Title

Competitive Model Selection in Algorithmic Targeting

Authors

Iyer, Ganesh
Ke, T. Tony

Publication Date

2023-04-03

DOI

10.26085/C32C7B

Series Name: WPS
WPS No. 221
Issue Date: 3/29/22

Competitive Model Selection in Algorithmic Targeting

Ganesh Iyer and T. Tony Ke

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

February 2023

*We thank Ying Zhu for her inspiring discussion that helped us initiate the research. This paper has also benefited from the comments of Anthony Dukes, Przemyslaw Jeziorski, Joe Perkins, Jeanine Miklos-Thal, Ran Spiegler and Jingbo Wang as well as seminar participants of UBC, UNC Chapel Hill, Peking University, PolyU, and the attendees of 2022 Marketing Science Doctoral Consortium, 2022 NBER Economics of AI Conference, 15th Digital Economics Conference at Toulouse, 2023 Hong Kong Quantitative Marketing Conference, and the 2023 UT Dallas Bass FORMS Conference. Wee Chaimanowong provided excellent research assistance. Ke acknowledges financial support from the General Research Fund provided by Hong Kong Research Grants Council under project number 14503122.

Competitive Model Selection in Algorithmic Targeting

Abstract

We study how market competition influences the algorithmic design choices of firms in the context of targeting. Firms face a general bias-variance trade-off when choosing the design of a supervised learning algorithm in terms of model complexity or the number of predictors. Each firm has a data analyst who uses the chosen algorithm to estimate demand for multiple consumer segments, based on which, it devises a targeting policy to maximize estimated profits. We show that competition induces firms to strategically choose simpler algorithms which involve more bias but lower variance. Therefore, more complex/flexible algorithms may have higher value for firms with greater monopoly power.

Keywords: *algorithmic competition, model selection, algorithmic bias, data analytics, targeting*

1 Introduction

The digital economy has made available unparalleled amount of consumer data to firms. Over the past decade firms have increasingly delegated many business decisions, such as pricing, advertising and targeting, to artificial intelligence (AI) algorithms, which utilize large amount of data on consumer characteristics and behaviors. 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 behaviors that is present can be comparable to the size of the dataset. Consequently, big-data environments can confront the firms with the classic over-fitting problem in statistical learning: the algorithm may use a large number of available consumer predictors and complex functions to map the data onto predictions of consumer behaviors. However, this increases the variance of the estimated predictions and thus reduces the precision of out-of-sample predictions. Alternatively, the algorithm can be regularized wherein the complex functions can be penalized leading to a selection of only the most relevant variables. This would reduce the variance of the estimated predictions but then may introduce biases in the estimates and thereby compromise prediction accuracy. This is the general bias-variance trade-off that underlies the design of any supervised learning algorithm. In this paper we examine this trade-off that underlies the design of the algorithm under market competition.

In particular, we recognize that the function of an AI algorithm is to make predictions (Agrawal et al. 2018) in order to facilitate decision making. In competitive market settings, when firms make decisions based on algorithmic predictions, the degree of market competition should have implications for their choice of the algorithmic design. In this paper we analyze how competition influences firms' algorithmic design choices that govern the bias variance trade-off in model selection. Algorithms that are more complex and flexible tend to have higher variance but lower bias, and therefore the bias variance trade-off also implies the interpretability versus flexibility trade-off in algorithmic design. We choose the important context of targeting (or targeted advertising) to study this trade-off. Targeting is a canonical business application of AI algorithms that leverages big data on consumers.[1]

We model firms' algorithmic decision making process as involving two stages. First, a firm chooses an algorithmic design in terms of a statistical model and then fits the model to the available data. Second, it makes competitive targeting decisions based on the model estimates. This two-stage setup is intended to represent common industry practices and to ensure that the firms' choices of algorithmic designs will impact their strategic targeting decisions. However, this setup also inevitably departs from the standard Bayesian approach, because for Bayesian decision makers, data are informative signals that always update their belief by Bayes' rule, and consequently, there is no active role that an algorithmic design could play. Nevertheless in practice firms use statistical algorithms on data to target consumers. Our setup rationalizes the inconsistency between observed practice with the Bayesian approach by considering delegation of a firm's data analytics process to an analyst in an incomplete contracting framework. In other words, we view a data analytics algorithm as one way of representing consumer data for the firms' decision making so that different algorithmic designs amount to different representations of consumer information, which induces the firms to make different decisions.

We operationalize the firms' algorithmic design problem of model selection by using the example of a well-known supervised learning algorithm—the Lasso regression, which selects variables via the penalization of variable coefficients (Tibshirani 1996). The penalization governs the extent of the prediction accuracy and model interpretability. The Lasso regression is a natural choice for our purpose because the bias-variance tradeoff is directly modulated by the degree of penalization, or the choice of a hyperparameter. The model is relevant for practical usage while at the same time allowing for analytical tractability.

Specifically, consider a market in which two firms compete by targeting consumers who are heterogeneous in some characteristic. Targeting is costly and acts as a form of informative advertising (Butters 1977). Firms observe consumer characteristics from their data but are uncertain about the profitability of

[1] Using AI algorithms to automate targeting decisions has been the focus of several recent empirical studies (e.g., Hitsch and Misra 2018; Simester et al. 2020; Rafieian and Yoganarasimhan 2021).

different consumer types. Firms can estimate this profitability via a statistical algorithm. To this end, in the first stage the firms strategically choose the algorithmic design which is operationalized as the degree of penalization or the extent of complexity of a Lasso regression. Based on the design choice, the firms in the second stage delegate the task of running predictive algorithm to a data analyst who has the capability of running the chosen Lasso regression model on the available data. Lastly, based on the model estimates reported by one's analyst, each firm chooses the targeting strategy to maximize the estimated profit.

We first analyze the monopoly benchmark and show that it is optimal for the firm to choose zero penalization. In other words, a monopoly firm prefers a more complex or flexible algorithmic design which admits greater variance but has lower bias. This enables the firm to achieve greater market coverage in the sense that it allows it to target the more profitable consumer segment with greater likelihood. We proceed to analyze the competitive market and find that in equilibrium, at least one firm will choose positive penalization. This introduces bias while reducing variance by selecting fewer variables for the predictive model. In other words, competition favors simpler models for consumer targeting in equilibrium. Under competition firms have two incentives: i) to correctly target the more profitable segment, and ii) to avoid competition and the overlap in targeting. Simpler models which involve bias while reducing variance lead to more uniform targeting, which helps to reduce overlap and soften competition. Overall, the suggestion of our analysis is that more flexible and complex algorithms such as deep learning are likely to be of higher value to firms with greater monopoly power.

2 Related Research

Our paper is broadly related to the emerging literature which examines strategic interactions and incentives with algorithms. One strand of research tackles the problem of algorithmic design for a principal when faced with strategic agents who can manipulate the information that is provided to the algorithm. For example, Eliaz and Spiegler (2019) examines a statistical algorithm faced with an agent who strategically self-reports her personal data and highlights the role of model selection and the incentive-compatibility issues in truthful reporting that it creates for the agent. In a similar vein, 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 given the costs of manipulating different behaviors. Our paper examines the model selection problem in a competitive market where firms choose the equilibrium design of their consumer targeting algorithms. Thus here the extent to which firms choose more or less flexible algorithms and the associated bias-variance trade-off is governed by the equilibrium consumer targeting incentives of competing firms.

There is a stream of research 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. (2022) 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. Here, in contrast, we focus on the strategic choice of algorithms in competitive environments.

There is also recent research on how algorithmic decision making affects market competition, a question complementary to ours. For example, Miklós-Thal and Tucker (2019) and O'Connor and Wilson (2021) model the effect of AI algorithms as better demand forecasting and show that algorithms could impede or facilitate tacit price collusion. Calvano et al. (2020) examine firms endowed with Q-learning algorithms in repeated interactions to show that they can robustly learn to cooperate to charge supra-competitive prices without 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; Levin and Milgrom 2010; 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 $x \in \{1, 0\}$. A fraction ϕ 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 decide 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. This targeting can be also interpreted as a form of costly informative advertising that informs consumers of 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. To focus the exposition on the effects of algorithmic targeting, we have abstracted away the firms' decisions on prices.[2]

Given that x is binary, it is without loss of generality to write down $\pi_j(x)$ as

[2] If price discrimination based on targeting outcomes is allowed, we may 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 without explicit consideration of prices.

the following linear function,

$$\pi_j(x) = \alpha_j + \beta_j x.$$

Firm j does not know α_j, β_j a priori. We assume a common prior for α_j, β_j, which follow differentiable 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 analyst who is equipped with the prediction algorithm. 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. This is the algorithmic design decision. 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

$$y_j^l = \pi_j(x^l) + \varepsilon_j^l = \alpha_j + \beta_j x^l + \varepsilon_j^l.$$

The error term, ε_j^l is i.i.d. across j and l and follows a differentiable distribution function G, which is symmetric around zero and supported in $[-\overline{\varepsilon}, \overline{\varepsilon}]$. Further define $\Delta\varepsilon_j \equiv \varepsilon_j^1 - \varepsilon_j^0$, which follows distribution function $\widetilde{G}$, where $\widetilde{G}(e) = \Pr(\varepsilon_j^1 - \varepsilon_j^0 \leq e) = \int_{-\overline{\varepsilon}}^{\overline{\varepsilon}} G(e' + e) dG(e')$. We make the following assumption.

Assumption 1. *$\widetilde{G}$ is a unimodal distribution; that is, $\widetilde{G}'(e)$ weakly decreases (increases) with e for $e > 0$ ($e < 0$).*

This is a regularity condition which ensures that the firms' ex-ante expected profit functions are well behaved. Note that the dataset that each firm uses for targeting is assumed to be exogenous and independent of the ensuing market competition. One interpretation of this setup is that each firm is able to experiment/test-market with the data analytics algorithm on consumers in its monopolistic sub-markets (such as geographic regions or sales channels). This would generate a monopolistic private dataset for each firm. In Section 6, we will describe an alternative setting in which the dataset results from market competition, and argue that it would nevertheless generate results that are qualitatively similar to that in the main model.

Based on the data, the analyst runs a Lasso regression, which is represented by the following minimization problem:

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

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 model flexibility and choosing all the available predictor variables. This will imply estimated parameters which are unbiased but which will have maximum variance. 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. In this case the variance of the estimated parameters would be lowered but at the cost of introducing bias.

From the corresponding first- and second-order optimality conditions, we can solve the data analyst's estimation problem in equation (1):

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

$$\hat{\beta}_j(\lambda_j) = \begin{cases} \max\{y_j^1 - y_j^0 - \lambda_j, 0\}, & \text{if } y_j^1 - y_j^0 \geq 0, \\ \min\{y_j^1 - y_j^0 + \lambda_j, 0\}, & \text{otherwise.} \end{cases} \tag{3}$$

The expression of $\hat{\alpha}_j(\lambda_j)$ in equation (2) 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 for any $\lambda_j \geq 0$. Formally,

Assumption 2. $\underline{\alpha} > \overline{\beta}/2 + \bar{\varepsilon}$.

This guarantees that firm j always prefers to target as many consumers as possible in the market. That is, the constraint of a total number of θ consumers to target will always be binding so that the firm's targeting decision boils down to which type(s) of consumers to target. 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. When $0 < \lambda_j < |y_j^1 - y_j^0|$, then $\hat{\beta}_j(\lambda_j)$ will have the same sign as $y_j^1 - y_j^0$ but is 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 gets a private dataset (x^l, y_j^l) for $l = 0, 1$, based on which, firm j's analyst generates the estimates $\hat{\alpha}_j(\lambda_j)$ and $\hat{\beta}_j(\lambda_j)$ by employing a Lasso regression. Lastly, each firm devises the targeting strategy to maximize the estimated profit. Figure 1 summarizes the timeline of the game. Because each firm has a private dataset before deciding its targeting decision, we are dealing with a Bayesian game. However, because the two firms' datasets are independent, one firm does not need to update its belief about the other firm's dataset based on the observation of its own dataset; instead, it just uses the prior belief. Before we proceed to analyze the game, we elaborate on the rationale and interpretation of our modeling choices.

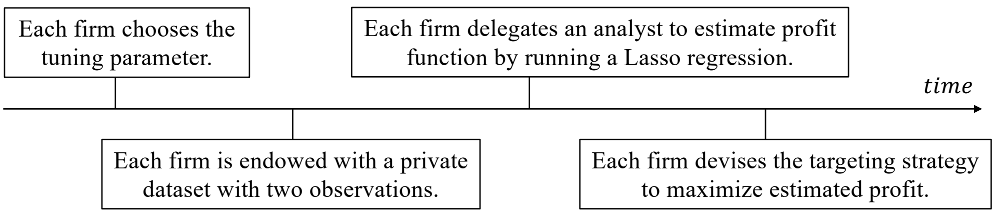


Figure 1: Timeline of the competitive algorithmic targeting game.

First, the reader may wonder that the simple setup above with a data set of just two observations and a binary characteristic ($x \in (1, 0)$) is a far cry from the big data situations confronting firms. Machine learning models are typically high dimensional and complex involving numerous dimensions available in big data. Nevertheless, as also previously argued by Eliaz and Spiegler (2019) the setup is designed to handle the crucial aspects of the "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 dataset but would have high variance and poor predictive performance compared to an estimator with shrinkage. However, a model with shrinkage may be subject to the introduction of bias in the estimated coefficients. The model with the Lasso regression 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, it endogenizes the model selection to the equilibrium incentives of the firms.

Second, the firms choose the tuning parameters before getting the dataset. In other words, firms make the model selection choice in anticipation of the possible dataset realizations. This may also be seen as 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 viewed 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 their flexibility and their degree of interpretability, as shown by Figure 2. Typically, those with higher flexibility (and lower interpretability) 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 and higher bias and may also be more easily interpretable when compared to OLS. Therefore, the choice of λ_j may also represent the relative complexity versus interpretability of the algorithm. Also by this understanding, the Lasso regression does not necessarily need to represent a "machine-

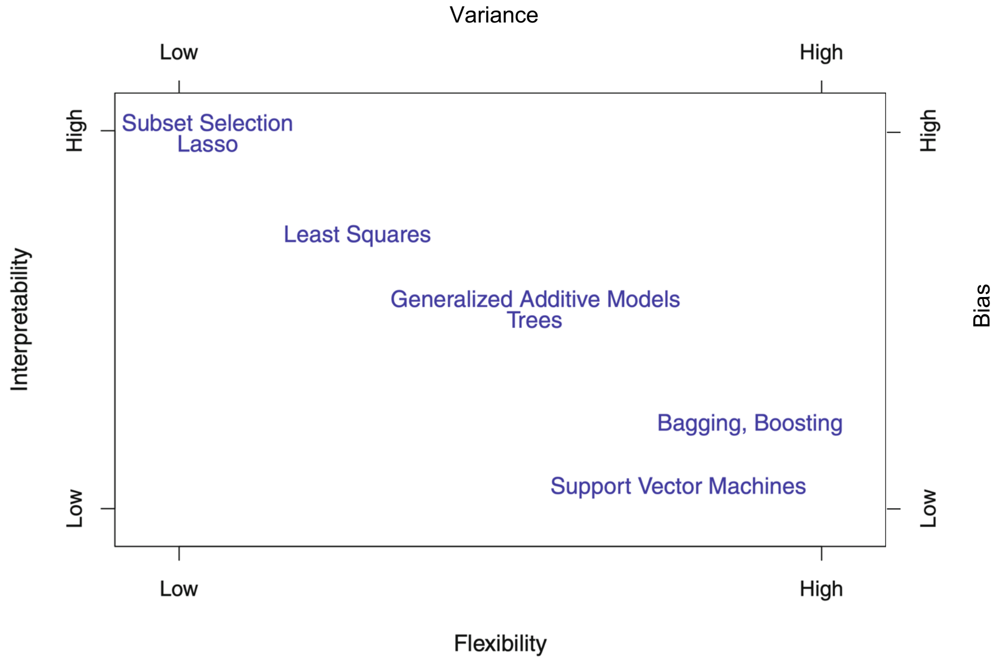


Figure 2: Tradeoff between flexibility and interpretability and tradeoff between bias and variance across different statistical learning methods (excerpted from James et al. (2013) page 25 and adapted).

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 rely on data analysts for running the estimation procedure on the data and that complete contracts are not available between a firm and its analyst. This assumption maps onto common practices in companies where managers rely on analysis by data analytics groups for decision-making. 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 es-

timate, $\hat{\alpha}_j(\lambda_j)$ 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 (1) are non-bayesian procedures. By separating the estimation problems from the firms, and delegating it to analysts who are agents, we are able to rationalize the reality of data-driven decision making in firms. Methodologically this feature of our framework can be seen as a representation of algorithmic decision making in firms.[3]

2. Our assumption that the data analyst performs the estimation procedure implies the minimization of mean squared error instead of profit maximization as the objective in estimating the parameters in the second stage. This assumption is made to reflect actual observed industry practices and their implications. First, minimization of mean squared error is available and used by companies in standard statistical packages. 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 find it hard to gather all relevant information. Finally, using standard statistical/algorithmic packages on the data may be seen as computationally and cognitively easier for analysts in a firm than performing Bayesian updating based on the data to compute expected profits.

We begin with the analysis of the monopoly setting with only one firm in

[3]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 (1) (Tibshirani 1996). But there are two caveats to this Bayesian approach. First, the tuning parameter λ_j is not firm j's choice but rather, a model primitive that is exogenously given. To endogenize the firm's choice of λ_j would be equivalent to let the firm choose its prior distribution. Second, the point estimates generated by the Lasso regression, $\hat{\alpha}(\lambda_j)$ and $\hat{\beta}(\lambda_j)$ in equations (2) and (3) are mode instead of mean of the posterior belief of α_j and β_j (Hastie et al. 2009). However, to calculate expected profit, we will be mostly concerned with the posterior mean instead of the mode.

the market as the benchmark, and then proceed to study the main model with competition.

4 Monopoly Benchmark

Given only one firm, we will drop the subscript j. We solve the game by backward induction. Suppose the firm decides to target $k \in [0, \phi]$ consumers with $x = 1$ and $\theta - k \in [0, 1-\phi]$ consumers with $x = 0$, which imply that

$$\max\{0, \theta + \phi - 1\} \leq k \leq \min\{\theta, \phi\}.$$

Given $\hat{\alpha}(\lambda)$ and $\hat{\beta}(\lambda)$, we have the estimated profit from a targeted consumer to be $\hat{\pi}(x) = \hat{\alpha}(\lambda) + \hat{\beta}(\lambda)x$. The firm chooses k to maximize the estimated profit. If $\hat{\beta}(\lambda) > 0$, it is optimal for the firm to target as many consumers with $x = 1$ as possible, so we have the firm's optimal choice of k as $k^* = \min\{\theta, \phi\}$. Similarly, if $\hat{\beta}(\lambda) < 0$, it is optimal for to target as many consumers with $x = 0$ as possible, and thus, $k^* = \max\{0, \theta + \phi - 1\}$. Lastly, if $\hat{\beta}(\lambda) = 0$, the firm is indifferent between the two types of consumers, and it is assumed that it will target $k \in [0, \theta]$ consumers with $x = 1$.

A priori, before obtaining the dataset, the firm chooses λ to maximize the expected profit from all consumers:

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

$$=\theta\mathrm{E}[\alpha] + \min\{\theta, 1-\theta, \phi, 1-\phi\} \int_{-2\bar{\varepsilon}}^{2\bar{\varepsilon}} d\widetilde{G}(e) \int_{\lambda - e}^{\overline{\beta}} b dB(b).$$

To get the third equation above, notice that $\hat{\beta}(\lambda) > 0 \Leftrightarrow \beta + \Delta\varepsilon > \lambda$, $\hat{\beta}(\lambda) < 0 \Leftrightarrow \beta + \Delta\varepsilon < -\lambda$, and $\hat{\beta}(\lambda) = 0 \Leftrightarrow |\beta + \Delta\varepsilon| \leq \lambda$, 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. To get the last equation, we have again utilized the symmetry of $\widetilde{G}$ and B. We the have the following proposition for a monopoly firm:

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

Proof.

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

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

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

where, to get the second equality above, we have changed the variable $e = \lambda - z$ for the second integral from $2\lambda - 2\bar{\varepsilon}$ to λ, and $e = \lambda + z$ for the third integral from λ to $2\bar{\varepsilon}$; moreover, we have utilized $B'(z) = B'(-z)$. To get the last inequality, notice that given the assumption $\widetilde{G}$ being unimodal 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 in this setup prefers the OLS regression to a Lasso. The intuition is as follows: Given that the monopoly firm chooses the tuning parameter in anticipation of the subsequent data realizations, the optimal model choice is the OLS estimator. This is because the OLS estimator is the most unbiased choice and thus enables the firm to target the more profitable segment correctly in expectation. The qualitative implication is that a monopolist optimally prefers a more flexible/complex 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 this would come at the expense of increased variance. This result serves as benchmark and motivates our analysis below of the competitive incentives for algorithmic targeting.

We have utilized the unimodality of $\widetilde{G}$ in the proof. Notice that we do not require B to be a unimodal distribution. If $\widetilde{G}$ were not unimodal either, it is possible that a monopoly might want set a positive λ that "estimates" β more accurately by leveraging the non-unimodality of both B and $\widetilde{G}$. (One can construct examples by considering B and $\widetilde{G}$ close to two-point distributions.) But then this behavior would be driven by the nature of the distributional characteristics. Our assumption of unimodal $\widetilde{G}$ serves to maintain a clean benchmark that highlights the comparison between monopoly and competition.

5 Competitive Targeting

Now we analyze the main model with competition between two firms and solve for the equilibrium by backward induction.

5.1 Targeting Decision

Given firm j's choice of the tuning parameter as λ_j and its private dataset, the firm's analyst's estimates, $\hat{\alpha}(\lambda_j)$ and $\hat{\beta}(\lambda_j)$ are given by equation (3). Suppose firm j decides to target k_j consumers with $x = 1$ and $\theta - k_j$ consumers with $x = 0$ for $j = 1, 2$. As before, we have $\max\{0, \theta + \phi - 1\} \leq k_j \leq \min\{\theta, \phi\}$.

Firm j does not observe the rival's choice of the tuning parameter nor its 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 equilibrium choice of k^*_{-j} depends on the realization of its private dataset and thus is a random variable, which is denoted as $\widetilde{k}^*_{-j}$. Let's calculate firm j's estimated profit:

$$\begin{aligned}\Pi_j(k_j,\widetilde{k}^*_{-j}) =& k_j\left(\frac{\widetilde{k}^*_{-j}}{\phi}\cdot\frac{1}{2}+1-\frac{\widetilde{k}^*_{-j}}{\phi}\right)\left(\hat{\alpha}_j(\lambda_j)+\hat{\beta}_j(\lambda_j)\right)\\ &+(\theta-k_j)\left(\frac{\theta-\widetilde{k}^*_{-j}}{1-\phi}\cdot\frac{1}{2}+1-\frac{\theta-\widetilde{k}^*_{-j}}{1-\phi}\right)\hat{\alpha}_j(\lambda_j)\\ =&\theta\left(1-\frac{\theta-\widetilde{k}^*_{-j}}{2(1-\phi)}\right)\hat{\alpha}_j(\lambda_j)\\ &+k_j\left(\frac{\phi\theta-\widetilde{k}^*_{-j}}{2\phi(1-\phi)}\hat{\alpha}_j(\lambda_j)+\left(1-\frac{\widetilde{k}^*_{-j}}{2\phi}\right)\hat{\beta}_j(\lambda_j)\right).\end{aligned} \tag{4}$$

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, $\mathrm{E}\left[\Pi_j(k_j,\widetilde{k}^*_{-j})\right]=\Pi_j(k_j,\mathrm{E}[\widetilde{k}^*_{-j}])$, where we have utilized the observation that $\Pi_j(k_j,\widetilde{k}^*_{-j})$ is linear in $\widetilde{k}^*_{-j}$.[4] Furthermore, notice that $\Pi_j(k_j,\mathrm{E}[\widetilde{k}^*_{-j}])$ is linear in k_j with

$$\frac{\partial\Pi_j(k_j,\mathrm{E}[\widetilde{k}^*_{-j}])}{\partial k_j}=\underbrace{\frac{\phi\theta-\mathrm{E}[\widetilde{k}^*_{-j}]}{2\phi(1-\phi)}\hat{\alpha}_j(\lambda_j)}_{\text{to avoid competition}}+\underbrace{\left(1-\frac{\mathrm{E}[\widetilde{k}^*_{-j}]}{2\phi}\right)\hat{\beta}_j(\lambda_j)}_{\text{to target the more profitable segment}}\equiv\eta_j(\lambda_j). \tag{5}$$

Consider the expression for $\partial_{k_j}\Pi_j(k_j,\mathrm{E}[\widetilde{k}^*_{-j}])$ in equation (5): The second term

[4] Notice that as α_1, β_1, α_2 and β_2 are independently distributed, firm j's private dataset provides no information on α_{-j} and β_{-j}. Therefore, $\mathrm{E}[\widetilde{k}^*_{-j}|\text{firm } j\text{'s dataset}]=\mathrm{E}[\widetilde{k}^*_{-j}]$.

plays a similar role as 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 would 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 would target proportionally more consumers with $x = 1$.

$\Pi_j(k_j, \mathrm{E}[\widetilde{k}^*_{-j}])$ being linear in k_j immediately implies that the firm's optimal targeting decision takes corner solutions. Specifically, if $\eta_j(\lambda_j) > 0$, firm j should set $k^*_j = \min\{\theta, \phi\}$ to target as many consumers with $x = 1$ as possible; if $\eta_j(\lambda_j) < 0$, the firm 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 as long as $\mathrm{E}[\widetilde{k}^*_{-j}] \neq \phi\theta$, $\eta_j(\lambda_j) = 0$ is a knife-edge case that happens with zero probability; consequently, the tie-breaking rule for which consumer to target at $\eta_j(\lambda_j) = 0$ has no consequence. On the other hand, if $\mathrm{E}[\widetilde{k}^*_{-j}] = \phi\theta$, we have $\eta_j(\lambda_j) = 0 \Leftrightarrow \hat{\beta}_j(\lambda_j) = 0$, at which, we have shown for the monopoly case above, the tie-breaking rule has no consequence either.

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:

$$
\begin{aligned}
p_{-j} \equiv & \Pr\left(\widetilde{k}^*_{-j} = \min\{\theta, \phi\}\right) \\
= & \Pr\left(\frac{\phi\theta - \mathrm{E}[\widetilde{k}^*_j]}{2\phi(1-\phi)}\hat{\alpha}_{-j}(\lambda^*_{-j}) + \left(1 - \frac{\mathrm{E}[\widetilde{k}^*_j]}{2\phi}\right)\hat{\beta}_{-j}(\lambda^*_{-j}) > 0\right) \\
= & \Pr\left(\frac{\phi\theta - (\max\{0, \theta+\phi-1\} + p_j \min\{\theta, 1-\theta, \phi, 1-\phi\})}{2\phi(1-\phi)}\hat{\alpha}_{-j}(\lambda^*_{-j})\right. \\
& \left. + \left(1 - \frac{\max\{0, \theta+\phi-1\} + p_j \min\{\theta, 1-\theta, \phi, 1-\phi\}}{2\phi}\right)\hat{\beta}_{-j}(\lambda^*_{-j}) > 0\right), \quad (6)
\end{aligned}
$$

where, to get the last equality in (6), we have utilized that

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

which is firm j's expectation of firm $-j$'s expectation of firm j's equilibrium choice of k_j^*, and thus $\mathrm{E}[\widetilde{k}_j^*]$ depends on λ_j^* (via p_j) instead of λ_j. By combining equation (6) for $j = 1, 2$, we should be able to solve p_1 and p_2, which depend on λ_1^* and λ_2^* (but not on λ_1 or λ_2).

Next, we determine λ_j^* 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 (4) 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.

$$\begin{aligned}\Pi_j(\lambda_j) \equiv &\mathrm{E}\left[\theta\left(1 - \frac{\theta - \widetilde{k}_{-j}^*}{2(1-\phi)}\right)\alpha_j + k_j^*\left(\frac{\phi\theta - \widetilde{k}_{-j}^*}{2\phi(1-\phi)}\alpha_j + \left(1 - \frac{\widetilde{k}_{-j}^*}{2\phi}\right)\beta_j\right)\right] \\ =&\theta\left(1 - \frac{\theta - \mathrm{E}[\widetilde{k}_{-j}^*]}{2(1-\phi)}\right)\mathrm{E}[\alpha_j] \\ &+ \min\{\theta, \phi\} \Pr\left(\eta_j(\lambda_j) > 0\right) \\ &\times \mathrm{E}\left[\frac{\phi\theta - \mathrm{E}[\widetilde{k}_{-j}^*]}{2\phi(1-\phi)}\alpha_j + \left(1 - \frac{\mathrm{E}[\widetilde{k}_{-j}^*]}{2\phi}\right)\beta_j \middle| \eta_j(\lambda_j) > 0\right] \\ &+ \max\{0, \theta + \phi - 1\} \Pr\left(\eta_j(\lambda_j) < 0\right) \\ &\times \mathrm{E}\left[\frac{\phi\theta - \mathrm{E}[\widetilde{k}_{-j}^*]}{2\phi(1-\phi)}\alpha_j + \left(1 - \frac{\mathrm{E}[\widetilde{k}_{-j}^*]}{2\phi}\right)\beta_j \middle| \eta_j(\lambda_j) < 0\right].\end{aligned} \tag{7}$$

In the calculation, we have utilized the independence between α_j, β_j and $\widetilde{k}_{-j}^*$. $\Pi_j(\lambda_j)$ depends on λ_j via $\eta_j(\lambda_j)$ and depends on λ_{-j}^* via $\widetilde{k}_{-j}^*$. That is, at the model selection stage, firm j has an expectation of firm $-j$'s choice of the tuning parameter, λ_{-j}^*, which will influence firm $-j$'s targeting decision and thus in turn influences firm j's expected profit. In expectation, each firm's choice should be consistent with the other firm's expectation:

$$\lambda_j^* = \arg\max_{\lambda_j} \Pi_j(\lambda_j), \text{ for } j = 1, 2. \tag{8}$$

To summarize, the Bayesian Nash equilibrium will be pinned down by the two sets of equations (6) and (8), where we have four equations to determine four variables: p_1, p_2, λ_1^* and λ_2^*. The main result of this paper is presented next.

5.3 Main Result

Proposition 2. *If a pure-strategy Bayesian Nash equilibrium exists, $\phi \neq 1/2$, $\theta \neq 1/2$, and $\bar{\varepsilon}$ is sufficiently high, then, we must have $\lambda_j^* > 0$ for at least one of $j = 1, 2$.*

Proposition 2 does not provide an explicit condition on when a pure-strategy equilibrium exists, which would require additional assumptions on the distribution functions, A, B and G to ensure firm j's profit function, $\Pi_j(\lambda_j)$ is quasi-concave for $j = 1, 2$. Nevertheless, notice that if pure-strategy equilibria do not exist, Nash's celebrated theorem immediately implies that there must exist a mixed-strategy equilibrium, where trivially, we must have $\Pr(\lambda_j^* > 0) > 0$ for at least one of $j = 1, 2$ (otherwise, we have $\lambda_j^* = 0$ for $j = 1, 2$, which is not a mixed-strategy equilibrium). Therefore, even if a pure-strategy equilibrium does not exist, we will end up with a result that is qualitatively similar in spirit with Proposition 2. Let's prove Proposition 2 next. Without loss of generality it is assumed that $\phi \in (0, 1/2)$. The other case with $\phi \in (1/2, 1)$ can be obtained by symmetry.

Proof. Let's first argue that given any λ_1^* and λ_2^*, there must exist a solution of (p_1, p_2) to equation (6) for $j = 1, 2$. In fact, the right-hand side of equation (6) for $j = 1, 2$ is a continuous map on a convex compact set $[0, 1]^2$ on to itself, and by Brouwer fixed-point theorem, a fixed point must exist. Next, we calculate $\Pi_j(\lambda_j)$ in equation (7). There are three cases to consider.

(i) $\mathrm{E}[\widetilde{k}_{-j}^*] < \phi\theta$, given which, there are two observations. First, Assumption 2 implies that $\hat{\alpha}_j(\lambda_j) > 0$. This further implies that if $\hat{\beta}_j(\lambda_j) \geq 0$, we must have $\eta_j(\lambda_j) > 0$ by the definition of $\eta_j(\lambda_j)$ in equation (5). Second, $\hat{\beta}_j(\lambda_j) < 0$ implies that $\hat{\beta}_j(\lambda_j) = \beta_j + \Delta\varepsilon_j + \lambda_j$ and $\hat{\alpha}_j(\lambda_j) = \alpha_j + \varepsilon_j^0 - \lambda_j/2$ by equations (2) and (3), based on which, we have

$$\eta_j(\lambda_j) < 0 \Leftrightarrow \alpha_j < -C\lambda_j + F(\beta_j, \varepsilon_j^0, \varepsilon_j^1), \text{ where}$$
$$C \equiv \frac{2\phi(1-\phi)}{\left|\phi\theta - \mathrm{E}[\widetilde{k}^*_{-j}]\right|}\left(1 - \frac{\mathrm{E}[\widetilde{k}^*_{-j}]}{2\phi} - \frac{\phi\theta - \mathrm{E}[\widetilde{k}^*_{-j}]}{4\phi(1-\phi)}\right),$$
$$F(\beta_j, \varepsilon_j^0, \varepsilon_j^1) \equiv -\frac{(1-\phi)(2\phi - \mathrm{E}[\widetilde{k}^*_{-j}])}{\phi\theta - \mathrm{E}[\widetilde{k}^*_{-j}]}\left(\beta_j + \varepsilon_j^1 - \varepsilon_j^0\right) - \varepsilon_j^0.$$

C is well defined given $\mathrm{E}[\widetilde{k}^*_{-j}] \neq \phi\theta$. It is easy to show that

$$C > 0 \Leftrightarrow 1 - \frac{\mathrm{E}[\widetilde{k}^*_{-j}]}{2\phi} - \frac{\phi\theta - \mathrm{E}[\widetilde{k}^*_{-j}]}{4\phi(1-\phi)} > 0 \Leftrightarrow (1-2\phi) + (1-\theta) + \mathrm{E}[\widetilde{k}^*_{-j}] > 0,$$

which always holds regardless of the comparison between $\mathrm{E}[\widetilde{k}^*_{-j}]$ and $\phi\theta$.

Putting the two observations above together, we have

$$\begin{aligned}
&\Pr\left(\eta_j(\lambda_j) < 0\right)\mathrm{E}\left[\alpha_j|\eta_j(\lambda_j) < 0\right] \\
&= \Pr\left(\eta_j(\lambda_j) < 0 \text{ and } \hat{\beta}_j(\lambda_j) < 0\right)\mathrm{E}\left[\alpha_j\middle|\eta_j(\lambda_j) < 0 \text{ and } \hat{\beta}_j(\lambda_j) < 0\right] \\
&\quad + \Pr\left(\eta_j(\lambda_j) < 0 \text{ and } \hat{\beta}_j(\lambda_j) \geq 0\right)\mathrm{E}\left[\alpha_j\middle|\eta_j(\lambda_j) < 0 \text{ and } \hat{\beta}_j(\lambda_j) \geq 0\right] \\
&= \Pr\left(\alpha_j < -C\lambda_j + F(\beta_j, \varepsilon_j^0, \varepsilon_j^1)\right)\mathrm{E}\left[\alpha_j\middle|\alpha_j < -C\lambda_j + F(\beta_j, \varepsilon_j^0, \varepsilon_j^1)\right] \\
&= \int_{-\overline{\varepsilon}}^{\overline{\varepsilon}}\int_{-\overline{\varepsilon}}^{\overline{\varepsilon}}\int_{-\overline{\beta}}^{\overline{\beta}}\int_{\underline{\alpha}}^{\min\left\{\max\left\{-C\lambda_j + F(\beta_j, \varepsilon_j^0, \varepsilon_j^1), \underline{\alpha}\right\}, \overline{\alpha}\right\}} \alpha_j dA(\alpha_j)dB(\beta_j)dG(\varepsilon_j^0)dG(\varepsilon_j^1),
\end{aligned}$$

where to get the first equality above, we have used the definition of conditional probabilities and the law of total probability. Moreover, we have argued $\Pr\left(\eta_j(\lambda_j) = 0\right) = 0$ above, which implies that,

$$\Pr\left(\eta_j(\lambda_j) > 0\right)\mathrm{E}[\alpha_j|\eta_j(\lambda_j) > 0] = \mathrm{E}[\alpha_j] - \Pr\left(\eta_j(\lambda_j) < 0\right)\mathrm{E}[\alpha_j|\eta_j(\lambda_j) < 0].$$

Similarly, we can write down the expressions for $\Pr(\eta_j(\lambda_j) > 0)\mathrm{E}[\beta_j|\eta_j(\lambda_j) > 0]$ and $\Pr(\eta_j(\lambda_j) < 0)\mathrm{E}[\beta_j|\eta_j(\lambda_j) < 0]$. By substituting these back to $\Pi_j(\lambda_j)$ in equation (7), we find:

$$\begin{aligned}
\Pi_j(\lambda_j) =&\theta\left(1-\frac{\theta-\mathrm{E}[\widetilde{k}^*_{-j}]}{2(1-\phi)}\right)\mathrm{E}[\alpha_j]\\
&+\min\{\theta,\phi\}\left(\frac{\phi\theta-\mathrm{E}[\widetilde{k}^*_{-j}]}{2\phi(1-\phi)}\mathrm{E}[\alpha_j]+\left(1-\frac{\mathrm{E}[\widetilde{k}^*_{-j}]}{2\phi}\right)\mathrm{E}[\beta_j]\right)\\
&-\min\{\theta,1-\theta,\phi,1-\phi\}\int_{-\overline{\varepsilon}}^{\overline{\varepsilon}}\int_{-\overline{\varepsilon}}^{\overline{\varepsilon}}\int_{-\overline{\beta}}^{\overline{\beta}}\int_{\underline{\alpha}}^{\min\left\{\max\left\{-C\lambda_j+F(\beta_j,\varepsilon_j^0,\varepsilon_j^1),\underline{\alpha}\right\},\overline{\alpha}\right\}}\\
&\quad\left(\frac{\phi\theta-\mathrm{E}[\widetilde{k}^*_{-j}]}{2\phi(1-\phi)}\alpha_j+\left(1-\frac{\mathrm{E}[\widetilde{k}^*_{-j}]}{2\phi}\right)\beta_j\right)dA(\alpha_j)dB(\beta_j)dG(\varepsilon_j^0)dG(\varepsilon_j^1).
\end{aligned}$$

Let's compute the derivative of $\Pi_j(\lambda_j)$ at $\lambda_j=0$:

$$\begin{aligned}
\Pi_j'(0) =&\min\{\theta,1-\theta,\phi,1-\phi\}C\\
&\times\iiint_{\underline{\alpha}\le F(\beta_j,\varepsilon_j^0,\varepsilon_j^1)\le\overline{\alpha}}\left(\frac{\phi\theta-\mathrm{E}[\widetilde{k}^*_{-j}]}{2\phi(1-\phi)}F(\beta_j,\varepsilon_j^0,\varepsilon_j^1)+\left(1-\frac{\mathrm{E}[\widetilde{k}^*_{-j}]}{2\phi}\right)\beta_j\right)\\
&\times A'(F(\beta_j,\varepsilon_j^0,\varepsilon_j^1))dB(\beta_j)dG(\varepsilon_j^0)dG(\varepsilon_j^1)\\
\ge&\min\{\theta,1-\theta,\phi,1-\phi\}C\left(\frac{\phi\theta-\mathrm{E}[\widetilde{k}^*_{-j}]}{2\phi(1-\phi)}\underline{\alpha}-\left(1-\frac{\mathrm{E}[\widetilde{k}^*_{-j}]}{2\phi}\right)\overline{\beta}\right)\\
&\times\iiint_{\underline{\alpha}\le F(\beta_j,\varepsilon_j^0,\varepsilon_j^1)\le\overline{\alpha}}A'(F(\beta_j,\varepsilon_j^0,\varepsilon_j^1))dB(\beta_j)dG(\varepsilon_j^0)dG(\varepsilon_j^1).
\end{aligned}$$

When $\overline{\varepsilon}$ is sufficiently large, Assumption 2 implies that $\underline{\alpha}$ is sufficiently large so that

$$\frac{\phi\theta-\mathrm{E}[\widetilde{k}^*_{-j}]}{2\phi(1-\phi)}\underline{\alpha}-\left(1-\frac{\mathrm{E}[\widetilde{k}^*_{-j}]}{2\phi}\right)\overline{\beta}>0;$$

moreover, $F(\beta_j,\varepsilon_j^0,\varepsilon_j^1)$ by definition is symmetrically distributed around zero and when $\overline{\varepsilon}$ is sufficiently large, $\Pr(\underline{\alpha}\le F(\beta_j,\varepsilon_j^0,\varepsilon_j^1)\le\overline{\alpha})>0$. Therefore, we have $\Pi_j'(0)>0$, which implies that $\lambda_j^*>0$.

(ii) $\mathrm{E}[\widetilde{k}^*_{-j}]>\phi\theta$, given which, there are similarly two observations. First, $\hat{\beta}_j(\lambda_j)\le 0$ implies $\eta_j(\lambda_j)<0$. Second, $\hat{\beta}_j(\lambda_j)>0$ implies that $\hat{\beta}_j(\lambda_j)=\beta_j+\Delta\varepsilon_j-\lambda_j$ and $\hat{\alpha}_j(\lambda_j)=\alpha_j+\varepsilon_j^0+\lambda_j/2$ by equations (2) and (3), based on which, we have

$$\eta_j(\lambda_j) > 0 \Leftrightarrow \alpha_j < -C\lambda_j + F(\beta_j, \varepsilon_j^0, \varepsilon_j^1),$$

the same as that in case (i). Putting the two observations together, we have that

$$\begin{aligned}
&\Pr\left(\eta_j(\lambda_j) > 0\right) \mathrm{E}\left[\alpha_j | \eta_j(\lambda_j) > 0\right] \\
&= \Pr\left(-C\lambda_j + F(\beta_j, \varepsilon_j^0, \varepsilon_j^1)\right) \mathrm{E}\left[\alpha_j \middle| -C\lambda_j + F(\beta_j, \varepsilon_j^0, \varepsilon_j^1)\right] \\
&= \int_{-\overline{\varepsilon}}^{\overline{\varepsilon}} \int_{-\overline{\varepsilon}}^{\overline{\varepsilon}} \int_{-\overline{\beta}}^{\overline{\beta}} \int_{\underline{\alpha}}^{\min\left\{\max\left\{-C\lambda_j + F(\beta_j, \varepsilon_j^0, \varepsilon_j^1), \underline{\alpha}\right\}, \overline{\alpha}\right\}} \alpha_j dA(\alpha_j) dB(\beta_j) dG(\varepsilon_j^0) dG(\varepsilon_j^1), \\
&\Pr\left(\eta_j(\lambda_j) < 0\right) \mathrm{E}[\alpha_j | \eta_j(\lambda_j) < 0] = \mathrm{E}[\alpha_j] - \Pr\left(\eta_j(\lambda_j) > 0\right) \mathrm{E}[\alpha_j | \eta_j(\lambda_j) > 0].
\end{aligned}$$

Similarly, we can write down $\Pi_j(\lambda_j)$:

$$\begin{aligned}
\Pi_j(\lambda_j) =& \theta \left(1 - \frac{\theta - \mathrm{E}[\widetilde{k}_{-j}^*]}{2(1-\phi)}\right) \mathrm{E}[\alpha_j] \\
&+ \max\{0, \theta + \phi - 1\} \left(\frac{\phi\theta - \mathrm{E}[\widetilde{k}_{-j}^*]}{2\phi(1-\phi)} \mathrm{E}[\alpha_j] + \left(1 - \frac{\mathrm{E}[\widetilde{k}_{-j}^*]}{2\phi}\right) \mathrm{E}[\beta_j]\right) \\
&- \min\{\theta, 1-\theta, \phi, 1-\phi\} \int_{-\overline{\varepsilon}}^{\overline{\varepsilon}} \int_{-\overline{\varepsilon}}^{\overline{\varepsilon}} \int_{-\overline{\beta}}^{\overline{\beta}} \int_{\underline{\alpha}}^{\min\left\{\max\left\{-C\lambda_j + F(\beta_j, \varepsilon_j^0, \varepsilon_j^1), \underline{\alpha}\right\}, \overline{\alpha}\right\}} \\
&\quad \left(\frac{\mathrm{E}[\widetilde{k}_{-j}^*] - \phi\theta}{2\phi(1-\phi)} \alpha_j - \left(1 - \frac{\mathrm{E}[\widetilde{k}_{-j}^*]}{2\phi}\right) \beta_j\right) dA(\alpha_j) dB(\beta_j) dG(\varepsilon_j^0) dG(\varepsilon_j^1).
\end{aligned}$$

Similarly, we can compute:

$$\begin{aligned}
\Pi_j'(0) =& \min\{\theta, 1-\theta, \phi, 1-\phi\} C \\
&\times \iiint_{\underline{\alpha} \le F(\beta_j, \varepsilon_j^0, \varepsilon_j^1) \le \overline{\alpha}} \left(\frac{\mathrm{E}[\widetilde{k}_{-j}^*] - \phi\theta}{2\phi(1-\phi)} F(\beta_j, \varepsilon_j^0, \varepsilon_j^1) - \left(1 - \frac{\mathrm{E}[\widetilde{k}_{-j}^*]}{2\phi}\right) \beta_j\right) \\
&\times A'(F(\beta_j, \varepsilon_j^0, \varepsilon_j^1)) dB(\beta_j) dG(\varepsilon_j^0) dG(\varepsilon_j^1) \\
\ge& \min\{\theta, 1-\theta, \phi, 1-\phi\} C \left(\frac{\mathrm{E}[\widetilde{k}_{-j}^*] - \phi\theta}{2\phi(1-\phi)} \underline{\alpha} - \left(1 - \frac{\mathrm{E}[\widetilde{k}_{-j}^*]}{2\phi}\right) \overline{\beta}\right) \\
&\times \iiint_{\underline{\alpha} \le F(\beta_j, \varepsilon_j^0, \varepsilon_j^1) \le \overline{\alpha}} A'(F(\beta_j, \varepsilon_j^0, \varepsilon_j^1)) dB(\beta_j) dG(\varepsilon_j^0) dG(\varepsilon_j^1).
\end{aligned}$$

The same argument as in Case (ii) shows that when $\bar{\varepsilon}$ is sufficiently large, $\lambda_j^* > 0$.

(iii) $\mathrm{E}[\widetilde{k}_{-j}^*] = \phi\theta$. If $\lambda_j^* > 0$, we have proved the proposition; otherwise, suppose $\lambda_j^* = 0$. We have $p^j = \Pr(\hat{\beta}_j(\lambda_j^*) > 0) = \Pr(\beta_j + \Delta\varepsilon_j > 0) = 1/2$. Correspondingly,

$$\mathrm{E}[\widetilde{k}_j^*] = \frac{1}{2}\left(\min\{\theta, \phi\} + \max\{0, \theta + \phi - 1\}\right) \neq \theta\phi.$$

In fact, for $0 < \phi < 1/2$, $\mathrm{E}[\widetilde{k}_j^*] = \theta\phi$ if and only if $\theta = 0, 1/2, 1$, which we have excluded by assumption. Therefore, it must be that $\mathrm{E}[\widetilde{k}_j^*] < \theta\phi$ or $\mathrm{E}[\widetilde{k}_j^*] > \theta\phi$. In either case, we can repeat the proof above with j and $-j$ switched to conclude that $\lambda_{-j}^* > 0$. $\square$

In contrast to Proposition 1, Proposition 2 shows that competition drives at least one firm to choose positive penalization. In other words, competition favors a simpler algorithmic design that reduces variance but at the cost of introducing bias. We provide below the economic intuition for this result.

Because the two consumer segments are of different sizes (by the assumption that $\phi \neq 1/2$), the one which is smaller will be ex-ante more competitive because when both firms target this segment, there will be higher expected overlap of the targeted consumers. Compared with the OLS estimator which induces a firm to concentrate targeting in one consumer segment (the one with higher estimated profitability), the penalization in the Lasso regression tends to induce the firm to target consumers across the two segments more evenly. When $\theta = 1/2$, the OLS and the Lasso will generate the same targeting outcome, because it amounts to the same 50% targeting probability on every consumer regardless of whether the firm targets the two consumer segments evenly or targets all the consumers evenly. Therefore, as long as $\theta \neq 1/2$, the penalization in the Lasso regression that induces more uniform targeting across consumers will reduce a firm's concentration of targeting on one particular consumer segment, which in turn reduces the expected overlap between the two firms' targeted consumers and thus softens competition. This can also be seen from equation (5), where a higher λ_j penalizes $\hat{\beta}_j(\lambda_j)$ towards zero and consequently, the competition avoidance incentive as captured by $(\phi\theta - \mathrm{E}[\widetilde{k}_{-j}^*])/(2\phi(1-\phi))$ has a relatively bigger impact on $\eta_j(\lambda_j)$ which determines firm j's targeting decision.

In fact, the competition avoidance incentive for firm j is present whenever $\mathrm{E}[\widetilde{k}^*_{-j}] \neq \phi\theta$—that is, when the competitor does not target all consumers equally. This provides firm j the strategic incentive to introduce bias to reduce the overlap in the targeting. In fact, as shown in the proof of Proposition 2 above, as long as $\mathrm{E}[\widetilde{k}^*_{-j}] \neq \phi\theta$, firm j will choose $\lambda^*_j > 0$ in equilibrium to lessen competition.

It is worthwhile to reiterate that in our modeling approach, different choices of λ_j by firm j determines different algorithmic designs, which amounts to different ways of representing consumer information for decision making on targeting. Proposition 2 implies that competition favors a positive penalization that leads to more precise but less accurate information about consumer profitability. In fact, $\hat{\beta}_j(\lambda_j)$ will be non-zero only if the profit difference between two segments of consumers is big enough to compensate the profit loss from more intense competition resulting from more concentrated targeting. In other words, compared with the OLS estimator, the estimator of $\hat{\beta}_j(\lambda_j)$ will be not very accurate when $|\beta_j|$ is close to zero but more precise.

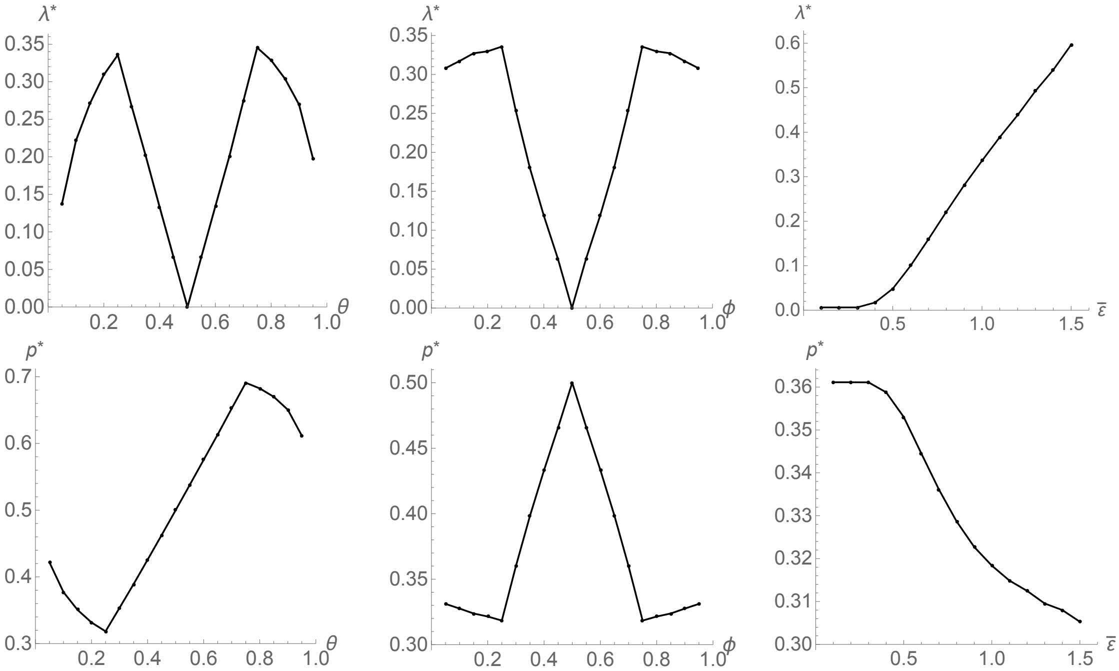
Lastly, we also require $\bar{\varepsilon}$ to be sufficiently high. With enough noise in the data, the risk of over-fitting becomes consequential. Moreover, a higher $\bar{\varepsilon}$ also implies a higher $\underline{\alpha}$ by Assumption 2, which translates into a higher incentive to avoid competition by equation (5). Both considerations make a positive penalization in the Lasso regression and the equilibrium choice of algorithmic bias more desirable.

5.4 Symmetric Equilibrium

Given our symmetric setup, it is natural to consider the symmetric equilibrium with $\lambda^*_1 = \lambda^*_2 = \lambda^*$. The corollary below is obvious from Proposition 2.

Corollary 1. *If a symmetric pure-strategy Bayesian Nash equilibrium exists, $\phi \neq 1/2$, $\theta \neq 1/2$ and $\bar{\varepsilon}$ is sufficiently high, then, we must have $\lambda^* > 0$.*

Figure 3 provides some some numerical examples of the equilibrium under uniform distributions and also examines the comparative statics. For all the parameter settings in Figure 3 a pure-strategy symmetric equilibrium exists.



(a) $\phi = 0.25$ and $\bar{\varepsilon} = 1$ (b) $\theta = 0.25$ and $\bar{\varepsilon} = 1$ (c) $\phi = \theta = 0.25$

Figure 3: Equilibrium λ^* and p^* given $A \sim$ Unif$[2, 4]$, $B \sim$ Unif$[-1, 1]$, $G \sim$ Unif$[-\bar{\varepsilon}, \bar{\varepsilon}]$.

There are several observations to make. First, notice that when $\theta < 1/2$ (as in the left half of panel (a) as well as the entire region of panels (b) and (c)), p^* decreases with λ^*. This is very intuitive—as the penalization gets higher, firms tend to target more evenly across the two consumer segments, which means reducing targeting probability in the more competitive segment—segment $x = 1$ in this case due to $\theta < 1/2$. On the other hand, when $\theta > 1/2$ (as in the right half of panel (a)), segment $x = 0$ is more competitive, so as λ^* increases, the firms allocate more targeting probability to the less competitive segment of $x = 1$, which means raising p^*. Second, we find that indeed for the two knife-edge cases of $\theta = 1/2$ and $\phi = 1/2$, $\lambda^* = 0$, as shown by panels (a) and (b); correspondingly, $p^* = 1/2$ in these cases. Thirdly, we find that the firms choose the maximum penalization when $\theta = \phi$ or $\theta = 1 - \phi$, as shown by panels (b) and (c). In fact, by calculation, one can show that these are the cases when the firms achieve the maximum reduction in consumer overlap by switching from targeting two segments evenly to targeting every consumer evenly. Therefore, these are the cases when the firms have the highest incentive to set a high penalization. Lastly, consistent with the standard statistical learning theory, as $\bar{\varepsilon}$ increases, the data gets noisier, and consequently, the firms choose a higher penalization to avoid the over-fitting problem, as shown by panel (c).

6 Summary and Discussion

In this paper, we examine how competitive firms employ algorithms to estimate demand and based on the estimates, make strategic consumer targeting decisions to maximize expected profit. Algorithmic design essentially implies different model selection strategies, which involve different bias and variance trade-offs under the general framework of supervised learning. Essentially we can view model selection as the choice of the consumer information structure that the firm strategically uses for targeting decisions. This bias-variance trade-off also implies the extent of model flexibility that the firm would like to optimally use for targeting. From this perspective, our paper studies firms' competitive model selection for algorithmic targeting and explores how competition moderates individual firms' bias-variance trade-off choices through the degree of complexity of the algorithm that is adopted. The central finding is that targeting under competition favors simpler models that reduce variance but which introduce bias. There is therefore the suggestion that more flexible algorithms like deep learning are more likely to be valuable for firms with monopoly power.

We focus on a specific decision of the firms—targeting. Thanks to large advertising platforms such as Facebook or Google, there is an ongoing trend of advertising targeting decisions being automated by algorithms for real-time advertising deployment based on rich customer behavior data on browsing, purchase, sharing, observed social connections, etc. Targeting is therefore a natural context to study algorithmic competition and our model and payoff function is designed to represent the classic competitive targeting problem. Within this context, our result that competition favors algorithmic bias holds for quite general distributional assumptions about the prior beliefs. In our current analysis, we have used an overall payoff function without explicit consideration for pricing or other decision variables. This helps us to highlight the targeting problem and the link between competitive targeting incentives and model selection. As next steps it would be interesting to explore a general class of oligopoly games with strategic firm decisions such as pricing, advertising or product design. The implications may depend on whether the firms' decisions are strategic substitutes or complements (Bulow et al. 1985).

We can also consider generalizing some aspects of the model in this paper. For example, we could analyze the impact of larger datasets with more observations. Larger datasets should reduce the variance of the estimated parameters and the extent of the over-fitting problem. This could likely lead competitive firms to have the incentive to choose lower penalization. In contrast, a model with greater number of predictor variables may lead firms to impose higher penalization. The results may also depend upon the extent of consumer information available to the firms and the presence of horizontal firm differentiation. Suppose firms can identify consumer types and suppose the profitability of the segments are negatively correlated across firms. Then this could reduce the incentive for penalization by competing firms.

Endogenous Data

We conclude by describing a setup which allows the targeting dataset to be generated from market competition. In the paper we have assumed that each firm is endowed with an exogenous dataset. To allow for the datasets to be endogenously generated from market interaction we should require the firms to compete in the targeting decisions at least twice, where the first-time competition generates the data, which is then utilized by the firms to devise their subsequent targeting strategies. Specifically, suppose that the game analyzed in the paper is modified through the following timeline. At time 0, the two firms simultaneously choose the tuning parameters. At time 1 when the first period begins, each firm decides on the consumers to target, who upon being targeted, decide whether to make a purchase. Each firm observes a noisy signal of the profit from each consumer who made a purchase. That is, we interpret $\pi_j(x)$ in the main model as firm j's average profit from an x-type consumer, and the firm's profit from an individual x-type consumer who made a purchase is $\pi_j(x)$ plus some idiosyncratic error (analogous to y_j^l in the main model). Based on the data, as before each firm delegates an analyst to estimate profit by running a Lasso regression. Based on the estimates, each firm devises the targeting strategy to maximize the estimated profit in the second period.

In this modified game, each firm makes targeting decisions in the first period based on its prior belief. Given that β_j is distributed symmetrically around zero, it is optimal for each firm to target randomly. Notice that if a consumer is targeted by both firms, she makes a random choice between the two. This implies that observation of a targeted consumer's purchase decision does not give the firm any extra information for estimating the consumer profitability. Consequently, each firm's first-period actions result in a dataset of "$\pi_j(x_i)$ plus some idiosyncratic errors", where i is the consumer index that spans across all consumers who made a purchase from firm j in the first period. Even though this dataset is generated from the first period market interaction, it is qualitatively similar to that in the main model and could be equivalently seen as being generated from a monopoly market, with the one caveat that the size of the data is less than θ, the size of the data for the monopoly market. Moreover, notice that the firms' choice in tuning parameters at time 0 has no impact on their profits in the first period, so when choosing λ_j, each firm j effectively only takes into account the impact on its profit in the second period, which essentially is the same decision problem as in the main model.

To summarize, this extended two-period model that allows for the datasets to be endogenous to the first period interaction is almost identical to our main model with exogenous datasets, except that i) for each dataset, the number and types of consumers observed can be different; ii) compared with the monopoly benchmark, each firm has a smaller dataset under competition. However, intuitively the first difference would not qualitatively alter the main result pertaining to the effect of competition on model selection; the second difference could potentially strengthen the main result, because a smaller dataset implies a higher level of noise, which would lead to a higher equilibrium choice of penalization in competition as compared with the monopoly benchmark.

References

Ajay Agrawal, Joshua Gans, and Avi Goldfarb. *Prediction machines: the simple economics of artificial intelligence*. Harvard Business Press, 2018.

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.

Trevor Hastie, Robert Tibshirani, Jerome H Friedman, and Jerome H Friedman. *The elements of statistical learning: data mining, inference, and prediction*, volume 2. Springer, 2009.

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.

Jonathan Levin and Paul Milgrom. Online advertising: Heterogeneity and conflation in market design. *American Economic Review*, 100(2):603–607, 2010.

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

Jeanine Miklós-Thal and Catherine Tucker. Collusion by algorithm: Does better demand prediction facilitate coordination between sellers? *Management Science*, 65(4):1552–1561, 2019.

Jason O'Connor and Nathan E Wilson. Reduced demand uncertainty and the sustainability of collusion: How ai could affect competition. *Information Economics and Policy*, 54:100882, 2021.

José Luis Montiel Olea, Pietro Ortoleva, Mallesh M Pai, and Andrea Prat. Competing models. *Quarterly Journal of Economics*, 137(4):2419–2457, 2022.

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.