the general framework of supervised learning. From this perspective, our paper essentially studies firms' competition in model selections for algorithmic targeting and explores how competition moderates individual firms' bias-variance tradeoff. The main finding is that competition favors simpler models that reduce variance but introduce bias.

We focus on one specific decision of the firms—targeting. Historically, this is mostly done through market research. Recently, partly thanks to large advertising platforms such as Facebook, there is an ongoing trend that targeting decisions have been automated by algorithms in real time based on real customer behavioral data on browse and purchase, etc. Moreover, it has been the focus of several recent studies to develop new research designs that combine machine learning algorithms and randomized experiments for targeting (e.g., Hitsch and Misra 2018; Simester et al. 2020; Rafieian and Yoganarasimhan 2021). Therefore, we consider targeting to be a great starting point to study algorithmic competition. Building on this research, it would be interesting to explore a general class of strategic decisions of the firms, such as pricing, advertising and product design, etc. The implications may depend on whether the firms' decisions are strategic substitutes or complements (Bulow et al. 1985).

References

Dirk Bergemann and Alessandro Bonatti. Targeting in advertising markets: implications for offline versus online media. *The RAND Journal of Economics*, 42(3):417–443, 2011.

Daniel Björkegren, Joshua E Blumenstock, and Samsun Knight. Manipulation-proof machine learning. *arXiv preprint arXiv:2004.03865*, 2020.

Jeremy I Bulow, John D Geanakoplos, and Paul D Klemperer. Multimarket oligopoly: Strategic substitutes and complements. *Journal of Political economy*, 93(3):488–511, 1985.

Gerard R Butters. Equilibrium distributions of sales and advertising prices. *The Review of Economic Studies*, pages 465–491, 1977.

Emilio Calvano, Giacomo Calzolari, Vincenzo Denicolo, and Sergio Pastorello. Artificial intelligence, algorithmic pricing, and collusion. *American Economic Review*, 110 (10):3267–97, 2020.

Yuxin Chen, Chakravarthi Narasimhan, and Z John Zhang. Individual marketing with imperfect targetability. *Marketing Science*, 20(1):23–41, 2001.

Kfir Eliaz and Ran Spiegler. The model selection curse. *American Economic Review: Insights*, 1(2):127–40, 2019.

Jerome Friedman, Trevor Hastie, Robert Tibshirani, et al. *The elements of statistical learning*, volume 1. Springer series in statistics New York, 2001.

Günter J Hitsch and Sanjog Misra. Heterogeneous treatment effects and optimal targeting policy evaluation. *Available at SSRN 3111957*, 2018.

Ganesh Iyer, David Soberman, and J Miguel Villas-Boas. The targeting of advertising. *Marketing Science*, 24(3):461–476, 2005.

Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. *An introduction to statistical learning*, volume 112. Springer, 2013.

Annie Liang. Games of incomplete information played by statisticians. *Available at SSRN 2931873*, 2020.

José Luis Montiel Olea, Pietro Ortoleva, Mallesh M Pai, and Andrea Prat. Competing models. *arXiv preprint arXiv:1907.03809*, 2019.

Omid Rafieian and Hema Yoganarasimhan. Targeting and privacy in mobile advertising. *Marketing Science*, 40(2):193–218, 2021.

Yuval Salant and Josh Cherry. Statistical inference in games. *Econometrica*, 88(4):1725–1752, 2020.

Greg Shaffer and Z John Zhang. Competitive coupon targeting. *Marketing Science*, 14 (4):395–416, 1995.

Duncan Simester, Artem Timoshenko, and Spyros I Zoumpoulis. Targeting prospective customers: Robustness of machine-learning methods to typical data challenges. *Management Science*, 66(6):2495–2522, 2020.

Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288, 1996.