Measuring Precision of Statistical Inference on Partially Identified Parameters*

Aleksey Tetenov†

January 31, 2012

Abstract

Planners of surveys and experiments that partially identify parameters of interest face trade offs between using limited resources to reduce sampling error and using them to reduce the extent of partial identification. I evaluate these trade offs in a simple statistical problem with normally distributed sample data and interval partial identification using different frequentist measures of inference precision (length of confidence intervals, minimax mean sqaured error and mean absolute deviation, minimax regret for treatment choice) and analogous Bayes measures with a flat prior. The relative value of collecting data with better identification properties (e.g., increasing response rates in surveys) depends crucially on the choice of the measure of precision. When the extent of partial identification is significant in comparison to sampling error, the length of confidence intervals, which has been used most often, assigns the lowest value to improving identification among the measures considered.

Keywords: statistical treatment choice; survey planning; nonresponse; mean squared error; mean absolute deviation; minimax regret

JEL classification: C21; C44; C83

*I am grateful to Chuck Manski, Elie Tamer, Keisuke Hirano and Luca Anderlini for their helpful comments. I have also benefitted from the opportunity to present this work at Collegio Carlo Alberto, LAMES 2008, ICEEE 2009, UCL, Cambridge and Toulouse.

†Collegio Carlo Alberto, Moncalieri (Torino), Italy. aleksey.tetenov@carloalberto.org

1 Introduction

Many types of statistical data only partially identify parameters of interest as simple as population means, meaning that they 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. To inform these choices, I attempt here to evaluate 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 selective non-response to the study's questions, they 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 this 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 non-response 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 for the identified interval. 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).

Length of a confidence interval for the identified interval is not the only measure of precision. In this paper I show that other reasonable measures yield qualitatively different conclusions about the relative merits of reducing sampling error and reducing the extent of partial identification. First, I consider the minimax mean squared error (MSE) and minimax mean absolute deviation (MAD) of point estimates around the true value of the parameter, which have been widely used to measure the precision of estimators for point identified parameters. In addition to the minimax measures, I also consider the average risk for these loss functions with a flat prior.

Another measure considered in this paper is the minimax regret of statistical treatment rules for choosing between two alternative policies (or treatments) when the parameter of interest is the difference in average returns of the two treatments. Regret 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 2009; Stoye 2009; Schlag 2007; Manski and Tetenov 2007) and when it is partially identified (Manski 2007a, 2007b, 2008, 2009; Stoye 2007, 2012). I also consider the average welfare loss with a flat prior.

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 $2P$. When P/σ is relatively large (i.e., partial identification is a significant issue), all of the considered measures of precision put a higher value on reducing the extent of partial identification than the confidence interval measure.

The problem is deliberately simplified to demonstrate in an analytically tractable setting the qualitative differences between the conclusions about the relative benefits of reducing sampling error vs. narrowing the identified interval based on alternative measures of precision. I do not develop here a formal asymptotic argument extending the solution to more realistic data generating processes. Song (2010) formally establishes that midpoint of an estimated identified interval is asymptotically minimax for absolute and square loss functions by considering sequences of problems with $P \to 0$, $\sigma \to 0$ and P/σ converging to a constant. An extension of Song's analysis may also be applicable to the problem considered here.

The paper proceeds as follows. Section 2 describes the statistical problem and considers the length of confidence intervals as a measure of precision. In section 3 I derive minimax estimators of θ under square loss, absolute loss and regret loss for treatment choice, and consider their minimax losses as measures of precision. Section 4 considers instead the average risk with an improper flat prior on (σ, P). Section 5 offers a numerical illustration applying the results to survey non-response. Section 6 concludes and the appendix collects proofs of the propositions.

2 Statistical Setting

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 $X \sim \mathcal{N}\left(\theta_O, \sigma^2\right)$ with standard error σ. θ_U is a partially identified (unobservable) component, which is only known to lie in a bounded interval $\theta_U \in [-P, P]$ of length $2P$ (setting P to be the half-length simplifies notation throughout the paper). The restriction that θ_U lies in a symmetric interval around zero is without loss of generality.

Survey non-response is a leading example to keep in mind. Suppose we're interested in the population mean of Y and survey N individuals. Let r be the proportion of respondents (denote them by $D = 1$), who may have a different distribution of Y than non-respondents, then $\theta = EY$ is identified up an interval

$$[rE(Y|D=1) + (1-r)Y_L, rE(Y|D=1) + (1-r)Y_H], \tag{1}$$

where Y_L and Y_H are the bounds on feasible values of Y (Manski 1995, 2007a). The length of the identified interval is $2P = (1-r)(Y_H - Y_L)$. It could be reduced at a cost (for example, by driving up in person to a household that does not respond to phone calls or by offering stronger incentives to respondents). The midpoint $\theta_O = rE(Y|D=1) + (1-r)(Y_H - Y_L)/2$ of the identified interval could be estimated with standard error σ proportional to $N^{-1/2}$, which could be reduced by sampling more households. The statistical setup with normally distributed X is not an exact representation of this problem, but is analytically tractable and more informative than solving the "correct" problem computationally

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. 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

$$\max_{\sigma,P} \left[b\left(M\left(\sigma,P\right)\right) - c\left(\sigma,P\right)\right]. \tag{2}$$

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

$$\frac{M_\sigma\left(\sigma^*,P^*\right)}{M_P\left(\sigma^*,P^*\right)} = \frac{c_\sigma\left(\sigma^*,P^*\right)}{c_P\left(\sigma^*,P^*\right)}. \tag{3}$$

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 M_σ/M_P ratios for them. Survey and experiment planners could compare these ratios to the marginal cost ratio c_σ/c_P and see whether 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 the budget.

2.1 Length of Confidence Intervals

First, let's consider using the length of a $1-\alpha$ level confidence interval for the identified interval as the measure of precision. In this model, the identified set for θ is

$$\theta \in \left[\theta_O - P, \theta_O + P\right]. \tag{4}$$

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

$$\left[X - P - \sigma\Phi^{-1}\left(1-\alpha/2\right), X + P + \sigma\Phi^{-1}\left(1-\alpha/2\right)\right] \tag{5}$$

contains the identified set (4) exactly with probability $1-\alpha$. Φ denotes the standard normal CDF. The precision of inference from an experiment with parameters (σ, P), as measured by the length of a $1-\alpha$ confidence interval then equals

$$M^{CS(\alpha)}\left(\sigma,P\right) \equiv 2P + 2\sigma\Phi^{-1}\left(1-\alpha/2\right). \tag{6}$$

The marginal effects of changes in σ and P (partial derivatives of $M^{CI(\alpha)}$) equal $M_\sigma^{CS(\alpha)} = 2\Phi^{-1}(1-\alpha/2)$ and $M_P^{CS(\alpha)} = 2$. The ratio of these marginal effects equals

$$\frac{M_\sigma^{CS(\alpha)}}{M_P^{CS(\alpha)}} = \Phi^{-1}(1-\alpha/2)\,. \tag{7}$$

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 identified interval by 1.96ε. The evaluation of the relative effects of reducing the sampling error and the extent of partial identification depends on the chosen confidence level $1-\alpha$. Thus, using 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.

Imbens and Manski (2004) proposed an alternative type of confidence interval which covers each point in the identified set with probability $1-\alpha$, but may cover the whole identified set with a smaller probability (see Stoye 2008 for more details). In the present problem, the shortest Imbens-Manski confidence interval is $\left[X - M^{CP(\alpha)}(\sigma,P), X + M^{CP(\alpha)}(\sigma,P)\right]$ where $M^{CP(\alpha)}(\sigma,P) > 0$ is the solution to

$$\Phi\left(\frac{P + M^{CP(\alpha)}(\sigma,P)}{\sigma}\right) - \Phi\left(\frac{P - M^{CP(\alpha)}(\sigma,P)}{\sigma}\right) = 1-\alpha. \tag{8}$$

The ratio of its partial derivatives then equals

$$\frac{M_\sigma^{CP(\alpha)}}{M_P^{CP(\alpha)}} = \frac{\frac{d}{d\sigma}\left[\Phi\left(\frac{P+M}{\sigma}\right) - \Phi\left(\frac{P-M}{\sigma}\right)\right]}{\frac{d}{dP}\left[\Phi\left(\frac{P+M}{\sigma}\right) - \Phi\left(\frac{P-M}{\sigma}\right)\right]} = \frac{-\frac{P+M}{\sigma}\phi\left(\frac{P+M}{\sigma}\right) + \frac{P-M}{\sigma}\phi\left(\frac{P-M}{\sigma}\right)}{\phi\left(\frac{P+M}{\sigma}\right) - \phi\left(\frac{P-M}{\sigma}\right)}. \tag{9}$$

As Figure 1 illustrates, this ratio is close $M_\sigma^{CS(2\alpha)}/M_P^{CS(2\alpha)}$ for small values of σ/P.

3 Minimax Measures of Precision

3.1 Absolute and Square Loss

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

$$E_X L\left(\hat{\theta}(X)-\theta\right), \tag{10}$$

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 $\Theta=\{\theta_O\in\mathbb{R},\theta_U\in[-P,P]\}$:

$$M^L(\sigma,P)\equiv\sup_{\theta_O,\theta_U} E_X L\left(\hat{\theta}(X)-(\theta_O+\theta_U)\right) \tag{11}$$

could be used as a conservative measure of the precision of $\hat{\theta}(X)$. Since $\hat{\theta}(X)$ is optimal in the sense of minimizing (11), $M^L(\sigma,P)$ is also a measure of precision of the experimental data itself.

Proposition 2 shows that a simple estimator $\hat{\theta}^*(X)=X$ minimizes maximum expected loss (11) for a broad class of symmetric convex loss functions. This class includes square loss and absolute loss, for which I derive more specific results later. Formally, suppose that the loss function L 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 (a)-(d) imply that L is continuous, non-negative, and non-decreasing on $[0,+\infty)$, while (e) ensures that expected loss is finite with normally distributed X.

Proposition 2 *If loss function L satisfies Condition 1, $\theta_O\in\mathbb{R}$, $\theta_U\in[-P,P]$, and $X\sim\mathcal{N}\left(\theta_O,\sigma^2\right)$, then the estimator $\hat{\theta}^*(X)=X$ minimizes maximum expected loss (11), which equals*

$$M^L(\sigma,P)=\begin{cases}\int_{-\infty}^{+\infty}L(t)\frac{1}{\sigma}\phi\left(\frac{t-P}{\sigma}\right)dt & \text{for } \sigma>0,\\ L(P) & \text{for } \sigma=0.\end{cases} \tag{12}$$

For square and absolute loss functions it is possible to evaluate (12) and its partial derivatives analytically. In case of square loss $L(t)=t^2$, the maximum mean squared error of $\hat{\theta}^*$ equals

$$M^{MSE}(\sigma,P)=\int_{-\infty}^{+\infty}t^2\frac{1}{\sigma}\phi\left(\frac{t-P}{\sigma}\right)dt=\int_{-\infty}^{+\infty}(s\sigma+P)^2\phi(s)\,ds=\sigma^2+P^2. \tag{13}$$

The marginal effects of changes in σ and P on the minimax MSE equal $M_{\sigma}^{MSE} = 2\sigma$ and $M_{P}^{MSE} = 2P$. The ratio of these marginal effects equals

$$\frac{M_{\sigma}^{MSE}}{M_{P}^{MSE}} = \frac{\sigma}{P}. \tag{14}$$

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 $M^{CS(\alpha)}$, which does not depend on σ/P. Whenever $\sigma/P < \Phi^{-1}(1-\alpha/2)$, $M_{\sigma}^{MSE}/M_{P}^{MSE} < M_{\sigma}^{CI(\alpha)}/M_{P}^{CI(\alpha)}$ and the minimax MSE measure of precision implies that further reducing sampling error is not as important as the length of confidence interval measure would suggest. In any experiment or survey with the standard error smaller than the length of the identified 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 a 95% confidence interval. The difference between the "marginal rates of substitution" produced by the two methods could be particularly striking when considering large sample surveys in which the extent of partial identification could greatly exceed sampling error in magnitude.

For the absolute loss function $L(t) = |t|$, the minimax mean absolute deviation (MAD) of $\hat{\theta}^*$ equals

$$\begin{aligned} M^{MAD}(\sigma, P) &= \int_{-\infty}^{+\infty} |t| \frac{1}{\sigma}\phi\left(\frac{t-P}{\sigma}\right) dt = \int_{-\infty}^{+\infty} |s\sigma + P| \phi(s) \, ds = \\ &= \int_{-P/\sigma}^{+\infty} (s\sigma + P)\phi(s)\, ds - \int_{-\infty}^{-P/\sigma} (s\sigma + P)\phi(s)\, ds = \\ &= \sigma\phi(P/\sigma) + P\Phi(P/\sigma) + \sigma\phi(P/\sigma) - P\Phi(-P/\sigma) = \\ &= 2\sigma\phi(P/\sigma) + P\left[\Phi(P/\sigma) - \Phi(-P/\sigma)\right] \end{aligned} \tag{15}$$

since $\int s\phi(s)\,\partial s = -\phi(s)$.The marginal effects of changes in σ and P on the minimax MAD are $M_{\sigma}^{MAD} = 2\phi(P/\sigma)$ and $M_{P}^{MAD} = \Phi(P/\sigma) - \Phi(-P/\sigma)$. The ratio of these marginal effects equals

$$\frac{M_{\sigma}^{MAD}(\sigma, P)}{M_{P}^{MAD}(\sigma, P)} = \frac{2\phi(P/\sigma)}{\Phi(P/\sigma) - \Phi(-P/\sigma)}. \tag{16}$$

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

Similarly to the minimax MSE, for sufficiently large values of P/σ the minimax 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 $M_{\sigma}^{MAD}/M_{P}^{MAD} < M_{\sigma}^{CS(.05)}/M_{P}^{CS(.05)}$ whenever $\sigma < 2.11P$. MAD and MSE measures yield similar conclusions about the relative benefits of reducing σ and P for small values of P/σ, since (16) $\approx \sigma/P$ when $P/\sigma \to 0$.

3.2 Regret Loss in Treatment Choice Problems

The next measure of precision - minimax regret - is motivated by considering the economic loss resulting from incorrect inference about θ when it is the difference in average returns of two alternative policy decisions and the aim of inference is to choose which policy to implement.

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 $r_2 - r_1 = \theta$. If policy choice is denoted by $a = 1$ for the second policy and $a = 0$ for the first (the choice could also be randomized with $a \in (0, 1)$), then the *regret loss* function is

$$
\begin{aligned}
L(a, \theta) &= \begin{cases} \theta\,[1 - a] & \text{if } \theta > 0, \\ -\theta a & \text{if } \theta \le 0, \end{cases} \\
&= \theta\,(\mathrm{I}\,[\theta > 0] - a)
\end{aligned} \tag{17}
$$

The method by which the decision maker chooses which policy to implement based on experimental data could be summarized by a *statistical treatment rule* $\delta(X)$, which is a function mapping feasible realizations of $X \in \mathbb{R}$ into actions in the $[0, 1]$ interval. The regret of statistical treatment rule δ is the average (over the distribution of X) regret loss incurred by the decision maker using rule δ. It is a function of θ_O and θ_U, and in this problem equals

$$
R(\delta, (\theta_O, \theta_U)) \equiv \begin{cases} \theta\,[1 - E_{\theta_O}\delta(X)] & \text{if } \theta > 0, \\ -\theta E_{\theta_O}\delta(X) & \text{if } \theta \le 0, \end{cases} \tag{18}
$$

where $E_{\theta_O}\delta(X)$ denotes the average value of $\delta(X)$. For $\theta < 0$, $E_{\theta_O}\delta(X)$ is the probability of mistakenly choosing the inferior second policy, while for $\theta > 0$, the probability of error is $1 - E_{\theta_O}\delta(X)$.

Minimizing maximum regret is a criterion suggested by Savage (1951) as a clarification of Wald's *minimax principle* (1950). Regret is a natural reparametrization for loss functions that are not minimized at zero by any action. A number of recent papers in Econometrics, starting with Manski (2004), applied the criterion to treatment choice problem. Similar loss function could also be motivated by the problem of eliciting valuation of public projects from a sample of individuals (an "economic jury") for the purpose of deciding whether they're efficient (McFadden 2012). Selective non-participation in such juries leads to partial identification of the valuation of the project and the analysis in this paper applies to the trade off between increasing participation and increasing sample size in such juries.

The following Proposition derives statistical treatment rules that minimize maximum regret for given experimental parameters (σ, P) and their minimax regret.

Proposition 3 *a) For $\sigma > 2P\phi(0)$, the unique minimax regret statistical treatment rule is $\delta^*(X) \equiv \mathrm{I}\left|X > 0\right|$. Its maximum regret equals*

$$M^{MMR}(\sigma, P) = \max_{h>0}\left[h\Phi\left(\frac{P-h}{\sigma}\right)\right], \tag{19}$$

which is greater than $P/2$ and is a strictly increasing function of σ for any given P.
b) For $\sigma \leq 2P\phi(0)$, statistical treatment rules

$$\delta_{M(\sigma,P)}(X) \equiv \begin{cases} \mathrm{I}\left|X > 0\right| & \text{if } \sigma = 2P\phi(0), \\ \Phi\left(X/\sqrt{(2P\phi(0))^2 - \sigma^2}\right) & \text{if } \sigma < 2P\phi(0), \end{cases} \tag{20}$$

minimize maximum regret, which equals $M^{MMR}(\sigma, P) = P/2$.

The results of Proposition 3 are qualitatively similar to those obtained by Stoye (2012), 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 the same result holds, the minimax regret statistical treatment rule δ^* is identical for all values of $P \leq \sigma/(2\phi(0))$, 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.

Statistical treatment rule in part (b) of Proposition 3 may not be the only one that minimizes maximum regret, but deriving one is sufficient to make conclusions about the minimax regret value, and thus about the precision of inference from the data for treatment choice.

Precision of inference generated by an experiment with parameters (σ, P), as measured by minimax regret for treatment choice, is

$$M^{MMR}(\sigma, P) = \begin{cases} \max_{h>0} \left\{ h\Phi\left(\frac{P-h}{\sigma}\right)\right\} & \text{if } \sigma > 2P\phi(0), \\ \frac{P}{2} & \text{if } \sigma \leq 2P\phi(0), \end{cases} \tag{21}$$

This measure of precision could yield the most drastic conclusions about the relative benefits of reducing the extent of partial identification since $M_{\sigma}^{MMR} = 0$ for $\sigma/P \leq 2\phi(0) \approx 0.8$, implying that reducing the extent of partial identification is not only important, but is the only way to reduce minimax regret and improve the inferential precision of experimental or survey data for treatment choice.

4 Average Risk Measures

One of the concerns with using minimax measures of risk is that they overemphasize extreme parameter values. For all the measures considered above, the risk is maximized at $\theta_U = \pm P$, which may overemphasize the relative benefits of reducing partial identification. For comparison, I consider in this section measures of precision based on average risk (for the same loss functions) with a Uniform$[-P, P]$ prior on θ_U and an independent improper flat prior on θ_O (Lebesgue measure on $\mathbb{R}$). Bayesian inference in partially identified models is more sensitive to the choice of prior (since some of its features are not "diluted" by any amount of sample data) and the uniform prior is considered here for its conventionality (Hirano and Porter (2009) considered average risk with a flat prior for regret loss in treatment choice problems with point identified θ). The results would clearly be very different, for example, with a prior that places point mass on $\theta_U = 0$ (implying that there is no partial identification problem) or on $\theta_O = 0$ (implying that there is no benefit to sampling). While the prior is improper, identical results could be obtained by considering a sequence of proper $\mathcal{N}\left(0,\tau^2\right)$ priors on θ_O with $\tau \longrightarrow \infty$.

With the prior on θ_O placing uniform measure on the real line and $X \sim \mathcal{N}\left(\theta_O,\sigma^2\right)$, the posterior measure is a proper normal distribution $\theta_O|X \sim \mathcal{N}\left(X,\sigma^2\right)$. The posterior measure on θ_U remains Uniform$[-P,P]$. The posterior MSE and MAD are minimized by $\hat{\theta}_B = X$, which is both the mean and the median of the posterior distribution of θ. The posterior mean squared error equals

$$\begin{aligned}
& \int_{-P}^{P}\left[\int_{\mathbb{R}}(\theta_O+\theta_U-X)^2\phi\left(\frac{\theta_O-X}{\sigma}\right)d\theta_O\right]\frac{1}{2P}d\theta_U \\
= & \int_{-P}^{P}\left[\int_{\mathbb{R}}(t_O+\theta_U)^2\phi\left(\frac{t_O}{\sigma}\right)dt_O\right]\frac{1}{2P}d\theta_U = \\
= & \int_{-P}^{P}\left[\sigma^2+\theta_U^2\right]\frac{1}{2P}d\theta_U = \sigma^2+\frac{P^2}{3}. \qquad (22)
\end{aligned}$$

Since it is constant in X, the average MSE (with respect to the prior) also equals $M^{AMSE}(\sigma,P) = \sigma^2 + P^2/3$, and the ratio of its partial derivatives is

$$\frac{M_\sigma^{AMSE}}{M_P^{AMSE}} = 3\frac{\sigma}{P}. \qquad (23)$$

For regret loss (17), it is optimal to adopt the second policy if the posterior mean is positive, which happens if $X > 0$, hence $\delta^*(X) \equiv \mathrm{I}\left|X>0\right|$. The average regret with respect to the prior then equals

$$\begin{aligned}
M^{AR} &= \int_{\mathbb{R}}\int_{-P}^{P}(\theta_O+\theta_U)\left(\mathrm{I}\left[\theta_O+\theta_U>0\right]-E_{\theta_O}\delta\left(X\right)\right)\frac{1}{2P}d\theta_U d\theta_O \\
&= \int_{\mathbb{R}}\int_{-P}^{P}(\theta_O+\theta_U)\left(\mathrm{I}\left[\theta_O+\theta_U>0\right]-\Phi\left(\frac{\theta_O}{\sigma}\right)\right)\frac{1}{2P}d\theta_U d\theta_O. \qquad (24)
\end{aligned}$$

The partial derivatives of M^{AR} equal

$$M_\sigma^{AR} = \sigma \text{ and } M_P^{AR} = P/3 \qquad (25)$$

(see Appendix), hence their ratio $M_\sigma^{AR}/M_P^{AR} = 3\sigma/P$ is the same as for average square loss, while the two loss functions yield starkly different results when minimax measures are considered. Predictably, the average regret loss always implies that reducing P and σ always has some positive value, while the minimax regret value didn't vary with σ at all for small values of σ/P.

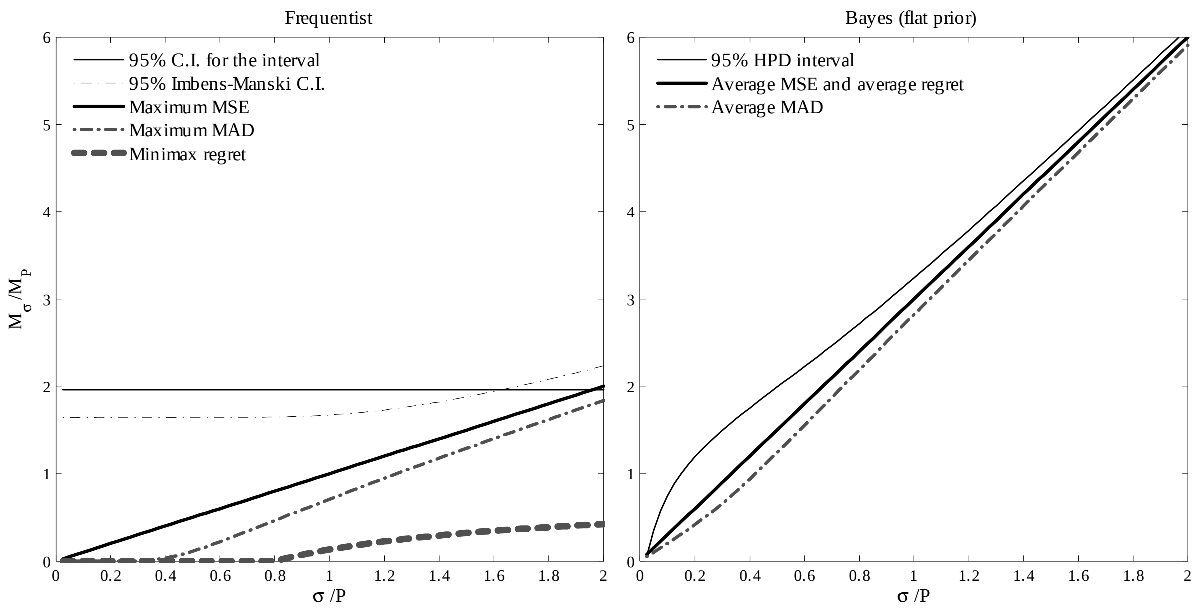

Figure 1: M_σ/M_P as a function of σ/P under different measures of precision.

Figure 1 summarizes the findings, displaying the ratios M_σ/M_P as a function of σ/P for different measures of precision. The left pane shows the ratios when the length of frequentist confidence intervals, minimax MSE, MAD and regret loss are used as measures of precision. The right pane displays the same ratios for Bayesian precision measures with a flat prior. The ratios for measures based on the length of the 95% Highest Posterior Density interval and absolute loss are computed numerically. The graph shows the extent to which conclusions depend on the measure of precision used. The disagreement in conclusions for frequentist measures could be infinite when partial identification is relatively important (σ/P is small), when the length of the confidence interval responds to changes in σ more than any other measure, while minimax regret does not change in σ at all. While all the measures based on the flat prior converge for large σ/P is large, when σ/P is smaller they also display substantial disagreement (up to 3.8 times between 95% HPD interval length and average absolute loss). The following section illustrates what range of values σ/P may be relevant in practice for problems of survey non-response.

5 Illustration: Survey Non-Response

The statistical setup of the previous sections is purposefully simplified to get clean analytical results. Here we will consider it as an informal approximation to the problem of survey non-response. Suppose that a survey samples N individuals, has response rate r and the variance

Response rate r	90%	90%	90%	90%	80%	80%	80%
Total sample size N	100	500	2000	10000	500	2000	10000
σ/P	1.054	0.471	0.236	0.105	0.25	0.125	0.056
Measures of precision:							
Length of 95% confidence interval	10.7	22.6	44.3	97.8	21.4	41.8	92.3
Length of 95% Imbens-Manski C.I.	12.3	26.8	52.6	116.4	25.3	49.6	109.8
Minimax MSE	19	91	361	1801	161	641	3201
Minimax MAD	25.5	488.4	$> 10^5$	$> 10^5$	$> 10^5$	$> 10^5$	$> 10^5$
Minimax regret (treatment choice)	119.8	$+\infty$	$+\infty$	$+\infty$	$+\infty$	$+\infty$	$+\infty$
Measures with flat prior:							
Length of 95% HPD interval	6.6	23.1	65.8	244.4	30.6	91.4	439.1
Average MSE or average regret	7	31	121	601	54.3	214.3	1067
Average MAD	7.3	37.8	171.1	890.4	76	316	1596

Table 1: Value of "converting" one non-respondent relative to an additional observation from the same subpopulation of respondents according to different measures of precision.

of the measured outcome among respondents is σ_0^2. Let θ_O stand for the mean outcome among respondents, which could be estimated with variance $\sigma^2 \approx \sigma_0^2/(rN)$. Let θ_U be the mean outcome among non-respondents, which is partially identified up to an interval of length $2P = (1-r)(Y_H - Y_L)$, where Y_H and Y_L are the highest and lowest feasible outcome values. Then

$$\frac{P}{\sigma} \approx \frac{Y_H - Y_L}{2\sigma_0}\sqrt{rN}\,(1-r)\,. \tag{26}$$

If outcomes of interest are binary, then $Y_H - Y_L = 1$ and if the mean among respondents equals $\theta_O = 1/2$, then $\sigma_0 = 1/2$ (it isn't very different for a large range of values of θ_O). Then $(Y_H - Y_L)/(2\sigma_0) = 1$ and $P/\sigma \approx \sqrt{rN}\,(1-r)$.

To make the differences in M_σ/M_P under various criteria more concrete, they could be translated into the ratio of values of two alternative marginal changes to the sample. One option is increase the response rate by $1/N$, that is, to "convert" one non-respondent to a respondent, which would reduce σ by $\Delta\sigma \approx \frac{d}{dN}\left(\sigma_0/\sqrt{rN}\right) = -\sigma_0/\left(2N\sqrt{rN}\right)$ and reduce P by $\Delta P \approx -(Y_H - Y_L)/(2N)$. An alternative option is to increase the overall sample size by $1/r$, thus adding one more respondent, which would reduce σ by $\Delta\sigma$ but leave P unchanged. The value of one non-respondent observation is then

$$\frac{\Delta\sigma M_\sigma + \Delta P M_P}{\Delta\sigma M_\sigma} = 1 + \frac{\Delta P}{\Delta\sigma}\frac{M_P}{M_\sigma} = 1 + \sqrt{rN}\frac{Y_H - Y_L}{\sigma_0}\frac{M_P}{M_\sigma} \tag{27}$$

times higher than the value of one from the existing population of respondents. For binary outcomes this ratio simplifies to $1 + 2\sqrt{rN}M_P/M_\sigma$.

Table 1 shows the value of this ratio under different measures of precision for a range of common sample sizes and response rates. Response rates of 80%-90% are common in major surveys like the NLSY, the HRS and the CPS. While their national sample sizes are greater than those in Table 1, the sample sizes for subpopulations in which researchers may be interested are smaller. The relative value of non-respondents rises with sample size regardless of the measure of precision used, because ΔP is proportional to N^{-1}, while $\Delta\sigma$ is proportional to $N^{-3/2}$, hence optimal response rates should be increasing with sample size. The value of non-respondents also rises with sample size and with the non-response rate because of the changes in M_P/M_σ under all measures of precision except for frequentist confidence interval lengths.

6 Conclusion

I have compared what different measures of inferential precision for partially identified parameters about optimal economic trade off between reducing sampling error and reducing the extent of partial identification in the data when the researcher could control both (for example, when choosing the size of a study and the level of effort to reduce non-response). The length of confidence intervals for the identified interval is the most apparent measure of precision, but it turns out to be an outlier. When the extent of partial identification is relatively important, all other measures of precision considered here (mean sqaured error, mean absolute deviation, regret loss for treatment choice) would lead the researcher to reallocate the budget more strongly towards reducing the identification problem at the expense of sampling error.

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 problems with similar structure and as a useful indicator of potential findings for future research that considers partial identification problems in greater generality.

7 Appendix: Proofs

Proof of Proposition 2

The proof relies on a well-known result (e.g., Berger 1985, p. 350, Theorem 18) that decision rule δ^* is minimax if there exists a sequence of proper priors $\{\pi_k\}$ such that $R(\delta^*, \theta) \leq \lim_{k\to\infty} r(\pi_k) < \infty$, where $r(\pi_k)$ is the Bayes risk of Bayes rule δ_k with respect to prior π_k. I will consider a sequence of such priors with $\pi_k(\theta_O) = \mathcal{N}(0, k^2)$, $\pi_k(\theta_U) = .5\mathrm{I}[|\theta_U| = P]$, and $\theta_O \perp \theta_U$.

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

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

where $c_k = k^2/(k^2 + \sigma^2)$, $\lim_{k\to\infty} c_k = 1$. Since the loss function L is convex and symmetric, the posterior Bayes estimator of θ is $\hat{\theta}_k(X) = c_k X$. Conditional on X, the variable $y = c_k X - \theta_O$ has a $\mathcal{N}(0, c_k \sigma^2)$ distribution. The posterior risk of $\hat{\theta}_k(X)$, then, equals

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

The third equality holds because L and $\pi_k(y|X)$ are symmetric. Condition 1(e) ensures that this and other improper integrals in the proof are finite. Since the posterior risk is constant in X, it is equal to the Bayes risk with prior π_k:

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

Functions $L(t)\left(\sigma\sqrt{c_k}\right)^{-1} \phi\left((t - P)/\left(\sigma\sqrt{c_k}\right)\right)$ converge pointwise in t to $L(t)\, \sigma^{-1} \phi((t - P)/\sigma)$

as $k \to \infty$. Due to Condition 1(e), Lebesgue dominated convergence theorem applies and

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

It remains to verify that the maximum risk of $\hat{\theta}^*$ over (θ_O, θ_U) is equal to this limit. The risk equals

$$\begin{aligned} R\left(\hat{\theta}^*, (\theta_O, \theta_U)\right) &= E_X L(X - (\theta_O + \theta_U)) = \\ = \int_{-\infty}^{+\infty} & L(x - \theta_O - \theta_U) \frac{1}{\sigma} \phi\left(\frac{x - \theta_O}{\sigma}\right) dx = \\ = \int_{-\infty}^{+\infty} & L(y - \theta_U) \frac{1}{\sigma} \phi\left(\frac{y}{\sigma}\right) dy = \\ = \int_{0}^{+\infty} & \left[L(y - \theta_U) + L(y + \theta_U)\right] \frac{1}{\sigma} \phi\left(\frac{y}{\sigma}\right) dy, \end{aligned} \tag{28}$$

where the last equality is due to symmetry of L and ϕ. The sum $L(y - \theta_U) + L(y + \theta_U)$ is non-decreasing in θ_U for $\theta_U > 0$ due to convexity of L, hence the risk is maximized at $|\theta_U| = P$ for any value of θ_O. Substituting $\theta_U = -P$ and $t = y + P$ in (28) yields

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

thus $\hat{\theta}^*(X) = X$ is a minimax estimator of θ under loss function L with maximum expected loss equal to $\int_{-\infty}^{+\infty} L(t) \frac{1}{\sigma} \phi\left(\frac{t-P}{\sigma}\right) dt$. □

Proof of Proposition 3(a)

The proof of part (a) relies on a well known result (e.g., Berger 1985, p. 350, Theorem 17) that decision rule δ^* is minimax if it is Bayes with respect to some proper prior π^* and for all $(\theta_O, \theta_U) \in \Theta$

$$R(\delta^*, (\theta_O, \theta_U)) \le r(\pi^*) = \int R(\delta^*, (\theta_O, \theta_U)) \partial \pi^*(\theta_O, \theta_U). \tag{29}$$

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

Decision rule $\delta^*(X) = \mathrm{I}[X > 0]$ is Bayes with respect to any symmetric two-point prior distribution π with $\pi(\theta_O^*, \theta_U^*) = 1/2$ and $\pi(-\theta_O^*, -\theta_U^*) = 1/2$, if $\theta_O^* > 0$ and $\theta_O^* + \theta_U^* > 0$. We will first find values of (θ_O, θ_U) that maximize $R(\delta^*, (\theta_O, \theta_U))$, then verify that δ^* is Bayes with respect to a two-point prior using these values, and then verify that equation (29) holds.

If $\theta = \theta_O + \theta_U \ge 0$, regret equals $R(\delta^*, (\theta_O, \theta_U)) = (\theta_O + \theta_U)[1 - E_{\theta_O} \delta^*(X)]$. For any value of θ_O it is maximized at $\theta_U^* = P$, since $E_{\theta_O}\delta^*(X)$ doesn't depend on θ_U. Since $E_{\theta_O}\delta^*(X) = 1 - \Phi(-\theta_O/\sigma)$, maximum regret of δ^* over $\theta \geq 0$ then equals (with the substitution $h = \theta_O + P$)

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

The maximum is attained at

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

Similarly, maximum regret over $\theta < 0$ also equals $\max_{h>0}\left[h\Phi\left((P-h)/\sigma\right)\right]$ and it is attained at $\theta_O = -\theta_O^*$ and $\theta_U = -P$.

The function $h\Phi\left(\frac{P-h}{\sigma}\right)$ is continuous in h with derivative

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

At $h = 0$, $\frac{d}{dh}\left[h\Phi\left(\frac{P-h}{\sigma}\right)\right] = \Phi\left(\frac{P}{\sigma}\right) > 0$. The function $\frac{\phi(y)}{\Phi(y)}$ is positive and decreasing in y, hence $\frac{h}{\sigma}\frac{\phi((P-h)/\sigma)}{\Phi((P-h)/\sigma)}$ is positive and increasing in h for $h \geq 0$ with $\lim_{h\to\infty}\frac{h}{\sigma}\frac{\phi((P-h)/\sigma)}{\Phi((P-h)/\sigma)} = +\infty$, therefore $h\Phi\left(\frac{P-h}{\sigma}\right)$ has a unique maximum over $h > 0$ at some point h^*.

The derivative of $h\Phi\left(\frac{P-h}{\sigma}\right)$ at $h = P$ equals $\Phi(0) - \frac{P}{\sigma}\phi(0) = \frac{1}{2} - \frac{P\phi(0)}{\sigma}$. By assumption $\sigma > 2P\phi(0)$, hence it is positive, $h^* > P$ and $\theta_O^* = h^* - P > 0$.

Since $\theta_O^* > 0$ and $\theta_O^* + P > 0$, δ^* is a Bayes rule with respect to prior π^* with $\pi^*(\theta_O^*, P) = 1/2$, $\pi^*(-\theta_O^*, -P) = 1/2$ and

$$r(\pi^*) = \frac{1}{2}R(\delta^*, (\theta_O^*, P)) + \frac{1}{2}R(\delta^*, (-\theta_O^*, -P)) = \max_{\theta_O,\theta_U} R(\delta^*, (\theta_O, \theta_U)),$$

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.

Maximum regret of δ^* exceeds $\frac{P}{2}$ because $h\Phi\left(\frac{P-h}{\sigma}\right)$ is not maximized at $h = P$, hence

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

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, 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, thus $\max_{h>P}\left[h\Phi\left(\frac{P-h}{\sigma}\right)\right]$ is also strictly decreasing in σ. □

Proof of Proposition 3(b)

Maximum regret of any decision rule is at least $P/2$ because

$$\begin{aligned}\max_{\theta_O,\theta_U} R\left(\delta,\left(\theta_O,\theta_U\right)\right) &\geq \max\left(R\left(\delta,(0,P)\right),R\left(\delta,(0,-P)\right)\right)= \\ &= \max\left(P\left(1-E_{\theta_O=0}\delta\left(X\right)\right),PE_{\theta_O=0}\delta\left(X\right)\right)\geq\frac{P}{2}.\end{aligned}$$

I will first show that any rule δ for which $E_{\theta_O}\delta\left(X\right)$ lies within the bounds

$$\begin{aligned} E_{\theta_O}\delta\left(X\right)\geq 1-\tfrac{P}{2(P+\theta_O)} \quad &\text{for } \theta_O\geq-\tfrac{P}{2}, \\ \text{and } E_{\theta_O}\delta\left(X\right)\leq \tfrac{P}{2(P-\theta_O)} \quad &\text{for } \theta_O\leq\tfrac{P}{2}\end{aligned} \tag{30}$$

has maximum regret of $P/2$ (hence minimizes maximum regret), then show that $\delta_{M(\sigma,P)}\left(X\right)$ satisfies this condition.

For all (θ_O,θ_U) such that $\theta\geq 0$, $R\left(\delta,\left(\theta_O,\theta_U\right)\right)=\left(\theta_O+\theta_U\right)\left(1-E_{\theta_O}\delta\left(X\right)\right)$ is increasing in θ_U, so

$$R\left(\delta,\left(\theta_O,\theta_U\right)\right)\leq R\left(\delta,\left(\theta_O,P\right)\right)=\left(\theta_O+P\right)\left(1-E_{\theta_O}\delta\left(X\right)\right).$$

If $\theta_O\geq -P/2$, the lower bound in (30) implies that

$$\left(\theta_O+P\right)\left(1-E_{\theta_O}\delta\left(X\right)\right)\leq\left(\theta_O+P\right)\frac{P}{2\left(P+\theta_O\right)}=\frac{P}{2}.$$

If $\theta_O\in[-P,-P/2]$, then since $\left(1-E_{\theta_O}\delta\left(X\right)\right)\leq 1$,

$$\left(\theta_O+P\right)\left(1-E_{\theta_O}\delta\left(X\right)\right)\leq\theta_O+P\leq\frac{P}{2}.$$

Hence for $\theta\geq 0$, $R\left(\delta,\left(\theta_O,\theta_U\right)\right)\leq 1/2$. The proof for $\theta<0$ is analogous.

Second, I will show that any decision rule δ with $E_{\theta_O}\delta(X) = q^*(\theta_O)$, where

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

satisfies inequalities (30). The proof verifies this for $\theta_O \geq 0$, it is symmetric for $\theta_O < 0$.

When $\theta_O = 0$, $q^*(0) = \Phi(0) = 1/2$, which coincides with both bounds. For $\theta_O \in [0, P/2]$, q^* satisfies the upper bound (30) because

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

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

The proof that q^* satisfies the lower bound is split into two cases: $\theta_O \in [0, P]$ and $\theta_O \geq P$.

For $\theta_O \in [0, P]$, I will prove that q^* has a higher derivative than the bound $1 - \frac{P}{2(P+\theta_O)}$, which implies the needed inequality since both are equal at $\theta_O = 0$. The derivative of q^* is

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

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

$$\begin{aligned}\frac{1}{2P}\exp\left(-\frac{\pi}{4}\left(\frac{\theta_O}{P}\right)^2\right) &\geq \frac{1}{2P}\frac{1}{1+\frac{\pi}{2}(\theta_O/P)^2} = \\ = \frac{P}{2P^2+\pi\theta_O^2} &\geq \frac{P}{2P^2+4P\theta_O+2\theta_O^2} = \frac{P}{2(P+\theta_O)^2}.\end{aligned}$$

The second inequality uses the fact that $\pi\theta_O \leq 4P$ and $2\theta_O^2 \geq 0$. On the other hand,

$$\frac{d}{d\theta_O}\left[1 - \frac{P}{2(P+\theta_O)}\right] = \frac{P}{2(P+\theta_O)^2},$$

hence q^* grows faster than the bound.

For $\theta_O \geq P$, I will use two inequalities. First, for the normal distribution $\Phi(t) > 1 - \frac{\phi(t)}{t}$ for $t > 0$ (cf. Feller 1968, Chapter VII, Lemma 2). Second, for $y \geq 0$, $\exp(y) \geq 1 + y$, thus for

$y \leq 0$, $\exp(y) \leq \frac{1}{1-y}$.

$$\begin{aligned}
q^*(\theta_O) &= \Phi\left(\frac{\theta_O}{2P\phi(0)}\right) > 1 - \frac{2P\phi(0)}{\theta_O}\phi\left(\frac{\theta_O}{2P\phi(0)}\right) = \\
&= 1 - \frac{P}{\pi\theta_O}\exp\left(-\frac{\pi}{4}\left(\frac{\theta_O}{P}\right)^2\right) > 1 - \frac{P}{\pi\theta_O}\frac{1}{1+\frac{\pi}{4}(\theta_O/P)^2} = \\
&= 1 - \frac{P}{\pi\theta_O + \frac{\pi^2}{4}\theta_O(\theta_O/P)^2} > 1 - \frac{P}{2(P+\theta_O)}.
\end{aligned}$$

The last inequality relies on observations that $\pi\theta_O > 2P$, $\pi^2/4 > 2$, and $(\theta_O/P)^2 \geq 1$.

It remains to show that the decision rule $\delta_{M(\sigma,P)}$ satisfies $E_{\theta_O}\delta_{M(\sigma,P)}(X) = q^*$, hence has maximum regret $P/2$.

For $\sigma = 2P\phi(0)$, $\delta_{M(\sigma,P)}(X) = \mathrm{I}\,|X > 0|$, thus $E_{\theta_O}\delta_{M(\sigma,P)}(X) = \Phi\left(\frac{\theta_O}{2P\phi(0)}\right) = q^*(\theta_O)$.

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

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

independent of the observed outcome $X \sim \mathcal{N}(\theta_O, \sigma^2)$. Then $X + Y \sim \mathcal{N}\left(\theta_O, (2P\phi(0))^2\right)$. Define the randomized statistical treatment rule $\tilde{\delta}(X,Y)$ as a function of both X and Y

$$\tilde{\delta}(X,Y) \equiv \mathrm{I}\,|X + Y > 0|,$$

then clearly

$$E_{\theta_O}\delta_{M(\sigma,P)}(X) = \Phi\left(\frac{\theta_O}{2P\phi(0)}\right) = q^*(\theta_O).$$

Integrating $\tilde{\delta}(X,Y)$ with respect to the distribution of Y yields

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

which thus satisfies $E_{\theta_O}\delta_{M(\sigma,P)}(X) = q^*(\theta_O)$ by construction and minimizes maximum regret, which equals $P/2$. □

Proof of Equation 25

Let's denote the inner integral over θ_U by $J(\theta_O, \sigma, P)$ and evaluate it

$$
\begin{aligned}
J(\theta_O, \sigma, P) &\equiv \frac{1}{2P}\int_{-P}^{P}(\theta_O+\theta_U)\left(\mathrm{I}[\theta_O+\theta_U>0]-\Phi\left(\frac{\theta_O}{\sigma}\right)\right)d\theta_U \\
&= \frac{1}{2P}\left(-\Phi\left(\frac{\theta_O}{\sigma}\right)\int_{-P}^{P}(\theta_O+\theta_U)\,d\theta_U+\int_{-P}^{P}(\theta_O+\theta_U)\,\mathrm{I}[\theta_U>-\theta_O]\,d\theta_U\right) \\
&= -\theta_O\Phi\left(\frac{\theta_O}{\sigma}\right)+\begin{cases} 0 & \text{for } \theta_O<-P, \\ \frac{P}{4}+\frac{\theta_O}{2}+\frac{\theta_O^2}{4P} & \text{for } \theta_O\in[-P,P], \\ \theta_O & \text{for } \theta_O>P. \end{cases}
\end{aligned}
$$

Differentiating it with respect to σ and P for a given θ_O yields

$$
\begin{aligned}
\frac{d}{dP}J(\theta_O, \sigma, P) &= \mathrm{I}[|\theta_O|\leq P]\cdot\left[\frac{1}{4}-\frac{\theta_O^2}{4P^2}\right], \\
\frac{d}{d\sigma}J(\theta_O, \sigma, P) &= \left(\frac{\theta_O}{\sigma}\right)^2\phi\left(\frac{\theta_O}{\sigma}\right).
\end{aligned}
$$

Integrating the derivatives over θ_O, we get

$$
\begin{aligned}
\frac{d}{dP}\int_{\mathbb{R}}J(\theta_O, \sigma, P)\,d\theta_O &= \int_{-P}^{P}\left[\frac{1}{4}-\frac{\theta_O^2}{4P^2}\right]d\theta_O=\frac{P}{3}, \\
\frac{d}{d\sigma}\int_{\mathbb{R}}J(\theta_O, \sigma, P)\,d\theta_O &= \int_{\mathbb{R}}\left(\frac{\theta_O}{\sigma}\right)^2\phi\left(\frac{\theta_O}{\sigma}\right)d\theta_O=\sigma.
\end{aligned}
$$

□

References

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

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

[3] Feller, W. (1968), *An Introduction to Probability and Its Applications Vol. 1 (3rd ed.)*, New York: Wiley.

[4] Hirano, K., and Porter, J.R. (2009), "Asymptotics for Statistical Treatment Rules," *Econometrica,* 77, 1683–1701.

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

[6] Imbens, G.W., and Manski, C.F. (2004), "Confidence Intervals for Partially Identified Parameters," *Econometrica*, 72, 1845-1857.

[7] Manski, C.F. (1995), *Identification Problems in the Social Sciences,* Cambridge, MA: Harvard University Press.

[8] Manski, C.F. (2004), "Statistical Treatment Rules for Heterogeneous Populations," *Econometrica,* 72, 1221–1246.

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

[10] Manski, C.F. (2007a), *Identification for Prediction and Decision.* Cambridge, MA: Harvard University Press.

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

[12] Manski, C.F. (2008), "Adaptive Partial Drug Approval: A Health Policy Proposal," *The Economists' Voice*, Vol. 6: Iss. 4, Article 9.

[13] Manski, C.F. (2009), "Diversified Treatment Under Ambiguity," *International Economic Review,* 50, 1013–1041.

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

[15] McFadden, D. (2012), "Economic Juries and Public Project Provision," *Journal of Econometrics,* 166, 116–126.

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

[17] Schlag, K.H. (2007), "Eleven – Designing Randomized Experiments Under Minimax Regret," working paper, European University Institute.

[18] Song, K. (2010), "Point Decisions for Interval-Identified Parameters," working paper, University of Pennsylvania.

[19] Stoye, J. (2007), "Minimax Regret Treatment Choice with Incomplete Data and Many Treatments," *Econometric Theory,* 23, 190–199.

[20] Stoye, J. (2008), "More on Confidence Intervals for Partially Identified Parameters," *Econometrica*, 77, 1299-1315.

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

[22] Stoye, J. (2012), "Minimax Regret Treatment Choice with Limited Validity of Experiments or with Covariates," *Journal of Econometrics,* 166, 138–156.

[23] Wald, A. (1950), *Statistical Decision Functions.* New York: Wiley.