Asymptotics for Statistical Treatment Rules

Keisuke Hirano
Department of Economics
University of Arizona
hirano@u.arizona.edu

Jack R. Porter
Department of Economics
University of Wisconsin
jrporter@ssc.wisc.edu

May 10, 2005

Abstract

This paper develops asymptotic optimality theory for statistical treatment rules. Manski (2000, 2002, 2004) and Dehejia (2005) have argued that the problem of choosing treatments to maximize social welfare is distinct from the point estimation and hypothesis testing problems usually considered in the treatment effects literature, and advocate formal analysis of decision procedures that map empirical data into policy recommendations. Motivated by the difficulty of obtaining useful finite-sample results in complex settings, we consider large-sample approximations to statistical treatment assignment problems. Using the Le Cam limits of experiments framework, we derive an asymptotic minmax regret bound on social welfare, and a minmax risk bound for a two-point loss function. We show that certain natural treatment assignment rules attain these bounds. We develop results for both smooth parametric models and semiparametric models.

1 Introduction

One major goal of treatment evaluation in the social and medical sciences is to provide guidance on how to assign individuals to treatments. For example, a number of studies have examined the problem of "profiling" individuals to identify those likely to benefit from a social program; see for example Worden (1993), O'Leary, Decker, and Wandner (1998), Berger, Black, and Smith (2001), Black, Smith, Berger, and Noel (2003), and O'Leary, Decker, and Wandner (2005). Similarly, in evaluating medical therapies, it may be important to provide guidelines on what types of patients should receive a particular therapy. Manski (2000, 2002, 2004) and Dehejia (2005) point out that the problem of assigning individuals to treatments, based on empirical data, is distinct from the problem of estimating the treatment effect efficiently or testing hypotheses about a treatment effect, and advocate formal analysis of decision *procedures* which map empirical data into policy recommendations. In this paper, we develop asymptotic approximation results to guide construction of such statistical treatment assignment rules. We obtain local asymptotic minmax bounds on risks of treatment assignment rules, and show how to construct rules that achieve these bounds.

We follow Manski (2004) by considering statistical approaches to treatment assignment within Wald's statistical decision theory framework.[1] This approach starts with a specification of the loss to the social planner when a particular action is taken in a particular state of nature. Decision rules are compared on the basis of their expected loss—their risk—under different parameter values. We focus on finding minmax rules, which have the smallest worst-case risk. Unfortunately, it is often extremely difficult to compute minmax decision rules when the class of potential rules is large. For example, Manski (2004) considers a randomized experiment with a discrete covariate, and a bounded continuous outcome. In this setting, it appears to be infeasible to solve for minmax rules without limiting the class of possible treatment rules. Manski develops bounds on risks over a restricted class of "empirical conditional success" rules.

Motivated by the difficulty of obtaining exact optimality results in many empirically relevant settings, we consider large-sample approximations to statistical treatment assignment problems. We focus on obtaining statistical decision rules that are approximately minmax for certain loss functions that arise naturally within the treatment assignment setting. We consider three loss functions: a two-point loss function that has been used in hypothesis testing theory, a "social welfare" loss function, and a regret version of the social welfare loss that has been proposed by Manski.

We use Le Cam's asymptotic extension of the Wald (1950) theory. The key idea is to obtain an asymptotic approximation to the entire statistical decision problem, not just a particular estimator or test statistic. Often, the approximate decision problem is considerably simpler, and can be solved exactly. Then, one finds a sequence of rules in the original problem that asymptotically matches the optimal rule in the simple limiting version of the decision problem. In regular parametric models, the treatment assignment problem is asymptotically equivalent to a simpler problem, in which one observes a single draw from a multivariate normal distribution with unknown mean and known variance matrix, and must decide whether a linear combination of the elements of the mean vector is greater than zero. Not surprisingly, there is a close connection between the treatment assignment problem and a one-sided hypothesis testing problem. In the limiting version of the treatment assignment problem, we appeal to an essential complete class theorem due to Karlin and Rubin (1956), which allows us to restrict attention to relatively simple types of rules on one-dimensional subspaces of the parameter space. Using this "slicing" of the parameter space, we obtain exact minmax bounds and the minmax rules for the three loss functions we are considering. It turns out that the same rule is minmax over all subspaces, leading to a minmax result over the entire parameter space. Finally, we use these exact results in the simple multivariate normal case to provide asymptotic minmax bounds, and sequences of decision rules which achieve these

[1] For other economic applications of minmax statistical decision theory, see Chamberlain (2000).

bounds, in the original sequence of decision problems. For a symmetric version of the two-point loss, and for the minmax regret criterion, a simple rule based on an asymptotically efficient parameter estimator (such as the maximum likelihood estimator) is asymptotically minmax. Although this rule has a very natural form, it implies less conservative decision making than hypothesis testing at conventional significance levels.

We then consider semiparametric settings, where the welfare gain of the treatment can be expressed as a given functional of the unknown distribution. Under regularity conditions that permit estimation of the welfare gain at a $\sqrt{n}$ rate, we derive similar results to the parametric case. In this case, the limit experiment can be expressed as a single draw from a countable sequence of independent normal distributions. As in the parametric case, we solve the problem along certain one-dimensional subspaces and then show that the same rule is optimal for all such subspaces. Optimal rules can be formed based on asymptotically efficient point estimators of the welfare gain functional. This finding is related to the uniformly most powerful property of certain semiparametric one-sided tests developed by Choi, Hall, and Schick (1996). As an example, we consider Manski's conditional empirical success rules, and show that they are asymptotically minmax regret rules when the model for outcomes is essentially unrestricted.

Our results are local asymptotic results: we reparametrize the models so that the problem of determining whether to assign the treatment (conditional on a specific covariate value) does not become trivial as the sample size increases. This is the same parameter localization commonly used in hypothesis testing theory, where one considers Pitman alternatives to the null hypothesis. Alternatively, one could examine asymptotic minmaxity from a large deviations perspective: see for example Puhalskii and Spokoiny (1998).

In the next section, we set up the basic statistical treatment assignment problem. In Section 3, motivated by the difficulty of solving the problem exactly, we take a local parameter approach and show the limiting Gaussian form of the treatment assignment problem. In Section 4, we solve the approximate treatment assignment problem according to the minmax criterion, and then apply the solution to obtain asymptotic minmax bounds on risk and asymptotic minmax rules in the original sequence of decision problems. Section 5 then develops the semiparametric version of the argument.

2 Statistical Treatment Assignment Problem

2.1 Known Outcome Distributions

Following Manski (2000, 2002, 2004), we consider a social planner, who assigns individuals to different treatments based on their observed background variables. Suppose that individuals have covariates X, a random variable on some space $\mathcal{X}$, with marginal distribution F_X. For simplicity the set of possible treatment values is $\mathcal{T} = \{0, 1\}$. (The analysis can be generalized to a finite number of possible treatments.) The planner observes $X = x$, and assigns the individual to treatment 1 according to a treatment rule

$$\delta(x) = Pr(T = 1|X = x).$$

Let Y_0 and Y_1 denote potential outcomes for the individual, and let their distribution functions conditional on $X = x$ be denoted $F_0(\cdot|x)$ and $F_1(\cdot|x)$ respectively. Given the rule δ, the outcome distribution conditional on $X = x$ is

$$F_\delta(\cdot|x) = \delta(x)F_1(\cdot|x) + (1 - \delta(x))F_0(\cdot|x).$$

For a given outcome distribution F, let the *social welfare* be a functional $W(F)$. We define

$$W_0(x) = W(F_0(\cdot|x)), \quad W_1(x) = W(F_1(\cdot|x)).$$

For example, we could consider a utilitarian social welfare function

$$W(F) = \int w(y)dF(y)$$

where $w : \mathbb{R} \to \mathbb{R}$ is a strictly increasing, concave function. Then

$$W_0(x) = \int w(y)dF_0(y|x), \qquad W_1(x) = \int w(y)dF_1(y|x).$$

Suppose that F_0 and F_1 are known. Then, the optimal rule would have, for F_x-almost all x,

$$\delta^*(x) = \begin{cases} 1 & \text{if } W_0(x) < W_1(x) \\ 0 & \text{if } W_0(x) > W_1(x) \end{cases}$$

(For x such that $W_0(x) = W_1(x)$, any value of $\delta^*(x)$ is optimal.)

2.2 Unknown Outcome Distributions

If F_0 and F_1 are not known, the optimal rule described above is not feasible. Suppose that F_0 and F_1 can be characterized by a parameter $\theta \in \Theta$, and let $w_0(x, \theta)$ and $w_1(x, \theta)$ denote the values for $W_0(x)$ and $W_1(x)$ when F_0 and F_1 follow θ. It will be convenient to work with the welfare contrast

$$g(x, \theta) := w_1(x, \theta) - w_0(x, \theta).$$

We assume that w_0 and g are continuously differentiable in θ for F_X-almost all x.[2]

Suppose we have some data that are informative about θ. For example, we might have run a

[2]For a discussion of the relationship between the net social welfare and traditional measures of effects of treatments, such as the average treatment effect, see Dehejia (2003).

randomized experiment in the past that is informative about the treatment effect. We can express this as $Z_n \sim P_\theta^n$, where $\{P_\theta^n, \theta \in \Theta\}$ is a collection of probability measures on some space $\mathcal{Z}^n$. Here, we interpret n as the sample size, and we will consider below a sequence of experiments $\mathcal{E}_n = \{P_\theta^n, \theta \in \Theta\}$ as the sample size grows.

Example 1 *Dehejia (2005) uses data from a randomized evaluation which compared the Greater Avenues for Independence (GAIN) program to the standard AFDC program for welfare recipients in Alameda County, California. The two possible treatments are the GAIN program ($T = 1$) and the standard AFDC program ($T = 0$). The outcome of interest is individual earnings in various quarters after the program. Since many welfare recipients had zero earnings even after the program, Dehejia used a Tobit model. A simplified version of Dehejia's model is:*

$$Y_i = \max\{0, \alpha_1' X_i + \alpha_2 T_i + \alpha_3' X_i \cdot T_i + \epsilon_i\},$$

where the ϵ_i are IID $N(0, \sigma^2)$. Dehejia estimated this model using the n experimental subjects, and then produced predictive distributions for a hypothetical $(n+1)$th subject to assess different treatment assignment rules.

In our notation, the parameter vector is $\theta = (\alpha_1, \alpha_2, \alpha_3, \sigma)$, and the data informative about θ are

$$Z_n = \{(T_i, X_i, Y_i) : i = 1, \ldots, n\}.$$

For a simple utilitarian social welfare measure that takes the average earnings of individuals, we would have

$$\begin{aligned} w_0(x, \theta) &= E_\theta[Y_{n+1} | X_{n+1} = x, T_{n+1} = 0]; \\ w_1(x, \theta) &= E_\theta[Y_{n+1} | X_{n+1} = x, T_{n+1} = 1]; \\ g(x, \theta) &= E_\theta[Y_{n+1} | X_{n+1} = x, T_{n+1} = 1] - E_\theta[Y_{n+1} | X_{n+1} = x, T_{n+1} = 0]. \end{aligned}$$

□

A <u>randomized statistical treatment rule</u> is a mapping $\delta : \mathcal{X} \times \mathcal{Z}^n \to [0, 1]$. We interpret it as the probability of assigning a (future) individual with covariate $X = x$ to treatment, given past data $Z_n = z$:

$$\delta(x, z) = Pr(T = 1 | X = x, Z_n = z).$$

In order to implement the Wald statistical decision theory approach, we need to specify a loss function which connects actions with consequences. We consider three possible loss functions. The first is taken from standard hypothesis testing theory, and penalizes making the wrong choice by an amount that depends only on whether the optimal assignment is treatment or control:

Loss A:

$$L^A(\delta,\theta,x) = \begin{cases} K_0 \cdot (1-\delta) & \text{if } g(x,\theta) > 0 \\ K_1 \cdot \delta & \text{if } g(x,\theta) \leq 0 \end{cases}$$

$$K_0 > 0, \qquad K_1 > 0$$

The next loss function corresponds to maximizing social welfare:

Loss B:

$$\begin{aligned} L^B(\delta,\theta,x) &= -[\delta W_1(x) + (1-\delta)W_0(x)] \\ &= -W_0(x) - \delta[W_1(x) - W_0(x)] \\ &= -w_0(x,\theta) - \delta \cdot g(x,\theta). \end{aligned}$$

Unfortunately, when combined with the minmax criterion introduced below, loss B typically leads to degenerate minmax solutions. This degeneracy arises because the loss may be unbounded in some region of the parameter space for each rule. This problem was pointed out by Savage (1951) and motivated his introduction of the minmax regret criterion, which compares the welfare loss to the welfare loss of the infeasible optimal rule. We follow Savage (1951) and Manski (2004) in focusing on minmax regret, and relegate the results for Loss B to Appendix B.

The minmax regret criterion can be implemented by modifying the loss function. Recall that the infeasible optimal treatment rule is $\delta^*(x) = 1(g(x,\theta) > 0)$. The regret is the welfare loss of a rule, compared with the welfare loss of the infeasible optimal rule:

Loss C:

$$\begin{aligned} L^C &= L^B(\delta,\theta,x) - L^B(\delta^*,\theta,x) \\ &= g(x,\theta)[1(g(x,\theta) > 0) - \delta]. \end{aligned}$$

Note that loss A and C do not depend on $w_0(x,\theta)$, so that only the welfare contrast $g(x,\theta)$ is relevant for the decision problem.

The risk of a rule $\delta(x,z)$ under loss L and given θ, is

$$\begin{aligned} R(\delta,\theta) &= EL(\delta(X,Z),\theta,X) \\ &= \int\int L(\delta(x,z),\theta,x)dP_\theta^n(z)dF_X(x) \end{aligned}$$

A minmax decision rule over some class Δ of decision rules, solves

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

For the local asymptotic theory to follow, it is more convenient to consider a "pointwise-in-X" version of the minmax problem:

$$\inf_{\delta(x,\cdot) \in \Delta} \sup_{\theta \in \Theta} \int L(\delta(x, z), \theta, x) dP_\theta^n(z).$$

In general, this can lead to different minmax rules than the global minmax problem.

3 Regular Parametric Models

We first consider regular parametric models, where the unknown parameter θ is finite-dimensional and satisfies conventional smoothness conditions. To develop asymptotic approximations, we adopt a local parametrization approach, as is standard in the literature on efficiency of estimators and test statistics. We use the local asymptotic representation of regular parametric models by a Gaussian shift experiment to derive a simple approximate characterization of the decision problem.

3.1 Plug-in Rules and Local Parametrization

Suppose that Θ is an open subset of $\mathbb{R}^k$, and that the $\{P_\theta^n, \theta \in \Theta\}$ are dominated by some measure μ^n and satisfy conventional regularity conditions. Then a natural estimator of θ is the maximum likelihood estimator

$$\hat{\theta}(Z_n) = \arg\max_\theta \frac{dP_\theta^n}{d\mu^n}(Z_n),$$

and one possible treatment assignment rule is the "plug-in" rule

$$\hat{\delta}(x, Z_n) = 1(g(x, \hat{\theta}(Z_n)) > 0),$$

with associated risk

$$\begin{aligned} R_x(\hat{\delta}, \theta) &= \int L(\hat{\delta}(x, z), \theta, x) dP_\theta^n(z) \\ &= \int L(1(g(x, \hat{\theta}(z)) > 0), \theta, x) dP_\theta^n(z). \end{aligned}$$

Except in certain special cases, the exact distribution of the MLE $\hat{\theta}$ under θ_0 cannot be easily obtained, and as a consequence it is difficult to simply calculate the risk of a given decision rule, much less find the rule which maximizes the worst-case risk. Although the exact distribution of the MLE is rarely known in a useful form, many asymptotic approximation results are available for the MLE and other estimators. This suggests that for reasonably large sample sizes, we may be able to use such approximations to study the corresponding decision rules.

First, consider the issue of consistency: As $n \to \infty$, the MLE and many other estimators satisfy $\hat{\theta} \xrightarrow{\text{p}} \theta_0$. This implies that the rule $\hat{\delta} = 1(g(x, \hat{\theta}) > 0)$ will be consistent, in the sense that if $g(x, \theta_0) > 0$, $\Pr(\hat{\delta} = 1) \longrightarrow 1$, and if $g(x, \theta_0) < 0$, then $\Pr(\hat{\delta} = 0) \longrightarrow 1$.

Although this is a useful first step, it does not permit us to distinguish between plug-in rules based on different consistent estimators, or to consider more general rules that do not have the plug-in form. We therefore focus on developing local asymptotic approximations to the distributions of decision rules.

For the MLE, under suitable regularity conditions we have

$$\sqrt{n}(\hat{\theta} - \theta_0) \overset{\theta_0}{\rightsquigarrow} N(0, I_{\theta_0}^{-1}),$$

where we use $\overset{\theta_0}{\rightsquigarrow}$ to denote convergence in distribution (weak convergence) under θ_0, and I_{θ_0} is the Fisher information matrix. By the delta method, we have that

$$\sqrt{n}(g(x, \hat{\theta}) - g(x, \theta_0)) \overset{\theta_0}{\rightsquigarrow} N(0, \dot{g}' I_{\theta_0}^{-1} \dot{g}),$$

where $\dot{g} = \frac{\partial}{\partial \theta} g(x, \theta_0)$. Consider an alternative estimator $\tilde{\theta}$ with a larger asymptotic variance:

$$\sqrt{n}(\tilde{\theta} - \theta_0) \overset{\theta_0}{\rightsquigarrow} N(0, V),$$

where $V - I_{\theta_0}^{-1}$ is positive definite. Since $\tilde{\theta}$ is "noisier" than the MLE, we might expect that the plug-in rule $\tilde{\delta} = 1(g(x, \tilde{\theta}) > 0)$ should do worse than $\hat{\delta}$. One way to make this reasoning formal, is to adopt the local parametrization (Pitman alternative) approach, which is commonly used in asymptotic analysis of hypothesis tests.[3] In our setting, this means considering values for θ such that $g(x, \theta)$ is "close" to 0, so that there is a nontrivial difficulty in distinguishing between the effects of the two treatments as sample size grows. Specifically, assume that θ_0 is such that

$$g(x, \theta_0) = 0,$$

and consider parameter sequences of the form $\theta_0 + \frac{h}{\sqrt{n}}$, for $h \in \mathbb{R}^k$. For the MLE, it can typically be shown that

$$\sqrt{n}(\hat{\theta} - \theta_0 - h/\sqrt{n}) \overset{\theta_0 + h/\sqrt{n}}{\rightsquigarrow} N(0, I_{\theta_0}^{-1}),$$

[3]Alternatively, we could use large-deviations asymptotics, in analogy with Bahadur efficiency of hypothesis tests. Manski (2003) uses finite-sample large-deviations results to bound the risk properties of certain types of treatment assignment rules in a binary-outcome randomized experiment. Puhalskii and Spokoiny (1998) develop a large-deviations version of asymptotic statistical decision theory and apply it to estimation and hypothesis testing.

where $\overset{\theta_0+h/\sqrt{n}}{\rightsquigarrow}$ denotes weak convergence under the sequence of probability measures $P_{\theta_0+h/\sqrt{n}}$. We will sometimes abbreviate this as $\overset{h}{\rightsquigarrow}$.

By assumption, for all $h \in \mathbb{R}^k$, $\sqrt{n}(g(x,\theta_0+h/\sqrt{n})-g(x,\theta_0)) \to \dot{g}'h$. Then by standard calculations

$$P_{\theta_0+h/\sqrt{n}}(g(x,\hat{\theta})>0) \to 1-\Phi\left(\frac{-\dot{g}'h}{\sqrt{\dot{g}'I_{\theta_0}^{-1}\dot{g}}}\right)=\Phi\left(\frac{\dot{g}'h}{\sqrt{\dot{g}'I_{\theta_0}^{-1}\dot{g}}}\right).$$

As an illustration, consider Loss C, the welfare regret loss function. In order to keep the loss from degenerating to 0 as sample size increases, we scale it up by the factor $\sqrt{n}$, leading to a loss function

$$L^C(\delta,h,x) = -\sqrt{n}\cdot g(x,\theta_0+h/\sqrt{n})[1(g(x,\theta_0+h/\sqrt{n})>0)-\delta(x,z)].$$

Then the risk can be written as

$$\begin{aligned} R^C(\delta,h,x) &= E_h L^C(\delta(x,Z),h,x) \\ &= \sqrt{n}\cdot g(x,\theta_0+h/\sqrt{n})\cdot[1(g(x,\theta_0+h/\sqrt{n})>0)-E_h(\delta(x,Z))] \end{aligned}$$

For the MLE plug-in rule,

$$\begin{aligned} R^C(\hat{\delta},h,x) &= \sqrt{n}g(x,\theta_0+h/\sqrt{n})\cdot[1(\sqrt{n}g(x,\theta_0+h/\sqrt{n})>0)-P_h(g(x,\hat{\theta})>0)] \\ &\to \dot{g}'h\cdot\left[1(\dot{g}'h>0)-\Phi\left(\frac{\dot{g}'h}{\sqrt{\dot{g}'I_{\theta_0}^{-1}\dot{g}}}\right)\right] \end{aligned}$$

Now consider another estimator $\tilde{\theta}$ with

$$\sqrt{n}(\tilde{\theta}-\theta_0-h/\sqrt{n}) \overset{\theta_0+h/\sqrt{n}}{\rightsquigarrow} N(0,V),$$

where $V-I_{\theta_0}^{-1}$ is positive semidefinite, and let $\tilde{\delta}=1(g(x,\tilde{\theta})>0)$. By straightforward calculations, it can be shown that for all h,

$$\lim_{n\to\infty} R^C(\tilde{\delta},h,x) \geq \lim_{n\to\infty} R^C(\hat{\delta},h,x).$$

Thus, $\hat{\delta}$ asymptotically dominates $\tilde{\delta}$, and the plug-in rule using the MLE is minmax among plug-in rules based on estimators which are asymptotically normal and unbiased. Loss A also yields the same conclusion. However, this result is severely limited in scope. For example, a conventional hypothesis testing approach might choose the treatment only if $g(x,\hat{\theta})$ is greater than a strictly positive constant c, rather than 0. In the next section, we examine the problem of finding asymptotically minmax decision rules without strong restrictions on the class of possible rules.

3.2 Limits of Experiments

In this section we use the Le Cam limits of experiments framework to examine the statistical treatment assignment problem. Although the Le Cam framework is typically applied to study point estimation and hypothesis testing, it applies much more broadly, to general statistical decision problems.

As before, we fix x, and let Θ be an open subset of $\mathbb{R}^k$. Let $\theta_0 \in \Theta$ satisfy $g(x, \theta_0) = 0$ for a given x. Assume that the sequence of experiments $\mathcal{E}_n = \{P_\theta^n, \theta \in \Theta\}$ satisfies local asymptotic normality: for all sequences $h_n \to h$ in $\mathbb{R}^k$,

$$\log \frac{dP^n_{\theta_0 + h_n/\sqrt{n}}}{dP^n_{\theta_0}} = h'\Delta_n - \frac{1}{2} h' I_{\theta_0} h + o_{P_{\theta_0}}(1),$$

where $\Delta \overset{\theta_0}{\rightsquigarrow} N(0, I_{\theta_0})$.[4] Further, assume that I_{θ_0} is nonsingular.

Then, by standard results, the experiments $\mathcal{E}_n$ converge weakly to the experiment

$$Z \sim N(h, I_{\theta_0}^{-1}).$$

By the asymptotic representation theorem, for any sequence of statistical decision rules δ_n that possesses limit distributions under every local parameter, there exists a feasible decision rule δ in the limit experiment such that $\delta_n \overset{h}{\rightsquigarrow} \delta$ for all h. In the case of statistical treatment rules, Theorem 9.4 and Corollary 9.5 of Van der Vaart (1998) can be specialized as follows:

Proposition 1 *Let Θ be an open subset of $\mathbb{R}^k$, with $\theta_0 \in \Theta$ such that $g(x, \theta_0) = 0$ for a given x, where $g(x, \theta)$ is differentiable at x, θ_0. Let the sequence of experiments $\mathcal{E}_n = \{P_\theta^n, \theta \in \Theta\}$ satisfy local asymptotic normality with nonsingular information matrix I_{θ_0}. Consider a sequence of treatment assignment rules $\delta_n(x, z_n)$ in the experiments $\mathcal{E}_n$, and let*

$$\pi_n(x, h) = E_h[\delta_n(x, Z_n)].$$

Suppose $\pi_n(x, h) \to \pi(x, h)$ for every h. Then there exists a function $\delta(x, z)$ such that

$$\begin{aligned} \pi(x, h) &= E_h[\delta(x, Z)] \\ &= \int \delta(x, z) dN(z | h, I_{\theta_0}^{-1}), \end{aligned}$$

where $dN(z|h, I_{\theta_0}^{-1})/dz$ is the pdf of a multivariate normal distribution with mean h and variance $I_{\theta_0}^{-1}$.

[4] In the case where the P_θ^n is the n-fold product measure corresponding to a random sample of size n from P_θ, then a sufficient condition for local asymptotic normality is differentiability in quadratic mean of the probability measures P_θ.

Proposition 1 shows that the simple multivariate normal shift experiment can be used to study the asymptotic behavior of treatment rules in parametric models. In particular, any asymptotic distribution of a sequence of treatment rules can be expressed as the (exact) distribution of a treatment rule in a simple Gaussian model with sample size one.

Before we consider the Gaussian limit experiment in detail, it is useful to examine the limiting behavior of the loss and risks functions, to provide heuristic guidance on the relevant forms of the loss functions in the limit experiment. Each loss function we consider can be written in the form

$$L(\delta, \theta, x) = L(0, \theta, x) + \delta[L(1, \theta, x) - L(0, \theta, x)].$$

This linearity in δ makes it possible to define the asymptotic risk functions to have essentially the same form as the original risk functions.

For Loss A, $K_0 - K_1$ loss, and an estimator sequence δ_n with $\pi_n(h, x) \to \pi(x, h)$, the associated risk function can be written in terms of the local parameter h as

$$\begin{aligned} R_n^A(\delta, h, x) &= E_h \left[L^A(\delta(Z_n), \theta_0 + \frac{h}{\sqrt{n}}, x) \right] \\ &= E_h \left[L^A(0, \theta_0 + \frac{h}{\sqrt{n}}, x) + \delta(Z_n) \left(L^A(1, \theta_0 + \frac{h}{\sqrt{n}}, x) - L^A(0, \theta_0 + \frac{h}{\sqrt{n}}, x) \right) \right] \end{aligned}$$

By differentiable continuity of g in θ,

$$\lim_{n \to \infty} 1(g(x, \theta_0 + \frac{h}{\sqrt{n}} > 0)) = 1(\dot{g}' h > 0)$$

and

$$\lim_{n \to \infty} 1(g(x, \theta_0 + \frac{h}{\sqrt{n}} < 0)) = 1(\dot{g}' h < 0).$$

The case $\dot{g}' h = 0$ presents a complication for taking limits as above, but since the loss function is bounded below by 0, we can express a lower bound on limiting risk as

$$\liminf_{n \to \infty} R_n^A(\delta, h, x) \geq K_0 \cdot 1(\dot{g}' h > 0) + \pi(x, h) \left[K_1 \cdot 1(\dot{g}' h < 0) - K_0 \cdot 1(\dot{g}' h > 0). \right]$$

This is the risk function for the same loss, but using $\dot{g}' h$ in place of $g(x, \theta_0 + h/\sqrt{n})$, and modified to be 0 when $\dot{g}' h = 0$. This suggests that analyzing the Gaussian shift limit experiment, with this modified version of the risk function, will yield asymptotic minmax bounds for the original treatment assignment problem.

For Loss C, the net welfare term $g(x, \theta_0 + \frac{h}{\sqrt{n}}) \to g(x, \theta_0) = 0$, so it is natural to renormalize the risk by a factor of $\sqrt{n}$ to keep the limiting risk nondegenerate. The behavior of the loss at

$\dot{g}'h = 0$ does not create a problem in this case, and by simple calculations we have

$$\lim_{n\to\infty} \sqrt{n} \cdot E_h \left[L^C \left(\delta(Z_n), \theta_0 + \frac{h}{\sqrt{n}}, x \right) \right] = (\dot{g}'h) \left[1(\dot{g}'h > 0) - \pi(x,h) \right].$$

The forms of the limiting risk functions are the same as the original ones, except with $g(x, \theta_0 + h/\sqrt{n})$ replaced by $\dot{g}'h$. This is intuitive: if $\dot{g}'h > 0$, then for sufficiently large n, the treatment effect $g(x, \theta_0 + h/\sqrt{n})$ will be positive, and likewise if $\dot{g}'h < 0$ the treatment effect will eventually be negative. Combined with Proposition 1, this suggests that we can study a simplified version of the original treatment assignment problem, in which the only data is a single draw from a multivariate normal distribution with unknown mean, and the treatment effect of interest is a simple linear function of the mean.

4 Minmax Treatment Rules in Gaussian Shift Experiments, and Asymptotic Minmax Rules

We have argued that the original sequence of treatment assignment problems can be approximated by an analogous treatment assignment problem in the simple Gaussian shift model. In this section, we consider the Gaussian shift experiment, and solve the minmax problem for the loss functions. This leads to a local asymptotic minmax theorem for treatment assignment rules.

Suppose that $Z \sim N(h, I_{\theta_0}^{-1}), h \in \mathbb{R}^k$. Let $\dot{g} \in \mathbb{R}^k$ satisfy $\dot{g}' I_{\theta_0}^{-1} \dot{g} > 0$. We wish to decide whether $\dot{g}'h$ is positive or negative. (Here, $\dot{g}$ corresponds to the derivative of the function $g(x, \theta_0)$ in the original sequence of experiments, but the results for the multivariate normal shift experiment simply treat it as a given constant vector.) The action space is $\mathcal{A} = \{0, 1\}$, and a randomized decision rule $\delta(\cdot)$ maps $\mathbb{R}^k$ into $[0,1]$ with the interpretation that $\delta(z) = Pr(T = 1|Z = z)$.

This situation is related to hypothesis testing problems with nuisance parameters. Here, interest centers on the scalar quantity $\dot{g}'h$. Our approach is to consider the problem along "slices" of the parameter space constructed in the following way: fix an h_0 such that $\dot{g}'h_0 = 0$, and for $b \in \mathbb{R}$, define

$$h_1(b, h_0) = h_0 + \frac{b}{\dot{g}' I_{\theta_0}^{-1} \dot{g}} I_{\theta_0}^{-1} \dot{g}.$$

In each slice, the quantity $\dot{g}'h_1 = b$ is of interest. In these one-dimensional subspaces, it is relatively easy to solve for minmax rules, and it turns out that the same rule is minmax over all the subspaces.

The following result says that rules of the form $\delta_c = 1(\dot{g}'Z > c)$, for $c \in \mathbb{R}$, form an essential complete class on each slice. For losses of the form we consider, this result simplifies the problem of finding minmax rules and bounds in the multivariate normal limit experiment, because we can limit our attention to the essential complete class on each subspace rather than have to search over all possible decision rules.

Proposition 2 *Let the loss $L(h,a)$ satisfy:*

$$[L(h,1) - L(h,0)]\,(\dot{g}'h) < 0$$

for all h such that $\dot{g}'h \neq 0$. For any randomized decision rule $\tilde{\delta}(z)$ and any fixed $h_0 \in \mathbb{R}^k$, there exists a rule of the form

$$\delta_c(z) = 1(\dot{g}'z > c)$$

which is at least as good as $\tilde{\delta}$ on the subspace $\{h_1(b,h_0) : b \in \mathbb{R}\}$.

Proof: see Appendix A. □

This result is a special case of the essential complete class theorem of Karlin and Rubin (1956), which applies to models with a scalar parameter satisfying the monotone likelihood ratio property. (See Schervish 1995, Theorem 4.68, p.244.) We present an elementary proof in Appendix A to highlight the role of the parametrization by $b = \dot{g}'h$.

We now turn to the minmax problem in the limit experiment. We want to calculate the minmax risk, and a corresponding minmax rule (in the class $\{\delta_c\}$), under the loss functions we are working with. All of the risk functions turn out to be linear in $E_h\delta$, so the following expression will play a key role in our risk computations.

$$E_h(\delta_c) = Pr_h\left(\dot{g}'Z > c\right) = Pr_h\left(\frac{\dot{g}'(Z-h)}{\sqrt{\dot{g}'I_{\theta_0}^{-1}\dot{g}}} > \frac{c - \dot{g}'h}{\sqrt{\dot{g}'I_{\theta_0}^{-1}\dot{g}}}\right) = 1 - \Phi\left(\frac{c - \dot{g}'h}{\sqrt{\dot{g}'I_{\theta_0}^{-1}\dot{g}}}\right)$$

For loss A in the limit experiment, the appropriate risk function is:

$$R^A(\delta,h,x) = 1(\dot{g}'h > 0)[1 - E_h(\delta)]K_0 + 1(\dot{g}'h < 0)E_h(\delta)K_1.$$

where $K_0, K_1 > 0$. The limit experiment risk for loss C is as described before:

$$R^C(\delta,h,x) = (\dot{g}'h)[1(\dot{g}'h > 0) - E_h(\delta)].$$

If $\dot{g} = 0$, then $R^A = R^C = 0$, so all rules are minimax and the bound is given by zero. In this case, the data are uninformative about the relative welfare of the treatments. The next result considers the more interesting case with $\dot{g} \neq 0$.

Proposition 3 *Suppose $Z \overset{h}{\sim} N(h, I_\theta^{-1})$ for $h \in \mathbb{R}^k$, and $\dot{g} \neq 0$. In each case below, the infimum is taken over all possible randomized decision rules, and δ^* denotes a rule which attains the given bound:*

(A) For Loss A,

$$\inf_{\delta} \sup_{h} R^A(\delta(Z,x),h,x) = \frac{K_0 K_1}{K_0+K_1},$$

$$\delta^* 1(\dot{g}'Z > c^*), \qquad c^* = \sqrt{\dot{g}' I_{\theta_0}^{-1} \dot{g}}\, \Phi^{-1}\left(\frac{K_1}{K_0+K_1}\right);$$

(C) For Loss C,

$$\inf_{\delta} \sup_{h} R^C(\delta(Z,x),h,x) = \tau^* \Phi(\tau^*) \sqrt{\dot{g}' I_{\theta_0}^{-1} \dot{g}},$$

$$\tau^* = \arg\max_{\tau} \tau \Phi(-\tau),$$

$$\delta^* = 1(\dot{g}'Z > 0).$$

Proof: see Appendix A. □

Loss A is well known from hypothesis testing theory, and the same bound is derived in the scalar normal case in Berger (1985), Section 5.3.2, Example 14. Our proof follows Berger's analysis along one-dimensional subspaces, showing that the same rule is optimal for each subspace, and then argues that the rule is optimal over the entire parameter space. The minmax result for Loss C appears to be new. As we noted earlier, the corresponding minmax analysis for Loss B (see Appendix B) leads to excessively conservative rules; thus we focus on the minmax regret approach, which corresponds to Loss C.

Since the multivariate shift model provides an asymptotic version of the original problem, in the sense that any sequence of decision rules in the original problem with limit distributions is matched by a decision rule in the limit experiment, an application of the Portmanteau lemma allows us to use the exact bounds developed in Proposition 3 as asymptotic bounds in the original problem. This leads to the following theorem, which is our main result for smooth parametric models:

Theorem 1 *Assume the conditions of Proposition 1, and suppose that δ_n is any sequence of treatment assignment rules that converge to limit distributions under $\theta_0 + \frac{h}{\sqrt{n}}$ for every $h \in \mathbb{R}^k$. Then, for Loss A,*

$$\liminf_{n\to\infty} \sup_{h\in\mathbb{R}^k} E_h\left[L^A(\delta_n(Z_n), \theta_0 + \frac{h}{\sqrt{n}}, x)\right] \geq \frac{K_0 K_1}{K_0+K_1},$$

and the bound is attained by the decision rules

$$\delta_n^* = 1(g(x,\hat{\theta}) > c^*), \qquad c^* = \left(\sqrt{\dot{g}' I_{\theta_0}^{-1} \dot{g}}\right) \Phi^{-1}\left(\frac{K_1}{K_0+K_1}\right),$$

where $\hat{\theta}$ is an estimator sequence that satisfies $\sqrt{n}(\hat{\theta} - \theta_0 - \frac{h}{\sqrt{n}}) \overset{h}{\rightsquigarrow} N(0, I_{\theta_0}^{-1})$ for every h.

For Loss C,

$$\liminf_{n\to\infty} \sup_{h\in\mathbb{R}^k} \sqrt{n}\cdot E_n\left[L^C\left(\delta_n(Z_n), \theta_0 + \frac{h}{\sqrt{n}}, x\right)\right] \geq \tau^*\Phi(\tau^*)\sqrt{\dot{g}' I_{\theta_0}^{-1}\dot{g}},$$

where $\tau^* = \arg\max_\tau \tau\Phi(-\tau)$. *The bound is attained by the rule* $\delta_n^* = 1(g(x,\hat{\theta}) > 0)$

Proof: see Appendix A. □

For Loss A, note that if $K_0 = K_1$, then $c^* = 0$ so that the optimal rule is the same as for Loss C. In particular, plugging in the maximum likelihood estimator (or any other efficient estimator, such as the Bayes estimator), leads to an optimal rule in this local asymptotic minmax risk sense. Although perhaps not surprising, this rule is distinct from the usual hypothesis testing approach, which would require that the estimated net effect be above some strictly positive cutoff determined by the level specified for the test.

5 Semiparametric Models

Next, we extend the results from the previous section to models with an infinite-dimensional parameter space. We use the local score representation of a general semiparametric model, as described in Van der Vaart (1991a). The limit experiment associated with this model is a Gaussian process experiment. As in the previous section, we argue along "slices" of the limiting version of the statistical decision problem, to obtain complete-class results and risk bounds.

Suppose Z_n consists of an i.i.d. sample of size n drawn from a probability measure $P \in \mathcal{P}$, where $\mathcal{P}$ is the set of probability measures defined by the underlying semiparametric model.[5] In some cases the set $\mathcal{P}$ will include all distributions satisfying certain weak conditions (so that the model is nonparametric); in other cases the form of the semiparametric model may restrict the feasible distributions in $\mathcal{P}$.

We fix $P \in \mathcal{P}$, and, following Van der Vaart (1991a), define a set of paths on $\mathcal{P}$ as follows. For a measurable real function h, suppose $P_{t,h} \in \mathcal{P}$ satisfies the differentiability in quadratic mean condition,

$$\int \left[\frac{1}{t}\left(dP_{t,h}^{1/2} - dP^{1/2}\right) - \frac{1}{2} h dP^{1/2}\right]^2 \longrightarrow 0 \quad \text{as } t \downarrow 0 \tag{1}$$

for $t \in (0, \eta)$, $\eta > 0$. Let $\mathcal{P}(P)$ denote the set of maps $t \to P_{t,h}$ satisfying (1). These maps are called paths and represent one-dimensional parametric submodels for P in $\mathcal{P}$. The functions h provide a parametrization for the set of probability measures we consider. This parametrization

[5]The i.i.d. assumption can be weakened; what is essential is that the limiting log-likelihood ratio process have the same limiting form.

is particularly convenient since, from (1), we can regard h as the score function for the submodel $\{P_{t,h} : t \in (0,\eta)\}$. Note that (1) implies $\int h dP = 0$ and $\int h^2 dP < \infty$. Hence, $h \in L_2(P)$, the Hilbert space of square-integrable functions with respect to P.[6] Let $\langle \cdot, \cdot \rangle$ and $\| \cdot \|$ denote the usual inner product and norm on this space. Then the Fisher information for the submodel $\{P_{t,h} : t \in (0,\eta)\}$ is given by $\|h\|^2$. Let $T(P) \subset L_2(P)$ denote the set of functions h satisfying (1). $T(P)$ is the tangent space which we will assume is a cone.

The limit experiment for this set of probability measures that localize P in the sense of (1) takes on a convenient form. Let $\tilde{h}_1, \tilde{h}_2, \ldots$ denote an orthonormal basis of a subspace of $L_2(P)$ that contains the closure of $\text{lin}(T(P))$. Any $h \in T(P)$ then satisfies $h = \sum_{j=1}^{\infty} \langle h, \tilde{h}_j \rangle \tilde{h}_j$. Consider a sequence of independent, normally distributed random variables $\Delta_1, \Delta_2, \ldots$ with $\Delta_j \sim N(\langle h, \tilde{h}_j \rangle, 1)$ under $h \in T(P)$. We can now construct the stochastic process $\{\Delta_h = \sum_{j=1}^{\infty} \langle h, \tilde{h}_j \rangle \Delta_j \ : \ h \in T(P)\}$, which under $h = 0$ has mean zero and covariance function $E\Delta_h \Delta_{h'} = \langle h, h' \rangle$. Let Q_h denote the law of this process under h. Then (1) implies

$$\ln \frac{dP_{1/\sqrt{n},h}}{dP} \overset{P}{\rightsquigarrow} \ln \frac{dQ_h}{dQ_0} = \Delta_h - \frac{1}{2}\|h\|^2.$$

It follows that the limit experiment corresponding to the semiparametric model consists of observing the sequence $(\Delta_1, \Delta_2, \ldots)$ distributed Q_h under $h \in T(P)$.

Again, we use g to denote the difference in social welfare $W_1(x) - W_0(x)$. For a probability measure $P_{t,h}$, we denote this welfare contrast by $g(x, P_{t,h})$. We assume functional differentiability of g: there exists a continuous linear map $\dot{g} : T(P) \to \mathbb{R}$ such that

$$\frac{1}{t}(g(x, P_{t,h}) - g(x,P)) \longrightarrow \dot{g}(h) \quad \text{as } t \downarrow 0 \tag{2}$$

for every path in $\mathcal{P}(P)$.[7] An immediate implication of this differentiability notion is that

$$\sqrt{n}(g(x, P_{1/\sqrt{n},h}) - g(x,P)) \to \dot{g}(h).$$

By the Riesz representation theorem, the functional $\dot{g}(\cdot)$ can be associated with an element $\dot{g} \in L_2(P)$ such that $\dot{g}(h) = \langle \dot{g}, h \rangle$ for all $h \in T(P)$. Assume $\|\dot{g}\|^2 = \langle \dot{g}, \dot{g} \rangle > 0$.

Note that $\Delta_{\dot{g}} = \sum_{j=1}^{\infty} \langle \dot{g}, \tilde{h}_j \rangle \Delta_j$. Assuming $\dot{g}$ is continuous, Van der Vaart (1989) shows that $\Delta_{\dot{g}}$ is an efficient estimator for $\dot{g}(h)$ in the limit experiment. From $\Delta_{\dot{g}} \sim N(0, \|\dot{g}\|^2)$ under $h = 0$, it follows that $\|\dot{g}\|^2$ provides the semiparametric efficiency bound for estimation of $g(P)$.

Given the limit experiment and the functional g, an analog to Proposition 1 follows from the same results in Van der Vaart (1998).

---

[6] Formally the h defined in (1) are elements of $\mathcal{L}_2(P)$. We work with equivalence classes of such functions with respect to the L_2 norm, so we can consider $h \in L_2(P)$. .

[7] Van der Vaart (1991b) contains a complete description of this differentiability notion, which is related to Hadamard differentiability.

Proposition 1′ *Suppose that $g(x, P) = 0$ for a given x, where g satisfies (2). Let the sequence of experiments $\mathcal{E}_n = \{P_{1/\sqrt{n},h} : h \in T(P)\}$ satisfy (1). Consider a sequence of treatment rules $\delta_n(x, z_n)$ in the experiments $\mathcal{E}_n$, and let $\pi_n(x, h) = E_h[\delta_n(x, Z_n)]$. Suppose $\pi_n(x, h) \longrightarrow \pi(x, h)$ for every h. Then there exists a function δ such that $\pi(x, h) = E_h[\delta(x, \Delta_1, \Delta_2, \ldots)]$ for $(\Delta_1, \Delta_2, \ldots)$ as defined above.*

For our statistical treatment rule problem, the loss functions considered will be the same as in the parametric case, with $g(x, P_{t,h})$ replacing $g(x, \theta)$. The limiting risk functions correspond exactly to the previous expressions. For δ_n and δ as in Proposition 1′,

$$\begin{aligned}\liminf_{n\to\infty} R_n^A(\delta, h, x) &= \liminf_{n\to\infty} E_h\left[L^A(\delta_n(x, Z_n), P_{1/\sqrt{n},h}, x)\right] \\ &\geq K_0 \cdot 1(\langle \dot{g}, h\rangle > 0) + \pi(x, h)\left[K_1 \cdot 1(\langle \dot{g}, h\rangle < 0) - K_0 \cdot 1(\langle \dot{g}, h\rangle > 0)\right] \\ &\quad (= R^A(\delta, h, x)),\end{aligned}$$

$$R^C(\delta, h, x) = \lim_{n\to\infty} \sqrt{n} \cdot E_h\left[L^C\left(\delta_n(x, Z_n), P_{1/\sqrt{n},h}, x\right)\right] = \langle \dot{g}, h\rangle \left[1(\langle \dot{g}, h\rangle > 0) - \pi(x, h)\right].$$

Next, we develop an analog of Proposition 2, the essential complete class theorem along slices, for the semiparametric limit experiment. Take $h_0 \in T(P)$ such that $\langle \dot{g}, h_0\rangle = 0$, and for $b \in \mathbb{R}$ let

$$h_1(b, h_0) = h_0 + \frac{b}{\|\dot{g}\|^2}\dot{g}.$$

Proposition 2′ *Let the loss $L(a, P_{t,h})$ satisfy:*

$$[L(1, P_{t,h}) - L(0, P_{t,h})]\langle \dot{g}, h\rangle < 0$$

for all h such that $\langle \dot{g}, h\rangle \neq 0$. For any randomized decision rule $\tilde{\delta}(\Delta_1, \Delta_2, \ldots)$ and any fixed $h_0 \in T(P)$, there exists a rule of the form

$$\delta_c(\Delta_1, \Delta_2, \ldots) = 1(\Delta_{\dot{g}} > c)$$

which is at least as good as $\tilde{\delta}$ on the subspace $\{h_1(b, h_0) : b \in \mathbb{R}\}$.

Proof: see Appendix A. □

Similar to the parametric case, Proposition 2′ can be used to obtain risk bounds for treatment assignment rules:

Proposition 3′ *Consider treatment rules on $(\Delta_1, \Delta_2, \ldots)$, as defined above, for $h \in T(P)$, and assume $\|\dot{g}\|^2 > 0$. In each case below, the infimum is taken over all possible randomized decision rules, and δ^* denotes a rule which attains the given bound:*

(A) For Loss A,

$$\inf_\delta \sup_h R^A(\delta, h, x) = \frac{K_0 K_1}{K_0 + K_1},$$

$$\delta^* 1(\Delta_{\dot{g}} > c^*), \qquad c^* = \sqrt{\|\dot{g}\|^2}\Phi^{-1}\left(\frac{K_1}{K_0 + K_1}\right);$$

(C) For Loss C,

$$\inf_\delta \sup_h R^C(\delta, h, x) = \tau^*\Phi(\tau^*)\sqrt{\|\dot{g}\|^2},$$

$$\tau^* = \arg\max_\tau \tau\Phi(-\tau),$$

$$\delta^* = 1(\Delta_{\dot{g}} > 0).$$

Proof: These bounds follow by the same argument as given for Proposition 3. □

Finally, using this result, we can characterize the asymptotically optimal treatment assignment rules in the semiparametric model:

Theorem 1′ *Assume the conditions of Proposition 1′, and let δ_n be any sequence of treatment assignment rules that converge to limit distributions under $P_{1/\sqrt{n},h}$ for every $h \in T(P)$. Suppose that the estimator sequence $\hat{g}_n(Z_n)$ attains the semiparametric efficiency bound for estimating $g(P)$.*

Then, for Loss A,

$$\liminf_{n\to\infty} \sup_{h\in T(P)} E_h\left[L^A(\delta_n(x, Z_n), P_{1/\sqrt{n},h}, x)\right] \geq \frac{K_0 K_1}{K_0 + K_1},$$

and the bound is attained by the decision rules

$$\delta_n^* = 1(\hat{g}_n(Z_n) > c^*), \qquad c^* = \left(\sqrt{\|\dot{g}\|^2}\right)\Phi^{-1}\left(\frac{K_1}{K_0 + K_1}\right).$$

For Loss C,

$$\liminf_{n\to\infty} \sup_{h\in T(P)} \sqrt{n}\cdot E_h\left[L^C\left(\delta_n(x, Z_n), P_{1/\sqrt{n},h}, x\right)\right] \geq \tau^*\Phi(\tau^*)\sqrt{\|\dot{g}\|^2},$$

where $\tau^* = \arg\max_\tau \tau\Phi(-\tau)$. The bound is attained by the rule $\delta_n^* = 1(\hat{g}_n(Z_n) > 0)$.

Proof: By the same argument as for Theorem 1. □

Thus, a plug-in rule based on a semiparametrically efficient estimator is optimal. This implies that the conditional empirical success rules studied by Manski (2004) are asymptotically optimal among all rules when the distribution of outcomes is essentially unrestricted:

Example 2 *(Conditional Empirical Success Rules)*

Suppose that $W_0(x) = 0$, and that we observe a random sample (X_i, Y_i), $i = 1, \ldots, n$, where X_i has a finitely supported distribution and $Y_i|X_i$ has conditional distribution $F_1(y|x)$. The social welfare contrast is the functional

$$g(x, F_1) = \int w(y) dF_1(y|x).$$

The conditional distribution function F_1 is unknown, and the set of possible CDFs $\mathcal{P}$ is the largest set satisfying

$$\sup_{F_1 \in \mathcal{P}} E[|w(Y)|^2 | X = x] < \infty.$$

The conditional empirical success rule of Manski (2004) can be expressed as

$$\hat{\delta}_n(x) = 1(\hat{g}_n(x) > 0),$$

where

$$\hat{g}_n(x) := \frac{\sum_{i=1}^n w(Y_i) \cdot 1(X_i = x)}{\sum_{i=1}^n 1(X_i = x)}.$$

The estimator $\hat{g}_n(x)$ is an asymptotically efficient estimator of $g(x, F_1)$ (Bickel, Klaasen, Ritov, and Wellner (1993), pp. 67-68). Therefore, $\hat{\delta}_n$ is asymptotically minmax for regret loss C.

This result extends easily to the case where $W_0(x)$ is not known; then $\hat{g}_n(x)$ would be a difference of conditional mean estimates for outcomes under treatments 1 and 0.

□

6 Conclusion

Le Cam's asymptotic statistical decision theory framework proposes to examine "limiting" versions of the original statistical decision problem, obtain optimal decision rules in the limit, and then find a sequence of decision rules in the original sequence of problems that matches the optimal rule asymptotically. This general approach has been used primarily for obtaining optimality results for point estimation and hypothesis testing, but it can be applied to a much wider range of statistical decision problems, such as the treatment assignment problem considered here.

We have examined asymptotic properties of treatment assignment rules in parametric and regular semiparametric settings, under standard regularity conditions. The limiting version of the decision problem is a treatment assignment problem involving a single observation from a Gaussian shift model. Using simple extensions of classic results from the theory of one-sided tests, it turns out to be possible to exactly solve for minmax rules in this simple setting. This leads to local asymptotic minmax bounds on risk in the original sequence of models. Our sharpest results are for the social welfare regret loss (loss C), which has been emphasized by Manski (2004). We show that a plug-in rule that uses an efficient estimator of the treatment effect, is locally asymptotically minmax. This rule is intuitive, but does lead to less conservative treatment assignment than typical applications of hypothesis testing, which would suggest to apply the treatment only if an estimator of the treatment effect was above some strictly positive cutoff.

Appendix A: Proofs

Proof of Proposition 2:
The result can be obtained as a special case of Karlin and Rubin, Theorem 1, but we present a simple proof that highlights the role of our restriction to subspaces $h \in \{h_1(b, h_0) : b \in \mathbb{R}\}$.

For a decision rule δ, the risk function is

$$\begin{aligned} R(h,\delta) &= \int [\delta(z)L(h,1) - (1-\delta(z))L(h,0)]\, f(z|h)dz \\ &= L(h,0) - [L(h,1) - L(h,0)] \int \delta(z) f(z|h)dz \end{aligned}$$

So, for any two rules δ_1, δ_2,

$$R(h,\delta_1) - R(h,\delta_2) = [L(h,1) - L(h,0)] \int [\delta_1(z) - \delta_2(z)]\, f(z|h)dz \quad (3)$$
$$= [L(h,1) - L(h,0)] \{E_h[\delta_1(Z)] - E_h[\delta_2(Z)]\}\,. \quad (4)$$

Thus the quantity $E_h[\delta(Z)]$ is key in comparing risks.

Note that if $\dot{g}'h_0 \neq 0$, then for $\tilde{h_0} = h_1(-\dot{g}'h_0, h_0)$, $\dot{g}'\tilde{h_0} = 0$. Since $\{h_1(b, h_0) : b \in \mathbb{R}\}$ $= \{h_1(b, \tilde{h}_0) : b \in \mathbb{R}\}$, we may assume without loss of generality that, in fact, $\dot{g}'h_0 = 0$.

Let $\tilde{\delta}$ be an arbitrary treatment assignment rule, and let c satisfy

$$E_{h_0}[\delta_c(Z)] = E_{h_0}[\tilde{\delta}(Z)].$$

Note that $\dot{g}'Z \sim N(0, \dot{g}'I_{\theta_0}^{-1}\dot{g})$ under h_0, so

$$\begin{aligned} E_{h_0}[\delta_c(Z)] &= \Pr_{h_0}(\dot{g}'Z > c) \\ &= 1 - \Phi\left(\frac{c}{\sqrt{\dot{g}'I_{\theta_0}^{-1}\dot{g}}}\right). \end{aligned}$$

It is easy to see that for any $\tilde{\delta}$, we can choose a c to satisfy the requirement above.

This part of the proof follows the method in the proof of van der Vaart, Proposition 15.2. Take some $b > 0$ and consider the test $H_0 : h = h_0$ against $H_1 : h = h_1(b, h_0)$ based on $Z \overset{h}{\sim} N(h, I_{\theta_0}^{-1})$. Note that $\dot{g}'h_1 = b > 0$. The likelihood ratio is:

$$LR = \frac{dN(h_1, I_{\theta_0}^{-1})}{dN(h_0, I_{\theta_0}^{-1})} = \exp\left(\frac{b}{\dot{g}'I_{\theta_0}^{-1}\dot{g}}\dot{g}'Z - \frac{b^2}{2\dot{g}'I_{\theta_0}^{-1}\dot{g}}\right).$$

By the Neyman-Pearson lemma, a most powerful test is based on rejecting for large values of $\dot{g}'Z$. Since the test δ_c has been defined to have the same size as $\tilde{\delta}$, $E_{h_1(b,h_0)}[\delta_c(Z)] \geq E_{h_1(b,h_0)}[\tilde{\delta}(Z)]$. This argument does not depend on which $b > 0$ is considered, so $E_{h_1(b,h_0)}[\delta_c(Z)] \geq E_{h_1(b,h_0)}[\tilde{\delta}(Z)]$ for all $b \geq 0$ (δ_c is more powerful than $\tilde{\delta}$ for $H_0 : h = h_0$ against $H_1 : h = h_1(b, h_0), b > 0$).

Next consider the case that $b < 0$. Note that $1 - \delta_c = 1(\dot{g}'Z \leq 0)$ is uniformly most powerful against $1 - \tilde{\delta}$ for $H_0 : h = h_0$ against $H_1 : h = h_1(b, h_0), b < 0$ by an analogous argument. Hence, $E_{h_1(b,h_0)}[1 - \delta_c(Z)] \geq E_{h_1(b,h_0)}[1 - \tilde{\delta}(Z)]$ for all $b \leq 0$. So $E_{h_1(b,h_0)}[\delta_c(Z)] \leq E_{h_1(b,h_0)}[\tilde{\delta}(Z)]$ for all $b \leq 0$.

By equation (3) and the assumptions on loss, it therefore follows that

$$R(h, \tilde{\delta}) \geq R(h, \delta_c)$$

for all $h \in \{h_1(b, h_0) : b \in \mathbb{R}\}$. $\square$

Proof of Proposition 3:

First, we will show that for $R = R^A$ or R^C, $R(\delta_c, h_1(b, h_0), x)$ does not depend on h_0. Note that from the definition of $h_1(b, h_0)$, $h_1(b, h_0) = h_1(b, 0) + h_0$. Since $\dot{g}'h_0 = 0$, $\dot{g}'h_1(b, h_0) \ = \dot{g}'h_1(b, 0)$ ($= b$). Further, $E_{h_1(b,h_0)}[\dot{g}'Z] = b \ = E_{h_1(b,0)}[\dot{g}'Z]$. It follows that under $h_1(b, h_0)$, $\dot{g}'Z \sim N(b, \dot{g}'I_{\theta_0}^{-1}\dot{g})$. That is, the distribution of $\dot{g}'Z$ under $h_1(b, h_0)$ does not depend on h_0. For $R = R^A$ or R^C, $R(\delta_c, h, x)$ depends on h only through two terms: $\dot{g}'h$ and $E_h(\delta_c) = \Pr_h(\dot{g}'Z > c)$. It follows then, that for any c, b, and x, $R(\delta_c, h_1(b, h_0), x) \ = R(\delta_c, h_1(b, 0), x)$.

Again let $R = R^A$ or R^C. Define δ_c^* as the solution to $\inf_c \sup_b R(\delta_c, h_1(b, 0), x)$. Below we will show that such a solution exists for each risk function. Now we have

$$\begin{aligned}
&\inf_c \sup_b R(\delta_c, h_1(b, 0), x) = \sup_b R(\delta_c^*, h_1(b, 0), x) = \sup_{h_0} \sup_b R(\delta_c^*, h_1(b, h_0), x) \\
&\quad \geq \ \inf_\delta \sup_{h_0} \sup_b R(\delta, h_1(b, h_0), x) \quad (\ = \inf_\delta \sup_h R(\delta, h, x) \ ) \\
&\quad \geq \ \inf_\delta \sup_b R(\delta, h_1(b, 0), x) = \inf_c \sup_b R(\delta_c, h_1(b, 0), x).
\end{aligned}$$

The first equality holds by the definition of δ_c^* and the second by the lack of dependence of $R(\delta_c, h_1(b, h_0), x)$ on h_0. The two inequalities follow by infimum properties. The final equality follows by Proposition 2. From these inequalities, $\inf_\delta \sup_h R(\delta, h, x) \ = \inf_c \sup_b R(\delta_c, h_1(b, 0), x)$, so it suffices to consider the latter term in the following computations.

(A) Take a rule δ_c, we want $\sup_b R^A(\delta_c, h_1(b, 0), x)$. Recall that $\dot{g}'h_1(b, 0) = b$.

$$\sup_{b:b>0} R^A(\delta_c, h_1(b, 0), x) = \sup_{h:b>0} [1 - E_{h_1(b,0)}(\delta_c)]K_0 = \sup_{b:b>0} K_0 \Phi \left(\frac{c - b}{\sqrt{\dot{g}'I_{\theta_0}^{-1}\dot{g}}} \right) = K_0 \Phi \left(\frac{c}{\sqrt{\dot{g}'I_{\theta_0}^{-1}\dot{g}}} \right)$$

$$\sup_{b:b<0} R^A(\delta_c, h_1(b,0), x) = \sup_{b:b<0} E_{h_1(b,0)}(\delta_c) K_1 = \sup_{b:b<0} K_1 \left[1 - \Phi\left(\frac{c-b}{\sqrt{\dot{g}' I_{\theta_0}^{-1} \dot{g}}}\right)\right] = K_1 \left[1 - \Phi\left(\frac{c}{\sqrt{\dot{g}' I_{\theta_0}^{-1} \dot{g}}}\right)\right]$$

For $b = 0$, $R^A(\delta_c, h_1(b,0), x) = 0$. Hence,

$$\sup_b R^A(\delta_c, h_1(b,0), x) = \max \left\{ K_0 \Phi\left(\frac{c}{\sqrt{\dot{g}' I_{\theta_0}^{-1} \dot{g}}}\right), K_1 \left[1 - \Phi\left(\frac{c}{\sqrt{\dot{g}' I_{\theta_0}^{-1} \dot{g}}}\right)\right] \right\}.$$

Then, $\inf_c \sup_b R^A(\delta_c, h_1(b,0), x)$ occurs when c is chosen to set $\Phi(c/\sqrt{\dot{g}' I_{\theta_0}^{-1} \dot{g}}) = K_1/(K_0 + K_1)$. Plugging in this minmax rule, the minmax value in the conclusion follows.

(C) Consider δ_c with $c \geq 0$. Fix b such that $b > 0$. Then

$$\begin{aligned} R^C(\delta_c, h_1(b,0), x) &= b\Phi\left(\frac{c-b}{\sqrt{\dot{g}' I_{\theta_0}^{-1} \dot{g}}}\right) \\ &\geq b\Phi\left(\frac{-c-b}{\sqrt{\dot{g}' I_{\theta_0}^{-1} \dot{g}}}\right) = b\left[1 - \Phi\left(\frac{c+b}{\sqrt{\dot{g}' I_{\theta_0}^{-1} \dot{g}}}\right)\right] = R^C(\delta_c, h_1(-b,0), x). \end{aligned}$$

So, $\sup_b R^C(\delta_c, h_1(b,0), x) = \sup_{b:b\geq 0} R^C(\delta_c, h_1(b,0), x)$.

Further, take $c > 0$ and any $b > 0$,

$$R^C(\delta_0, h_1(b,0), x) = b\Phi\left(\frac{-b}{\sqrt{\dot{g}' I_{\theta_0}^{-1} \dot{g}}}\right) < b\Phi\left(\frac{c-b}{\sqrt{\dot{g}' I_{\theta_0}^{-1} \dot{g}}}\right) = R^C(\delta_c, h_1(b,0), x)$$

Also, $R^C(\delta_0, 0, x) = 0 = R^C(\delta_c, 0, x)$ (corresponding to $b = 0$). So,

$$\sup_b R^C(\delta_0, h_1(b,0), x) = \sup_{b:b\geq 0} R^C(\delta_0, h_1(b,0), x) \leq \sup_{b:b>0} R^C(\delta_c, h_1(b,0), x) = \sup_b R^C(\delta_c, h_1(b,0), x),$$

which shows that δ_0 is minmax over all rules δ_c with $c \geq 0$. The analogous argument for $c \leq 0$ yields δ_0 as the minmax rule.

□

Proof of Theorem 1:

Consider Loss A. Let Q_h denote the limit distributions of δ_n under h. Note that L^A is lower semicontinuous, so by the Portmanteau Lemma (e.g. van der Vaart and Wellner, 1996, Theorem 1.3.4),

$$\liminf_{n\to\infty} \sup_{h\in\mathbb{R}^k} E_h \left[L^A\left(\delta_n(Z_n), \theta_0 + \frac{h}{\sqrt{n}}, x\right)\right] \geq \sup_{h\in\mathbb{R}^k} \int L(a,h,x) dQ_h(a).$$

By Proposition 1, Q_h is the distribution of a randomized decision rule in the $N(h, I_{\theta_0}^{-1})$ limit experiment. Hence the right hand side of the previous expression is bounded below by

$$\inf_T \sup_{h \in \mathbb{R}^k} \int L^A\left(T(z), h, x\right) dN(z|h, I_{\theta_0}^{-1}),$$

where the infimum is taken over all randomized decision rules T. Proposition 3 provides a lower bound on the maximum risk of a slightly modified version of Loss A, call it $\tilde{L}^A$, which satisfies $L^A \geq \tilde{L}^A$. Thus the previous expression is greater than

$$\inf_T \sup_{h \in \mathbb{R}^k} \int \tilde{L}^A\left(T(z), h, x\right) dN(z|h, I_{\theta_0}^{-1}) = \frac{K_0 K_1}{K_0 + K_1},$$

giving the local asymptotic minmax bound as stated. It is straightforward to verify that the rule $\delta_n^* = 1(g(x, \hat{\theta}) > c^*)$ achieves this bound.

For Loss C, consider

$$\liminf_{n \to \infty} \sup_{h \in \mathbb{R}^k} \sqrt{n} \cdot E_n\left[L^C\left(\delta_n(Z_n), \theta_0 + \frac{h}{\sqrt{n}}, x\right)\right].$$

Note that

$$\liminf_{n \to \infty} \sqrt{n} L^C(a_n, \theta_0 + \frac{h}{\sqrt{n}}, x) \geq L^C(a, h, x)$$

for every sequence $a_n \to a$. We can use this fact, and modify the proof of the Portmanteau lemma to obtain that

$$\liminf_{n \to \infty} \sup_{h \in \mathbb{R}^k} \sqrt{n} \cdot E_n\left[L^C\left(\delta_n(Z_n), \theta_0 + \frac{h}{\sqrt{n}}, x\right)\right] \geq \sup_{h \in \mathbb{R}^k} \int L^C\left(a, h, x\right) dQ_h(a).$$

The remainder of the proof is analogous to the case for Loss A.

□

Proof of Proposition 2′:

The proof follows the proof of Proposition 2.

Since $\Delta_{\dot{g}} \sim N(0, \|\dot{g}\|^2)$ under h_0, we can find c satisfying

$$E_{h_0}[\delta_c] = E_{h_0}[\tilde{\delta}].$$

As before, we use the Neyman-Pearson lemma to derive a most powerful test of $H_0 : h = h_0$ against

$H_1 : h = h_1(b, h_0)$ for some $b > 0$. A most powerful test rejects for large values of

$$\ln \frac{dQ_{h_1}}{dQ_{h_0}} = \Delta_{(h_1 - h_0)} - \frac{1}{2}\|h_1\|^2 + \frac{1}{2}\|h_0\|^2 = \left(\frac{b}{\|\dot{g}\|^2}\right)\Delta_{\dot{g}} - \frac{1}{2}\|h_1\|^2 + \frac{1}{2}\|h_0\|^2.$$

The last equality follows by

$$\Delta_{(h_1 - h_0)} = \frac{b}{\|\dot{g}\|^2}\sum_{j=1}\langle \dot{g}, \tilde{h}_j\rangle \Delta_j = \left(\frac{b}{\|\dot{g}\|^2}\right)\Delta_{\dot{g}}.$$

The rest of the proof then follows as in the proof of Proposition 2. □

Appendix B: Loss B

Recall that $L^B(\delta, \theta, x) = -w_0(x, \theta) - \delta g(x, \theta)$. As we did for Loss A and C, fix θ_0 such that $w_0(x, \theta_0) = g(x, \theta_0) = 0$ and scale Loss B by $\sqrt{n}$ to obtain the limiting loss and risk functions. Suppose $\pi(x, h) = \lim_{n\longrightarrow\infty} E_h\delta_n(x, Z_n)$, and denote $\dot{w}_0 := \partial w_0(x, \theta_0)/\partial\theta$. Then,

$$\lim_{n\longrightarrow\infty} \sqrt{n}E_h\left[L_B\left(\delta_n(x, Z_n), \theta_0 + \frac{h}{\sqrt{n}}, x\right)\right] = -\dot{w}_0' h - \pi(x, h)\ \dot{g}' h.$$

By Proposition 1, there is a rule δ in the limit experiment such that $\pi(x, h) = E_h\delta(x, Z)$, and the limiting risk takes on the following form:

$$R^B(\delta, h, x) = -\dot{w}_0' h - (E_h\delta)\ \dot{g}' h.$$

Now we state and prove the loss B result under the conditions of Proposition 3.

Proposition B3 *Suppose $Z \overset{h}{\sim} N(h, I_\theta^{-1})$ for $h \in \mathbb{R}^k$, and $\dot{g} \neq 0$. The infimum is taken over all possible randomized decision rules, and δ^* denotes a rule which attains the given bound.*
For Loss B,

(i) $\dot{w}_0 = a\dot{g}$ for some $a \in [-1, 0]$,

$$\inf_\delta \sup_h R^B(\delta(Z, x), h, x) = 0,$$

where δ^* is any rule with $E\delta^* = -a$;

(ii) $\dot{w}_0 \neq a\dot{g}$ for any $a \in [-1, 0]$,

$$\inf_\delta \sup_h R^B(\delta(Z, x), h, x) = \infty,$$

and all rules are minmax.

Proof: Let $H_0 = \{h : \dot{g}'h = 0\}$ and $H_0^{\perp} = \{v : v'h_0 = 0 \text{ for all } h_0 \in H_0\}$. It is straightforward to show that $H_0^{\perp} = \{a\dot{g} : a \in \mathbb{R}\}$. For each h, there exists a unique $h_0 \in H_0$ and $b \in \mathbb{R}$ such that $h = h_1(b, h_0)$. Suppose $\dot{w}_0 = a\dot{g}$ for some $a \in \mathbb{R}$. Then,

$$R^B(\delta, h, x) = R^B(\delta, h_1(b, h_0), x) = -(a\dot{g}'h_0) - b\frac{a\dot{g}'I_{\theta_0}^{-1}\dot{g}}{\dot{g}'I_{\theta_0}^{-1}\dot{g}} - (E\delta)b = -b(a + E\delta)$$

Case (i): $\dot{w}_0 = a\dot{g}$ for some $a \in [-1, 0]$.
Take δ^* such that $E\delta^* = -a$. Then $R^B(h, \delta^*) = 0$. Consider any δ with $E\delta \neq -a$. Then $a + E\delta \neq 0$ and for some b, $R^B(h_1(b, h_0), \delta) = -b(a + E\delta) > 0$. Hence, all such δ^* rules are minmax.

Case (ii)(a): $\dot{w}_0 = a\dot{g}$ for some $a \notin [-1, 0]$.
Then for all δ, $a + E\delta \neq 0$, and as $b \longrightarrow \infty$ or $-\infty$, $-b(a + E\delta) \longrightarrow \infty$. So, $\sup_h R^B(h, \delta) = \infty$ for all δ and all rules are minmax.
Case (ii)(b): $\dot{w}_0 \neq a\dot{g}$ for any $a \in \mathbb{R}$, ie $\dot{w}_0 \notin H_0^{\perp}$.
There exists $\tilde{h}_0 \in H_0$ such that $-\dot{w}_0'\tilde{h}_0 > 0$. Note that $a\tilde{h}_0 \in H_0$ for $a \in \mathbb{R}$. Also $-\dot{w}_0'(a\tilde{h}_0) \longrightarrow \infty$ as $a \longrightarrow \infty$.

$$\sup_h R^B(h, \delta) \geq \sup_{h_0 \in H_0} R^B(h_1(b = 0, h_0), \delta) \geq \sup_{a \in \mathbb{R}} R^B(h_1(b = 0, a\tilde{h}_0), \delta) = \sup_{a \in \mathbb{R}} -\dot{w}_0'(a\tilde{h}_0) = \infty$$

So the maximum risk of all rules is infinite, and all rules are minmax.

□

The extension of Theorem 1 to Loss B follows from the above proposition with the same bounds and optimal rules for each case. The proof follows the proof for Loss C. These parametric results for Loss B also carry over to the semiparametric case with the same bounds and optimal rules. This extension is straightforward.

References

Berger, J. O. (1985): *Statistical Decision Theory and Bayesian Analysis*. Springer-Verlag, New York.

Berger, M. C., D. Black, and J. Smith (2001): "Evaluating Profiling as a Means of Allocating Government Services," in *Econometric Evaluation of Labour Market Policies*, ed. by M. Lechner, and F. Pfeiffer, pp. 59–84. Physica Heidelberg.

Bickel, P. J., C. A. Klaasen, Y. Ritov, and J. A. Wellner (1993): *Efficient and Adaptive Estimation for Semiparametric Models*. Springer-Verlag, New York.

Black, D., J. Smith, M. Berger, and B. Noel (2003): "Is the Threat of Training More Effective than Training Itself? Experimental Evidence from the UI System," *American Economic Review*, 93(4), 1313–1327.

Chamberlain, G. (2000): "Econometrics and Decision Theory," *Journal of Econometrics*, 95, 255–283.

Choi, S., W. J. Hall, and A. Schick (1996): "Asymptotically Uniformly Most Powerful Tests in Parametric and Semiparametric Models," *The Annals of Statistics*, 24(2), 841–861.

Dehejia, R. (2003): "When is ATE Enough? Risk Aversion and Inequality Aversion in Evaluating Training Programs," working paper, Columbia University.

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

Karlin, S., and H. Rubin (1956): "The Theory of Decision Procedures for Distributions with Monotone Likelihood Ratio," *Annals of Mathematical Statistics*, 27, 272–299.

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

——— (2002): "Treatment Choice Under Ambiguity Induced by Inferential Problems," *Journal of Statistical Planning and Inference*, 105, 67–82.

——— (2003): "Statistical Treatment Rules for Heterogeneous Populations," working paper, Northwestern University.

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

O'Leary, C. J., P. T. Decker, and S. A. Wandner (1998): "Reemployment Bonuses and Profiling," Discussion Paper 98-51, W. E. Upjohn Institute for Employment Research.

——— (2005): "Cost-Effectiveness of Targeted Reemployment Bonuses," *The Journal of Human Resources*, 40(1), 270–279.

Puhalskii, A., and V. Spokoiny (1998): "On Large Deviation Efficiency in Statistical Inference," *Bernoulli*, 4(2), 203–272.

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

Schervish, M. J. (1995): *Theory of Statistics*. Springer-Verlag, New York.

Van der Vaart, A. W. (1989): "On the Asymptotic Information Bound," *The Annals of Statistics*, 17, 1487–1500.

——— (1991a): "An Asymptotic Representation Theorem," *International Statistical Review*, 59, 99–121.

——— (1991b): "On Differentiable Functionals," *The Annals of Statistics*, 19, 178–204.

——— (1998): *Asymptotic Statistics*. Cambridge University Press, New York.

Van der Vaart, A. W., and J. A. Wellner (1996): *Weak Convergence and Empirical Processes*. Springer-Verlag, New York.

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

Worden, K. (1993): "Profiling Dislocated Workers for Early Referral to Reemployment Services," Unpublished manuscript, U.S. Department of Labor.