Optimal Decision Rules Under Partial Identification

Kohei Yata*

November 8, 2021

Abstract

I consider a class of statistical decision problems in which the policy maker must decide between two alternative policies to maximize social welfare (e.g., the population mean of an outcome) based on a finite sample. The central assumption is that the underlying, possibly infinite-dimensional parameter, lies in a known convex set, potentially leading to partial identification of the welfare effect. An example of such restrictions is the smoothness of counterfactual outcome functions. As the main theoretical result, I obtain a finite-sample decision rule (i.e., a function that maps data to a decision) that is optimal under the minimax regret criterion. This rule is easy to compute, yet achieves optimality among all decision rules; no ad hoc restrictions are imposed on the class of decision rules. I apply my results to the problem of whether to change a policy eligibility cutoff in a regression discontinuity setup. I illustrate my approach in an empirical application to the BRIGHT school construction program in Burkina Faso (Kazianga, Levy, Linden and Sloan, 2013), where villages were selected to receive schools based on scores computed from their characteristics. Under reasonable restrictions on the smoothness of the counterfactual outcome function, the optimal decision rule implies that it is not cost-effective to expand the program. I empirically compare the performance of the optimal decision rule with alternative decision rules.

*Department of Economics, Yale University. email: kohei.yata@yale.edu. I am grateful to my advisors Yuichi Kitamura, Timothy Armstrong, and Yusuke Narita for their invaluable advice and encouragement. I also thank Donald Andrews, Toru Kitagawa, Jaewon Lee, Vitor Possebom, Kosuke Uetake, and participants at the Yale Econometrics Prospectus Lunch, Yale Econometrics Seminar, and Happy Hour Seminar for helpful comments and suggestions.

1 Introduction

A fundamental goal of empirical research in economics is to inform policy decisions. Evaluation of counterfactual policies often requires extrapolating from observables to unobservables. Without strong model restrictions such as functional form assumptions or exogeneity of an intervention, the performance of each counterfactual policy may be only partially determined by observed data. In such situations, policy decision-making is challenging since we have no clear understanding of which policy is the best.

For example, a regression discontinuity (RD) design only credibly estimates the impact of treatment on the individuals at the eligibility cutoff. Therefore, without restrictive assumptions such as constant treatment effects, whether or not to offer the treatment to those away from the cutoff is ambiguous. Even randomized controlled trials may provide only partial knowledge of the impact of a new intervention, as can happen if participants do not perfectly comply with their assigned treatment or if the experimental sample is an unrepresentative subset of the target population.

This paper develops an optimal way of using data to make policy decisions when the performance of counterfactual policies is only partially identified. Specifically, I solve a class of statistical decision problems. The setup is as follows. The policy maker must decide between two alternative policies, policy 1 and policy 0, to maximize social welfare. The difference in welfare between policy 1 and policy 0 is given by $L(\theta) \in \mathbb{R}$. $L(\cdot)$ is a linear function of an unknown, possibly infinite-dimensional parameter θ, where θ belongs to a known parameter space Θ. By construction, it is optimal to choose policy 1 if $L(\theta) \geq 0$ and to choose policy 0 if $L(\theta) < 0$. The policy maker makes a decision after observing a finite sample $(Y_1, ..., Y_n) \in \mathbb{R}^n$ whose expected value is given by $(m_1(\theta), ..., m_n(\theta)) \in \mathbb{R}^n$, where $m_i(\cdot)$'s are linear functions of θ.

A leading example of this setup is a choice between two treatment assignment policies based on data generated by nonparametric regression models, including an RD model. A treatment assignment policy specifies who would receive treatment based on an individual's observable covariates. In this example, the parameter θ is a conditional mean function of a counterfactual outcome given covariates and treatment. $m_i(\theta)$ is the conditional mean counterfactual outcome given individual i's observed covariates and treatment. The welfare difference $L(\theta)$ corresponds to the average treatment effect for the subpopulation that would be affected by the switch from the status quo to a new policy. The parameter space Θ is a class of conditional mean counterfactual outcome functions that satisfy, for example, some smoothness restrictions (e.g., bounds on derivatives or the linearity of a function). The welfare difference $L(\theta)$ may or may not be point identified, depending on which function class the policy maker imposes.[1]

As the main theoretical result, I obtain a finite-sample decision rule (i.e., a function that maps the sample $(Y_1, ..., Y_n)$ to a probability of choosing policy 1) that is optimal under the minimax

[1] In Section 2.1, I will discuss what I mean by identification in this finite-sample setup.

regret criterion, a standard criterion used in the literature on statistical treatment choice (e.g., Manski, 2004; Stoye, 2009; Kitagawa and Tetenov, 2018). The minimax regret criterion evaluates decision rules based on the maximum regret, that is, the maximum of the expected amount of welfare lost by choosing the worse policy over the parameter space. The decision rule derived in this paper minimizes the maximum regret over the class of all decision rules. This optimality result holds whether the welfare difference $L(\theta)$ is point or partially identified. To derive the optimality result, I assume that the sample $(Y_1, ..., Y_n)$ is normally distributed with a known variance and that the parameter space Θ is convex and symmetric with respect to the origin, as well as mild regularity conditions.

Importantly, I do not impose any restrictions on the class of decision rules, thus allowing for nonrandomized threshold rules based on a nonlinear function of the sample $(Y_1, ..., Y_n)$ and randomized rules, among others. Solving minimax problems over the class of all decision rules is generally a difficult task. The main tool that I use to solve the minimax regret problem is what is called the *modulus of continuity* (Donoho, 1994). The modulus of continuity at $\epsilon \geq 0$ is the largest possible welfare difference over the parameter space under the constraint that the Euclidean norm of the expected value of the sample $(Y_1, ..., Y_n)$ is at most ϵ, formally defined in Section 3. The minimax problem can be simplified into an optimization problem with respect to the modulus of continuity, which is analytically and computationally tractable.

The resulting decision rule is simple and thus easy to compute. It makes a decision based on a linear function of $(Y_1, ..., Y_n)$. The minimax regret rule may be randomized or nonrandomized, depending on the restrictions imposed on the parameter space. Specifically, it is a nonrandomized rule if the length of the identified set of the welfare difference $L(\theta)$ is short relative to the variance of the sample $(Y_1, ..., Y_n)$, including the case where $L(\theta)$ is point identified. Otherwise, it is a randomized rule, assigning a positive probability both to policies 1 and 0.

When the minimax regret rule is nonrandomized, it can be viewed as a rule that plugs a particular linear estimator of the welfare difference $L(\theta)$ into the optimal decision $\mathbf{1}\{L(\theta) \geq 0\}$. I compare this linear estimator with a linear minimax mean squared error (MSE) estimator of $L(\theta)$, which minimizes the maximum of the MSE over the parameter space within the class of all linear estimators. The two estimators are shown to be generally different, which suggests that the plug-in rule based on the linear minimax MSE estimator is not optimal under the minimax regret criterion. More precisely, the linear estimator used by the minimax regret rule places more importance on the bias than on the variance compared to the linear minimax MSE estimator.

This paper makes new contributions even under point identification in settings with restricted parameter spaces. When the welfare difference $L(\theta)$ is point identified, the minimax regret rule is insensitive to the choice of the restrictions imposed on the parameter space as long as the restrictions are weak enough. For example, consider linear regression models where $m_i(\theta) = x_i'\theta$, $x_i \in \mathbb{R}^k$ is unit i's fixed regressors, and $\theta \in \Theta \subset \mathbb{R}^k$. The minimax regret rule bases decisions on the sign of $L(\hat{\theta})$, where $\hat{\theta}$ is the best linear unbiased estimator of θ, if the parameter space

Θ is sufficiently large (e.g., if $\Theta = \mathbb{R}^k$). When the restrictions on Θ become strong enough, the minimax regret rule starts to use an estimator that optimally trades off the bias and variance.

I then apply my results to the problem of eligibility cutoff choice in an RD setup. In many policy domains, the eligibility for treatment is determined based on an individual's observable characteristics. One crucial policy question is whether we should change the eligibility criterion to achieve better outcomes (Dong and Lewbel, 2015). Specifically, I consider an RD setup and study the problem of whether or not to change the eligibility cutoff from a current value c_0 to a new value c_1. For an illustration of the results, I focus on the case where the new value is smaller than the current one (i.e., $c_1 < c_0$) and the conditional mean counterfactual outcome function belongs to the class of Lipschitz functions with a known Lipschitz constant C. The absolute value of the derivative of any differentiable function in this function class is bounded above by C. Under the Lipschitz constraint, the effect of the cutoff change on the population mean outcome is partially identified.

A closed-form expression for the minimax regret rule can be obtained in this application when the Lipschitz constant C is large enough. In such cases, the minimax regret rule is based on the mean outcome difference between the treated unit closest to the status quo cutoff c_0 and the untreated units between the two cutoffs c_0 and c_1. On the other hand, when C is not sufficiently large, the minimax regret rule may also use outcomes of other units, although it does not generally admit a closed form. I provide a simple procedure to numerically compute it for any choice of C.

Implementation of the minimax regret rule requires choosing the Lipschitz constant C. In principle, it is not possible to choose the Lipschitz constant C that applies to both sides of the status quo cutoff c_0 in a data-driven way since we only observe outcomes either under treatment or under no treatment on each side. It is, however, possible to estimate a lower bound on C. In practice, I recommend considering a range of plausible choices of C, including the estimated lower bound, to conduct a sensitivity analysis.

Finally, I illustrate my approach in an empirical application to the Burkinabé Response to Improve Girls' Chances to Succeed (BRIGHT) program, a school construction program in Burkina Faso (Kazianga *et al.*, 2013). Aiming to improve educational outcomes in rural villages, the program constructed primary schools in 132 villages from 2005 to 2008. To allocate schools, the Ministry of Education first computed a score summarizing village characteristics for each of the nominated 293 villages and then selected the highest-ranking villages to receive a school. This situation fits into an RD setup.

I ask whether we should expand this program or not. The more specific question considered in this analysis is whether or not to construct schools in the top 20% of previously ineligible villages. The analysis uses the enrollment rate as the welfare measure and assumes that the conditional mean counterfactual outcome function belongs to the class of Lipschitz functions with a known Lipschitz constant C. To consider policy costs, I assume that implementing the policy is optimal if it is better in terms of cost-effectiveness than a similar policy, whose cost-effectiveness is available from external studies. Given available estimates of the new policy costs, my approach can be used to consider this decision problem. For a plausible range of the Lipschitz constant C, the minimax regret rule implies that building schools in the top 20% of previously ineligible villages is not cost-effective.

I empirically compare the minimax regret rule with plug-in decision rules that make a decision according to the sign of a policy effect's estimator. The performance of the minimax regret rule is shown to be relatively robust to misspecification of the Lipschitz constant C toward zero, which suggests that the potential loss due to an optimistic choice of C may not be a major concern.

My approach is applicable to many other policy choice problems. One example is the problem of deciding whether to introduce a new policy based on data from a randomized experiment when the experiment has imperfect compliance or when the experimental sample is a selected subset of the target population.

1.1 Related Literature

This paper contributes to the literature on statistical treatment choice, which has been growing in econometrics since the work by Manski (2000, 2004). The literature has intensively studied optimal treatment assignment based on covariates in settings where social welfare under each policy is point identified (Manski, 2004; Dehejia, 2005; Hirano and Porter, 2009; Stoye, 2009, 2012; Bhattacharya and Dupas, 2012; Kitagawa and Tetenov, 2018, 2021; Athey and Wager, 2021; Mbakop and Tabord-Meehan, 2021).[2] In contrast, my approach can be applied to this problem even with partial identification if the choice set consists of two treatment assignment policies. Additionally, while many of the above papers provide finite-sample regret bounds or asymptotic optimality results, I derive a finite-sample optimality result.

This paper is more closely related to the area of treatment choice under partial identification. In particular, Stoye (2012) and Ishihara and Kitagawa (2021) consider binary treatment choice problems under Gaussian models and derive minimax regret rules. Stoye (2012) provides a special case of my result in a setting where the experiment has imperfect internal or external validity. Ishihara and Kitagawa (2021) consider the problem of deciding whether or not to introduce a new policy to a specific local population based on causal evidence of similar policies implemented in other populations. While they derive a minimax regret rule within the class of plug-in rules based on a linear function of the sample, I derive a minimax regret rule within the class of all decision rules. Other papers studying optimal policy under partial identification include Manski (2007, 2009, 2010, 2011a,b, 2020), Kasy (2016, 2018), Mo, Qi and Liu (2021), Russell (2020), Christensen, Moon and Schorfheide (2020), and Kallus and Zhou (2021) among others.

The Gaussian model used in this paper has been studied for the problem of optimal estimation

[2]This problem has also been actively studied in statistics and machine learning. A partial list includes Qian and Murphy (2011), Zhao, Zeng, Rush and Kosorok (2012), Swaminathan and Joachims (2015), and Kallus (2018).

and inference in nonparametric regression models. Donoho (1994) uses the modulus of continuity to characterize minimax optimal estimators and confidence intervals on linear functionals of a regression function. The derivation of my result and that of Donoho (1994)'s consist of similar steps. However, the proof of each step is nontrivially different since the problem of policy choice and that of estimation and inference are not nested by each other; specifically, the loss function and the action space are different. Recent work on estimation and inference using Donoho (1994)'s framework includes Armstrong and Kolesár (2018), Imbens and Wager (2019), Rambachan and Roth (2020), Armstrong and Kolesár (2021), and de Chaisemartin (2021).

In terms of an application to eligibility cutoff choice, this paper is also related to the growing literature on extrapolation away from the cutoff in RD designs, including Rokkanen (2015), Angrist and Rokkanen (2015), Dong and Lewbel (2015), Bertanha and Imbens (2020), Bertanha (2020), Bennett (2020), and Cattaneo, Keele, Titiunik and Vazquez-Bare (2020). Unlike these papers, I explicitly consider the decision problem of whether or not to change the cutoff and derive an optimal decision rule. To the best of my knowledge, there are no existing results for optimal policy decisions based on data generated by an RD model.

2 Setup, Optimality Criterion, and Motivating Example

In this section, I set up the policy maker's problem of deciding between two alternative policies. This setup allows for the case where the welfare difference between the two policies is only partially identified. I then introduce the minimax regret criterion to evaluate different procedures for using data to make decisions. To illustrate my framework and its applicability, I present the problem of eligibility cutoff choice in an RD setup as an example.

2.1 Setup

Data-generating Model. Suppose that the policy maker observes a sample $\boldsymbol{Y} = (Y_1, ..., Y_n)' \in \mathbb{R}^n$ of the form

$$\boldsymbol{Y} \sim \mathcal{N}(\boldsymbol{m}(\theta), \boldsymbol{\Sigma}), \tag{1}$$

where θ is an unknown parameter that lies in a known subset Θ of a vector space $\mathbb{V}$, $\boldsymbol{m} : \mathbb{V} \to \mathbb{R}^n$ is a known linear function, and $\boldsymbol{\Sigma}$ is a known, positive-definite $n \times n$ matrix.[3] I allow θ to be an infinite-dimensional parameter such as functions.

The linearity of $\boldsymbol{m}$ is not necessarily restrictive. If we specify θ so that it contains each of the expected values of $Y_1, ..., Y_n$ as its element, $\boldsymbol{m}$ is a function that extracts those expected values from θ, which is linear in θ.

[3] Donoho (1994), Low (1995), and Armstrong and Kolesár (2018) investigate optimal estimation and inference of a linear functional of θ in a slightly more general version of this model that allows $\boldsymbol{Y}$ to be infinite dimensional.

This model allows the expected value of $\boldsymbol{Y}$ to depend on other observed variables such as covariates and treatment by treating them as fixed and subsuming them into $\boldsymbol{m}$ and $\boldsymbol{\Sigma}$. For example, a regression model with fixed regressors

$$Y_i = f(x_i) + u_i, \quad u_i \sim \mathcal{N}(0, \sigma^2(x_i)) \text{ independent across } i$$

is a special case where $\boldsymbol{Y} = (Y_1, ..., Y_n)'$, $\theta = f$, Θ is a class of functions, $\boldsymbol{m}(f) = (f(x_1), ..., f(x_n))'$, and $\boldsymbol{\Sigma} = \text{diag}(\sigma^2(x_1), ..., \sigma^2(x_n))$.

The normality of $\boldsymbol{Y}$ and the assumption of known variance are restrictive, but are often imposed to deliver finite-sample optimality results for problems of estimation, inference, and treatment choice. In some cases, it is plausible to assume the normality of $\boldsymbol{Y}$. For example, suppose that unit i represents a group of individuals defined by place, time, and individual characteristics among others and that $Y_1, ..., Y_n$ are group-level mean outcomes. If the number of groups is fixed at n, the distribution of $\boldsymbol{Y}$ approaches a normal distribution as the size of each group grows to infinity by the central limit theorem. The normal model (1) can be viewed as an asymptotic approximation if each group is large enough.[4]

I assume that the parameter space Θ is convex and centrosymmetric (i.e., $\theta \in \Theta$ implies $-\theta \in \Theta$) throughout the paper. Typical parameter spaces considered in empirical analyses are convex. For example, in the regression model above, classes of functions with bounded derivatives (e.g., the class of Lipschitz functions with a known Lipschitz constant) are convex. The centrosymmetry simplifies the analysis, but rules out some shape restrictions. In the regression model above, Θ fails to be centrosymmetric if we assume the convexity or concavity of the regression function.[5]

Policy Choice Problem. Now, suppose that the policy maker is interested in choosing between two alternative policies, policy 1 and policy 0, to maximize social welfare. The class of binary policy decisions includes, for example, whether to introduce a program to a target population and whether to change a policy from the status quo to a new one. Suppose that the welfare resulting from implementing policy $a \in \{0, 1\}$ under θ is $W_a(\theta)$, where $W_a : \mathbb{V} \to \mathbb{R}$ is a known function specified by the policy maker. The welfare difference between policy 1 and

[4]More generally, suppose that the policy maker observes an n-dimensional vector of statistics of the original data and that it is an asymptotically normal estimator of its population counterpart. For example, the mean outcome difference between the treatment and control groups in a randomized experiment is a statistic that is asymptotically normal for the population mean difference. If we regard the n-dimensional vector of statistics as $\boldsymbol{Y}$, the normal model (1) can again be viewed as an asymptotic approximation (Stoye, 2012; Tetenov, 2012; Rambachan and Roth, 2020; Andrews, Kitagawa and McCloskey, 2021; Ishihara and Kitagawa, 2021).

[5]On the other hand, in some cases, it is possible to impose the monotonicity of the regression function by normalizing the sample $\boldsymbol{Y}$ so that the new parameter space is centrosymmetric. Suppose, for example, that $\Theta = \{f \in \mathcal{F}_{\text{Lip}}(C) : f(x) \text{ is nondecreasing in } x\}$, where $\mathcal{F}_{\text{Lip}}(C) = \{f : |f(x) - f(\tilde{x})| \leq C|x - \tilde{x}| \text{ for every } x, \tilde{x} \in \mathbb{R}\}$. $\mathcal{F}_{\text{Lip}}(C)$ is centrosymmetric while Θ is not. It is easy to show that $\Theta = \{\tilde{f} + f_0 : \tilde{f} \in \mathcal{F}_{\text{Lip}}(C/2)\}$, where $f_0(x) = \frac{C}{2}x$ for all $x \in \mathbb{R}$. Therefore, the model $\boldsymbol{Y} \sim \mathcal{N}(\boldsymbol{m}(f), \boldsymbol{\Sigma})$, $f \in \Theta$, is equivalent to the model $\tilde{\boldsymbol{Y}} \sim \mathcal{N}(\boldsymbol{m}(\tilde{f}), \boldsymbol{\Sigma})$, $\tilde{f} \in \mathcal{F}_{\text{Lip}}(C/2)$, where $\tilde{\boldsymbol{Y}} = \boldsymbol{Y} - \boldsymbol{m}(f_0) = (Y_1 - f_0(x_1), ..., Y_n - f_0(x_n))'$; the set of distributions of $\boldsymbol{Y}$ over $f \in \Theta$ is identical to the set of distributions of $\tilde{\boldsymbol{Y}} + \boldsymbol{m}(f_0)$ over $\tilde{f} \in \mathcal{F}_{\text{Lip}}(C/2)$.

policy 0 is given by

$$L(\theta) \coloneqq W_1(\theta) - W_0(\theta).$$

I assume that $L : \mathbb{V} \to \mathbb{R}$ is a linear function. The optimal policy under θ is policy 1 if $L(\theta) > 0$, policy 0 if $L(\theta) < 0$, and either of the two if $L(\theta) = 0$.

One example of a welfare criterion is a weighted average of an outcome across individuals. For example, suppose that a policy could change the outcome of each individual in the population. Suppose also that we specify $\theta = (f_1(\cdot), f_0(\cdot))$, where $f_a(x)$ represents the counterfactual mean outcome under policy a across individuals whose observed covariates are x. The welfare under policy a can be defined, for example, by the population mean outcome $W_a(\theta) = \int f_a(x)dP_X$, where P_X is the probability measure of covariates in the population and is assumed to be known. In this case, the welfare difference $L(\theta) = \int [f_1(x) - f_0(x)]dP_X$ is linear in $\theta = (f_1(\cdot), f_0(\cdot))$. If we are required to take the policy cost into account, we can incorporate it into the welfare by redefining the outcome to be the raw outcome minus the cost. On the other hand, the linearity of L may rule out welfare criteria that depend on the distribution of the counterfactual outcome.[6]

Importantly, this framework allows for cases in which $L(\theta)$ is not point identified in the sense that the identified set of $L(\theta)$ when $\boldsymbol{m}(\theta) = \boldsymbol{\mu}$, namely

$$\{L(\theta) : \boldsymbol{m}(\theta) = \boldsymbol{\mu}, \theta \in \Theta\},$$

is nonsingleton for some or all $\boldsymbol{\mu} \in \mathbb{R}^n$. This is the set of possible values of $L(\theta)$ consistent with the observed value of $\boldsymbol{Y} \in \mathbb{R}^n$ when there is no sampling uncertainty. If the identified set contains both positive and negative values, which policy we should choose is ambiguous even without sampling uncertainty. Whether $L(\theta)$ is point identified or not depends on the parameter space Θ.

This framework nests some existing setups of treatment choice, such as limit experiments under parametric models by Hirano and Porter (2009), Gaussian experiments with limited validity by Stoye (2012), and a setup of policy choice based on multiple studies by Ishihara and Kitagawa (2021).[7] One of the essential departures from these setups is that the parameter θ can be infinite dimensional, accommodating nonparametric regression models, for example.[8]

[6]One example of such a criterion is $\int_0^1 w(\tau)F^{-1}(\tau; f_a)d\tau$. Here, $w(\cdot)$ is a known weight function, $F(\cdot; f_a)$ is the distribution of the counterfactual outcome under policy a induced by the normal distribution with the conditional mean function f_a, and $F^{-1}(\tau; f_a)$ is the τ-th quantile of the distribution.

[7]Ishihara and Kitagawa (2021) do not impose the convexity of the parameter space to derive their results. However, the specific examples of the parameter space that they consider satisfy the convexity.

[8]Hirano and Porter (2009) consider a model with an infinite Gaussian sequence as a limit experiment under semiparametric models. They do not allow for a partially identified welfare difference—one of the crucial aspects of this paper's setup.

2.2 Optimality Criterion

What is the optimal procedure for using the sample $\boldsymbol{Y}$ to make a policy choice? This paper considers the minimax regret criterion as an optimality criterion, following existing treatment choice studies (e.g., Manski, 2004, 2007; Stoye, 2009, 2012; Kitagawa and Tetenov, 2018).[9]

I define a few concepts to introduce the minimax regret criterion. A *decision rule* is a measurable function $\delta : \mathbb{R}^n \to [0,1]$, where $\delta(\boldsymbol{y})$ represents the probability of choosing policy 1 when the realization of the sample $\boldsymbol{Y}$ is $\boldsymbol{y}$. The *welfare regret loss* for policy choice $a \in \{0,1\}$ is

$$l(a,\theta) \coloneqq \max_{a' \in \{0,1\}} W_{a'}(\theta) - W_a(\theta) = \begin{cases} L(\theta) \cdot (1-a) & \text{if } L(\theta) \geq 0, \\ -L(\theta) \cdot a & \text{if } L(\theta) < 0. \end{cases}$$

The welfare regret loss $l(a,\theta)$ is the difference between the welfare under the optimal policy and the welfare under policy a under θ. If the policy maker chooses the superior policy, they do not incur any loss; otherwise, they incur a loss of the absolute value of the welfare difference $L(\theta)$. The risk or *regret* of decision rule δ under θ is the expected welfare regret loss

$$R(\delta,\theta) \coloneqq \begin{cases} L(\theta)(1 - \mathbb{E}_\theta[\delta(\boldsymbol{Y})]) & \text{if } L(\theta) \geq 0, \\ -L(\theta)\mathbb{E}_\theta[\delta(\boldsymbol{Y})] & \text{if } L(\theta) < 0, \end{cases}$$

where $\mathbb{E}_\theta$ denotes the expectation taken with respect to $\boldsymbol{Y}$ under θ.

Given a particular choice of Θ, I evaluate decision rules based on the maximum regret over Θ, $\sup_{\theta \in \Theta} R(\delta,\theta)$. My goal is to derive a *minimax regret* decision rule, which achieves

$$\inf_\delta \sup_{\theta \in \Theta} R(\delta,\theta),$$

where the infimum is taken over the set of all possible decision rules. I do not impose any restrictions on the class of decision rules.

To sum up, the minimax regret criterion deals with the sampling uncertainty given θ by taking the expectation of the welfare regret loss with respect to the distribution of $\boldsymbol{Y}$. It then deals with the parameter θ by considering the worst-case expected welfare regret loss. It does not distinguish between the case where the welfare difference $L(\theta)$ is point identified and the case where it is not. Nevertheless, as I show in Section 3, the minimax regret rule behaves differently in each case.

[9] Alternative criteria include the maximin criterion, which solves $\sup_\delta \inf_{\theta \in \Theta} U(\delta,\theta)$, where $U(\delta,\theta) = W_1(\theta)\mathbb{E}_\theta[\delta(\boldsymbol{Y})] + W_0(\theta)(1 - \mathbb{E}_\theta[\delta(\boldsymbol{Y})])$ is the expected welfare under decision rule δ under θ. It has been pointed out that the maximin criterion is unreasonably pessimistic and can lead to pathological decision rules (Savage, 1951; Manski, 2004). Another approach is the Bayesian one, which solves $\sup_\delta \int U(\delta,\theta) d\pi(\theta)$, where π is a prior on the vector space $\mathbb{V}$ that θ belongs to. In practice, when it is difficult to make a prior, the minimax regret criterion is a reasonable choice.

2.3 Motivating Example: Eligibility Cutoff Choice in Regression Discontinuity Designs

In many policy domains, ranging from health to education to social programs, the eligibility for treatment is determined based on an individual's observable characteristics. A critical policy question is whether we should change the eligibility criterion to achieve better welfare (Dong and Lewbel, 2015).[10] My framework can be of use for policy makers interested in utilizing data to make such decisions.

Consider the following RD setup. For each unit $i = 1, .., n$, we observe a fixed running variable $x_i \in \mathbb{R}$, a binary treatment status $d_i \in \{0,1\}$, and an outcome $Y_i \in \mathbb{R}$. The eligibility for treatment is determined based on whether the running variable exceeds a specific cutoff $c_0 \in \mathbb{R}$, so that $d_i = \mathbf{1}\{x_i \geq c_0\}$. Suppose that the outcome Y_i is of the form

$$Y_i = f(x_i, d_i) + u_i, \quad u_i \sim \mathcal{N}(0, \sigma^2(x_i, d_i)) \text{ independent across } i, \tag{2}$$

where $f : \mathbb{R} \times \{0,1\} \to \mathbb{R}$ is an unknown function and the conditional variance $\sigma^2(x_i, d_i)$ is known for $i = 1, ..., n$.[11] We interpret $f(x, d)$ as the counterfactual mean outcome across individuals with running variable x if their treatment status is set to $d \in \{0,1\}$.[12] We can write the model in a vector form:

$$\boldsymbol{Y} \sim \mathcal{N}(\boldsymbol{m}(f), \boldsymbol{\Sigma}),$$

where $\boldsymbol{Y} = (Y_1, ..., Y_n)'$, $\boldsymbol{m}(f) = (f(x_1, d_1), ..., f(x_n, d_n))'$, and $\boldsymbol{\Sigma} = \text{diag}(\sigma^2(x_1, d_1), ..., \sigma^2(x_n, d_n))$. Here, f plays the role of the unknown parameter θ.

Now, suppose that we are interested in changing the eligibility cutoff from c_0 to a specific value c_1. For illustration purposes, I assume $c_1 < c_0$. Suppose that the welfare under the cutoff c_a, with $a \in \{0,1\}$, is an average of the counterfactual mean outcome across different values of the running variable

$$W_a(f) = \int [f(x,1)\mathbf{1}\{x \geq c_a\} + f(x,0)\mathbf{1}\{x < c_a\}] d\nu(x)$$

for some known measure ν. One choice of ν is an empirical measure, for which the welfare is the unweighted sample average: $W_a(f) = \frac{1}{n}\sum_{i=1}^{n}[f(x_i,1)\mathbf{1}\{x_i \geq c_a\} + f(x_i,0)\mathbf{1}\{x_i < c_a\}]$. The

[10] For example, there is a heated debate about whether to extend Medicare eligibility in the United States (Song, 2020).

[11] I make this assumption to deliver finite-sample optimality results. In practice, one replaces the true conditional variances with their consistent estimators. See Section 4.1 for possible estimators.

[12] This interpretation is established in a potential outcome model as follows (Armstrong and Kolesár, 2021). Suppose we observe a triple of the outcome, treatment status, and running variable (Y_i, D_i, X_i). The observed outcome is $Y_i = Y_i(1)D_i + Y_i(0)(1 - D_i)$, where $Y_i(1)$ and $Y_i(0)$ are potential outcomes under treatment and no treatment, respectively. Let $f(x, d) = \mathbb{E}[Y_i(d)|X_i = x]$, which is equal to $\mathbb{E}[Y_i|X_i = x, D_i = d]$ if D_i is a deterministic function of X_i. We obtain model (2) by conditioning on the realized values $\{(x_i, d_i)\}_{i=1}^{n}$ and assuming normal conditional errors.

welfare difference between the two cutoffs is

$$L(f) = W_1(f) - W_0(f) = \int \mathbf{1}\{c_1 \leq x < c_0\}[f(x,1) - f(x,0)]d\nu(x),$$

which is a linear function of f. $L(f)$ is a weighted sum of the conditional average treatment effect $f(x,1) - f(x,0)$ across different values of the running variable between the two cutoffs c_1 and c_0.

To conclude the problem's setup, suppose that $f \in \mathcal{F}$, where $\mathcal{F}$ is a known set of functions and plays the role of the parameter space Θ. For an illustration of the results and empirical application, I focus on the *Lipschitz class* with a known Lipschitz constant $C \geq 0$:

$$\mathcal{F}_{\text{Lip}}(C) = \{f : |f(x,d) - f(\tilde{x},d)| \leq C|x - \tilde{x}| \text{ for every } x, \tilde{x} \in \mathbb{R} \text{ and } d \in \{0,1\}\}.$$

The Lipschitz constraint bounds the maximum possible change in $f(x,d)$ in response to a shift in x by one unit. In other words, the absolute value of the derivative of $f(x,d)$ with respect to x must be at most C if f is differentiable. The Lipschitz class $\mathcal{F}_{\text{Lip}}(C)$ is both convex and centrosymmetric.

Imposing $f \in \mathcal{F}_{\text{Lip}}(C)$ is not strong enough to uniquely determine $L(f)$ from a given value of $\boldsymbol{m}(f) = (f(x_1,d_1), ..., f(x_n,d_n))'$. Nevertheless, it produces an informative identified set of $L(f)$ since it gives finite upper and lower bounds on $f(x,d)$ for every $(x,d) \in \mathbb{R} \times \{0,1\}$ from the knowledge of $(f(x_1,d_1), ..., f(x_n,d_n))$.[13]

In Section 4, I derive a minimax regret rule for this example when the welfare is the sample average outcome and $\mathcal{F} = \mathcal{F}_{\text{Lip}}(C)$.

This example can be easily generalized to a setup where the observed treatment is independent of counterfactual outcomes conditional on multidimensional covariates (i.e., the unconfoundedness assumption holds) and there is no or limited overlap in the covariate distribution between the treatment and control groups. This general setup covers the problem of whether to change an eligibility criterion based on multiple covariates. Limited overlap may also occur, for example, if the policy maker wishes to consider whether to introduce a policy using a composite dataset of treated units from a local randomized experiment and nonexperimental comparison units from national surveys (LaLonde, 1986; Dehejia and Wahba, 1999). See Appendix A.1 for details on the general setup.

3 Main Result

In this section, I derive a minimax regret rule among all possible decision rules and then discuss its interpretations and implications. For simplicity, I normalize $\boldsymbol{\Sigma} = \sigma^2 \boldsymbol{I}_n$ for some $\sigma > 0$, where

[13] Given a value of $(f(x_1,d_1), ..., f(x_n,d_n))$, the upper bound on $f(x,d)$ is $\min_{i:d_i=d}(f(x_i,d) + C|x_i - x|)$. The lower bound on $f(x,d)$ is $\max_{i:d_i=d}(f(x_i,d) - C|x_i - x|)$.

$\boldsymbol{I}_n$ is the identity matrix.[14]

3.1 Modulus of Continuity

Solving minimax problems over the class of all decision rules is generally a difficult task. The main tool that I use to solve the minimax regret problem is the *modulus of continuity*, defined as

$$\omega(\epsilon; L, \boldsymbol{m}, \Theta) := \sup\{L(\theta) : \|\boldsymbol{m}(\theta)\| \le \epsilon, \theta \in \Theta\}, \quad \epsilon \ge 0,$$

where $\|\cdot\|$ is the Euclidean norm. The modulus of continuity and its variants have been used in constructing minimax optimal estimators and confidence intervals on linear functionals in Gaussian models (Donoho, 1994; Low, 1995; Cai and Low, 2004; Armstrong and Kolesár, 2018).[15] It has not been used in deriving minimax regret rules for the problem of treatment choice.

By definition, $\omega(\epsilon; L, \boldsymbol{m}, \Theta)$ is nonnegative and nondecreasing in ϵ. Furthermore, $\omega(\epsilon; L, \boldsymbol{m}, \Theta)$ is concave in ϵ if Θ is convex.[16] I say that $\theta_\epsilon \in \Theta$ *attains the modulus of continuity at* ϵ if $L(\theta_\epsilon) = \omega(\epsilon; L, \boldsymbol{m}, \Theta)$ and $\|\boldsymbol{m}(\theta_\epsilon)\| \le \epsilon$, namely if $\theta_\epsilon \in \arg\max_{\theta\in\Theta} L(\theta)$ *s.t.* $\|\boldsymbol{m}(\theta)\| \le \epsilon$. Below, I suppress the arguments $L, \boldsymbol{m}$, and Θ if they are clear from the context.

In the context of this paper, the modulus of continuity at ϵ is the largest possible welfare difference under the constraint that the norm of $\boldsymbol{m}(\theta)$, namely the expected value of $\boldsymbol{Y}$, is less than or equal to ϵ. When $\epsilon = 0$ and hence the expected value of $\boldsymbol{Y}$ must be a vector of zeros, the sample $\boldsymbol{Y}$ is uninformative. When the norm constraint $\|\boldsymbol{m}(\theta)\| \le \epsilon$ is relaxed, the strength of $\boldsymbol{Y}$ as a signal for $L(\theta)$ may increase, which makes it easier for the policy maker to detect the optimal policy. At the same time, the largest potential welfare loss when choosing the inferior policy may increase since the flexibility of θ increases because of the weaker norm constraint. The modulus of continuity is used to trade off these two and find parameter values that are least favorable for the policy maker.

Here, I briefly formalize the above argument, deferring the statement of the necessary assumptions and the complete proof and discussion to Section 3.2 and Section 6, respectively.

I introduce some notation. I use $\mathcal{R}(\sigma; \Theta)$ to denote the *minimax risk* $\inf_\delta \sup_{\theta\in\Theta} R(\delta, \theta)$, which may depend on the standard deviation σ and on the choice of the parameter space Θ among others. Given any two parameter values $\tilde{\theta}, \bar{\theta} \in \mathbb{V}$, where $\mathbb{V}$ is the vector space that the parameter θ belongs to, I define a *one-dimensional subproblem* as the set of all convex combinations of $\tilde{\theta}$ and $\bar{\theta}$, denoted by $[\tilde{\theta}, \bar{\theta}] = \{(1-\lambda)\tilde{\theta} + \lambda\bar{\theta} : \lambda \in [0,1]\}$. Let $\mathcal{R}(\sigma; [\tilde{\theta}, \bar{\theta}])$ denote the minimax risk $\inf_\delta \sup_{\theta\in[\tilde{\theta},\bar{\theta}]} R(\delta, \theta)$ for the one-dimensional subproblem $[\tilde{\theta}, \bar{\theta}]$. Additionally, let Φ and ϕ denote the cumulative distribution function and the probability density function,

[14] This normalization is without loss of information in the following sense. If $\boldsymbol{\Sigma}$ is known, observing $\boldsymbol{Y} \sim \mathcal{N}(\boldsymbol{m}(\theta), \boldsymbol{\Sigma})$ is equivalent to observing $\tilde{\boldsymbol{Y}} \sim \mathcal{N}(\tilde{\boldsymbol{m}}(\theta), \sigma^2 \boldsymbol{I}_n)$ for any $\sigma > 0$, where $\tilde{\boldsymbol{Y}} = \sigma\boldsymbol{\Sigma}^{-1/2}\boldsymbol{Y}$ and $\tilde{\boldsymbol{m}}(\theta) = \sigma\boldsymbol{\Sigma}^{-1/2}\boldsymbol{m}(\theta)$.

[15] Donoho (1994) defines the modulus of continuity as $\tilde{\omega}(\epsilon) = \sup\{|L(\theta) - L(\tilde{\theta})| : \|\boldsymbol{m}(\theta - \tilde{\theta})\| \le \epsilon, \theta, \tilde{\theta} \in \Theta\}$. If Θ is convex and centrosymmetric, the relationship $\tilde{\omega}(\epsilon) = 2\omega(\epsilon/2)$ holds.

[16] See, for example, Donoho (1994, Lemma 3) and Armstrong and Kolesár (2018, Appendix A).

respectively, of a standard normal variable. Lastly, let $a^* \in \arg\max_{a\geq 0} a\Phi(-a)$, which is shown to be unique by Lemma B.1 in Appendix B.1.

Below, I first use the modulus of continuity to characterize the hardest one-dimensional subproblem of the form $[-\bar{\theta}, \bar{\theta}]$ with $\bar{\theta} \in \Theta$, namely the one that has the largest minimax risk $\mathcal{R}(\sigma; [-\bar{\theta}, \bar{\theta}])$ among all one-dimensional subproblems of this form. I then explain that the hardest one-dimensional subproblem is as hard as the original problem with the whole parameter space Θ, suggesting that the hardest one-dimensional subproblem consists of the least favorable parameter values in the original problem.

As shown in Lemmas 3 and 6 in Section 6, the minimax risk for the one-dimensional subproblem $[-\bar{\theta}, \bar{\theta}]$ with $L(\bar{\theta}) \geq 0$ is given by

$$\mathcal{R}(\sigma; [-\bar{\theta}, \bar{\theta}]) = \begin{cases} L(\bar{\theta})\Phi\left(-\frac{\|\boldsymbol{m}(\bar{\theta})\|}{\sigma}\right) & \text{if } \|\boldsymbol{m}(\bar{\theta})\| \leq a^*\sigma, \\ a^*\sigma\frac{L(\bar{\theta})}{\|\boldsymbol{m}(\bar{\theta})\|}\Phi\left(-a^*\right) & \text{if } \|\boldsymbol{m}(\bar{\theta})\| > a^*\sigma. \end{cases}$$

For example, if $L(\bar{\theta}) \geq 0$ and $\|\boldsymbol{m}(\bar{\theta})\| > 0$, the decision rule $\bar{\delta}(\boldsymbol{Y}) = \mathbf{1}\{\boldsymbol{m}(\bar{\theta})'\boldsymbol{Y} \geq 0\}$ is shown to be minimax regret for the subproblem $[-\bar{\theta}, \bar{\theta}]$. Computing the maximum regret $\sup_{\theta\in[-\bar{\theta},\bar{\theta}]} R(\bar{\delta}, \theta)$ yields the above display.

For simplicity, I assume only here that for each $\epsilon \geq 0$, there exists a value of θ that attains the modulus of continuity at ϵ with $\|\boldsymbol{m}(\theta)\| = \epsilon$. The minimax risk for the hardest one-dimensional subproblem can then be expressed in terms of the modulus of continuity:

$$\begin{aligned}
&\sup_{\bar{\theta}\in\Theta} \mathcal{R}(\sigma; [-\bar{\theta}, \bar{\theta}]) = \sup_{\epsilon\geq 0} \sup_{\bar{\theta}\in\Theta:\|\boldsymbol{m}(\bar{\theta})\|=\epsilon, L(\bar{\theta})\geq 0} \mathcal{R}(\sigma; [-\bar{\theta}, \bar{\theta}]) \\
&= \sup\left\{\sup_{\epsilon\in[0,a^*\sigma]} \sup_{\bar{\theta}\in\Theta:\|\boldsymbol{m}(\bar{\theta})\|=\epsilon} L(\bar{\theta})\Phi\left(-\frac{\|\boldsymbol{m}(\bar{\theta})\|}{\sigma}\right), \sup_{\epsilon>a^*\sigma} \sup_{\bar{\theta}\in\Theta:\|\boldsymbol{m}(\bar{\theta})\|=\epsilon} a^*\sigma\frac{L(\bar{\theta})}{\|\boldsymbol{m}(\bar{\theta})\|}\Phi\left(-a^*\right)\right\} \\
&= \sup\left\{\sup_{\epsilon\in[0,a^*\sigma]} \omega(\epsilon)\Phi(-\epsilon/\sigma), \sup_{\epsilon>a^*\sigma} a^*\sigma\frac{\omega(\epsilon)}{\epsilon}\Phi(-a^*)\right\},
\end{aligned}$$

where the first equality holds since restricting attention to $\bar{\theta}$ with $L(\bar{\theta}) \geq 0$ does not change the supremum by the centrosymmetry of Θ and the last equality follows from the definition of the modulus of continuity. Furthermore, since $\frac{\omega(\epsilon)}{\epsilon}$ is shown to be nonincreasing, $\sup_{\epsilon>a^*\sigma} a^*\sigma\frac{\omega(\epsilon)}{\epsilon}\Phi(-a^*) = \omega(a^*\sigma)\Phi(-a^*)$. The above expression can then be simplified into:

$$\sup_{\bar{\theta}\in\Theta} \mathcal{R}(\sigma; [-\bar{\theta}, \bar{\theta}]) = \sup_{\epsilon\in[0,a^*\sigma]} \omega(\epsilon)\Phi(-\epsilon/\sigma).$$

Now, let ϵ^* solve the maximization problem on the right-hand side. ϵ^* balances the potential welfare loss ($\omega(\epsilon)$) and the probability of incurring loss ($\Phi\left(-\epsilon/\sigma\right)$). The corresponding subproblem $[-\theta_{\epsilon^*}, \theta_{\epsilon^*}]$ has the largest minimax risk among all one-dimensional subproblems, where θ_{ϵ^*} attains the modulus of continuity at ϵ^*.

It turns out that the subproblem $[-\theta_{\epsilon^*}, \theta_{\epsilon^*}]$ is as hard as the original problem with the whole parameter space Θ. That is, the two problems have the same minimax risk:

$$\mathcal{R}(\sigma; [-\theta_{\epsilon^*}, \theta_{\epsilon^*}]) = \mathcal{R}(\sigma; \Theta),$$

as shown in Section 6. Thus, θ_{ϵ^*} and $-\theta_{\epsilon^*}$ are the least favorable parameter values for the policy maker.

In the next section, I derive a minimax regret rule. The rule protects against these worst cases, as I discuss in Sections 3.3 and 3.4.

3.2 Minimax Regret Rules

I now present a minimax regret rule. To derive the result, I impose the following restrictions on L, $\boldsymbol{m}$, and Θ. For expositional purposes, I assume that $a^*\sigma < \sup_{\theta\in\Theta} \|\boldsymbol{m}(\theta)\|$.[17]

Assumption 1 (Regularity)**.** *The following holds for some $\bar{\epsilon} > 0$.*

(a) For all $\epsilon \in [0, \bar{\epsilon}]$, there exists $\theta_\epsilon \in \Theta$ that attains the modulus of continuity at ϵ.

(b) There exists $\boldsymbol{w}^* \in \mathbb{R}^n$ such that $\lim_{\epsilon\to 0} \epsilon^{-1}\left(\boldsymbol{w}^* - \frac{\boldsymbol{m}(\theta_\epsilon)}{\|\boldsymbol{m}(\theta_\epsilon)\|}\right) = \mathbf{0}$.

(c) For all $\epsilon \in [0, \bar{\epsilon}]$, there exists $\iota \in \Theta$ such that $L(\iota) \neq 0$ and $\theta_\epsilon + c\iota \in \Theta$ for all c in a neighborhood of zero.

(d) $\omega(\cdot)$ is differentiable at any $\epsilon \in (0, a^*\sigma]$. Furthermore, $\rho(\cdot)$ is differentiable at any $\epsilon \in (\epsilon_1, \epsilon_2)$, where $\rho(\epsilon) = \sup\{L(\theta) : (\boldsymbol{w}^*)'\boldsymbol{m}(\theta) = \epsilon, \theta \in \Theta\}$ for $\epsilon \in \mathbb{R}$, $\epsilon_1 = \inf\{(\boldsymbol{w}^*)'\boldsymbol{m}(\theta) : \theta \in \Theta\}$, and $\epsilon_2 = \sup\{(\boldsymbol{w}^*)'\boldsymbol{m}(\theta) : \theta \in \Theta\}$.

Assumption 1(a) says that the modulus of continuity is attained for all sufficiently small $\epsilon \geq 0$, which typically holds if Θ is closed.[18] Assumption 1(b) requires that the unit vector $\frac{\boldsymbol{m}(\theta_\epsilon)}{\|\boldsymbol{m}(\theta_\epsilon)\|} \in \mathbb{R}^n$ converge to some constant $\boldsymbol{w}^*$ faster than ϵ as $\epsilon \to 0$. The limit $\boldsymbol{w}^*$ can be viewed as the direction at which the welfare difference $L(\theta)$ increases the most when we move $\boldsymbol{m}(\theta)$ from $\mathbf{0}$. In Section 4, I show that Assumption 1(b) holds for the example in Section 2.3 by calculating a closed-form expression for $\frac{\boldsymbol{m}(\theta_\epsilon)}{\|\boldsymbol{m}(\theta_\epsilon)\|}$ for any sufficiently small $\epsilon > 0$. In principle, it is possible to verify whether Assumption 1(b) holds or not by numerically computing the limit of $\frac{\boldsymbol{m}(\theta_\epsilon)}{\|\boldsymbol{m}(\theta_\epsilon)\|}$ as $\epsilon \to 0$ and its convergence rate.

Assumption 1(c) and (d) are mild regularity conditions. Assumption 1(c) says that θ_ϵ lies in Θ even after receiving a small perturbation in the direction of some ι such that $L(\iota) \neq 0$. Assumption 1(d) assumes the differentiability of $\omega(\epsilon)$ and $\rho(\epsilon) = \sup\{L(\theta) : (\boldsymbol{w}^*)'\boldsymbol{m}(\theta) = \epsilon, \theta \in \Theta\}$. I provide sufficient conditions for the differentiability in Appendix A.3. I make Assumption

[17]When $a^*\sigma \geq \sup_{\theta\in\Theta} \|\boldsymbol{m}(\theta)\|$, Theorem 1 holds with $a^*\sigma$ replaced with $\sup_{\theta\in\Theta} \|\boldsymbol{m}(\theta)\|$.
[18]See Donoho (1994, Lemma 2) for sufficient conditions.

1(c) and (d) to simplify the characterization of a minimax regret decision rule. In Section 6, I present the results under relaxed conditions.

The following theorem derives a minimax regret rule.

Theorem 1 (Minimax Regret Rule)**.** *Let* Θ *be convex and centrosymmetric, and suppose that Assumption 1 holds. Let*

$$\epsilon^* \in \arg\max_{\epsilon\in[0,a^*\sigma]} \omega(\epsilon)\Phi(-\epsilon/\sigma),$$

and suppose that there exists $\theta_{\epsilon^*} \in \Theta$ *that attains the modulus of continuity at* ϵ^**. Then, the following decision rule is minimax regret:*

$$\delta^*(\boldsymbol{Y}) = \begin{cases} \mathbf{1}\left\{\boldsymbol{m}(\theta_{\epsilon^*})'\boldsymbol{Y} \geq 0\right\} & \text{if } \sigma > 2\phi(0)\frac{\omega(0)}{\omega'(0)}, \\ \mathbf{1}\left\{(\boldsymbol{w}^*)'\boldsymbol{Y} \geq 0\right\} & \text{if } \sigma = 2\phi(0)\frac{\omega(0)}{\omega'(0)}, \\ \Phi\left(\dfrac{(\boldsymbol{w}^*)'\boldsymbol{Y}}{((2\phi(0)\omega(0)/\omega'(0))^2-\sigma^2)^{1/2}}\right) & \text{if } \sigma < 2\phi(0)\frac{\omega(0)}{\omega'(0)}, \end{cases}$$

where $\omega'(0)$ *is the right derivative of* $\omega(\cdot)$ *at* $\epsilon = 0$.[19] *Here,* $\boldsymbol{m}(\theta_{\epsilon^*})$ *does not depend on the choice of* θ_{ϵ^*} *among those that attain the modulus of continuity at* ϵ^**. The minimax risk is given by*

$$\mathcal{R}(\sigma;\Theta) = \omega(\epsilon^*)\Phi(-\epsilon^*/\sigma).$$

Proof. See Section 6. □

The minimax regret rule takes different forms for the case where $\sigma \geq 2\phi(0)\frac{\omega(0)}{\omega'(0)}$ and for the case where $\sigma < 2\phi(0)\frac{\omega(0)}{\omega'(0)}$. If $\sigma \geq 2\phi(0)\frac{\omega(0)}{\omega'(0)}$, the minimax regret rule is nonrandomized, making a choice according to the sign of a weighted sum of the sample $\boldsymbol{Y}$. If $\sigma < 2\phi(0)\frac{\omega(0)}{\omega'(0)}$, the minimax regret rule is randomized, assigning a positive probability both to policies 1 and 0.

Below, I discuss the interpretations and implications of Theorem 1, starting from the condition $\sigma \geq 2\phi(0)\frac{\omega(0)}{\omega'(0)}$.

3.3 When and Why Randomize?

The condition $\sigma \geq 2\phi(0)\frac{\omega(0)}{\omega'(0)}$ determines whether the minimax regret rule is randomized or not. This condition is related to the strength of the restrictions imposed on the parameter space Θ. To see this, note first that $\omega(0) = \sup\{L(\theta) : \boldsymbol{m}(\theta) = \mathbf{0}, \theta \in \Theta\}$ by definition. Since L and $\boldsymbol{m}$ are linear and Θ is convex and centrosymmetric, it is shown that the closure of the identified set

[19]The right derivative exists since $\omega(\cdot)$ is concave. Under Assumption 1, it is shown that $\omega'(0) > 0$. See Lemma 5 in Section 6.2 and its proof for the details.

of $L(\theta)$ when $\boldsymbol{m}(\theta) = \mathbf{0}$ is given by[20]

$$\mathrm{cl}(\{L(\theta) : \boldsymbol{m}(\theta) = \mathbf{0}, \theta \in \Theta\}) = [-\omega(0), \omega(0)].$$

We can thus interpret $\omega(0)$ as half of the length of the identified set of $L(\theta)$ when $\boldsymbol{m}(\theta) = \mathbf{0}$.

If $L(\theta)$ is identified, the length of the identified set is zero, which means that $\omega(0) = 0$. Since $\sigma \geq 2\phi(0)\frac{\omega(0)}{\omega'(0)}$, the minimax regret rule is always nonrandomized for cases where $L(\theta)$ is identified. In the example of Section 2.3, $L(\theta)$ is identified if, for example, we specify a polynomial model for f (see Section 5.2.1 for details).

On the other hand, if $L(\theta)$ is not identified, the length of the identified set is nonzero, which means that $\omega(0) > 0$. If the identified set is small relative to σ (holding $\omega'(0)$ fixed), the condition $\sigma \geq 2\phi(0)\frac{\omega(0)}{\omega'(0)}$ holds, and the minimax regret rule is nonrandomized. If the identified set is large relative to σ, the minimax regret rule is randomized. In Section 4, I show how this condition translates into one regarding the Lipschitz constant C in the example of Section 2.3.

For understanding why the policy maker should randomize their decisions when $\omega(0)$ is large relative to σ, it is useful to consider the problem of finding worst-case parameter values for a generic decision rule δ. The worst-case regret is attained at the parameter values that optimally trade off the potential welfare loss and the probability of incurring loss (i.e., $L(\theta)$ and $1-\mathbb{E}_\theta[\delta(\boldsymbol{Y})]$ when $L(\theta) \geq 0$). Suppose that $\omega(0)$ is large relative to σ and that the policy maker uses a nonrandomized rule. Since σ is small, $\boldsymbol{Y}$ does not vary much across repeated samples, which makes the policy maker's choice based on the nonrandomized rule too predictable. By exploiting it, it is easy to find a value of θ under which the policy maker chooses the inferior policy with a high probability. If $\omega(0)$ is large enough, such choice of θ is not likely associated with a small welfare loss, leading to a large expected welfare loss of the decision rule. The policy maker can avoid this by randomizing their decisions; randomization makes their choice less predictable and protects against the exploitation of predictable choices.[21]

3.4 Intuition for Minimax Regret Rules

I now provide intuition for and interpretations of the minimax regret rule separately for the case where the rule is nonrandomized and for the one where the rule is randomized.

Nonrandomized Rule. If $\sigma > 2\phi(0)\frac{\omega(0)}{\omega'(0)}$, we first compute $\epsilon^* \in \arg\max_{\epsilon\in[0,a^*\sigma]} \omega(\epsilon)\Phi(-\epsilon/\sigma)$ to construct the minimax regret rule. This maximization problem corresponds to that of finding the hardest one-dimensional subproblem, as discussed in Section 3.1. The corresponding pa-

[20] Since L and $\boldsymbol{m}$ are linear and Θ is centrosymmetric, $-\omega(0) = \inf\{L(\theta) : \boldsymbol{m}(\theta) = \mathbf{0}, \theta \in \Theta\}$. Moreover, for any $\alpha \in (-\omega(0), \omega(0))$, we can find $\theta \in \Theta$ such that $L(\theta) = \alpha$ and $\boldsymbol{m}(\theta) = \mathbf{0}$ by the linearity of L and $\boldsymbol{m}$ and the convexity of Θ.

[21] The fact that the minimax regret criterion may lead to a randomized rule under partial identification has been documented in the literature on treatment choice. See, for example, Manski (2007, 2009, 2011a,b) and Stoye (2012).

rameter values θ_{ϵ^*} and $-\theta_{\epsilon^*}$ are the least favorable for the policy maker, where θ_{ϵ^*} attains the modulus of continuity at ϵ^*.

The minimax regret rule δ^* protects against the worst cases. To see this, note that the optimal policy is policy 1 under θ_{ϵ^*} and policy 0 under $-\theta_{\epsilon^*}$. The decision rule δ^* chooses policy 1 if the signal $\boldsymbol{Y}$ agrees more with θ_{ϵ^*} (i.e., $\boldsymbol{m}(\theta_{\epsilon^*})'\boldsymbol{Y} > 0$) and chooses policy 0 if the signal $\boldsymbol{Y}$ agrees more with $-\theta_{\epsilon^*}$ (i.e., $\boldsymbol{m}(\theta_{\epsilon^*})'\boldsymbol{Y} < 0$). More specifically, $\boldsymbol{m}(\theta_{\epsilon^*})'\boldsymbol{Y}$ is shown to be a sufficient statistic of the sample $\boldsymbol{Y}$ for the parameter θ in the one-dimensional subproblem $[-\theta_{\epsilon^*}, \theta_{\epsilon^*}]$. The sign of $\boldsymbol{m}(\theta_{\epsilon^*})'\boldsymbol{Y}$ provides information about whether the true θ is closer to θ_{ϵ^*} or to $-\theta_{\epsilon^*}$.

Randomized Rule. If $\sigma \leq 2\phi(0)\frac{\omega(0)}{\omega'(0)}$, $\epsilon^* = 0$ as shown in Lemma 1, where $\epsilon^* \in \arg\max_{\epsilon\in[0,a^*\sigma]} \omega(\epsilon)\Phi(-\epsilon/\sigma)$. As a result, θ_0 and $-\theta_0$ are the least favorable parameter values for the policy maker, where θ_0 attains the modulus of continuity at $\epsilon = 0$. Unlike in the case where $\epsilon^* > 0$, the minimax regret rule does not base decisions on a weighted sum $\boldsymbol{m}(\theta_0)'\boldsymbol{Y}$, which is always zero since $\boldsymbol{m}(\theta_0) = \boldsymbol{0}$ by definition. The minimax regret rule instead uses $(\boldsymbol{w}^*)'\boldsymbol{Y}$, where $\boldsymbol{w}^* = \lim_{\epsilon\to 0} \frac{\boldsymbol{m}(\theta_\epsilon)}{\|\boldsymbol{m}(\theta_\epsilon)\|}$.

Simple calculations show that the randomized minimax regret rule is equivalent to

$$\delta^*(\boldsymbol{Y}) = \Pr_{\xi\sim\mathcal{N}(0,(2\phi(0)\omega(0)/\omega'(0))^2-\sigma^2)} \left((\boldsymbol{w}^*)'\boldsymbol{Y} + \xi \geq 0\right).$$

This rule is obtained through the following two-step procedure. We first add a noise $\xi \sim \mathcal{N}(0, (2\phi(0)\omega(0)/\omega'(0))^2 - \sigma^2)$ to $(\boldsymbol{w}^*)'\boldsymbol{Y}$. This addition artificially increases the standard deviation of $(\boldsymbol{w}^*)'\boldsymbol{Y}$ from σ to $2\phi(0)\frac{\omega(0)}{\omega'(0)}$, which is the threshold at which we switch from a nonrandomized rule to a randomized rule. We then make a decision according to the sign of $(\boldsymbol{w}^*)'\boldsymbol{Y} + \xi$.

The larger $\omega(0)$ is, the larger the variance of ξ is and the more dependent the choice is on the noise. As a result, given any realization of $\boldsymbol{Y}$, the probabilities of choosing policy 1 and policy 0 approach 1/2 as $\omega(0)$ increases, which suggests that the decisions become more mixed if we impose weaker restrictions on Θ.

3.5 Relation to Existing Results

Theorem 1 contains Proposition 7(iii) of Stoye (2012) as a special case. He considers a simple setup with a specific form of partial identification. In the notation of this paper, we observe a scalar sample $Y \sim \mathcal{N}(m(\theta), \sigma^2)$, $\theta = (\theta_1, \theta_2)' \in \mathbb{R}^2$, $m(\theta) = \theta_1$, $\Theta = \{(\theta_1, \theta_2)' \in [-1, 1]^2 : \theta_2 \in [a\theta_1 - b, a\theta_1 + b]\}$ for some known constants $a \in (0, 1]$ and $b \in (0, 1)$, and $L(\theta) = \theta_2$.[22] Stoye

[22] Strictly speaking, Stoye (2012) also covers the case where $b \geq 1$. This case is not covered by Theorem 1 since Assumption 1(c) does not hold; $\theta^* = (0, 1)'$ attains the modulus of continuity at $\epsilon = 0$, but there exists no $\theta \in \Theta$ such that $L(\theta) = \theta_2 \neq 0$ and $\theta^* + c\theta \in \Theta$ for any small $c > 0$. Theorem 3 in Section 6.2 covers this case.

(2012) shows that the following rule is minimax regret:

$$\delta^*(Y) = \begin{cases} \mathbf{1}\{Y \geq 0\} & \text{if } \sigma \geq 2\phi(0)\frac{b}{a}, \\ \Phi\left(\frac{Y}{((2\phi(0)b/a)^2 - \sigma^2)^{1/2}}\right) & \text{if } \sigma < 2\phi(0)\frac{b}{a}. \end{cases}$$

The condition $\sigma \geq 2\phi(0)\frac{b}{a}$ is equivalent to $\sigma \geq 2\phi(0)\frac{\omega(0)}{\omega'(0)}$ since $\omega(\epsilon) = \sup\{\theta_2 : \theta_1 \in [-\epsilon, \epsilon], (\theta_1, \theta_2)' \in \Theta\} = \min\{a\epsilon + b, 1\}$ in this setup. Note that the nonrandomized minimax regret rule $\delta^*(Y) = \mathbf{1}\{Y \geq 0\}$ is insensitive to any of σ, a, and b as long as $\sigma \geq 2\phi(0)\frac{b}{a}$.

Theorem 1 confirms that both nonrandomized and randomized rules can be minimax regret even in much more general setups than the above. At the same time, Theorem 1 suggests that the nonrandomized minimax regret rule $\mathbf{1}\{\boldsymbol{m}(\theta_{\epsilon^*})'\boldsymbol{Y} \geq 0\}$ may be sensitive to σ and Θ, since ϵ^* and θ_{ϵ^*} depend on them. Therefore, the robustness of the nonrandomized minimax regret rule to the error variance and to the parameter space is not a general property.[23]

In a special case of this paper's setup, Ishihara and Kitagawa (2021) characterize the decision rule that minimizes the maximum regret within the class of decision rules of the form $\delta(\boldsymbol{Y}) = \mathbf{1}\{\boldsymbol{w}'\boldsymbol{Y} \geq 0\}$, where $\boldsymbol{w} \in \mathbb{R}^n$. Theorem 1 shows that this restricted class contains the minimax regret rule when $\sigma \geq 2\phi(0)\frac{\omega(0)}{\omega'(0)}$ and does not when $\sigma < 2\phi(0)\frac{\omega(0)}{\omega'(0)}$.

4 Application to Eligibility Cutoff Choice

Theorem 1 provides a procedure to compute a minimax regret rule for the example in Section 2.3. In this section, I provide the formula of the minimax regret rule and discuss how the rule depends on the specification of the Lipschitz constant C and the new cutoff c_1.

I first normalize $\boldsymbol{Y}$ and $\boldsymbol{m}(\cdot)$ by left multiplying them by $\boldsymbol{\Sigma}^{-1/2}$ so that the variance-covariance matrix of the sample is a diagonal matrix:

$$\tilde{\boldsymbol{Y}} \sim \mathcal{N}(\tilde{\boldsymbol{m}}(f), \boldsymbol{I}_n),$$

where $\tilde{\boldsymbol{Y}} = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{Y} = (Y_1/\sigma(x_1, d_1), ..., Y_n/\sigma(x_n, d_n))'$, and $\tilde{\boldsymbol{m}}(f) = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{m}(f) = (f(x_1, d_1)/\sigma(x_1, d_1), ..., f(x_n, d_n)/\sigma(x_n, d_n))'$. For illustration, I focus on the Lipschitz class $\mathcal{F} = \mathcal{F}_{\text{Lip}}(C)$ and suppose that the welfare is the sample average of the expected outcome. The welfare difference is given by

$$L(f) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\{c_1 \leq x_i < c_0\}[f(x_i, 1) - f(x_i, 0)].$$

[23] It is possible to come up with an example where the nonrandomized rule depends on σ and Θ if $\boldsymbol{Y}$ is two or higher dimensional. This suggests that the robustness property is specific to the problem with a scalar sample, where a reasonable nonrandomized decision rule is only either $\mathbf{1}\{Y \geq 0\}$ or $\mathbf{1}\{Y < 0\}$.

Below, I first verify that Assumption 1 holds and then apply Theorem 1 to derive the minimax regret rule. Assumption 1(c) straightforwardly holds with $\iota(x, d) = d$ for all $x \in \mathbb{R}$, provided that Assumption 1(a) holds. Assumption 1(d) is shown to hold in Appendix A.4.

I verify Assumption 1(a) and (b) by deriving a closed-form expression for a value of f that attains the modulus of continuity at ϵ when ϵ is sufficiently small. The modulus of continuity $\omega(\epsilon; L, \tilde{\boldsymbol{m}}, \mathcal{F}_{\text{Lip}}(C))$ is computed by solving

$$\sup_{f \in \mathcal{F}_{\text{Lip}}(C)} \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}\{c_1 \leq x < c_0\}[f(x_i, 1) - f(x_i, 0)] \quad s.t. \quad \sum_{i=1}^{n} \frac{f(x_i, d_i)^2}{\sigma^2(x_i, d_i)} \leq \epsilon^2. \tag{3}$$

The unknown parameter f is infinite dimensional, but the objective and the norm constraint $\sum_{i=1}^{n} \frac{f(x_i,d_i)^2}{\sigma^2(x_i,d_i)} \leq \epsilon^2$ depend on f only through its values at $(x_1, 0), ..., (x_n, 0), (x_1, 1), ..., (x_n, 1)$. This optimization problem can be reduced to the following convex optimization problem with $2n$ unknowns and $1 + n(n-1)$ inequality constraints by a slight modification of Theorem 2.2 in Armstrong and Kolesár (2021):

$$\max_{(f(x_i,0), f(x_i,1))_{i=1,...,n} \in \mathbb{R}^{2n}} \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}\{c_1 \leq x_i < c_0\}[f(x_i, 1) - f(x_i, 0)] \tag{4}$$
$$s.t. \quad \sum_{i=1}^{n} \frac{f(x_i, d_i)^2}{\sigma^2(x_i, d_i)} \leq \epsilon^2, \;\; f(x_i, d) - f(x_j, d) \leq C|x_i - x_j|, \quad d \in \{0, 1\}, i, j \in \{1, ..., n\}.$$

Once we find a solution $(f(x_i, 0), f(x_i, 1))_{i=1,...,n}$ to (4), we can always find a function $f \in \mathcal{F}_{\text{Lip}}(C)$ that interpolates the points $(x_i, f(x_i, 0)), (x_i, f(x_i, 1))$, $i = 1, ..., n$ (Beliakov, 2006, Theorem 4), which is a solution to the original problem (3).

Now, I show that the problem (4) has a closed-form solution for any sufficiently small $\epsilon \geq 0$. The derivation utilizes the specific treatment assignment rule in the RD, namely $d_i = \mathbf{1}\{x_i \geq c_0\}$. Let $\tilde{n} = \sum_{i=1}^{n} \mathbf{1}\{c_1 \leq x_i < c_0\}$ denote the number of units whose treatment status would be changed if the cutoff were changed. Additionally, let $x_{+,\min} = \min\{x_i : x_i \geq c_0\}$ be the value of x of the treated unit closest to the original cutoff c_0, and let $\sigma^2_{+,\min} = \sigma^2(x_{+,\min}, 1)$. To simplify the exposition, I assume that $x_i \neq x_j$ for any $i \neq j$, $i, j = 1, ..., n$, in what follows.[24]

Proposition 1 (Solution to Modulus Problem for Cutoff Choice)**.** *Suppose that $d_i = \mathbf{1}\{x_i \geq c_0\}$ for all $i = 1, ..., n$ and that $x_i \neq x_j$ for any $i \neq j$, $i, j = 1, ..., n$. Then, there exists $\bar{\epsilon} > 0$ such*

[24] It is possible to obtain a closed-form solution without this assumption at the cost of making the presentation more complex.

that for any $\epsilon \in [0, \bar{\epsilon}]$, one solution to (4) is given by

$$f_\epsilon(x_i, 0) = \begin{cases} 0 & \text{if } x_i < c_1 \text{ or } x_i \geq c_0, \\ -\dfrac{\sigma^2(x_i, 0)\epsilon}{\bar{\sigma}} & \text{if } c_1 \leq x_i < c_0, \end{cases}$$

$$f_\epsilon(x_i, 1) = \begin{cases} 0 & \text{if } x_i > x_{+,\min}, \\ C(x_{+,\min} - x_i) + \dfrac{\tilde{n}\sigma^2_{+,\min}\epsilon}{\bar{\sigma}} & \text{if } x_i \leq x_{+,\min}, \end{cases}$$

and the modulus of continuity is given by

$$\omega(\epsilon; L, \tilde{\boldsymbol{m}}, \mathcal{F}_{\mathrm{Lip}}(C)) = C\frac{1}{n}\sum_{i=1}^n \mathbf{1}\{c_1 \leq x_i < c_0\}[x_{+,\min} - x_i] + \frac{\bar{\sigma}\epsilon}{n},$$

where $\bar{\sigma} = (\tilde{n}^2\sigma^2_{+,\min} + \sum_{i:c_1 \leq x_i < c_0} \sigma^2(x_i, 0))^{1/2}$.

Proof. See Appendix B.2. □

A brief explanation of this result is as follows. Since $d_i = \mathbf{1}\{x_i \geq c_0\}$, the norm constraint of (4) does not depend on $f(x_i, 1)$ for i with $x_i < c_0$. The upper bound on $f(x_i, 1)$ for such unit i is $C(x_{+,\min} - x_i) + f(x_{+,\min}, 1)$ under the Lipschitz constraint; $(x_i, f(x_i, 1))$ lies on the straight line with slope $-C$ that goes through $(x_{+,\min}, f(x_{+,\min}, 1))$. Given a value of $f(x_{+,\min}, 1)$, the objective of (4) then becomes $C\frac{1}{n}\sum_{i:c_1 \leq x_i < c_0}[x_{+,\min} - x_i] + \frac{\tilde{n}}{n} f(x_{+,\min}, 1) - \frac{1}{n}\sum_{i:c_1 \leq x_i < c_0} f(x_i, 0)$, which is a constant plus a weighted sum of $f(x_{+,\min}, 1)$ and $f(x_i, 0)$ for i with $c_1 \leq x_i < c_0$. By maximizing this under the norm constraint, we obtain the display of f_ϵ in Proposition 1, which turns out to satisfy the Lipschitz constraint for any sufficiently small ϵ.

Assumption 1(a) immediately follows from Proposition 1. Moreover, for any $\epsilon \in [0, \bar{\epsilon}]$, $\frac{\tilde{\boldsymbol{m}}(f_\epsilon)}{\|\tilde{\boldsymbol{m}}(f_\epsilon)\|} = \frac{\tilde{\boldsymbol{m}}(f_\epsilon)}{\epsilon}$ is constant and equal to $\boldsymbol{w}^* = (w_1^*, ..., w_n^*)'$, where

$$w_i^* = \begin{cases} 0 & \text{if } x_i < c_1 \text{ or } x_i > x_{+,\min}, \\ -\dfrac{\sigma(x_i, 0)}{\bar{\sigma}} & \text{if } c_1 \leq x_i < c_0, \\ \dfrac{\tilde{n}\sigma_{+,\min}}{\bar{\sigma}} & \text{if } x_i = x_{+,\min}. \end{cases} \tag{5}$$

Therefore, Assumption 1(b) holds.

Now, I apply Theorem 1 to derive the minimax regret rule. By Proposition 1, we obtain closed-form expressions for $\omega(0; L, \tilde{\boldsymbol{m}}, \mathcal{F}_{\mathrm{Lip}}(C))$ and $\omega'(0; L, \tilde{\boldsymbol{m}}, \mathcal{F}_{\mathrm{Lip}}(C))$:

$$\omega(0; L, \tilde{\boldsymbol{m}}, \mathcal{F}_{\mathrm{Lip}}(C)) = \frac{C}{n}\sum_{i=1}^n \mathbf{1}\{c_1 \leq x_i < c_0\}[x_{+,\min} - x_i],\ \omega'(0; L, \tilde{\boldsymbol{m}}, \mathcal{F}_{\mathrm{Lip}}(C)) = \frac{\bar{\sigma}}{n}.$$

Let

$$\sigma^* := 2\phi(0)\frac{\omega(0; L, \tilde{\boldsymbol{m}}, \mathcal{F}_{\text{Lip}}(C))}{\omega'(0; L, \tilde{\boldsymbol{m}}, \mathcal{F}_{\text{Lip}}(C))} = 2\phi(0)C\sum_{i=1}^{n}\mathbf{1}\{c_1 \le x_i < c_0\}[x_{+,\min} - x_i]/\bar{\sigma}. \tag{6}$$

Recall that $\tilde{\boldsymbol{Y}} = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{Y} = (Y_1/\sigma(x_1, d_1), ..., Y_n/\sigma(x_n, d_n))'$, $\tilde{\boldsymbol{m}}(f) = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{m}(f) = (f(x_1, d_1)/\sigma(x_1, d_1), ..., f(x_n, d_n)/\sigma(x_n, d_n))'$, and the variance of $\tilde{\boldsymbol{Y}}$ is $\boldsymbol{I}_n$. Let $\epsilon^* \in \arg\max_{\epsilon\in[0,a^*]} \omega(\epsilon; L, \tilde{\boldsymbol{m}}, \mathcal{F}_{\text{Lip}}(C))\Phi(-\epsilon)$ and $(f_{\epsilon^*}(x_i, 0), f_{\epsilon^*}(x_i, 1))_{i=1,...,n}$ solve the problem (4) for $\epsilon = \epsilon^*$. By Theorem 1, the following rule is minimax regret:

$$\delta^*(\boldsymbol{Y}) = \begin{cases} \mathbf{1}\left\{\sum_{i=1}^n f_{\epsilon^*}(x_i, d_i)Y_i/\sigma^2(x_i, d_i) \ge 0\right\} & \text{if } 1 > \sigma^*, \\ \mathbf{1}\left\{\sum_{i=1}^n w_i^* Y_i/\sigma(x_i, d_i) \ge 0\right\} & \text{if } 1 = \sigma^*, \\ \Phi\left(\dfrac{\sum_{i=1}^n w_i^* Y_i/\sigma(x_i, d_i)}{((\sigma^*)^2 - 1)^{1/2}}\right) & \text{if } 1 < \sigma^*. \end{cases} \tag{7}$$

The minimax regret rule makes a decision based on a weighted sum of $Y_1, ..., Y_n$.

To understand how the rule differs across different values of the Lipschitz constant C, suppose first that the magnitude of C is moderate so that σ^* is marginally smaller than 1. In this case, ϵ^* tends to be sufficiently small, which implies that $\frac{f_{\epsilon^*}(x_i,d_i)/\sigma(x_i,d_i)}{\epsilon^*} = w_i^*$ by Proposition 1. The minimax regret rule is given by

$$\begin{aligned} \delta^*(\boldsymbol{Y}) &= \mathbf{1}\left\{\sum_{i=1}^n f_{\epsilon^*}(x_i, d_i)Y_i/\sigma^2(x_i, d_i) \ge 0\right\} \\ &= \mathbf{1}\left\{\sum_{i=1}^n w_i^* Y_i/\sigma(x_i, d_i) \ge 0\right\} = \mathbf{1}\left\{Y_{+,\min} - \frac{1}{\tilde{n}}\sum_{i:c_1\le x_i<c_0} Y_i \ge 0\right\}, \end{aligned}$$

where $Y_{+,\min} = Y_i$ for i with $x_i = x_{+,\min}$. $Y_{+,\min} - \frac{1}{\tilde{n}}\sum_{i:c_1\le x_i<c_0} Y_i$ is the difference between the outcome of the treated unit closest to the status quo cutoff c_0 and the mean outcome across the untreated units between the two cutoffs c_0 and c_1. This difference can be interpreted as an estimator of the effect of the cutoff change. The outcomes of the other units are not used to construct the estimator. The minimax regret rule makes a decision according to its sign.

On the other hand, if the Lipschitz constant C is small enough so that σ^* is substantially smaller than 1, nonzero weights may be assigned to some of the other units, that is, $f_{\epsilon^*}(x_i, d_i)$ may be nonzero for some of the units with $x_i < c_1$ or $x_i > x_{+,\min}$. If the Lipschitz constant C is large enough so that $\sigma^* > 1$, the minimax regret rule is a randomized rule based on $Y_{+,\min} - \frac{1}{\tilde{n}}\sum_{i:c_1\le x_i<c_0} Y_i$.

Whether the minimax regret rule is randomized or not depends not only on the Lipschitz constant C but also on the cutoffs c_0 and c_1 and $\bar{\sigma} = (\tilde{n}^2\sigma_{+,\min}^2 + \sum_{i:c_1\le x_i<c_0}\sigma^2(x_i, 0))^{1/2}$. To investigate their relationships, suppose that $\sigma^2(x_i, d_i) = \sigma^2$ for all i for some $\sigma^2 > 0$ for simplicity.

In this situation, $\bar{\sigma} = (\tilde{n}^2 + \tilde{n})^{1/2}\sigma$, and

$$\sigma^* = \frac{2\phi(0)C\frac{1}{\tilde{n}}\sum_{i=1}^{n}\mathbf{1}\{c_1 \leq x_i < c_0\}[x_{+,\min} - x_i]}{(1+1/\tilde{n})^{1/2}\,\sigma}.$$

σ^* is nonincreasing in c_1 since $\frac{1}{\tilde{n}}\sum_{i=1}^{n}\mathbf{1}\{c_1 \leq x_i < c_0\}[x_{+,\min} - x_i]$ and $\tilde{n}$ are nonincreasing in c_1.[25] Furthermore, σ^* is decreasing in σ. Therefore, the minimax regret rule is nonrandomized when c_1 is large (i.e., when the cutoff change $c_0 - c_1$ is small) or σ is large. The minimax regret rule is randomized otherwise.

4.1 Practical Implementation

Here, I summarize the procedure for computing the minimax regret rule and discuss practical issues. Given the conditional variances $\sigma^2(x_i, d_i)$, $i = 1, ..., n$ and the Lipschitz constant C, the minimax regret rule is computed as follows.

1. Compute σ^* using the closed-form expression (6).
2. If $1 > \sigma^*$, find $\epsilon^* \in \arg\max_{\epsilon\in[0,a^*]}\omega(\epsilon)\Phi(-\epsilon)$ and compute f_{ϵ^*} that attains the modulus of continuity at ϵ^*. For each $\epsilon \geq 0$, $\omega(\epsilon)$ is computed by solving the convex optimization problem (4).[26] An efficient method for computing ϵ^* is provided in Appendix A.6.
3. If $1 \leq \sigma^*$, compute $\boldsymbol{w}^*$ using the closed-form expression (5).
4. Construct the decision rule according to (7).

In practice, the conditional variance $\sigma^2(x_i, d_i)$ is unknown. I suggest using a consistent estimator in place of the true $\sigma^2(x_i, d_i)$. The conditional variance can be estimated, for example, by applying a local linear regression to the squared residuals (Fan and Yao, 1998) or by the nearest-neighbor variance estimator (Abadie and Imbens, 2006). In the case where unit i represents a group of individuals and where Y_i is the sample mean outcome within group i, it is natural to use the conventional standard error of the sample mean as $\sigma(x_i, d_i)$.

Implementation of the minimax regret rule requires choosing the Lipschitz constant C. In principle, it is not possible to choose the Lipschitz constant C that applies to both sides of the cutoff c_0 in a data-driven way since we only observe outcomes either under treatment or under no treatment on each side. It is, however, possible to estimate a lower bound on C by using the following observation: if $f \in \mathcal{F}_{\mathrm{Lip}}(C)$ is differentiable, a lower bound on C is given by $\max\left\{\max_{\tilde{x}\geq c_0}\left|\frac{\partial f(\tilde{x},1)}{\partial x}\right|, \max_{\tilde{x}<c_0}\left|\frac{\partial f(\tilde{x},0)}{\partial x}\right|\right\}$ since $\left|\frac{\partial f(\tilde{x},d)}{\partial x}\right| \leq C$ for all $\tilde{x}$ and d. To estimate

[25] Whether σ^* is increasing in c_0 or not depends on the empirical distribution of x_i.

[26] In the empirical application in Section 7, I solve the convex optimization problem using CVXPY, a Python-embedded modeling language for convex optimization problems (Diamond and Boyd, 2016; Agrawal, Verschueren, Diamond and Boyd, 2018).

a lower bound, we could estimate the derivatives $\frac{\partial f(\tilde{x},1)}{\partial x}$ for $\tilde{x} \geq c_0$ and $\frac{\partial f(\tilde{x},0)}{\partial x}$ for $\tilde{x} < c_0$ by a local polynomial regression and then take the maximum of their absolute values over a plausible, relevant interval.[27] In practice, I recommend considering a range of plausible choices of C, including the estimated lower bound, to conduct a sensitivity analysis. I implement this approach for my empirical application in Section 7.

5 Additional Implications of the Main Result

In this section, I present two implications of Theorem 1. First, I discuss the difference between the minimax regret rule and a plug-in decision rule based on a linear minimax mean squared error (MSE) estimator. Second, I investigate the properties of the minimax regret rule when the welfare difference is point identified.

5.1 Comparison with a Plug-in Rule Based on a Linear Minimax Mean Squared Error Estimator

When $\sigma > 2\phi(0)\frac{\omega(0)}{\omega'(0)}$, the minimax regret rule is a nonrandomized rule that makes a choice according to the sign of a weighted sum of $\boldsymbol{Y}$. In this section, I compare the nonrandomized minimax regret rule with a plug-in rule based on a linear minimax MSE estimator.[28]

To define the alternative rule, let $\hat{L}_{\text{MSE}}(\boldsymbol{Y}) = \boldsymbol{w}'_{\text{MSE}}\boldsymbol{Y}$ be a *linear minimax MSE estimator*, where

$$\boldsymbol{w}_{\text{MSE}} \in \arg\min_{\boldsymbol{w}\in\mathbb{R}^n} \sup_{\theta\in\Theta} \mathbb{E}_\theta[(\boldsymbol{w}'\boldsymbol{Y} - L(\theta))^2].$$

$\hat{L}_{\text{MSE}}(\boldsymbol{Y})$ is an estimator of $L(\theta)$ that has the smallest worst-case MSE within the class of linear estimators. I define the *plug-in MSE rule* as $\delta_{\text{MSE}}(\boldsymbol{Y}) = \mathbf{1}\{\hat{L}_{\text{MSE}}(\boldsymbol{Y}) \geq 0\}$, which makes a choice according to the sign of the linear minimax MSE estimator of $L(\theta)$.

Donoho (1994) characterizes $\hat{L}_{\text{MSE}}(\boldsymbol{Y})$ using the modulus of continuity. For a simple statement of Donoho (1994)'s result, suppose that $\omega(\cdot)$ is differentiable and that Θ is convex and centrosymmetric. Let $\epsilon_{\text{MSE}} > 0$ solve

$$\frac{\epsilon^2}{\epsilon^2 + \sigma^2} = \frac{\omega'(\epsilon)\epsilon}{\omega(\epsilon)}.$$

The linear minimax MSE estimator is then given by $\hat{L}_{\text{MSE}}(\boldsymbol{Y}) = \frac{\omega'(\epsilon_{\text{MSE}})}{\epsilon_{\text{MSE}}}\boldsymbol{m}(\theta_{\epsilon_{\text{MSE}}})'\boldsymbol{Y}$, where $\theta_{\epsilon_{\text{MSE}}}$ attains the modulus of continuity at ϵ_{MSE} with $\|\boldsymbol{m}(\theta_{\epsilon_{\text{MSE}}})\| = \epsilon_{\text{MSE}}$. The plug-in MSE rule is $\delta_{\text{MSE}}(\boldsymbol{Y}) = \mathbf{1}\{\boldsymbol{m}(\theta_{\epsilon_{\text{MSE}}})'\boldsymbol{Y} \geq 0\}$.

[27] Simply taking the maximum of the estimated derivatives could raise a concern of upward bias. One could use the method for intersection bounds developed by Chernozhukov, Lee and Rosen (2013) to address it.

[28] In Appendix A.2, I also compare the minimax regret rule with a hypothesis testing rule that chooses policy 1 if a hypothesis that supports policy 0 is rejected.

Recall that the minimax regret rule is $\delta^*(\boldsymbol{Y}) = \mathbf{1}\{\boldsymbol{m}(\theta_{\epsilon^*})'\boldsymbol{Y} \geq 0\}$ if $\sigma > 2\phi(0)\frac{\omega(0)}{\omega'(0)}$, where ϵ^* solves $\max_{\epsilon\in[0,a^*\sigma]}\omega(\epsilon)\Phi(-\epsilon/\sigma)$.

Proposition 2 (Comparison with Plug-in MSE Rule)**.** *Suppose that $\omega(\cdot)$ is differentiable, and let ϵ_{MSE} solve $\frac{\epsilon^2}{\epsilon^2+\sigma^2} = \frac{\omega'(\epsilon)\epsilon}{\omega(\epsilon)}$ and ϵ^* solve $\max_{\epsilon\in[0,a^*\sigma]}\omega(\epsilon)\Phi(-\epsilon/\sigma)$. Then, $\epsilon^* < \epsilon_{\text{MSE}}$.*

Proof. See Appendix B.3. □

I provide an implication of Proposition 2 through the following result from Donoho (1994) and Low (1995): the optimal bias-variance frontier in the estimation of $L(\theta)$ can be traced out by a class of linear estimators $\{\hat{L}_\epsilon(\boldsymbol{Y})\}_{\epsilon>0}$ of the form $\hat{L}_\epsilon(\boldsymbol{Y}) = \frac{\omega'(\epsilon)}{\epsilon}\boldsymbol{m}(\theta_\epsilon)'\boldsymbol{Y}$. Here, θ_ϵ attains the modulus of continuity at ϵ with $\|\boldsymbol{m}(\theta_\epsilon)\| = \epsilon$. Specifically, for each $\epsilon > 0$, $\hat{L}_\epsilon(\boldsymbol{Y})$ minimizes the maximum bias among all linear estimators with variance bounded by $\text{Var}(\hat{L}_\epsilon(\boldsymbol{Y})) = (\sigma\omega'(\epsilon))^2$:

$$\frac{\omega'(\epsilon)}{\epsilon}\boldsymbol{m}(\theta_\epsilon) \in \arg\min_{\boldsymbol{w}\in\mathbb{R}^n} \overline{\text{Bias}}_\Theta(\boldsymbol{w}'\boldsymbol{Y}) \quad s.t. \quad \text{Var}(\boldsymbol{w}'\boldsymbol{Y}) \leq (\sigma\omega'(\epsilon))^2,$$

where $\overline{\text{Bias}}_\Theta(\boldsymbol{w}'\boldsymbol{Y}) = \sup_{\theta\in\Theta}\mathbb{E}_\theta[\boldsymbol{w}'\boldsymbol{Y} - L(\theta)]$ is the maximum bias of $\boldsymbol{w}'\boldsymbol{Y}$ over Θ. As ϵ increases, the maximum bias $\overline{\text{Bias}}_\Theta(\hat{L}_\epsilon(\boldsymbol{Y}))$ increases and the variance $\text{Var}(\hat{L}_\epsilon(\boldsymbol{Y})) = (\sigma\omega'(\epsilon))^2$ decreases. ϵ_{MSE} minimizes the worst-case MSE $\sup_{\theta\in\Theta}\mathbb{E}_\theta[(\hat{L}_\epsilon(\boldsymbol{Y}) - L(\theta))^2] = \overline{\text{Bias}}_\Theta(\hat{L}_\epsilon(\boldsymbol{Y}))^2 + \text{Var}(\hat{L}_\epsilon(\boldsymbol{Y}))$.

Since $\delta^*(\boldsymbol{Y}) = \mathbf{1}\{\boldsymbol{m}(\theta_{\epsilon^*})'\boldsymbol{Y} \geq 0\} = \mathbf{1}\{\hat{L}_{\epsilon^*}(\boldsymbol{Y}) \geq 0\}$, the minimax regret rule $\delta^*(\boldsymbol{Y})$ can be viewed as a rule that makes a choice according to the sign of the linear estimator $\hat{L}_{\epsilon^*}(\boldsymbol{Y})$. Proposition 2 implies that the corresponding linear estimator $\hat{L}_{\epsilon^*}(\boldsymbol{Y})$ places more importance on the bias than on the variance compared with the linear minimax MSE estimator $\hat{L}_{\text{MSE}}(\boldsymbol{Y})$. In other words, ϵ^* minimizes a particular weighted average of the squared bias and variance $\alpha \cdot \overline{\text{Bias}}_\Theta(\hat{L}_\epsilon(\boldsymbol{Y}))^2 + (1-\alpha)\cdot\text{Var}(\hat{L}_\epsilon(\boldsymbol{Y}))$ for some $\alpha \in [1/2, 1]$.[29] This result suggests that the plug-in MSE rule is not necessarily optimal under the minimax regret criterion.

5.2 Minimax Regret Rules Under Point Identification of Welfare Difference

Here, I discuss the properties of the minimax regret rule in special cases where the welfare difference $L(\theta)$ is point identified. The starting point is the following result, a corollary of Theorem 1. The minimax regret rule is locally insensitive to Θ in some special cases.

Corollary 1 (Robustness)**.** *Let Θ be convex and centrosymmetric, and let $\theta^* \in \mathbb{V}$ solve $\sup_{\theta\in\mathbb{V}:\|\boldsymbol{m}(\theta)\|\leq 1} L(\theta)$, where $\mathbb{V}$ is the vector space that θ belongs to. Suppose that $\{\epsilon\theta^* : 0 \leq \epsilon \leq a^*\sigma\} \subset \Theta$. Then, the decision rule*

$$\delta^*(\boldsymbol{Y}) = \mathbf{1}\left\{\boldsymbol{m}(\theta^*)'\boldsymbol{Y} \geq 0\right\}$$

is minimax regret. The minimax risk is given by $\mathcal{R}(\sigma;\Theta) = a^*\sigma L(\theta^*)\Phi(-a^*)$.

[29] In Figure A.3 in Appendix C, I report the weight $\alpha \in [1/2, 1]$ to empirically quantify how much importance the minimax regret rule places on the bias in the empirical illustration in Section 7.

Proof. See Appendix B.4. □

Note that θ^* depends on $\mathbb{V}$, not on Θ. Corollary 1 implies that the minimax regret rule and minimax risk are robust to the choice of Θ as long as Θ is large enough to contain $\{\epsilon\theta^* : 0 \leq \epsilon \leq a^*\sigma\}$.

Notably, the above result holds only for cases where the welfare difference $L(\theta)$ is identified when $\boldsymbol{m}(\theta) = \boldsymbol{0}$. This is because it is shown that $\omega(0) = 0$ under the conditions in Corollary 1 (see Appendix B.4), which means that $\{L(\theta) : \boldsymbol{m}(\theta) = \boldsymbol{0}, \theta \in \Theta\} = \{0\}$.

I illustrate this result through an example with linear regression models. This example nests the one in Section 2.3 when $\mathcal{F}$ is a class of polynomial functions. I show that the minimax regret rule is based on the best linear unbiased estimator of the parameter θ when Θ is sufficiently large.

5.2.1 Example: Linear Regression Models

Suppose it is known to the policy maker that the sample $\boldsymbol{Y} = (Y_1, ..., Y_n)'$ is generated by the following linear regression model:

$$\boldsymbol{Y} = \boldsymbol{X}\theta + \boldsymbol{U}, \quad \boldsymbol{U} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}),$$

where $\boldsymbol{X}$ is a fixed $n \times k$ design matrix that stacks k-dimensional covariate vectors of n units, $\theta \in \Theta \subset \mathbb{V} = \mathbb{R}^k$, and $\boldsymbol{\Sigma}$ is a known variance-covariance matrix. This model is a special case of the general one where $\boldsymbol{m}(\theta) = \boldsymbol{X}\theta$. Suppose also that the welfare difference is given by

$$L(\theta) = \boldsymbol{l}'\theta$$

for some known $\boldsymbol{l} \in \mathbb{R}^k$. This setup covers the example in Section 2.3 when $\mathcal{F}$ is a class of polynomial functions.[30]

I normalize $\boldsymbol{Y}$ and $\boldsymbol{m}(\cdot)$ by left multiplying them by $\boldsymbol{\Sigma}^{-1/2}$ so that the variance-covariance matrix of the sample is a diagonal matrix:

$$\tilde{\boldsymbol{Y}} \sim \mathcal{N}(\tilde{\boldsymbol{m}}(\theta), \boldsymbol{I}_n),$$

where $\tilde{\boldsymbol{Y}} = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{Y}$ and $\tilde{\boldsymbol{m}}(\theta) = \tilde{\boldsymbol{X}}\theta$ with $\tilde{\boldsymbol{X}} = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{X}$.

In this example, $L(\theta)$ is identified as long as the rank condition holds. To see this, suppose that $\tilde{\boldsymbol{m}}(\theta) = \boldsymbol{\mu}$ for some $\boldsymbol{\mu} \in \mathbb{R}^n$. If $\tilde{\boldsymbol{X}}$ is of rank k so that $\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}}$ is invertible, then $\theta = (\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}})^{-1}\tilde{\boldsymbol{X}}'\boldsymbol{\mu}$, and hence $L(\theta) = \boldsymbol{l}'(\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}})^{-1}\tilde{\boldsymbol{X}}'\boldsymbol{\mu}$.

[30] Suppose that $\mathcal{F} = \mathcal{F}_{\mathrm{Pol}}(p) \coloneqq \{f : f(x,d) = (\boldsymbol{x}', d\boldsymbol{x}')\theta \text{ for some } \theta \in \mathbb{R}^{2(p+1)}\}$, where $\boldsymbol{x} = (1, x, x^2, ..., x^p)' \in \mathbb{R}^{p+1}$. $\mathcal{F}_{\mathrm{Pol}}(p)$ is the set of functions such that $f(\cdot, d)$ is a polynomial function of degree at most p for each $d \in \{0,1\}$. Let $\boldsymbol{X}$ denote the $n \times 2(p+1)$ matrix whose i-th row is $(\boldsymbol{x}_i', d_i\boldsymbol{x}_i')$. The model $\boldsymbol{Y} \sim \mathcal{N}(\boldsymbol{m}(f), \boldsymbol{\Sigma})$ with $f \in \mathcal{F}_{\mathrm{Pol}}(p)$ then becomes $\boldsymbol{Y} \sim \mathcal{N}(\boldsymbol{X}\theta, \boldsymbol{\Sigma})$ with $\theta \in \mathbb{R}^{2(p+1)}$. Since $f(x,1) - f(x,0) = (\boldsymbol{x}', 1\cdot\boldsymbol{x}')\theta - (\boldsymbol{x}', 0\cdot\boldsymbol{x}')\theta$, the welfare difference is given by $L(\theta) = \boldsymbol{l}'\theta$, where $\boldsymbol{l} = \left(\int \boldsymbol{1}\{c_1 \leq x < c_0\}\left[(\boldsymbol{x}', 1\cdot\boldsymbol{x}') - (\boldsymbol{x}', 0\cdot\boldsymbol{x}')\right] d\nu(x)\right)' \in \mathbb{R}^{2(p+1)}$.

To apply Corollary 1, consider the following problem:

$$\sup_{\theta \in \mathbb{R}^k} \boldsymbol{l}'\theta \ s.t. \ \left(\theta' \tilde{\boldsymbol{X}}' \tilde{\boldsymbol{X}} \theta\right)^{1/2} \leq 1.$$

Simple calculations show that the solution is $\theta^* = \frac{(\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}})^{-1}\boldsymbol{l}}{(\boldsymbol{l}'(\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}})^{-1}\boldsymbol{l})^{1/2}}$.

By Corollary 1, if $\{\epsilon\theta^* : 0 \leq \epsilon \leq a^*\} \subset \Theta$, the minimax regret rule is given by

$$\delta^*(\boldsymbol{Y}) = \mathbf{1}\left\{(\tilde{\boldsymbol{X}}\theta^*)'\tilde{\boldsymbol{Y}} \geq 0\right\} = \mathbf{1}\left\{\boldsymbol{l}'\hat{\theta}_{\text{WLS}}(\boldsymbol{Y}) \geq 0\right\},$$

where $\hat{\theta}_{\text{WLS}}(\boldsymbol{Y}) = (\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{Y}$ is the weighted least squares (WLS) estimator of θ using $\boldsymbol{\Sigma}^{-1}$ as the weighting matrix. $\hat{\theta}_{\text{WLS}}(\boldsymbol{Y})$ is the best linear unbiased estimator of θ by the Gauss-Markov theorem. $\boldsymbol{l}'\hat{\theta}_{\text{WLS}}(\boldsymbol{Y})$ can be viewed as an estimator of $L(\theta) = \boldsymbol{l}'\theta$.

If Θ does not contain $\{\epsilon\theta^* : 0 \leq \epsilon \leq a^*\}$, the minimax regret rule does not generally admit a closed form. When a closed form is not available, we can directly use Theorem 1 to calculate the minimax regret rule. The discussion in Section 5.1 suggests that the minimax regret rule may use a linear estimator of $L(\theta) = \boldsymbol{l}'\theta$ that optimally trades off the bias and variance.

6 Proof of Theorem 1

In this section, I provide the proof of Theorem 1. I provide separate arguments for the nonrandomized and randomized rules. For each case, I first state an assumption weaker than the conditions in Theorem 1, present a result under the relaxed assumption, and then provide the proof for the more general result.

6.1 Nonrandomized Rule

Consider the following assumption.

Assumption 2 (Informative Worst Case)**.** *There exists a unique, nonzero solution to the maximization problem* $\max_{\epsilon \in [0, a^*\sigma]} \omega(\epsilon)\Phi(-\epsilon/\sigma)$.

An interpretation of this assumption is as follows. As discussed in Section 3.1, the maximization problem $\max_{\epsilon \in [0, a^*\sigma]} \omega(\epsilon)\Phi(-\epsilon/\sigma)$ corresponds to the problem of finding the hardest one-dimensional subproblem. The hardest one-dimensional subproblem is $[-\theta_{\epsilon^*}, \theta_{\epsilon^*}]$, where $\epsilon^* \in \arg\max_{\epsilon \in [0, a^*\sigma]} \omega(\epsilon)\Phi(-\epsilon/\sigma)$ and θ_{ϵ^*} attains the modulus of continuity at ϵ^*. If the constraint $\|\boldsymbol{m}(\theta_{\epsilon^*})\| \leq \epsilon^*$ of the modulus problem holds with equality and Assumption 2 holds, $\|\boldsymbol{m}(\theta_{\epsilon^*})\| > 0$. Equivalently, $\boldsymbol{m}(\theta_{\epsilon^*}) \neq \mathbf{0}$, which means that the signal $\boldsymbol{Y}$ under the worst-case parameter values θ_{ϵ^*} and $-\theta_{\epsilon^*}$ is informative.

The following lemma shows that Assumption 2 holds if $\sigma > 2\phi(0)\frac{\omega(0)}{\omega'(0)}$ under Assumption 1(d).

Lemma 1. *Suppose that $\omega(\cdot)$ is differentiable at any $\epsilon \in (0, a^*\sigma]$. Then, there exists a unique solution to the maximization problem $\max_{\epsilon\in[0,a^*\sigma]} \omega(\epsilon)\Phi(-\epsilon/\sigma)$. The solution is nonzero if and only if $\sigma > 2\phi(0)\frac{\omega(0)}{\omega'(0)}$.*

Proof. See Appendix B.5. □

I obtain a minimax regret rule under Assumption 2. The statement on the nonrandomized rule in Theorem 1 immediately follows from the result below.

Theorem 2 (Nonrandomized Minimax Regret Rule)**.** *Let Θ be convex and centrosymmetric, and suppose that Assumption 2 holds. Let $\epsilon^* \in \arg\max_{\epsilon\in[0,a^*\sigma]} \omega(\epsilon)\Phi(-\epsilon/\sigma)$, and suppose that there exists θ_{ϵ^*} that attains the modulus of continuity at ϵ^*. Then, the decision rule $\delta^*(\boldsymbol{Y}) = \mathbf{1}\{\boldsymbol{m}(\theta_{\epsilon^*})'\boldsymbol{Y} \geq 0\}$ is minimax regret. Here, $\boldsymbol{m}(\theta_{\epsilon^*})$ does not depend on the choice of θ_{ϵ^*} among those that attain the modulus of continuity at ϵ^*. The minimax risk is given by $\mathcal{R}(\sigma;\Theta) = \omega(\epsilon^*)\Phi(-\epsilon^*/\sigma)$.*

Now, I provide the proof of Theorem 2. The proof consists of four steps. First, I consider the simplest problem with a univariate sample and a bounded scalar parameter. Second, I use the result from the first step to solve one-dimensional subproblems, where the parameter space is restricted to a one-dimensional bounded submodel. Third, I characterize the hardest one-dimensional subproblem, that is, the one-dimensional subproblem that has the largest minimax risk. Lastly, I show that the minimax risk in the original problem is achieved by a minimax regret rule for the hardest one-dimensional subproblem. For the problem of estimation and inference, Donoho (1994) splits his proof into these steps. However, the proof of each step is nontrivially different since the problem of policy choice and that of estimation and inference are not nested by each other; specifically, the loss function and the action space are different.

Step 1. Minimax Regret Rules for Univariate Problems. I begin with a class of univariate problems. The parameter θ is a scalar and lies in $\Theta = [-\tau, \tau]$ for some $\tau > 0$. We observe a sample $Y \sim \mathcal{N}(\theta, \sigma^2)$. This setup is a special case of the general framework where $\boldsymbol{m}(\theta) = \theta$. The welfare difference is given by $L(\theta) = \theta$.

Lemma 2 (Univariate Problems)**.** *Suppose that $\Theta = [-\tau, \tau]$ for some $\tau > 0$, that $\boldsymbol{m}(\theta) = \theta$, and that $L(\theta) = \theta$. Then, the decision rule $\delta^*(Y) = \mathbf{1}\{Y \geq 0\}$ is minimax regret. The minimax risk is given by*

$$\mathcal{R}_{\text{uni}}(\sigma; [-\tau,\tau]) = \begin{cases} \tau\Phi(-\tau/\sigma) & \text{if } \tau \leq a^*\sigma, \\ a^*\sigma\Phi(-a^*) & \text{if } \tau > a^*\sigma. \end{cases}$$

Proof. See Appendix B.6. □

In univariate problems, the minimax regret rule makes a choice according to the sign of the sample Y. The minimax risk does not depend on τ as long as $\tau > a^*\sigma$.

Step 2. Minimax Regret Rules for One-dimensional Subproblems. Consider the original setup where θ resides in a vector space $\mathbb{V}$. Recall that $[\tilde{\theta}, \bar{\theta}] = \{(1-\lambda)\tilde{\theta} + \lambda\bar{\theta} : \lambda \in [0,1]\}$ for $\tilde{\theta}, \bar{\theta} \in \mathbb{V}$. I use Lemma 2 to derive minimax regret rules and the minimax risk for one-dimensional subproblems of the form $[-\bar{\theta}, \bar{\theta}]$ with $L(\bar{\theta}) > 0$ and $\boldsymbol{m}(\bar{\theta}) \neq \mathbf{0}$.

Lemma 3 (Informative One-dimensional Subproblems)**.** *Suppose that $\Theta = [-\bar{\theta}, \bar{\theta}]$, where $\bar{\theta} \in \mathbb{V}$, $L(\bar{\theta}) > 0$, and $\boldsymbol{m}(\bar{\theta}) \neq \mathbf{0}$. Then, the decision rule $\delta^*(\boldsymbol{Y}) = \mathbf{1}\left\{\boldsymbol{m}(\bar{\theta})'\boldsymbol{Y} \geq 0\right\}$ is minimax regret. The minimax risk is given by*

$$\mathcal{R}(\sigma; [-\bar{\theta}, \bar{\theta}]) = \frac{L(\bar{\theta})}{\|\boldsymbol{m}(\bar{\theta})\|}\mathcal{R}_{\text{uni}}(\sigma; [-\|\boldsymbol{m}(\bar{\theta})\|, \|\boldsymbol{m}(\bar{\theta})\|]).$$

Proof. See Appendix B.7. □

In one-dimensional subproblems, the minimax regret rule chooses policy 1 if the sample $\boldsymbol{Y}$ agrees more with $\bar{\theta}$ (or $\boldsymbol{m}(\bar{\theta})'\boldsymbol{Y} > 0$) and chooses policy 0 if the sample $\boldsymbol{Y}$ agrees more with $-\bar{\theta}$ (or $\boldsymbol{m}(\bar{\theta})'\boldsymbol{Y} < 0$).

Step 3. Hardest One-dimensional Subproblems. Using Lemma 3, I characterize the supremum of the minimax risk $\mathcal{R}(\sigma; [-\bar{\theta}, \bar{\theta}])$ over all one-dimensional subproblems of the form $[-\bar{\theta}, \bar{\theta}]$, where $\bar{\theta} \in \Theta$, $L(\bar{\theta}) > 0$, and $\boldsymbol{m}(\bar{\theta}) \neq \mathbf{0}$.

First, let ϵ^* be the unique solution to $\max_{\epsilon \in [0, a^*\sigma]} \omega(\epsilon)\Phi(-\epsilon/\sigma)$, which is positive under Assumption 2. Let θ_{ϵ^*} attain the modulus of continuity at ϵ^*. Lemma B.2 in Appendix B.1 shows that the constraint $\|\boldsymbol{m}(\theta_{\epsilon^*})\| \leq \epsilon^*$ of the modulus problem holds with equality. As a result, we obtain

$$\begin{aligned}
\mathcal{R}(\sigma; [-\theta_{\epsilon^*}, \theta_{\epsilon^*}]) &= \frac{L(\theta_{\epsilon^*})}{\|\boldsymbol{m}(\theta_{\epsilon^*})\|}\mathcal{R}_{\text{uni}}(\sigma; [-\|\boldsymbol{m}(\theta_{\epsilon^*})\|, \|\boldsymbol{m}(\theta_{\epsilon^*})\|]) \\
&= \frac{\omega(\epsilon^*)}{\epsilon^*}\mathcal{R}_{\text{uni}}(\sigma; [-\epsilon^*, \epsilon^*]) \\
&= \omega(\epsilon^*)\Phi(-\epsilon^*/\sigma),
\end{aligned}$$

where the first equality follows from Lemma 3 and the last follows from Lemma 2 and the fact that $\epsilon^* \leq a^*\sigma$.

Now, I use the modulus of continuity $\omega(\epsilon)$ to write

$$\begin{aligned}\sup_{\bar{\theta}\in\Theta:L(\bar{\theta})>0,\boldsymbol{m}(\bar{\theta})\neq\mathbf{0}}\mathcal{R}(\sigma;[-\bar{\theta},\bar{\theta}]) &= \sup_{\bar{\theta}\in\Theta:L(\bar{\theta})>0,\boldsymbol{m}(\bar{\theta})\neq\mathbf{0}}\frac{L(\bar{\theta})}{\|\boldsymbol{m}(\bar{\theta})\|}\mathcal{R}_{\text{uni}}(\sigma;[-\|\boldsymbol{m}(\bar{\theta})\|,\|\boldsymbol{m}(\bar{\theta})\|]) \\ &= \sup_{\epsilon>0}\left\{\sup_{\bar{\theta}\in\Theta:\|\boldsymbol{m}(\bar{\theta})\|=\epsilon}\frac{L(\bar{\theta})}{\|\boldsymbol{m}(\bar{\theta})\|}\mathcal{R}_{\text{uni}}(\sigma;[-\|\boldsymbol{m}(\bar{\theta})\|,\|\boldsymbol{m}(\bar{\theta})\|])\right\} \\ &= \sup_{\epsilon>0}\left\{\frac{\sup_{\bar{\theta}\in\Theta:\|\boldsymbol{m}(\bar{\theta})\|=\epsilon}L(\bar{\theta})}{\epsilon}\mathcal{R}_{\text{uni}}(\sigma;[-\epsilon,\epsilon])\right\} \\ &\leq \sup_{\epsilon>0}\frac{\omega(\epsilon)}{\epsilon}\mathcal{R}_{\text{uni}}(\sigma;[-\epsilon,\epsilon]),\end{aligned}$$

where the last inequality holds by the definition of $\omega(\epsilon)$. By Lemma 2,

$$\frac{\omega(\epsilon)}{\epsilon}\mathcal{R}_{\text{uni}}(\sigma;[-\epsilon,\epsilon]) = \begin{cases}\omega(\epsilon)\Phi(-\epsilon/\sigma) & \text{if } \epsilon\leq a^*\sigma, \\ \frac{\omega(\epsilon)}{\epsilon}a^*\sigma\Phi(-a^*) & \text{if } \epsilon > a^*\sigma.\end{cases}$$

Since $\omega(\epsilon)$ is concave, $\frac{\omega(\epsilon)}{\epsilon}$ is nonincreasing, so that $\sup_{\epsilon>a^*\sigma}\frac{\omega(\epsilon)}{\epsilon}a^*\sigma\Phi(-a^*) = \omega(a^*\sigma)\Phi(-a^*)$. Therefore,

$$\sup_{\epsilon>0}\frac{\omega(\epsilon)}{\epsilon}\mathcal{R}_{\text{uni}}(\sigma;[-\epsilon,\epsilon]) = \sup_{0<\epsilon\leq a^*\sigma}\omega(\epsilon)\Phi(-\epsilon/\sigma) = \omega(\epsilon^*)\Phi(-\epsilon^*/\sigma).$$

Hence, $\sup_{\bar{\theta}\in\Theta:L(\bar{\theta})>0,\boldsymbol{m}(\bar{\theta})\neq\mathbf{0}}\mathcal{R}(\sigma;[-\bar{\theta},\bar{\theta}])\leq\omega(\epsilon^*)\Phi(-\epsilon^*/\sigma)$.

Since $\mathcal{R}(\sigma;[-\theta_{\epsilon^*},\theta_{\epsilon^*}]) = \omega(\epsilon^*)\Phi(-\epsilon^*/\sigma)$, it follows that

$$\sup_{\bar{\theta}\in\Theta:L(\bar{\theta})>0,\boldsymbol{m}(\bar{\theta})\neq\mathbf{0}}\mathcal{R}(\sigma;[-\bar{\theta},\bar{\theta}]) = \mathcal{R}(\sigma;[-\theta_{\epsilon^*},\theta_{\epsilon^*}]) = \omega(\epsilon^*)\Phi(-\epsilon^*/\sigma).$$

Therefore, $[-\theta_{\epsilon^*},\theta_{\epsilon^*}]$ is one of the hardest one-dimensional subproblems. Its minimax risk is $\omega(\epsilon^*)\Phi(-\epsilon^*/\sigma)$.

Step 4. Minimax Regret Rules for the Original Problem. By Lemma 3, the decision rule $\delta^*(\boldsymbol{Y}) = \mathbf{1}\{\boldsymbol{m}(\theta_{\epsilon^*})'\boldsymbol{Y}\geq 0\}$ is minimax regret for the one-dimensional subproblem $[-\theta_{\epsilon^*},\theta_{\epsilon^*}]$. Since $\boldsymbol{m}(\theta_{\epsilon^*})'\boldsymbol{Y}\sim\mathcal{N}(\boldsymbol{m}(\theta_{\epsilon^*})'\boldsymbol{m}(\theta),\sigma^2\|\boldsymbol{m}(\theta_{\epsilon^*})\|^2)$ under θ, the maximum regret of δ^* over Θ is given by

$$\begin{aligned}\max_{\theta\in\Theta}R(\delta^*,\theta) &= \max_{\theta\in\Theta}\left[(L(\theta))^+\Phi\left(-\frac{\boldsymbol{m}(\theta_{\epsilon^*})'\boldsymbol{m}(\theta)}{\sigma\|\boldsymbol{m}(\theta_{\epsilon^*})\|}\right) + (-L(\theta))^+\left(1-\Phi\left(-\frac{\boldsymbol{m}(\theta_{\epsilon^*})'\boldsymbol{m}(\theta)}{\sigma\|\boldsymbol{m}(\theta_{\epsilon^*})\|}\right)\right)\right] \\ &= \max_{\theta\in\Theta:L(\theta)>0}L(\theta)\Phi\left(-\frac{\boldsymbol{m}(\theta_{\epsilon^*})'\boldsymbol{m}(\theta)}{\sigma\|\boldsymbol{m}(\theta_{\epsilon^*})\|}\right),\end{aligned}$$

where $x^+ = \max\{x,0\}$ and the second equality holds by the symmetry of the objective function and the centrosymmetry of Θ.

The following lemma is fundamental to characterizing minimax regret rules for the original problem.

Lemma 4 (Worst Case for Nonrandomized Rule)**.** *Under the conditions in Theorem 2,*

$$\theta_{\epsilon^*} \in \arg\max_{\theta\in\Theta: L(\theta)>0} L(\theta)\Phi\left(-\frac{\boldsymbol{m}(\theta_{\epsilon^*})'\boldsymbol{m}(\theta)}{\sigma\|\boldsymbol{m}(\theta_{\epsilon^*})\|}\right).$$

Proof. See Appendix B.8. □

By Lemma 4, the maximum regret of the decision rule δ^* over Θ is attained at θ_{ϵ^*}. Therefore,

$$\max_{\theta\in\Theta} R(\delta^*,\theta) = \max_{\theta\in[-\theta_{\epsilon^*},\theta_{\epsilon^*}]} R(\delta^*,\theta) = \mathcal{R}(\sigma;[-\theta_{\epsilon^*},\theta_{\epsilon^*}]),$$

where the last equality holds since δ^* is minimax regret for $[-\theta_{\epsilon^*},\theta_{\epsilon^*}]$. However, by definition,

$$\max_{\theta\in\Theta} R(\delta^*,\theta) \geq \mathcal{R}(\sigma;\Theta) \geq \mathcal{R}(\sigma;[-\theta_{\epsilon^*},\theta_{\epsilon^*}]).$$

It follows that $\max_{\theta\in\Theta} R(\delta^*,\theta) = \mathcal{R}(\sigma;\Theta) = \mathcal{R}(\sigma;[-\theta_{\epsilon^*},\theta_{\epsilon^*}])$, and hence δ^* is minimax regret for Θ. The minimax risk for the original problem is the same as that for the hardest one-dimensional subproblem $[-\theta_{\epsilon^*},\theta_{\epsilon^*}]$.

Lastly, Lemma B.2 in Appendix B.1 shows that $\boldsymbol{m}(\theta_{\epsilon^*})$ does not depend on the choice of θ_{ϵ^*} among those that attain the modulus of continuity at ϵ^*, which completes the proof of Theorem 2.

6.2 Randomized Rule

Consider the following assumption.

Assumption 3 (Regularity for Randomized Rule)**.** *The following holds for some* $\bar{\epsilon} > 0$.

(a) For all $\epsilon \in [0,\bar{\epsilon}]$*, there exists* $\theta_\epsilon \in \Theta$ *that attains the modulus of continuity at* ϵ *with* $\|\boldsymbol{m}(\theta_\epsilon)\| = \epsilon$.

(b) There exists $\boldsymbol{w}^* \in \mathbb{R}^n$ *such that* $\lim_{\epsilon\to 0}\epsilon^{-1}\left(\boldsymbol{w}^* - \frac{\boldsymbol{m}(\theta_\epsilon)}{\|\boldsymbol{m}(\theta_\epsilon)\|}\right) = \boldsymbol{0}$.

(c) There exists $\sigma^* \geq \sigma$ *such that* $0 \in \arg\max_{\epsilon\in\mathbb{R}} \rho(\epsilon)\Phi(-\epsilon/\sigma^*)$*, where* $\rho(\epsilon) := \sup\{L(\theta) : (\boldsymbol{w}^*)'\boldsymbol{m}(\theta) = \epsilon, \theta\in\Theta\}$ *for* $\epsilon \in \mathbb{R}$.[31]

Assumption 3(a) is slightly stronger than Assumption 1(a) since it requires that the constraint $\|\boldsymbol{m}(\theta_\epsilon)\| \leq \epsilon$ of the modulus problem hold with equality. Assumption 3(b) is the same as Assumption 1(b).

[31] I allow the search space of σ^* to contain ∞, letting $\Phi(x/\infty) = 1/2$ for all $x \in \mathbb{R}$. Assumption 3(c) then holds with $\sigma^* = \infty$ in Stoye (2012)'s setup described in Section 3.5 when $b \geq 1$.

Given σ^*, the maximization problem $\max_{\epsilon \in \mathbb{R}} \rho(\epsilon)\Phi\left(-\epsilon/\sigma^*\right)$ described in Assumption 3(c) corresponds to the problem of finding the worst-case parameter values for a randomized decision rule $\delta(\boldsymbol{Y}) = \Pr_{\xi \sim \mathcal{N}(0,(\sigma^*)^2-\sigma^2)}\left((\boldsymbol{w}^*)'\boldsymbol{Y} + \xi \geq 0\right)$. Later, I will show that, under Assumption 3(c), we can find the variance of the artificial noise ξ such that the maximum regret of δ is attained at a value of θ that attains the modulus of continuity at $\epsilon = 0$.

The following lemma shows that these conditions hold if $\sigma \leq 2\phi(0)\frac{\omega(0)}{\omega'(0)}$ under Assumption 1.

Lemma 5. *Let Θ be convex and centrosymmetric, and suppose that Assumption 1 holds. Then, Assumption 3(a) holds. Moreover, $0 \in \arg\max_{\epsilon \in \mathbb{R}} \rho(\epsilon)\Phi\left(-\epsilon/\sigma^*\right)$ with $\sigma^* = 2\phi(0)\frac{\omega(0)}{\omega'(0)}$.*

Proof. See Appendix B.9. □

I obtain a minimax regret rule under Assumption 3. The statement on the randomized rule in Theorem 1 immediately follows from the following result.

Theorem 3 (Randomized Minimax Regret Rule)**.** *Let Θ be convex and centrosymmetric, and suppose that Assumption 3 holds. Then, the following decision rule is minimax regret:*

$$\delta^*(\boldsymbol{Y}) = \begin{cases} \mathbf{1}\{(\boldsymbol{w}^*)'\boldsymbol{Y} \geq 0\} & \text{if } \sigma^* = \sigma, \\ \Phi\left(\dfrac{(\boldsymbol{w}^*)'\boldsymbol{Y}}{((\sigma^*)^2 - \sigma^2)^{1/2}}\right) & \text{if } \sigma^* > \sigma. \end{cases}$$

The minimax risk is given by $\mathcal{R}(\sigma;\Theta) = \omega(0)/2$.

Note that if $\omega(\cdot)$ is differentiable at any $\epsilon \in (0, a^*\sigma]$ and $\sigma \leq 2\phi(0)\frac{\omega(0)}{\omega'(0)}$, $\epsilon^* = 0$ by Lemma 1, where $\epsilon^* \in \arg\max_{\epsilon \in [0,a^*\sigma]} \omega(\epsilon)\Phi(-\epsilon/\sigma)$. The minimax risk $\omega(0)/2$ can then be written as $\omega(\epsilon^*)\Phi(-\epsilon^*/\sigma)$, leading to the expression in Theorem 1.

Below, I provide the proof of Theorem 3. Note first that $\boldsymbol{w}^*$ is a unit vector by construction. Hence, $(\boldsymbol{w}^*)'\boldsymbol{Y} \sim \mathcal{N}\left((\boldsymbol{w}^*)'\boldsymbol{m}(\theta), \sigma^2\right)$. Simple calculations show that δ^* is equivalent to $\delta^*(\boldsymbol{Y}) = \Pr_{\xi \sim \mathcal{N}(0,(\sigma^*)^2-\sigma^2)}\left((\boldsymbol{w}^*)'\boldsymbol{Y} + \xi \geq 0\right)$. Since $(\boldsymbol{w}^*)'\boldsymbol{Y} + \xi \sim \mathcal{N}\left(0, (\sigma^*)^2\right)$ if $\boldsymbol{Y} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \boldsymbol{I}_n)$, it follows that $\mathbb{E}_{\boldsymbol{Y} \sim \mathcal{N}(\mathbf{0},\sigma^2\boldsymbol{I}_n)}[\delta^*(\boldsymbol{Y})] = \frac{1}{2}$. The following lemma shows that δ^* is minimax regret for the one-dimensional subproblem $[-\theta_0, \theta_0]$, where θ_0 attains the modulus of continuity at 0 and hence $\boldsymbol{m}(\theta_0) = \mathbf{0}$.

Lemma 6 (Uninformative One-dimensional Subproblems)**.** *Suppose that $\Theta = [-\bar{\theta}, \bar{\theta}]$, where $\bar{\theta} \in \mathbb{V}$, $L(\bar{\theta}) \geq 0$, and $\boldsymbol{m}(\bar{\theta}) = \mathbf{0}$. Then, any decision rule δ^* such that $\mathbb{E}_{\boldsymbol{Y} \sim \mathcal{N}(\mathbf{0},\sigma^2\boldsymbol{I}_n)}[\delta^*(\boldsymbol{Y})] = \frac{1}{2}$ is minimax regret. The minimax risk is given by $\mathcal{R}(\sigma; [-\bar{\theta}, \bar{\theta}]) = L(\bar{\theta})/2$.*

Proof. See Appendix B.10. □

If $\boldsymbol{m}(\bar{\theta}) = \mathbf{0}$, $\boldsymbol{Y} \sim \mathcal{N}(\mathbf{0}, \sigma^2\boldsymbol{I}_n)$ under any $\theta \in [-\bar{\theta}, \bar{\theta}]$. Lemma 6 shows that choosing each policy with probability one half over the distribution of $\boldsymbol{Y}$ is minimax regret for the subproblem $[-\bar{\theta}, \bar{\theta}]$.

Since $(\boldsymbol{w}^*)'\boldsymbol{Y} + \xi \sim \mathcal{N}\left((\boldsymbol{w}^*)'\boldsymbol{m}(\theta), (\sigma^*)^2\right)$ under θ, the maximum regret of δ^* over Θ is given by

$$\begin{aligned}\max_{\theta\in\Theta} R(\delta^*,\theta) &= \max_{\theta\in\Theta}\left[(L(\theta))^+\Phi\left(-\frac{(\boldsymbol{w}^*)'\boldsymbol{m}(\theta)}{\sigma^*}\right) + (-L(\theta))^+\left(1-\Phi\left(-\frac{(\boldsymbol{w}^*)'\boldsymbol{m}(\theta)}{\sigma^*}\right)\right)\right] \\ &= \max_{\theta\in\Theta} L(\theta)\Phi\left(-\frac{(\boldsymbol{w}^*)'\boldsymbol{m}(\theta)}{\sigma^*}\right),\end{aligned}$$

where the second equality holds by the symmetry of the objective function and the centrosymmetry of Θ. The following lemma shows that the maximum regret is attained at θ_0.

Lemma 7 (Worst Case for Randomized Rule)**.** *Under the conditions in Theorem 3,*

$$\theta_0 \in \arg\max_{\theta\in\Theta} L(\theta)\Phi\left(-\frac{(\boldsymbol{w}^*)'\boldsymbol{m}(\theta)}{\sigma^*}\right).$$

Proof. See Appendix B.11. □

From the above results, we obtain

$$\max_{\theta\in\Theta} R(\delta^*,\theta) = \max_{\theta\in[-\theta_0,\theta_0]} R(\delta^*,\theta) = \mathcal{R}(\sigma;[-\theta_0,\theta_0]),$$

where the last equality holds since δ^* is minimax regret for $[-\theta_0,\theta_0]$. However, by definition,

$$\max_{\theta\in\Theta} R(\delta^*,\theta) \geq \mathcal{R}(\sigma;\Theta) \geq \mathcal{R}(\sigma;[-\theta_0,\theta_0]).$$

It follows that $\max_{\theta\in\Theta} R(\delta^*,\theta) = \mathcal{R}(\sigma;\Theta) = \mathcal{R}(\sigma;[-\theta_0,\theta_0])$, and hence δ^* is minimax regret for Θ. The minimax risk is given by $\mathcal{R}(\sigma;[-\theta_0,\theta_0]) = L(\theta_0)/2 = \omega(0)/2$.

7 Empirical Policy Application

I now illustrate my approach in an empirical application to the Burkinabé Response to Improve Girls' Chances to Succeed (BRIGHT) program in Burkina Faso. I consider the hypothetical problem of whether or not to expand the program and empirically compare the performance of the minimax regret rule with alternative decision rules.

7.1 Background and Data

The goal of the BRIGHT program was to improve children's and especially girls' educational outcomes in rural villages by constructing well-resourced village-based schools. The program was funded by the Millennium Challenge Corporation, a U.S. government agency, and implemented by a consortium of non-governmental organizations. The program constructed primary schools with three classrooms for grades 1 to 3 in 132 villages from 47 departments during the period from 2005 to 2008. The Ministry of Education determined the villages where schools would be built through the following process.

1. 293 villages were nominated based on low school enrollment rates.

2. The Ministry administered a survey in each village and assigned each village a score using a set formula. The formula attached large weight to the estimated number of children to be served from the nominated and neighboring villages, giving additional weight to girls.

3. The Ministry ranked villages within each department and selected the top half of the villages to receive a school.

For further details on the BRIGHT program and allocation process, see Levy, Sloan, Linden and Kazianga (2009) and Kazianga *et al.* (2013).

Since the school allocation was determined at department level, the cutoff score for the program eligibility was different across departments. Following Kazianga *et al.* (2013), I define the *relative score* as the score for each village minus the cutoff score for the department that the village belongs to. As a result, a village is eligible for the program when the relative score is larger than zero. Kazianga *et al.* (2013) use the relative score as a running variable and evaluate the causal effect of the program on educational outcomes using a regression discontinuity design. Figure 1 reports the distribution of the relative score.

I use the replication data for Kazianga *et al.* (2013)'s results (Kazianga, Levy, Linden and Sloan, 2019) and consider whether we should expand the program or not. I explain the details of the counterfactual policy in Section 7.2. The dataset contains survey results about 30 households from 287 nominated villages, yielding a total sample of 23,282 children between the ages of 5 and 12. The survey was conducted in 2008, namely 2.5 years after the start of the program. Table 1 reports summary statistics about child educational outcomes and characteristics. Children in eligible villages are more likely to attend school, achieve higher test scores, and complete a higher grade. Household heads in eligible villages completed slightly more years of schooling. Furthermore, households in eligible villages tend to have more assets such as basic roofing and motorbikes.

I consider school enrollment as the target outcome. Since the score and program eligibility are determined at village level, I use the village-level mean outcome, namely the enrollment rate for each village. This setting fits into the setup in Section 2.3, where i represents a village, Y_i is the observed enrollment rate of village i, d_i is the program eligibility, and x_i is the relative score. The original cutoff is $c_0 = 0$, that is, $d_i = \mathbf{1}\{x_i \geq 0\}$. The parameter is a function $f : \mathbb{R} \times \{0, 1\} \to \mathbb{R}$, where $f(x, d)$ represents the counterfactual mean of the enrollment rate conditional on the relative score if the eligibility status were set to $d \in \{0, 1\}$. Since Y_i is a village-level sample mean, it is plausible to assume that Y_i is approximately normally distributed. I use the conventional standard error of the sample mean Y_i as the standard deviation of Y_i.[32]

[32] The observed enrollment rate is zero in 21 out of 287 villages. I exclude these villages from the analysis since

7.2 Hypothetical Policy Choice Problem

I ask whether we should scale up the program and build BRIGHT schools in other villages. Specifically, I consider the following decision problem. The counterfactual policy is to build BRIGHT schools in previously ineligible villages whose relative scores are in the top 20%, which corresponds to lowering the cutoff from 0 to -0.256.[33] I use the average enrollment rate across villages as the welfare criterion, so that the welfare effect of this policy relative to the status quo is

$$L(f) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\{-0.256 \leq x_i < 0\}[f(x_i, 1) - f(x_i, 0)].$$

When deciding whether to implement the policy, it is important to consider the benefit relative to the cost. Kazianga *et al.* (2013) provide an estimate of the cost of constructing a BRIGHT school, which is \$4,758 per village.[34] To incorporate the cost into the decision problem, I suppose that the policy maker cares about the cost-effectiveness of this new policy relative to similar programs. Cost-effectiveness is defined as the ratio of the policy cost to the increase in the target outcome, namely the enrollment in the current context. I assume that it is optimal to implement the policy if its cost-effectiveness is smaller than \$83.77, which is the cost-effectiveness of a school construction program in Indonesia (Duflo, 2001; Kazianga *et al.*, 2013). Specifically, it is optimal to implement the policy if

$$\frac{\$4{,}758}{416 \cdot \frac{1}{\tilde{n}}\sum_{i=1}^{n} \mathbf{1}\{-0.256 \leq x_i < 0\}[f(x_i, 1) - f(x_i, 0)]} \leq \$83.77,$$

where 416 is the number of children per village and $\tilde{n} = \sum_{i=1}^{n} \mathbf{1}\{-0.256 \leq x_i < 0\}$ is the number of villages that would receive a school under the new policy.[35] The denominator represents the increase in the average enrollment across villages that would receive a BRIGHT school under the new policy.

Simple calculations show that the above condition is equivalent to

$$\frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\{-0.256 \leq x_i < 0\}[f(x_i, 1) - 0.137 - f(x_i, 0)] \geq 0.$$

My method can be used to consider this decision problem by setting the outcome to $Y_i - 0.137d_i$,

---

the standard error of Y_i is zero.

[33] Figure A.4 in Appendix C reports the results when I use 10% and 30% instead of 20%. As predicted by the result in Section 4, the minimax regret rule switches from a nonrandomized rule to a randomized rule at a smaller Lipschitz constant C when the fraction of the target villages is larger.

[34] I assume that the cost estimate is a known quantity and is constant across villages. It is, however, natural to think of the policy cost as unknown and heterogeneous across villages and to introduce the cost model on top of the outcome model. I leave this for future work.

[35] The cost per village and the cost-effectiveness of a school construction program in Indonesia are found in Tables A18 and A20, respectively, in Online Appendix of Kazianga *et al.* (2013). I compute the number of children per village by dividing the total enrollment by the enrollment rate reported in Table A17 in Online Appendix of Kazianga *et al.* (2013).

where 0.137 can be viewed as the policy cost measured in the unit of the enrollment rate. I present the results for this scenario with the cost of 0.137 as well as for the scenario where we ignore the policy cost.

I implement my method assuming that the counterfactual outcome function f belongs to the Lipschitz class $\mathcal{F}_{\text{Lip}}(C)$. Since the relative score x_i is computed based on several village-level characteristics, it is difficult to interpret and specify the Lipschitz constant C using domain-specific knowledge. To obtain a reasonable range of C, I estimate a lower bound on C using the method described in Section 4.1, which yields the lower bound estimate of 0.149.[36] I present the results for $C \in \{0.05, 0.1, ..., 0.95, 1\}$ and examine their sensitivity to the choice of C.

7.3 Results

Figure 2 plots $\delta^*(\boldsymbol{Y})$, the probability of choosing the new policy computed by the minimax regret rule, against the Lipschitz constant C. When $C < 0.6$, the minimax regret rule is nonrandomized. It chooses the new policy in the no-cost scenario and maintains the status quo in the scenario where the policy cost is 0.137. When $C \geq 0.6$, on the other hand, the minimax regret rule is randomized. The decisions become more mixed as C increases.

Given that the estimate of the lower bound on C is 0.149, the minimax regret rule is nonrandomized when C is less than four times the estimated lower bound. Under this reasonable range of C, the optimal decision is the same in each scenario. If the policy maker wants to be more conservative about the choice of C, they need to randomize their decisions.

The above analysis considers the scenario where the policy cost is fixed at 0.137. To examine the sensitivity of the result to the policy cost, I compute the maximum of the cost values under which the minimax regret rule chooses the new policy with probability one, reported in Figure 3. If the policy cost is less than this value, it is optimal to choose the new policy; otherwise, it is optimal to maintain the status quo. The result shows that, when C is above its lower bound 0.149, it is optimal to maintain the status quo as long as the policy cost is higher than 0.10.

If the minimax regret rule is nonrandomized, the rule is of the form $\delta^*(\boldsymbol{Y}) = \mathbf{1}\{\sum_{i=1}^n w_i Y_i \geq 0\}$ for some weights w_i's. Panels (a) and (b) of Figure 4 plot the weight w_i attached to each village against the relative score x_i for $C = 0.1$ and $C = 0.5$, respectively. In the plots, the size of circles is proportional to the inverse of the standard error of the enrollment rate Y_i. For both $C = 0.1$ and $C = 0.5$, a few treated units just above the original cutoff (the solid vertical line) receive a positive weight, the untreated units between the original cutoff and the new cutoff (the dashed vertical line) receive a negative weight, and no other units receive any weight. When $C = 0.1$, the weight tends to be larger for units with a smaller standard error. When $C = 0.5$, a positive weight is attached only to the treated unit closest to the original cutoff. Additionally,

[36] I estimate $\frac{\partial f(x,0)}{\partial x}$ at $x \in \{-2.5, -2.45, ..., -0.05\}$ and $\frac{\partial f(x,1)}{\partial x}$ at $x \in \{0.05, 0.1, ..., 2.5\}$ by local quadratic regression and take the maximum of their absolute values. For local quadratic regression, I use the MSE-optimal bandwidth selection procedure by Calonico, Cattaneo and Farrell (2018), which can be implemented by R package "nprobust."

the weights on the untreated units between the two cutoffs are almost identical. This situation corresponds to the minimax regret rule of the form $\delta^*(\boldsymbol{Y}) = \mathbf{1}\left\{Y_{+,\min} - \frac{1}{\tilde{n}}\sum_{i:c_1 \le x_i < c_0} Y_i \ge 0\right\}$ discussed in Section 4.

7.3.1 Comparison with Plug-in Rules

I compare the minimax regret rule with plug-in decision rules that make a decision according to the sign of an estimator of the policy effect. I consider three estimators of the policy effect.

1. The linear minimax MSE estimator (Donoho, 1994), described in Section 5.1, under the Lipschitz class $\mathcal{F}_{\text{Lip}}(C)$.

2. The linear minimax MSE estimator under the additional assumption of constant conditional treatment effects. In other words, I construct the estimator assuming that $\mathcal{F} = \{f \in \mathcal{F}_{\text{Lip}}(C) : f(x,1) - f(x,0) = f(\tilde{x},1) - f(\tilde{x},0) \text{ for all } x, \tilde{x}\}$. This estimation corresponds to first nonparametrically estimating the average treatment effect at the original cutoff and then extrapolating the effects on the units between the two cutoffs by the constant effects assumption.

3. The polynomial regression estimator (Kazianga *et al.*, 2013).[37] Given the degree of polynomial p, I first estimate the model $f(x,d) = \alpha_0 + \alpha_1 x + \cdots + \alpha_p x^p + \beta_0 d + \beta_1 d \cdot x + \cdots \beta_p d \cdot x^p$ by the weighted least squares regression using $1/\sigma^2(x_i,d_i)$ as the weight.[38] I then estimate $L(f)$ by $\frac{1}{n}\sum_{i=1}^n \mathbf{1}\{-0.256 \le x_i < 0\}[\hat{f}(x_i,1) - \hat{f}(x_i,0)]$, where $\hat{f}$ is the estimated polynomial function. This estimator relies on the functional form of f to extrapolate $f(x_i,1)$ for the untreated units.

Panel (a) of Figure 5 reports the estimated policy effects from the linear minimax MSE estimators with and without constant conditional treatment effects. Overall, these two estimators exhibit a similar pattern. While the estimated policy effects are larger than the policy cost when C is close to zero, they are smaller than the policy cost when C is moderate or large. For $C \ge 0.2$, the resulting decisions about whether to choose the new policy are the same as the decision made by the minimax regret rule until C reaches 0.6, where the minimax regret rule starts to randomize. In contrast, the estimated policy effects from the polynomial regression estimators of degrees 1 to 5 exceed the policy cost, as reported in Panel (b) of Figure 5. The estimates appear to be close to the simple mean outcome difference between eligible and ineligible villages that can be computed from Table 1. The resulting decisions are different from the decision made by the minimax regret rule.[39]

[37] Kazianga *et al.* (2013) estimate the treatment effect at the cutoff, not the effect on the units away from the cutoff. They apply global polynomial regression RD estimators to child-level data.

[38] This is equivalent to the OLS regression of $Y_i/\sigma(x_i,d_i)$ on $(1, x, ..., x^p, d, d \cdot x, ..., d \cdot x^p)'/\sigma(x_i,d_i)$.

[39] The estimators presented here can be written as $\sum_{i=1}^n w_i Y_i$ for some weights w_i's. See Figure A.1 in Appendix C for the plots of these weights. While the linear minimax MSE estimators attach weights to units just above

The above estimates and resulting decisions are computed from a particular realization of the sample. To assess the ex ante performance of different decision rules, I compute the maximum regret of these rules when the true function class is $\mathcal{F}_{\text{Lip}}(C)$.[40] Figure 6 reports the result for the minimax regret rule and the plug-in rules based on the linear minimax MSE estimators with and without constant conditional treatment effects.[41] The maximum regret of the plug-in MSE rule with constant conditional treatment effects is much larger than that of the other two, especially when the Lipschitz constant C is large. The plug-in MSE rule without constant conditional treatment effects performs worse than the minimax regret rule, as predicted by the theoretical analysis. The ratio of the maximum regret between the two rules is maximized at $C = 0.6$, where the minimax regret rule starts to randomize.

7.3.2 Sensitivity to Misspecification of Lipschitz Constant C

So far, I have constructed decision rules assuming that the Lipschitz constant C is known, which is a crucial assumption in my theoretical analysis. To assess the sensitivity of the performance to misspecification of C, I construct decision rules assuming $C = 0.3$ and then compute their maximum regret when the true value of C lies in $\{0.05, 0.1, ..., 0.95, 1\}$.

Figure 7 reports the result. The solid line indicates the "oracle" maximum regret, which can be achieved if we correctly specify C. The result shows that the plug-in MSE rule without constant conditional treatment effects performs slightly better than the minimax regret rule when the true C is close to zero. On the other hand, the minimax regret rule outperforms the plug-in MSE rule with nonnegligible differences for any value of the true C greater than 0.2. The result suggests that the minimax regret rule is more robust to misspecification of C toward zero than the plug-in MSE rule.

The potential superiority of the minimax regret rule seems consistent with the theoretical results in the following way. As shown in Section 4, when the true value of C is large, the oracle minimax regret rule only uses the treated units just above the original cutoff and the untreated units between the original and new cutoffs (see Panel (b) of Figure 4). If the specified C is smaller than the true value, the resulting minimax regret rule is closer to the oracle rule than the plug-in MSE rule since the minimax regret rule places more importance on the bias than the plug-in MSE rule as discussed in Section 5.1. Therefore, it is expected that the minimax regret rule performs better than the plug-in MSE rule under misspecification of C toward zero.

the original cutoff and to units between the two cutoffs, polynomial regression estimators attach weights even to units further away from the cutoffs.

[40] I compute the maximum regret of the minimax regret rule using the formula in Theorem 1. For the other rules, I adapt the approach by Ishihara and Kitagawa (2021) to numerically calculate the maximum regret in this setup.

[41] See Figure A.2 in Appendix C for the result for the plug-in rules based on polynomial regression estimators. The maximum regret of these rules is significantly larger than that of alternative rules.

8 Conclusion and Future Directions

This paper develops an optimal procedure for using data to make policy decisions in settings where social welfare under each counterfactual policy is only partially identified. I derive a decision rule that achieves the minimax regret optimality in finite samples and within the class of all decision rules. I apply the result to the problem of eligibility cutoff choice and illustrate it in an empirical application to a school construction program in Burkina Faso.

While my application focuses on eligibility rules based on a scalar variable, it is possible to apply my approach to a choice of treatment assignment policy based on multiple covariates. My method can also be applied to the problem of deciding whether to introduce a new policy using data from a randomized experiment when the experiment has imperfect compliance or when the experimental sample is a selected subset of the target population. I plan to apply my general result to these scenarios and provide an empirical illustration.

Several extensions of my work are possible. First, my result relies on the assumption that the sample is normally distributed with a known variance. It is challenging but natural to conside the asymptotic optimality without the distributional assumption, for example, by extending the limits of experiments framework of Hirano and Porter (2009) to setups with partial identification and restricted parameter spaces. Second, my approach only covers a binary choice problem. It is both theoretically and practically important to extend the analysis to a multiple or continuous policy space. Lastly, while this work focuses on one-shot decision-making, it may in practice be possible to make a policy change again after observing the result of a previous policy choice for a certain period of time. It would be interesting to consider such sequential decision problems.

References

Abadie, A. and Imbens, G. W. (2006). Large Sample Properties of Matching Estimators for Average Treatment Effects. *Econometrica*, **74** (1), 235–267.

Agrawal, A., Verschueren, R., Diamond, S. and Boyd, S. (2018). A Rewriting System for Convex Optimization Problems. *Journal of Control and Decision*, **5** (1), 42–60.

Andrews, I., Kitagawa, T. and McCloskey, A. (2021). Inference on Winners. *Working Paper*.

Angrist, J. D. and Rokkanen, M. (2015). Wanna Get Away? Regression Discontinuity Estimation of Exam School Effects Away From the Cutoff. *Journal of the American Statistical Association*, **110** (512), 1331–1344.

Armstrong, T. B. and Kolesár, M. (2018). Optimal Inference in a Class of Regression Models. *Econometrica*, **86** (2), 655–683.

— and Kolesár, M. (2021). Finite-Sample Optimal Estimation and Inference on Average Treatment Effects Under Unconfoundedness. *Econometrica*, **89** (3), 1141–1177.

Athey, S. and Wager, S. (2021). Policy Learning With Observational Data. *Econometrica*, **89** (1), 133–161.

Beliakov, G. (2006). Interpolation of Lipschitz Functions. *Journal of Computational and Applied Mathematics*, **196** (1), 20–44.

Bennett, M. (2020). How Far is Too Far? Estimation of an Interval for Generalization of a Regression Discontinuity Design Away from the Cutoff. *Working Paper*.

Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis; 2nd ed.* Springer Series in Statistics, New York: Springer.

Bertanha, M. (2020). Regression Discontinuity Design with Many Thresholds. *Journal of Econometrics*, **218** (1), 216–241.

— and Imbens, G. W. (2020). External Validity in Fuzzy Regression Discontinuity Designs. *Journal of Business & Economic Statistics*, **38** (3), 593–612.

Bhattacharya, D. and Dupas, P. (2012). Inferring Welfare Maximizing Treatment Assignment Under Budget Constraints. *Journal of Econometrics*, **167** (1), 168–196.

Cai, T. T. and Low, M. G. (2004). An Adaptation Theory for Nonparametric Confidence Intervals. *The Annals of Statistics*, **32** (5), 1805–1840.

Calonico, S., Cattaneo, M. D. and Farrell, M. H. (2018). On the Effect of Bias Estimation on Coverage Accuracy in Nonparametric Inference. *Journal of the American Statistical Association*, **113** (522), 767–779.

Cattaneo, M. D., Keele, L., Titiunik, R. and Vazquez-Bare, G. (2020). Extrapolating Treatment Effects in Multi-Cutoff Regression Discontinuity Designs. *Journal of the American Statistical Association*, **0** (0), 1–12.

Chernozhukov, V., Lee, S. and Rosen, A. M. (2013). Intersection Bounds: Estimation and Inference. *Econometrica*, **81** (2), 667–737.

Christensen, T., Moon, H. R. and Schorfheide, F. (2020). Robust Forecasting. *arXiv:2011.03153*.

de Chaisemartin, C. (2021). The Minimax Estimator of the Average Treatment Effect, among Linear Combinations of Estimators of Bounded Conditional Average Treatment Effects. *arXiv:2105.08766*.

DEHEJIA, R. H. (2005). Program Evaluation as a Decision Problem. *Journal of Econometrics*, **125** (1), 141–173.

— and WAHBA, S. (1999). Causal Effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Programs. *Journal of the American Statistical Association*, **94** (448), 1053–1062.

DIAMOND, S. and BOYD, S. (2016). CVXPY: A Python-embedded Modeling Language for Convex Optimization. *Journal of Machine Learning Research*, **17** (83), 1–5.

DONG, Y. and LEWBEL, A. (2015). Identifying The Effect of Changing The Policy Threshold in Regression Discontinuity Models. *The Review of Economics and Statistics*, **97** (5), 1081–1092.

DONOHO, D. L. (1994). Statistical Estimation and Optimal Recovery. *The Annals of Statistics*, **22** (1), 238–270.

DUFLO, E. (2001). Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment. *American Economic Review*, **91** (4), 795–813.

FAN, J. and YAO, Q. (1998). Efficient Estimation of Conditional Variance Functions in Stochastic Regression. *Biometrika*, **85** (3), 645–660.

HIRANO, K. and PORTER, J. R. (2009). Asymptotics for Statistical Treatment Rules. *Econometrica*, **77** (5), 1683–1701.

IMBENS, G. and WAGER, S. (2019). Optimized Regression Discontinuity Designs. *The Review of Economics and Statistics*, **101** (2), 264–278.

ISHIHARA, T. and KITAGAWA, T. (2021). Evidence Aggregation for Treatment Choice. *arXiv:2108.06473*.

KALLUS, N. (2018). Balanced Policy Evaluation and Learning. In *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, pp. 8909–8920.

— and ZHOU, A. (2021). Minimax-Optimal Policy Learning Under Unobserved Confounding. *Management Science*, **67** (5), 2870–2890.

KARLIN, S. and RUBIN, H. (1956). The Theory of Decision Procedures for Distributions with Monotone Likelihood Ratio. *The Annals of Mathematical Statistics*, **27** (2), 272–299.

KASY, M. (2016). Partial Identification, Distributional Preferences, and the Welfare Ranking of Policies. *The Review of Economics and Statistics*, **98** (1), 111–131.

— (2018). Optimal Taxation and Insurance Using Machine Learning — Sufficient Statistics and Beyond. *Journal of Public Economics*, **167**, 205–219.

Kazianga, H., Levy, D., Linden, L. L. and Sloan, M. (2013). The Effects of "Girl-Friendly" Schools: Evidence from the BRIGHT School Construction Program in Burkina Faso. *American Economic Journal: Applied Economics*, **5** (3), 41–62.

—, —, — and — (2019). Replication Data for: The Effects of "Girl-Friendly" Schools: Evidence from the BRIGHT School Construction Program in Burkina Faso, TN: American Economic Association [publisher], 2013. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2019-10-12. https://doi.org/10.3886/E113862V1.

Kitagawa, T. and Tetenov, A. (2018). Who Should Be Treated? Empirical Welfare Maximization Methods for Treatment Choice. *Econometrica*, **86** (2), 591–616.

— and — (2021). Equality-Minded Treatment Choice. *Journal of Business & Economic Statistics*, **39** (2), 561–574.

LaLonde, R. J. (1986). Evaluating the Econometric Evaluations of Training Programs with Experimental Data. *American Economic Review*, **76** (4), 604–620.

Levy, D., Sloan, M., Linden, L. L. and Kazianga, H. (2009). Impact Evaluation of Burkina Faso's BRIGHT Program: Final Report. Mathematica Policy Research, Washington, DC, USA.

Low, M. G. (1995). Bias-Variance Tradeoffs in Functional Estimation Problems. *The Annals of Statistics*, **23** (3), 824–835.

Manski, C. F. (2000). Identification Problems and Decisions Under Ambiguity: Empirical Analysis of Treatment Response and Normative Analysis of Treatment Choice. *Journal of Econometrics*, **95** (2), 415–442.

— (2004). Statistical Treatment Rules for Heterogeneous Populations. *Econometrica*, **72** (4), 1221–1246.

— (2007). Minimax-regret Treatment Choice with Missing Outcome Data. *Journal of Econometrics*, **139** (1), 105–115.

— (2009). The 2009 Lawrence R. Klein Lecture: Diversified Treatment Under Ambiguity. *International Economic Review*, **50** (4), 1013–1041.

— (2010). Vaccination with Partial Knowledge of External Effectiveness. *Proceedings of the National Academy of Sciences*, **107** (9), 3953–3960.

— (2011a). Choosing Treatment Policies Under Ambiguity. *Annual Review of Economics*, **3** (1), 25–49.

— (2011b). Policy Choice with Partial Knowledge of Policy Effectiveness. *Journal of Experimental Criminology*, **7**, 111–125.

— (2020). Econometrics For Decision Making: Building Foundations Sketched By Haavelmo And Wald. *Working Paper*.

Mbakop, E. and Tabord-Meehan, M. (2021). Model Selection for Treatment Choice: Penalized Welfare Maximization. *Econometrica*, **89** (2), 825–848.

Mo, W., Qi, Z. and Liu, Y. (2021). Learning Optimal Distributionally Robust Individualized Treatment Rules. *Journal of the American Statistical Association*, **116** (534), 659–674.

Qian, M. and Murphy, S. A. (2011). Performance Guarantees for Individualized Treatment Rules. *The Annals of Statistics*, **39** (2), 1180–1210.

Rambachan, A. and Roth, J. (2020). An Honest Approach to Parallel Trends. *Working Paper*.

Rokkanen, M. (2015). Exam Schools, Ability, and the Effects of Affirmative Action: Latent Factor Extrapolation in the Regression Discontinuity Design. Discussion Paper 1415-03, Columbia University, Department of Economics, New York, NY.

Russell, T. M. (2020). Policy Transforms and Learning Optimal Policies. *arXiv:2012.11046*.

Savage, L. J. (1951). The Theory of Statistical Decision. *Journal of the American Statistical Association*, **46** (253), 55–67.

Song, Z. (2020). Potential Implications of Lowering the Medicare Eligibility Age to 60. *Journal of the American Medical Association (JAMA)*, **323** (24), 2472–2473.

Stoye, J. (2009). Minimax Regret Treatment Choice with Finite Samples. *Journal of Econometrics*, **151** (1), 70–81.

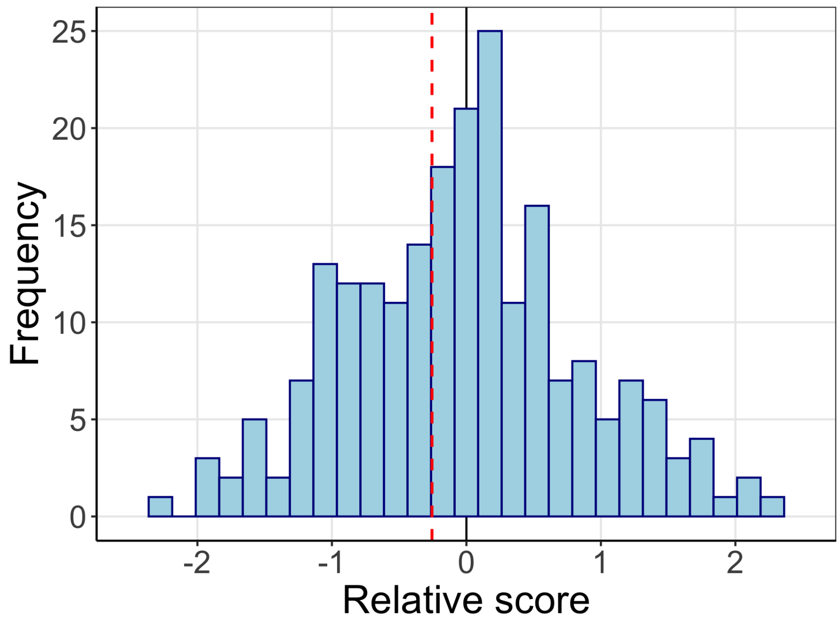
— (2012). Minimax Regret Treatment Choice with Covariates or with Limited Validity of Experiments. *Journal of Econometrics*, **166** (1), 138–156.

Swaminathan, A. and Joachims, T. (2015). Counterfactual Risk Minimization: Learning from Logged Bandit Feedback. In *Proceedings of the 32nd International Conference on International Conference on Machine Learning*, vol. 37, pp. 814–823.

Tetenov, A. (2012). Statistical Treatment Choice Based on Asymmetric Minimax Regret Criteria. *Journal of Econometrics*, **166** (1), 157–165.

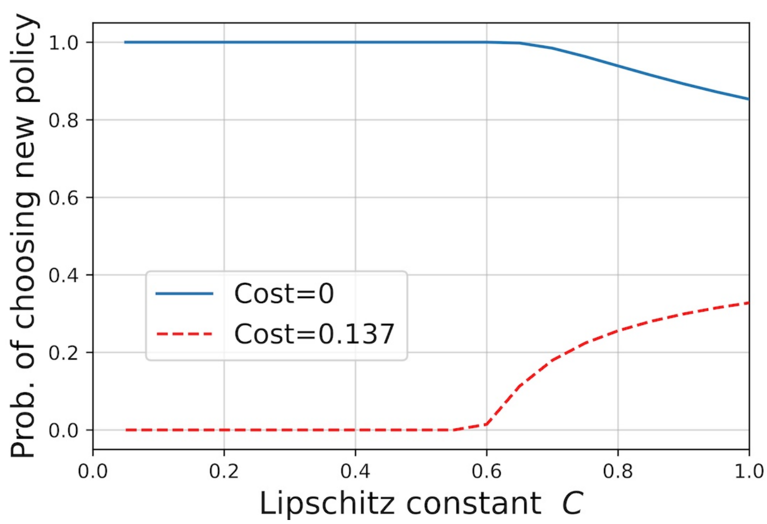
Zhao, Y., Zeng, D., Rush, A. J. and Kosorok, M. R. (2012). Estimating Individualized Treatment Rules Using Outcome Weighted Learning. *Journal of the American Statistical Association*, **107** (499), 1106–1118.

Figure 1: Distribution of Relative Score



Notes: This figure shows the histogram of the relative score of villages on the interval $[-2.5, 2.5]$. The vertical dashed line indicates the new cutoff -0.256, which corresponds to the hypothetical policy of constructing schools in previously ineligible villages whose relative scores are in the top 20%. The villages with zero observed enrollment rates are excluded.

Figure 2: Optimal Decisions: Probability of Choosing the New Policy



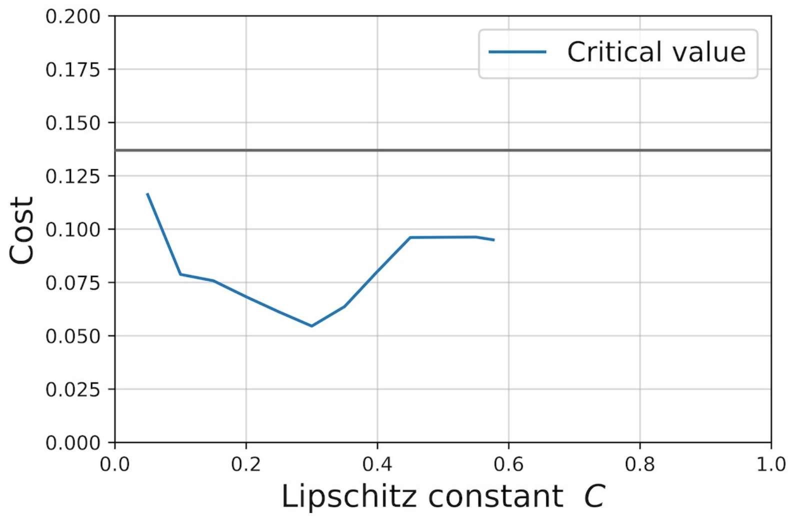
Notes: This figure shows the probability of choosing the new policy computed by the minimax regret rule. The new policy is to construct BRIGHT schools in previously ineligible villages whose relative scores are in the top 20%. The solid line shows the results for the scenario where we ignore the policy cost. The dashed line shows the results for the scenario where the policy cost measured in the unit of the enrollment rate is 0.137. I report the results for the range $[0.05, 0.1, ..., 0.95, 1]$ of the Lipschitz constant C.

Table 1: Child Educational Outcomes and Characteristics

	All (1)	Eligible villages (2)	Ineligible villages (3)
Panel A. Educational outcomes (child-level means)			
Enrollment	0.366	0.494	0.259
Normalized total test scores	0.000	0.248	−0.209
Highest grade child has achieved	0.876	1.132	0.636
Panel B. Child and household characteristics (child-level means)			
Child's age	8.121	8.174	8.071
Child is female	0.503	0.476	0.525
Head's age	47.653	47.387	47.904
Head years of schooling	0.156	0.198	0.117
Number of members	10.812	10.815	10.808
Number of children	5.971	6.098	5.850
Muslim	0.587	0.576	0.597
Basic roofing	0.516	0.534	0.500
Number of motorbikes	0.299	0.319	0.279
Number of phones	0.185	0.199	0.172
Total number of children	23,282	10,645	12,637
Total number of villages	287	136	151

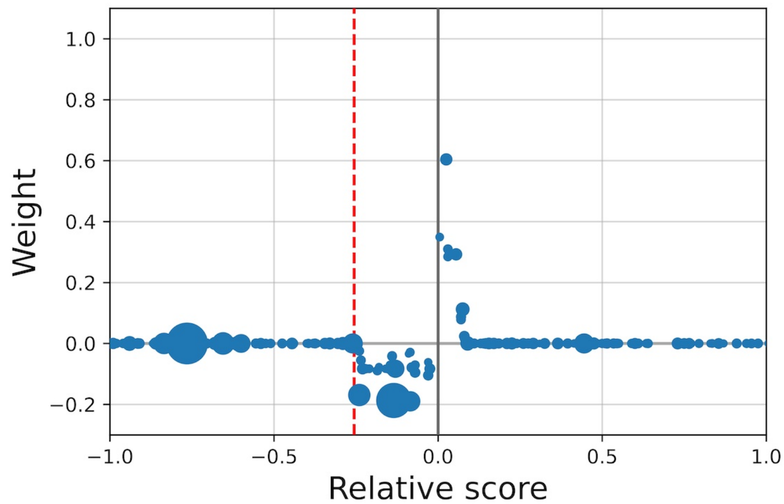
Notes: This table reports child-level averages of educational outcomes and characteristics by program eligibility in the year 2008, namely 2.5 years after the start of the BRIGHT program. Panel A reports the educational outcomes' means. Panel B reports the means of child and household characteristics. Column (1) shows the means for children in all villages. Columns (2) and (3) show the means for children in villages selected for BRIGHT school and in unselected villages, respectively.

Figure 3: Maximum of Cost Values Under Which Choosing the New Policy is Optimal

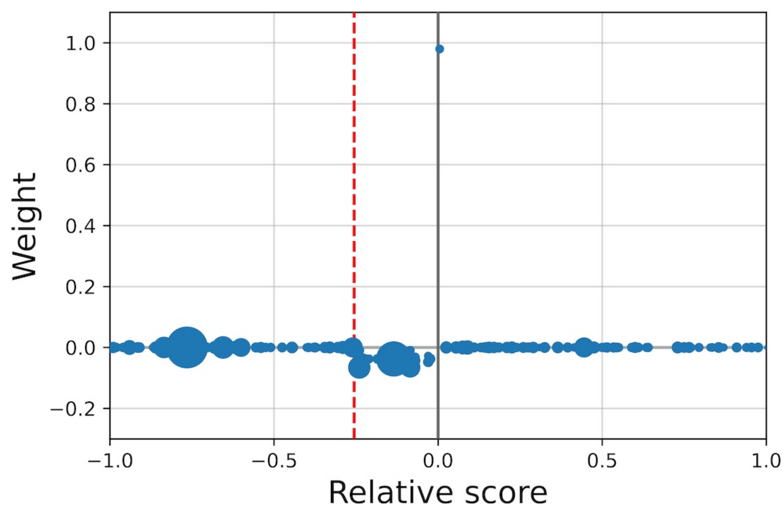


Notes: This figure shows the maximum of the cost values under which the minimax regret rule chooses the new policy with probability one. The horizontal line shows the cost of 0.137, which is my main specification of the policy cost. I only report the results for the Lipschitz constant $C < 0.6$ since the minimax regret rule is randomized for $C \geq 0.6$.

Figure 4: Weight to Each Village Attached by Minimax Regret Rule



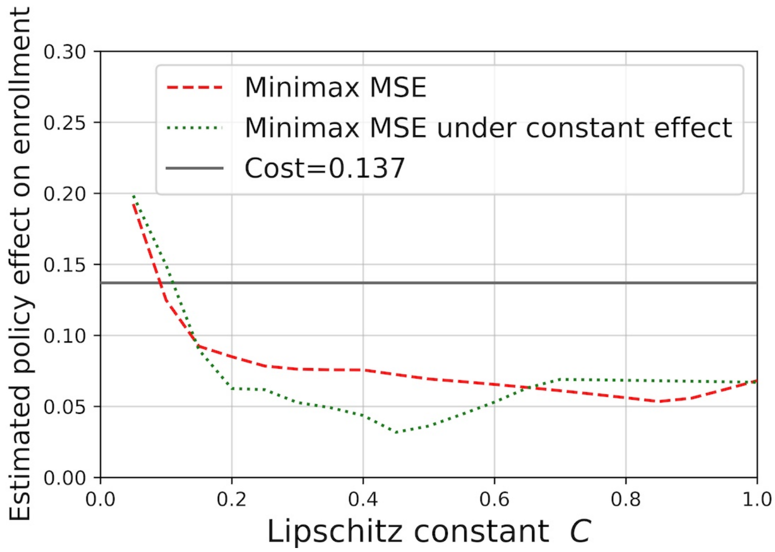
(a) $C = 0.1$



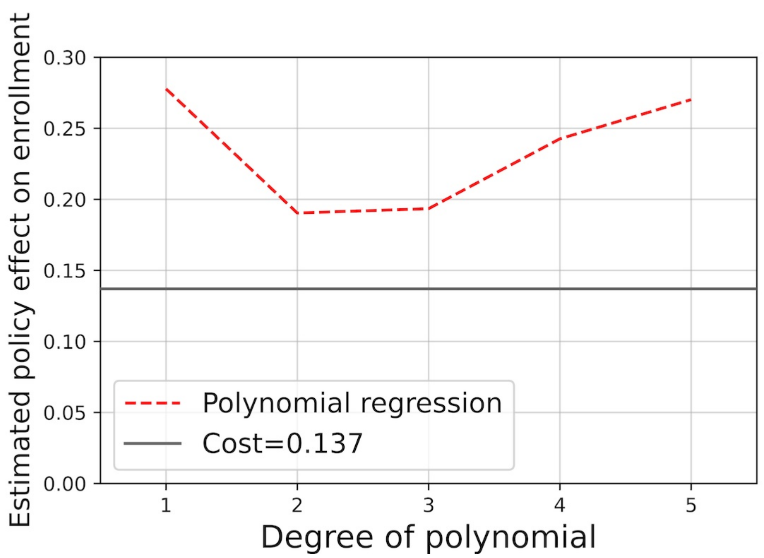
(b) $C = 0.5$

Notes: This figure shows the weight w_i attached to each village by the minimax regret rule of the form $\delta^*(\boldsymbol{Y}) = \mathbf{1}\{\sum_{i=1}^n w_i Y_i \geq 0\}$. The weights are normalized so that $\sum_{i=1}^n w_i^2 = 1$. The horizontal axis indicates the relative score of each village. Each circle corresponds to each village. The size of circles is proportional to the inverse of the standard error of the enrollment rate Y_i. The vertical dashed line corresponds to the new cutoff -0.256. Panels (a) and (b) show the results when the Lipschitz constant C is 0.1 and 0.5, respectively.

Figure 5: Estimated Effects of New Policy on Enrollment Rate



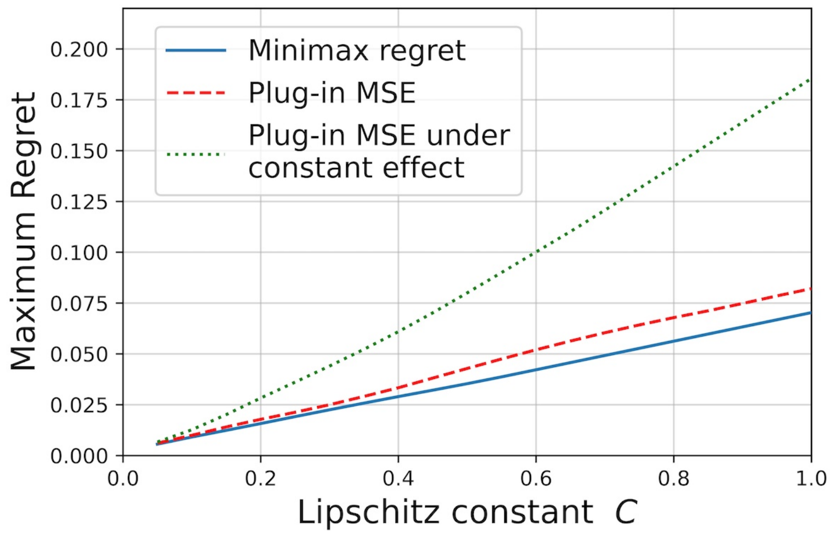
(a) Minimax MSE Estimator Under Lipschitz Class



(b) Polynomial Regression Estimator

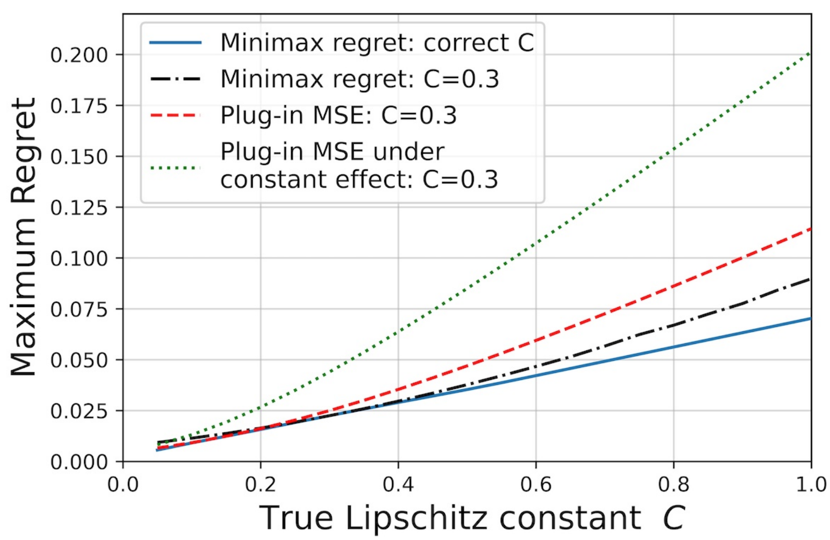
Notes: This figure shows the average effect of the new policy on the enrollment rate across the villages that would receive a school under the new policy. Panel (a) reports the estimates from the linear minimax MSE estimators with and without the assumption of constant conditional treatment effects. I report the results for the range $[0.05, 0.1, ..., 0.95, 1]$ of the Lipschitz constant C. Panel (b) reports the estimates from the polynomial regression estimators of degrees 1 to 5. The horizontal line shows the cost of 0.137, which is my main specification of the policy cost.

Figure 6: Maximum Regret of Minimax Regret Rule and Plug-in MSE Rules



Notes: This figure shows the maximum regret of the minimax regret rule and the plug-in rules based on the linear minimax MSE estimators with and without the assumption of constant conditional treatment effects. The maximum regret is normalized so that the unit is the same as that of the enrollment rate. I report the results for the range $[0.05, 0.1, ..., 0.95, 1]$ of the Lipschitz constant C.

Figure 7: Maximum Regret Under Misspecification of Lipschitz Constant C



Notes: This figure shows the maximum regret of the minimax regret rule and the plug-in rules based on the linear minimax MSE estimators with and without the assumption of constant conditional treatment effects, which are constructed assuming that the Lipschitz constant C is 0.3. The maximum regret is computed by setting the true C to the value on the horizontal axis. The solid line indicates the maximum regret that can be achieved if C is correctly specified. The maximum regret is normalized so that the unit is the same as that of the enrollment rate. I report the results for the range $[0.05, 0.1, ..., 0.95, 1]$ of the true Lipschitz constant C.

A Additional Results and Details

A.1 Example: Optimal Treatment Assignment Policy Under Unconfoundedness

The basic setup is the same as the one in Section 2.3. I generalize it in three ways. First, the covariates x_i are k dimensional, where $k \geq 1$. Second, I remove the assumption that $d_i = \mathbf{1}\{x_i \geq 0\}$ and instead assumes the unconfoundedness (i.e., the observed treatment is independent of potential outcomes conditional on covariates). These first two do not change the notation of the data-generating process:

$$\boldsymbol{Y} \sim \mathcal{N}(\boldsymbol{m}(f), \boldsymbol{\Sigma}),$$

where $\boldsymbol{Y} = (Y_1, ..., Y_n)'$, $\boldsymbol{m}(f) = (f(x_1, d_1), ..., f(x_n, d_n))'$, and $\boldsymbol{\Sigma} = \text{diag}(\sigma^2(x_1, d_1), ..., \sigma^2(x_n, d_n))$. Under the unconfoundedness, we can interpret $f(x, d)$ as the counterfactual mean outcome for those with covariates x under treatment status d (see footnote 12).

Third, two alternative policies are functions $\pi_a : \mathbb{R}^k \to [0, 1]$, $a \in \{0, 1\}$, where $\pi_a(x)$ is the probability of assigning treatment to individuals whose covariates are $x \in \mathbb{R}^k$. Suppose that the welfare under policy a, $a \in \{0, 1\}$, is an average of the counterfactual mean outcome across different values of covariates

$$W_a(f) = \int [f(x, 1)\pi_a(x) + f(x, 0)(1 - \pi_a(x))]d\nu(x)$$

for some known measure ν. The welfare difference between the two policies is

$$L(f) = W_1(f) - W_0(f) = \int (\pi_1(x) - \pi_0(x))(f(x, 1) - f(x, 0))d\nu(x).$$

One example of the function class $\mathcal{F}$ is the Lipschitz class with a known Lipschitz constant $C \geq 0$:

$$\mathcal{F}_{\text{Lip}}(C) = \{f : |f(x, d) - f(\tilde{x}, d)| \leq C\|x - \tilde{x}\| \text{ for every } x, \tilde{x} \in \mathbb{R}^k \text{ and } d \in \{0, 1\}\}.$$

A.2 Comparison with Hypothesis Testing Rules

Hypothesis testing can be viewed as an alternative procedure for deciding between two policies. Here, I compare the minimax regret rule with a class of hypothesis testing rules.

To define it, suppose Θ is convex and centrosymmetric, and consider testing

$$H_0 : L(\theta) \leq -b \text{ and } \theta \in \Theta \quad \text{vs.} \quad H_1 : L(\theta) \geq b \text{ and } \theta \in \Theta$$

for some $b > 0$. Let $\theta^{(b)}$ solve $\inf_{\theta\in\Theta:L(\theta)\geq b}\|\boldsymbol{m}(\theta)\|$. For any level $\alpha > 0$, the minimax test, which has the largest minimum power under H_1, is given by the Neyman-Pearson test of H_0 : $\theta = -\theta^{(b)}$ vs. $H_1 : \theta = \theta^{(b)}$ (Armstrong and Kolesár, 2018, Lemma A.2). It rejects H_0 if the test statistic $\boldsymbol{m}(\theta^{(b)})'\boldsymbol{Y}$ is greater than its $1-\alpha$ quantile under $-\theta^{(b)}$. Since $\boldsymbol{m}(\theta^{(b)})'\boldsymbol{Y} \sim \mathcal{N}(-\|\boldsymbol{m}(\theta^{(b)})\|^2, \sigma^2\|\boldsymbol{m}(\theta^{(b)})\|^2)$ under $-\theta^{(b)}$, the critical value is $-\|\boldsymbol{m}(\theta^{(b)})\|^2 + z_{1-\alpha}\sigma\|\boldsymbol{m}(\theta^{(b)})\|$, where $z_{1-\alpha}$ is the $1-\alpha$ quantile of a standard normal variable. The level-α minimax test is then given by

$$\delta_{\alpha,b}(\boldsymbol{Y}) = \mathbf{1}\{\boldsymbol{m}(\theta^{(b)})'\boldsymbol{Y} \geq -\|\boldsymbol{m}(\theta^{(b)})\|^2 + z_{1-\alpha}\sigma\|\boldsymbol{m}(\theta^{(b)})\|\}.$$

I call such tests *hypothesis testing rules.*

Are there any hypothesis testing rules that exactly match the minimax regret rule? Let $\epsilon^* > 0$ solve $\max_{\epsilon\in[0,a^*\sigma]}\omega(\epsilon)\Phi(-\epsilon/\sigma)$, and let θ_{ϵ^*} solve the modulus of continuity at ϵ^* with $\|\boldsymbol{m}(\theta_{\epsilon^*})\| = \epsilon^*$. By the duality of the problem, θ_{ϵ^*} also solves $\inf_{\theta\in\Theta:L(\theta)\geq b^*}\|\boldsymbol{m}(\theta)\|$, where $b^* = \omega(\epsilon^*)$. Let α^* satisfy $-\|\boldsymbol{m}(\theta_{\epsilon^*})\| + z_{1-\alpha^*}\sigma = 0$, i.e., $\alpha^* = \Phi(-\epsilon^*/\sigma)$, so that the critical value is zero. For this choice of α^* and b^*, the hypothesis testing rule is

$$\delta_{\alpha^*,b^*}(\boldsymbol{Y}) = \mathbf{1}\{\boldsymbol{m}(\theta_{\epsilon^*})'\boldsymbol{Y} \geq 0\},$$

which is identical to the minimax regret rule. Since $\epsilon^* \leq a^*\sigma$, we can obtain a lower bound on α^*: $\alpha^* = \Phi(-\epsilon^*/\sigma) \geq \Phi(-a^*) \approx 0.226$. Therefore, the minimax regret rule is less conservative in rejection of the null hypothesis than hypothesis testing rules that use conventional levels such as 0.01 and 0.05. This is consistent with the fact that the minimax regret criterion takes into consideration the potential welfare loss as well as the probability of making a wrong choice.

A.3 Sufficient Conditions for Differentiability of $\omega(\cdot)$ and $\rho(\cdot)$

The result below follows from Lemma D.1 in Supplemental Appendix D of Armstrong and Kolesár (2018) in the case where $\mathcal{F} = \mathcal{G}$ in their notation. Note that their definition of the modulus of continuity when $\mathcal{F} = \mathcal{G}$ is the same as Donoho (1994)'s definition, which is different from my definition. See Appendix A.5 for the relationship between their definition and mine.

Lemma A.1. *Let Θ be convex. Let θ_ϵ attain the modulus of continuity at $\epsilon > 0$ with $\|\boldsymbol{m}(\theta_\epsilon)\| = \epsilon$, and suppose that there exists $\iota \in \Theta$ such that $L(\iota) = 1$ and $\theta_\epsilon + c\iota \in \Theta$ for all c in a neighborhood of zero. Then, $\omega(\cdot)$ is differentiable at ϵ with $\omega'(\epsilon) = \frac{\epsilon}{\boldsymbol{m}(\iota)'\boldsymbol{m}(\theta_\epsilon)}$.*

The result below follows from arguments similar to the proof of Lemma D.1 in Armstrong and Kolesár (2018).

Lemma A.2. *Let Θ be convex. Let θ_ϵ satisfy $L(\theta_\epsilon) = \rho(\epsilon)$ and $(\boldsymbol{w}^*)'\boldsymbol{m}(\theta_\epsilon) = \epsilon$, and suppose that there exists $\iota \in \Theta$ such that $L(\iota) = 1$ and $\theta_\epsilon + c\iota \in \Theta$ for all c in a neighborhood of zero. Then, $\rho(\cdot)$ is differentiable at ϵ with $\rho'(\epsilon) = \frac{1}{(\boldsymbol{w}^*)'\boldsymbol{m}(\iota)}$.*

Proof. First, I show that $\rho(\cdot)$ is concave on (ϵ_1, ϵ_2), where $\epsilon_1 = \inf\{(\boldsymbol{w}^*)'\boldsymbol{m}(\theta) : \theta \in \Theta\}$ and $\epsilon_2 = \sup\{(\boldsymbol{w}^*)'\boldsymbol{m}(\theta) : \theta \in \Theta\}$. Pick any $\epsilon, \epsilon' \in (\epsilon_1, \epsilon_2)$, and let $\{\theta_{\epsilon,n}\}_{n=1}^{\infty}$ and $\{\theta_{\epsilon',n}\}_{n=1}^{\infty}$ be sequences in Θ such that $(\boldsymbol{w}^*)'\boldsymbol{m}(\theta_{\epsilon,n}) = \epsilon$ and $(\boldsymbol{w}^*)'\boldsymbol{m}(\theta_{\epsilon',n}) = \epsilon'$ for all $n \geq 1$ and that $\lim_{n\to\infty} L(\theta_{\epsilon,n}) = \rho(\epsilon)$ and $\lim_{n\to\infty} L(\theta_{\epsilon',n}) = \rho(\epsilon')$. Then, for each $\lambda \in [0,1]$, $\lambda\theta_{\epsilon,n} + (1-\lambda)\theta_{\epsilon',n} \in \Theta$ by the convexity of Θ, and $(\boldsymbol{w}^*)'\boldsymbol{m}(\lambda\theta_{\epsilon,n} + (1-\lambda)\theta_{\epsilon',n}) = \lambda\epsilon + (1-\lambda)\epsilon'$ so that

$$\rho(\lambda\epsilon + (1-\lambda)\epsilon') \geq L(\lambda\theta_{\epsilon,n} + (1-\lambda)\theta_{\epsilon',n})$$

by the definition of ρ. Taking the limit of the right-hand side as $n \to \infty$ gives

$$\rho(\lambda\epsilon + (1-\lambda)\epsilon') \geq \lambda\rho(\epsilon) + (1-\lambda)\rho(\epsilon').$$

Therefore, $\rho(\cdot)$ is concave.

Since $\rho(\cdot)$ is concave, the superdifferential of $\rho(\cdot)$ at ϵ,

$$\partial\rho(\epsilon) = \{d : \rho(\eta) \leq \rho(\epsilon) + d(\eta - \epsilon) \text{ for all } \eta \in \mathbb{R}\},$$

is nonempty for all $\epsilon \in (\epsilon_1, \epsilon_2)$.

Now, let θ_ϵ satisfy $L(\theta_\epsilon) = \rho(\epsilon)$ and $(\boldsymbol{w}^*)'\boldsymbol{m}(\theta_\epsilon) = \epsilon$ for some ϵ, and suppose that there exists $\iota \in \Theta$ such that $L(\iota) = 1$ and $\theta_\epsilon + c\iota \in \Theta$ for all c in a neighborhood of zero. Then, for any $d \in \partial\rho(\epsilon)$ and for any c in a neighborhood of zero such that $\theta_\epsilon + c\iota \in \Theta$,

$$\rho(\epsilon) + d[(\boldsymbol{w}^*)'\boldsymbol{m}(\theta_\epsilon + c\iota) - \epsilon] \geq \rho((\boldsymbol{w}^*)'\boldsymbol{m}(\theta_\epsilon + c\iota)) \geq L(\theta_\epsilon + c\iota) = L(\theta_\epsilon) + c = \rho(\epsilon) + c,$$

where the first inequality follows since $d \in \partial\rho(\epsilon)$, and the second inequality follows from the definition of ρ. Since $(\boldsymbol{w}^*)'\boldsymbol{m}(\theta_\epsilon + c\iota) = \epsilon + c(\boldsymbol{w}^*)'\boldsymbol{m}(\iota)$, it follows that $cd(\boldsymbol{w}^*)'\boldsymbol{m}(\iota) \geq c$ for all c in a neighborhood of zero. This implies that $d(\boldsymbol{w}^*)'\boldsymbol{m}(\iota) = 1$. The result then follows. □

A.4 Differentiability of $\omega(\cdot)$ and $\rho(\cdot)$ for Example in Section 4

I apply Lemma A.1 to show the differentiability of $\omega(\cdot)$. Consider the problem (4). There exists a solution to this problem for any $\epsilon > 0$, since the objective is continuous, and the set of the vectors of $2n$ unknowns that satisfy the constraints is closed and bounded. The norm constraint must hold with equality, for otherwise we can increase the objective by increasing $f(x_i, 1)$ for all i by a small amount. The differentiability of $\omega(\cdot)$ then follows from Lemma A.1.

I show the differentiability of $\rho(\cdot)$ at any ϵ by deriving its closed-form expression. Observe that $\rho(\epsilon)$ is obtained by solving

$$\max_{(f(x_i,0),f(x_i,1))_{i=1,\dots,n}\in\mathbb{R}^{2n}} \frac{1}{n}\sum_{i=1}^{n}\mathbf{1}\{c_1\le x_i<c_0\}[f(x_i,1)-f(x_i,0)] \tag{A.1}$$

$$s.t.\quad \sum_{i=1}^{n} w_i^*\frac{f(x_i,d_i)}{\sigma(x_i,d_i)}=\epsilon,\ f(x_i,d)-f(x_j,d)\le C|x_i-x_j|,\quad d\in\{0,1\},i,j\in\{1,...,n\},$$

where w_i^* is the ith element of $\boldsymbol{w}^*$ given in Section 4. By the same logic explained in Section 4 for the problem for $\omega(\epsilon)$, it is sufficient to check the Lipschitz constraint among x's in the sample.

The constraint $\sum_{i=1}^{n} w_i^*\frac{f(x_i,d_i)}{\sigma(x_i,d_i)}=\epsilon$ is equivalent to

$$-\sum_{i:c_1\le x_i<c_0} f(x_i,0)+\tilde{n}f(x_{+,\min},1)=\bar{\sigma}\epsilon, \tag{A.2}$$

where $\bar{\sigma}=(\tilde{n}^2\sigma_{+,\min}^2+\sum_{i:c_1\le x_i<c_0}\sigma^2(x_i,0))^{1/2}$. The objective of (A.1) and the constraint (A.2) depend only on $((f(x_i,0),f(x_i,1))_{i:c_1\le x_i<0},f(x_{+,\min},1))$, so we can set the other values arbitrarily as long as the Lipschitz constraint holds. Also, given a value of $f(x_{+,\min},1)$, the objective of (A.1) is maximized only when $f(x_i,1)=C(x_{+,\min}-x_i)+f(x_{+,\min},1)$ for all i with $c_1\le x_i<c_0$ under the Lipschitz constraint. Therefore, the problem (A.1) reduces to

$$\begin{aligned}\max_{((f(x_i,0))_{i:c_1\le x_i<c_0},f(x_{+,\min},1))\in\mathbb{R}^{\tilde{n}+1}} &\frac{1}{n}\sum_{i=1}^{n}\mathbf{1}\{c_1\le x_i<c_0\}[C(x_{+,\min}-x_i)\\ &\qquad\qquad +f(x_{+,\min},1)-f(x_i,0)]\\ s.t. -\sum_{i:c_1\le x_i<c_0} &f(x_i,0)+\tilde{n}f(x_{+,\min},1)=\bar{\sigma}\epsilon.\end{aligned}$$

By plugging the constraint to the objective, we obtain that the maximized value is

$$\rho(\epsilon)=C\frac{1}{n}\sum_{i=1}^{n}\mathbf{1}\{c_1\le x_i<c_0\}[x_{+,\min}-x_i]+\frac{\bar{\sigma}\epsilon}{n}.$$

Therefore, $\rho(\cdot)$ is differentiable for all $\epsilon\in\mathbb{R}$.

A.5 Linear Minimax MSE Estimator and Optimal Bias-Variance Tradeoff

Donoho (1994) defines the modulus of continuity as $\tilde{\omega}(\epsilon)=\sup\{|L(\theta)-L(\tilde{\theta})|:\|\boldsymbol{m}(\theta-\tilde{\theta})\|\le\epsilon,\theta,\tilde{\theta}\in\Theta\}$. I first discuss some relationships between this definition and the definition in this paper. If Θ is convex and centrosymmetric, the relationship $\tilde{\omega}(\epsilon)=2\omega(\epsilon/2)$ holds. Also, if $\tilde{\omega}(\cdot)$ is differentiable, $\tilde{\omega}'(\epsilon)=\omega'(\epsilon/2)$. Let $(-\tilde{\theta}_{\tilde{\epsilon}},\tilde{\theta}_{\tilde{\epsilon}})$ solve the modulus problem $\sup\{|L(\theta)-L(\tilde{\theta})|:\|\boldsymbol{m}(\theta-\tilde{\theta})\|\le\tilde{\epsilon},\theta,\tilde{\theta}\in\Theta\}$, i.e., $\tilde{\omega}(\tilde{\epsilon})=2L(\tilde{\theta}_{\tilde{\epsilon}})$ and $2\|\boldsymbol{m}(\tilde{\theta}_{\tilde{\epsilon}})\|\le\tilde{\epsilon}$. Note that $\tilde{\theta}_{\tilde{\epsilon}}$ solves $\sup\{2L(\theta):\|2\boldsymbol{m}(\theta)\|\le\tilde{\epsilon},\theta\in\Theta\}$, or $\sup\{L(\theta):\|\boldsymbol{m}(\theta)\|\le\tilde{\epsilon}/2,\theta\in\Theta\}$, so that $\tilde{\theta}_{\tilde{\epsilon}}=\theta_{\tilde{\epsilon}/2}$, where

θ_ϵ solves $\sup\{L(\theta) : \|\boldsymbol{m}(\theta)\| \le \epsilon, \theta \in \Theta\}$ as in the main text.

Linear Minimax MSE Estimators. Let $\tilde{\epsilon}_{\text{MSE}}$ solve

$$\frac{(\epsilon/2)^2}{(\epsilon/2)^2 + \sigma^2} = \frac{\epsilon\tilde{\omega}'(\epsilon)}{\tilde{\omega}(\epsilon)}.$$

The linear minimax MSE estimator of $L(\theta)$ is then given by $\hat{L}_{\text{MSE}}(\boldsymbol{Y}) = \boldsymbol{w}_{\text{MSE}}'\boldsymbol{Y}$ (Donoho, 1994), where

$$\boldsymbol{w}_{\text{MSE}} = \frac{2\tilde{\omega}'(\tilde{\epsilon}_{\text{MSE}})\boldsymbol{m}(\tilde{\theta}_{\tilde{\epsilon}_{\text{MSE}}})}{\tilde{\epsilon}_{\text{MSE}}}.$$

Now, let $\epsilon_{\text{MSE}} = \tilde{\epsilon}_{\text{MSE}}/2$, so that

$$\frac{(\epsilon_{\text{MSE}})^2}{(\epsilon_{\text{MSE}})^2 + \sigma^2} = \frac{2\epsilon_{\text{MSE}}\tilde{\omega}'(2\epsilon_{\text{MSE}})}{\tilde{\omega}(2\epsilon_{\text{MSE}})},$$

which is equivalent to

$$\frac{(\epsilon_{\text{MSE}})^2}{(\epsilon_{\text{MSE}})^2 + \sigma^2} = \frac{\epsilon_{\text{MSE}}\omega'(\epsilon_{\text{MSE}})}{\omega(\epsilon_{\text{MSE}})}.$$

We also have

$$\boldsymbol{w}_{\text{MSE}} = \frac{2\tilde{\omega}'(2\epsilon_{\text{MSE}})\boldsymbol{m}(\tilde{\theta}_{2\epsilon_{\text{MSE}}})}{2\epsilon_{\text{MSE}}} = \frac{\omega'(\epsilon_{\text{MSE}})\boldsymbol{m}(\theta_{\epsilon_{\text{MSE}}})}{\epsilon_{\text{MSE}}}.$$

Optimal Bias-Variance Frontier. The optimal bias-variance frontier in estimation of $L(\theta)$ can be traced out by a class of linear estimators $\{\tilde{L}_{\tilde{\epsilon}}(\boldsymbol{Y})\}_{\tilde{\epsilon}>0}$, where for each $\tilde{\epsilon} > 0$,

$$\tilde{L}_{\tilde{\epsilon}}(\boldsymbol{Y}) = \frac{2\tilde{\omega}'(\tilde{\epsilon})\boldsymbol{m}(\tilde{\theta}_{\tilde{\epsilon}})'}{\tilde{\epsilon}}\boldsymbol{Y}.$$

The maximum bias of $\tilde{L}_{\tilde{\epsilon}}$ is

$$\overline{\text{Bias}}_\Theta(\tilde{L}_{\tilde{\epsilon}}(\boldsymbol{Y})) = \frac{1}{2}(\tilde{\omega}(\tilde{\epsilon}) - \tilde{\epsilon}\tilde{\omega}'(\tilde{\epsilon})),$$

and the variance is $(\sigma\tilde{\omega}'(\tilde{\epsilon}))^2$.

For each $\epsilon > 0$, let $\hat{L}_\epsilon(\boldsymbol{Y}) = \tilde{L}_{2\epsilon}(\boldsymbol{Y})$. Then,

$$\hat{L}_\epsilon(\boldsymbol{Y}) = \frac{2\tilde{\omega}'(2\epsilon)\boldsymbol{m}(\tilde{\theta}_{2\epsilon})'}{2\epsilon}\boldsymbol{Y} = \frac{\omega'(\epsilon)\boldsymbol{m}(\theta_\epsilon)'}{\epsilon}\boldsymbol{Y}.$$

Therefore, the class of estimators $\{\tilde{L}_{\tilde{\epsilon}}(\boldsymbol{Y})\}_{\tilde{\epsilon}>0}$ is the same as a class of linear estimators $\{\hat{L}_\epsilon(\boldsymbol{Y})\}_{\epsilon>0}$. The maximum bias of $\hat{L}_\epsilon$ is

$$\overline{\text{Bias}}_\Theta(\hat{L}_\epsilon(\boldsymbol{Y})) = \frac{1}{2}(\tilde{\omega}(2\epsilon) - 2\epsilon\tilde{\omega}'(2\epsilon)) = \omega(\epsilon) - \epsilon\omega'(\epsilon),$$

and the variance is $(\sigma\omega'(\epsilon))^2$.

Example A.1 (Eligibility Cutoff Choice (Cont.))**.** Consider the setup in Section 4. Let $\iota \in \mathcal{F}_{\text{Lip}}(C)$ such that $\iota(x, 0) = 0$ for all x and $\iota(x, 1) = \frac{n}{\tilde{n}}$. Then, $L(\iota) = 1$, and $f + c\iota \in \mathcal{F}_{\text{Lip}}(C)$ for all $c \in \mathbb{R}$ and $f \in \mathcal{F}_{\text{Lip}}(C)$. By Lemma A.1 in Appendix A.3, we obtain

$$\omega'(\epsilon; L, \tilde{\boldsymbol{m}}, \mathcal{F}_{\text{Lip}}(C)) = \frac{\epsilon}{\tilde{\boldsymbol{m}}(\iota)'\tilde{\boldsymbol{m}}(f_\epsilon)} = \frac{\epsilon}{\frac{n}{\tilde{n}}\sum_{i=1}^n d_i f_\epsilon(x_i, d_i)/\sigma^2(x_i, d_i)}, \tag{A.3}$$

where f_ϵ solves the modulus problem in this paper's definition. Then,

$$\hat{L}_\epsilon(\boldsymbol{Y}) = \sum_{i=1}^n \frac{f_\epsilon(x_i, d_i)/\sigma^2(x_i, d_i)}{\frac{n}{\tilde{n}}\sum_{j=1}^n d_i f_\epsilon(x_j, d_j)/\sigma^2(x_j, d_j)} Y_i.$$

The maximum bias of $\hat{L}_\epsilon$ is

$$\begin{aligned}&\overline{\text{Bias}}_{\mathcal{F}_{\text{Lip}}(C)}(\hat{L}_\epsilon(\boldsymbol{Y}))\\&= \frac{1}{n}\sum_{i=1}^n \mathbf{1}\{c_1 \le x_i < c_0\}[f_\epsilon(x_i, 1) - f_\epsilon(x_i, 0)] - \frac{\epsilon^2}{\frac{n}{\tilde{n}}\sum_{i=1}^n d_i f_\epsilon(x_i, d_i)/\sigma^2(x_i, d_i)},\end{aligned}$$

and the variance is $\text{Var}(\hat{L}_\epsilon(\boldsymbol{Y})) = \epsilon^2/\left(\frac{n}{\tilde{n}}\sum_{i=1}^n d_i f_\epsilon(x_i, d_i)/\sigma^2(x_i, d_i)\right)^2$.

□

A.6 Computing ϵ^* for Example in Section 4

Here, I provide a procedure for computing $\epsilon^* \in \arg\max_{\epsilon \in [0, a^*]} \omega(\epsilon)\Phi(-\epsilon)$ for the example of eligibility cutoff choice under the Lipschitz class.

The procedure is based on the first-order condition. By differentiating $\omega(\epsilon)\Phi(-\epsilon)$, we have

$$\omega'(\epsilon)\Phi(-\epsilon) - \omega(\epsilon)\phi(-\epsilon) = \left[\frac{1 - \Phi(\epsilon)}{\phi(\epsilon)} - \frac{\omega(\epsilon)}{\omega'(\epsilon)}\right]\omega'(\epsilon)\phi(\epsilon),$$

where the equality holds since $\Phi(x) = 1 - \Phi(-x)$ and $\phi(x) = \phi(-x)$. $\frac{1-\Phi(\epsilon)}{\phi(\epsilon)}$ is the Mills ratio of a standard normal variable, which is strictly decreasing in ϵ. Since $\omega(\epsilon)$ is nondecreasing and concave, $\frac{\omega(\epsilon)}{\omega'(\epsilon)}$ is nondecreasing in ϵ. Therefore, $\frac{1-\Phi(\epsilon)}{\phi(\epsilon)} - \frac{\omega(\epsilon)}{\omega'(\epsilon)}$ is strictly decreasing in ϵ.

I suggest using the following procedure to compute ϵ^*.

1. If $\frac{1-\Phi(a^*)}{\phi(a^*)} - \frac{\omega(a^*)}{\omega'(a^*)} > 0$, $\epsilon^* = a^*$.

2. If not, use the bisection method to find $\epsilon^* \in [0, a^*]$ that solves $\frac{1-\Phi(\epsilon)}{\phi(\epsilon)} - \frac{\omega(\epsilon)}{\omega'(\epsilon)} = 0$.

Note that, for each ϵ, once we solve the convex optimization problem (4) to compute $\omega(\epsilon)$ and $(f_\epsilon(x_i, 0), f_\epsilon(x_i, 1))$, $i = 1, ..., n$, we can compute $\omega'(\epsilon)$ using the closed-form expression (A.3) in Appendix A.5.

B Proofs

B.1 Auxiliary Lemmas

Lemma B.1. *Let $g(t) = h(t)\Phi\left(\frac{b-t}{a}\right)$, where $h(t)$ is nonconstant, nondecreasing, concave, and differentiable on $[\underline{t}, \bar{t}]$, $a > 0$, and $b \in \mathbb{R}$. If $a\frac{1-\Phi\left(\frac{\underline{t}-b}{a}\right)}{\phi\left(\frac{\underline{t}-b}{a}\right)} \leq \frac{h(\underline{t})}{h'(\underline{t})}$, then $g(t)$ is strictly decreasing on $[\underline{t}, \bar{t}]$. If $a\frac{1-\Phi\left(\frac{\bar{t}-b}{a}\right)}{\phi\left(\frac{\bar{t}-b}{a}\right)} \geq \frac{h(\bar{t})}{h'(\bar{t})}$, then $g(t)$ is strictly increasing on $[\underline{t}, \bar{t}]$. If $a\frac{1-\Phi\left(\frac{\underline{t}-b}{a}\right)}{\phi\left(\frac{\underline{t}-b}{a}\right)} > \frac{h(\underline{t})}{h'(\underline{t})}$ and $a\frac{1-\Phi\left(\frac{\bar{t}-b}{a}\right)}{\phi\left(\frac{\bar{t}-b}{a}\right)} < \frac{h(\bar{t})}{h'(\bar{t})}$, then there exists a unique $t^* \in [\underline{t}, \bar{t}]$ such that $g(t)$ is strictly increasing on $[\underline{t}, t^*)$ and strictly decreasing on $(t^*, \bar{t}]$. t^* is the solution to $a\frac{1-\Phi\left(\frac{t-b}{a}\right)}{\phi\left(\frac{t-b}{a}\right)} = \frac{h(t)}{h'(t)}$ if $h'(t)$ is continuous.*

Proof. Note first that $h'(\underline{t}) > 0$; if $h'(\underline{t}) \leq 0$, then $h'(t) = 0$ for all $t \in [\underline{t}, \bar{t}]$ since $h(t)$ is nondecreasing and concave, but this contradicts the assumption that $h(t)$ is nonconstant.

By differentiating $g(t)$, we have for $t \in [\underline{t}, \bar{t}]$,

$$g'(t) = h'(t)\Phi\left(\frac{b-t}{a}\right) - h(t)\phi\left(\frac{b-t}{a}\right)/a = \left[a\frac{1-\Phi\left(\frac{t-b}{a}\right)}{\phi\left(\frac{t-b}{a}\right)} - \frac{h(t)}{h'(t)}\right]h'(t)\phi\left(\frac{t-b}{a}\right)/a,$$

where the second equality holds since $\Phi(x) = 1 - \Phi(-x)$ and $\phi(x) = \phi(-x)$. By the fact that the Mills ratio $\frac{1-\Phi(x)}{\phi(x)}$ of a standard normal variable is strictly decreasing, $a\frac{1-\Phi\left(\frac{t-b}{a}\right)}{\phi\left(\frac{t-b}{a}\right)}$ is strictly decreasing in t. In addition, $a\frac{1-\Phi\left(\frac{t-b}{a}\right)}{\phi\left(\frac{t-b}{a}\right)}$ is continuous. Furthermore, since $h(t)$ is nondecreasing and concave on $[\underline{t}, \bar{t}]$, $\frac{h(t)}{h'(t)}$ is nondecreasing on $[\underline{t}, \bar{t}]$. Therefore, if $a\frac{1-\Phi\left(\frac{\underline{t}-b}{a}\right)}{\phi\left(\frac{\underline{t}-b}{a}\right)} \leq \frac{h(\underline{t})}{h'(\underline{t})}$, then $g'(t) < 0$ for all $t \in (\underline{t}, \bar{t}]$. If $a\frac{1-\Phi\left(\frac{\bar{t}-b}{a}\right)}{\phi\left(\frac{\bar{t}-b}{a}\right)} \geq \frac{h(\bar{t})}{h'(\bar{t})}$, then $g'(t) > 0$ for all $t \in [\underline{t}, \bar{t})$. If $a\frac{1-\Phi\left(\frac{\underline{t}-b}{a}\right)}{\phi\left(\frac{\underline{t}-b}{a}\right)} > \frac{h(\underline{t})}{h'(\underline{t})}$ and $a\frac{1-\Phi\left(\frac{\bar{t}-b}{a}\right)}{\phi\left(\frac{\bar{t}-b}{a}\right)} < \frac{h(\bar{t})}{h'(\bar{t})}$, then $g'(t) > 0$ for $t \in [\underline{t}, t^*)$ and $g'(t) < 0$ for $t \in (t^*, \bar{t}]$, where $t^* = \sup\{t \in [\underline{t}, \bar{t}] : g'(t) \geq 0\}$. t^* is the solution to $a\frac{1-\Phi\left(\frac{t-b}{a}\right)}{\phi\left(\frac{t-b}{a}\right)} = \frac{h(t)}{h'(t)}$ if $h'(t)$ is continuous. The conclusion then follows. □

Lemma B.2. *Suppose that the conditions in Theorem 2 hold, and let $\Theta_{\epsilon^*} = \arg\max_{\theta\in\Theta:\|\boldsymbol{m}(\theta)\|\leq\epsilon^*} L(\theta)$. Then, $\|\boldsymbol{m}(\theta)\| = \epsilon^*$ for any $\theta \in \Theta_{\epsilon^*}$, and $\boldsymbol{m}(\theta) = \boldsymbol{m}(\tilde{\theta})$ for any $\theta, \tilde{\theta} \in \Theta_{\epsilon^*}$.*

Proof. First, pick any $\theta \in \Theta_{\epsilon^*}$. Since θ attains the modulus of continuity at ϵ^*, it also attains the modulus at $\|\boldsymbol{m}(\theta)\|$, so that $\omega(\|\boldsymbol{m}(\theta)\|) = \omega(\epsilon^*)$. It follows that $\|\boldsymbol{m}(\theta)\| = \epsilon^*$, since if $\|\boldsymbol{m}(\theta)\| < \epsilon^*$,

$$\omega(\|\boldsymbol{m}(\theta)\|)\Phi(-\|\boldsymbol{m}(\theta)\|/\sigma) = \omega(\epsilon^*)\Phi(-\|\boldsymbol{m}(\theta)\|/\sigma) > \omega(\epsilon^*)\Phi(-\epsilon^*/\sigma),$$

which contradicts the assumption that ϵ^* maximizes $\omega(\epsilon)\Phi(-\epsilon/\sigma)$ over $[0, a^*\sigma]$.

Now, pick any $\theta, \tilde{\theta} \in \Theta_{\epsilon^*}$. By the above argument, $\|\boldsymbol{m}(\theta)\| = \|\boldsymbol{m}(\tilde{\theta})\| = \epsilon^*$, and hence $\boldsymbol{m}(\theta), \boldsymbol{m}(\tilde{\theta}) \in \{\boldsymbol{\beta} \in \mathbb{R}^n : \|\boldsymbol{\beta}\| \leq \epsilon^*\}$. Suppose that $\boldsymbol{m}(\theta) \neq \boldsymbol{m}(\tilde{\theta})$, and let $\bar{\theta} = \lambda\theta + (1-\lambda)\tilde{\theta}$ for some $\lambda \in (0,1)$. Then $L(\bar{\theta}) = \lambda L(\theta) + (1-\lambda)L(\tilde{\theta}) = \omega(\epsilon^*)$. By the convexity of Θ, $\bar{\theta} \in \Theta$. Furthermore, since $\{\boldsymbol{\beta} \in \mathbb{R}^n : \|\boldsymbol{\beta}\| \leq \epsilon^*\}$ is strictly convex, $\boldsymbol{m}(\bar{\theta}) = \lambda\boldsymbol{m}(\theta) + (1-\lambda)\boldsymbol{m}(\tilde{\theta})$ is an interior point of $\{\boldsymbol{\beta} \in \mathbb{R}^n : \|\boldsymbol{\beta}\| \leq \epsilon^*\}$, which implies that $\|\boldsymbol{m}(\bar{\theta})\| < \epsilon^*$. Thus, $\bar{\theta}$ attains the modulus at ϵ^*, but then it must be the case that $\|\boldsymbol{m}(\bar{\theta})\| = \epsilon^*$. □

Lemma B.3. *Let $\psi(a,b) = a\Phi(-b)$. Then, $\psi(a,b)$ is strictly quasi-concave on $(0,\infty) \times \mathbb{R}$.*

Proof. Take any $a_0, a_1 > 0$ and $b_0, b_1 \in \mathbb{R}$ such that $(a_0, b_0) \neq (a_1, b_1)$. I show that $\psi(a_0 + \lambda(a_1 - a_0), b_0 + \lambda(b_1 - b_0)) > \min\{\psi(a_0, b_0), \psi(a_1, b_1)\}$ for all $\lambda \in (0,1)$.

First, suppose that $a_0 \leq a_1$ and $b_0 \geq b_1$. Since either $a_0 < a_1$ or $b_0 > b_1$ or both must hold, $\psi(a_0 + \lambda(a_1 - a_0), b_0 + \lambda(b_1 - b_0)) = (a_0 + \lambda(a_1 - a_0))\Phi(-b_0 - \lambda(b_1 - b_0))$ is strictly increasing in λ. It then follows that $\psi(a_0 + \lambda(a_1 - a_0), b_0 + \lambda(b_1 - b_0)) > \psi(a_0, b_0)$. Likewise, if $a_0 \geq a_1$ and $b_0 \leq b_1$, then $\psi(a_0 + \lambda(a_1 - a_0), b_0 + \lambda(b_1 - b_0)) > \psi(a_1, b_1)$.

Now suppose that $a_0 < a_1$ and $b_0 < b_1$. Note that the set $\{(a_0, b_0) + \lambda(a_1 - a_0, b_1 - b_0) : \lambda \in (0,1)\}$ is equivalent to

$$\left\{\left(0, b_0 - a_0\frac{b_1 - b_0}{a_1 - a_0}\right) + t\left(\frac{a_1 - a_0}{b_1 - b_0}, 1\right) : t \in \left(a_0\frac{b_1 - b_0}{a_1 - a_0}, a_1\frac{b_1 - b_0}{a_1 - a_0}\right)\right\}.$$

We have

$$\begin{aligned}\psi\left(\left(0, b_0 - a_0\frac{b_1 - b_0}{a_1 - a_0}\right) + t\left(\frac{a_1 - a_0}{b_1 - b_0}, 1\right)\right) &= t\left(\frac{a_1 - a_0}{b_1 - b_0}\right)\Phi\left(-b_0 + a_0\frac{b_1 - b_0}{a_1 - a_0} - t\right) \\ &= \left(\frac{a_1 - a_0}{b_1 - b_0}\right) g(t),\end{aligned}$$

where $g(t) = t\Phi\left(-b_0 + a_0\frac{b_1 - b_0}{a_1 - a_0} - t\right)$. Lemma B.1 implies that the minimum of $g(t)$ over an interval $[t_0, t_1]$ is attained only at t_0 or t_1 or both. Hence, for all $t \in \left(a_0\frac{b_1 - b_0}{a_1 - a_0}, a_1\frac{b_1 - b_0}{a_1 - a_0}\right)$,

$$g(t) > \min\left\{g\left(a_0\frac{b_1 - b_0}{a_1 - a_0}\right), g\left(a_1\frac{b_1 - b_0}{a_1 - a_0}\right)\right\}.$$

Thus, for all $t \in \left(a_0\frac{b_1 - b_0}{a_1 - a_0}, a_1\frac{b_1 - b_0}{a_1 - a_0}\right)$,

$$\begin{aligned}&\psi\left(\left(0, b_0 - a_0\frac{b_1 - b_0}{a_1 - a_0}\right) + t\left(\frac{a_1 - a_0}{b_1 - b_0}, 1\right)\right) \\ &> \left(\frac{a_1 - a_0}{b_1 - b_0}\right)\min\left\{g\left(a_0\frac{b_1 - b_0}{a_1 - a_0}\right), g\left(a_1\frac{b_1 - b_0}{a_1 - a_0}\right)\right\} \\ &= \min\{\psi(a_0, b_0), \psi(a_1, b_1)\}.\end{aligned}$$

Therefore, $\psi(a_0+\lambda(a_1-a_0), b_0+\lambda(b_1-b_0)) > \min\{\psi(a_0,b_0),\psi(a_1,b_1)\}$ for all $\lambda \in (0,1)$. The same argument holds for the case where $a_0 > a_1$ and $b_0 > b_1$. □

B.2 Proof of Proposition 1

The problem (4) is equivalent to

$$\max_{(f(x_i,0),f(x_i,1))_{i=1,...,n}\in\mathbb{R}^{2n}} \frac{1}{n}\sum_{i:c_1\le x_i<c_0}[f(x_i,1)-f(x_i,0)] \tag{B.1}$$

$$s.t. \sum_{i:x_i<c_1}\frac{f(x_i,0)^2}{\sigma^2(x_i,0)} + \sum_{i:c_1\le x_i<c_0}\frac{f(x_i,0)^2}{\sigma^2(x_i,0)} + \frac{f(x_{+,\min},1)^2}{\sigma^2_{+,\min}} + \sum_{i:x_i>x_{+,\min}}\frac{f(x_i,1)^2}{\sigma^2(x_i,1)} \le \epsilon^2, \tag{B.2}$$

$$f(x_i,d)-f(x_j,d)\le C|x_i-x_j|,\quad d\in\{0,1\}, i,j\in\{1,...,n\}. \tag{B.3}$$

First, consider the case where $\epsilon = 0$. Since the left-hand side of (B.2) must be zero, any solution satisfies $f(x_i,0)=0$ if $x_i<c_0$ and $f(x_i,1)=0$ if $x_i\ge c_0$. The objective (B.1) and constraint (B.2) do not depend on $(f(x_i,0))_{i:x_i\ge c_0}$, so we can set these values arbitrarily as long as the Lipschitz constraint (B.3) holds. I set all of them to 0, so that $f(x_i,0)=0$ for all i. The above problem then reduces to solving

$$\max_{(f(x_i,1))_{i:x_i<c_0}\in\mathbb{R}^n} \frac{1}{n}\sum_{i:c_1\le x_i<c_0} f(x_i,1) \tag{B.4}$$

$$s.t.\ \ f(x_i,1)-f(x_j,1)\le C|x_i-x_j|,\quad i,j\in\{1,...,n\}, \tag{B.5}$$

where $f(x_i,1)=0$ for any i with $x_i\ge c_0$. The Lipschitz constraint (B.5) implies that $f(x_i,1)\le C(x_{+,\min}-x_i)$ for any i with $c_1\le x_i<c_0$. Therefore, the value of the objective (B.4) is at most $\frac{1}{n}\sum_{i:c_1\le x_i<c_0} C(x_{+,\min}-x_i)$. This value is attained by setting $f(x_i,1)=C(x_{+,\min}-x_i)$ for all i with $x_i<c_0$, which satisfies the Lipschitz constraint (B.5). In sum, when $\epsilon=0$, one solution to problem (B.1)–(B.3) is given by

$$f_0(x_i,0)=0,\ \ i=1,...,n,\quad f_0(x_i,1)=\begin{cases}0 & \text{if } x_i>x_{+,\min},\\ C(x_{+,\min}-x_i) & \text{if } x_i\le x_{+,\min}.\end{cases}$$

Now, let $\bar\epsilon=(C/\bar\sigma)\min_{x,\tilde x\in\mathcal{X},x\ne\tilde x}|x-\tilde x|$, where $\bar\sigma=\max_i\sigma(x_i,d_i)$ and $\mathcal{X}=\{x_i: i=1,...,n\}$ is the set of points of x in the sample. Consider any $\epsilon\in(0,\bar\epsilon]$. I claim that problem (B.1)–(B.3) reduces to solving

$$\max_{((f(x_i,0),f(x_i,1))_{i:c_1\leq x_i<c_0},f(x_{+,\min},1))\in\mathbb{R}^{2\tilde{n}+1}} \frac{1}{n}\sum_{i:c_1\leq x_i<c_0}[f(x_i,1)-f(x_i,0)] \tag{B.6}$$

$$s.t. \quad \sum_{i:c_1\leq x_i<c_0}\frac{f(x_i,0)^2}{\sigma^2(x_i,0)}+\frac{f(x_{+,\min},1)^2}{\sigma^2_{+,\min}}\leq \epsilon^2, \tag{B.7}$$

$$f(x_i,1)-f(x_j,1)\leq C|x_i-x_j|, \quad i,j\in\{k: c_1\leq x_k\leq x_{+,\min}\}. \tag{B.8}$$

To see this, suppose that $((f_\epsilon(x_i,0),f_\epsilon(x_i,1))_{i:c_1\leq x_i<c_0},f_\epsilon(x_{+,\min},1))$ is a solution to the above problem. First, it must be the case that $f_\epsilon(x_i,1)\geq 0$ for any i with $c_1\leq x_i<c_0$, since if $f_\epsilon(x_i,1)<0$ for some i with $c_1\leq x_i<c_0$, it is possible to strictly increase the objective (B.6) without violating the constraints (B.7) and (B.8) by changing f_ϵ to $\tilde{f}_\epsilon$ such that $\tilde{f}_\epsilon(x_i,1)=\max\{f(x_i,1),0\}$ for any i with $c_1\leq x_i<c_0$. By similar arguments, it must be the case that $f_\epsilon(x_{+,\min},1)\geq 0$ and $f_\epsilon(x_i,0)\leq 0$ for i with $c_1\leq x_i<c_0$.

Next, given $((f_\epsilon(x_i,0),f_\epsilon(x_i,1))_{i:c_1\leq x_i<c_0},f_\epsilon(x_{+,\min},1))$, set

$$f_\epsilon(x_i,0)=0 \quad \text{if } x_i<c_1 \text{ or } x_i\geq c_0, \tag{B.9}$$

$$f_\epsilon(x_i,1)=\begin{cases}0 & \text{if } x_i>x_{+,\min},\\ C(x_{-,\min}-x_i)+f_\epsilon(x_{-,\min},1) & \text{if } x_i<c_1,\end{cases} \tag{B.10}$$

where $x_{-,\min}=\min\{x_i: c_1\leq x_i<c_0\}$ is the smallest values of x among those whose treatment status would be changed if the cutoff were changed to c_1 in the sample. I show that $(f_\epsilon(x_i,0),f_\epsilon(x_i,1))_{i=1,\ldots,n}$ is a solution to the original problem (B.1)–(B.3). Clearly, it satisfies the constraint (B.2). To see that the Lipschitz constraint (B.3) is satisfied for $d=0$, it suffices to check that $|f_\epsilon(x_i,0)-f_\epsilon(x_j,0)|\leq C|x_i-x_j|$ for any i,j with $x_i,x_j\leq x_{+,\min}$, given that the Lipschitz constraint for any i,j with $x_i,x_j\geq x_{+,\min}$ holds by construction. Observe that for any i,j with $x_i,x_j\leq x_{+,\min}$ and $x_i\neq x_j$,

$$\begin{aligned}|f_\epsilon(x_i,0)-f_\epsilon(x_j,0)|^2 &= f_\epsilon(x_i,0)^2+f_\epsilon(x_j,0)^2-2f_\epsilon(x_i,0)f_\epsilon(x_j,0)\\ &\leq f_\epsilon(x_i,0)^2+f_\epsilon(x_j,0)^2\\ &\leq \bar{\sigma}^2\bar{\epsilon}^2\\ &= C^2\min_{x,\tilde{x}\in\mathcal{X},x\neq\tilde{x}}|x-\tilde{x}|^2\\ &\leq C^2|x_i-x_j|^2,\end{aligned}$$

where the inequality in the second line holds since $f_\epsilon(x_i,0)\leq 0$ for all i, the inequality in the third line follows from the constraint (B.7), and the equality in the fourth line from the definition of $\bar{\epsilon}$. For $d=1$, it is sufficient to check that $|f_\epsilon(x_i,1)-f_\epsilon(x_{+,\min},1)|\leq C(x_i-x_{+,\min})$ for any i with $x_i>x_{+,\min}$ and that $|f_\epsilon(x_i,1)-f_\epsilon(x_{-,\min},1)|\leq C(x_{-,\min}-x_i)$ for any i with $x_i<c_1$.

The latter immediately follows from the construction of f_ϵ. Regarding the former, for any i with $x_i > x_{+,\min}$,

$$|f_\epsilon(x_i,1) - f_\epsilon(x_{+,\min},1)|^2 = f_\epsilon(x_{+,\min},1)^2 \le \bar{\sigma}^2\bar{\epsilon}^2 \le C^2|x_i - x_{+,\min}|^2.$$

Therefore, $(f_\epsilon(x_i,0), f_\epsilon(x_i,1))_{i=1,\ldots,n}$ satisfies constraint (B.3). Since the value of the original problem (B.1)–(B.3) at $(f_\epsilon(x_i,0), f_\epsilon(x_i,1))_{i=1,\ldots,n}$ is equal to the maximized value of the less constrained problem (B.6)–(B.8), $(f_\epsilon(x_i,0), f_\epsilon(x_i,1))_{i=1,\ldots,n}$ is the solution to problem (B.1)–(B.3).

Now I derive a solution to (B.6)–(B.8). Note that, given a value of $f(x_{+,\min},1)$, the objective is maximized only when $f(x_i,1) = C(x_{+,\min} - x_i) + f(x_{+,\min},1)$ under the constraints (B.7) and (B.8). Plugging this into (B.6)–(B.8), one can further simplify the problem to

$$\max_{((f(x_i,0))_{i:c_1\le x_i<c_0}, f(x_{+,\min},1))\in\mathbb{R}^{\tilde{n}+1}} \frac{C}{n}\sum_{i:c_1\le x_i<c_0}(x_{+,\min} - x_i) + \frac{\tilde{n}}{n}f(x_{+,\min},1) - \frac{1}{n}\sum_{i:c_1\le x_i<c_0} f(x_i,0)$$

$$s.t. \quad \sum_{i:c_1\le x_i<c_0}\frac{f(x_i,0)^2}{\sigma^2(x_i,0)} + \frac{f(x_{+,\min},1)^2}{\sigma^2_{+,\min}} \le \epsilon^2.$$

This is a convex optimization problem that maximizes a weighted sum of $\tilde{n}+1$ unknowns under the constraint on the upper bound on a weighted Euclidean norm of the unknowns. Simple calculations show that the solution is given by $f(x_i,0) = -\frac{\sigma^2(x_i,0)\epsilon}{(\tilde{n}^2\sigma^2_{+,\min} + \sum_{i:c_1\le x_i<c_0}\sigma^2(x_i,0))^{1/2}}$ for any i with $c_1 \le x_i < c_0$ and $f(x_{+,\min},1) = \frac{\tilde{n}\sigma^2_{+,\min}\epsilon}{(\tilde{n}^2\sigma^2_{+,\min} + \sum_{i:c_1\le x_i<c_0}\sigma^2(x_i,d_i))^{1/2}}$. From (B.9) and (B.10), one solution to the original problem (B.1)–(B.3) is then given by

$$f_\epsilon(x_i,0) = \begin{cases} 0 & \text{if } x_i < c_1 \text{ or } x_i \ge c_0, \\ -\frac{\sigma^2(x_i,0)\epsilon}{(\tilde{n}^2\sigma^2_{+,\min} + \sum_{i:c_1\le x_i<c_0}\sigma^2(x_i,0))^{1/2}} & \text{if } c_1 \le x_i < c_0, \end{cases}$$

$$f_\epsilon(x_i,1) = \begin{cases} 0 & \text{if } x_i > x_{+,\min}, \\ C(x_{+,\min} - x_i) + \frac{\tilde{n}\sigma^2_{+,\min}\epsilon}{(\tilde{n}^2\sigma^2_{+,\min} + \sum_{i:c_1\le x_i<c_0}\sigma^2(x_i,0))^{1/2}} & \text{if } x_i \le x_{+,\min}. \end{cases}$$

The modulus of continuity is given by

$$\omega(\epsilon) = C\frac{1}{n}\sum_{i=1}^n \mathbf{1}\{c_1 \le x_i < c_0\}[x_{+,\min} - x_i] + \frac{1}{n}\left(\tilde{n}^2\sigma^2_{+,\min} + \sum_{i:c_1\le x_i<c_0}\sigma^2(x_i,0)\right)^{1/2}\epsilon.$$

□

B.3 Proof of Proposition 2

I first show that $\sigma\frac{1-\Phi(\epsilon/\sigma)}{\phi(\epsilon/\sigma)} \geq \frac{\omega(\epsilon)}{\omega'(\epsilon)}$ for all $\epsilon \leq \epsilon^*$. From the arguments in the proof of Lemma 1, if $\sigma\frac{1-\Phi(a^*)}{\phi(a^*)} \geq \frac{\omega(a^*\sigma)}{\omega'(a^*\sigma)}$, then $\epsilon^* = a^*\sigma$, and the above statement holds. Suppose that $\sigma\frac{1-\Phi(a^*)}{\phi(a^*)} < \frac{\omega(a^*\sigma)}{\omega'(a^*\sigma)}$. Again from the arguments in the proof of Lemma 1, $\sigma\frac{1-\Phi(\epsilon/\sigma)}{\phi(\epsilon/\sigma)} > \frac{\omega(\epsilon)}{\omega'(\epsilon)}$ for $\epsilon < \epsilon^*$ and $\sigma\frac{1-\Phi(\epsilon/\sigma)}{\phi(\epsilon/\sigma)} < \frac{\omega(\epsilon)}{\omega'(\epsilon)}$ for $\epsilon > \epsilon^*$. Since the left-hand side is continuous, and the curve $(\epsilon, \frac{\omega(\epsilon)}{\omega'(\epsilon)})_{\epsilon>0}$ is connected by Lemma 3 of Donoho (1994), $\sigma\frac{1-\Phi(\epsilon^*/\sigma)}{\phi(\epsilon^*/\sigma)} = \frac{\omega(\epsilon^*)}{\omega'(\epsilon^*)}$. This implies that $\sigma\frac{1-\Phi(\epsilon/\sigma)}{\phi(\epsilon/\sigma)} \geq \frac{\omega(\epsilon)}{\omega'(\epsilon)}$ for all $\epsilon \leq \epsilon^*$.

Now, note that ϵ_{MSE} solves $\frac{\epsilon^2+\sigma^2}{\epsilon} = \frac{\omega(\epsilon)}{\omega'(\epsilon)}$. If $\frac{\epsilon^2+\sigma^2}{\epsilon} > \sigma\frac{1-\Phi(\epsilon/\sigma)}{\phi(\epsilon/\sigma)}$ for all $\epsilon > 0$, then $\frac{\epsilon^2+\sigma^2}{\epsilon} > \frac{\omega(\epsilon)}{\omega'(\epsilon)}$ for all $\epsilon \leq \epsilon^*$, which implies that $\epsilon^* < \epsilon_{\text{MSE}}$. Below, I show that $\frac{\epsilon^2+\sigma^2}{\epsilon} > \sigma\frac{1-\Phi(\epsilon/\sigma)}{\phi(\epsilon/\sigma)}$ for all $\epsilon > 0$. Let $g(\epsilon) = \frac{\epsilon^2+\sigma^2}{\epsilon}$. We have $g'(\epsilon) = 1-\frac{\sigma^2}{\epsilon^2}$, which is strictly increasing in ϵ. Therefore, $g(\epsilon)$ is minimized at $\epsilon = \sigma$, at which $g'(\epsilon) = 0$. For any $\epsilon > 0$, $g(\epsilon) \geq g(\sigma) = 2\sigma > \sigma\frac{1-\Phi(0)}{\phi(0)} > \sigma\frac{1-\Phi(\epsilon/\sigma)}{\phi(\epsilon/\sigma)}$, where the second inequality holds since $\frac{1-\Phi(0)}{\phi(0)} \approx 1.253$, and the last holds since $\frac{1-\Phi(\epsilon/\sigma)}{\phi(\epsilon/\sigma)}$ is strictly decreasing. □

B.4 Proof of Corollary 1

I first show that $\omega(\epsilon) = L(\epsilon\theta^*)$ for any $\epsilon \in (0, a^*\sigma]$. Since $\omega(\epsilon) \leq \sup_{\theta\in\mathbb{V}:\|\boldsymbol{m}(\theta)\|\leq\epsilon} L(\theta)$ by definition, and $\epsilon\theta^* \in \Theta$ and $\|\boldsymbol{m}(\epsilon\theta^*)\| \leq \epsilon$ for all $\epsilon \in (0, a^*\sigma]$ by assumption, we have

$$L(\epsilon\theta^*) \leq \omega(\epsilon) \leq \sup_{\theta\in\mathbb{V}:\|\boldsymbol{m}(\theta)\|\leq\epsilon} L(\theta), \quad \epsilon \in (0, a^*\sigma].$$

Therefore, it suffices to show that $\sup_{\theta\in\mathbb{V}:\|\boldsymbol{m}(\theta)\|\leq\epsilon} L(\theta) \leq L(\epsilon\theta^*)$ for all $\epsilon \in (0, a^*\sigma]$. Suppose that $\sup_{\theta\in\mathbb{V}:\|\boldsymbol{m}(\theta)\|\leq\epsilon} L(\theta) > L(\epsilon\theta^*)$ for some $\epsilon \in (0, a^*\sigma]$. Then, there exists $\theta \in \mathbb{V}$ such that $\|\boldsymbol{m}(\theta)\| \leq \epsilon$ and $L(\theta) > L(\epsilon\theta^*)$. It follows that $\|\boldsymbol{m}(\theta/\epsilon)\| \leq 1$ and $L(\theta/\epsilon) > L(\theta^*)$, which contradicts the assumption that θ^* solves $\sup_{\theta\in\mathbb{V}:\|\boldsymbol{m}(\theta)\|\leq 1} L(\theta)$.

By the definition of a^*, we have

$$\arg\max_{0<\epsilon\leq a^*\sigma} \omega(\epsilon)\Phi(-\epsilon/\sigma) = \arg\max_{0<\epsilon\leq a^*\sigma} L(\theta^*)\epsilon\Phi(-\epsilon/\sigma) = \{a^*\sigma\}.$$

It is straightforward to show that Assumptions 1 and 2 hold. Applying Theorem 1 or Theorem 2, it is shown that the minimax regret decision rule is

$$\delta^*(\boldsymbol{Y}) = \mathbf{1}\left\{\boldsymbol{m}(a^*\sigma\theta^*)'\boldsymbol{Y} \geq 0\right\} = \mathbf{1}\left\{\boldsymbol{m}(\theta^*)'\boldsymbol{Y} \geq 0\right\},$$

and the minimax risk is $\mathcal{R}(\sigma; \Theta) = a^*\sigma L(\theta^*)\Phi(-a^*)$. □

B.5 Proof of Lemma 1

Let $g(\epsilon) = \omega(\epsilon)\Phi(-\epsilon/\sigma)$. As in the proof of Lemma B.1, differentiating g (from the right) at $\epsilon \geq 0$ gives

$$g'(\epsilon) = \left[\sigma \frac{1 - \Phi\left(\frac{\epsilon}{\sigma}\right)}{\phi\left(\frac{\epsilon}{\sigma}\right)} - \frac{\omega(\epsilon)}{\omega'(\epsilon)}\right] \omega'(\epsilon)\phi\left(\frac{\epsilon}{\sigma}\right)/\sigma.$$

By the fact that the Mills ratio $\frac{1-\Phi(x)}{\phi(x)}$ of a standard normal variable is strictly decreasing, $\sigma\frac{1-\Phi\left(\frac{\epsilon}{\sigma}\right)}{\phi\left(\frac{\epsilon}{\sigma}\right)}$ is strictly decreasing in ϵ. In addition, $\sigma\frac{1-\Phi\left(\frac{\epsilon}{\sigma}\right)}{\phi\left(\frac{\epsilon}{\sigma}\right)}$ is continuous. Furthermore, since $\omega(\epsilon)$ is nondecreasing and concave, $\frac{\omega(\epsilon)}{\omega'(\epsilon)}$ is nondecreasing.

Suppose that $\sigma > 2\phi(0)\frac{\omega(0)}{\omega'(0)}$. Then, $g'(0) > 0$. This implies that $g(\epsilon) > g(0)$ for any sufficiently small $\epsilon > 0$. If $g'(a^*\sigma) > 0$, $g(\epsilon)$ is strictly increasing on $[0, a^*\sigma]$, so g is uniquely maximized at $a^*\sigma$ over $[0, a^*\sigma]$. If $g'(a^*\sigma) \leq 0$, $g'(\epsilon) > 0$ for $\epsilon \in [0, \epsilon^*)$ and $g'(\epsilon) < 0$ for $\epsilon \in (\epsilon^*, a^*\sigma]$, where $\epsilon^* = \sup\{\epsilon \in [0, a^*\sigma] : g'(\epsilon) \geq 0\}$. Then g is uniquely maximized at ϵ^* over $[0, a^*\sigma]$.

Suppose that $\sigma \leq 2\phi(0)\frac{\omega(0)}{\omega'(0)}$. Then, $g'(0) \leq 0$. Since $\sigma\frac{1-\Phi\left(\frac{\epsilon}{\sigma}\right)}{\phi\left(\frac{\epsilon}{\sigma}\right)} - \frac{\omega(\epsilon)}{\omega'(\epsilon)}$ is strictly decreasing, $g'(\epsilon) < 0$ for any $\epsilon > 0$. By the mean value theorem, for every $\epsilon > 0$, $g(\epsilon) = g(0) + g'(\ddot{\epsilon})\epsilon$ for some $\ddot{\epsilon} \in (0, \epsilon)$, which implies $g(\epsilon) < g(0)$. Therefore, g is uniquely maximized at 0 over $[0, a^*\sigma]$. □

B.6 Proof of Lemma 2

Here, I prove the following result, which covers Lemma 2 as a special case.

Lemma B.4 (Univariate Problems (General)). *Suppose that $\Theta = [-\tau_1, -\tau_0] \cup [\tau_0, \tau_1]$ for some τ_0, τ_1 such that $\tau_1 \geq \tau_0 \geq 0$ and $\tau_1 > 0$, that $\boldsymbol{m}(\theta) = \theta$, and that $L(\theta) = \theta$. Then, the decision rule $\delta^*(Y) = \mathbf{1}\{Y \geq 0\}$ is minimax regret. The minimax risk is given by*

$$\mathcal{R}_{\text{uni}}(\sigma; [-\tau_1, -\tau_0] \cup [\tau_0, \tau_1]) = \begin{cases} \tau_1\Phi(-\tau_1/\sigma) & \text{if } \tau_1 < a^*\sigma, \\ a^*\sigma\Phi(-a^*) & \text{if } a^*\sigma \in [\tau_0, \tau_1], \\ \tau_0\Phi(-\tau_0/\sigma) & \text{if } a^*\sigma < \tau_0. \end{cases}$$

Following Stoye (2009), I use a statistical game to solve the minimax regret problem. Consider the following two-person zero-sum game between the decision maker and nature. The strategy space for the decision maker is $\mathcal{D}$, the set of all decision rules. The strategy space for nature is $\Delta(\Theta)$, the set of probability distributions on $\Theta = [-\tau_1, -\tau_0] \cup [\tau_0, \tau_1]$. If the decision maker chooses $\delta \in \mathcal{D}$ and nature chooses $\pi \in \Delta(\Theta)$, nature's expected payoff (and the decision maker's expected loss) is given by $r(\delta, \pi) = \int R(\delta, \theta)d\pi(\theta)$, the Bayes risk of δ with respect to prior π.

By Theorem 17 in Chapter 5 of Berger (1985), if (δ^*, π^*) satisfies

$$\delta^* \in \arg\min_{\delta\in\mathcal{D}} r(\delta, \pi^*), \text{ and } R(\delta^*, \theta) \le r(\delta^*, \pi^*) \text{ for all } \theta \in \Theta,$$

then δ^* is a minimax regret rule and π^* is a least favorable prior. Below I construct (δ^*, π^*) that satisfies the above conditions.

I first restrict the search space of decision rules to an essentially complete class of decision rules, following Tetenov (2012).[42] Since Y has monotone likelihood ratio and the loss function satisfies $l(1,\theta) - l(0,\theta) \ge 0$ if $\theta < 0$ and $l(1,\theta) - l(0,\theta) \le 0$ if $\theta > 0$, it follows from Theorem 5 in Chapter 8 of Berger (1985) (which is originally from Karlin and Rubin (1956)) that the class of monotone decision rules

$$\delta(Y) = \begin{cases} 0 & \text{if } Y < t, \\ \lambda & \text{if } Y = t, \\ 1 & \text{if } Y > t, \end{cases}$$

where $t \in \mathbb{R}$ and $\lambda \in [0,1]$, is essentially complete. Furthermore, since $\mathbb{P}_\theta(Y = t) = 0$, a smaller class of threshold decision rules $\delta(Y) = \mathbf{1}\{Y \ge t\}$, $t \in \mathbb{R}$, is also essentially complete.

Let δ_t denote the threshold rule with threshold t. Since $Y \sim \mathcal{N}(\theta, \sigma^2)$,

$$R(\delta_t, \theta) = \begin{cases} \theta\Phi(\sigma^{-1}(t-\theta)) & \text{if } \theta \ge 0, \\ (-\theta)(1 - \Phi(\sigma^{-1}(t-\theta))) & \text{if } \theta < 0. \end{cases}$$

Let $\bar{R}_0(t, \tau_0, \tau_1) = \max_{\theta\in[-\tau_1,-\tau_0]} R(\delta_t, \theta) = \max_{\theta\in[-\tau_1,-\tau_0]} -\theta(1 - \Phi(\sigma^{-1}(t - \theta)))$ and $\bar{R}_1(t, \tau_0, \tau_1) = \max_{\theta\in[\tau_0,\tau_1]} R(\delta_t, \theta) = \max_{\theta\in[\tau_0,\tau_1]} \theta\Phi(\sigma^{-1}(t-\theta))$. By symmetry of $\bar{R}_0(t, \tau_0, \tau_1)$ and $\bar{R}_1(t, \tau_0, \tau_1)$, $\bar{R}_0(0, \tau_0, \tau_1) = \bar{R}_1(0, \tau_0, \tau_1)$.

Now let $\theta_0^* \in \arg\max_{\theta\in[-\tau_1,-\tau_0]} R(\delta_0, \theta)$ and $\theta_1^* \in \arg\max_{\theta\in[\tau_0,\tau_1]} R(\delta_0, \theta)$, where $\delta_0(Y) = \mathbf{1}\{Y \ge 0\}$. By symmetry, $\theta_0^* = -\theta_1^*$. Let $\pi^* \in \Delta(\Theta)$ be such that

$$\pi^*(\theta_1^*) = \frac{-\theta_0^*\phi(\sigma^{-1}(-\theta_0^*))}{-\theta_0^*\phi(\sigma^{-1}(-\theta_0^*)) + \theta_1^*\phi(\sigma^{-1}(-\theta_1^*))} = \frac{1}{2},$$

and $\pi^*(\theta_0^*) = 1 - \pi^*(\theta_1^*) = \frac{1}{2}$. Since $\bar{R}_0(0, \tau_0, \tau_1) = \bar{R}_1(0, \tau_0, \tau_1) = R(\delta_0, \theta_0^*) = R(\delta_0, \theta_1^*)$, $r(\delta_0, \pi^*) = R(\delta_0, \theta_0^*) = R(\delta_0, \theta_1^*) \ge R(\delta_0, \theta)$ for all $\theta \in [-\tau_1, -\tau_0] \cup [\tau_0, \tau_1]$.

Since the class of threshold rules is essentially complete, $\delta_0 \in \arg\min_{\delta\in\mathcal{D}} r(\delta, \pi^*)$ if $0 \in$

[42] A class $\mathcal{C}$ of decision rules is essentially complete if, for any decision rule $\delta \notin \mathcal{C}$, there is a decision rule $\delta' \in \mathcal{C}$ such that $R(\delta, \theta) \ge R(\delta', \theta)$ for all $\theta \in \Theta$.

$\arg\min_{t\in\mathbb{R}} r(\delta_t, \pi^*)$. Observe

$$\begin{aligned} r(\delta_t, \pi^*) &= R(\delta_t, \theta_0^*)(1 - \pi^*(\theta_1^*)) + R(\delta_t, \theta_1^*)\pi^*(\theta_1^*) \\ &= -\theta_0^*(1 - \Phi(\sigma^{-1}(t - \theta_0^*)))(1 - \pi^*(\theta_1^*)) + \theta_1^*\Phi(\sigma^{-1}(t - \theta_1^*))\pi^*(\theta_1^*), \end{aligned}$$

and

$$\begin{aligned} \frac{\partial r(\delta_t, \pi^*)}{\partial t} &= \sigma^{-1}\theta_0^*\phi(\sigma^{-1}(t - \theta_0^*))(1 - \pi^*(\theta_1^*)) + \sigma^{-1}\theta_1^*\phi(\sigma^{-1}(t - \theta_1^*))\pi^*(\theta_1^*) \\ &= \sigma^{-1}\phi(\sigma^{-1}(t - \theta_0^*)) \left[\theta_0^*(1 - \pi^*(\theta_1^*)) + \theta_1^*\pi^*(\theta_1^*)\frac{\phi(\sigma^{-1}(t - \theta_1^*))}{\phi(\sigma^{-1}(t - \theta_0^*))}\right]. \end{aligned}$$

Since $\frac{\phi(\sigma^{-1}(t-\theta_1^*))}{\phi(\sigma^{-1}(t-\theta_0^*))}$ is increasing in t by the monotone likelihood ratio property, and $\theta_0^*(1-\pi^*(\theta_1^*))+\theta_1^*\pi^*(\theta_1^*)\frac{\phi(\sigma^{-1}(t-\theta_1^*))}{\phi(\sigma^{-1}(t-\theta_0^*))}$ is equal to zero at $t = 0$ by construction of π^*, it follows that

$$\frac{\partial r(\delta_t, \pi^*)}{\partial t} \begin{cases} > 0 & \text{if } t > 0, \\ = 0 & \text{if } t = 0, \\ < 0 & \text{if } t < 0. \end{cases}$$

Therefore, $r(\delta_t, \pi^*)$ is minimized at $t = 0$. Thus, δ_0 is minimax regret.

The minimax risk is given by

$$\begin{aligned} \mathcal{R}_{\text{uni}}(\sigma; [-\tau_1, -\tau_0] \cup [\tau_0, \tau_1]) &= \max_{\theta\in[\tau_0,\tau_1]} \theta\Phi(-\theta/\sigma) \\ &= \max_{a\in[\tau_0/\sigma,\tau_1/\sigma]} \sigma a\Phi(-a). \end{aligned}$$

By Lemma B.1, $a\Phi(-a)$ has a unique maximizer a^* over $[0, \infty)$ and $a\Phi(-a)$ is strictly increasing on $[0, a^*)$ and strictly decreasing on (a^*, ∞). Therefore,

$$\mathcal{R}_{\text{uni}}(\sigma; [-\tau_1, -\tau_0] \cup [\tau_0, \tau_1]) = \begin{cases} \tau_1\Phi(-\tau_1/\sigma) & \text{if } \tau_1 < \sigma a^*, \\ \sigma a^*\Phi(-a^*) & \text{if } \sigma a^* \in [\tau_0, \tau_1], \\ \tau_0\Phi(-\tau_0/\sigma) & \text{if } \sigma a^* < \tau_0. \end{cases}$$

□

B.7 Proof of Lemma 3

Here, I prove the following result, which covers Lemma 3 as a special case.

Lemma B.5 (Informative One-dimensional Subproblems (General))**.** *Suppose that* $\Theta = [-\bar{\theta}, -t\bar{\theta}] \cup [t\bar{\theta}, \bar{\theta}]$*, where* $\bar{\theta} \in \mathbb{V}$*,* $L(\bar{\theta}) > 0$*,* $\boldsymbol{m}(\bar{\theta}) \neq \mathbf{0}$*, and* $t \in [0, 1]$*. Then, the decision* *rule* $\delta^*(\boldsymbol{Y}) = \mathbf{1}\left\{\boldsymbol{m}(\bar{\theta})'\boldsymbol{Y} \geq 0\right\}$ *is minimax regret. The minimax risk is given by*

$$\mathcal{R}(\sigma; [-\bar{\theta}, -t\bar{\theta}] \cup [t\bar{\theta}, \bar{\theta}]) = \frac{L(\bar{\theta})}{\|\boldsymbol{m}(\bar{\theta})\|}\mathcal{R}_{\text{uni}}(\sigma; [-\|\boldsymbol{m}(\bar{\theta})\|, -t\|\boldsymbol{m}(\bar{\theta})\|] \cup [t\|\boldsymbol{m}(\bar{\theta})\|, \|\boldsymbol{m}(\bar{\theta})\|]).$$

Fix $\bar{\theta} \in \mathbb{V}$, where $L(\bar{\theta}) > 0$ and $\boldsymbol{m}(\bar{\theta}) \neq \mathbf{0}$, and $t \in [0, 1]$. We can write $[-\bar{\theta}, -t\bar{\theta}] \cup [t\bar{\theta}, \bar{\theta}] = \{\lambda\bar{\theta} : \lambda \in [-1, -t] \cup [t, 1]\}$. For $\lambda \in [-1, -t] \cup [t, 1]$, the regret of decision rule δ under $\lambda\bar{\theta}$ equals

$$\begin{aligned} R(\delta, \lambda\bar{\theta}) &= (L(\lambda\bar{\theta}))^+(1 - \mathbb{E}_{\lambda\bar{\theta}}[\delta(\boldsymbol{Y})]) + (-L(\lambda\bar{\theta}))^+\mathbb{E}_{\lambda\bar{\theta}}[\delta(\boldsymbol{Y})] \\ &= L(\bar{\theta})\left(\lambda^+(1 - \mathbb{E}_{\lambda\bar{\theta}}[\delta(\boldsymbol{Y})]) + (-\lambda)^+\mathbb{E}_{\lambda\bar{\theta}}[\delta(\boldsymbol{Y})]\right), \end{aligned}$$

where $x^+ = \max\{x, 0\}$. Minimax regret decision rules thus solve

$$\inf_{\delta} \sup_{\lambda \in [-1, -t] \cup [t, 1]} \left(\lambda^+(1 - \mathbb{E}_{\lambda\bar{\theta}}[\delta(\boldsymbol{Y})]) + (-\lambda)^+\mathbb{E}_{\lambda\bar{\theta}}[\delta(\boldsymbol{Y})]\right).$$

Viewing λ as a parameter, I derive a sufficient statistic of $\boldsymbol{Y}$ for λ. For $\lambda \in [-1, -t] \cup [t, 1]$, $\boldsymbol{Y} \sim \mathcal{N}(\lambda\boldsymbol{m}(\bar{\theta}), \sigma^2\boldsymbol{I}_n)$ under $\lambda\bar{\theta}$. It follows that the probability density of $\boldsymbol{Y}$ is

$$\begin{aligned} p(\boldsymbol{y}) &= \frac{1}{\sqrt{(2\pi)^n}\sigma^n}\exp\left(-\frac{1}{2\sigma^2}\|\boldsymbol{y} - \lambda\boldsymbol{m}(\bar{\theta})\|^2\right) \\ &= \frac{1}{\sqrt{(2\pi)^n}\sigma^n}\exp\left(-\frac{1}{2\sigma^2}(\|\boldsymbol{y}\|^2 - 2\lambda\boldsymbol{m}(\bar{\theta})'\boldsymbol{y} + \lambda^2\|\boldsymbol{m}(\bar{\theta})\|^2)\right) \\ &= h(\boldsymbol{y})g(T(\boldsymbol{y}), \lambda), \end{aligned}$$

where $h(\boldsymbol{y}) = \frac{1}{\sqrt{(2\pi)^n}\sigma^n}\exp(-\frac{1}{2\sigma^2}\|\boldsymbol{y}\|^2)$, $g(t, \lambda) = \exp(-\frac{1}{2\sigma^2}(-2\lambda t + \lambda^2)\|\boldsymbol{m}(\bar{\theta})\|^2)$, and $T(\boldsymbol{y}) = \frac{\boldsymbol{m}(\bar{\theta})'\boldsymbol{y}}{\|\boldsymbol{m}(\bar{\theta})\|^2}$. By the factorization theorem, $T(\boldsymbol{Y})$ is a sufficient statistic for λ.

It follows from Theorem 1 in Chapter 1 of Berger (1985) that the class of decision rules that only depend on $T(\boldsymbol{Y})$ is essentially complete. Since $T(\boldsymbol{Y}) \sim \mathcal{N}(\lambda, \frac{\sigma^2}{\|\boldsymbol{m}(\bar{\theta})\|^2})$ under $\lambda\bar{\theta}$, minimax regret decision rules that only depend on $T(\boldsymbol{Y})$ solve

$$\inf_{\delta} \sup_{\lambda \in [-1, -t] \cup [t, 1]} \left(\lambda^+(1 - \mathbb{E}_{\lambda}[\delta(T)]) + (-\lambda)^+\mathbb{E}_{\lambda}[\delta(T)]\right),$$

where the expectation is taken with respect to $T \sim \mathcal{N}(\lambda, \frac{\sigma^2}{\|\boldsymbol{m}(\bar{\theta})\|^2})$. This problem is equivalent to the univariate problem where $\Theta = [-1, -t] \cup [t, 1]$, $\boldsymbol{m}(\theta) = \theta$, $L(\theta) = \theta$, and the variance of the observed normal variable is $\frac{\sigma^2}{\|\boldsymbol{m}(\bar{\theta})\|^2}$. Thus, by Lemma B.4, the decision rule

$$\delta^*(\boldsymbol{Y}) = \mathbf{1}\{T(\boldsymbol{Y}) \geq 0\} = \mathbf{1}\left\{\boldsymbol{m}(\bar{\theta})'\boldsymbol{Y} \geq 0\right\}$$

is minimax regret. The minimax risk is given by

$$\begin{aligned}\mathcal{R}(\sigma;[-\bar\theta,-t\bar\theta]\cup[t\bar\theta,\bar\theta]) &= L(\bar\theta)\mathcal{R}_{\text{uni}}\left(\frac{\sigma}{\|\boldsymbol{m}(\bar\theta)\|};[-1,-t]\cup[t,1]\right)\\ &= \frac{L(\bar\theta)}{\|\boldsymbol{m}(\bar\theta)\|}\mathcal{R}_{\text{uni}}\left(\sigma;[-\|\boldsymbol{m}(\bar\theta)\|-t\|\boldsymbol{m}(\bar\theta)\|]\cup[t\|\boldsymbol{m}(\bar\theta)\|,\|\boldsymbol{m}(\bar\theta)\|]\right),\end{aligned}$$

where the second equality follows from the fact that $\mathcal{R}_{\text{uni}}(\alpha\sigma;[-\alpha\tau_1,-\alpha\tau_0]\cup[\alpha\tau_0,\alpha\tau_1]) = \alpha\mathcal{R}_{\text{uni}}(\sigma;[-\tau_1,-\tau_0]\cup[\tau_0,\tau_1])$ for all $\alpha>0$. □

B.8 Proof of Lemma 4

Let ϵ^* be the unique nonzero solution to $\max_{0\le\epsilon\le a^*\sigma}\omega(\epsilon)\Phi(-\epsilon/\sigma)$, and let θ_{ϵ^*} attain the modulus of continuity at ϵ^*. By Lemma B.2, $\|\boldsymbol{m}(\theta_{\epsilon^*})\|=\epsilon^*$. I first introduce some notation. Pick any $\underline{\eta}\in(0,\min\{\omega(\epsilon^*)\Phi(-\epsilon^*/\sigma),\epsilon^*\})$ and any $\bar\epsilon>\epsilon^*$, and define

$$\Gamma_{+,\underline{\eta},\bar\epsilon}=\left\{(L(\theta),\boldsymbol{m}(\theta)')'\in\mathbb{R}^{n+1}:\theta\in\Theta,L(\theta)\ge\underline{\eta},\frac{\boldsymbol{m}(\theta_{\epsilon^*})'\boldsymbol{m}(\theta)}{\|\boldsymbol{m}(\theta_{\epsilon^*})\|}\ge\underline{\eta},\|\boldsymbol{m}(\theta)\|\le\bar\epsilon\right\}.$$

Since $\omega(\epsilon^*)$ is finite, $\omega(\bar\epsilon)$ is also finite by the convexity of $\omega(\epsilon)$. Note that $\Gamma_{+,\underline{\eta},\bar\epsilon}$ is bounded, since $\underline{\eta}\le\alpha\le\omega(\bar\epsilon)$ and $\|\boldsymbol{\beta}\|\le\bar\epsilon$ for all $\boldsymbol{\gamma}=(\alpha,\boldsymbol{\beta}')'\in\Gamma_{+,\underline{\eta},\bar\epsilon}$. Let

$$\Theta_{+,\underline{\eta},\bar\epsilon}=\left\{\theta\in\Theta:L(\theta)\ge\underline{\eta},\frac{\boldsymbol{m}(\theta_{\epsilon^*})'\boldsymbol{m}(\theta)}{\|\boldsymbol{m}(\theta_{\epsilon^*})\|}\ge\underline{\eta},\|\boldsymbol{m}(\theta)\|\le\bar\epsilon\right\}$$

and

$$\begin{aligned}\Theta_{\underline{\eta},\bar\epsilon}&=\{\theta\in\Theta:(\theta\in\Theta_{+,\underline{\eta},\bar\epsilon})\text{ or }(-\theta\in\Theta_{+,\underline{\eta},\bar\epsilon})\}\\ &=\left\{\theta\in\Theta:\left[\left(L(\theta)\ge\underline{\eta},\frac{\boldsymbol{m}(\theta_{\epsilon^*})'\boldsymbol{m}(\theta)}{\|\boldsymbol{m}(\theta_{\epsilon^*})\|}\ge\underline{\eta}\right)\text{ or }\left(L(\theta)\le-\underline{\eta},\frac{\boldsymbol{m}(\theta_{\epsilon^*})'\boldsymbol{m}(\theta)}{\|\boldsymbol{m}(\theta_{\epsilon^*})\|}\le-\underline{\eta}\right)\right]\right.\\ &\qquad\left.\text{and }\|\boldsymbol{m}(\theta)\|\le\bar\epsilon\right\}.\end{aligned}$$

We can then write $\Gamma_{+,\underline{\eta},\bar\epsilon}=\{(L(\theta),\boldsymbol{m}(\theta)')'\in\mathbb{R}^{n+1}:\theta\in\Theta_{+,\underline{\eta},\bar\epsilon}\}$.

Now, let $\bar\Gamma_{+,\underline{\eta},\bar\epsilon}$ denote the closure of $\Gamma_{+,\underline{\eta},\bar\epsilon}$. Define a set-valued function $\Psi:\bar\Gamma_{+,\underline{\eta},\bar\epsilon}\to2^{\bar\Gamma_{+,\underline{\eta},\bar\epsilon}}$ as follows: for $\boldsymbol{\gamma}=(\alpha,\boldsymbol{\beta}')'\in\bar\Gamma_{+,\underline{\eta},\bar\epsilon}$,

$$\Psi(\boldsymbol{\gamma})=\arg\max_{\bar{\boldsymbol{\gamma}}=(\bar\alpha,\bar{\boldsymbol{\beta}}')'\in\bar\Gamma_{+,\underline{\eta},\bar\epsilon}}\bar\alpha\Phi\left(-\frac{\boldsymbol{\beta}'\bar{\boldsymbol{\beta}}}{\sigma\|\boldsymbol{\beta}\|}\right).$$

Note that $\alpha>0$ and $\boldsymbol{\beta}\ne0$ for all $(\alpha,\boldsymbol{\beta}')'\in\bar\Gamma_{+,\underline{\eta},\bar\epsilon}$, since $\tilde\alpha\ge\underline{\eta}$ and $\frac{\boldsymbol{m}(\theta_{\epsilon^*})'\tilde{\boldsymbol{\beta}}}{\|\boldsymbol{m}(\theta_{\epsilon^*})\|}\ge\underline{\eta}$ for all $\tilde{\boldsymbol{\gamma}}=(\tilde\alpha,\tilde{\boldsymbol{\beta}}')'\in\Gamma_{+,\underline{\eta},\bar\epsilon}$, and $(\alpha,\boldsymbol{\beta}')'$ is a point or a limit point of $\Gamma_{+,\underline{\eta},\bar\epsilon}$.

The proof consists of six steps.

Step 1. *Ψ has a fixed point, i.e., there exists $\boldsymbol{\gamma} \in \bar{\Gamma}_{+,\underline{\eta},\bar{\epsilon}}$ such that $\boldsymbol{\gamma} \in \Psi(\boldsymbol{\gamma})$.*

Proof. I apply Kakutani's fixed point theorem. First of all, $\bar{\Gamma}_{+,\underline{\eta},\bar{\epsilon}}$ is nonempty, since $(L(\theta_{\epsilon^*}), \boldsymbol{m}(\theta_{\epsilon^*}))')' \in \bar{\Gamma}_{+,\underline{\eta},\bar{\epsilon}}$. Furthermore, $\bar{\Gamma}_{+,\underline{\eta},\bar{\epsilon}}$ is closed and bounded by construction.

I now show that $\bar{\Gamma}_{+,\underline{\eta},\bar{\epsilon}}$ is convex. It suffices to show that $\Gamma_{+,\underline{\eta},\bar{\epsilon}}$ is convex, since the closure of a convex subset of $\mathbb{R}^{n+1}$ is convex. Pick any $\boldsymbol{\gamma}, \tilde{\boldsymbol{\gamma}} \in \Gamma_{+,\underline{\eta},\bar{\epsilon}}$. Let $\theta, \tilde{\theta} \in \Theta_{+,\underline{\eta},\bar{\epsilon}}$ be such that $(L(\theta), \boldsymbol{m}(\theta)')' = \boldsymbol{\gamma}$ and $(L(\tilde{\theta}), \boldsymbol{m}(\tilde{\theta})')' = \tilde{\boldsymbol{\gamma}}$. Fix $\lambda \in [0,1]$. By the linearity of L and $\boldsymbol{m}$, $\lambda\boldsymbol{\gamma} + (1-\lambda)\tilde{\boldsymbol{\gamma}} = (L(\lambda\theta + (1-\lambda)\tilde{\theta}), \boldsymbol{m}(\lambda\theta+(1-\lambda)\tilde{\theta})')'$. Since $\lambda\theta + (1-\lambda)\tilde{\theta} \in \Theta$ by the convexity of Θ, $L(\lambda\theta + (1-\lambda)\tilde{\theta}) = \lambda L(\theta) + (1-\lambda)L(\tilde{\theta}) \geq \underline{\eta}$, $\frac{\boldsymbol{m}(\theta_{\epsilon^*})'\boldsymbol{m}(\lambda\theta+(1-\lambda)\tilde{\theta})}{\|\boldsymbol{m}(\theta_{\epsilon^*})\|} = \lambda\frac{\boldsymbol{m}(\theta_{\epsilon^*})'\boldsymbol{m}(\theta)}{\|\boldsymbol{m}(\theta_{\epsilon^*})\|} + (1-\lambda)\frac{\boldsymbol{m}(\theta_{\epsilon^*})'\boldsymbol{m}(\tilde{\theta})}{\|\boldsymbol{m}(\theta_{\epsilon^*})\|} \geq \underline{\eta}$, and $\|\boldsymbol{m}(\lambda\theta + (1-\lambda)\tilde{\theta})\| \leq \|\boldsymbol{m}(\lambda\theta)\| + \|\boldsymbol{m}((1-\lambda)\tilde{\theta})\| = \lambda\|\boldsymbol{m}(\theta)\| + (1-\lambda)\|\boldsymbol{m}(\tilde{\theta})\| \leq \bar{\epsilon}$, it follows that $\lambda\theta + (1-\lambda)\tilde{\theta} \in \Theta_{+,\underline{\eta},\bar{\epsilon}}$. Therefore, $\lambda\boldsymbol{\gamma} + (1-\lambda)\tilde{\boldsymbol{\gamma}} \in \Gamma_{+,\underline{\eta},\bar{\epsilon}}$.

Next, I show that $\Psi(\boldsymbol{\gamma})$ is nonempty and convex for all $\boldsymbol{\gamma} \in \bar{\Gamma}_{+,\underline{\eta},\bar{\epsilon}}$. Fix $\boldsymbol{\gamma} = (\alpha, \boldsymbol{\beta}')' \in \bar{\Gamma}_{+,\underline{\eta},\bar{\epsilon}}$. Let

$$S_{\boldsymbol{\beta}} = \left\{ \left(\bar{\alpha}, \frac{\boldsymbol{\beta}'\bar{\boldsymbol{\beta}}}{\sigma\|\boldsymbol{\beta}\|} \right) \in \mathbb{R}^2 : \bar{\boldsymbol{\gamma}} = (\bar{\alpha}, \bar{\boldsymbol{\beta}}')' \in \bar{\Gamma}_{+,\underline{\eta},\bar{\epsilon}} \right\},$$

which is a subset of $(0,\infty) \times \mathbb{R}$. Using $S_{\boldsymbol{\beta}}$, we can write

$$\Psi(\boldsymbol{\gamma}) = \left\{ (\bar{\alpha}, \bar{\boldsymbol{\beta}}')' \in \bar{\Gamma}_{+,\underline{\eta},\bar{\epsilon}} : \left(\bar{\alpha}, \frac{\boldsymbol{\beta}'\bar{\boldsymbol{\beta}}}{\sigma\|\boldsymbol{\beta}\|} \right) \in \arg\max_{(a,b)\in S_{\boldsymbol{\beta}}} a\Phi(-b) \right\}.$$

Since the mapping $\bar{\boldsymbol{\gamma}} \mapsto \left(\bar{\alpha}, \frac{\boldsymbol{\beta}'\bar{\boldsymbol{\beta}}}{\sigma\|\boldsymbol{\beta}\|}\right)$ is continuous and $\bar{\Gamma}_{+,\underline{\eta},\bar{\epsilon}}$ is compact, $S_{\boldsymbol{\beta}}$ is compact. Furthermore, since $\bar{\boldsymbol{\gamma}} \mapsto \left(\bar{\alpha}, \frac{\boldsymbol{\beta}'\bar{\boldsymbol{\beta}}}{\sigma\|\boldsymbol{\beta}\|}\right)$ is linear and $\bar{\Gamma}_{+,\underline{\eta},\bar{\epsilon}}$ is convex, $S_{\boldsymbol{\beta}}$ is convex. It then follows that $\arg\max_{(a,b)\in S_{\boldsymbol{\beta}}} a\Phi(-b)$ is nonempty and singleton, since $a\Phi(-b)$ is continuous and is strictly quasi-concave on $(0,\infty)\times\mathbb{R}$ by Lemma B.3. Let $(a^*_{\boldsymbol{\beta}}, b^*_{\boldsymbol{\beta}}) \in \arg\max_{(a,b)\in S_{\boldsymbol{\beta}}} a\Phi(-b)$. We can then write

$$\Psi(\boldsymbol{\gamma}) = \left\{ (\bar{\alpha}, \bar{\boldsymbol{\beta}}')' \in \bar{\Gamma}_{+,\underline{\eta},\bar{\epsilon}} : \bar{\alpha} = a^*_{\boldsymbol{\beta}}, \frac{\boldsymbol{\beta}'\bar{\boldsymbol{\beta}}}{\sigma\|\boldsymbol{\beta}\|} = b^*_{\boldsymbol{\beta}} \right\},$$

which is nonempty and convex.

Lastly, I show that Ψ has a closed graph. Let $f(\bar{\boldsymbol{\gamma}}, \boldsymbol{\gamma}) = \bar{\alpha}\Phi\left(-\frac{\boldsymbol{\beta}'\bar{\boldsymbol{\beta}}}{\sigma\|\boldsymbol{\beta}\|}\right)$. Take any sequence $\{(\boldsymbol{\gamma}_n, \boldsymbol{\gamma}^*_n)\}_{n=1}^{\infty}$ such that $\boldsymbol{\gamma}_n, \boldsymbol{\gamma}^*_n \in \bar{\Gamma}_{+,\underline{\eta},\bar{\epsilon}}$ for all n, $(\boldsymbol{\gamma}_n, \boldsymbol{\gamma}^*_n) \to (\boldsymbol{\gamma}, \boldsymbol{\gamma}^*)$, and $\boldsymbol{\gamma}^*_n \in \Psi(\boldsymbol{\gamma}_n)$ for all n. Since $\boldsymbol{\gamma}_n, \boldsymbol{\gamma}^*_n \in \bar{\Gamma}_{+,\underline{\eta},\bar{\epsilon}}$ and $\bar{\Gamma}_{+,\underline{\eta},\bar{\epsilon}}$ is closed, it follows that $\boldsymbol{\gamma}, \boldsymbol{\gamma}^* \in \bar{\Gamma}_{+,\underline{\eta},\bar{\epsilon}}$. This implies that $\boldsymbol{\beta}^* \neq 0$.

I show that $\boldsymbol{\gamma}^* \in \Psi(\boldsymbol{\gamma})$. Suppose this does not hold. Then there exist $\boldsymbol{\gamma}^{**} \in \bar{\Gamma}_{+,\underline{\eta},\bar{\epsilon}}$ and $\epsilon > 0$ such that $f(\boldsymbol{\gamma}^{**}, \boldsymbol{\gamma}) > f(\boldsymbol{\gamma}^*, \boldsymbol{\gamma}) + 3\epsilon$. Also, since f is continuous on $\bar{\Gamma}_{+,\underline{\eta},\bar{\epsilon}} \times \bar{\Gamma}_{+,\underline{\eta},\bar{\epsilon}}$ and $(\boldsymbol{\gamma}_n, \boldsymbol{\gamma}^*_n) \to (\boldsymbol{\gamma}, \boldsymbol{\gamma}^*)$, we have $f(\boldsymbol{\gamma}^{**}, \boldsymbol{\gamma}_n) > f(\boldsymbol{\gamma}^{**}, \boldsymbol{\gamma}) - \epsilon$ and $f(\boldsymbol{\gamma}^*, \boldsymbol{\gamma}) > f(\boldsymbol{\gamma}^*_n, \boldsymbol{\gamma}_n) - \epsilon$ for any sufficiently large n. Combining the preceding inequalities, we obtain for any sufficiently large n,

$$f(\boldsymbol{\gamma}^{**}, \boldsymbol{\gamma}_n) > f(\boldsymbol{\gamma}^*, \boldsymbol{\gamma}) + 2\epsilon > f(\boldsymbol{\gamma}^*_n, \boldsymbol{\gamma}_n) + \epsilon.$$

This contradicts the assumption that $\boldsymbol{\gamma}_n^* \in \Psi(\boldsymbol{\gamma}_n)$ for all n.

Application of Kakutani's fixed point theorem proves the statement. $\square$

Let $\boldsymbol{\gamma}^* = (\alpha^*, (\boldsymbol{\beta}^*)')'$ be a fixed point of Ψ. In Steps 2–5, I prove that $L(\theta_{\epsilon^*}) = \alpha^*$ and $\boldsymbol{m}(\theta_{\epsilon^*}) = \boldsymbol{\beta}^*$.

$\boldsymbol{\gamma}^*$ may not be an element of $\Gamma_{+,\underline{\eta},\bar{\epsilon}}$, but since it is an element of $\bar{\Gamma}_{+,\underline{\eta},\bar{\epsilon}}$, we can take sequences $\{\boldsymbol{\gamma}_n = (\alpha_n, \boldsymbol{\beta}_n')'\}_{n=1}^\infty$ and $\{\theta_n\}_{n=1}^\infty$ such that $\theta_n \in \Theta_{+,\underline{\eta},\bar{\epsilon}}$ and $\boldsymbol{\gamma}_n = (L(\theta_n), \boldsymbol{m}(\theta_n)')' \in \Gamma_{+,\underline{\eta},\bar{\epsilon}}$ for all $n \geq 1$ and that $\lim_{n\to\infty} \boldsymbol{\gamma}_n = \boldsymbol{\gamma}^*$. Let $\tilde{\delta}(\boldsymbol{Y}) = \mathbf{1}\{(\boldsymbol{\beta}^*)'\boldsymbol{Y} \geq 0\}$. Below, I suppress the argument σ of the minimax risk $\mathcal{R}(\sigma;\cdot)$ for the notational brevity.

Step 2. $\sup_{\theta\in\Theta_{\underline{\eta},\bar{\epsilon}}} R(\tilde{\delta},\theta) = \sup_{\theta\in\Theta_{+,\underline{\eta},\bar{\epsilon}}} \mathcal{R}([-\theta,\theta]\cap\Theta_{\underline{\eta},\bar{\epsilon}}) = \lim_{n\to\infty}\mathcal{R}([-\theta_n,\theta_n]\cap\Theta_{\underline{\eta},\bar{\epsilon}}) = \alpha^*\Phi\left(-\frac{\|\boldsymbol{\beta}^*\|}{\sigma}\right)$.

Proof. Since $\boldsymbol{\gamma}^* \in \arg\max_{\boldsymbol{\gamma}=(\alpha,\boldsymbol{\beta}')'\in\bar{\Gamma}_{+,\underline{\eta},\bar{\epsilon}}} \alpha\Phi\left(-\frac{(\boldsymbol{\beta}^*)'\boldsymbol{\beta}}{\sigma\|\boldsymbol{\beta}^*\|}\right)$,

$$
\begin{aligned}
\max_{\boldsymbol{\gamma}=(\alpha,\boldsymbol{\beta}')'\in\bar{\Gamma}_{+,\underline{\eta},\bar{\epsilon}}} \alpha\Phi\left(-\frac{(\boldsymbol{\beta}^*)'\boldsymbol{\beta}}{\sigma\|\boldsymbol{\beta}^*\|}\right) &= \alpha^*\Phi\left(-\frac{(\boldsymbol{\beta}^*)'(\boldsymbol{\beta}^*)}{\sigma\|\boldsymbol{\beta}^*\|}\right)\\
&= \lim_{n\to\infty}\alpha_n\Phi\left(-\frac{(\boldsymbol{\beta}^*)'\boldsymbol{\beta}_n}{\sigma\|\boldsymbol{\beta}^*\|}\right)\\
&= \lim_{n\to\infty}L(\theta_n)\Phi\left(-\frac{(\boldsymbol{\beta}^*)'\boldsymbol{m}(\theta_n)}{\sigma\|\boldsymbol{\beta}^*\|}\right)\\
&= \lim_{n\to\infty}R(\tilde{\delta},\theta_n),
\end{aligned}
$$

where the second equality follows by the fact that the mapping $\boldsymbol{\gamma}\mapsto\alpha\Phi\left(-\frac{(\boldsymbol{\beta}^*)'\boldsymbol{\beta}}{\sigma\|\boldsymbol{\beta}^*\|}\right)$ is continuous. Also, by continuity of the mapping $\boldsymbol{\gamma}\mapsto\alpha\Phi\left(-\frac{(\boldsymbol{\beta})'\boldsymbol{\beta}}{\sigma\|\boldsymbol{\beta}\|}\right)$,

$$
\begin{aligned}
\alpha^*\Phi\left(-\frac{(\boldsymbol{\beta}^*)'(\boldsymbol{\beta}^*)}{\sigma\|\boldsymbol{\beta}^*\|}\right) &= \lim_{n\to\infty}\alpha_n\Phi\left(-\frac{\boldsymbol{\beta}_n'\boldsymbol{\beta}_n}{\sigma\|\boldsymbol{\beta}_n\|}\right)\\
&= \lim_{n\to\infty}L(\theta_n)\Phi\left(-\frac{\boldsymbol{m}(\theta_n)'\boldsymbol{m}(\theta_n)}{\sigma\|\boldsymbol{m}(\theta_n)\|}\right)\\
&= \lim_{n\to\infty}R(\delta_n,\theta_n),
\end{aligned}
$$

where $\delta_n(\boldsymbol{Y}) = \mathbf{1}\{\boldsymbol{m}(\theta_n)'\boldsymbol{Y} \geq 0\}$ for all n.

On the other hand, by definition,

$$\sup_{\boldsymbol{\gamma}=(\alpha,\boldsymbol{\beta}')'\in\bar{\Gamma}_{+,\underline{\eta},\bar{\epsilon}}} \alpha\Phi\left(-\frac{(\boldsymbol{\beta}^*)'\boldsymbol{\beta}}{\sigma\|\boldsymbol{\beta}^*\|}\right) \geq \sup_{\boldsymbol{\gamma}=(\alpha,\boldsymbol{\beta}')'\in\Gamma_{+,\underline{\eta},\bar{\epsilon}}} \alpha\Phi\left(-\frac{(\boldsymbol{\beta}^*)'\boldsymbol{\beta}}{\sigma\|\boldsymbol{\beta}^*\|}\right)$$
$$= \sup_{\theta\in\Theta_{+,\underline{\eta},\bar{\epsilon}}} L(\theta)\Phi\left(-\frac{(\boldsymbol{\beta}^*)'\boldsymbol{m}(\theta)}{\sigma\|\boldsymbol{\beta}^*\|}\right)$$
$$= \sup_{\theta\in\Theta_{+,\underline{\eta},\bar{\epsilon}}} R(\tilde{\delta},\theta)$$
$$\geq \lim_{n\to\infty} R(\tilde{\delta},\theta_n).$$

Therefore,

$$\sup_{\theta\in\Theta_{+,\underline{\eta},\bar{\epsilon}}} R(\tilde{\delta},\theta) = \lim_{n\to\infty} R(\tilde{\delta},\theta_n) = \lim_{n\to\infty} R(\delta_n,\theta_n) = \alpha^*\Phi\left(-\frac{(\boldsymbol{\beta}^*)'(\boldsymbol{\beta}^*)}{\sigma\|\boldsymbol{\beta}^*\|}\right).$$

Note that, $\sup_{\theta\in\Theta_{+,\underline{\eta},\bar{\epsilon}}} R(\tilde{\delta},\theta) = \sup_{\theta\in\Theta_{\underline{\eta},\bar{\epsilon}}} R(\tilde{\delta},\theta)$ by the symmetry of the regret function and the centrosymmetry of $\Theta_{\underline{\eta},\bar{\epsilon}}$.

We also have

$$\lim_{n\to\infty} R(\delta_n,\theta_n) \leq \lim_{n\to\infty} \sup_{\theta\in[-\theta_n,\theta_n]\cap\Theta_{\underline{\eta},\bar{\epsilon}}} R(\delta_n,\theta)$$
$$= \lim_{n\to\infty} \mathcal{R}([-\theta_n,\theta_n]\cap\Theta_{\underline{\eta},\bar{\epsilon}})$$
$$\leq \sup_{\theta\in\Theta_{+,\underline{\eta},\bar{\epsilon}}} \mathcal{R}([-\theta,\theta]\cap\Theta_{\underline{\eta},\bar{\epsilon}}),$$

where the equality in the second line holds since $[-\theta_n,\theta_n]\cap\Theta_{\underline{\eta},\bar{\epsilon}} = [-\theta_n,-t_n\theta_n]\cup[t_n\theta_n,\theta_n]$ with $t_n = \max\{\frac{\eta}{L(\theta_n)}, \underline{\eta}\frac{\|\boldsymbol{m}(\theta_{\epsilon^*})\|}{\boldsymbol{m}(\theta_{\epsilon^*})'\boldsymbol{m}(\theta_n)}\}$, and δ_n is minimax regret for $[-\theta_n,-t_n\theta_n]\cup[t_n\theta_n,\theta_n]$ by Lemma B.5. However, by definition,

$$\sup_{\theta\in\Theta_{\underline{\eta},\bar{\epsilon}}} R(\tilde{\delta},\theta) \geq \mathcal{R}(\Theta_{\underline{\eta},\bar{\epsilon}}) \geq \sup_{\theta\in\Theta_{+,\underline{\eta},\bar{\epsilon}}} \mathcal{R}([-\theta,\theta]\cap\Theta_{\underline{\eta},\bar{\epsilon}}).$$

It follows that

$$\sup_{\theta\in\Theta_{\underline{\eta},\bar{\epsilon}}} R(\tilde{\delta},\theta) = \sup_{\theta\in\Theta_{+,\underline{\eta},\bar{\epsilon}}} \mathcal{R}([-\theta,\theta]\cap\Theta_{\underline{\eta},\bar{\epsilon}}) = \lim_{n\to\infty} \mathcal{R}([-\theta_n,\theta_n]\cap\Theta_{\underline{\eta},\bar{\epsilon}}) = \alpha^*\Phi\left(-\frac{(\boldsymbol{\beta}^*)'(\boldsymbol{\beta}^*)}{\sigma\|\boldsymbol{\beta}^*\|}\right).$$

□

Step 3. $\sup_{\theta\in\Theta_{+,\underline{\eta},\bar{\epsilon}}} \mathcal{R}([-\theta,\theta]\cap\Theta_{\underline{\eta},\bar{\epsilon}}) = \omega(\epsilon^*)\Phi(-\epsilon^*/\sigma)$.

Proof. By Lemma B.5,

$$\mathcal{R}([-\theta_{\epsilon^*},\theta_{\epsilon^*}]) = \frac{L(\theta_{\epsilon^*})}{\|\boldsymbol{m}(\theta_{\epsilon^*})\|}\mathcal{R}_{\text{uni}}\left(\sigma;[-\|\boldsymbol{m}(\theta_{\epsilon^*})\|,\|\boldsymbol{m}(\theta_{\epsilon^*})\|]\right) = \frac{\omega(\epsilon^*)}{\epsilon^*}\mathcal{R}_{\text{uni}}\left(\sigma;[-\epsilon^*,\epsilon^*]\right).$$

On the other hand, with $t=\max\{\frac{\underline{\eta}}{L(\theta_{\epsilon^*})},\frac{\underline{\eta}}{\epsilon^*}\}$,

$$\begin{aligned}\mathcal{R}([-\theta_{\epsilon^*},\theta_{\epsilon^*}]\cap\Theta_{\underline{\eta},\bar{\epsilon}}) &= \mathcal{R}([-\theta_{\epsilon^*},-t\theta_{\epsilon^*}]\cup[t\theta_{\epsilon^*},\theta_{\epsilon^*}])\\ &= \frac{L(\theta_{\epsilon^*})}{\|\boldsymbol{m}(\theta_{\epsilon^*})\|}\mathcal{R}_{\text{uni}}\left(\sigma;[-\|\boldsymbol{m}(\theta_{\epsilon^*})\|,-t\|\boldsymbol{m}(\theta_{\epsilon^*})\|]\cup[t\|\boldsymbol{m}(\theta_{\epsilon^*})\|,\|\boldsymbol{m}(\theta_{\epsilon^*})\|]\right)\\ &= \frac{\omega(\epsilon^*)}{\epsilon^*}\mathcal{R}_{\text{uni}}\left(\sigma;[-\epsilon^*,-t\epsilon^*]\cup[t\epsilon^*,\epsilon^*]\right).\end{aligned}$$

Since $\epsilon^*\le a^*\sigma$, $\mathcal{R}_{\text{uni}}(\sigma;[-\epsilon^*,\epsilon^*])=\mathcal{R}_{\text{uni}}(\sigma;[-\epsilon^*,-t\epsilon^*]\cup[t\epsilon^*,\epsilon^*])$ by Lemma B.4. Therefore, $\mathcal{R}([-\theta_{\epsilon^*},\theta_{\epsilon^*}])=\mathcal{R}([-\theta_{\epsilon^*},\theta_{\epsilon^*}]\cap\Theta_{\underline{\eta},\bar{\epsilon}})$. Note that this equals $\sup_{\theta\in\Theta:L(\theta)>0,\boldsymbol{m}(\theta)\neq\boldsymbol{0}}\mathcal{R}([-\theta,\theta])=\omega(\epsilon^*)\Phi(-\epsilon^*/\sigma)$ as discussed in Step 3 in Section 6.1.

Now, since $\theta_{\epsilon^*}\in\Theta_{+,\underline{\eta},\bar{\epsilon}}$,

$$\sup_{\theta\in\Theta_{+,\underline{\eta},\bar{\epsilon}}}\mathcal{R}([-\theta,\theta]\cap\Theta_{\underline{\eta},\bar{\epsilon}})\ge\mathcal{R}([-\theta_{\epsilon^*},\theta_{\epsilon^*}]\cap\Theta_{\underline{\eta},\bar{\epsilon}}).$$

However, by definition and the above result,

$$\sup_{\theta\in\Theta_{+,\underline{\eta},\bar{\epsilon}}}\mathcal{R}([-\theta,\theta]\cap\Theta_{\underline{\eta},\bar{\epsilon}})\le\sup_{\theta\in\Theta_{+,\underline{\eta},\bar{\epsilon}}}\mathcal{R}([-\theta,\theta])\le\sup_{\theta\in\Theta:L(\theta)>0,\boldsymbol{m}(\theta)\neq\boldsymbol{0}}\mathcal{R}([-\theta,\theta])=\mathcal{R}([-\theta_{\epsilon^*},\theta_{\epsilon^*}]\cap\Theta_{\underline{\eta},\bar{\epsilon}}).$$

Therefore, I obtain

$$\sup_{\theta\in\Theta_{+,\underline{\eta},\bar{\epsilon}}}\mathcal{R}([-\theta,\theta]\cap\Theta_{\underline{\eta},\bar{\epsilon}})=\sup_{\theta\in\Theta_{+,\underline{\eta},\bar{\epsilon}}}\mathcal{R}([-\theta,\theta])=\omega(\epsilon^*)\Phi(-\epsilon^*/\sigma).$$

□

Step 4. *$\alpha^*=\omega(\epsilon^*)$ and $\|\boldsymbol{\beta}^*\|=\epsilon^*$.*

Proof. First, with $t_n=\max\{\frac{\underline{\eta}}{L(\theta_n)},\underline{\eta}\frac{\|\boldsymbol{m}(\theta_{\epsilon^*})\|}{\boldsymbol{m}(\theta_{\epsilon^*})'\boldsymbol{m}(\theta_n)}\}$ and $t^*=\lim_{n\to\infty}t_n=\max\{\frac{\underline{\eta}}{\alpha^*},\underline{\eta}\frac{\|\boldsymbol{m}(\theta_{\epsilon^*})\|}{\boldsymbol{m}(\theta_{\epsilon^*})'\boldsymbol{\beta}^*}\}$,

$$\begin{aligned}&\lim_{n\to\infty}\mathcal{R}([-\theta_n,\theta_n]\cap\Theta_{\underline{\eta},\bar{\epsilon}})\\ =&\lim_{n\to\infty}\mathcal{R}([-\theta_n,-t_n\theta_n]\cup[t_n\theta_n,\theta_n])\\ =&\lim_{n\to\infty}\frac{L(\theta_n)}{\|\boldsymbol{m}(\theta_n)\|}\mathcal{R}_{\text{uni}}\left(\sigma;[-\|\boldsymbol{m}(\theta_n)\|,-t_n\|\boldsymbol{m}(\theta_n)\|]\cup[t_n\|\boldsymbol{m}(\theta_n)\|,\|\boldsymbol{m}(\theta_n)\|]\right)\\ =&\frac{\alpha^*}{\|\boldsymbol{\beta}^*\|}\mathcal{R}_{\text{uni}}\left(\sigma;[-\|\boldsymbol{\beta}^*\|,-t^*\|\boldsymbol{\beta}^*\|]\cup[t^*\|\boldsymbol{\beta}^*\|,\|\boldsymbol{\beta}^*\|]\right),\end{aligned}$$

where the last equality follows from the fact that the mapping $(\tau, t) \mapsto \mathcal{R}_{\text{uni}}(\sigma; [-\tau, -t\tau] \cup [t\tau, \tau])$ is continuous. By Steps 2–3, $\omega(\epsilon^*)\Phi(-\epsilon^*/\sigma) = \lim_{n\to\infty} \mathcal{R}([-\theta_n, \theta_n] \cap \Theta_{\underline{\eta},\bar{\epsilon}})$. Therefore,

$$\begin{aligned}\omega(\epsilon^*)\Phi(-\epsilon^*/\sigma) &= \frac{\alpha^*}{\|\boldsymbol{\beta}^*\|}\mathcal{R}_{\text{uni}}\left(\sigma; [-\|\boldsymbol{\beta}^*\|, -t^*\|\boldsymbol{\beta}^*\|] \cup [t^*\|\boldsymbol{\beta}^*\|, \|\boldsymbol{\beta}^*\|]\right)\\ &\le \frac{\alpha^*}{\|\boldsymbol{\beta}^*\|}\mathcal{R}_{\text{uni}}\left(\sigma; [-\|\boldsymbol{\beta}^*\|, \|\boldsymbol{\beta}^*\|]\right).\end{aligned}$$

Note that $\alpha_n \le \omega(\|\boldsymbol{\beta}_n\|)$ for all $n \ge 1$ by the definition of $\omega(\cdot)$. Since $\omega(\cdot)$ is continuous by the concavity, taking the limit of both sides yields $\alpha^* \le \omega(\|\boldsymbol{\beta}^*\|)$. It follows that

$$\omega(\epsilon^*)\Phi(-\epsilon^*/\sigma) \le \frac{\omega(\|\boldsymbol{\beta}^*\|)}{\|\boldsymbol{\beta}^*\|}\mathcal{R}_{\text{uni}}\left(\sigma; [-\|\boldsymbol{\beta}^*\|, \|\boldsymbol{\beta}^*\|]\right).$$

On the other hand, as discussed in Step 3 in Section 6.1,

$$\omega(\epsilon^*)\Phi(-\epsilon^*/\sigma) = \max_{\epsilon>0}\frac{\omega(\epsilon)}{\epsilon}\mathcal{R}_{\text{uni}}\left(\sigma; [-\epsilon, \epsilon]\right) \ge \frac{\omega(\|\boldsymbol{\beta}^*\|)}{\|\boldsymbol{\beta}^*\|}\mathcal{R}_{\text{uni}}\left(\sigma; [-\|\boldsymbol{\beta}^*\|, \|\boldsymbol{\beta}^*\|]\right).$$

It follows that

$$\begin{aligned}\max_{\epsilon>0}\frac{\omega(\epsilon)}{\epsilon}\mathcal{R}_{\text{uni}}\left(\sigma; [-\epsilon, \epsilon]\right) &= \frac{\alpha^*}{\|\boldsymbol{\beta}^*\|}\mathcal{R}_{\text{uni}}\left(\sigma; [-\|\boldsymbol{\beta}^*\|, -t^*\|\boldsymbol{\beta}^*\|] \cup [t^*\|\boldsymbol{\beta}^*\|, \|\boldsymbol{\beta}^*\|]\right)\\ &= \frac{\alpha^*}{\|\boldsymbol{\beta}^*\|}\mathcal{R}_{\text{uni}}\left(\sigma; [-\|\boldsymbol{\beta}^*\|, \|\boldsymbol{\beta}^*\|]\right)\\ &= \frac{\omega(\|\boldsymbol{\beta}^*\|)}{\|\boldsymbol{\beta}^*\|}\mathcal{R}_{\text{uni}}\left(\sigma; [-\|\boldsymbol{\beta}^*\|, \|\boldsymbol{\beta}^*\|]\right).\end{aligned}$$

Therefore, $\alpha^* = \omega(\|\boldsymbol{\beta}^*\|)$ and $\|\boldsymbol{\beta}^*\| \in \arg\max_{\epsilon>0}\frac{\omega(\epsilon)}{\epsilon}\mathcal{R}_{\text{uni}}\left(\sigma; [-\epsilon, \epsilon]\right)$. Furthermore,

$$\mathcal{R}_{\text{uni}}\left(\sigma; [-\|\boldsymbol{\beta}^*\|, -t^*\|\boldsymbol{\beta}^*\|] \cup [t^*\|\boldsymbol{\beta}^*\|, \|\boldsymbol{\beta}^*\|]\right) = \mathcal{R}_{\text{uni}}\left(\sigma; [-\|\boldsymbol{\beta}^*\|, \|\boldsymbol{\beta}^*\|]\right). \tag{B.11}$$

If it is shown that $\|\boldsymbol{\beta}^*\| \le a^*\sigma$, then

$$\|\boldsymbol{\beta}^*\| \in \arg\max_{0<\epsilon\le a^*\sigma}\frac{\omega(\epsilon)}{\epsilon}\mathcal{R}_{\text{uni}}\left(\sigma; [-\epsilon, \epsilon]\right) = \arg\max_{0<\epsilon\le a^*\sigma}\omega(\epsilon)\Phi(-\epsilon/\sigma) = \{\epsilon^*\}.$$

Suppose to the contrary that $\|\boldsymbol{\beta}^*\| > a^*\sigma$. By inspection of the form of $\mathcal{R}_{\text{uni}}$ given in Lemma B.4, it is necessary that $t^*\|\boldsymbol{\beta}^*\| \le a^*\sigma$ for Eq. (B.11) to hold. Therefore, $t^* \le \frac{a^*\sigma}{\|\boldsymbol{\beta}^*\|} < 1$ since $\|\boldsymbol{\beta}^*\| > a^*\sigma$. It follows that $t^* = \max\{\frac{\underline{\eta}}{\alpha^*}, \underline{\eta}\frac{\|\boldsymbol{m}(\theta_{\epsilon^*})\|}{\boldsymbol{m}(\theta_{\epsilon^*})'\boldsymbol{\beta}^*}\} < 1$, so $\underline{\eta} < \frac{\boldsymbol{m}(\theta_{\epsilon^*})'\boldsymbol{\beta}^*}{\|\boldsymbol{m}(\theta_{\epsilon^*})\|}$. Note also that

$$\alpha^* > \alpha^*\Phi\left(-\frac{\|\boldsymbol{\beta}^*\|}{\sigma}\right) = \omega(\epsilon^*)\Phi(-\epsilon^*/\sigma) > \underline{\eta},$$

where the equality follows from Steps 2–3, and the last inequality by the choice of $\underline{\eta}$. We can pick

$\underline{t} \in (0,1)$ sufficiently close to 1 so that for all $t \in [\underline{t}, 1]$, $\underline{\eta} < \frac{\boldsymbol{m}(\theta_{\epsilon^*})' t\boldsymbol{\beta}^*}{\|\boldsymbol{m}(\theta_{\epsilon^*})\|} = \lim_{n\to\infty} \frac{\boldsymbol{m}(\theta_{\epsilon^*})'\boldsymbol{m}(t\theta_n)}{\|\boldsymbol{m}(\theta_{\epsilon^*})\|}$ and $\underline{\eta} < t\alpha^* = \lim_{n\to\infty} L(t\theta_n)$. It follows that $t\boldsymbol{\gamma}^* = (t\alpha^*, t(\boldsymbol{\beta}^*)')' = \lim_{n\to\infty}(L(t\theta_n), \boldsymbol{m}(t\theta_n)')' \in \bar{\Gamma}_{+,\underline{\eta},\bar{\epsilon}}$ for all $t \in [\underline{t}, 1]$. Since $\boldsymbol{\gamma}^* \in \arg\max_{\boldsymbol{\gamma}=(\alpha,\boldsymbol{\beta}')'\in\bar{\Gamma}_{+,\underline{\eta},\bar{\epsilon}}} \alpha\Phi\left(-\frac{(\boldsymbol{\beta}^*)'\boldsymbol{\beta}}{\sigma\|\boldsymbol{\beta}^*\|}\right)$, we have

$$\boldsymbol{\gamma}^* \in \arg\max_{\boldsymbol{\gamma}\in[\underline{t}\boldsymbol{\gamma}^*,\boldsymbol{\gamma}^*]} \alpha\Phi\left(-\frac{(\boldsymbol{\beta}^*)'\boldsymbol{\beta}}{\sigma\|\boldsymbol{\beta}^*\|}\right).$$

This implies that

$$1 \in \arg\max_{t\in[\underline{t},1]} t\alpha^*\Phi\left(-\frac{(\boldsymbol{\beta}^*)'t\boldsymbol{\beta}^*}{\sigma\|\boldsymbol{\beta}^*\|}\right) = \arg\max_{t\in[\underline{t},1]} t\Phi\left(-\frac{t\|\boldsymbol{\beta}^*\|}{\sigma}\right).$$

By Lemma B.1, $t \mapsto t\Phi\left(-\frac{t\|\boldsymbol{\beta}^*\|}{\sigma}\right)$ is strictly increasing on $[0, \frac{a^*\sigma}{\|\boldsymbol{\beta}^*\|})$ and strictly decreasing on $(\frac{a^*\sigma}{\|\boldsymbol{\beta}^*\|}, \infty)$. Since $\frac{a^*\sigma}{\|\boldsymbol{\beta}^*\|} < 1$ by the hypothesis, $1 \notin \arg\max_{t\in[\underline{t},1]} t\Phi\left(-\frac{t\|\boldsymbol{\beta}^*\|}{\sigma}\right)$, which is a contradiction. □

Step 5. $\boldsymbol{m}(\theta_{\epsilon^*}) = \boldsymbol{\beta}^*$.

Proof. Suppose that $\boldsymbol{m}(\theta_{\epsilon^*}) \neq \boldsymbol{\beta}^*$. Pick any $\lambda \in (0,1)$, and consider the sequence $\{\theta_{\lambda,n}\}_{n=1}^{\infty}$, where $\theta_{\lambda,n} = \lambda\theta_{\epsilon^*} + (1-\lambda)\theta_n$. Since $\Theta_{+,\underline{\eta},\bar{\epsilon}}$ is convex, $\theta_{\lambda,n} \in \Theta_{+,\underline{\eta},\bar{\epsilon}}$ for all n. We have

$$\lim_{n\to\infty} L(\theta_{\lambda,n}) = \lambda L(\theta_{\epsilon^*}) + (1-\lambda)\lim_{n\to\infty} L(\theta_n) = \lambda\omega(\epsilon^*) + (1-\lambda)\alpha^* = \omega(\epsilon^*),$$

and

$$\lim_{n\to\infty} \|\boldsymbol{m}(\theta_{\lambda,n})\| = \|\lambda\boldsymbol{m}(\theta_{\epsilon^*}) + (1-\lambda)\lim_{n\to\infty} \boldsymbol{m}(\theta_n)\| = \|\lambda\boldsymbol{m}(\theta_{\epsilon^*}) + (1-\lambda)\boldsymbol{\beta}^*\| < \epsilon^*,$$

where the last inequality holds since $\{\boldsymbol{\beta} \in \mathbb{R}^n : \|\boldsymbol{\beta}\| \leq \epsilon^*\}$ is strictly convex. These imply that

$$\omega(\tilde{\epsilon}) = \sup\{L(\theta) : \theta \in \Theta, \|\boldsymbol{m}(\theta)\| \leq \tilde{\epsilon}\} = \omega(\epsilon^*)$$

for any $\tilde{\epsilon} \in (\lim_{n\to\infty} \|\boldsymbol{m}(\theta_{\lambda,n})\|, \epsilon^*)$. It follows that

$$\omega(\tilde{\epsilon})\Phi(-\tilde{\epsilon}/\sigma) = \omega(\epsilon^*)\Phi(-\tilde{\epsilon}/\sigma) > \omega(\epsilon^*)\Phi(-\epsilon^*/\sigma),$$

which contradicts the fact that ϵ^* maximizes $\omega(\epsilon)\Phi(-\epsilon/\sigma)$ over $[0, a^*\sigma]$. □

Step 6. $\theta_{\epsilon^*} \in \arg\max_{\theta\in\Theta:L(\theta)>0} L(\theta)\Phi\left(-\frac{\boldsymbol{m}(\theta_{\epsilon^*})'\boldsymbol{m}(\theta)}{\sigma\|\boldsymbol{m}(\theta_{\epsilon^*})\|}\right)$.

Proof. Since $\alpha^* = L(\theta_{\epsilon^*})$, $\boldsymbol{\beta}^* = \boldsymbol{m}(\theta_{\epsilon^*})$, and $\boldsymbol{\gamma}^* \in \arg\max_{\boldsymbol{\gamma}=(\alpha,\boldsymbol{\beta}')'\in\bar{\Gamma}_{+,\underline{\eta},\bar{\epsilon}}} \alpha\Phi\left(-\frac{(\boldsymbol{\beta}^*)'\boldsymbol{\beta}}{\sigma\|\boldsymbol{\beta}^*\|}\right)$,

$$\boldsymbol{\gamma}^* = (L(\theta_{\epsilon^*}), \boldsymbol{m}(\theta_{\epsilon^*})')' \in \arg\max_{\boldsymbol{\gamma}=(\alpha,\boldsymbol{\beta}')'\in\bar{\Gamma}_{+,\underline{\eta},\bar{\epsilon}}} \alpha\Phi\left(-\frac{\boldsymbol{m}(\theta_{\epsilon^*})'\boldsymbol{\beta}}{\sigma\|\boldsymbol{m}(\theta_{\epsilon^*})\|}\right), \tag{B.12}$$

which implies

$$\theta_{\epsilon^*} \in \arg \max_{\theta \in \Theta_{+,\underline{\eta},\bar{\epsilon}}} L(\theta)\Phi\left(-\frac{\boldsymbol{m}(\theta_{\epsilon^*})'\boldsymbol{m}(\theta)}{\sigma\|\boldsymbol{m}(\theta_{\epsilon^*})\|}\right).$$

Pick any $\theta \in \Theta$ such that $L(\theta) > 0$ and $\theta \notin \Theta_{+,\underline{\eta},\bar{\epsilon}}$. Let $\boldsymbol{\gamma} = (\alpha, \boldsymbol{\beta}')' = (L(\theta), \boldsymbol{m}(\theta)')'$. Since $\underline{\eta} < \omega(\epsilon^*) = \alpha^*$ and $\underline{\eta} < \epsilon^* < \bar{\epsilon}$ by the choice of $\underline{\eta}$ and $\bar{\epsilon}$, we can pick $t \in (0,1)$ sufficiently close to 1 so that

$$L((1-t)\theta + t\theta_{\epsilon^*}) = (1-t)\alpha + t\alpha^* > \underline{\eta},$$

$$\frac{\boldsymbol{m}(\theta_{\epsilon^*})'\boldsymbol{m}((1-t)\theta + t\theta_{\epsilon^*})}{\|\boldsymbol{m}(\theta_{\epsilon^*})\|} = \frac{\boldsymbol{m}(\theta_{\epsilon^*})'((1-t)\boldsymbol{\beta} + t\boldsymbol{\beta}^*)}{\|\boldsymbol{m}(\theta_{\epsilon^*})\|} = (1-t)\frac{\boldsymbol{m}(\theta_{\epsilon^*})'\boldsymbol{\beta}}{\|\boldsymbol{m}(\theta_{\epsilon^*})\|} + t\epsilon^* > \underline{\eta},$$

and

$$\|\boldsymbol{m}((1-t)\theta + t\theta_{\epsilon^*})\| = \|(1-t)\boldsymbol{\beta} + t\boldsymbol{\beta}^*\| \leq (1-t)\|\boldsymbol{\beta}\| + t\epsilon^* < \bar{\epsilon}.$$

It follows that $(1-t)\theta + t\theta_{\epsilon^*} \in \Theta_{+,\underline{\eta},\bar{\epsilon}}$ and $(1-t)\boldsymbol{\gamma} + t\boldsymbol{\gamma}^* \in \Gamma_{+,\underline{\eta},\bar{\epsilon}}$. By Eq. (B.12), this implies that

$$\alpha^*\Phi\left(-\frac{\boldsymbol{m}(\theta_{\epsilon^*})'\boldsymbol{\beta}^*}{\sigma\|\boldsymbol{m}(\theta_{\epsilon^*})\|}\right) \geq [(1-t)\alpha + t\alpha^*]\Phi\left(-(1-t)\frac{\boldsymbol{m}(\theta_{\epsilon^*})'\boldsymbol{\beta}}{\sigma\|\boldsymbol{m}(\theta_{\epsilon^*})\|} - t\frac{\boldsymbol{m}(\theta_{\epsilon^*})'\boldsymbol{\beta}^*}{\sigma\|\boldsymbol{m}(\theta_{\epsilon^*})\|}\right).$$

Since the function $(a, b) \mapsto a\Phi(-b)$ is strictly quasi-concave on $(0, \infty) \times \mathbb{R}$ by Lemma B.3,

$$\begin{aligned}&[(1-t)\alpha + t\alpha^*]\Phi\left(-(1-t)\frac{\boldsymbol{m}(\theta_{\epsilon^*})'\boldsymbol{\beta}}{\sigma\|\boldsymbol{m}(\theta_{\epsilon^*})\|} - t\frac{\boldsymbol{m}(\theta_{\epsilon^*})'\boldsymbol{\beta}^*}{\sigma\|\boldsymbol{m}(\theta_{\epsilon^*})\|}\right)\\ &> \min\left\{\alpha\Phi\left(-\frac{\boldsymbol{m}(\theta_{\epsilon^*})'\boldsymbol{\beta}}{\sigma\|\boldsymbol{m}(\theta_{\epsilon^*})\|}\right), \alpha^*\Phi\left(-\frac{\boldsymbol{m}(\theta_{\epsilon^*})'\boldsymbol{\beta}^*}{\sigma\|\boldsymbol{m}(\theta_{\epsilon^*})\|}\right)\right\}.\end{aligned}$$

Therefore,

$$\alpha^*\Phi\left(-\frac{\boldsymbol{m}(\theta_{\epsilon^*})'\boldsymbol{\beta}^*}{\sigma\|\boldsymbol{m}(\theta_{\epsilon^*})\|}\right) > \alpha\Phi\left(-\frac{\boldsymbol{m}(\theta_{\epsilon^*})'\boldsymbol{\beta}}{\sigma\|\boldsymbol{m}(\theta_{\epsilon^*})\|}\right),$$

which implies that

$$L(\theta_{\epsilon^*})\Phi\left(-\frac{\boldsymbol{m}(\theta_{\epsilon^*})'\boldsymbol{m}(\theta_{\epsilon^*})}{\sigma\|\boldsymbol{m}(\theta_{\epsilon^*})\|}\right) > L(\theta)\Phi\left(-\frac{\boldsymbol{m}(\theta_{\epsilon^*})'\boldsymbol{m}(\theta)}{\sigma\|\boldsymbol{m}(\theta_{\epsilon^*})\|}\right).$$

The conclusion then follows.

□

B.9 Proof of Lemma 5

Let θ_ϵ attain the modulus of continuity at $\epsilon \in [0, \bar{\epsilon}]$. Also, let $\iota \in \Theta$ be a parameter value that satisfies Assumption 1(c). Without loss of generality, I normalize ι so that $L(\iota) = 1$.

First, Assumption 3(a) holds under Assumption 1(a) and (c), since if $\|\boldsymbol{m}(\theta_\epsilon)\| < \epsilon$, then there exists $c > 0$ such that $\theta_\epsilon + c\iota \in \Theta$, $L(\theta_\epsilon + c\iota) > L(\theta_\epsilon)$, and $\|\boldsymbol{m}(\theta_\epsilon + c\iota)\| < \epsilon$, which contradicts the definition of θ_ϵ.

Under Assumption 1(b), θ_0 satisfies $L(\theta_0) = \omega(0) = \rho(0)$ as shown in the proof of Lemma 7. Additionally, $(\boldsymbol{w}^*)'\boldsymbol{m}(\theta_0) = 0$ by construction. Applying Lemma A.2, we have that $\rho'(0) = \frac{1}{(\boldsymbol{w}^*)'\boldsymbol{m}(\iota)}$.

By Lemma A.1, for any sufficiently small $\epsilon > 0$, $\omega'(\epsilon) = \frac{\epsilon}{\boldsymbol{m}(\iota)'\boldsymbol{m}(\theta_\epsilon)}$. Since $\omega(\cdot)$ is differentiable and concave, it is continuously differentiable. Therefore,

$$\omega'(0) = \lim_{\epsilon \to 0} \omega'(\epsilon) = \lim_{\epsilon \to 0} \frac{1}{\boldsymbol{m}(\iota)'\boldsymbol{m}(\theta_\epsilon/\epsilon)} = \frac{1}{\boldsymbol{m}(\iota)'\boldsymbol{w}^*} = \rho'(0),$$

where the second last equality holds by Assumption 1(b) and the fact that $\|\boldsymbol{m}(\theta_\epsilon)\| = \epsilon$. Since $\omega'(0)$ is nonnegative by the definition of the modulus and $\boldsymbol{m}(\iota)'\boldsymbol{w}^*$ is finite, $\omega'(0)$ and $\rho'(0)$ must be positive.

Let $\sigma^* = 2\phi(0)\frac{\rho(0)}{\rho'(0)} = 2\phi(0)\frac{\omega(0)}{\omega'(0)}$ and $g(\epsilon) = \rho(\epsilon)\Phi\left(-\frac{\epsilon}{\sigma^*}\right)$. By differentiating $g(\cdot)$, we have

$$g'(\epsilon) = \rho'(\epsilon)\Phi\left(-\frac{\epsilon}{\sigma^*}\right) - \rho(\epsilon)\phi\left(-\frac{\epsilon}{\sigma^*}\right)/\sigma^*.$$

$g'(0) = 0$ by the choice of σ^*. Since $\rho(\cdot)$ is concave (as shown in the proof of Lemma A.2) and differentiable, $\rho'(\cdot)$ is continuous. Let $\epsilon_1 = \inf\{(\boldsymbol{w}^*)'\boldsymbol{m}(\theta) : \theta \in \Theta\}$, $\epsilon_1' = \sup\{\epsilon : \rho'(\epsilon) \geq 0\}$, and $\epsilon_2' = \sup\{\epsilon : \rho(\epsilon) \geq 0\}$. Since $\rho(\cdot)$ is concave and hence $\rho'(\cdot)$ is nonincreasing, $\rho(\cdot)$ is nondecreasing on $(\epsilon_1, \epsilon_1']$. Also, since $\rho'(0) > 0$, we have $\epsilon_1' > 0$, and $\rho(\cdot)$ is nonconstant on $(\epsilon_1, \epsilon_1']$. By Lemma B.1, $g(\cdot)$ is maximized at 0 over $(\epsilon_1, \epsilon_1']$. For $\epsilon \in (\epsilon_1', \epsilon_2']$, $\rho'(\epsilon) \leq 0$ and $\rho(\epsilon) \geq 0$, so that $g'(\epsilon) \leq 0$. Therefore, $g(\cdot)$ is maximized at 0 over $(\epsilon_1, \epsilon_2']$. Finally, $g(\epsilon) \leq 0$ for all $\epsilon > \epsilon_2'$. Since $g(0) = \rho(0)/2 = \omega(0)/2 \geq 0$, $g(\cdot)$ is maximized at 0 globally. □

B.10 Proof of Lemma 6

Since $\boldsymbol{Y} \sim \mathcal{N}(0, \sigma^2 \boldsymbol{I}_n)$ under θ for all $\theta \in [-\bar{\theta}, \bar{\theta}]$, $\mathbb{E}_\theta[\delta(\boldsymbol{Y})] = \mathbb{E}_{\boldsymbol{Y}\sim\mathcal{N}(\boldsymbol{0},\sigma^2\boldsymbol{I}_n)}[\delta(\boldsymbol{Y})]$ is constant over $\theta \in [-\bar{\theta}, \bar{\theta}]$. Therefore, the maximum regret of decision rule δ is

$$\sup_{\theta \in [-\bar{\theta}, \bar{\theta}]} R(\delta, \theta) = \begin{cases} L(\bar{\theta})(1 - \mathbb{E}_{\boldsymbol{Y}\sim\mathcal{N}(\boldsymbol{0},\sigma^2\boldsymbol{I}_n)}[\delta(\boldsymbol{Y})]) & \text{if } \mathbb{E}_{\boldsymbol{Y}\sim\mathcal{N}(\boldsymbol{0},\sigma^2\boldsymbol{I}_n)}[\delta(\boldsymbol{Y})] < 1/2, \\ L(\bar{\theta})/2 & \text{if } \mathbb{E}_{\boldsymbol{Y}\sim\mathcal{N}(\boldsymbol{0},\sigma^2\boldsymbol{I}_n)}[\delta(\boldsymbol{Y})] = 1/2, \\ (-L(-\bar{\theta}))\mathbb{E}_{\boldsymbol{Y}\sim\mathcal{N}(\boldsymbol{0},\sigma^2\boldsymbol{I}_n)}[\delta(\boldsymbol{Y})] & \text{if } \mathbb{E}_{\boldsymbol{Y}\sim\mathcal{N}(\boldsymbol{0},\sigma^2\boldsymbol{I}_n)}[\delta(\boldsymbol{Y})] > 1/2. \end{cases}$$

Thus, any decision rule δ^* such that $\mathbb{E}_{\boldsymbol{Y}\sim\mathcal{N}(\boldsymbol{0},\sigma^2\boldsymbol{I}_n)}[\delta^*(\boldsymbol{Y})] = \frac{1}{2}$ is minimax regret. The minimax risk is given by

$$\mathcal{R}(\sigma; [-\bar{\theta}, \bar{\theta}]) = \frac{L(\bar{\theta})}{2}.$$

□

B.11 Proof of Lemma 7

The maximum regret of δ^* over Θ is

$$
\begin{aligned}
\sup_{\theta\in\Theta} R(\delta^*,\theta) &= \sup_{\theta\in\Theta} L(\theta)\Phi\left(-\frac{(\boldsymbol{w}^*)'\boldsymbol{m}(\theta)}{\sigma^*}\right) \\
&= \sup_{\epsilon\in\mathbb{R}} \sup_{\theta\in\Theta:(\boldsymbol{w}^*)'\boldsymbol{m}(\theta)=\epsilon} L(\theta)\Phi\left(-\frac{\epsilon}{\sigma^*}\right) \\
&= \sup_{\epsilon\in\mathbb{R}} \rho(\epsilon)\Phi\left(-\frac{\epsilon}{\sigma^*}\right) \\
&= \frac{1}{2}\rho(0),
\end{aligned}
$$

where the last equality holds by Assumption 3(c).

Since $(\boldsymbol{w}^*)'\boldsymbol{m}(\theta) = 0$ for any $\theta \in \Theta$ such that $\boldsymbol{m}(\theta) = \mathbf{0}$, it follows that $\rho(0) \geq \omega(0)$ by the definition of $\omega(\cdot)$ and $\rho(\cdot)$. If $\rho(0) = \omega(0)$, then $\sup_{\theta\in\Theta} R(\delta^*,\theta)$ is attained at θ_0, and $\sup_{\theta\in\Theta} R(\delta^*,\theta) = \frac{1}{2}\omega(0)$. Below, I show that $\rho(0) = \omega(0)$. Suppose to the contrary that $\rho(0) > \omega(0)$. Then there exists $\theta \in \Theta$ such that $(\boldsymbol{w}^*)'\boldsymbol{m}(\theta) = 0$, $\boldsymbol{m}(\theta) \neq \mathbf{0}$, and $L(\theta) > \omega(0)$. For $\epsilon \in (0,\bar{\epsilon}]$, by the Cauchy-Schwarz inequality,

$$
\epsilon^{-1}\left|\frac{\boldsymbol{m}(\theta_\epsilon)'}{\epsilon}\boldsymbol{m}(\theta)\right| = \epsilon^{-1}\left|\left(\frac{\boldsymbol{m}(\theta_\epsilon)}{\|\boldsymbol{m}(\theta_\epsilon)\|} - \boldsymbol{w}^*\right)'\boldsymbol{m}(\theta)\right| \leq \epsilon^{-1}\left\|\frac{\boldsymbol{m}(\theta_\epsilon)}{\|\boldsymbol{m}(\theta_\epsilon)\|} - \boldsymbol{w}^*\right\|\|\boldsymbol{m}(\theta)\|,
$$

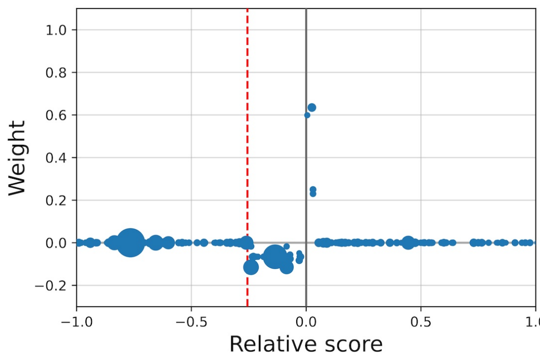
where θ_ϵ attains the modulus of continuity at ϵ, and the first equality follows from Assumption 3(a) and the fact that $(\boldsymbol{w}^*)'\boldsymbol{m}(\theta) = 0$. The right-hand side converges to zero as $\epsilon \to 0$ by Assumption 3(b). Also, $\omega(\cdot)$ is concave and hence is continuous. Therefore, $\epsilon^{-1}\frac{\boldsymbol{m}(\theta_\epsilon)'}{\epsilon}\boldsymbol{m}(\theta) \leq 1/2$ and $L(\theta) > \omega(\epsilon)$ for any sufficiently small $\epsilon > 0$. Pick such an $\epsilon > 0$, and let $\theta_\lambda = \lambda\theta_\epsilon + (1-\lambda)\theta$ for $\lambda \in \mathbb{R}$. By simple algebra,

$$
\begin{aligned}
\|\boldsymbol{m}(\theta_\lambda)\|^2 &= \lambda^2\|\boldsymbol{m}(\theta_\epsilon)\|^2 + 2\lambda(1-\lambda)\boldsymbol{m}(\theta_\epsilon)'\boldsymbol{m}(\theta) + (1-\lambda)^2\|\boldsymbol{m}(\theta)\|^2 \\
&\leq \lambda^2\epsilon^2 + \lambda(1-\lambda)\epsilon^2 + (1-\lambda)^2\|\boldsymbol{m}(\theta)\|^2 \\
&= \|\boldsymbol{m}(\theta)\|^2\lambda^2 - (2\|\boldsymbol{m}(\theta)\|^2 - \epsilon^2)\lambda + \|\boldsymbol{m}(\theta)\|^2.
\end{aligned}
$$

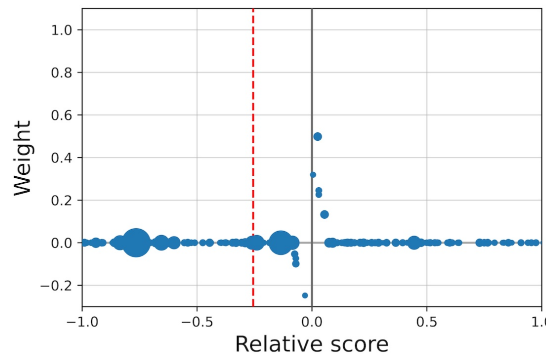
Observe that the right-hand side is quadratic in λ, minimized at $\lambda = \frac{2\|\boldsymbol{m}(\theta)\|^2-\epsilon^2}{2\|\boldsymbol{m}(\theta)\|^2} < 1$, and equal to ϵ^2 when $\lambda = 1$. This implies that $\|\boldsymbol{m}(\theta_\lambda)\|^2 < \epsilon^2$ for any λ close to one. However, $L(\theta_\lambda) = \lambda L(\theta_\epsilon) + (1-\lambda)L(\theta) > \omega(\epsilon)$ for all $\lambda \in (0,1)$, which contradicts the assumption that θ_ϵ attains the modulus of continuity at ϵ. □

C Empirical Policy Application: Additional Figures

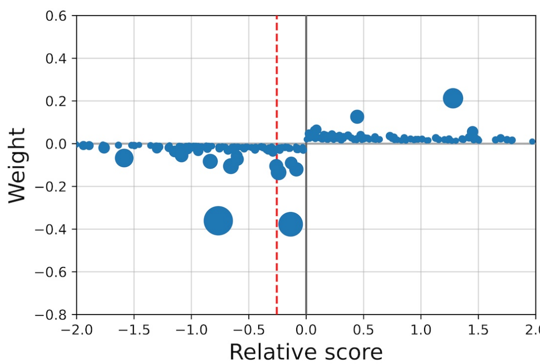
Figure A.1: Weight to Each Village Attached by Plug-in Rules



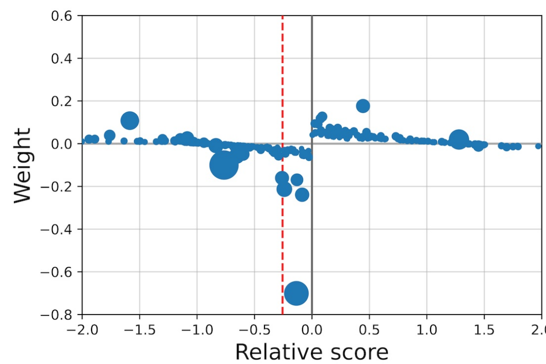
(a) Plug-in MSE Rule, $C = 0.5$



(b) Plug-in MSE Rule Under Constant Effect, $C = 0.5$



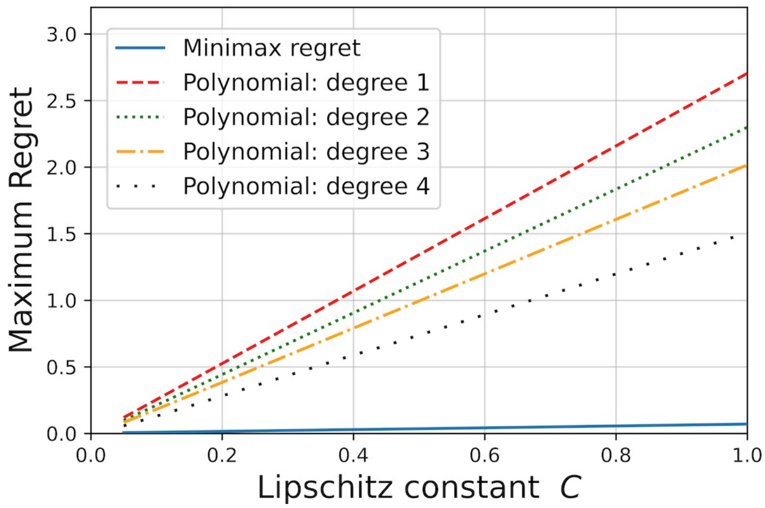
(c) Polynomial of Degree 2



(d) Polynomial of Degree 4

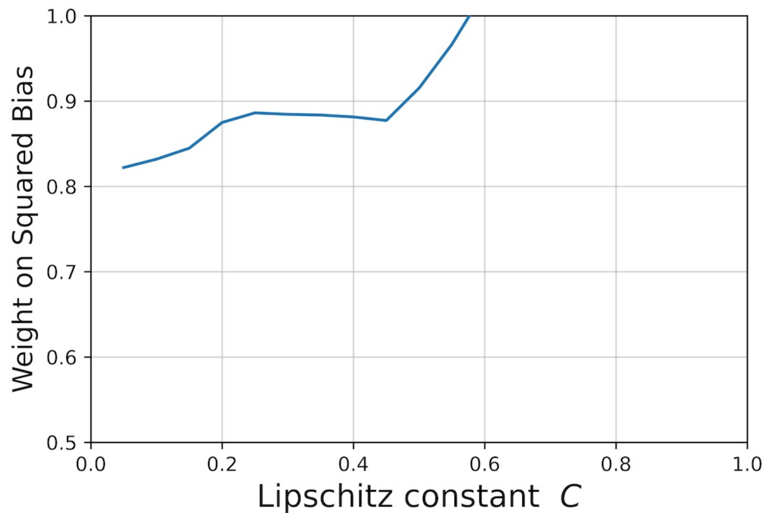
Notes: This figure shows the weight w_i attached to each village by the plug-in decision rules of the form $\delta^*(\boldsymbol{Y}) = \mathbf{1}\{\sum_{i=1}^n w_i Y_i \geq 0\}$. The weights are normalized so that $\sum_{i=1}^n w_i^2 = 1$. The horizontal axis indicates the relative score of each village. Each circle corresponds to each village. The size of circles is proportional to the inverse of the standard error of the enrollment rate Y_i. The vertical dashed line corresponds to the new cutoff -0.256. Panels (a) and (b) show the results for the plug-in rules based on the linear minimax MSE estimators with or without the assumption of constant conditional treatment effects when the Lipschitz constant C is 0.5. Panels (c) and (d) show the results for the plug-in rules based on the polynomial regression estimators of degrees 2 and 4, respectively.

Figure A.2: Maximum Regret of Minimax Regret Rule and Plug-in Rules Based on Polynomial Regression Estimators



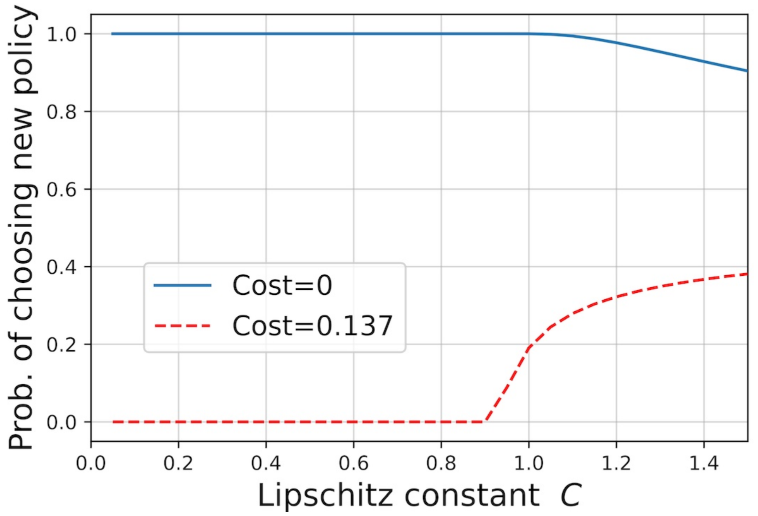
Notes: This figure shows the maximum regret of the minimax regret rule and the plug-in rules based on the polynomial regression estimators of degrees 1 to 4. The maximum regret is computed by setting the true function class of the counterfactual outcome function to the Lipschitz class. The maximum regret is normalized so that the unit is the same as that of the enrollment rate. I report the results for the range $[0.05, 0.1, ..., 0.95, 1]$ of the Lipschitz constant C.

Figure A.3: Weight on Bias Placed by Minimax Regret Rule

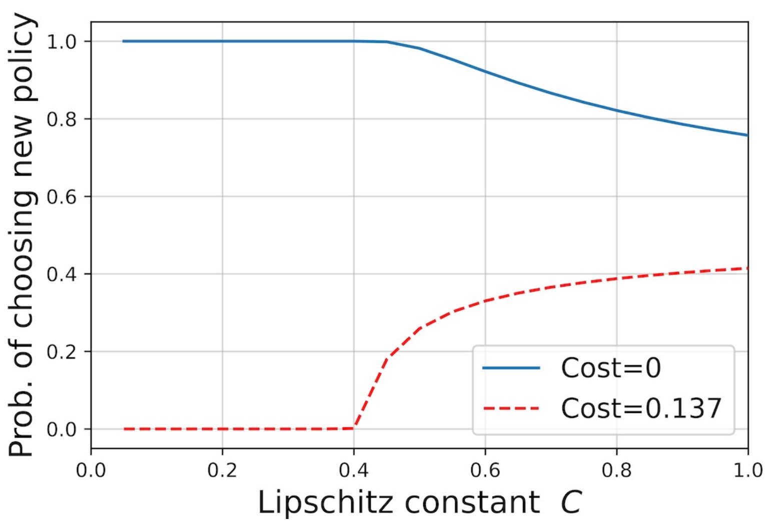


Notes: This figure shows the weight $\alpha \in [1/2, 1]$ on the squared worst-case bias placed by the minimax regret rule of the form $\delta^*(\boldsymbol{Y}) = \mathbf{1}\{\tilde{\boldsymbol{w}}'\boldsymbol{Y} \geq 0\}$, where $\tilde{\boldsymbol{w}} \in \arg\min_{\boldsymbol{w} \in \mathbb{R}^n} \left\{\alpha \cdot \left(\sup_{f \in \mathcal{F}_{\mathrm{Lip}}(C)} \mathbb{E}_f[\boldsymbol{w}'\boldsymbol{Y} - L(f)]\right)^2 + (1-\alpha) \cdot \mathrm{Var}(\boldsymbol{w}'\boldsymbol{Y})\right\}$. I only report the results for the Lipschitz constant $C < 0.6$ since the minimax regret rule is randomized for $C \geq 0.6$.

Figure A.4: Optimal Decisions for Alternative New Policies



(a) New Policy of Constructing Schools in 10% of Villages



(b) New Policy of Constructing Schools in 30% of Villages

Notes: This figure shows the probability of choosing the new policy computed by the minimax regret rule. The new policy is to construct BRIGHT schools in previously ineligible villages whose relative scores are in the top 10% (Panel (a)) or in the top 30% (Panel (b)). The solid line shows the results for the scenario where we ignore the policy cost. The dashed line shows the results for the scenario where the policy cost measured in the unit of the enrollment rate is 0.137. I report the results for the range $[0.05, 0.1, ..., 1.45, 1.5]$ of the Lipschitz constant C in Panel (a) and for the range $[0.05, 0.1, ..., 0.95, 1]$ in Panel (b).