Model Selection for Treatment Choice: Penalized Welfare Maximization

Eric Mbakop
Department of Economics
University of Calgary
eric.mbakop@ucalgary.ca

Max Tabord-Meehan
Department of Economics
University of Chicago
maxtm@uchicago.edu *

28th September 2020

Abstract

This paper studies a penalized statistical decision rule for the treatment assignment problem. Consider the setting of a utilitarian policy maker who must use sample data to allocate a binary treatment to members of a population, based on their observable characteristics. We model this problem as a statistical decision problem where the policy maker must choose a subset of the covariate space to assign to treatment, out of a class of potential subsets. We focus on settings in which the policy maker may want to select amongst a *collection* of constrained subset classes: examples include choosing the number of covariates over which to perform best-subset selection, and model selection when approximating a complicated class via a sieve. We adapt and extend results from statistical learning to develop the Penalized Welfare Maximization (PWM) rule. We establish an oracle inequality for the regret of the PWM rule which shows that it is able to perform model selection over the collection of available classes. We then use this oracle inequality to derive relevant bounds on maximum regret for PWM. An important consequence of our results is that we are able to formalize model-selection using a "hold-out" procedure, where the policy maker would first estimate various policies using half of the data, and then select the policy which performs the best when evaluated on the other half of the data.

KEYWORDS: Treatment Choice, Minimax-Regret, Statistical Learning
JEL classification codes: C01, C14, C44, C52

*We are grateful for advice and encouragement from Ivan Canay, Joel Horowitz, Chuck Manski and Alex Torgovitsky. We would also like to thank Toru Kitagawa, Azeem Shaikh, Alex Tetenov, Stefan Wager, the editor, anonymous referees, and seminar participants at Northwestern University, (continued on next page)

1 Introduction

This paper develops a new statistical decision rule for the treatment assignment problem. A major goal of treatment evaluation is to provide policy makers with guidance on how to assign individuals to treatment, given experimental or quasi-experimental data. Following the literature inspired by Manski (2004) (a partial list in econometrics includes Armstrong and Shen, 2015; Athey and Wager, 2017; Bhattacharya and Dupas, 2012; Chamberlain, 2011; Dehejia, 2005; Hirano and Porter, 2009; Kitagawa and Tetenov, 2018; Kock and Thyrsgaard, 2017; Rai, 2018; Schlag, 2007; Stoye, 2009, 2012; Tetenov, 2012; Viviano, 2019), we treat the treatment assignment problem as a statistical decision problem of maximizing population welfare. Like many of the above papers, we evaluate our decision rule by its maximum regret.

The rule we develop, the Penalized Welfare Maximization (PWM) rule, is designed to address situations in which the policy maker can choose amongst a *collection* of constrained classes of allocations. To be concrete, suppose we have two treatments, and we represent assignment into these treatments by partitioning the covariate space into two pieces. We can then think of constraints on assignment as constraints on the allowable subsets that we can consider for the partitions. For example, policy makers may face exogenous constraints on how they can use covariates for legal, ethical, or political reasons. Even in cases where policy makers have leeway in how they assign treatment, plausible modeling assumptions may imply certain restrictions on assignment. Kitagawa and Tetenov (2018) develop what they call the Empirical Welfare Maximization (or EWM) rule, whose primary feature is its ability to solve the treatment choice problem when certain exogenous constraints are placed on assignment. Kitagawa and Tetenov (2018) focus on deriving bounds on maximum regret of the EWM rule for a *fixed* class of subsets of finite VC dimension (see Györfi et al., 1996, for a definition). In this paper, however, we consider settings where the class of allowable subsets is "large". We approach the problem by approximating our class of allowable allocations by a sequence of subclasses of finite VC dimension. We establish an oracle inequality for the regret of the PWM rule which shows that it behaves as if we knew the "correct" class to use in the sequence. We then use this result to derive bounds on the maximum regret of the PWM rule in two empirically relevant settings.

The main setting that we consider is one where the class of feasible allocations has infinite VC dimension. In particular, we argue that economic modeling assumptions may sometimes put restrictions on the unconstrained optimum that naturally generate classes of infinite VC dimension. For example, plausible assumptions may only impose shape restrictions on the optimal allocation. To solve the optimal welfare assignment problem in this setting, we approximate these large classes

[0](continued from previous page), NASMES 2017, and the Bristol Econometrics Study Group for helpful comments, as well as Nitish Keskar for help in implementing EWM. This research was supported in part through the computational resources and staff contributions provided for the Social Sciences Computing Cluster (SSCC), and the Quest high performance computing facility at Northwestern University. All mistakes are our own.

of feasible allocations by sequences of classes of finite VC dimension. The strength of the PWM rule in this setting will then be to provide a data-driven method by which to select an "appropriate" approximating class. In doing so we will derive bounds on the maximum regret of the PWM rule for a large set of classes of infinite VC dimension.

We also consider the setting where the class of feasible allocations may have large VC dimension relative to the sample size. This could arise, for example, if the planner has many covariates on which to base assignment. As is shown in Kitagawa and Tetenov (2018), when the constraints placed on assignment are too flexible relative to the sample size available, the EWM rule may suffer from overfitting, which can result in inflated values of regret. By the same mechanism that allows PWM to select an appropriate approximating class in our first application, we can use PWM in order to select amongst simpler subclasses in this setting as well, in a way that improves the performance of the allocation rule in finite samples.

The PWM rule is heavily inspired by the literature on model selection in classification: see for example the seminal work of Vapnik and Chervonenkis (1974), as well as Györfi et al. (1996), Koltchinskii (2001), Bartlett et al. (2002), Boucheron et al. (2005), Scott and Nowak (2006), Bartlett (2008), Koltchinskii (2008) among many others. The theoretical contribution of our paper is to modify and extend some of these tools to the setting of treatment choice. In deciding which tools to extend, we have attempted to strike a balance between ease of use for practitioners, theoretical appeal, and performance in simulations. An important consequence of our results is that we are able to formalize model-selection using a "hold-out" procedure, where the policy maker would first estimate various policies using half of the data, and then select the policy which performs the best when evaluated on the other half of the data. The connection between classification and treatment choice has been explored in various fields, including machine learning, under the label of *policy learning* (see Beygelzimer and Langford, 2009; Kallus, 2016; Swaminathan and Joachims, 2015; Zadrozny, 2003, among others), and in epidemiology under the label of *individualized treatment rules* (examples include Qian and Murphy, 2011; Zhao et al., 2012). Kitagawa and Tetenov (2018) and Athey and Wager (2017) provide a discussion on the link between these various literatures.

The remainder of the paper is organized as follows. In Section 2, we set up the notation and formally define the problem that the policy maker (i.e. social planner) is attempting to solve. In Section 3, we introduce the PWM rule, present general results about its maximum regret, and explain how our results allow us to study the properties of the "hold-out" model-selection procedure. In Section 4 we derive bounds on maximum regret of the PWM rule when the planner is constrained to what we call *monotone* allocations, and then apply PWM in an application using data from the Job Training Partnership Act (JTPA) study.

2 Setup and Notation

Let Y_i denote the observed outcome of a unit i, and let D_i be a binary variable which denotes the treatment received by unit i. Let $Y_i(1)$ denote the potential outcome of unit i under treatment 1 (which we will refer to as "the treatment"), and let $Y_i(0)$ denote the potential outcome of unit i under treatment 0 (which we will refer to as "the control"). The observed outcome for each unit is related to their potential outcomes through the expression:

$$Y_i = Y_i(1)D_i + Y_i(0)(1 - D_i) \ . \tag{1}$$

Let $X_i \in \mathcal{X} \subset \mathbb{R}^{d_x}$ denote a vector of observed covariates for unit i. Let Q denote the distribution of $(Y_i(0), Y_i(1), D_i, X_i)$, then we assume that the planner observes a size n random sample

$$(Y_i, D_i, X_i)_{i=1}^n \sim P^n \ ,$$

where P is jointly determined by Q, and the expression in (1). Throughout the paper we will assume unconfoundedness, i.e.

Assumption 2.1. *(Unconfoundedness) The distribution Q satisfies:*

$$\Big((Y(1), Y(0)) \perp D\Big) \Big| X \ .$$

This assumption asserts that, once we condition on the observable covariates, the treatment is exogenous. This assumption will hold in a randomized controlled trial (RCT), which is our primary application of interest, since the treatment is exogenous by construction.

The planner's goal is to optimally assign the treatment to the population. The objective function we consider is utilitarian welfare, which is defined by the average of the individual outcomes in the population:

$$E_Q[Y(1)\mathbf{1}\{X \in G\} + Y(0)\mathbf{1}\{X \notin G\}] \ ,$$

where $G \subset \mathcal{X}$ represents the set of covariate values of the individuals assigned to treatment. The planner is tasked with choosing a *treatment allocation* $G \subset \mathcal{X}$ using the empirical data. Using Assumption 2.1, we can rewrite the welfare criterion as:

$$E_Q[Y(0)] + E_P\Big[\Big(\frac{YD}{e(X)} - \frac{Y(1-D)}{1-e(X)}\Big)\mathbf{1}\{X \in G\}\Big] \ ,$$

where $e(X) = E_P[D|X]$ is the propensity score. Since the first term of this expression does not depend on G, we define the planner's objective function given a choice of treatment allocation G as:

$$W(G) := E_P\Big[\Big(\frac{YD}{e(X)} - \frac{Y(1-D)}{1-e(X)}\Big)\mathbf{1}\{X \in G\}\Big] \ .$$

Let $\mathcal{G}$ be the class of all feasible treatment allocations. Here, we consider the possibility that the planner may be restricted in what type of allocations she can (or wants to) consider. These restrictions may arise from legal, ethical, or political considerations, or could arise as natural constraints from an economic model. Consider the following three examples of $\mathcal{G}$:

Example 2.1. $\mathcal{G}$ could be the set of all measurable subsets of $\mathcal{X}$. This is the largest possible class of admissible allocations. It is straightforward to show that the optimal allocation in this case is as follows: define

$$\tau(x) := E_Q[Y(1) - Y(0)|X = x] ,$$

then the optimal allocation is given by

$$G^*_{FB} := \{x \in \mathcal{X} : \tau(x) \geq 0\} ,$$

which assigns an individual with covariate x to treatment or control depending on whether the conditional average treatment effect at x is non-negative. ■

Example 2.2. Suppose $\mathcal{X} \subset \mathbb{R}$, and consider the class of *threshold allocations*:

$$\mathcal{G} = \{G : G = (-\infty, x] \cap \mathcal{X} \text{ or } G = [x, \infty) \cap \mathcal{X}, \text{ for } x \in \mathcal{X}\} .$$

Such a class $\mathcal{G}$ would be reasonable, for example, when assigning scholarships to students: suppose the only covariate available to the planner is a student's GPA, then it may be school policy that only threshold-type rules are to be considered. ■

Example 2.3. Let $\mathcal{X} = \mathcal{X}_1 \times \mathcal{X}_2 \subset \mathbb{R}^2$, and consider the class of *monotone allocations*:

$$\mathcal{G} = \left\{G : G = \{(x_1, x_2) \in \mathcal{X} |\ x_2 \geq f(x_1) \text{ for } f : \mathcal{X}_1 \to \mathcal{X}_2 \text{ increasing}\}\right\} .$$

As an example, consider again the setting of assigning scholarships to students (Example 2.2), but now suppose that the covariates available to the planner are parental income (x_1) and a student's GPA (x_2). The allocation rules considered in $\mathcal{G}$ are such that the GPA requirement for scholarship eligibility increases with parental income. Such a restriction could be imposed exogenously or could potentially arise as a shape restriction from an economic model. ■

Given a feasible class $\mathcal{G}$, we denote the highest attainable welfare by:

$$W^*_{\mathcal{G}} := \sup_{G \in \mathcal{G}} W(G) .$$

A *decision rule* is a function $\hat{G}$ from the observed data $\{(Y_i, D_i, X_i)\}_{i=1}^n$ into the set of admissible allocations $\mathcal{G}$. We call the rule that we develop and study in this paper the *Penalized Welfare Maximization* (or PWM) rule. As in much of the literature that follows the work of Manski (2004), we assume that the planner is interested in rules $\hat{G}$ that, on average, are close to the highest attainable welfare. To that end, the criterion by which we evaluate a decision rule is given by what we call maximum *$\mathcal{G}$-regret*:

$$\sup_P E_{P^n}[W^*_{\mathcal{G}} - W(\hat{G})] .$$

3 Penalized Welfare Maximization

In this section, we present the main results of our paper. In Section 3.1, we review some properties of the empirical welfare maximization (EWM) rule of Kitagawa and Tetenov (2018), which will motivate the PWM rule and serve as an important building block in its construction. In Section 3.2, we define the penalized welfare maximization rule and present bounds on its maximum $\mathcal{G}$-regret for general penalties. In Section 3.3 we illustrate these results by applying them to some specific penalties, and in particular we show that a standard "hold-out" procedure can be formalized as a penalty which satisfies our assumptions. In Section 3.4 we present results for a modification of the PWM rule for quasi-experimental settings where the propensity score is not known and must be estimated.

3.1 Empirical Welfare Maximization: a Review and Some Motivation

The EWM rule solves a sample analog of the population welfare maximization problem:

$$\hat{G}_{EWM} \in \arg\max_{G\in\mathcal{G}} W_n(G) \ ,$$

where

$$W_n(G) := \frac{1}{n}\sum_{i=1}^{n}\tau_i \mathbf{1}\{X_i \in G\} := \frac{1}{n}\sum_{i=1}^{n}\Big[\Big(\frac{Y_iD_i}{e(X_i)} - \frac{Y_i(1-D_i)}{1-e(X_i)}\Big)\mathbf{1}\{X_i \in G\}\Big] \ . \tag{2}$$

Kitagawa and Tetenov (2018) show how to formulate this problem as a Mixed Integer Linear Program (MILP) for many classes $\mathcal{G}$ of practical interest (see Appendix C for examples). Alternatively, direct parameter search has been shown to be very effective at solving the welfare maximization problem in certain applications as well: see for example Zhou et al. (2018). Note that to solve this optimization problem, the planner must know the propensity score $e(\cdot)$. This assumption is reasonable if the data comes from a randomized experiment, but clearly could not be made in a setting where the planner is using observational data. Kitagawa and Tetenov (2018) derive results for a modified version of the EWM rule where the propensity score is estimated, which we will review in Section 3.4.

To derive their non-asymptotic bounds on the maximum $\mathcal{G}$-regret of the EWM rule, Kitagawa and Tetenov (2018) make the following assumptions, which we will also maintain in our results:

Assumption 3.1. *(Bounded Outcomes and Strict Overlap) The set of distributions $\mathcal{P}(M,\kappa)$ has the following properties:*

- *There exists some $M < \infty$ such that the support of the outcome variable Y is contained in $[-\frac{M}{2}, \frac{M}{2}]$.*
- *There exists some $\kappa \in (0, 0.5)$ such that $e(x) \in [\kappa, 1-\kappa]$ for all x.*

In order to derive their results, Kitagawa and Tetenov (2018) also make the following assumption, which we will *not* require:

Assumption 3.2. *(Finite VC Dimension): $\mathcal{G}$ has finite VC dimension $V < \infty$.*

Such an assumption may or may not be restrictive depending on the application in question. Consider Example 2.2, the class of threshold allocations on $\mathbb{R}$. This class has VC dimension 2, thus Assumption 3.2 holds. On the other hand, it can be shown that the class of monotone allocations on $[0, 1]^2$ that was introduced in Example 2.3 has infinite VC dimension (see Györfi et al., 1996).

Given Assumptions 3.1 and 3.2, Kitagawa and Tetenov (2018) derive the following non-asymptotic upper bound on the maximum $\mathcal{G}$-regret of the EWM rule:

$$\sup_{P \in \mathcal{P}(M,\kappa)} E_{P^n}[W^*_{\mathcal{G}} - W(\hat{G}_{EWM})] \leq C\frac{M}{\kappa}\sqrt{\frac{V}{n}} \ , \tag{3}$$

for some universal constant C. Moreover, when X has sufficiently "large" support, they derive the following *lower* bound: for *any* decision rule $\hat{G}$,

$$\sup_{P \in \mathcal{P}(M,\kappa)} E_{P^n}[W^*_{\mathcal{G}} - W(\hat{G})] \geq RM\sqrt{\frac{V-1}{n}} \ , \tag{4}$$

for R a universal constant and for all sufficiently large n. This shows that the rate of convergence of maximum $\mathcal{G}$-regret implied by (3) is the best possible, i.e. that no other decision rule could achieve a faster rate without imposing additional assumptions.

Remark 3.1. Theorem 2.2 in Kitagawa and Tetenov (2018), which establishes (4), implies that if X has "large" support and we do not impose additional restrictions on the set of distributions, then it is *impossible* to derive a uniform rate of convergence of maximum $\mathcal{G}$-regret for *any* rule, for classes $\mathcal{G}$ of infinite VC dimension. A related result is derived in Stoye (2009), who shows that in a setting with a continuous covariate, and for any sample size, flipping a coin to assign individuals to treatment is minimax-regret optimal. Hence we will require additional restrictions on the set of distributions when deriving bounds on maximum regret for classes $\mathcal{G}$ of infinite VC dimension. ■

3.2 Penalized Welfare Maximization: General Results

We now consider a setting where the class $\mathcal{G}$ of admissible rules is "large", but can be "approximated" by a sequence of less complex subclasses $\mathcal{G}_k$:[1]

$$\mathcal{G}_1 \subseteq \mathcal{G}_2 \subseteq \mathcal{G}_3 \subseteq \cdots \subseteq \mathcal{G}_k \subseteq \cdots \subseteq \mathcal{G} \ .$$

Let $\hat{G}_{n,k}$ be the EWM rule in the class $\mathcal{G}_k$. Then we can decompose the $\mathcal{G}$-regret of the rule $\hat{G}_{n,k}$ as follows:

$$E_{P^n}[W^*_{\mathcal{G}} - W(\hat{G}_{n,k})] = E_{P^n}[W^*_{\mathcal{G}_k} - W(\hat{G}_{n,k})] + W^*_{\mathcal{G}} - W^*_{\mathcal{G}_k} \ .$$

[1] As can be seen from the proofs, the results we present below remain valid even if the sequence $\{\mathcal{G}_k\}_k$ is not nested.

Given this decomposition, we call

$$E_{P^n}[W^*_{\mathcal{G}_k} - W(\hat{G}_{n,k})] \ ,$$

the *estimation* error of the rule $\hat{G}_{n,k}$ in the class $\mathcal{G}_k$, and we call

$$W^*_{\mathcal{G}} - W^*_{\mathcal{G}_k} \ ,$$

the *approximation* error (or bias) of the class $\mathcal{G}_k$. Note that since the classes $\{\mathcal{G}_k\}_k$ are nested, the estimation error (respectively approximation bias) is non-decreasing (resp. non-increasing) with respect to k. If one had sharp uniform bounds on these errors, then an appropriate choice of k would be one that minimizes the sum of these bounds. In Theorem 3.1, we derive an oracle inequality which shows that PWM selects such a k, in a data-driven fashion. We use this feature of PWM to derive bounds on maximum regret in two settings of empirical interest.

The first setting we consider is one where $\mathcal{G}$ has infinite VC dimension (see Examples 2.1 and 2.3). In this setting, performing EWM on the whole class $\mathcal{G}$ may be undesirable. For example, regret may not converge to zero, or may converge to zero at a suboptimal rate (see Birgé and Massart, 1993, for related results in a regression context), or it may simply be the case that maximization over $\mathcal{G}$ is computationally difficult. Instead, we apply EWM to an approximating class $\mathcal{G}_k$, and we allow the complexity of the approximating class to grow as the sample size increases. We present examples of relevant approximating classes in Examples 3.2 and 3.3 below. In Corollary 3.1 we establish a bound on maximum regret in this setting.

The second setting that we consider is one where the class $\mathcal{G}$ has finite but large VC dimension relative to the sample size. This situation can arise, for instance, in applications where the planner has a large set of covariates on which to base treatment, and where the feasible allocations are threshold allocations (see Example 3.1 below). The bound on regret given by (3) increases with the VC dimension V of $\mathcal{G}$, so that EWM tends to "overfit"' the data when V is large relative to the sample size. In this situation, it may be beneficial to perform EWM in a subclass $\mathcal{G}'$ of smaller VC dimension, and hence we face the same tradeoff between estimation and approximation error that was noted above. In Corollary 3.2 we specialize Theorem 3.1 to a finite collection of approximating classes, and then in Corollary 3.3 establish a bound for the PWM rule which shows that it behaves as if we knew the correct class $\mathcal{G}'$ to use ex-ante, in the special case where the optimal allocation in $\mathcal{G}$ is contained in $\mathcal{G}'$.

We impose the following assumption on our sequence of classes, which we call a *sieve* of $\mathcal{G}$:

Assumption 3.3. *The sequence of classes*

$$\mathcal{G}_1 \subseteq \mathcal{G}_2 \subseteq \mathcal{G}_3 \subseteq \cdots \subseteq \mathcal{G}_k \subseteq \cdots \subseteq \mathcal{G}$$

is such that each class $\mathcal{G}_k$ has VC dimension V_k, which is finite.[2]

We present three examples of sieves $\mathcal{G}_k$ in Examples 3.1, 3.2, 3.3 below. Given a sieve $\{\mathcal{G}_k\}_k$, let

$$\hat{G}_{n,k} := \arg\max_{G \in \mathcal{G}_k} W_n(G) \ ,$$

be the EWM rule in the class $\mathcal{G}_k$. Our goal is to select the appropriate class k^* in which to perform EWM. We do this by selecting the class k^* in the following way: for each class $\mathcal{G}_k$, suppose we had some (potentially data-dependent) measure $C_n(k)$ of the amount of "overfitting" that results from using the rule $\hat{G}_{n,k}$ (Assumption 3.4 specifies our precise conditions on $C_n(k)$). Given such a measure $C_n(k)$, let $\{t_k\}_{k=1}^{\infty}$ be an increasing sequence of real numbers, and define the following penalized objective function:

$$R_{n,k}(G) := W_n(G) - C_n(k) - \sqrt{\frac{t_k}{n}} \ . \tag{5}$$

Then the *penalized welfare maximization* rule $\hat{G}_n$ is defined as follows:

$$\hat{G}_n := \hat{G}_{n,\hat{k}^*} \ ,$$

where

$$\hat{k}^* := \arg\max_{k} R_{n,k}(\hat{G}_{n,k}) \ .$$

In words, the PWM rule selects an allocation which maximizes a penalized version of the empirical welfare, with the penalty for allocations in $\mathcal{G}_k$ given by the term $C_n(k)$ (plus the auxiliary term $\sqrt{t_k/n}$).

Remark 3.2. Note that the PWM objective function $R_{n,k}(\cdot)$ includes the term: $\sqrt{t_k/n}$. This component of the objective is a technical device that is used to ensure that the classes get penalized at a sufficiently fast rate as k increases. The dependence of the penalty term on the sequence $\{t_k\}_k$ is somewhat undesirable, as it implies that the size of the penalty term for a given class depends on the specific choice of the sequence $\{t_k\}_k$. This technical device seems—however—unavoidable, and similar terms are pervasive throughout the literature on model selection in classification: see Koltchinskii (2001), Bartlett et al. (2002), Boucheron et al. (2005), Koltchinskii (2008). Nevertheless, as we will show, our results hold for many choices of $\{t_k\}_{k=1}^{\infty}$ (including our preferred choice $t_k = k$), and the choice is reflected explicitly in the bounds that we derive. Moreover, if one is only interested in using PWM in settings where the sequence of classes is finite, then we will show that the $\sqrt{t_k/n}$ term is not required. For simplicity, and unless otherwise specified, we will present all of our results with the specific choice $t_k = k$; in practice we find that overall performance of our procedure is essentially unaffected by this decision. ■

[2] Kitagawa and Tetenov (2018) additionally assume that their class $\mathcal{G}$ is countable so as to avoid potential measurability concerns. We instead choose not to address these concerns explicitly, as is done in most of the literature on classification. See Van Der Vaart and Wellner (1996) for a discussion of possible resolutions to this issue.

Remark 3.3. As noted by Kitagawa and Tetenov (2018), given a sieve $\{\mathcal{G}_k\}_k$, one can use their results to derive uniform (w.r.t $\mathcal{P}(M,\kappa)$) bounds on the estimation error. If one has in addition uniform bounds on the approximation bias, then one can construct a decision rule $\hat{G}_{n,k(n)}$, where $k(n)$ minimizes sum of these bounds. However, the merit of such an approach would depend on obtaining "good" computable bounds for the estimation and approximation error, which may be difficult to do in practice. For instance, the uniform bounds on the estimation error from Kitagawa and Tetenov (2018) depend on the VC dimension of the classes $\{\mathcal{G}_k\}_k$ which may be hard to bound precisely. Furthermore, a deterministic choice of $k(n)$ may lead to suboptimal rates if the true DGP satisfies additional regularity conditions which may be unknown to the econometrician. Given these challenges, PWM displays two advantages. First, PWM will perform−in a data-driven way−the optimal tradeoff between the approximation and estimation error, without relying on explicit bounds for these quantities. Second, PWM will select the subclass $\hat{k}$ over which to perform EWM in a way that adapts to additional "regularities" that may be satisfied by the true DGP. ■

We present three examples of sieves $\{\mathcal{G}_k\}_k$:

Example 3.1. Recall the class of threshold allocations in one dimension introduced in Example 2.2. Now we introduce the class of threshold allocations in K dimensions. Let $x = (x_1, ..., x_k) \in \mathcal{X} \subset \mathbb{R}^K$, and consider the following class $\mathcal{G}$:

$$\mathcal{G} = \left\{G \subset \mathcal{X} : G = \{x \in \mathcal{X} : s_k x_k \leq \bar{x}_k \text{ for } k \in \{1, ..., K\}\}, \bar{x} \in \mathbb{R}^K, s \in \{-1, 1\}^K\right\}$$

For large K, the VC dimension of $\mathcal{G}$ can become large relative to the sample size, and we may want to base treatment only on a smaller subset of the covariates. This is a variant of the best-subset selection problem (see Chen and Lee (2016) for related results in a classification context). However, the question still remains as to *how many* covariates to consider. Consider the sieve sequence $\{\mathcal{G}_k\}_{k=1}^K$, where $\mathcal{G}_k$ corresponds to the class of threshold allocations that uses $(k-1)$ out of K covariates, then PWM applied to this sieve can determine, in a data-driven way, the number of covariates to use for treatment assignment In Appendix B.3 we illustrate PWM's ability to reduce regret in this context in a simulation study. ■

Example 3.2. Recall the class of monotone allocations introduced in Example 2.3. Suppose that $\mathcal{X} = [0,1]^2$, so that $\mathcal{G}$ has infinite VC dimension (see Györfi et al., 1996, for a proof of this fact). We will construct a sieve for $\mathcal{G}$ where we approximate sets in $\mathcal{G}$ with sets that feature monotone, piecewise-linear boundaries. We proceed in three steps.

First define, for T an integer and $0 \leq j \leq T$, the following function $\psi_{T,j} : [0,1] \to [0,1]$:

$$\psi_{T,j}(x) = \begin{cases} 1 - |Tx - j|, & x \in [\frac{j-1}{T}, \frac{j+1}{T}] \cap [0,1] \\ 0, & \text{otherwise} . \end{cases}$$

The function $\psi_{T,j}(\cdot)$ is simply a triangular kernel whose base shifts with j and is scaled by T. Next, using these functions, define the following classes $\mathcal{S}_k$:

$$\mathcal{S}_k = \left\{ G : G = \{x = (x_1, x_2) \in \mathcal{X} | \sum_{j=0}^{T} \theta_j \psi_{T,j}(x_1) + x_2 \geq 0\} \text{ for } \theta_j \in \mathbb{R}, \, \forall \, 0 \leq j \leq T \right\} ,$$

where $T = 2^{k-1}$. It can be shown using results in Dudley (1999) that $\mathcal{S}_k$ has VC dimension $T + 2$. In words, the sets in $\mathcal{S}_k$ divide the covariate space into treatment and control such that the resulting boundary is a piecewise linear curve.

Finally, to construct our approximating class $\mathcal{G}_k$, we modify the class $\mathcal{S}_k$ to ensure that the resulting treatment allocations are monotone. For T an integer, let D_T be the following $T \times (T+1)$ "difference" matrix:

$$D_T := \begin{bmatrix} -1 & 1 & 0 & \dots & 0 & 0 \\ 0 & -1 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & -1 & 1 \end{bmatrix} .$$

Then $\mathcal{G}_k$ is defined as follows:

$$\mathcal{G}_k = \left\{ G : G \in \mathcal{S}_k \text{ and } D_T \Theta_T \geq 0 \, , \, \Theta_T = [\theta_0 \cdots \theta_T]' \right\} ,$$

for $T = 2^{k-1}$. This construction, which we borrow from Beresteanu (2004), is useful as it imposes monotonicity through a *linear* constraint, which is ideal for our implementation. In Section 4, we use this sequence of approximating classes in an application to the JTPA study, and then derive bounds on the maximum regret of PWM when $\mathcal{X} = [0, 1]^2$; Proposition 4.1 provides a uniform rate at which $W^*_{\mathcal{G}_k} \to W^*_{\mathcal{G}}$ under some additional regularity conditions, and Corollary 4.1 derives the corresponding bound on maximum $\mathcal{G}$-regret. ■

Example 3.3. Decision-tree based policy classes have recently become popular in treatment choice (see for example Athey and Wager, 2017; Kallus, 2016). Suppose the planner faces no restrictions on treatment assignment, so that $\mathcal{G}$ is the class of all measurable subsets of $\mathcal{X}$. In this case we could consider approximating $\mathcal{G}$ via decision trees of *increasing depth*. PWM could them be used to select the appropriate depth to use in practice. ■

We are now prepared to state the main results of the paper. We require the following high-level condition on the penalty $C_n(k)$:

Assumption 3.4. *There exist positive constants c_0 and c_1 such that $C_n(k)$ satisfies the following tail inequality for every n, k, and for every $\epsilon > 0$:*

$$\sup_{P \in \mathcal{P}(M,\kappa)} P^n(W_n(\hat{G}_{n,k}) - W(\hat{G}_{n,k}) - C_n(k) > \epsilon) \leq c_1 e^{-2c_0 n \epsilon^2} .$$

Let us provide some intuition for this assumption. Given an EWM rule $\hat{G}_{n,k}$, the value of the empirical welfare is given by $W_n(\hat{G}_{n,k})$. To evaluate $\mathcal{G}$-regret, we would ideally like to know the value of population welfare $W(\hat{G}_{n,k})$. Although the latter quantity is unknown, if we could define the (infeasible) penalty $C_n(k)$ as $W_n(\hat{G}_{n,k}) - W(\hat{G}_{n,k})$, then the penalized objective $W_n(\hat{G}_{n,k}) - C_n(k)$ would be exactly equal to $W(\hat{G}_{n,k})$. Since implementing such a $C_n(k)$ is impossible, our assumption requires that our feasible penalty be a good (empirical) upper bound on $W_n(\hat{G}_{n,k}) - W(\hat{G}_{n,k})$. In Section 3.3, we provide some specific examples of penalties that satisfy this assumption. In particular, we show that a standard "hold-out" procedure can be formalized as such a penalty. We are now ready to state our main workhorse result: an *oracle inequality* that characterizes the $\mathcal{G}$-regret of the PWM rule.

Theorem 3.1. *Suppose that Assumptions 2.1, 3.1, 3.3 and 3.4 hold, and set $t_k = k$ in (5). Then there exist constants Δ and c_0 such that for every $P \in \mathcal{P}(M, \kappa)$:*

$$E_{P^n}[W^*_{\mathcal{G}} - W(\hat{G}_n)] \leq \inf_k \left[E_{P^n}[C_n(k)] + \left(W^*_{\mathcal{G}} - W^*_{\mathcal{G}_k}\right) + \sqrt{\frac{k}{n}}\right] + \sqrt{\frac{\log(\Delta e)}{2c_0 n}} \ .$$

Theorem 3.1 forms the basis of all the results we present in Sections 3.2 and 3.3. It says that, at least from the perspective of *pointwise* (as opposed to maximum) $\mathcal{G}$-regret, the PWM rule is able to balance the tradeoff between $E_{P^n}[C_n(k)]$ and the approximation error, at the cost of adding two additional terms that are $O(1/\sqrt{n})$. The relative importance of these terms is hard to quantify at this level of generality, and we will attempt to shed some light on them, for specific penalties, in Section 3.3. Note that this result does not *quite* accomplish our initial goal of balancing the estimation and approximation error along our sieve: it is possible to choose a $C_n(k)$ that satisfies Assumption 3.4 for which $E_{P^n}[C_n(k)]$ is too large a bound for the estimation error. For this reason, we also impose the requirement that any penalty we consider should have the following additional property:

Assumption 3.5. *There exists a positive constant C_1 such that, for every n, $C_n(k)$ satisfies*

$$\sup_{P \in \mathcal{P}(M,\kappa)} E_{P^n}[C_n(k)] \leq C_1 \sqrt{\frac{V_k}{n}} \ ,$$

where V_k is the VC dimension of $\mathcal{G}_k$.

This assumption ensures that $E_{P^n}[C_n(k)]$ is comparable to the estimation error for EWM derived in (3), which was shown to be rate-optimal (for the class $\mathcal{P}(M, \kappa)$) in (4).

The next result we present is a bound on maximum regret for our first setting of interest: choosing the appropriate approximating class when $\mathcal{G}$ has infinite VC dimension. As discussed in Remark 3.1, a bound on maximum regret may not exist unless we impose additional regularity conditions on the family of DGPs under consideration. Hence we make the additional assumption that we restrict ourselves to a set of distributions $\mathcal{P}_r$ for which there exists a uniform bound on the approximation error. Note however that we do not assume that the rate of decay of the approximation bias is necessarily known to the econometrician, thus illustrating the "oracle" nature of our results.

Assumption 3.6. *Let $\mathcal{P}_r$ be a set of distributions such that*

$$\sup_{P \in \mathcal{P}_r} W^*_{\mathcal{G}} - W^*_{\mathcal{G}_k} = O(\gamma_k) \ ,$$

$$\sup_{P \in \mathcal{P}_r \cap \mathcal{P}(M,\kappa)} E_{P^n}[C_n(k)] = O(\zeta(k,n)) \ ,$$

for a sequence $\gamma_k \to 0$, and $\zeta(k,n)$ non-decreasing in k, $\zeta(k,n) \to 0$ as $n \to \infty$.

The first assumption asserts that we have a uniform bound on the approximation error. We present an example of such a uniform bound for our application in Section 4. The second assumption is made to highlight the following possibility: although Assumption 3.5 guarantees that we can satisfy this restriction with $\zeta(k,n) = \sqrt{V_k/n}$, it is possible that, once we have imposed that P must lie in $\mathcal{P}_r$, an even *tighter* bound may exist on $C_n(k)$ (see for example the discussion which follows Corollary 4.1 below). We emphasize that PWM will balance the tradeoff between the estimation and approximation error according to the *tighest* possible bounds on $E_P[C_n(k)]$ and $W^*_{\mathcal{G}} - W^*_{\mathcal{G}_k}$, regardless of whether or not we know these bounds in a given application.

Remark 3.4. A well-known restriction on the class of distributions which may lead to faster rates $\zeta(k,n)$ for certain choices of $C_n(k)$ is the *margin assumption* (see Kitagawa and Tetenov, 2018, for a formal definition in the context of treatment choice). Roughly, the margin assumption imposes restrictions on the behavior of $\tau(\cdot)$ near zero, and thus allows for faster than root-n rates of convergence. Although the study of margin-adaptive penalties is beyond the scope of our paper, Massart (2007) argues (in a classification context) that the hold-out penalty is margin-adaptive. We introduce this penalty in Section 3.3 below. ■

Given Assumption 3.6, we immediately obtain our first corollary:

Corollary 3.1. *Under Assumptions 2.1, 3.1, 3.3, 3.4, and 3.6, we have that*

$$\sup_{P \in \mathcal{P}_r \cap \mathcal{P}(M,\kappa)} E_{P^n}[W^*_{\mathcal{G}} - W(\hat{G}_n)] \leq \inf_k \left[O(\zeta(k,n)) + O(\gamma_k) + \sqrt{\frac{k}{n}} \right] + \sqrt{\frac{\log(\Delta e)}{2c_0 n}} \ .$$

As mentioned in Remark 3.3, if $\{\zeta(k,n)\}_{k,n}$ and $\{\gamma_k\}_k$ were known, then we could achieve such a result with a deterministic sequence $k(n)$. The strength of the PWM rule then is that we achieve the *same* behavior for any class $\mathcal{G}$ and approximating sequence $\{\mathcal{G}_k\}_k$ without having to know these quantities in practice. We present an application of this result in Section 4, in the setting of Example 3.2.

The second Corollary we present specializes Theorem 3.1 to our second setting of interest: the appropriate selection of a subclass when the VC-dimension of $\mathcal{G}$ is finite and large relative to the sample size. The result highlights two important points. First, it shows that by balancing the trade-off between the approximation and estimation error, PWM can potentially lead to a reduction in regret (relative to EWM) for values of the sample size that are comparable in magnitude to the VC-dimension of $\mathcal{G}$. Second, it illustrates how our bound changes when the sieve is finite and we drop the auxiliary $\sqrt{k/n}$ component of our penalty.

Corollary 3.2. *Suppose that Assumptions 2.1, 3.1, 3.3, 3.4, and 3.5 hold, and that $\mathcal{G}_K = \mathcal{G}$ for some finite K. Furthemore, suppose that in our definition of the penalty we omit the term $\sqrt{k/n}$. Then we have that*

$$E_{P^n}[W^*_{\mathcal{G}} - W(\hat{G}_n)] \leq \inf_{1\leq k\leq K} \left[C_1\sqrt{\frac{V_k}{n}} + \left(W^*_{\mathcal{G}} - W^*_{\mathcal{G}_k}\right)\right] + \sqrt{\frac{\log(Kc_1e)}{2c_0n}} \ .$$

Note that if the above bound is minimized at $k = K$, then the approximation error $W^*_{\mathcal{G}} - W^*_{\mathcal{G}_k}$ is zero and the resulting bound is comparable to the one derived in (3), with one additional term. In Section 3.3 we argue that for specific choices of $C_n(k)$ this term can be quantified more precisely.

Our final corollary of Section 3.2 considers the particular setting in which the constrained optimum $W^*_{\mathcal{G}}$ over the class $\mathcal{G}$ is *achieved* in $\mathcal{G}_{k_0}$, for some k_0, but that this class is unknown to the econometrician. The result shows that the resulting upper bound on maximum regret for PWM is as if we had performed EWM in the appropriate class $\mathcal{G}_{k_0}$.

Corollary 3.3. *Suppose that Assumptions 2.1, 3.1, 3.3, 3.4, and 3.5 hold, and let $\mathcal{P}_k \subset \mathcal{P}(M,\kappa)$ be the set of distributions such that $G^* \in \mathcal{G}_k$, then*

$$\sup_{P\in\mathcal{P}_k} E_{P^n}[W^*_{\mathcal{G}} - W(\hat{G}_n)] \leq C_1\sqrt{\frac{V_k}{n}} + \sqrt{\frac{k}{n}} + \sqrt{\frac{\log(\Delta e)}{2c_0n}} \ .$$

Furthermore, if $\{\mathcal{G}_k\}_{k=1}^K$ is finite, and we do not include the $\sqrt{k/n}$ term as discussed in Remark 3.2, then we have that:

$$\sup_{P\in\mathcal{P}_k} E_{P^n}[W^*_{\mathcal{G}} - W(\hat{G}_n)] \leq C_1\sqrt{\frac{V_k}{n}} + \sqrt{\frac{\log(Kc_1e)}{2c_0n}} \ ,$$

where c_0, c_1 are as in Assumption 3.4.

3.3 Penalized Welfare Maximization: Some Examples of Penalties

This section illustrates the results of Section 3.2 with two concrete choices for the penalty $C_n(k)$, and quantifies the size of the auxiliary term in the bound of Theorem 3.1 for these penalties. The first penalty we present, the holdout penalty, formalizes a sample-splitting procedure. The second penalty, the Rademacher penalty, does not involve sample-splitting but could be computationally burdensome in practice. Both of the penalties share the property that they do not require that the practitioner have precise bounds on the VC dimensions V_k of the approximating classes (only that they are finite), which we feel is important to make the method broadly applicable.

3.3.1 The Holdout Penalty

The first penalty we introduce is motivated by the following idea: fix some number $\ell \in (0,1)$ such that $m := n(1-\ell)$ (for expositional simplicity suppose that m is an integer), and let $r := n - m$. Given our original sample $S_n = \{(Y_i, D_i, X_i)\}_{i=1}^n$, let $S_n^E := \{(Y_i, D_i, X_i)\}_{i=1}^m$ denote what we call the *estimating* sample, and let $S_n^T := \{(Y_i, D_i, X_i)\}_{i=m+1}^n$ denote the *testing* sample. Now, using S_n^E, compute $\hat{G}_{m,k}$ for each k. Intuitively, we could get a sense of the efficacy of $\hat{G}_{m,k}$ by applying this rule to the subsample S_n^T and computing the empirical welfare $W_r(\hat{G}_{m,k})$. We could then select the class k that results in the highest empirical welfare $W_r(\hat{G}_{m,k})$.

This idea can be formalized in our framework by treating it as a PWM-rule on the estimating sample, with the following penalty: for each EWM rule $\hat{G}_{m,k}$ estimated on S_n^E, let

$$W_m(\hat{G}_{m,k}) = \frac{1}{m}\sum_{i=1}^{m} \tau_i \mathbf{1}\{X_i \in \hat{G}_{m,k}\} \ ,$$

be the empirical welfare of the rule $\hat{G}_{m,k}$ on S_n^E and let

$$W_r(\hat{G}_{m,k}) = \frac{1}{r}\sum_{i=m+1}^{n} \tau_i \mathbf{1}\{X_i \in \hat{G}_{m,k}\} \ ,$$

be the empirical welfare of the rule $\hat{G}_{m,k}$ on S_n^T. We define the *holdout* penalty to be

$$C_m(k) := W_m(\hat{G}_{m,k}) - W_r(\hat{G}_{m,k}) \ .$$

Now, recall that the PWM rule is given by

$$\hat{G}_m = \arg\max_k \left[W_m(\hat{G}_{m,k}) - C_m(k) - \sqrt{\frac{k}{m}} \right] \ ,$$

which, given the definition of $C_m(k)$, simplifies to

$$\hat{G}_m = \arg\max_k \left[W_r(\hat{G}_{m,k}) - \sqrt{\frac{k}{m}} \right] \ .$$

Hence we see that the PWM rule with the holdout penalty reproduces the intuition presented above (with the usual addition of the $\sqrt{k/m}$ term; see Remark 3.2).

We check the conditions of Assumptions 3.4 and 3.5:

Lemma 3.1. *Consider Assumptions 2.1, 3.1, 3.3. Suppose we have a sample of size n and recall that $m = n(1-\ell)$ and $r = n - m$. Let $C_m(k)$ be the holdout penalty as defined above. Then we have that*

$$P^n(W_m(\hat{G}_{m,k}) - W(\hat{G}_{m,k}) - C_m(k) > \epsilon) \leq \exp\Big(-2\Big(\frac{\kappa}{M}\Big)^2 n\ell\epsilon^2\Big),$$

and

$$E_{P^n}[C_m(k)] \leq C\frac{M}{\kappa\sqrt{(1-\ell)}}\sqrt{\frac{V_k}{n}}\ ,$$

where C is the same universal constant that appears in equation (3).

With Lemma 3.1 established, Theorem 3.1 becomes:

Proposition 3.1. *Consider Assumptions 2.1, 3.1, 3.3. Suppose we have a sample of size n, and let $m = n(1-\ell)$, $r = n-m$. Let $C_m(k)$ be the holdout penalty as defined above. Then we have that for every $P \in \mathcal{P}(M,\kappa)$:*

$$E_{P^n}[W^*_{\mathcal{G}} - W(\hat{G}_m)] \leq \inf_k \Big[E_{P^n}[C_n(k)] + \big(W^*_{\mathcal{G}} - W^*_{\mathcal{G}_k}\big) + \sqrt{\frac{k}{n}}\Big] + g(M,\kappa,\ell)\frac{M}{\kappa\sqrt{\ell}}\sqrt{\frac{1}{n}}\ ,$$

with

$$E_{P^n}[C_n(k)] \leq C\frac{M}{\kappa\sqrt{(1-\ell)}}\sqrt{\frac{V_k}{n}}\ ,$$

where C is the same universal constant as that in equation (3) and

$$g(M,\kappa,\ell) := 2\sqrt{\log\Big(\sqrt{\frac{e}{2\ell}}\frac{M}{\kappa}\Big)}\ .$$

As we show in the next section, the bound in Proposition 3.1 is similar to what we derive for the Rademacher penalty, but with larger constants which reflect the fact that we split the sample. However, a major benefit of the holdout penalty lies in the fact that it is simple to implement. The only remaining issue is how to split the data. Although we do not study this problem formally, we have found that it is more important to focus on accurate estimation of the rule $\hat{G}_{m,k}$ than on the computation of $W_r(\hat{G}_{m,k})$. In other words, we recommend that the estimating sample S^E_n be a large proportion of the original sample S_n. Throughout the rest of the paper we designate three quarters of the sample as the estimating sample.

3.3.2 The Rademacher Penalty

The second penalty we present is attractive in that it does not introduce sample splitting, but may be computationally burdensome when compared to the holdout procedure. Let $S_n := \{(Y_i, D_i, X_i)\}_{i=1}^n$ be the observed data. Then the *Rademacher* penalty is given by

$$C_n(k) = E_\sigma\Big[\sup_{G\in\mathcal{G}_k}\frac{2}{n}\sum_{i=1}^n \sigma_i\tau_i\mathbf{1}\{X_i \in G\} \mid S_n\Big]\ ,$$

where τ_i is defined as in equation (2), and $\{\sigma_1, ..., \sigma_n\}$ are a sequence of i.i.d Rademacher variables, i.e. they take on the values $\{-1, 1\}$, each with probability half.

To clarify the origin of this penalty, recall that $C_n(k)$ must be a good upper bound on $W_n(\hat{G}_{n,k}) - W(\hat{G}_{n,k})$, which is the requirement of Assumption 3.4. Bounding such quantities is common in the study of empirical processes, and the usual first step is to use what is known as *symmetrization*, which gives the following bound:

$$E_{P^n}[\sup_{G \in \mathcal{G}} W_n(G) - W(G)] \leq E_{P^n}\Big[E_\sigma\big[\sup_{G \in \mathcal{G}} \frac{2}{n} \sum_{i=1}^{n} \sigma_i \tau_i \mathbf{1}\{X_i \in G\} \mid S_n\big]\Big] \ .$$

It is thus this inequality that inspires the definition of $C_n(k)$. The concept of Rademacher complexity[4] is pervasive throughout the statistical learning literature (see for example Bartlett et al., 2002; Bartlett and Mendelson, 2002; Koltchinskii, 2001). Intuitively, it measures a notion of complexity that is finer than that of VC dimension, and is at the same time computable from the data at hand.

First we prove that the conditions of Assumptions 3.4 and 3.5 hold for the Rademacher penalty:

Lemma 3.2. *Consider Assumptions 2.1, 3.1, 3.3. Let $C_n(k)$ be the Rademacher penalty as defined above. Then we have that*

$$P^n(W_n(\hat{G}_{n,k}) - W(\hat{G}_{n,k}) - C_n(k) > \epsilon) \leq \exp\Big(-2\Big(\frac{\kappa}{3M}\Big)^2 n\epsilon^2\Big),$$

and

$$E_{P^n}[C_n(k)] \leq C\frac{M}{\kappa}\sqrt{\frac{V_k}{n}} \ ,$$

where C is the same universal constant that appears in equation (3).

We are thus able to refine Theorem 3.1 to the case of the Rademacher penalty.

Proposition 3.2. *Consider Assumptions 2.1, 3.1, 3.3. Let $C_n(k)$ be the Rademacher penalty as defined above. Then we have that for every $P \in \mathcal{P}(M, \kappa)$:*

$$E_{P^n}[W^*_{\mathcal{G}} - W(\hat{G}_n)] \leq \inf_k \Big[E_{P^n}[C_n(k)] + \big(W^*_{\mathcal{G}} - W^*_{\mathcal{G}_k}\big) + \sqrt{\frac{k}{n}}\Big] + g(M, \kappa)\frac{M}{\kappa}\sqrt{\frac{1}{n}} \ ,$$

with $E_{P^n}[C_n(k)] \leq C\frac{M}{\kappa}\sqrt{\frac{V_k}{n}}$, where C is the same universal constant as that in equation (3) and

$$g(M, \kappa) := 6\sqrt{\log\Big(\frac{3\sqrt{e}}{\sqrt{2}}\frac{M}{\kappa}\Big)} \ .$$

Remark 3.5. In Appendix B we perform a back-of-the-envelope calculation that provides insight into the size of $g(M, \kappa)$, and compares it to the size of the universal constant C derived in Kitagawa and Tetenov (2018). ■

[4]Note that the definition of Rademacher complexity is slightly different than the definition of our penalty. Here we follow Bartlett et al. (2002) and do not include the absolute value in our definition of the penalty.

3.4 Penalized Welfare Maximization: Estimated Propensity Score

In this section we present a modification of the PWM rule where the propensity score is not known and must be estimated from the data. This situation would arise if the planner had access to observational data instead of data from a randomized experiment. Before describing our modification of the PWM rule, we first review results about the corresponding modification of the EWM rule. The modification we consider here is what Kitagawa and Tetenov (2018) call the *e-hybrid* EWM rule. Recall the EWM objective function as defined in equation (2). To define the e-hybrid EWM rule we modify this objective function by replacing τ_i with

$$\hat{\tau}_i := \Big[\frac{Y_i D_i}{\hat{e}(X_i)} - \frac{Y_i(1-D_i)}{1-\hat{e}(X_i)}\Big]\mathbf{1}\{\epsilon_n \le \hat{e}(X_i) \le 1-\epsilon_n\} ,$$

where $\hat{e}(\cdot)$ is an estimator of the propensity score, and ϵ_n is a trimming parameter such that $\epsilon_n = O(n^{-\alpha})$ for some $\alpha > 0$. The e-hybrid EWM objective function is defined as follows:

$$W_n^e(G) := \frac{1}{n}\sum_{i=1}^{n} \hat{\tau}_i \mathbf{1}\{X_i \in G\} .$$

Since we are now estimating the propensity score, we must impose additional regularity conditions on P to guarantee a uniform rate of convergence. We impose the following high level assumption:

Assumption 3.7. *Given an estimator $\hat{e}(\cdot)$, let $\mathcal{P}_e$ be a class of data generating processes such that*

$$\sup_{P\in\mathcal{P}_e} E_{P^n}\Big[\frac{1}{n}\sum_{i=1}^{n}|\hat{\tau}_i - \tau_i|\Big] = O(\phi_n^{-1}) ,$$

where $\phi_n \to \infty$.

Although we do not explore low-level conditions that satisfy this assumption here, Kitagawa and Tetenov (2018) do so in their paper. Let $\hat{G}_{e-hybrid}$ be the solution to the e-hybrid problem in a class $\mathcal{G}$ of finite VC dimension, then Kitagawa and Tetenov (2018) derive the following bound on maximum $\mathcal{G}$-regret:

$$\sup_{P\in\mathcal{P}_e\cap\mathcal{P}(M,\kappa)} E_{P^n}\Big[W_{\mathcal{G}}^* - W(\hat{G}_{e-hybrid})\Big] \le O(\phi_n^{-1} \vee n^{-1/2}) . \tag{6}$$

With a non-parametric estimator of $e(\cdot)$, ϕ_n will generally be slower than $\sqrt{n}$ and hence determine the rate of convergence. In a recent paper, Athey and Wager (2017) argue that more sophisticated estimators of the welfare objective can improve performance relative to the e-hybrid rule, and derive corresponding bounds on the maximum regret of their procedure. Importantly, by exploiting an orthogonal moments construction, the procedure in Athey and Wager (2017) converges at a $\sqrt{n}$-rate even when the propensity score is estimated non-parametrically. Modifying our method using their techniques would be an interesting direction for future work.

We now present the construction of the corresponding e-hybrid PWM estimator. Let $\mathcal{G}$ be an arbitrary class of allocations, and let $\{\mathcal{G}_k\}_k$ be some approximating sequence for $\mathcal{G}$. Let $\hat{G}^e_{n,k}$ be the hybrid EWM rule in the class $\mathcal{G}_k$. Let $C^e_n(k)$ be our penalty for the hybrid PWM rule. We require that the penalty satisfies the following properties:

Assumption 3.8. *(Assumptions on $C^e_n(k)$)*
In addition to making assumptions about $C^e_n(k)$, we assume there exists an "infeasible penalty" $\tilde{C}_n(k)$ with the following properties:

- *There exist positive constants c_0 and c_1 such that $\tilde{C}_n(k)$ satisfies the following tail inequality for every n, k and for every $\epsilon > 0$:*

$$\sup_{P \in \mathcal{P}_e \cap \mathcal{P}(M,\kappa)} P^n(W_n(\hat{G}^e_{n,k}) - W(\hat{G}^e_{n,k}) - \tilde{C}_n(k) > \epsilon) \leq c_1 e^{-2c_0 n \epsilon^2}$$

- *There exists a positive constant C_1 such that, for every n, $\tilde{C}_n(k)$ satisfies*

$$\sup_{P \in \mathcal{P}_e \cap \mathcal{P}(M,\kappa)} E_{P^n}[\tilde{C}_n(k)] \leq C_1 \sqrt{\frac{V_k}{n}} \ ,$$

  where V_k is the VC dimension of $\mathcal{G}_k$.

- *$\tilde{C}_n(k)$ and $C^e_n(k)$ are such that*

$$\sup_{P \in \mathcal{P}_e \cap \mathcal{P}(M,\kappa)} E_{P^n}\left[\sup_k \left| C^e_n(k) - \tilde{C}_n(k) \right| \right] = O(\phi_n^{-1}) \ .$$

Note that we have introduced an object $\tilde{C}_n(k)$ which we call an *infeasible penalty.* The first bullet point asserts that the infeasible penalty obeys a similar tail inequality to $C_n(k)$, except that $\tilde{C}_n(k)$ satisfies this assumption with respect to the *e-hybrid* EWM rule. However, we evaluate the hybrid rule through the empirical objective $W_n(\cdot)$, *which is the objective when the propensity score is known.* This is our motivation for calling $\tilde{C}_n(k)$ an infeasible penalty. Luckily, $\tilde{C}_n(k)$ is purely a theoretical device and does not serve a role in the actual implementation of PWM. We provide an example of such an infeasible penalty in the setting of the holdout penalty below. The second bullet point is the same as Assumption 3.5, but now with respect to the infeasible penalty $\tilde{C}_n(k)$. The third bullet simply links the true penalty $C^e_n(k)$ to the infeasible penalty $\tilde{C}_n(k)$ in such a way that both should agree asymptotically and do so at an appropriate rate.

Given this, we obtain the following analogue to Theorem 3.1:

Theorem 3.2. *Given assumptions 2.1, 3.1, 3.3, 3.7 and 3.8, there exist constants Δ and c_0 such that for every $P \in \mathcal{P}_e \cap \mathcal{P}(M,\kappa)$:*

$$E_{P^n}[W^*_{\mathcal{G}} - W(\hat{G}^e_n)] \leq \inf_k \left[E_{P^n}[\tilde{C}_n(k)] + \left(W^*_{\mathcal{G}} - W^*_{\mathcal{G}_k}\right) + \sqrt{\frac{k}{n}} \right] + O(\phi_n^{-1}) + \sqrt{\frac{\log(\Delta e)}{2c_0 n}} \ .$$

As we can see, the only difference between this bound and the bound derived in Theorem 3.1 is that there is an additional term of order ϕ_n^{-1}.

Next, we check that the conditions in Assumption 3.8 are satisfied with a modified version of the holdout penalty (the results for the Rademacher penalty follow similarly). To define the *hybrid holdout* penalty, let $\hat{e}^E(\cdot)$ be the propensity estimated on S_n^E, and let $\hat{e}^T(\cdot)$ be the propensity estimated on S_n^T. Define

$$W_m^e(G) := \frac{1}{m}\sum_{i=1}^{m} \hat{\tau}_i{}^E \mathbf{1}\{X_i \in G\} ,$$

where

$$\hat{\tau}_i{}^E = \Big[\frac{Y_i D_i}{\hat{e}^E(X_i)} - \frac{Y_i(1-D_i)}{1-\hat{e}^E(X_i)}\Big]\mathbf{1}\{\epsilon_n \leq \hat{e}^E(X_i) \leq 1-\epsilon_n\} .$$

Define $W_r^e(G)$ on the testing sample analogously. Letting $\hat{G}_{m,k}^e$ be the hybrid EWM rule computed on the estimating sample in the class $\mathcal{G}_k$, the hybrid holdout penalty is defined as:

$$C_m^e(k) := W_m^e(\hat{G}_{m,k}^e) - W_r^e(\hat{G}_{m,k}^e) .$$

We must also assert the existence of an infeasible penalty $\tilde{C}_m(k)$ that satisfies our assumptions. The infeasible penalty we consider is given by

$$\tilde{C}_m(k) := W_m(\hat{G}_{m,k}^e) - W_r(\hat{G}_{m,k}^e) ,$$

where $W_m(\cdot)$ and $W_r(\cdot)$ are defined as in Section 3.3, that is, they are computed *as if the propensity score were known.* Lemma 3.1 verifies Assumption 3.8 for the hybrid holdout penalty:

Lemma 3.3. *Assume Assumptions 2.1, 3.1, 3.3, and 3.7. Suppose we have a sample of size n and recall that $m = n(1-\ell)$ and $r = n - m$. Let $C_m^e(k)$ be the hybrid holdout penalty and $\tilde{C}_m(k)$ be the infeasible penalty as defined above. Then we have that*

$$P^n(W_m(\hat{G}_{m,k}^e) - W(\hat{G}_{m,k}^e) - \tilde{C}_m(k) > \epsilon) \leq \exp\Big(-2\Big(\frac{\kappa}{M}\Big)^2 n\ell\epsilon^2\Big),$$

$$E_{P^n}[\tilde{C}_m(k)] \leq C\frac{M}{\kappa\sqrt{(1-\ell)}}\sqrt{\frac{V_k}{n}} ,$$

and

$$\sup_{P\in\mathcal{P}_e} E_{P^n}\big[\sup_k |C_m^e(k) - \tilde{C}_m(k)|\big] = O(\phi_n^{-1}) ,$$

where C is the same universal constant as that in equation (3).

We thus obtain an analogous result to Proposition 3.1 for PWM with the hybrid holdout penalty.

4 An Application using Monotone Allocations

In this section we apply the PWM rule to the sieve we constructed in Example 3.2 for monotone allocations. First, we apply our method to experimental data from the Job Training Partnership Act (JTPA) Study. Then, we derive bounds on maximum regret in a setting where our class has infinite VC dimension.

The JTPA study was a randomized controlled trial whose purpose was to measure the benefits and costs of employment and training programs. The study randomized whether applicants would be eligible to receive a collection of services provided by the JTPA related to job training, for a period of 18 months. The study collected background information about the applicants prior to the experiment, as well as data on applicants' earnings for 30 months following assignment (for a detailed description of the study, see Bloom et al., 1997).[5]

We revisit the setup in Kitagawa and Tetenov (2018), which has frequently been considered in recent related papers. The outcome that we consider is total individual earnings in the 30 months following program assignment. The covariates on which we define our treatment allocations are the individual's years of education and their earnings in the year prior to the assignment. The set of allocations we consider is the set of monotone allocations defined in Example 2.3, but with a *non-increasing* monotone function. To be precise, let $\mathcal{X}_1$ be the covariate set of years of education, and let $\mathcal{X}_2$ be the covariate set of previous earnings, then the set of allocations we consider is given by:

$$\mathcal{G} = \left\{ G : G = \{(x_1, x_2) \in \mathcal{X} | \, x_2 \leq f(x_1) \text{ for } f : \mathcal{X}_1 \to \mathcal{X}_2 \text{ non-increasing}\} \right\} .$$

In the context of this application, this restriction imposes that, the less education you have, the more accessible is the program based on your previous earnings. It is plausible that such a restriction may be exogenously imposed on the planner for political reasons; after all, it may not be politically viable to implement a job-training program where only those with high levels of education or income are accepted.

The approximating sequence we consider is the one described in Example 3.2, but now with a non-increasing monotonicity constraint. Recall that this was a sequence such that the resulting allocations partitioned the covariate space with a progressively refined, piecewise-linear, monotone boundary. Given any fixed class in this sequence, we can perform EWM in that class. Figure 1 below illustrates the result of performing EWM on the simplest class in the approximating sequence. This class is equivalent to the class of linear treatment rules from Kitagawa and Tetenov (2018), but with an additional slope constraint.

[5] The sample we use is the same as that in Abadie et al. (2018), which we downloaded from ideas.repec.org/c/boc/bocode/s457801.html. We supplemented this dataset with education data from the `expbif.dta` dataset available at the W.E. Upjohn Institute website. Observations with years of education coded as '99' were dropped.

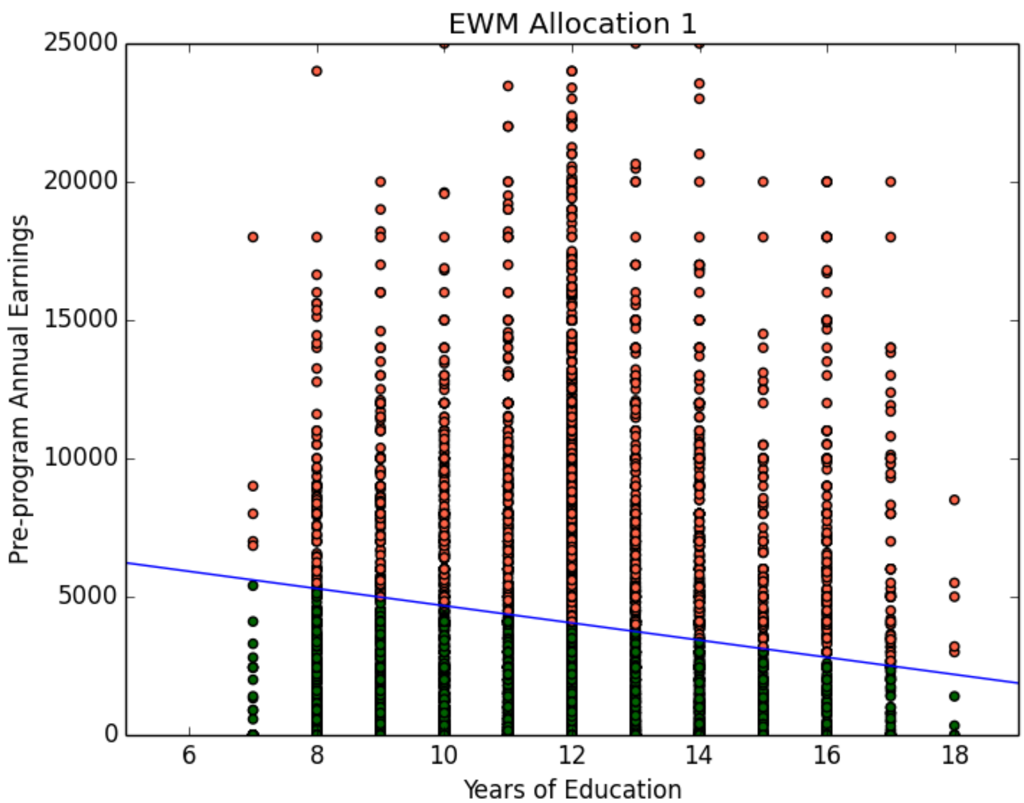

Figure 1: The resulting treatment allocation from performing EWM in $\mathcal{G}_1$.
Each point represents a covariate pair in the sample. The region shaded in green (dark) is the prescribed treatment region, the region shaded in red (light) is the prescribed control region.

At the other end of the spectrum, we could consider performing EWM in the most complicated class in our approximating sequence: this class corresponds to allocations that stipulate a threshold for previous income at every level of education (note that such a class exists here because years of education is discrete with finite support). Figure 2 below illustrates the result of performing EWM in this class.

As we can see, the resulting allocation in the simplest class and in the most complicated class look quite different, and given the option to choose any class from our sequence, it is not obvious which one should be chosen given the size of the experiment. Given that we have a finite sieve in this application, we can view the use of PWM in this context through the lenses of Corollaries 3.2 or 3.3. In Figure 3, we illustrate the result of performing PWM on our sequence of classes, where we used 3/4 of our sample for estimation. In Appendix C we discuss the computational details of our implementation. Note that PWM selects the allocation from the second class in our sequence, which corresponds to a piecewise-linear allocation with one possible "kink".

Remark 4.1. In Appendix B we perform a sample splitting exercise to estimate the welfare per-

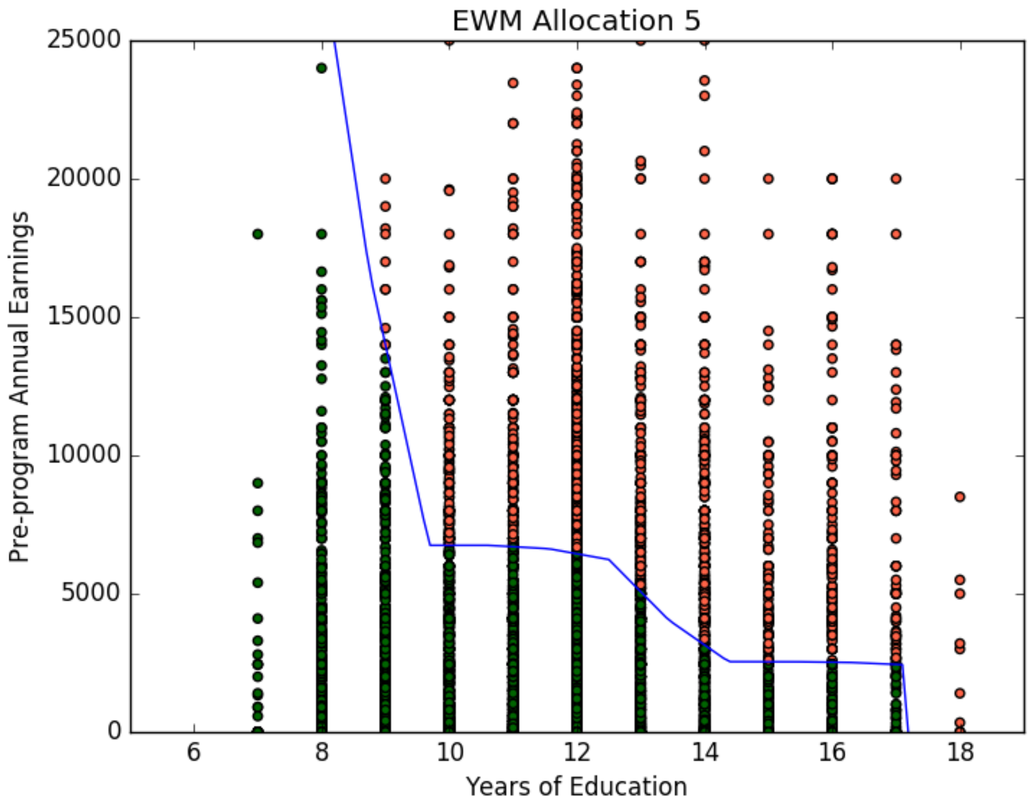

Figure 2: The resulting treatment allocation from performing EWM in $\mathcal{G}_5$.
Each point represents a covariate pair in the sample. The region shaded in green (dark) is the prescribed treatment region, the region shaded in red (light) is the prescribed control region.

formance of various decision rules on the JTPA data. In summary, we find that PWM can obtain higher estimated welfare than EWM in this application. However, we emphasize that this difference was not found to be statistically significant. ■

Next, we derive a bound on maximum regret in the setting where $\mathcal{X} = [0,1]^2$, so that our class has infinite VC dimension. We consider the following restriction on the class of distributions:

Assumption 4.1. *Let $\mathcal{P}_r$ be a set of DGPs such that there exists some constant $A > 0$, where for every distribution in $\mathcal{P}_r$, the marginal distribution of $X = (X_1, X_2)$ can be decomposed as follows:*

$$P_X(M_1 \times M_2) = \int_{M_2} P_{X_1|x_2}(M_1) dP_{X_2} \ ,$$

where M_1 and M_2 are measurable subsets of $[0,1]$, and $P_{X_1|x_2}$ is continuous with density bounded above by A, for all $x_2 \in [0,1]$.

In words, Assumption 4.1 requires that the conditional distribution of X_1 given X_2 is continuous

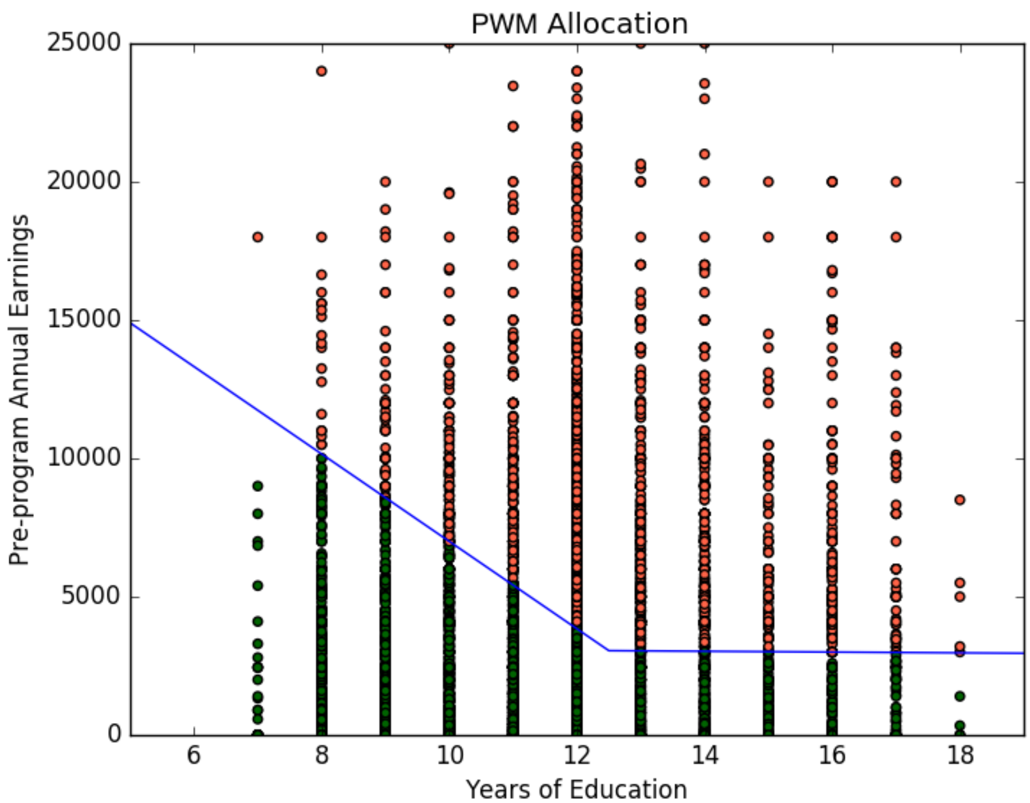

Figure 3: The resulting treatment allocation from performing PWM on the approximating sequence $\{\mathcal{G}_k\}_{k=1}^{5}$. Each point represents a covariate pair in the sample. The region shaded in green (dark) is the prescribed treatment region, the region shaded in red (light) is the prescribed control region.

with a uniformly bounded density. With this regularity condition imposed, we are able to derive the following uniform bound on the approximation bias $W^*_{\mathcal{G}} - W^*_{\mathcal{G}_k}$:

Proposition 4.1. *Under Assumption 4.1, the approximation bias of the approximating sequence $\{\mathcal{G}_k\}_{k=1}^{\infty}$ from Example 3.2 satisfies*

$$\sup_{P \in \mathcal{P}_r \cap \mathcal{P}(M,\kappa)} W^*_{\mathcal{G}} - W^*_{\mathcal{G}_k} \leq A \frac{M}{\kappa} 2^{-k} \ ,$$

To illustrate the use of Proposition 4.1 in our setting, we derive a bound on maximum regret for monotone allocations. Proposition 4.1 and Corollary 3.1, along with the bound on V_k given in Example 3.2 allow us to conclude that:

Corollary 4.1. *Let $C_n(k)$ be the Rademacher or holdout penalty. Under Assumptions 2.1, 3.1, 3.3, and 4.1, we have that*

$$\sup_{P \in \mathcal{P}_r \cap \mathcal{P}(M,\kappa)} E_{P^n}[W^*_{\mathcal{G}} - W(\hat{G}_n)] = O\left(n^{-\frac{1}{3}}\right) \ .$$

Corollary 4.1 establishes a polynomial rate of convergence for PWM. In contrast, we are not aware of any results which would allow us to derive a bound on maximum regret for EWM (or other related methods which do not employ a sieve construction) under Assumption 4.1.

In Appendix B.4, we derive a series of results on the behavior of EWM under suitable entropy restrictions on the class $\mathcal{G}$, which show that under the stronger assumption that X is continuous with a bounded density, EWM in fact achieves a root-n rate (up to a log factor) in this example, and that this rate is optimal. As we explain in Remark B.2, PWM can also achieve the same optimal rate of convergence in this setting, which would *not* be the case for a deterministic $k(n)$ chosen to obtain the rate derived in Corollary 4.1. This further reinforces the observation made in Remark 3.3 about the adaptation of PWM to additional regularities.

References

Abadie, Alberto, Matthew M Chingos, and Martin R West (2018), "Endogenous stratification in randomized experiments." *Review of Economics and Statistics*, 100, 567–580.

Armstrong, Timothy and Shu Shen (2015), "Inference on optimal treatment assignments."

Athey, Susan and Stefan Wager (2017), "Efficient policy learning." *arXiv preprint arXiv:1702.02896*.

Bartlett, Peter L (2008), "Fast rates for estimation error and oracle inequalities for model selection." *Econometric Theory*, 24, 545–552.

Bartlett, Peter L, Stéphane Boucheron, and Gábor Lugosi (2002), "Model selection and error estimation." *Machine Learning*, 48, 85–113.

Bartlett, Peter L and Shahar Mendelson (2002), "Rademacher and gaussian complexities: Risk bounds and structural results." *Journal of Machine Learning Research*, 3, 463–482.

Beresteanu, Arie (2004), "Nonparametric estimation of regression functions under restrictions on partial derivatives." Technical report, Duke University, Department of Economics.

Beygelzimer, Alina and John Langford (2009), "The offset tree for learning with partial labels." In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, 129–138, ACM.

Bhattacharya, Debopam and Pascaline Dupas (2012), "Inferring welfare maximizing treatment assignment under budget constraints." *Journal of Econometrics*, 167, 168–196.

Birgé, Lucien and Pascal Massart (1993), "Rates of convergence for minimum contrast estimators." *Probability Theory and Related Fields*, 97, 113–150.

Bloom, Howard S, Larry L Orr, Stephen H Bell, George Cave, Fred Doolittle, Winston Lin, and Johannes M Bos (1997), "The benefits and costs of jtpa title ii-a programs: Key findings from the national job training partnership act study." *Journal of human resources*, 549–576.

Boucheron, Stéphane, Olivier Bousquet, and Gábor Lugosi (2005), "Theory of classification: A survey of some recent advances." *ESAIM: probability and statistics*, 9, 323–375.

Chamberlain, Gary (2011), "Bayesian aspects of treatment choice." *The Oxford Handbook of Bayesian Econometrics*, 11–39.

Chen, Le-Yu and Sokbae Lee (2016), "Best subset binary prediction." *arXiv preprint arXiv:1610.02738*.

Dehejia, Rajeev H (2005), “Program evaluation as a decision problem.” *Journal of Econometrics*, 125, 141–173.

Dudley, Richard M (1999), *Uniform central limit theorems*, volume 23. Cambridge Univ Press.

Györfi, L, L Devroye, and G Lugosi (1996), *A probabilistic theory of pattern recognition*. Springer-Verlag.

Hirano, Keisuke and Jack R Porter (2009), “Asymptotics for statistical treatment rules.” *Econometrica*, 77, 1683–1701.

Kallus, Nathan (2016), “Learning to personalize from observational data.” *arXiv preprint arXiv:1608.08925*.

Kitagawa, Toru and Aleksey Tetenov (2018), “Who should be treated? empirical welfare maximization methods for treatment choice.” *Econometrica*, 86, 591–616.

Kock, Anders Bredahl and Martin Thyrsgaard (2017), “Optimal sequential treatment allocation.”

Koltchinskii, V (2008), “Oracle inequalities in empirical risk minimization and sparse recovery problems: Lecture notes.” Technical report, Technical report, Ecole d’ete de Probabilités de Saint-Flour, 2008. 12.6.

Koltchinskii, Vladimir (2001), “Rademacher penalties and structural risk minimization.” *IEEE Transactions on Information Theory*, 47, 1902–1914.

Manski, Charles F (2004), “Statistical treatment rules for heterogeneous populations.” *Econometrica*, 72, 1221–1246.

Massart, Pascal (2007), “Concentration inequalities and model selection.”

Qian, Min and Susan A Murphy (2011), “Performance guarantees for individualized treatment rules.” *Annals of statistics*, 39, 1180.

Rai, Yoshiyasu (2018), “Statistical inference for treatment assignment policies.”

Schlag, Karl H (2007), “Eleven - designing randomized experiments under minimax regret.” *Unpublished manuscript, European University Institute.*

Scott, Clayton and Robert D Nowak (2006), “Minimax-optimal classification with dyadic decision trees.” *IEEE transactions on information theory*, 52, 1335–1353.

Stoye, Jörg (2009), “Minimax regret treatment choice with finite samples.” *Journal of Econometrics*, 151, 70–81.

Stoye, Jörg (2012), “Minimax regret treatment choice with covariates or with limited validity of experiments.” *Journal of Econometrics*, 166, 138–156.

Swaminathan, Adith and Thorsten Joachims (2015), “Batch learning from logged bandit feedback through counterfactual risk minimization.” *Journal of Machine Learning Research*, 16, 1731–1755.

Tetenov, Aleksey (2012), “Statistical treatment choice based on asymmetric minimax regret criteria.” *Journal of Econometrics*, 166, 157–165.

Van Der Vaart, Aad W and Jon A Wellner (1996), “Weak convergence.” In *Weak Convergence and Empirical Processes*, 16–28, Springer.

Vapnik, Vladimir N and Alexey J Chervonenkis (1974), “Theory of pattern recognition.”

Viviano, Davide (2019), “Policy targeting under network interference.” *arXiv preprint arXiv:1906.10258*.

Zadrozny, Bianca (2003), “Policy mining: Learning decision policies from fixed sets of data.”

Zhao, Yingqi, Donglin Zeng, A John Rush, and Michael R Kosorok (2012), “Estimating individualized treatment rules using outcome weighted learning.” *Journal of the American Statistical Association*, 107, 1106–1118.

Zhou, Zhengyuan, Susan Athey, and Stefan Wager (2018), “Offline multi-action policy learning: Generalization and optimization.” *arXiv preprint arXiv:1810.04778*.