Measuring Precision of Statistical Inference on Partially Identified Parameters

Aleksey Tetenov[1]

April 14, 2009

[1]Collegio Carlo Alberto, aleksey.tetenov@carloalberto.org. I am grateful to Chuck Manski, Elie Tamer, Luca Anderlini and Keisuke Hirano for their helpful comments. I have also benefitted from the opportunity to present this work at Collegio Carlo Alberto, LAMES 2008 and ICEEE 2009.

Abstract

Planners of surveys and experiments that partially identify parameters of interest face trade offs between using limited resources to reduce sampling error or to reduce the extent of partial identification. Researchers who previously attempted evaluating these trade offs used the length of confidence intervals for the identification region to measure the precision of inference. I show that other reasonable measures of statistical precision yield qualitatively different conclusions, often implying higher value to reducing the extent of partial identification. I consider three alternative measures - maximum mean squared error, maximum mean absolute deviation, and maximum regret (applicable when the purpose of estimation is binary treatment choice). I analytically derive and compare estimation precision and tradeoffs implied by these measures in a simple statistical problem with normally distributed sample data and interval partial identification.

JEL Classification: C21, C44, C83.

Keywords: partial identification, statistical treatment choice, mean absolute error, mean squared error, minimax regret, survey planning

1 Introduction

It has become widely recognized that many types of statistical data only partially identify the parameters of interest as simple as population means, meaning that the parameters cannot be estimated with arbitrary precision simply by increasing the sample size. Statisticians designing surveys and experiments which generate such data could use limited resources either to reduce the extent of partial identification or to reduce sampling error. The former can be accomplished, for example, by putting more effort into pursuing sampled population members who did not respond to a survey. The latter by increasing sample size or improving measurement precision. To inform these choices, it is useful to analytically derive the relative effects of both margins of planning on the precision of inference, which the planner could then compare to their relative costs.

The problem was first considered in the Cochran-Mosteller-Tukey report on the Kinsey study published in 1954. Concerned with nonrandom nonresponse to the study's questions, CMT advocated a conservative approach to inference that sets limits on population parameters by allowing for any values of the variable in the part of the population that was not sampled or refused to respond. A variety of applications of the same approach, now known as *partial identification*, has been developed by Manski (1995, 2007a) and other researchers. CMT calculated for different sample sizes and refusal rates the relative effects of reducing nonresponse or increasing the sample size on the precision of inference about the population means. They judged the precision of inference by the length of a 95% confidence interval around the estimated identification region. The same measure of precision has been used to illustrate the effects of missing data on the precision of inference by Horowitz and Manski (1998) and McFadden (2006).

Length of a confidence interval for the identification region is not the only reasonable way to measure the precision of inference on the parameter of interest. In this paper I consider other measures of precision and show that they yield qualitatively different conclusions about the relative merits of reducing sampling error and reducing the extent of partial identification. First, I consider the maximum mean squared error (MSE) of the point estimate around the true value of the parameter, which has often been used by statisticians to measure the precision of estimators of point identified parameters. I also consider the maximum mean absolute deviation (MAD) of the estimate around the true parameter value, a popular alternative to the squared error.

Another measure considered in this paper is the maximum regret of a statistical treatment rule. It is applicable when the parameter of interest is the difference in average returns of two mutually exclusive policies or treatments for a population of interest and the goal of inference is to decide which one should be used. Regret, then, is the average welfare loss incurred from choosing an inferior treatment for the population based on the observed statistical data. In recent years, econometricians started studying statistical treatment rules that minimize maximum regret both when the average treatment effect of interest is point identified (Manski 2004, 2005; Hirano and Porter 2008; Stoye 2007b; Schlag 2007; Manski and Tetenov 2007) and when it is partially identified (Manski 2007a, 2007b, 2008a, 2008b; Stoye 2007a, 2007c).

I apply these measures of precision to the following partial identification problem. Let the real-valued parameter of interest $\theta = \theta_O + \theta_U$ be the sum of a point identified component θ_O and a partially identified component θ_U. For the point identified component θ_O, the statistician observes an unbiased normally distributed estimate with known standard error σ. The partially identified component θ_U is only known to lie in a given bounded interval of length $2P$. The problem is deliberately simplified to demostrate in an analytically tractable setting the qualitative differences between the conclusions about the relative benefits of reducing sampling error vs. narrowing the identification region drawn based on alternative measures of precision. I derive minimax estimators of θ under maximum MSE and maximum MAD criteria and a minimax regret statistical treatment rule under the maximum regret criterion. I show that for sufficiently small values of $\frac{\sigma}{P}$, all of the considered measures of precision imply greater relative importance of addressing the partial identificaton problem than measuring the length of confidence intervals suggests. For maximum regret, the result is particularly strong. If the standard error σ falls below a certain proportion of the width of the identification region $2P$, then reducing it even further does not reduce maximum regret. Thus, more precise inference for treatment choice could be made only by reducing the width of the identification region. The same effect has been shown by Stoye (2007c) in a problem of treatment choice based on random samples of binary treatment outcomes with missing data.

The paper proceeds as follows. Section 2 describes the statistical problem and reviews the results of measuring precision of inference by the length of confidence intervals. In section 3, I derive estimators of θ that minimize maximum MSE and maximum MAD and evaluate the effect of changing the parameters of the problem on its minimax MSE and MAD. In section 4, I consider the problem from a statistical treatment choice perspective, derive a minimax regret

statistical treatment rule and evaluate the effects of changing the parameters of the problem (σ and P) on its minimax regret. Section 5 concludes and the appendix collects all proofs.

2 Statistical Setting and the Confidence Interval Approach

I will consider the following partial identification problem. The parameter of interest to the statistician is

$$\theta = \theta_O + \theta_U.$$

$\theta_O \in \mathbb{R}$ is a point identified (observable) component, for which the statistician could obtain an unbiased normally distributed estimate X with standard error σ:

$$X \sim \mathcal{N}\left(\theta_O, \sigma^2\right).$$

θ_U is a partially identified (unobservable) component, which is only known to lie in a bounded interval of length $2P$:

$$\theta_U \in [-P, P].$$

The restriction that θ_U lies in a symmetric interval around zero is without loss of generality.

For example, θ could be the difference between average potential outcomes of two alternative treatments on a population of interest. Suppose that θ_O is the average outcome of one treatment, which is point identified by experimental data generated by assigning that treatment to a random sample of population members, while $-\theta_U$ is the average outcome of the second treatment, which is known to lie within a given interval based on observational data.

Alternatively, θ_O could be the average difference in potential outcomes of the two treatments point identified by an experiment that randomly assigned one of two treatments to members of the population of interest, while $-\theta_U$ is the difference between future costs of the two treatments that the randomized experiment does not reveal.

In this setting, the pair (σ, P) describes the experimental design parameters. The main question of this chapter is how do these design parameters affect the precision of inference on θ that the statistician could carry out based on the results of the experiment (observation of X). Formally, let the function

$$M(\sigma, P) \geq 0$$

be a particular measure of maximum precision with which the statistician can carry out inference on θ based on the data from an experiment with design parameters (σ, P). Lower values of $M(\sigma, P)$ will correspond to more precise inference and $M(\sigma, P) = 0$ will correspond to perfect precision (for example, $M(0, 0) = 0$). Let a differentiable function $b(\cdot) \geq 0$, $b' < 0$ denote the economic benefit of inference with a given level of precision and let a differentiable function $c(\sigma, P), c_\sigma < 0, c_P < 0$ denote the costs of conducting an experiment with design parameters σ and P. Then the statistical planning problem is to maximize the net benefit of the experiment

$$\max_{\sigma, P} \left[b\left(M\left(\sigma, P \right) \right) - c\left(\sigma, P \right) \right].$$

If M is differentiable with partial derivatives $M_\sigma > 0$ and $M_P > 0$, a necessary condition for a pair (σ^*, P^*) with $\sigma^* > 0$ and $P^* > 0$ to be a solution to the planning problem is that

$$\frac{M_\sigma\left(\sigma^*, P^* \right)}{M_P\left(\sigma^*, P^* \right)} = \frac{c_\sigma\left(\sigma^*, P^* \right)}{c_P\left(\sigma^*, P^* \right)}.$$

If these ratios are unequal, then it is possible to adjust σ and P in a way that improves precision without increasing costs. I will evaluate a few functions $M(\sigma, P)$ based on different criteria of precision and derive the $\frac{M_\sigma}{M_P}$ ratios for them. Survey and experiment planners could compare these ratios to the marginal cost ratio $\frac{c_\sigma}{c_P}$ and see whether a a proposed allocation of resources maximizes the precision of inference for a given budget. These conclusions could be made without specifying the benefit function $b(\cdot)$. Knowledge of $b(\cdot)$ is required, however, to determine the optimal size of a survey or experiment's budget.

First, let's consider using the length of a $100(1-\alpha)\,\%$ confidence interval for the identification interval as the measure of precision. In this model, the identification set for the parameter of interest θ is

$$\theta \in \left[\theta_O - P, \theta_O + P \right]. \tag{1}$$

Given that the random experimental outcome X is normally distributed with mean θ_O and standard error σ, the confidence interval

$$\left[X - P - \Phi^{-1}\left(1 - \alpha/2 \right) \sigma, X + P + \Phi^{-1}\left(1 - \alpha/2 \right) \sigma \right] \tag{2}$$

contains the identification set (1) exactly with probability $1-\alpha$. Φ denotes the standard normal c.d.f., so for the conventional 95% confidence intervals, for example, $\Phi^{-1}(1 - \alpha/2) \approx 1.96$. The

precision of inference from an experiment with parameters (σ, P), as measured by the length of a $100\,(1-\alpha)\,\%$ confidence interval then equals

$$M^{CI(\alpha)}(\sigma, P) \equiv 2\Phi^{-1}(1-\alpha/2)\,\sigma + 2P.$$

The marginal effects of changes in σ and P (partial derivatives of $M^{CI(\alpha)}$) equal

$$\begin{aligned} M_{\sigma}^{CI(\alpha)} &= 2\Phi^{-1}(1-\alpha/2), \\ \text{and } M_{P}^{CI(\alpha)} &= 2. \end{aligned}$$

The ratio of these marginal effects equals

$$\frac{M_{\sigma}^{CI(\alpha)}}{M_{P}^{CI(\alpha)}} = \Phi^{-1}(1-\alpha/2). \tag{3}$$

Thus, if the length of conventional 95% confidence intervals is used as a measure of precision, then a reduction of the standard error σ by ε always brings the same improvement as a reduction of the half-length P of the identification interval by 1.96ε. Note that the evaluation of the relative effects of reducing the sampling error and the extent of partial identification depends on the chosen confidence level $100\,(1-\alpha)\,\%$. Thus, using a 99% confidence level instead of 95% would imply a relatively higher value of reducing the standard error instead of reducing the extent of partial identification.

3 Minimax Mean Squared Error and Mean Absolute Deviation

Suppose, now, that instead of an interval the statistician is asked to provide a single point estimate of θ. Let the estimator $\hat{\theta}(X)$ be a function mapping the observed experimental outcome X into the estimate that the statistician provides upon observing X. There is a long tradition in statistics of measuring the precision of point estimators by their expected loss

$$E^{X} L\left(\hat{\theta}(X) - \theta\right), \tag{4}$$

where the expectation is taken with respect to the distribution of X for fixed values of θ_O and θ_U. Expected loss differs across values of θ_O and θ_U, its maximum value over the parameter

space

$$\sup_{\substack{\theta_O\in\mathbb{R},\\ \theta_U\in[-P,P]}} E^X L\left(\hat{\theta}(X)-(\theta_O+\theta_U)\right) \tag{5}$$

is a conservative measure of the precision of estimator $\hat{\theta}(X)$. If $\hat{\theta}(X)$ is optimal in the sense of minimizing (5), then its maximum expected loss could be used as a measure of precision of the experiment

$$M^L(\sigma,P)\equiv \sup_{\substack{\theta_O\in\mathbb{R},\\ \theta_U\in[-P,P]}} E^X L\left(\hat{\theta}(X)-(\theta_O+\theta_U)\right).$$

Proposition 2 shows that a simple estimator $\hat{\theta}^*(X)=X$ minimizes maximum expected loss (5) for a broad class of symmetric convex loss functions. This class includes commonly used.square loss $L(t)=t^2$ and absolute loss $L(t)=|t|$, for which I derive additional specific results afterwards. Formally, suppose that the loss function $L:\mathbb{R}\to\mathbb{R}_+$ satisfies the following conditions:

Condition 1 *a) L is symmetric ($L(t)=L(-t)$),*

b) L is convex,

c) $L(0)=0$,

d) $L(t)>0$ for some $t>0$,

e) $L(t)\le q\cdot\exp(rt)$ for all $t\ge 0$ and some constants $q>0, r>0$.

Then L is also continuous and non-decreasing on $[0,+\infty)$.

Proposition 2 *If loss function L satisfies Condition 1, $\theta_O\in\mathbb{R}$, $\theta_U\in[-P,P]$, and $X\sim\mathcal{N}(\theta_O,\sigma^2)$, then the estimator $\hat{\theta}^*(X)=X$ minimizes maximum expected loss (5), which for $\sigma>0$ and $P\ge 0$ equals*

$$M^L(\sigma,P)=\int_{-\infty}^{+\infty} L(t)\frac{1}{\sigma}\phi\left(\frac{t-P}{\sigma}\right)dt \tag{6}$$

and for $\sigma=0$ and $P\ge 0$, $M^L(0,P)=L(P)$.

Both square and absolute loss functions satisfy Condition 1. For them, (6) and its partial derivatives could be derived in closed form. In case of square loss $L(t)=t^2$, the maximum mean squared error

$$\sup_{\theta_O,\theta_U} E^X\left(\hat{\theta}^*(X)-(\theta_O+\theta_U)\right)^2 \tag{7}$$

of $\hat{\theta}^{*}(X) = X$ equals

$$M^{MSE}(\sigma, P) = \sigma^2 + P^2. \tag{8}$$

The marginal effects of changes in σ and P on the maximum mean squared error equal

$$\begin{aligned} M_{\sigma}^{MSE} &= 2\sigma, \\ \text{and } M_{P}^{MSE} &= 2P. \end{aligned}$$

The ratio of these marginal effects equals

$$\frac{M_{\sigma}^{MSE}}{M_{P}^{MSE}} = \frac{\sigma}{P}. \tag{9}$$

This ratio shows that using M^{MSE} as a measure of precision yields qualitatively different conclusions about the optimal choices of σ and P than using $M^{CI(\alpha)}$. Whenever $\frac{\sigma}{P} < \Phi^{-1}(1-\alpha/2)$,

$$\frac{M_{\sigma}^{MSE}}{M_{P}^{MSE}} < \frac{M_{\sigma}^{CI(\alpha)}}{M_{P}^{CI(\alpha)}}.$$

The maximum MSE measure of precision implies lower importance of further reducing standard errors than does the length of confidence interval measure. For the conventional 95% confidence intervals $\Phi^{-1}(.975) \approx 1.96$. Thus, in evaluating any proposed experiment or survey in which the standard error is going to be smaller than the length of the identification interval ($\sigma < 1.96P$) a planner using the maximum MSE measure of precision would allocate more resources to reducing the extent of partial identification than a planner measuring precision by the length of the confidence interval. The difference between the "marginal rates of substitution" produced by the two methods could be particularly large when considering large sample surveys and experiments in which the extent of partial identification could greatly exceed sampling error.

For the absolute loss function $L(t) = |t|$, the maximum mean absolute deviation (MAD)

$$\sup_{\theta_O, \theta_U} E^{X}\left|\hat{\theta}(X) - (\theta_O + \theta_U)\right| \tag{10}$$

of $\hat{\theta}^{*}(X) = X$ equals

$$M^{MAD}(\sigma, P) = 2\sigma\phi\left(\frac{P}{\sigma}\right) + 2P\left[\Phi\left(\frac{P}{\sigma}\right) - \Phi(0)\right]. \tag{11}$$

The marginal effects of changes in σ and P on the maximum MAD equal

$$\begin{aligned} M_{\sigma}^{MAD} &= 2\phi\left(\frac{P}{\sigma}\right), \\ \text{and } M_{P}^{MAD} &= 2\left[\Phi\left(\frac{P}{\sigma}\right) - \Phi\left(0\right)\right]. \end{aligned} \tag{12}$$

The ratio of these marginal effects equals

$$\frac{M_{\sigma}^{MAD}\left(\sigma, P\right)}{M_{P}^{MAD}\left(\sigma, P\right)} = \frac{\phi\left(\frac{P}{\sigma}\right)}{\Phi\left(\frac{P}{\sigma}\right) - \Phi\left(0\right)}. \tag{13}$$

This is a continuous decreasing function of $\frac{P}{\sigma}$, which goes to infinity as $\frac{P}{\sigma} \to 0$ and to zero as $\frac{P}{\sigma} \to \infty$.

Similarly to the maximum MSE, for sufficiently large values of $\frac{P}{\sigma}$ the maximum MAD measure of precision implies greater importance of reducing the scope of partial identification than does the confidence interval measure. For conventional 95% confidence intervals, calculations show that $\frac{M_{\sigma}^{MAD}}{M_{P}^{MAD}} < \frac{M_{\sigma}^{CI(.05)}}{M_{P}^{CI(.05)}}$ whenever $\sigma < 2.11P$. MAD and MSE measures yield similar conclusions about the relative benefits of reducing σ and P for small values of $\frac{P}{\sigma}$, since $\frac{\phi\left(\frac{P}{\sigma}\right)}{\Phi\left(\frac{P}{\sigma}\right)-\Phi(0)} \approx \frac{\sigma}{P}$ when $\frac{P}{\sigma} \to 0$.

4 Minimax Regret Approach

The third measure of precision - minimax regret - is motivated by directly considering the economic loss resulting from incorrect inference about θ when θ is the difference in average returns of two alternative policy decisions and the ultimate aim of inference about θ is to choose which policy to implement. For example, the policies may be two proposed cancer therapies, with θ measuring the average difference in the welfare of cancer patients from a target population net of the average difference between the costs of these two therapies.

Let $\theta = r_2 - r_1$, where r_1 is the average return from implementing the first policy and r_2 the average return from implementing the second policy. Then the economic loss from choosing the second policy when, in fact, $r_1 > r_2$ ($\theta < 0$) equals $r_1 - r_2 = -\theta$. The economic loss from choosing to implement the first policy when, in fact, $r_1 < r_2$ ($\theta > 0$) equals $r_2 - r_1 = \theta$. The method by which the decision maker chooses which policy to implement based on experimental data X could be summarized by a *statistical treatment rule* $\delta\left(X\right)$, which is a function mapping feasible realizations of $X \in \mathbb{R}$ into the $[0, 1]$ interval. $\delta\left(\bar{X}\right) = 0$ if the decision maker implements

the first policy when outcome $\bar{X}$ is observed, $\delta\left(\bar{X}\right) = 1$ if she implements the second policy. $\delta\left(\bar{X}\right)$ could takes values between 0 and 1 if the decision maker could implement either policy with some probability upon observing outcome $\bar{X}$.

The regret of statistical treatment rule δ is the average (over the probability distribution of outcome X) economic loss incurred by the decision maker using δ. It is a function of θ_O and θ_U, and in this problem equals

$$R\left(\delta,\left(\theta_O,\theta_U\right)\right) \equiv \begin{cases} \theta \cdot \left[1 - E_{\theta_O}\delta\left(X\right)\right] & \text{if } \theta > 0, \\ -\theta \cdot E_{\theta_O}\delta\left(X\right) & \text{if } \theta \leq 0, \end{cases} \tag{14}$$

where $E_{\theta_O}\delta\left(X\right)$ denotes the average value of $\delta\left(X\right)$ given that $X \sim \mathcal{N}\left(\theta_O, \sigma^2\right)$. When $\theta > 0$, the first policy is inferior and $\left[1 - E_{\theta_O}\delta\left(X\right)\right]$ is the probability with which the decision maker would mistakenly choose it based on observation of the random experimental outcome X. When $\theta < 0$, the second policy is inferior and $E_{\theta_O}\delta\left(X\right)$ is the probability of choosing it.

Minimizing maximum regret was a criterion suggested by Savage (1951) as a clarification of Wald's *minimax principle* (1950). For a more detailed discussion on applying minimax regret criterion to statistical treatment choice problems see Manski (2004, 2007a, Chapter 11).

To measure the precision of inference by the maximum regret of a statistical treatment rule we first ought to find statistical treatment rules that minimize maximum regret for given experimental parameters (σ, P). Proposition 3 derives such rules and their maximum regret.

Proposition 3 *a) For $\sigma > 2P \cdot \phi\left(0\right)$, the unique minimax regret statistical treatment rule is*

$$\delta_{M(\sigma,P)}(X) \equiv 1\left|X > 0\right|. \tag{15}$$

Its maximum regret equals

$$\sup_{\substack{\theta_O \in \mathbb{R}, \\ \theta_U \in [-P,P]}} R\left(\delta_{M(\sigma,P)}, \left(\theta_O, \theta_U\right)\right) = \max_{h>0}\left[h\Phi\left(\frac{P-h}{\sigma}\right)\right] > \frac{P}{2},$$

which is a strictly increasing function of σ for any given P.

b) For $\sigma \leq 2P \cdot \phi\left(0\right)$, statistical treatment rules

$$\delta_{M(\sigma,P)}\left(X\right) \equiv \begin{cases} 1\left|X > 0\right| & \text{if } \sigma = 2P \cdot \phi\left(0\right), \\ \Phi\left(\left[\left(2P \cdot \phi\left(0\right)\right)^2 - \sigma^2\right]^{-1/2} X\right) & \text{if } \sigma < 2P \cdot \phi\left(0\right), \end{cases} \tag{16}$$

minimize maximum regret, which equals $\frac{P}{2}$.

Two features of Proposition 3 are qualitatively similar to results obtained by Stoye (2007c), who studied minimax regret statistical treatment rules based on binary outcome data from an experiment with randomized treatment assignment in which the outcomes are missing with some probability.

First, when the extent of partial identification (in Stoye's problem, the maximum feasible proportion of missing outcomes) is below some threshold relative to the sampling error, the minimax regret statistical treatment rule is the same as it would be with point identification. In Proposition 3.2 (part a) the same result holds, the minimax regret statistical treatment rule (15) is identical for all values of $P \leq \frac{\sigma}{2\phi(0)}$, including the point identified case $P = 0$.

The second qualitative similarity is that maximum regret of the minimax regret statistical treatment rule becomes constant with respect to the sampling error once the sampling error falls below some threshold relative to the extent of partial identification. Thus, reducing the sampling error below that threshold (reducing σ in this chapter, increasing sample size in Stoye's) could not further reduce minimax regret.

Since this second result could appear counterintuitive, it deserves further explanation. Let

$$q(\delta, \theta_O) \equiv E_{\theta_O} \delta(X)$$

denote the average probability (with respect to the distribution of X) with which the decision maker using statistical treatment δ will choose the second policy. Then for a given value of P, $\frac{P}{2}$ is the lower bound on maximum regret attainable by any statistical treatment rule for any value of σ. This could be seen by considering maximum regret over the subset $\{\theta_O = 0, \theta_U \in [-P, P]\}$

$$\sup_{\substack{\theta_O \in \mathbb{R}, \\ \theta_U \in [-P,P]}} R(\delta, (\theta_O, \theta_U)) \geq \max_{\theta_U \in [-P,P]} R(\delta, (0, \theta_U)) = \max(P \cdot q(\delta, 0), P \cdot (1 - q(\delta, 0))) \geq \frac{P}{2}.$$

In order to attain this lower bound, the statistical treatment rule δ must satisfy $q(\delta, 0) = \frac{1}{2}$. For values $\bar{\theta}_O \neq 0$, however, there is a range of values of $q\left(\delta, \bar{\theta}_O\right)$ for which

$$\max_{\theta_U \in [-P,P]} R\left(\delta, \left(\bar{\theta}_O, \theta_U\right)\right) \leq \frac{P}{2}.$$

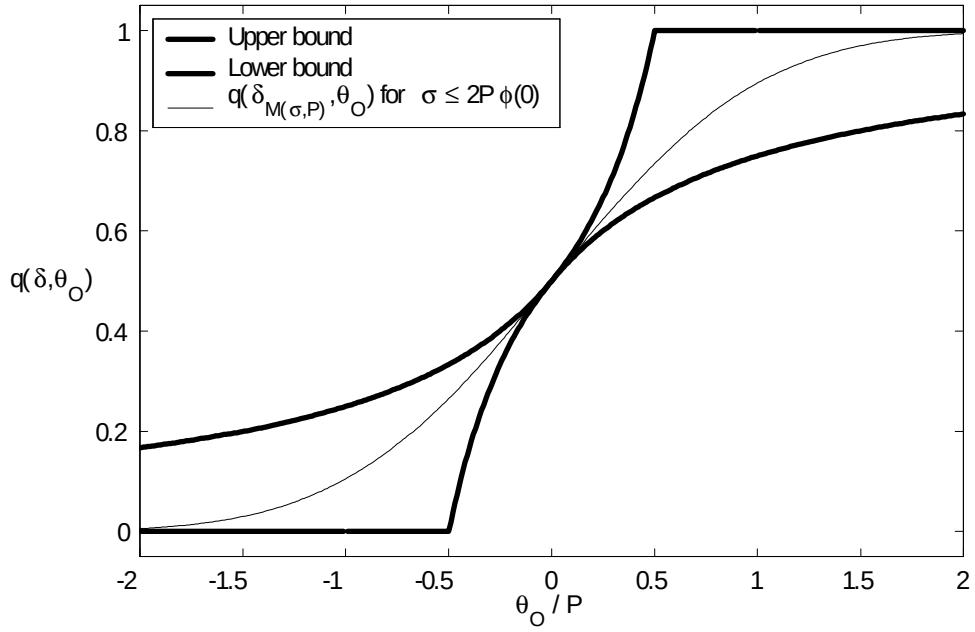

Figure 1: Bounds on $q(\delta, \theta_O)$ that guarantee attaining the lower bound on maximum regret $(P/2)$.

This range is given by the inequalities

$$\begin{array}{ll} q(\delta, \theta_O) \geq 1 - \frac{P}{2(P+\theta_O)} & \text{for } \theta_O \geq -\frac{P}{2}, \\ q(\delta, \theta_O) \leq \frac{P}{2(P-\theta_O)} & \text{for } \theta_O \leq \frac{P}{2}. \end{array} \tag{17}$$

As shown in Proposition 3, for $\sigma \leq 2P \cdot \phi(0)$, it is possible to construct statistical treatment rules that satisfy these inequalities for all $\theta_O \in \mathbb{R}$. Figure 1 displays in bold lines the bounds (17) and shows that the function $q\left(\delta_{M(\sigma,P)}, \theta_O\right) = \Phi\left(\frac{\theta_O}{2P \cdot \phi(0)}\right)$, which is identical for all minimax regret rules defined by (16), fits within these bounds.

Statistical treatment rules derived in part b of Proposition 3 may not be the only ones that minimize maximum regret, but deriving one class of minimax regret rules is sufficient to make conclusions about the minimum value of maximum regret, and thus about the precision of inference from the data for treatment choice.

If the planner chooses minimax regret to measure inferential precision, then the precision of inference generated by an experiment with parameters (σ, P) is

$$M^{MMR}(\sigma, P) = \begin{cases} \max\limits_{h>0} \left\{ h\Phi\left(\frac{P-h}{\sigma}\right) \right\} & \text{if } \sigma > 2P \cdot \phi(0), \\ \frac{P}{2} & \text{if } \sigma \leq 2P \cdot \phi(0), \end{cases}$$

This measure of precision could yield drastically different conclusions about the relative benefits of reducing the extent of partial identification and reducing sampling error than confidence interval, maximum MSE, and maximum MAD approaches, since for $\frac{\sigma}{P} \leq 2\phi(0) \approx 0.8$

$$\begin{aligned} M_{\sigma}^{MMR} &= 0, \\ M_{P}^{MMR} &= \frac{1}{2}, \end{aligned}$$

implying that reducing the extent of partial identification is not only relatively more important than reducing sampling error, it is the only way to reduce minimax regret and improve the inferential precision of experimental or survey data for treatment choice.

5 Conclusion

In this paper, I considered alternative measures of inferential precision for partially identified parameteres in addition to the length of 95% confidence interval, which is the primary measure previously considered by other researchers. These measures yield qualitatively different conclusions about the relative merits of reducing sampling error and reducing the extent of partial identification in the data. Both the maximum mean squared error, the maximum mean absolute deviation, and minimax regret (applicable when inference is carried out on the average treatment effect with the goal of choosing the best treatment) emphasize greater value of reducing the extent of partial identification compared to the confidence interval measure if the sampling error is relatively small compared to the width of the identification interval.

The statistical problem with a normal sampling distribution considered in the paper is simple in comparison to many practical problems. However, it is sufficiently rich to capture some of the main features of partial identification problems and to concisely illustrate how choosing different criteria for measuring the precision of inference qualitatively impacts the conclusions about the relative value of reducing the extent of partial identification and reducing sampling error. The results could serve both as a rough practical approximation for partial identification problems with similar structure and as a useful indicator of potential findings for future research that considers more complex practical partial identification problems.

6 Appendix: Proofs

The proof of Proposition 2 will use the following theorem (e.g., Berger 1985, p. 350), in which $r(\pi_n) \equiv \int R(\delta_n, \theta)\, \partial\pi_n(\theta)$ denotes the Bayes risk of a Bayes decision rule $\delta_n = \arg\min_{\delta} \int R(\delta, \theta)\, \partial\pi_n(\theta)$ under the prior distribution π_n.

Theorem 4 *If $\{\pi_n\}$ is a sequence of proper prior distributions on the parameter space Θ and for all $\theta \in \Theta$*

$$R(\delta^*, \theta) \leq \lim_{n\to\infty} r(\pi_n) < \infty$$

then the decision rule δ^* is minimax.

Proof of Proposition 2

To show that $\hat{\theta}^*(X) = X$ is a minimax expected loss estimator of $\theta = \theta_O + \theta_U$, I will consider a sequence $\{\pi_k\}$ of proper prior distributions on the parameters (θ_O, θ_U), such that θ_O and θ_U are independent under each π_k, θ_O is distributed $\mathcal{N}(0, k^2)$, while θ_U has an equal probability of being equal to $\pm P$:

$$\begin{aligned} \pi_k(\theta_O) &= \mathcal{N}(0, k^2), \\ \pi_k(\theta_U) &= .5 \cdot I[|\theta_U| = P], \\ \theta_O &\perp \theta_U. \end{aligned}$$

Since $X \sim \mathcal{N}(\theta_O, \sigma^2)$, the posterior distributions of θ_O and θ_U conditional on observing X are

$$\begin{aligned} \pi_k(\theta_O|X) &= \mathcal{N}(c_k X, c_k \sigma^2), \\ \text{and } \pi_k(\theta_U|X) &= .5 \cdot I[|\theta_U| = P], \end{aligned}$$

where $c_k = \frac{k^2}{k^2+\sigma^2}$, $\lim_{k\to\infty} c_k = 1$. Since the loss function L is convex and symmetric, $\hat{\theta}_k(X) = c_k X$ is the Bayes estimator of θ that minimizes the posterior risk

$$\int L\left(\hat{\theta}_k(X) - (\theta_O + \theta_U)\right) d\pi_k(\theta_O, \theta_U|X).$$

Given the posterior distribution $\pi_k(\theta_O|X)$, the random variable $y = c_k X - \theta_O$ has a $\mathcal{N}\left(0, c_k\sigma^2\right)$ distribution. The posterior risk of $\hat{\theta}_k(X)$, then, equals

$$\begin{aligned}
& \int L\left(c_k X - (\theta_O + \theta_U)\right) d\pi_k\left(\theta_O, \theta_U|X\right) = \\
= & \int \left[\frac{1}{2}L\left(c_k X - \theta_O - P\right) + \frac{1}{2}L\left(c_k X - \theta_O + P\right)\right] d\pi_k\left(\theta_O|X\right) = \\
= & \int \left[\frac{1}{2}L\left(y - P\right) + \frac{1}{2}L\left(y + P\right)\right] d\pi_k\left(y|X\right) = \int L\left(y + P\right) d\pi_k\left(y|X\right) = \\
= & \int_{-\infty}^{+\infty} L\left(y + P\right) \frac{1}{\sqrt{c_k}\sigma}\phi\left(\frac{y}{\sqrt{c_k}\sigma}\right) dy = \int_{-\infty}^{+\infty} L\left(z\right) \frac{1}{\sqrt{c_k}\sigma}\phi\left(\frac{z - P}{\sqrt{c_k}\sigma}\right) dz
\end{aligned}$$

The second equality holds because L and $\pi_k(y|X)$ are symmetric. Condition 1(e) guarantees that this and other improper integrals in this proof are well defined. The posterior risk is constant for all possible values of X, thus the Bayes risk with prior π_k is also equal

$$r\left(\pi_k\right) = \int_{-\infty}^{+\infty} L\left(z\right) \frac{1}{\sqrt{c_k}\sigma}\phi\left(\frac{z - P}{\sqrt{c_k}\sigma}\right) dz.$$

The functions $L(z)\frac{1}{\sqrt{c_k}\sigma}\phi\left(\frac{z-P}{\sqrt{c_k}\sigma}\right)$ converge pointwise to $L(z)\frac{1}{\sigma}\phi\left(\frac{z-P}{\sigma}\right)$ as $k \to \infty$. Due to Condition 1(e), Lebesgue dominated convergence theorem applies and

$$\lim_{k\to\infty} r\left(\pi_k\right) = \int_{-\infty}^{+\infty} L\left(z\right) \frac{1}{\sigma}\phi\left(\frac{z - P}{\sigma}\right) dz.$$

The risk of $\hat{\theta}^*$ equals

$$\begin{aligned}
R\left(\hat{\theta}^*, (\theta_O, \theta_U)\right) &= E^X L\left(X - (\theta_O + \theta_U)\right) \\
&= \int_{-\infty}^{+\infty} L\left(x - \theta_O - \theta_U\right) \frac{1}{\sigma}\phi\left(\frac{x - \theta_O}{\sigma}\right) dx \\
&= \int_{-\infty}^{+\infty} L\left(z\right) \frac{1}{\sigma}\phi\left(\frac{z + \theta_U}{\sigma}\right) dz.
\end{aligned}$$

It is maximized for $|\theta_U| = P$, thus

$$R\left(\hat{\theta}^*, (\theta_O, \theta_U)\right) \leq \int_{-\infty}^{+\infty} L\left(z\right) \frac{1}{\sigma}\phi\left(\frac{z - P}{\sigma}\right) dz = \lim_{k\to\infty} r\left(\pi_k\right),$$

thus Theorem 4 applies and $\hat{\theta}^*(X) = X$ is a minimax estimator of $\theta = \theta_O + \theta_U$ under loss function L with maximum expected loss equal to $\int_{-\infty}^{+\infty} L(z)\frac{1}{\sigma}\phi\left(\frac{z-P}{\sigma}\right) dz$. □

Proof of Equation 8

$$M^{MSE}(\sigma, P) = \int_{-\infty}^{+\infty} t^2 \frac{1}{\sigma}\phi\left(\frac{t-P}{\sigma}\right) dt = \int_{-\infty}^{+\infty} (s\sigma + P)^2 \phi(s)\, ds = \sigma^2 + P^2,$$

since $\int s\phi(s)\, ds = 0$ and $\int s^2\phi(s)\, ds = 1$ are the mean and variance of the standard normal distribution.

Proof of Equation 11 Since $\phi'(z) = -z\phi(z)$, $\int_a^b z\phi(z)\, \partial z = \phi(a) - \phi(b)$.

$$\begin{aligned}
M^{MAD}(\sigma, P) &= \int_{-\infty}^{+\infty} |t| \frac{1}{\sigma}\phi\left(\frac{t-P}{\sigma}\right) dt = \int_{-\infty}^{+\infty} |s\sigma + P| \phi(s)\, ds = \\
&= \int_{-\frac{P}{\sigma}}^{+\infty} (s\sigma + P)\phi(s)\, ds - \int_{-\infty}^{-\frac{P}{\sigma}} (s\sigma + P)\phi(s)\, ds = \\
&= \sigma\phi\left(\frac{P}{\sigma}\right) + P\Phi\left(\frac{P}{\sigma}\right) + \sigma\phi\left(\frac{P}{\sigma}\right) - P\Phi\left(-\frac{P}{\sigma}\right) = \\
&= 2\sigma\phi\left(\frac{P}{\sigma}\right) + P\left[\Phi\left(\frac{P}{\sigma}\right) - \Phi\left(-\frac{P}{\sigma}\right)\right] = \\
&= 2\sigma\phi\left(\frac{P}{\sigma}\right) + 2P\left[\Phi\left(\frac{P}{\sigma}\right) - \Phi(0)\right].
\end{aligned}$$

Proof of Proposition 3(a)

Let $(\theta_O, \theta_U) \in \Theta, \Theta = \mathbb{R} \times [-P, P]$. The proof of part a relies on a well known result (e.g., Berger 1985, p. 350) that if π^* is a proper prior distribution on Θ, the decision rule δ^* is Bayes with respect to π^*, and for all $(\theta_O, \theta_U) \in \Theta$

$$R(\delta^*, (\theta_O, \theta_U)) \leq \int R(\delta, (\theta_O, \theta_U))\, \partial\pi^*(\theta_O, \theta_U),$$

then the decision rule δ^* is minimax. This result applies as well when R denotes regret, then δ^* is a minimax-regret rule.

Decision rule

$$\delta^*(X) \equiv 1\,|X > 0|$$

is Bayes with respect to any symmetric two-point prior distribution π with $\pi(\theta_O^*, \theta_U^*) = .5$ and $\pi(-\theta_O^*, -\theta_U^*) = .5$, if $\theta_O^* > 0$ and $\theta_O^* + \theta_U^* > 0$.

When $\theta > 0$, for a given value of θ_O, regret $R(\delta^*, (\theta_O, \theta_U)) = (\theta_O + \theta_U) \cdot [1 - E_{\theta_O}\delta^*(X)]$ is largest at $\theta_U = P$, since the first term is increasing in θ_U and the second term is positive and doesn't depend on θ_U. Since $E_{\theta_O}\delta^*(X) = 1 - \Phi\left(-\frac{\theta_O}{\sigma}\right)$, maximum regret of δ^* over $\theta > 0$ then

equals (with the substitution $h = \theta_O + P$)

$$\max_{\substack{\theta_O,\theta_U\in\Theta,\\ \theta_O+\theta_U>0}} R\left(\delta^*,(\theta_O,\theta_U)\right) = \max_{\theta_O>-P}\left[(\theta_O+P)\cdot\Phi\left(-\frac{\theta_O}{\sigma}\right)\right] = \max_{h>0}\left[h\Phi\left(\frac{P-h}{\sigma}\right)\right].$$

The maximum is attained at

$$\theta_O^* = \arg\max_{h>0}\left[h\Phi\left(\frac{P-h}{\sigma}\right)\right] - P.$$

When $\theta < 0$, regret $R\left(\delta^*,(\theta_O,\theta_U)\right) = -\left(\theta_O+\theta_U\right)\cdot E_{\theta_O}\delta^*\left(X\right)$ is maximized at $\theta_U = -P$ for a given θ_O, and equals (with the substitution $h = -\left(\theta_O - P\right)$)

$$\max_{\substack{\theta_O,\theta_U\in\Theta,\\ \theta_O+\theta_U<0}} R\left(\delta^*,(\theta_O,\theta_U)\right) = \max_{\theta_O<P}\left[-\left(\theta_O-P\right)\cdot\Phi\left(\frac{\theta_O}{\sigma}\right)\right] = \max_{h>0}\left[h\Phi\left(\frac{P-h}{\sigma}\right)\right].$$

Let's differentiate $h\Phi\left(\frac{P-h}{\sigma}\right)$ with respect to h

$$\frac{\partial}{\partial h}\left[h\Phi\left(\frac{P-h}{\sigma}\right)\right] = \Phi\left(\frac{P-h}{\sigma}\right) - \frac{h}{\sigma}\phi\left(\frac{P-h}{\sigma}\right) = \Phi\left(\frac{P-h}{\sigma}\right)\left[1 - \frac{h}{\sigma}\frac{\phi\left(\frac{P-h}{\sigma}\right)}{\Phi\left(\frac{P-h}{\sigma}\right)}\right].$$

At $h = 0$, $\frac{\partial}{\partial h}\left[h\Phi\left(\frac{P-h}{\sigma}\right)\right] = \Phi\left(\frac{P}{\sigma}\right) > 0$. The function $\frac{\phi(y)}{\Phi(y)} > 0$ is strictly decreasing with $\lim_{y\to-\infty}\frac{\phi(y)}{\Phi(y)} = +\infty$, thus $\frac{h}{\sigma}\frac{\phi\left(\frac{P-h}{\sigma}\right)}{\Phi\left(\frac{P-h}{\sigma}\right)}$ is strictly increasing in h over $h > 0$ and $\lim_{h\to\infty}\left[1 - \frac{h}{\sigma}\frac{\phi\left(\frac{P-h}{\sigma}\right)}{\Phi\left(\frac{P-h}{\sigma}\right)}\right] = -\infty$. It follows that $\frac{\partial}{\partial h}\left[h\Phi\left(\frac{P-h}{\sigma}\right)\right]$ changes sign once over $h > 0$ from positive to negative at h^* given by $\frac{h^*}{\sigma}\frac{\phi\left(\frac{P-h^*}{\sigma}\right)}{\Phi\left(\frac{P-h^*}{\sigma}\right)} = 1$, thus $h\Phi\left(\frac{P-h}{\sigma}\right)$ attains its maximum over $h > 0$ at h^*.

When $\sigma > 2P\cdot\phi\left(0\right)$, $h^* > P$. To see this, evaluate $\frac{\partial}{\partial h}\left[h\Phi\left(\frac{P-h}{\sigma}\right)\right]$ at $h = P$:

$$\left.\frac{\partial}{\partial h}\left[h\Phi\left(\frac{P-h}{\sigma}\right)\right]\right|_{h=P} = \Phi\left(0\right) - \frac{P}{\sigma}\phi\left(0\right) = \frac{1}{2} - \frac{P\cdot\phi\left(0\right)}{\sigma} > 0.$$

Thus $\frac{\partial}{\partial h}\left[h\Phi\left(\frac{P-h}{\sigma}\right)\right]$ changes sign at $h^* > P$. Since $h^* > P$, maximum regret is attained at (θ_O^*, P) and $(-\theta_O^*, -P)$, where $\theta_O^* = h^* - P > 0$.

Maximum regret of δ^* exceeds $\frac{P}{2}$ because $\frac{\partial}{\partial h}\left[h\Phi\left(\frac{P-h}{\sigma}\right)\right] > 0$ for $P \le h < h^*$, therefore

$$\max_{h>0}\left[h\Phi\left(\frac{P-h}{\sigma}\right)\right] = h^*\Phi\left(\frac{P-h^*}{\sigma}\right) > P\Phi\left(\frac{P-P}{\sigma}\right) = \frac{P}{2}.$$

Since δ^* is a Bayes rule with respect to prior π^* with $\pi^*(\theta_O^*, P) = .5$ and $\pi^*(-\theta_O^*, -P) = .5$ and

$$\int R(\delta^*, (\theta_O, \theta_U))\, \partial\pi^*(\theta_O, \theta_U) = R(\delta^*, (\theta_O^*, P)) = \max_{\substack{\theta_O \in \mathbb{R}, \\ \theta_U \in [-P,P]}} R(\delta^*, (\theta_O, \theta_U)),$$

δ^* minimizes maximum regret. Furthermore, since δ^* is a unique Bayes rule up to randomization at $X = 0$, which does not affect $R(\delta, (\theta_O, \theta_U))$ for any values of (θ_O, θ_U), it is admissible.

To verify that minimax regret $\max_{h>0} \left[h\Phi\left(\frac{P-h}{\sigma}\right)\right]$ is a decreasing function of σ for a given P and $\sigma > 2P \cdot \phi(0)$, observe that since $h^* > P$,

$$\max_{h>0} \left[h\Phi\left(\frac{P-h}{\sigma}\right)\right] = \max_{h>P} \left[h\Phi\left(\frac{P-h}{\sigma}\right)\right].$$

For any $h > P$, $h\Phi\left(\frac{P-h}{\sigma}\right)$ is strictly decreasing in σ and has a unique maximum over $h > P$ for a given σ, thus $\max_{h>P} \left[h\Phi\left(\frac{P-h}{\sigma}\right)\right]$ is strictly decreasing in σ. □

Proof of Proposition 3(b)

First, I will show that any rule δ for which $q(\delta, \theta_O) = E_{\theta_O}\delta(X)$ lies within the bounds (17) has maximum regret of $\frac{P}{2}$. The lower bound

$$q(\delta, \theta_O) \geq 1 - \tfrac{P}{2(P+\theta_O)} \quad \text{for } \theta_O \geq -\tfrac{P}{2},$$

guarantees that $R(\delta, (\theta_O, \theta_U)) \leq \frac{P}{2}$ over $\theta > 0$. Since $R(\delta, (\theta_O, \theta_U))$ is increasing in θ_U when $\theta > 0$,

$$\max_{\substack{\theta_O, \theta_U \in \Theta, \\ \theta_O + \theta_U > 0}} R(\delta, (\theta_O, \theta_U)) = \max_{\theta_O > -P} R(\delta, (\theta_O, P)) = \max_{\theta_O > -P} \left[(\theta_O + P) \cdot [1 - q(\delta, \theta_O)]\right].$$

For $\theta_O \geq -\frac{P}{2}$, if $q(\delta, \theta_O) \geq 1 - \frac{P}{2(P+\theta_O)} \geq 0$, then

$$(\theta_O + P) \cdot [1 - q(\delta, \theta_O)] \leq (\theta_O + P) \cdot \frac{P}{2(P + \theta_O)} = \frac{P}{2}.$$

For $\theta_O \in \left[-P, -\frac{P}{2}\right]$,

$$(\theta_O + P) \cdot [1 - q(\delta, \theta_O)] \leq \theta_O + P \leq \frac{P}{2}.$$

The proof for the upper bound, which ensures that $R(\delta,(\theta_O,\theta_U)) \leq \frac{P}{2}$ for $\theta < 0$, is analogous. Both the lower and the upper bound are equal to $\frac{1}{2}$ at $\theta_O = 0$, thus $q(\delta,0) = \frac{1}{2}$ and

$$\max_{\theta_O,\theta_U\in\Theta} R(\delta,(\theta_O,\theta_U)) \geq \max_{\theta_U\in[-P,P]} R(\delta,(0,\theta_U)) = \frac{P}{2}.$$

Thus the maximum regret of δ equals $\frac{P}{2}$ if $q(\delta,\theta_O)$ satisfies inequalities (17).

Second, I will show that the function

$$q^*(\theta_O) \equiv \Phi\left(\frac{\theta_O}{2P\cdot\phi(0)}\right)$$

lies within the bounds (17). The proof will verify this for $\theta_O \geq 0$, it is analogous for $\theta_O < 0$.

When $\theta_O = 0$, $q^*(0) = \Phi(0) = \frac{1}{2}$, which coincides with both bounds. $q^*(\theta_O)$ satisfies the upper bound because for $\theta_O \in \left[0, \frac{P}{2}\right]$

$$\Phi\left(\frac{\theta_O}{2P\cdot\phi(0)}\right) \leq \frac{1}{2} + \frac{\theta_O}{2P\cdot\phi(0)}\cdot\phi(0) = \frac{P+\theta_O}{2P} \leq \frac{P}{2(P-\theta_O)}.$$

The first inequality follows from using $\phi(0)$ as an upper bound on the derivative of Φ. The second one follows from $(P+\theta_O)(P-\theta_O) = P^2 - \theta_O^2 \leq P^2$.

The proof that $q^*(\theta_O) \geq 1 - \frac{P}{2(P+\theta_O)}$ for all $\theta_O \geq 0$ is split into two cases, $\theta_O \in [0,P]$ and $\theta_O \geq P$.

Case 1. For $\theta_O \in [0,P]$, I will prove that $q^*(\theta_O)$ increases faster than the lower bound, which guarantees that $q^*(\theta_O) \geq 1 - \frac{P}{2(P+\theta_O)}$, since both are equal at $\theta_O = 0$. It will be sufficient to consider $P = 1$, to simplify notation, and thus $\theta_O \in [0,1]$. For $P = 1$, $q^*(\theta_O) = \Phi\left(\frac{\theta_O}{2\phi(0)}\right)$, $\phi(y) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}y^2\right)$, with $2\phi(0) = \sqrt{\frac{2}{\pi}}$, thus

$$\frac{\partial}{\partial\theta_O}q^*(\theta_O) = \frac{1}{2\phi(0)}\phi\left(\frac{\theta_O}{2\phi(0)}\right) = \sqrt{\frac{\pi}{2}}\cdot\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}\left(\sqrt{\frac{\pi}{2}}\theta_O\right)^2\right) = \frac{1}{2}\exp\left(-\frac{\pi}{4}\theta_O^2\right).$$

Since the function $e(y)$ is convex with $e(0) = 1$ and $e(1) < 3$, $e(y) \leq 1 + 2y$ for $y \in [0,1]$, therefore $e(y) \geq \frac{1}{1-2y}$ for $y \in [-1,0]$. Since $\frac{\pi}{4} < 1$ and $\theta_O^2 < 1$,

$$\frac{1}{2}\exp\left(-\frac{\pi}{4}\theta_O^2\right) \geq \frac{1}{2}\cdot\frac{1}{1+\frac{\pi}{2}\theta_O^2} = \frac{1}{2+\pi\theta_O^2}.$$

For $\theta \in [0,1]$, $\pi\theta_O < 4$, thus $\pi\theta_O^2 \leq 4\theta_O$ and $2 + \pi\theta_O^2 \leq 2 + 4\theta_O + 2\theta_O^2 = 2(1+\theta_O)^2$, therefore $\frac{1}{2+\pi\theta_O^2} \geq \frac{1}{2(1+\theta_O)^2}$, and

$$\frac{\partial}{\partial\theta_O} q^*(\theta_O) \geq \frac{1}{2+\pi\theta_O^2} \geq \frac{1}{2(1+\theta_O)^2} = \frac{\partial}{\partial\theta_O}\left[1 - \frac{1}{2(1+\theta_O)}\right].$$

Case 2. For $\theta_O \geq P$, I will also use $P = 1$ to simpify notation, so the aim is to prove that $q^*(\theta_O) = \Phi\left(\frac{\theta_O}{2\phi(0)}\right) \geq 1 - \frac{1}{2(1+\theta_O)}$. For $y > 0$, $1 - \Phi(y) < \frac{\phi(y)}{y}$, which implies

$$q^*(\theta_O) = \Phi\left(\frac{\theta_O}{2\phi(0)}\right) > 1 - \frac{2\phi(0)}{\theta_O}\phi\left(\frac{\theta_O}{2\phi(0)}\right) = 1 - \frac{1}{\pi\theta_O}\exp\left(-\frac{\pi}{4}\theta_O^2\right).$$

For $y \geq 0$, $e(y) \geq 1 + y$, thus for $y \leq 0$, $e(y) \leq \frac{1}{1-y}$. Using this inequality yields

$$q^*(\theta_O) > 1 - \frac{1}{\pi\theta_O} \cdot \frac{1}{1+\frac{\pi}{4}\theta_O^2} = 1 - \frac{1}{\pi\theta_O + \frac{\pi^2}{4}\theta_O^3} > 1 - \frac{1}{2(1+\theta_O)}$$

where the last inequality follows from observation that $\frac{\pi^2}{4} > 2$, and for $\theta_O \geq 1$, $\pi\theta_O > 2$ and $\theta_O^3 \geq \theta_O$.

Since $q^*(\theta_O)$ satisfies the inequalities (17), any statistical treatment rule with $q(\delta, \theta_O) = q^*(\theta_O)$ has maximum regret of $\frac{P}{2}$. It remains to show that this holds for statistical treatment rules (16) defined in part b of Proposition 3.

For $\sigma = 2P \cdot \phi(0)$, $\delta_{M(\sigma,P)}(X) = 1|X > 0|$, thus $q\left(\delta_{M(\sigma,P)}, \theta_O\right) = \Phi\left(\frac{\theta_O}{2P\cdot\phi(0)}\right) = q^*(\theta_O)$ and the rule minimizes maximum regret, which equals $\frac{P}{2}$.

For $\sigma < 2P \cdot \phi(0)$, it is simplest to derive $\delta_{M(\sigma,P)}(X)$ using the following construction[1]. Let $\sigma_0 = 2P \cdot \phi(0)$. Define an auxiliary random variable

$$Y \sim \mathcal{N}\left(0, \sigma_0^2 - \sigma^2\right),$$

independent of the observed outcome $X \sim \mathcal{N}\left(\theta_O, \sigma^2\right)$. Then $X + Y \sim \mathcal{N}\left(\theta_O, \sigma_0^2\right)$. Define the statistical treatment rule $\delta^*_{M(\sigma,P)}(X, Y)$ as

$$\delta^*_{M(\sigma,P)}(X, Y) \equiv 1|X + Y > 0|,$$

[1]This proof technique is similar to Schlag's (2007) *binomial average*, in that both algebraically simplify the problem by adding some noise to the observed outcomes.

then clearly

$$q\left(\delta^*_{(\sigma,P)},\theta_O\right)=\Phi\left(\frac{\theta_O}{2P\cdot\phi\left(0\right)}\right)=q^*\left(\theta_O\right).$$

Integrating $\delta^*_{M(\sigma,P)}\left(X,Y\right)$ with respect to the distribution of Y yields

$$\delta_{M(\sigma,P)}\left(X\right)\equiv E\left(1\left|X+Y>0\right|\right)=1-\Phi\left(-\left(\sigma_0^2-\sigma^2\right)^{-1/2}X\right)=\Phi\left(\left(\sigma_0^2-\sigma^2\right)^{-1/2}X\right),$$

which thus satisfies $q\left(\delta_{M(\sigma,P)},\theta_O\right)=q^*\left(\theta_O\right)$ by construction and.minimizes maximum regret, which equals $\frac{P}{2}$. $\square$

References

[1] Berger, J.O. 1985. *Statistical Decision Theory and Bayesian Analysis (2nd Edition).* New York: Springer Verlag.

[2] Cochran, W., F. Mosteller, and J. Tukey. 1954. *Statistical Problems of the Kinsey Report on Sexual Behavior in the Human Male.* Washington, DC: American Statistical Association.

[3] Hirano, K., and J.R. Porter. 2008. Asymptotics for Statistical Treatment Rules. *Econometrica* (forthcoming).

[4] Horowitz, J.L., and C.F. Manski. 1998. Censoring of Outcomes and Regressors Due to Survey Nonresponse: Identification and Estimation Using Weights and Imputations. *Journal of Econometrics* 84: 37–58.

[5] Manski, C.F. 1995. *Identification Problems in the Social Sciences.* Cambridge, Massachusetts: Harvard University Press.

[6] Manski, C.F. 2004. Statistical Treatment Rules for Heterogeneous Populations. *Econometrica* 72: 1221–1246.

[7] Manski, C.F. 2005. *Social Choice with Partial Knowledge of Treatment Response.* Princeton: Princeton University Press.

[8] Manski, C.F. 2007a. *Identification for Prediction and Decision.* Cambridge, Massachusetts: Harvard University Press.

[9] Manski, C.F. 2007b. Minimax-Regret Treatment Choice with Missing Outcome Data. *Journal of Econometrics* 139: 105–115.

[10] Manski, C.F. 2008a. Adaptive Partial Drug Approval. Working paper, Northwestern University.

[11] Manski, C.F. 2008b. Adaptive Partial Policy Innovation: Coping with Ambiguity through Diversification. Working paper, Northwestern University.

[12] Manski, C.F., and A. Tetenov. 2007. Admissible Treatment Rules for a Risk-Averse Planner with Experimental Data on an Innovation. *Journal of Statistical Planning and Inference* 137: 1998–2010.

[13] McFadden, D. 2006. How Consumers Respond to Incentives. *Jean-Jacques Laffont Lecture.* Toulouse, France. http://idei.fr/doc/conf/annual/paper_2006.pdf.

[14] Savage, L.J. 1951. The Theory of Statistical Decision. *Journal of the American Statistical Association* 46: 55–67.

[15] Schlag, K.H. 2007. Eleven - Designing Randomized Experiments under Minimax Regret. Working paper, European University Institute.

[16] Stoye, J. 2007a. Minimax Regret Treatment Choice with Incomplete Data and Many Treatments. *Econometric Theory* 23: 190–199.

[17] Stoye, J. 2007b. Minimax Regret Treatment Choice with Finite Samples. Working paper, New York University.

[18] Stoye, J. 2007c. Minimax Regret Treatment Choice with Finite Samples and Missing Outcome Data. In *Proceedings of the Fifth International Symposium on Imprecise Probability: Theories and Applications*, ed. G. de Cooman, J. Veinarová, and M. Zaffalon. Prague, Czech Republic.

[19] Wald, A. 1950. *Statistical Decision Functions.* New York: Wiley.