Robust Estimation of Some Nonregular Parameters

Kyungchul Song[1]
Department of Economics, University of Pennsylvania

June 17, 2010

Abstract

This paper develops optimal estimation of a potentially nondifferentiable functional $\Gamma(\beta)$ of a regular parameter β, when Γ satisfies certain conditions. Primary examples are min or max functionals that frequently appear in the analysis of partially identified models. This paper investigates both the average risk approach and the minimax approach. The average risk approach considers average local asymptotic risk with a weight function π over $\beta - q(\beta)$ for a fixed location-scale equivariant map q, and the minimax approach searches for a robust decision that minimizes the local asymptotic maximal risk. In both approaches, optimal decisions are proposed. Certainly, the average risk approach is preferable to the minimax approach when one has fairly accurate information of $\beta - q(\beta)$. When one does not, one may ask whether the average risk decision with a certain weight function π is as robust as the minimax decision. This paper specifies conditions for Γ such that the answer is negative. This paper discusses some results from Monte Carlo simulation studies.

Keywords: Local Asymptotic Minimax Estimation, Average Risks, Limit Experiments, Nondifferentiable Functionals, Partial Identification

JEL Codes: C10, C13, C14, C44.

1 Introduction

When one imposes inequality constraints on a parameter, the parameter is often rendered nonregular, i.e. made to behave nonsmoothly as the underlying probability is locally perturbed. For example, when a parameter β is known to take nonnegative values, the object

[1] I thank Frank Schorfheide for valuable comments. All errors are mine. Correspondence to: Kyungchul Song, 528 McNeil Building, 3718 Locust Walk, Philadelphia, PA 19104. Email address: kysong@sas.upenn.edu

of interest naturally takes the form of $\theta = \max\{\beta, 0\}$, a nondifferentiable transform of β. Nonregular parameters also frequently arise in partially identified models. Suppose for instance that the parameter of interest is interval identified in two different intervals $[\beta_{L,1}, \beta_{U,1}]$ and $[\beta_{L,2}, \beta_{U,2}]$. Then the identified interval becomes $[\max\{\beta_{L,1}, \beta_{L,2}\}, \min\{\beta_{U,1}, \beta_{U,2}\}]$ with nonregular bound parameters.

In contrast to the ease with which such parameters arise in the literature, a formal analysis of the estimation problem remains a challenging task. Among others, there does not exist an asymptotically unbiased estimator or a regular estimator for such parameters (e.g. Hirano and Porter (2009b) and references therein.) Furthermore, elimination of bias through a bias correction method entails infinite variance. (Doss and Sethuraman (1989)). One might ask what would be the optimal balance between bias and variance. The standard theory of semiparametric efficiency offers no answer in this regard, because there does not exist an influence function for the parameter to begin with.

This paper offers a partial answer by imposing a particular structure on the way that θ becomes nonregular. Suppose that a data generating process P of observations identifies a regular parameter $\beta \in \mathbf{R}^d$. It is assumed that the object of interest is not β but a certain functional Γ of β, i.e., $\theta = \Gamma(\beta)$. This paper focuses on a particular class of maps Γ, by requiring that Γ be a composition of a contraction map φ that satisfies a certain condition and a location-scale equivariant map ψ. Despite its seeming restrictiveness, a large class of nonregular parameters fall into this paper's framework.

Example 1: (Intersection Bounds): In partially identified models, the identified set of the reduced form parameters often takes the form of a rectangle or an interval. When there is a multiple number of identified rectangles, one often takes the intersection of these rectangles to obtain a tighter identified set. The resulting bounds typically involve min or max functions. For example, Haile and Tamer (2003) studied an English auction model and showed that the optimal reserve price is identified in such an interval. Other examples are found in the literature on bounds of treatment effects (Manski (1989, 1990, 1997), and Manski and Pepper (2000)), where the treatment effect bounds involve min or max transforms over values of exogenous variables. When these exogenous variables are discretized, the problem of estimating the bounds falls into this paper's scope. (See Manski (2008), Chernozhukov, Lee and Rosen (2009) and Hirano and Porter (2009b) for more examples.)

Example 2: (Fréchet-Hoeffding Bounds): Fréchet-Hoeffding bounds provide upper and lower bounds for the joint distribution function of a random vector when the marginal distributions are identified. For example, let F be the joint distribution function of X_1 and X_2 whose marginal distributions are uniform on $[0, 1]$. Then the joint distribution function

F lies between Fréchet-Hoeffding bounds: for $(x_1, x_2) \in [0,1]^2$,

$$\max\{x_1 + x_2 - 1, 0\} \leq F(x_1, x_2) \leq \min\{x_1, x_2\}.$$

Fan and Wu (2009) used these bounds in deriving the identified set for distributional treatment effects. Recently, Hoderlein and Stoye (2009) obtained similar bounds for the probability of the weak axiom of revealed preference being violated.

Example 3: (Sign Restrictions): In various econometric models, certain parameters have known sign restrictions due to the nature of the parameter or certain prior information, and the object of interest is a sign-restricted parameter, i.e., $\Gamma(\beta) = \max\{\beta, 0\}$ or $\Gamma(\beta) = \min\{\beta, 0\}$. A natural estimator $\Gamma(\hat{\beta})$ using an asymptotically unbiased estimator $\hat{\beta}$ of β suffers from asymptotic bias. Then one may ask whether there is an estimator that performs better than $\Gamma(\hat{\beta})$ in terms of the mean squared error, for example, by using an *asymptotically biased* estimator of β. The results of this paper address such questions in a much broader context.

Example 4: (Measuring the Best Possible Performance of a Set of Predictive Models): When there are multiple sets of predictive models available, one may be interested in estimating the maximum or minimum mean square prediction error over different predictive models. The minimum mean square prediction error measures the best possible performance of models in the set, and the maximum mean square error prediction the worst possible performance. These performance measures are nonregular parameters due to nondifferentiable transform $\Gamma(\cdot) = \min(\cdot)$ or $\Gamma(\cdot) = \max(\cdot)$.

The theory of optimal decisions in this paper is developed along two different approaches. The first approach focuses on the local asymptotic average risk, where one considers a weighted risk over the difference $\beta - q(\beta)$ for a location-scale equivariant functional q. This approach is relevant, for example, when $\Gamma(\beta) = |\max\{\beta_1, \beta_2\}|$ and one has reliable information of $\beta_1 - \beta_2$. This paper shows that the optimal decision minimizing the average risk takes the form of

$$\varphi\left(a'\tilde{\beta} + c/\sqrt{n}\right), \tag{1}$$

where $c \in \mathbf{R}$ is a bias-adjustment term that depends on the weight function π, $a \in \mathbf{R}^d$ is a certain vector, and $\tilde{\beta}$ is a semiparametrically efficient estimator of β. In this paper we call this decision an *average risk decision*.

The second approach considers a minimax approach, where one seeks a robust procedure that performs reasonably well regardless of the values of β. In this case, an estimator of the form:

$$\Gamma(\tilde{\beta} + w/\sqrt{n}), \tag{2}$$

with bias-adjustment term $w \in \mathbf{R}^d$, is shown to be robust in the sense of local asymptotic minimaxity. For example, when $\Gamma(\beta) = \max\{\beta_1, \beta_2\}$, the result implies that the minimax decision takes the form of $\max\{\tilde{\beta}_1 + w_1/\sqrt{n}, \tilde{\beta}_2 + w_2/\sqrt{n}\}$. When Γ is linear so that $\theta = \Gamma(\beta)$ is a regular parameter, the decision in (2) is reduced to a semiparametric efficient estimator of $\theta = \Gamma(\beta)$, confirming the continuity of this paper's approach with the standard literature of semiparametric efficiency.

In several examples of Γ, it is found that it suffices to set $w = 0$. For example, when $\Gamma(\beta) = \max\{\beta' b, s\}$, $\Gamma(\beta) = \max\{|\beta' b|, s\}$, or $\Gamma(\beta) = |\beta' b|$ with $b \in \mathbf{R}^d$ and s being a known vector and a scalar, the local asymptotic minimax decision in (2) does not require bias-adjustment. In these examples, the estimator $\Gamma(\tilde{\beta})$ is the local asymptotic minimax decision.

It is interesting to observe that when the candidate decisions are appropriately restricted, the optimal estimator in (2) is reduced to $\max\{\tilde{\beta}_1, \cdots, \tilde{\beta}_d\} + v/\sqrt{n}$, with bias-adjustment quantity v. This is a form that is similar to a bias-reduced decision proposed by Chernozhukov, Lee and Rosen (2009) recently. Their major analysis centers around the case of infinite-dimensional β, its primary focus being on improved inference on $\Gamma(\beta)$ not on its optimal estimation. In contrast to their method, this paper's bias-adjustment term v is adaptive to the given decision-theoretic environment such as loss functions. Therefore, when bias-adjustment tends to do more harm than good in terms of the local asymptotic maximal risk, the bias-adjustment term v is automatically set to be close to zero.

A natural question that arises is whether one can robustify the average risk decision in (1) by employing a highly "uninformative" weight function π such as a uniform density over a large area. It turns out that when ψ is nondifferentiable, there exists no weight function π for which the average risk decision achieves the minimax risk. While this is already hinted from the fact that the decision (1) cannot be reduced to that of (2) for any π, the result is proved formally. Therefore, the average risk approach with an uninformative weight function has limitation in delivering a robust decision when ψ is nondifferentiable.

Inference in nonregular models has long received attention in the literature. For example, estimation of a normal mean under bound restrictions has been studied by Casella and Strawderman (1981), Bickel (1981), and Moors (1981), and estimation of parameters under order restrictions, by Blumenthal and Cohen (1968b), and Lovell and Prescott (1970). Andrews (1999, 2001) proposed general asymptotic inference procedures when parameters potentially lie on the boundary of the parameter space. Estimating a parameter from a family of nonregular densities has also been investigated in the literature (See Pflug (1983), Ghosh, Ghosal and Samanta (1994), and references therein.) Hirano and Porter (2003) and Chernozhukov and Hong (2003) studied likelihood models that have a parameter-dependent support. See also Ploberger and Phillips (2010) for optimal estimation under nonregular models with dependent observations.

A research that is closest to this paper is Blumenthal and Cohen (1968a) who studied a generalized Bayes estimator and a maximum likelihood estimator of $\max\{\beta_1, \beta_2\}$, when two independent sets of i.i.d. observations from two location families are available. Chernozhukov, Lee and Rosen (2009) recently proposed and analyzed a bias-reduction method for inference procedures of parameters such as $\min(\beta)$ or $\max(\beta)$ when β is finite dimensional or infinite dimensional.

The implication of a nondifferentiable transform for a regular parameter has been noted by Hirano and Porter (2009b). In particular, they pointed out that for a parameter that is not differentiable in the underlying probability, there exists no asymptotic unbiased estimator. See also van der Vaart (1991) and Theorem 9 of Lehmann (1986), p.55, for a related result.[2] See Doss and Sethuraman (1989) for implications of bias-correction when the parameter has no unbiased estimator.

The next section defines the decision-theoretic environment in general terms, introducing loss functions and risk. Section 3 investigates optimal decisions based on average risks, and Section 4, the maximal risks. At the end of Section 4, this paper discusses nonminimaxity of average risk decisions. Section 5 concludes. Technical proofs are relegated to the Appendix.

2 Parameter of Interest, Loss and Risk

The loss function represents the decision-maker's preference over various pairs of the decision and the object of interest. As for the decision space and the loss function, we consider the following.

Assumption 1: (i) The decision space D is given by $D = \mathbf{R}$.
(ii) The loss function is given as follows: for $d \in D$ and $\theta \in \mathbf{R}$,

$$L(d, \theta) = \tau(|d - \theta|) \tag{3}$$

where $\tau : [0, \infty) \to [0, \infty)$ is increasing on $[0, \infty)$, $\tau(0) = 0$, $\tau(y) \to \infty$ as $y \to \infty$, and for each $M > 0$, $\min\{\tau(\cdot), M\}$ is Lipschitz continuous.

The decision space is a real line and the loss function is an increasing function of the

[2] I thank Marcelo Moreira for pointing me to the latter reference.

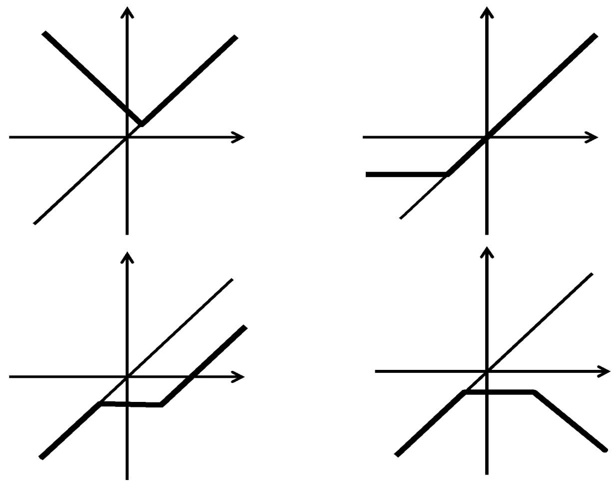

Figure 1: Some Examples of $\varphi(y)$

difference between the object of interest θ and the decision d. The condition that $\tau(y) \to \infty$ as $y \to \infty$ are used only in Theorem 5 in Section 4.2.1. later.

We introduce some notations. Let $\mathbf{1}_d$ be a $d \times 1$ vector of ones. For a vector $x \in \mathbf{R}^d$ and a scalar c, we simply write $x + c = x + \mathbf{1}_d c$. We define $S_1 \equiv \{x \in \mathbf{R}^d : x'\mathbf{1}_d = 1\}$, where the notation $\equiv$ indicates definition. When $x \in \mathbf{R}^d$, the notation $\max(x)$ (or $\min(x)$) means the maximum (or the minimum) over the entries of the vector x. When $x_1, \cdots, x_n$ are scalars, we also use the notations $\max\{x_1, \cdots, x_n\}$ and $\min\{x_1, \cdots, x_n\}$ whose meaning is obvious.

As for the parameter of interest θ, this paper assumes that

$$\theta = \Gamma(\beta) \equiv (\varphi \circ \psi)(\beta)$$

where $\beta \in \mathbf{R}^d$ is a regular parameter and $\varphi \circ \psi$ is the composite map of φ and ψ. (The regularity condition for β is specified in Assumption 4 below.) As for the maps ψ and φ, we assume the following.

Assumption 2: (i) The map $\psi : \mathbf{R}^d \to \mathbf{R}$ is Lipschitz continuous, and satisfies the following.
(a) (Location Equivariance) For each $c \in \mathbf{R}$ and $x \in \mathbf{R}^d$, $\psi(x + c) = \psi(x) + c$.
(b) (Scale Equivariance) For each $u \geq 0$ and $x \in \mathbf{R}^d$, $\psi(ux) = u\psi(x)$.
(ii) The map $\varphi : \mathbf{R} \to \mathbf{R}$ satisfies the following.
(a) (Contraction) For any $y_1, y_2 \in \mathbf{R}$, $|\varphi(y_1) - \varphi(y_2)| \leq |y_1 - y_2|$.

(b) (Identity on a Scanning Set) For some $k_0 \in \mathbf{R}$, $\varphi(y) = y$ for all $y \in (-\infty, k_0]$ or $\varphi(y) = y$ for all $y \in [k_0, \infty)$.

Assumption 2 essentially defines the scope of this paper. Some examples of ψ and φ are as follows. (See Figure 1 also for some examples of φ.)

EXAMPLES 5: (i)(a) $\psi(x) = b'x$, where $b \in S_1$.
(b) $\psi(x) = \max(x)$ or $\psi(x) = \min(x)$.
(c) $\psi(x) = \max\{\min(x_1), x_2\}$ where x_1 and x_2 are (possibly overlapping) subvectors of x.
(d) $\psi(x) = \max(x_1) + \max(x_2)$, $\psi(x) = \min(x_1) + \min(x_2)$, $\psi(x) = \max(x_1) + \min(x_2)$, or $\psi(x) = \max(x_1) + b'x_2$ with $b \in S_1$.
(ii)(a) $\varphi(y) = y$, $\varphi(y) = \max\{y, s\}$ or $\varphi(y) = \min\{y, s\}$ for some known constant $s \in \mathbf{R}$.
(b) $\varphi(y) = |y|$.
(c) $\varphi(y) = \max\{|y|, s\}$ or $\varphi(y) = \min\{-|y|, s\}$ for some known constant $s \in \mathbf{R}$.

When we take $\psi(x) = b'x$ as in Example 5(i)(a) and $\varphi(y) = y$, the parameter $\Gamma(\beta)$ becomes a regular one to which the existing theory of asymptotically optimal estimation applies. This example is used to confirm that the results of this paper are consistent with the existing theory.

Many examples of nondifferentiable maps Γ are written as $\varphi \circ \psi$ or $a(\varphi \circ \psi) + b$ for some known constants $a > 0$ and $b \in \mathbf{R}$. For intersection bounds of the form $\max(\beta)$ or $\min(\beta)$, we can simply take $\Gamma(\beta) = (\varphi \circ \psi)(\beta)$ with φ being an identity map and $\psi(\beta)$ being $\max(\beta)$ or $\min(\beta)$. In the case of the Fréchet-Hoeffding lower bound, i.e., $\Gamma(\beta) = \max\{\beta_1 + \beta_2 - 1, 0\}$, we write it as $2\tilde{\Gamma}(\beta) - 1$, where

$$\tilde{\Gamma}(\beta) = \max\{(\beta_1 + \beta_2)/2, 1/2\}.$$

Hence it suffices to produce an optimal decision $\hat{\theta}^*$ on $\tilde{\Gamma}(\beta)$ and take $2\hat{\theta}^* - 1$ as our optimal decision for $\Gamma(\beta)$. By taking $\varphi(y) = \max\{y, 1/2\}$ and $\psi(x) = (x_1 + x_2)/2$, the functional $\tilde{\Gamma}(\beta)$ is written as the composite map of φ and ψ. Assumptions 2(i) and (ii) are satisfied by ψ and φ respectively.

We introduce two assumptions for β and the underlying probability model that identifies β (Assumptions 3 and 4.) These two assumptions are standard, whose eventual consequence is the existence of a well defined semiparametric efficiency bound for the parameter β. The formulation of regularity conditions for β below is taken from Song (2009), which is originally adapted from van der Vaart (1991) and van der Vaart and Wellner (1996). Let $\mathcal{B}$ be the Borel σ-field of $\mathbf{R}^d$ and $(H, \langle\cdot,\cdot\rangle)$ be a linear subspace of a separable Hilbert space with $\bar{H}$ denoting its completion. Let $\mathbb{N}$ be the collection of natural numbers. For each $n \in \mathbb{N}$ and $h \in H$, let $P_{n,h}$ be a probability on $(\mathbf{R}^d, \mathcal{B})$ indexed by $h \in H$, so that $\mathcal{E}_n = (\mathbf{R}^d, \mathcal{B}, P_{n,h}; h \in H)$ constitutes a sequence of experiments. As for $\mathcal{E}_n$, we assume local asymptotic normality as follows.

Assumption 3: (Local Asymptotic Normality) For each $h \in H$,

$$\log \frac{dP_{n,h}}{dP_{n,0}} = \zeta_n(h) - \frac{1}{2}\langle h, h \rangle,$$

where for each $h \in H$, $\zeta_n(h) \rightsquigarrow \zeta(h)$ (weak convergence under $\{P_{n,0}\}$) and $\zeta(\cdot)$ is a centered Gaussian process on H with covariance function $\mathbf{E}[\zeta(h_1)\zeta(h_2)] = \langle h_1, h_2 \rangle$.

Local asymptotic normality reduces the decision problem to one in which an optimal decision is sought under a single Gaussian shift experiment $\mathcal{E} = (\mathbf{R}^d, \mathcal{B}, P_h; h \in H)$, where P_h is such that $\log dP_h/dP_0 = \zeta(h) - \frac{1}{2}\langle h, h \rangle$. The local asymptotic normality is ensured, for example, when $P_{n,h} = P_h^n$ and P_h is Hellinger-differentiable (Begun, Hall, Huang, and Wellner (1983).) The space H is a tangent space for β associated with the space of probability sequences $\{\{P_{n,h}\}_{n \geq 1} : h \in H\}$ (van der Vaart (1991).) Taking β as an $\mathbf{R}^d$-valued map on $\{\{P_{n,h}\}_{n \geq 1} : h \in H\}$, we can regard the map as a sequence of $\mathbf{R}^d$-valued maps on H and write it as $\beta_n(h)$.

Assumption 4: (Regular Parameter) There exists a continuous linear $\mathbf{R}^d$-valued map on H, $\dot{\beta}$, such that

$$\sqrt{n}(\beta_n(h) - \beta_n(0)) \to \dot{\beta}(h)$$

as $n \to \infty$.

Assumption 4 says that the sequence of parameters $\beta_n(h)$ are *differentiable* in the sense of van der Vaart (1991). The continuous linear map $\dot{\beta}$ is associated with the semiparametric efficiency bound of the boundary parameter in the following way. Let $\dot{\beta}^* \in \bar{H}$ be such that for each $b \in \mathbf{R}^d$ and each $h \in H$, $b'\dot{\beta}(h) = \langle b'\dot{\beta}^*, h \rangle$. Then for any $b \in \mathbf{R}^d$, $||b'\dot{\beta}^*||^2$ represents the asymptotic variance bound of the parameter $b'\beta$. The map $\dot{\beta}^*$ is called the efficient influence function of β in the literature (e.g. van der Vaart (1991)). For future references, we define

$$\Sigma \equiv \langle \dot{\beta}^*, \dot{\beta}^{*\prime} \rangle. \tag{4}$$

As for Σ, we assume the following:

Assumption 5: Σ is invertible.

The inverse of matrix Σ is the semiparametric efficiency bound for β.

3 Optimal Decisions based on Average Risks

Suppose that one has prior information of $\beta - q(\beta)$ for some functional q on $\mathbf{R}^d$ such that q satisfies location and scale equivariance conditions of Assumption 2(i)(a)(b). This is the situation, for example, where $\Gamma(\beta) = \max\{\beta_1, \beta_2\}$ and one has information of $\beta_1 - \beta_2$. (See Example 6 below.) One can always translate information of $\beta - q(\beta)$ into that of $\beta - a'\beta$ for any vector $a \in S_1$ as follows:

$$\beta - a'\beta = U_a(\beta - q(\beta)), \tag{5}$$

where $U_a = I_d - \mathbf{1}_d a'$ and I_d is the d dimensional identity matrix. The following example illustrates how we translate information of $\beta - q(\beta)$ into that of $\beta - a'\beta$ when we have information of $\beta - q(\beta)$ in terms of a prior density π_1.

Example 6: Suppose that $\beta = [\beta_1, \beta_2]' \in \mathbf{R}^2$. At the current sample size n, suppose that one has prior information of $s = \sqrt{n}(\beta_2 - \beta_1)$ that is represented by density function $\pi_1(s) = \frac{1}{\sqrt{2\pi}} \exp(-(s - \bar{r})^2/2)$ for some known constant $\bar{r} \in \mathbf{R}$. This is equivalent to saying that we have information of $\beta - q(\beta)$ with $q(\beta) = (\beta_1 + \beta_2)/2$ because $\sqrt{n}(\beta - q(\beta)) = [-s/2, s/2]'$. Now, for any choice of $a \in S_1$,

$$\sqrt{n}\,(\beta - a'\beta) = \begin{bmatrix} -(1 - a_1)\sqrt{n}(\beta_2 - \beta_1) \\ a_1\sqrt{n}(\beta_2 - \beta_1) \end{bmatrix},$$

where a_1 is the first component of a. The weight function π for $\sqrt{n}\,(\beta - a'\beta)$ is taken to be the density of $[-(1 - a_1)W, a_1 W]'$, where $W \sim N(\bar{r}, 1)$. ■

Since we can translate information of $\beta - q(\beta)$ into that of $\beta - a'\beta$ for any vector $a \in S_1$, we lose no generality by fixing $a \in S_1$ that is convenient for our purpose. It is convenient, as this paper does, if we choose a as

$$a = \frac{\Sigma^{-1}\mathbf{1}_d}{\mathbf{1}_d'\Sigma^{-1}\mathbf{1}_d}, \tag{6}$$

so that the constraint $\dot{\beta} - a'\dot{\beta} = 0$ is made *ancillary* for the efficient estimation of the regular component $a'\beta$. This *does not* mean that this paper's procedure renders information of $\beta - q(\beta)$ irrelevant by choosing a as in (6). (See Section 5.2.2 for simulation results that reflect advantage of such information.) Choice of such a is merely a normalization in which we translate knowledge of $\beta - q(\beta)$ into that of $\beta - a'\beta$ so that after the translation the constraint $\dot{\beta} - a'\dot{\beta} = 0$ does not interfere with efficient estimation of the regular component $a'\beta$.

Let us consider the following *subclass maximal risk*: for each $r \in S(a)$,

$$\mathcal{R}_n^{\varepsilon}(\hat{\theta}; r) = \sup_{h \in H_n^{\varepsilon}(r)} \mathbf{E}_h \left[\tau(|\sqrt{n}\{\hat{\theta} - \theta\}|) \right], \text{ with } \theta = \Gamma(\beta_n(h)), \tag{7}$$

where $H_n^{\varepsilon}(r) \equiv \{h \in H : ||\sqrt{n}\{\beta_n(h) - a'\beta_n(h)\} - r|| \leq \varepsilon\}$ for $\varepsilon > 0$. The set $H_n^{\varepsilon}(r)$ is a collection of h such that $\beta - a'\beta$ is approximately equal to $r/\sqrt{n}$. Given a nonnegative weight function π over r, we consider the average risk:

$$\int \mathcal{R}_n^{\varepsilon}(\hat{\theta}; r)\pi(r)dr. \tag{8}$$

The approach of average risks allows one to incorporate prior information of r into the decision-making process.[3]

The theorem below establishes an asymptotic average risk bound. Let $Z \in \mathbf{R}^d$ be a normal random vector such that

$$Z \sim N(0, \Sigma), \tag{9}$$

where Σ is as defined in (4).

Theorem 1: *Suppose that* Assumptions 1-5 *hold and* $\int \mathbf{E}\left[\tau(|a'Z - \psi(r)|)\right]\pi(r)dr < \infty$. *Then, for any sequence of estimators* $\hat{\theta}$,

$$\lim_{\varepsilon \to 0} \liminf_{n \to \infty} \int \mathcal{R}_n^{\varepsilon}(\hat{\theta}; r)\pi(r)dr \geq \inf_{c \in \mathbf{R}} \int \mathbf{E}\left[\tau\left(|a'Z + c - \psi(r)|\right)\right]\pi(r)dr. \tag{10}$$

Note that the lower bound does not involve the map φ. When $\psi(\beta) = b'\beta$ and π is symmetric around 0, the infimum over $c \in \mathbf{R}$ in the lower bound of (10) is achieved by $c = 0$ due to Anderson's Lemma (e.g. Strasser (1985).) However, in general, the infimum is achieved by a nonzero c.

Let us define an optimal solution that achieves the bound. The solution involves two components: a semiparametrically efficient estimator $\tilde{\beta}$ of β and a *bias-adjustment term* c^* that solves the minimization in the risk bound in (10). As for $\tilde{\beta}$, we assume that

$$\sqrt{n}\{\tilde{\beta} - \beta\} \to_d Z \sim N(0, \Sigma).$$

As for estimation of c^*, we consider the following procedure. Let $M > 0$ be a fixed large

[3] One might view the weight function π as playing the role of a prior in the Bayesian approach. It should be noted though that the average risk approach here is not a fully Bayesian approach because the "prior" is imposed only over the index r that represents $\sqrt{n}\{\beta - a'\beta\}$ in the limit, not over the whole index $h \in H$ of the likelihood process.

number and

$$\tau_M(\cdot) \equiv \min\{\tau(\cdot), M\}. \tag{11}$$

Define $\hat{\Sigma}$ to be a consistent estimator of Σ and let

$$\hat{a} = \frac{\hat{\Sigma}^{-1}\mathbf{1}_d}{\mathbf{1}_d'\hat{\Sigma}^{-1}\mathbf{1}_d}.$$

Let $\{\xi_i\}_{i=1}^L$ be i.i.d. draws from $N(0, I_d)$ and

$$\tilde{Q}_\pi(c) = \int \frac{1}{L}\sum_{i=1}^{L} \tau_M\left(|\hat{a}'\xi_i + c - \psi(r)|\right)\pi(r)dr. \tag{12}$$

The integration over r in the above can be done using a numerical integration method. Then define

$$\tilde{c}_\pi^* = \frac{1}{2}\left\{\max \tilde{E}_\pi + \min \tilde{E}_\pi\right\}, \tag{13}$$

where $\tilde{E}_\pi = \{c \in [-M, M] : \tilde{Q}_\pi(c) \leq \inf_{c\in[-M,M]} \tilde{Q}_\pi(c) + \eta_{n,L}\}$ with $\eta_{n,L} \to 0$ and $\eta_{n,L}\sqrt{n} \to \infty$ as $n \to \infty$ and $\eta_{n,L}\sqrt{L} \to \infty$ as $L \to \infty$.

The optimization for obtaining $\tilde{c}_\pi^*$ does not entail much computational cost as c is a scalar regardless of the dimension d. The formulation of $\tilde{c}_\pi^*$ in (13) is designed to facilitate the proof of the result. In practice, it suffices to take an infimum of $\tilde{Q}_\pi(c)$ over $c \in [-M, M]$.

When one knows $\beta - a'\beta = \bar{r}/\sqrt{n}$ with certainty for some known vector $\bar{r} \in \mathbf{R}^d$ and accordingly adopts $\pi(\cdot)$ as a point mass at $\bar{r}$, we can take

$$\tilde{c}_\pi^* = \psi(\bar{r}).$$

Hence we do not have to go through a numerical step in this case.

As for $\hat{\Sigma}$ and $\tilde{\beta}$, we assume the following.

Assumption 6: (i) For each $\varepsilon > 0$, there exists $M > 0$ such that

$$\text{limsup}_{n\to\infty}\text{sup}_{h\in H} P_{n,h}\{\sqrt{n}||\hat{\Sigma} - \Sigma|| > M\} < \varepsilon.$$

(ii) For each $t \in \mathbf{R}^d$, $\sup_{h\in H}\left|P_{n,h}\{\sqrt{n}(\tilde{\beta} - \beta_n(h)) \leq t\} - P\{Z \leq t\}\right| \to 0$ as $n \to \infty$.

Assumption 6 imposes $\sqrt{n}$-consistency of $\hat{\Sigma}$ and convergence in distribution of $\sqrt{n}(\tilde{\beta} - \beta_n(h))$ both uniform over $h \in H$. The uniform convergence can be proved in various ways. (e.g. Lemma 2.1 of Giné and Zinn (1991).)

Now we are prepared to introduce an optimal decision. Let

$$\tilde{\theta}_\pi = \varphi\left(\hat{a}'\tilde{\beta} + \frac{\tilde{c}_\pi^*}{\sqrt{n}}\right). \tag{14}$$

The solution depends on the given weight function π through $\tilde{c}_\pi^*$. Verifying the optimality of $\tilde{\theta}_\pi$ may involve proving the uniform integrability condition for a sequence of the decisions. To dispense with such a nuisance, we follow the suggestion by Strasser (1985) (p.480) and consider instead

$$\mathcal{R}_{n,M}^\varepsilon(\hat{\theta}; r) \equiv \sup_{h \in H_n^\varepsilon(r)} \mathbf{E}_h\left[\tau_M(|\sqrt{n}\{\hat{\theta} - \theta\}|)\right] \tag{15}$$

with $\theta = \Gamma(\beta_n(h))$ and with τ_M defined in (11). The following theorem establishes that the solution $\tilde{\theta}_\pi$ is optimal.

Theorem 2: *Suppose that* Assumptions 1-6 *hold and* $\int \mathbf{E}\left[\tau(|a'Z - \psi(r)|)\right]\pi(r)dr < \infty$. *Then,*

$$\lim_{M,L\to\infty} \lim_{\varepsilon\to 0} \limsup_{n\to\infty} \int \mathcal{R}_{n,M}^\varepsilon(\tilde{\theta}_\pi; r)\pi(r)dr \leq \inf_{c\in\mathbf{R}} \int \mathbf{E}\left[\tau\left(|a'Z + c - \psi(r)|\right)\right]\pi(r)dr.$$

When $\psi(\beta) = b'\beta$ for $b \in S_1$ and π is symmetric around zero, the bias-adjustment term c^* is zero, so that the optimal decision in this case becomes simply

$$\tilde{\theta}_\pi = \varphi(\hat{a}'\tilde{\beta}).$$

This yields the following results.

Example 7: (a) When $\Gamma(\beta) = \beta'b$ for a known vector $b \in S_1$, $\tilde{\theta}_\pi = \hat{a}'\tilde{\beta}$. Interestingly, the optimal decision does not depend on b. This is because when $\beta \approx a'\beta + r/\sqrt{n}$, we have $b'\beta \approx a'\beta + b'r/\sqrt{n}$ so that the rotation vector b is involved only in the constant drift component $b'r/\sqrt{n}$ and hence in $\tilde{c}_\pi^*$ in (14). As long as π is symmetric around 0, Anderson's Lemma makes the role of b neutral, because regardless of b, we can set $\tilde{c}_\pi^* = 0$.
(b) When $\Gamma(\beta) = \max\{\beta'b, s\}$ for a known vector $b \in S_1$ and a known constant s, $\tilde{\theta}_\pi = \max\{\hat{a}'\tilde{\beta}, s\}$.
(c) When $\Gamma(\beta) = |\beta|$ for a scalar parameter β, $\tilde{\theta}_\pi = |\tilde{\beta}|$. This result is reminiscent of a result of Blumenthal and Cohen (1968a) that the minimax estimator of $|\beta|$ from a single observation $Y \sim N(\beta, 1)$ is $|Y|$.
(d) When $\Gamma(\beta) = \max\{\beta_1 + \beta_2 - 1, 0\}$, $\tilde{\theta}_\pi = \max\{2\hat{a}'\tilde{\beta} - 1, 0\}$. ■

The following example investigates whether the solution in (14) reduces to a semiparametrically efficient estimator when θ is regular.

EXAMPLE 8: Consider the case of Example 7(a), where one knows for certain that $\beta = b'\beta$ so that $\beta - a'\beta = 0$. Then $\tilde{\theta}_\pi = \hat{a}'\tilde{\beta}$ is a well-known efficient estimator of $\Gamma(\beta)$. To see this, let $B = \Sigma A'(A\Sigma A')^{-1}A$ with $A = -(I_{d-1} - \mathbf{1}_{d-1}a_2')[\mathbf{1}_{d-1}; -I_{d-1}]$ and a_2 is a $(d-1) \times 1$ subvector of a with the first entry a_1 excluded. One can show that the limiting distribution of

$$\sqrt{n}\{\tilde{\theta}_\pi - b'\beta\}$$

is equal to that of $\sqrt{n}\{\hat{\theta} - b'\beta\}$ with $\hat{\theta} = b'(I-B)\tilde{\beta}$. The estimator $\hat{\theta}$ is an efficient estimator of $b'\beta$ under the constraint that $\beta - b'\beta = 0$. ■

4 Robust Decisions based on Maximal Risks

4.1 Local Asymptotic Minimax Decisions

When one does not have prior information of r and the decision is sensitive to the choice of a weight function π, one may pursue a robust procedure instead. In this section, we consider a minimax approach.

A typical approach to find a minimax decision searches for a least favorable prior whose Bayes risk is equal to the minimax risk. Finding a least favorable prior is often complicated when the parameter of interest is constrained or required to satisfy certain order restrictions. This is true even if the parameter of focus is a point on the real line and observations follow a normal family of distributions. This paper takes a different approach that makes full use of the conditions for the map ψ in Assumption 2.

We define the *local maximal risk*:

$$\mathcal{R}_n(\hat{\theta}) = \sup_{h \in H} \mathbf{E}_h \left[\tau(|\sqrt{n}\{\hat{\theta} - \theta\}|) \right],$$

where $\theta = \Gamma(\beta_n(h))$. The situation here is different from the average risk decision. In the case of the average risk decision, the optimality result is uniform over the limit values of $\sqrt{n}\{\beta_n(0) - \psi(\beta_n(0))\}$ due to the use of the subclass system based on (6). In the minimax approach, the limit values matter. For example, when $\psi(\beta) = \max\{\beta_1, \beta_2\}$, the limit of the risk $\mathcal{R}_n(\hat{\theta})$ changes depending on whether β_1 is close to β_2 or not. The local asymptotic minimax approach that this paper develops pursues a robust decision that does not assume knowledge of $\sqrt{n}\{\beta_n(0) - \psi(\beta_n(0))\}$ and focuses on a supremum of the limit of the risk $\mathcal{R}_n(\hat{\theta})$ where the supremum is over all the possible limit values of $\sqrt{n}\{\beta_n(0) - \psi(\beta_n(0)\}$.

Let $\Psi \equiv \{\delta \in [-\infty, \infty]^d : \psi(\delta) = 0\}$. For each $\delta \in \Psi$ and $\varepsilon > 0$, let $N(\delta; \varepsilon) \equiv \{n \in \mathbb{N} : ||\sqrt{n}\{\beta_n(0) - \psi(\beta_n(0))\} - \delta|| \leq \varepsilon\}$. We present the following local asymptotic minimax risk bound.

Theorem 3: *Suppose that* Assumptions 1-5 *hold and* $\sup_{r \in \mathbf{R}^d} \mathbf{E}\left[\tau(|\psi(Z+r) - \psi(r)|)\right] < \infty$. *Then for any sequence of estimators* $\hat{\theta}$,

$$\sup_{\delta \in \Psi} \lim_{\varepsilon \to 0} \liminf_{n \in N(\delta;\varepsilon)} \mathcal{R}_n(\hat{\theta}) \geq \inf_{w \in \mathbf{R}^d} \sup_{r \in \mathbf{R}^d} \mathbf{E}\left[\tau(|\psi(Z+r+w) - \psi(r)|)\right].$$

As in Theorem 1, the lower bound does not depend on φ that constitutes Γ. The main feature of the lower bound in Theorem 3 is that it involves infimum over a *finite dimensional space* $\mathbf{R}^d$ in its risk bound. In general, as a consequence of the generalized convolution theorem (e.g. Theorem 2.2 of van der Vaart (1989)), the risk bound involves infimum over the infinite dimensional space of probability measures over $\mathbf{R}^d$. In a standard situation with $\Gamma(\beta) = b'\beta$ with $b \in S_1$, this infimum poses no difficulty because the infimum is achieved by a probability measure with a point mass at zero due to Anderson's Lemma. However for a general class of Γ as is the focus of this paper, the infimum over probability measures in the lower bound complicates the computation of the optimal decision. To avoid this difficulty, Theorem 3 makes use of the classic purification result in the game theory (Dvoretsky, Wald, and Wolfowitz (1951).)

We construct a minimax decision as follows. Draw $\{\xi_i\}_{i=1}^L$ i.i.d. from $N(0, I_d)$ as before, and let

$$\tilde{Q}_{mx}(w) = \sup_{r \in \mathbf{R}^d : r'\mathbf{1}_d = 0} \frac{1}{L} \sum_{i=1}^{L} \tau_M\left(|\psi(\hat{\Sigma}^{1/2}\xi_i + r + w) - \psi(r)|\right).$$

Then for large $M > 0$,.we obtain

$$\tilde{w}^*_{mx} = \frac{1}{2}\left\{\max \tilde{E}_{mx} + \min \tilde{E}_{mx}\right\},$$

where $\tilde{E}_{mx} = \{w \in [-M, M]^d : \tilde{Q}_{mx}(w) \leq \inf_{w \in [-M,M]^d} \tilde{Q}_{mx}(w) + \eta_{n,L}\}$. Here $\max \tilde{E}_{mx}$ is the collection of vectors of coordinatewise maximizers, i.e. $e \in \tilde{E}_{mx}$ if and only if $e_j \geq \max\{\tilde{e}_j : \tilde{e} \in \tilde{E}_{mx}\}$ for all $j = 1, \cdot\cdot\cdot, d$. Similarly we define $\min \tilde{E}_{mx}$. This specification of $\tilde{w}^*_{mx}$ facilitates the proof of the result. In practice, it suffices to take $\tilde{w}^*_{mx}$ as a minimizer of $\tilde{Q}_{mx}(w)$ over $w \in [-M, M]^d$ for a large number M.

We introduce a local asymptotic minimax decision. Let $\tilde{\beta}$ be a semiparametrically efficient estimator of β as in Theorem 2, and let

$$\tilde{\theta}_{mx} = \Gamma\left(\tilde{\beta} + \frac{\tilde{w}^*_{mx}}{\sqrt{n}}\right). \tag{16}$$

In the following, we establish that $\tilde{\theta}_{mx}$ is local asymptotic minimax.

THEOREM 4: *Suppose that* Assumptions 1-6 *hold and* $\sup_{r\in\mathbf{R}^d}\mathbf{E}\left[\tau(|\psi(Z+r)-\psi(r)|)\right]<\infty$. *Then,*

$$\lim_{M\to\infty}\sup_{\delta\in\Psi}\lim_{\varepsilon\to 0}\limsup_{n\in N(\delta;\varepsilon)}\mathcal{R}_{n,M}(\tilde{\theta}_{mx}) \leq \inf_{w\in\mathbf{R}^d}\sup_{r\in\mathbf{R}^d}\mathbf{E}\left[\tau(|\psi(Z+r+w)-\psi(r)|)\right],$$

where $\mathcal{R}_{n,M}(\cdot)$ is $\mathcal{R}_n(\cdot)$ except that $\tau(\cdot)$ is replaced by $\tau_M(\cdot)$.

When $\psi(\beta)$ is a regular parameter, taking the form of $\psi(\beta) = b'\beta$ with $b \in S_1$, the local asymptotic minimax risk bound becomes

$$\inf_{w\in\mathbf{R}^d}\mathbf{E}\left[\tau\left(|\psi(Z+w)|\right)\right] = \mathbf{E}\left[\tau\left(|b'Z|\right)\right].$$

Hence one does not need to compute $\tilde{w}^*_{mx}$, because the bias-adjustment term w^* (i.e. a solution of the above minimization over $w \in \mathbf{R}^d$) is zero. Hence in this case, we simply set $\tilde{w}^*_{mx} = 0$ so that the minimax decision becomes simply

$$\tilde{\theta}_{mx} = \varphi(\tilde{\beta}'b). \tag{17}$$

This has the following consequences.

EXAMPLE 9: (a) When $\Gamma(\beta) = \beta'b$ for a known vector $b \in S_1$, $\tilde{\theta}_{mx} = \tilde{\beta}'b$. Therefore, the decision in (17) reduces to the well-known semiparametric efficient estimator of $\beta'b$.
(b) When $\Gamma(\beta) = \max\{\beta'b, s\}$ for a known vector $b \in S_1$ and a known constant s, $\tilde{\theta}_{mx} = \max\{\tilde{\beta}'b, s\}$.
(c) When $\Gamma(\beta) = |\beta|$ for a scalar parameter β, $\tilde{\theta}_{mx} = |\tilde{\beta}|$. This decision coincides with the average risk decision with π symmetric around 0 (Example 7(c)).
(d) When $\Gamma(\beta) = \max\{\beta_1 + \beta_2 - 1, 0\}$, $\tilde{\theta}_{mx} = \max\{\tilde{\beta}_1 + \tilde{\beta}_2 - 1, 0\}$. ■

The examples of (b)-(d) involve nondifferentiable transform Γ, and hence estimators of $\Gamma(\beta)$ for a regular parameter β are asymptotically biased. However, the result of this paper tells that the natural estimator $\Gamma(\tilde{\beta})$ that does not involve any bias-reduction is local asymptotic minimax.

4.2 Discussions

4.2.1 Nonminimaxity of Average Risk Decisions $\tilde{\theta}_\pi$ for Nondifferentiable ψ

When information of $\beta - q(\beta)$ is not available for any q, one may ask whether one can still achieve the robustness of the minimax decision by using the average risk decision $\tilde{\theta}_\pi$ with a "least favorable prior" π. When ψ is nondifferentiable, the answer is negative as we shall see now. Let Π be the set of all the nonnegative functions on $\mathbf{R}^d$. For each $M > 0$, let $\mathcal{D}_{n,M}^{AV}$ be the collection of decisions $\tilde{\theta}_\pi$ as given in (14) with π running in Π.

If there exists decision $\tilde{\theta}_\pi \in \mathcal{D}_{n,M}^{AV}$ for some $\pi \in \Pi$ such that

$$\lim_{M\to\infty} \sup_{\delta\in\Psi} \lim_{\varepsilon\to 0} \limsup_{n\in N(\delta;\varepsilon)} \mathcal{R}_{n,M}^{\varepsilon}(\tilde{\theta}_\pi) \leq \inf_{w\in\mathbf{R}^d} \sup_{r\in\mathbf{R}^d} \mathbf{E}\left[\tau(|\psi(Z+r+w)-\psi(r)|)\right],$$

then we can say that the decision $\tilde{\theta}_\pi$ is as robust as the minimax decision $\tilde{\theta}_{mx}$. Indeed, from Examples 6 and 8 that when $\Gamma(\beta) = \max\{\beta, s\}$ or $\Gamma(\beta) = |\beta|$, for a scalar parameter β, or $\Gamma(\beta) = \beta' b$ with $b \in S_1$ for a vector parameter β, taking π to be symmetric around 0 makes the average risk decision a minimax decision. These examples have a common feature that $\psi(\beta) = \beta' b$ with $b \in S_1$. When ψ is nondifferentiable, there does not exist a decision in $\mathcal{D}_{n,M}^{AV}$ with a potential for a minimax decision as the following theorem shows.

Theorem 5: *Suppose that* Assumptions 1-5 *hold. Furthermore, assume that ψ is nondifferentiable. Then, there exists no decision sequence $\{\hat{\theta}_n\}_{n=1}^{\infty}$ such that for some $M > 0$, $\hat{\theta}_n \in \mathcal{D}_{n,M}^{AV}$ for all $n \geq 1$ and $\{\hat{\theta}_n\}_{n=1}^{\infty}$ achieves the local asymptotic minimax risk bound in* Theorem 3.

The practical implication of Theorem 5 is that the approach of average risk that uses decisions of the form (14) has a limitation in attaining the robustness of a minimax decision, if ψ is nondifferentiable. Hence when one is concerned about the robustness of the decision, it is better to use the minimax decision than to use an average risk decision with an uninformative weight function.

4.2.2 Using a Given Inefficient Estimator of β

When a semiparametric efficient estimator of β is hard to find or compute, one may want to use an inefficient estimator $\hat{\beta}$ which is easy to compute. In this case, one may search for a functional δ such that $\delta(\hat{\beta})$ has good properties. By imposing restrictions on the space of candidate decisions, we propose an estimator that satisfies a weaker notion of optimality and yet computationlly attractive when d is large.

Suppose that we are given with $\hat{\beta}$ such that for each $t \in \mathbf{R}^d$,

$$\sup_{h \in H} \left| P_{n,h}\{\sqrt{n}(\hat{\beta} - \beta_n(h)) \leq t\} - P\{V \leq t\} \right| \to 0, \text{ as } n \to \infty,$$

for some random vector $V \in \mathbf{R}^d$. We assume that the distribution of V does not depend on $h \in H$. We consider the following collection of candidate decisions.

Definition 1: Let $\mathcal{D}_n(\hat{\beta})$ be the set of decisions of form $\hat{\theta}_n = \kappa(\delta_n(\hat{\beta}) + \hat{v}/\sqrt{n})$, where κ is a functional satisfying Assumption 2(ii) and $\delta_n : \mathbf{R}^d \to \mathbf{R}$ is a functional such that for each $h \in H$, $s \in \mathbf{R}^d$ and $\varepsilon, \eta > 0$,

$$\limsup_{n \to \infty} \sup_{s_1 \in \mathbf{R}^d : ||s - s_1|| < \eta} P_{n,h} \left\{ \left| \sqrt{n} \left\{ \delta_n \left(\hat{\beta} - \beta_n(h) + \frac{s_1}{\sqrt{n}} \right) \right\} - \delta(V + s) \right| > \varepsilon \right\} < \eta,$$

for some map $\delta : \mathbf{R}^d \to \mathbf{R}$, and for each $\varepsilon > 0$, $\sup_{h \in H} P_{n,h}\{|\hat{v} - v| > \varepsilon\} \to 0$, for some nonrandom number $v \in \mathbf{R}$.

The optimality notion based on $\mathcal{D}_n(\hat{\beta})$ is weaker than that of the previous section. Nevertheless, this decision can still be a reasonable choice in practice when a semiparametrically efficient estimator of β is hard to find or compute. One can show that under the assumptions of Theorem 3, the following analogous results hold.

Corollary 1: *Suppose that the conditions of* Theorem 3 *hold. Then for any* $\hat{\theta} \in \mathcal{D}_n(\hat{\beta})$,

$$\sup_{\delta \in \Psi} \lim_{\varepsilon \to 0} \liminf_{n \in N(\delta, \varepsilon)} \mathcal{R}_n(\hat{\theta}) \geq \inf_{v \in \mathbf{R}} \sup_{r \in \mathbf{R}^d} \mathbf{E}\left[\tau(|\psi(V + r) - \psi(r) + v|)\right].$$

This suggests the following way to obtain optimal decisions. Let $\{\hat{V}_i\}_{i=1}^L$ be i.i.d. draws from a distribution that converges to the distribution of V as $n \to \infty$. This can be immediately done when V is a centered normal random vector whose covariance matrix we can estimate consistently. Take a large number $M > 0$ and define

$$\begin{aligned} \hat{Q}_{mx}(v) &= \sup_{r \in \mathbf{R}^d : r'\mathbf{1}_d = 0} \frac{1}{L} \sum_{i=1}^L \tau\left(|\psi(\hat{V}_i + r) + v - \psi(r)| \right) \text{ and} \\ \hat{E}_{mx} &= \left\{ v \in [-M, M] : \hat{Q}_{mx}(v) \leq \inf_{v \in [-M, M]} \hat{Q}_{mx}(v) + \eta_{n,L} \right\}. \end{aligned}$$

Let $\hat{v}^*_{mx} = \frac{1}{2}\left\{\max \hat{E}_{mx} + \min \hat{E}_{mx}\right\}$. We are ready to introduce the minimax decisions:

$$\hat{\theta}_{mx} = \varphi\left(\psi(\hat{\beta}) + \frac{\hat{v}^*_{mx}}{\sqrt{n}}\right). \tag{18}$$

Therefore, using the decision of the form $\hat{\theta}_{mx}$ is still a reasonable choice although it satisfies a weaker notion of optimality.

5 Monte Carlo Simulations

5.1 Simulation Designs

In the simulation studies, we considered the following data generating process. Let $\{X_i\}_{i=1}^n$ be i.i.d random vectors in $\mathbf{R}^2$ where $X_1 \sim N(\beta, \Sigma)$, where

$$\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} 0 \\ \delta_0/\sqrt{n} \end{bmatrix} \text{ and } \Sigma = \begin{bmatrix} 2 & 1/2 \\ 1/2 & \kappa_0 \end{bmatrix}, \tag{19}$$

where κ_0 is a constant to be determined later, and δ_0 is chosen from grid points in $[0, 5]$. Note that $\delta_0 = \sqrt{n}(\beta_2 - \beta_1)$. We chose δ_0 from a grid from 0 to 5. The parameter of interest was taken to be $\Gamma(\beta) = (\varphi \circ \psi)(\beta)$ with φ being an identity map and $\psi(\beta) = \max(\beta)$. When δ_0 is close to zero, it means that β_1 and β_2 are close to the kink points of $\Gamma(\beta)$. However, when δ_0 is away from zero, $\Gamma(\beta)$ becomes more like a regular parameter (i.e., β_2 in this simulation set-up). We take $\tilde{\beta} = \frac{1}{n}\sum_{i=1}^n X_i$ as the estimator of β. As for the finite sample risk, we adopted the squared error loss and considered the following:

$$\mathbf{E}\left[\left(\hat{\theta} - \Gamma(\beta)\right)^2\right],$$

where $\hat{\theta}$ is a candidate estimator. We evaluated the risk using Monte Carlo simulations. The sample size was 300. The Monte Carlo simulation number was set to be 500.

In the simulation study, we investigate the finite sample risk profile of decisions by varying δ_0. It is worth remembering that both the average risk decision and the minimax decision are not necessarily optimal *uniformly* over δ_0. Therefore, for some values of δ_0, there can be other decisions that strictly dominate these decisions. By investigating the risk profile for each δ_0, we can discern the characteristics of each decision.

As for the average risk decisions, we computed $\tilde{c}^*_\pi$ by applying a grid search over $c \in [-15, 15]$ with grid size 0.05, and took as $\tilde{c}^*_\pi$ one that minimizes $\tilde{Q}_\pi(c)$ defined in (12). (We

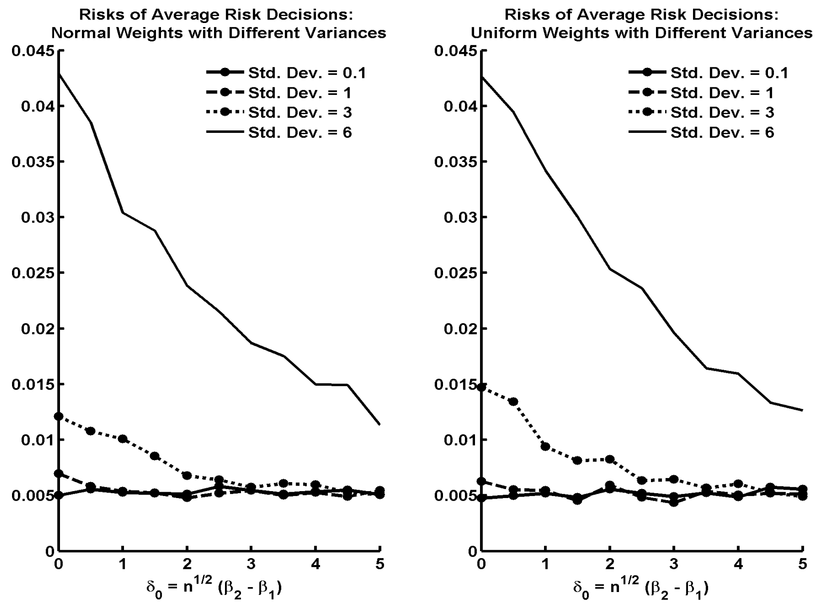

Figure 2: Instability of Average Risk Decisions: Performance of average risk decisions deteriorates near the kink points ($\delta_0 \approx 0$) when the weight function has high dispersion.

did not truncate the loss function.) To compute the average risk, we generated 2000 number of random numbers with density π (the specification of π is explained in the next subsection) and computed the sample mean of the risks.

5.2 Results

5.2.1 Instability of Average Risk Decisions near the Points of Nondifferentiability

In this subsection, we check how the quality of average risk decisions depends on the accuracy of prior information over r. We represent the accuracy of this information using weights π with different variances. First we define

$$\bar{r} \equiv \sqrt{n}(\beta - \hat{a}'\beta) = \begin{bmatrix} -\hat{a}_2\delta_0 \\ \delta_0 - \hat{a}_2\delta_0 \end{bmatrix},$$

where $\hat{a}_2$ is the second component of $\hat{a} = \hat{\Sigma}^{-1}\mathbf{1}_2/(\mathbf{1}_2'\hat{\Sigma}^{-1}\mathbf{1}_2)$ and $\hat{\Sigma}$ is the sample analogue estimator of Σ.

We consider the following weight functions: for $\sigma > 0$, let

$$\begin{aligned} \pi_1 &: \text{ the density of } \sigma N(0, I_2) + \bar{r} \\ \pi_2 &: \text{ the density of } \sqrt{12}\sigma U - \sqrt{12}\sigma/2 + \bar{r}, \end{aligned}$$

where U is a random vector in $\mathbf{R}^2$ whose entries are independent $Uniform[0, 1]$. The parameter σ represents the standard deviation of π_1 and π_2. The magnitude of σ hence represents the accuracy of prior information. In this exercise, we mainly focus on the role of σ while having π_1 and π_2 centered at the correct value of $\bar{r}$. (Later we will investigate its robustness property when the weight functions are not centered around the true value $\bar{r}$.) The variance parameter κ_0 in (19) was set to be 4.

Figure 1 reports the finite sample risk of average risk decisions $\tilde{\theta}_\pi$ using different weight functions π_1 and π_2 with standard deviation σ chosen from $\{0.1, 1, 3, 6\}$. The x-axis represents δ_0, which governs the discrepancy between β_1 and β_2. The finite sample risk profiles for uniform weights and for normal weights are similar. When σ is small, it is as if one knows well the difference $\beta_2 - \beta_1$, and with this knowledge, the decision problem becomes like one focusing on a regular parameter. This is true as long as one has fairly accurate information of $\beta_2 - \beta_1$, regardless of what the actual difference $\beta_2 - \beta_1$ is. This is reflected by the flatness of the risk profiles with $\sigma = 0.1$. However, this is no longer the case when σ is large, say, $\sigma = 6$. In this case, it matters whether β is close to the kink points of $\Gamma(\beta)$ or not. When δ_0 is close to zero so that β is close to the kink points of $\Gamma(\beta)$, the risk (with $\sigma = 6$) is very high. On the other hand, when β is away from the kink points (i.e. δ_0 is away from zero), the risk is attenuated. This shows that the choice of σ for the weight function becomes increasingly important as β becomes close to the kink points. Hence when β is close to the points of nondifferentiability, the risk is not robust to the choice of the weight functions even if their centers are correctly chosen.

5.2.2 Advantage of Prior Information under Correctly Centered Weights

In the preliminary simulation studies of minimax decisions, we find that decisions $\tilde{\theta}_{mx}$ and $\hat{\theta}_{mx}$ do not make much difference in terms of finite sample risks in our simulation set-up. Hence as for the minimax decisions, we report only the performance of the decision $\hat{\theta}_{mx}$ that is computationally faster.

Figure 2 compares the minimax decision and the average risk decision, when the average risk decision is obtained with correctly centered weights. Recall that in the case of correctly centered weights, the weight function centers around the true value of $\sqrt{n}(\beta - a'\beta)$.

When $\sigma = 0.1$, the risk profile dominates that of minimax decision. This attests to the

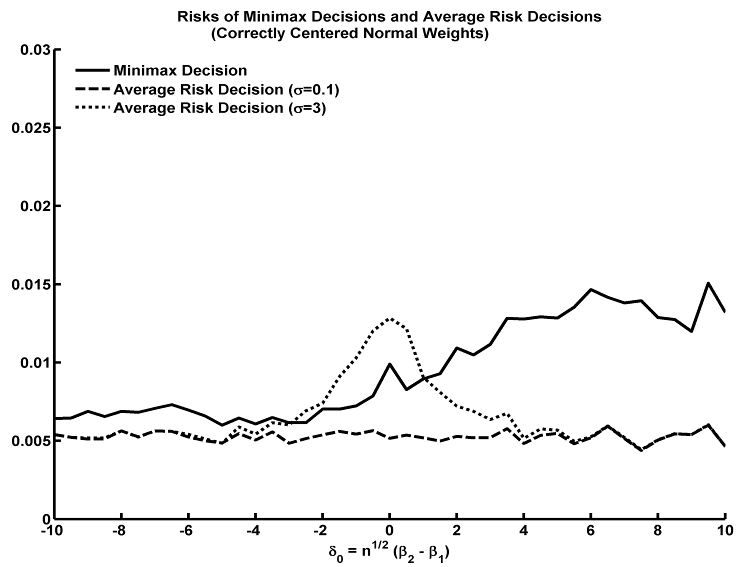

Figure 3: Advantage of Prior Information: Average risk decisions with correctly centered weights having small variance perform better than the minimax decisions.

benefit of additional information of $\beta - a'\beta$ in the decision making. When this information is subject to uncertainty so that we have now $\sigma = 3$ or 6, the average risk decisions do not dominate the minimax decision uniformly over δ_0. As shown in Figure 1, this is because the average risk decisions behave unstably with $\beta - a'\beta$ close to zero.

5.2.3 Nonrobustness of Average Risk Decisions with Weights with Misspecified Centers

The study of average risk decisions so far has assumed that the weights have correctly specified centers. The question that we ask here is whether the performance of the average risk decisions is robust to the misspecification of the centers, and whether the performance can be made robust by choosing π with high variance. The design with misspecified centers places the center of the weight π away from the true value of $\beta - a'\beta$.

The results are shown in Figure 3. The left panel shows results with the center $[0, \gamma]'$ of the weight π set to be $[0, 0.1]'$, i.e. close to the kink points of $\psi(\beta)$ and the right panel results with the center $[0, \gamma]'$ of the weight equal to $[0, 2]'$ away from zero. In both cases, the risk profile of the average risk decision performs conspicuously worse than the minimax approach except for certain local areas of δ_0. Note that the performance is not quite robustified even if we increase σ from 0.1 to 3. In other words, using highly uninformative weight function

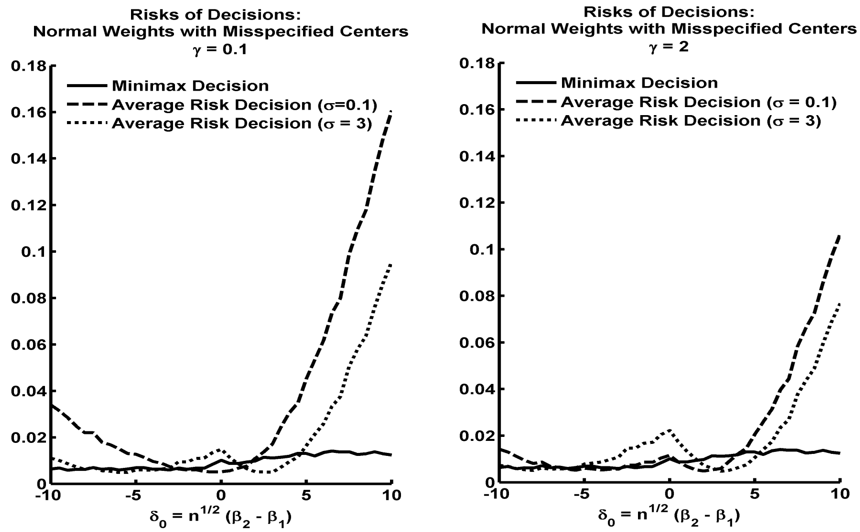

Figure 4: Nonrobustness of Average Risk Decisions: Average risk decisions with weights having misspecified centers are not robust to δ_0, while the minimax decisions show robustness.

does not alleviate the problem of nonrobustness. When we increase σ further, the risk profile of the average risk decision deteriorates further around the kink points of $\psi(\beta)$ over a larger area, preventing the decision from achieving robustness. This performance of average risk decisions makes sharp contrast with the minimax decisions. Regardless of whether the data generating process is close to the kink points or not, the finite sample risk profile shows the stable performance of the minimax decision. This result is consistent with what we found from Theorem 5.

5.2.4 Minimax Decision and Bias Reduction

It is well-known that the estimator of the type $\max(\hat{\beta})$ is asymptotically biased and many researches have proposed bias-reduction methods to address this problem. (e.g. Manski and Pepper (2000), Haile and Tamer (2003), and Chernozhukov, Lee, and Rosen (2009)). However, it is not yet clear whether a bias reduction method does the estimator harm or good from a decision-theoretic point of view, when $\theta = \Gamma(\beta)$ for nondifferentiable map Γ.

In this section, we consider estimators obtained through certain primitive methods of bias reduction and compare their properties with the minimax decision proposed in this paper. In our simulation set-up, the term $b_{F,n} = \mathbf{E}\left[\max\{X_{11} - \beta_1, X_{12} - \beta_2\}\right]$ becomes the asymptotic bias of the estimator $\max(\hat{\beta})$ when $\beta_1 = \beta_2$. One may consider the following estimator of

$b_{F,n}$:

$$\hat{b}_{F,n} = \frac{1}{L}\sum_{i=1}^{L} \max\left(\hat{\Sigma}^{1/2}\xi_i\right),$$

where ξ_i is drawn i.i.d. from $N(0, I_2)$. This adjustment term $\hat{b}_{F,n}$ is fixed over different values of $\beta_2 - \beta_1$ (in large samples). Since the bias of $\max(\hat{\beta})$ becomes prominent only when β_1 is close to β_2, one may consider performing bias adjustment only when the estimated difference $|\beta_2 - \beta_1|$ is close to zero. In the simulation study, we also consider the following estimated adjustment term:

$$\hat{b}_{S,n} = \left(\frac{1}{L}\sum_{i=1}^{L} \max\left(\hat{\Sigma}^{1/2}\xi_i\right)\right) 1\left\{\left|\hat{\beta}_2 - \hat{\beta}_1\right| < 1.7/n^{1/3}\right\}.$$

Then, we compare the following two estimators with the minimax decision $\hat{\theta}_{mx}$:

$$\begin{aligned}\hat{\theta}_F &= \max(\hat{\beta}) - \hat{b}_{F,n}/\sqrt{n} \text{ and} \\ \hat{\theta}_S &= \max(\hat{\beta}) - \hat{b}_{S,n}/\sqrt{n}.\end{aligned}$$

We call $\hat{\theta}_F$ the estimator with fixed bias-reduction and $\hat{\theta}_S$ the estimator with selective bias-reduction. The results are reported in Figure 5.

The finite sample risks of $\hat{\theta}_F$ are better than the minimax decision $\hat{\theta}_{mx}$ only around $\delta_0 = 0$. The bias reduction using $\hat{b}_{F,n}$ improves the estimator's performance in this case. However, for other values of δ_0, the bias correction does more harm than good because it lowers the bias when it is better not to. This is seen in the right-hand panel of Figure 5 which presents the finite sample bias of the estimators. With δ_0 close to zero, the estimator with fixed bias-reduction eliminates the bias almost entirely. However, for other values of δ_0, this bias correction induces negative bias, deteriorating the risk performances.

The estimator $\hat{\theta}_S$ with selective bias-reduction is designed to be hybrid between the two extremes of $\hat{\theta}_F$ and $\hat{\theta}_{mx}$. When $\beta_2 - \beta_1$ is estimated to be close to zero, the estimator performs like $\hat{\theta}_F$ and when it is away from zero, it performs like $\max(\hat{\beta})$. As expected, the bias of the estimator $\hat{\theta}_S$ is better than that of $\hat{\theta}_F$ while successfully eliminating nearly the entire bias when δ_0 is close to zero. Nevertheless, it is remarkable that the estimator shows highly unstable finite sample risk properties overall. When δ_0 is away from zero and around 3 to 7, the performance is worse than the other estimators. This result illuminates the fact that successful reduction of bias does not always imply a better risk performance.

The minimax decision shows finite sample risks that are robust over the values of δ_0. In fact, the estimated bias adjustment term $\hat{v}^*_{mx}$ of the minimax decision is close to zero. This

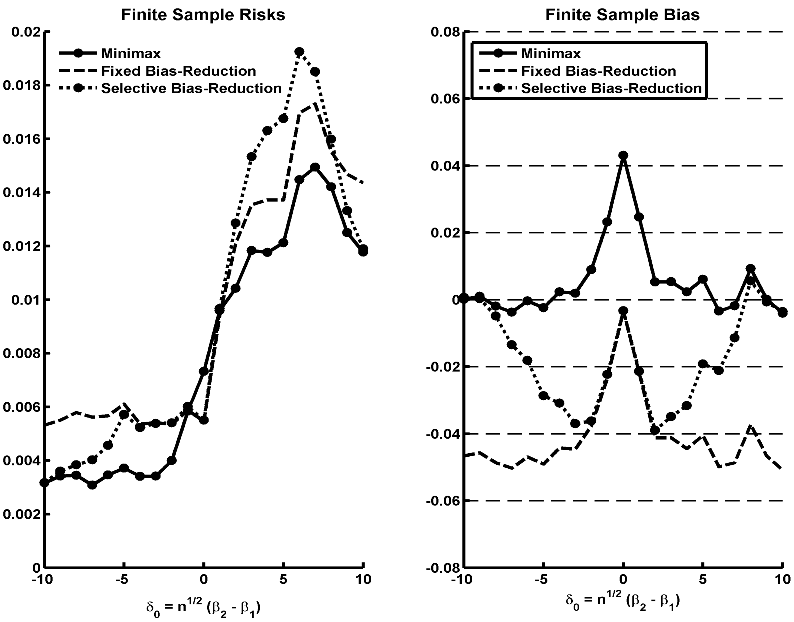

Figure 5: Comparison of the Minimax Decisions with Estimators obtained through Bias-Reduction Methods

means that the estimator $\hat{\theta}_{mx}$ involves almost zero bias adjustment, due to the concern for its robust performance. In terms of finite sample bias, the minimax estimator suffers from a substantially positive bias as compared to the other two estimators, when δ_0 is close to zero. The minimax decision tolerates this bias because by doing so, it can maintain robust performance for other cases where bias reduction is not needed. The minimax estimator is ultimately concerned with the overall risk properties, not just a bias component of the estimator, and as the left-hand panel of Figure 5 shows, it performs more reliably over various values of δ_0 relative to the other two estimators.

6 Conclusion

This paper has investigated the problem of optimal estimation of certain nonregular parameters, when a nonregular parameter is a nondifferentiable transform of a regular parameter. This paper demonstrates that we can define and find an average risk decision and a minimax decision, modifying the standard local asymptotic minimax theorem. While these results are new to the best of the author's knowledge, they nonetheless fall short of providing a complete picture of optimal estimation of nonregular parameters in general.

One interesting finding of this paper is that when the functional ψ is nondifferentiable (within the context of this paper), there exists no weight function that makes the average risk decision of the form in the paper a minimax decision. This seems to suggest the divergence of the Bayesian approach and the frequentist (or minimax) approach in the case of nonregular parameters. If this divergence is indeed a general phenomenon, it is conjectured that minimax decisions for nonregular parameters in this paper are not asymptotically admissible, which eventually means that one may be able to obtain other minimax decisions that improve on the decisions given in this paper.[4] This issue is left to a future research.

7 Appendix: Mathematical Proofs

We begin by presenting a lemma which is a generalization of Lemma A5 of Song (2009). Let $[-\infty, \infty]$ and $[-\infty, \infty]^d$ be one-point compactifications of $\mathbf{R}$ and $\mathbf{R}^d$ respectively. Convergence in distribution $\to_d$ in the proof is viewed as in the one-point compactifications. We assume the environment of Theorem 1. Choose $\{h_i\}_{i=1}^m$ from an orthonormal basis $\{h_i\}_{i=1}^\infty$ of H. For $p \in \mathbf{R}^m$, we consider $h(p) \equiv \Sigma_{i=1}^m p_i h_i$ so that $\dot{\beta}_j(h(p)) = \sum_{i=1}^m \dot{\beta}_j(h_i) p_i$, where $\dot{\beta}_j$ is the j-th element of $\dot{\beta}$. Given a $d \times k$ full column rank matrix ($d \geq k$) Λ, let $\bar{\beta}$ and $\bar{\gamma}$ be $m \times d$ and $m \times k$ matrices such that

$$\bar{\beta} \equiv \begin{bmatrix} \dot{\beta}_1(h_1) & \dot{\beta}_2(h_1) & \cdots & \dot{\beta}_d(h_1) \\ \dot{\beta}_1(h_2) & \dot{\beta}_2(h_2) & \cdots & \dot{\beta}_d(h_2) \\ \vdots & \vdots & & \vdots \\ \dot{\beta}_1(h_m) & \dot{\beta}_2(h_m) & \cdots & \dot{\beta}_d(h_m) \end{bmatrix} \text{ and } \bar{\gamma} \equiv \begin{bmatrix} \dot{\beta}(h_1)'\Lambda \\ \dot{\beta}(h_2)'\Lambda \\ \vdots \\ \dot{\beta}(h_m)'\Lambda \end{bmatrix}. \tag{20}$$

We also define $\bar{\zeta} \equiv (\zeta(h_1), \cdots, \zeta(h_m))'$, where ζ is the Gaussian process that appears in Assumption 3. We assume that $m \geq d$ and $\bar{\beta}$ is full column rank. We fix $\lambda > 0$ and let A_λ be a $m \times 1$ random vector with distribution $N(0, I_m/\lambda)$ and let $F_{\lambda,q}(\cdot)$ be the cdf of $\bar{\Sigma} A_\lambda + \bar{\mu}_q$, where

$$\bar{\mu}_q \equiv \bar{\gamma}(\bar{\gamma}'\bar{\gamma})^{-1} q \text{ and } \bar{\Sigma} \equiv I_m - \bar{\gamma}(\bar{\gamma}'\bar{\gamma})^{-1}\bar{\gamma}',$$

and $q \in \mathbf{R}^k$. Then, it is easy to check that for almost all realizations of A_λ, for each $j = 1, \cdots, k$,

$$\Lambda'\dot{\beta}\left(\left(\bar{\Sigma} A_\lambda + \bar{\mu}_q\right)' \bar{h}\right) = q,$$

[4] Note that in the case of estimating truncated normal means, Moors (1981) showed that the constrainted MLE is not admissible. Charras and Eeden (1991) established general conditions that so-called "boundary" estimators are inadmissible.

where $\bar{h} = (h_1, \cdots, h_m)'$. Suppose that $\hat{\beta}$ is a sequence of estimators such that for each $h \in H$,

$$V_n^h \equiv \sqrt{n}\left\{\hat{\beta} - \beta_n(h)\right\} \to_d \mathcal{L}^h,$$

where $\mathcal{L}^h$ is a potentially deficient distribution. Finally let $Z_\lambda^{(m)}(q) \in \mathbf{R}^d$ be a random vector distributed as $N(\bar{\beta}'(I_m - \bar{\Sigma}\bar{\Sigma}_\lambda^{-1}\bar{\Sigma})\bar{\mu}_q, \bar{\beta}'\bar{\Sigma}\bar{\Sigma}_\lambda^{-1}\bar{\Sigma}\bar{\beta})$, where $\bar{\Sigma}_\lambda = \bar{\Sigma} + \lambda I_m$. The following result is a conditional version of the convolution theorem that appears in Theorem 2.2 of van der Vaart (1989). (See also Theorem 2.7 of van der Vaart (1988).)

LEMMA A1: (i) *For any $\lambda > 0$ and $q \in \mathbf{R}^k$,*

$$\int \mathcal{L}^{h(p)} dF_{\lambda,q}(p) = \mathcal{Z}_{\lambda,q}^{(m)} * \mathcal{M}_{\lambda,q}^{(m)},$$

where $\mathcal{Z}_{\lambda,q}^{(m)}$ denotes the distribution of $Z_\lambda^{(m)}(q)$, $\mathcal{M}_{\lambda,q}^{(m)}$ a potentially deficient distribution on $\mathbf{R}^d$, and $*$ the convolution of distributions.
(ii) *Furthermore, as first $\lambda \to 0$ and then $m \to \infty$, $Z_\lambda^{(m)}(q)$ weakly converges to the conditional distribution of Z given $\Lambda' Z = q$.*

PROOF: The proof is essentially the same as that of Lemma A5 of Song (2009). ■

We assume the situation of Lemma A1. Suppose that $\hat{\theta} \in \mathbf{R}$ is a sequence of estimators such that along $\{P_{n,0}\}$

$$\begin{aligned}
\sqrt{n}\{\hat{\theta} - \psi(\beta_n(0))\} &\to_d V \text{ and} \\
\sqrt{n}\{\hat{\theta} - \psi(\beta_n(h))\} &\to_d V - \psi(\dot{\beta}(h) + \beta_\psi)
\end{aligned}$$

for some nonstochastic vector $\beta_\psi \in [-\infty, \infty]^d$ such that $\psi(\beta_\psi) = 0$, and $V \in [-\infty, \infty]^d$ is a random vector having a potentially deficient distribution. Let $F_\lambda(\cdot)$ be the cdf of A_λ. Let $\mathcal{L}_\psi^h$ be the limiting distribution of $\sqrt{n}\{\hat{\theta} - \psi(\beta_n(h))\}$ in $[-\infty, \infty]^d$ along $\{P_{n,h}\}$ for each $h \in H$. Then the following holds.

LEMMA A2: *For any $\lambda > 0$, the distribution $\int \mathcal{L}_\psi^{h(p)} dF_\lambda(p)$ is equal to that of $\psi(Z_\lambda^{(m)}(q) + W_\lambda^{(m)}(q) + \beta_\psi)$, where $W_\lambda^m(q) \in [-\infty, \infty]^d$ is a random vector having a potentially deficient distribution independent of $Z_\lambda^{(m)}(q)$.*

PROOF: Applying Le Cam's third lemma, we find that for all $B \in \mathcal{B}_1$, the Borel σ-field of

$[-\infty, \infty]$,

$$\begin{aligned}
\mathcal{L}_{\psi}^{h(p)}(B) &= \int \mathbf{E}\left[1_B(v - \psi(\bar{\beta}'p + \beta_\psi))e^{p'\bar{\zeta} - \frac{1}{2}||p||^2}\right] d\mathcal{L}_{\psi}^0(v) \\
&= \int \mathbf{E}\left[1_{\psi^{-1}(-B)}(\bar{\beta}'p + \beta_\psi - v)e^{p'\bar{\zeta} - \frac{1}{2}||p||^2}\right] d\mathcal{L}_{\psi}^0(v),
\end{aligned}$$

where $\psi^{-1}(-B) = \{x \in [-\infty, \infty]^d : \psi(x) \in -B\}$. Following the proof of Theorem 2.7 of van der Vaart (1988) and going through some algebra, we obtain the wanted result. ■

Proof of Theorem 1: We first solve for the case where φ is an identity map. Suppose that $\hat{\theta} \in \mathbf{R}$ is a sequence of estimators. By Prohorov's Theorem, for any subsequence of $\{n\}$, there exists a further subsequence $\{n'\}$ such that along $\{P_{n',0}\}$

$$\begin{aligned}
\sqrt{n'}\{\hat{\theta} - a'\beta_{n'}(0)\} &\to_d V_a, \\
\sqrt{n'}\{\hat{\theta} - a'\beta_{n'}(h)\} &\to_d V_a - a'\dot{\beta}(h), \\
\sqrt{n'}\{\beta_n(0) - a'\beta_{n'}(0)\} &\to \beta_a,
\end{aligned}$$

for some nonstochastic vector $\beta_a \in [-\infty, \infty]^d$ such that $a'\beta_a = 0$, and a random vector $V_0 \in [-\infty, \infty]^d$ having a potentially deficient distribution, and along $\{P_{n',h}\}$ for each $h \in H$,

$$\sqrt{n'}\{\hat{\theta} - a'\beta_{n'}(h)\} \to_d V_{h,a}$$

for some random vector $V_{h,a}$ in $[-\infty, \infty]^d$. For the rest of the proofs, we focus on such a subsequence and write n instead of cumbersome n'. Let $\dot{\Delta}_{n,a}(h) \equiv \sqrt{n}\{\beta_n(h) - a'\beta_n(h)\}$. Then,

$$\dot{\Delta}_{n,a}(h) \to \dot{\Delta}_a(h),$$

where $\dot{\Delta}_a(h) \equiv \dot{\beta}(h) - a'\dot{\beta}(h) + \beta_a$. We define for $r \in \mathbf{R}^d$,

$$H(r) = \left\{h \in H : \dot{\Delta}_a(h) = r\right\}.$$

Without loss of generality, we assume that a_1, the first element of a is not zero, and let $a = [a_1, a_2']'$ where $a_1 \in \mathbf{R}$ and $a_2 \in \mathbf{R}^{d-1}$ and similarly write $r = [r_1, r_2']'$ and $\beta_a = [\beta_{a,1}, \beta_{a,2}']'$. Define $A_1 = I_d - \mathbf{1}_d a'$. From the fact that $a'\mathbf{1}_d = 1$, it turns out that the restriction $A_1 Z = r - \beta_a$ is equivalent to the restriction that

$$AZ = r_2 - \beta_{a,2}$$

for $A = -(I_{d-1} - \mathbf{1}_{d-1}a_2')[\mathbf{1}_{d-1}; -I_{d-1}]$.

Using similar arguments in the proof of Theorem 1 of Song (2009), we deduce that

$$\begin{aligned}
& \liminf_{n\to\infty} \sup_{h\in H_n^\varepsilon(r)} \int \mathbf{E}_h \left[\tau\left(\left|\sqrt{n}\{\hat{\theta} - \psi(\beta_n(h))\}\right|\right)\right] \pi(r)dr \\
= & \liminf_{n\to\infty} \sup_{h\in H_n^\varepsilon(r)} \int \mathbf{E}_h \left[\tau\left(\left|\sqrt{n}\{\hat{\theta} - a'\beta_n(h) - \psi(\beta_n(h) - a'\beta_n(h))\}\right|\right)\right] \pi(r)dr \\
\geq & \sup_{h\in H(r)} \int \mathbf{E}_h \left[\tau\left(|V_{h,a} - \psi(r)\}|\right)\right] \pi(r)dr.
\end{aligned}$$

Since $a'\beta$ is a regular parameter, we use Lemma A1 and follow the proof of Theorem 1 of Song (2009) to bound the above supremum from below by

$$\int\int \mathbf{E}\left[\tau_M\left(|a'Z + w - \psi(r)|\right)|AZ = r_2 - \beta_{a,2}\right] dF(w)\pi(r)dr, \tag{21}$$

where F is the (potentially defective) distribution. Using the joint normality of Z and computing the covariance matrix, one can easily show that $a'Z$ and AZ are independent, so that we can write the above double integral as (by Fubini Theorem)

$$\begin{aligned}
& \int\int \mathbf{E}\left[\tau_M\left(|a'Z + w - \psi(r)|\right)\right] \pi(r)drdF(w) && (22) \\
\geq & \inf_{c\in[-\infty,\infty]} \int \mathbf{E}\left[\tau_M\left(|a'Z + c - \psi(r)|\right)\right] \pi(r)dr.
\end{aligned}$$

The integral is bounded and continuous in $c \in [-\infty, \infty]$. Hence the integral remains the same when we replace $\inf_{c\in[-\infty,\infty]}$ by $\inf_{c\in\mathbf{R}}$. By increasing $M \uparrow \infty$, we establish that

$$\lim_{\varepsilon\to 0} \liminf_{n\to\infty} \int \mathcal{R}_n^\varepsilon(\hat{\theta}; r)\pi(r)dr \geq \inf_{c\in\mathbf{R}} \int \mathbf{E}\left[\tau\left(|a'Z + c - \psi(r)|\right)\right] \pi(r)dr.$$

Now, consider the general case where φ is not an identity map but a contraction map such that, without loss of generality, $\varphi(y) = y$ for all $y \in [k_0, \infty)$ for some $k_0 \in \mathbf{R}$. Fix arbitrary $s \in \mathbf{R}$ and define $s_n = s + \sqrt{n}(k_0 + \varepsilon)$. Define $H_s(r) = \{h \in H(r) : \limsup_{n\to\infty}\{\sqrt{n}a'\beta_n(h) - s_n\} < -\varepsilon\}$ and $H_n^\varepsilon(r; s) = \{h \in H_n^\varepsilon(r) : \psi(\beta_n(h)) \in [k_0, \infty)\} \cap H_s(r)$. Then,

$$\begin{aligned}
& \lim_{\varepsilon\to 0} \liminf_{n\to\infty} \int \mathcal{R}_n^\varepsilon(\hat{\theta}; r)\pi(r)dr && (23) \\
\geq & \lim_{\varepsilon\to 0} \liminf_{n\to\infty} \int \sup_{h\in H_n^\varepsilon(r)} \mathbf{E}_h \left[\tau_M\left(\left|\sqrt{n}\{\hat{\theta} - (\varphi\circ\psi)(\beta_n(h))\}\right|\right)\right] \pi(r)dr \\
\geq & \lim_{\varepsilon\to 0} \liminf_{n\to\infty} \int \sup_{h\in H_n^\varepsilon(r;s)} \mathbf{E}_h \left[\tau_M\left(\left|\sqrt{n}\{\hat{\theta} - \psi(\beta_n(h))\}\right|\right)\right] \pi(r)dr.
\end{aligned}$$

Note that $h \in H_n^{\varepsilon}(r;s)$ for all $n \geq 1$ if and only if

$$\sqrt{n}\psi(\beta_n(h)) - \sqrt{n}a'\beta_n(h) \to \psi(\dot{\Delta}_a(h)) = \psi(r).$$

This means that when r is such that $\sqrt{n}(k_0 + \varepsilon) - s_n < \psi(r)$ (i.e. $-s < \psi(r)$), for each $h \in H_s(r)$, $\psi(\beta_n(h)) \in [k_0, \infty)$ from some large n on. Hence from some large n on, $H_s(r) \subset H_n^{\varepsilon}(r;s)$, whenever r is such that $-s < \psi(r)$. We bound the last term in (23) from below by

$$\begin{aligned}
&\lim_{\varepsilon \to 0} \liminf_{n \to \infty} \int \sup_{h \in H_s(r)} \mathbf{E}_h \left[\tau_M \left(\left| \sqrt{n}\{\hat{\theta} - \psi(\beta_n(h))\} \right| \right) \right] 1\{-s < \psi(r)\} \pi(r) dr \\
\geq\ & \int \sup_{h \in H_s(r)} \mathbf{E}_h \left[\tau_M \left(|V_{h,a} - \psi(r)| \right) \right] 1\{-s < \psi(r)\} \pi(r) dr.
\end{aligned}$$

The integral is monotone increasing in s. Hence by sending $s \to \infty$, we obtain the bound

$$\int \sup_{h \in H(r)} \mathbf{E}_h \left[\tau_M \left(|V_{h,a} - \psi(r)| \right) \right] \pi(r) dr.$$

Following the previous arguments, we can obtain the wanted bound. ■

For a given $M > 0$, define

$$E_\pi = \left\{ c \in [-M, M] : Q_\pi(c) \leq \inf_{c \in [-M,M]} Q_\pi(c) \right\},$$

where $Q_\pi(c) = \int \mathbf{E}\left[\tau_M \left(|a'Z + c - \psi(r)| \right) \right] \pi(r) dr$. Define c_π^* to be such that

$$c_\pi^* = \frac{1}{2}\left\{ \max E_\pi + \min E_\pi \right\}.$$

The quantity c_π^* is a population version of $\tilde{c}_\pi^*$.

Lemma A3: *Suppose that Assumptions* 1-2 *hold. Then for any* $\varepsilon > 0$,

$$\sup_{h \in H} P_{n,h} \left\{ |\tilde{c}_\pi^* - c_\pi^*| > \varepsilon \right\} \to 0,$$

as $n \to \infty$ *and* $L \to \infty$.

Proof: Let the Hausdorf distance between the two sets E_1 and E_2 in $\mathbf{R}$ denoted by $d_H(E_1, E_2)$. First we show that $d_H(E_\pi, \tilde{E}_\pi) \to_P 0$ as $n \to \infty$ and $L \to \infty$ uniformly over $h \in H$. Let $E_\pi^{\varepsilon} = \{x \in [-M, M] : \sup_{y \in E_\pi} |x - y| \leq \varepsilon\}$. The proof can be proceeded as in the proof of Theorem 3.1 of Chernozhukov, Hong and Tamer (2007), where it suffices to show that for any $\varepsilon > 0$,

$$\begin{aligned}
&\text{(a) } \inf_{h\in H} P_{n,h}\left\{\sup_{c\in E_\pi}\tilde{Q}_\pi(c) \leq \inf_{c\in[-M,M]}\tilde{Q}_\pi(c)+\eta_{n,L}\right\} \to 1 \\
&\text{(b) } \sup_{c\in\tilde{E}_\pi} Q_\pi(c) < \inf_{c\in[-M,M]\backslash E_\pi^\varepsilon} Q_\pi(c) + o_P(1),
\end{aligned}$$

as $n \to \infty$ and $L \to \infty$, where the last term $o_P(1)$ is uniform over $h \in H$.

We first consider (a). Define

$$\begin{aligned}
\tilde{Q}_\pi(c,\bar{a}) &= \frac{1}{L}\sum_{i=1}^{L}\int \tau_M\left(|\bar{a}'\xi_i + c - \psi(r)|\right)\pi(r)dr \text{ and} \\
Q_\pi(c,\bar{a}) &= \mathbf{E}\left[\int \tau_M\left(|\bar{a}'\xi_i + c - \psi(r)|\right)\pi(r)dr\right]
\end{aligned}$$

Let $\varphi(\xi;\bar{a},c) = \int \tau_M\left(|\bar{a}'\xi + c - \psi(r)|\right)\pi(r)dr$ and $\mathcal{T}_M = \{\varphi(\cdot;\bar{a},c) : (\bar{a},c) \in S_1 \times [-M,M]\}$. The class $\mathcal{T}_M$ is uniformly bounded, and by Lemma 22(ii) of Nolan and Pollard (1987), it is Euclidean, and hence P-Donsker. Therefore

$$\sup_{(\bar{a},c)\in S_1\times[-M,M]}\{\tilde{Q}_\pi(c,\bar{a}) - Q_\pi(c,\bar{a})\} = O_P(1/\sqrt{L}) \text{ as } L \to \infty.$$

The randomness of $\tilde{Q}_\pi(c,\bar{a})$ has nothing to do with $h \in H$ because it is with regard to the simulated draws $\{\xi_i\}$. Hence the convergence above is uniform over $h \in H$.

By Assumptions 5 and 6(i), we have $\hat{a}_n = a + O_P(n^{-1/2})$ uniformly over $h \in H$. From the Lipschitz continuity of τ and ψ, we conclude that

$$\sup_{c\in[-M,M]}\{\tilde{Q}_\pi(c,\hat{a}) - Q_\pi(c)\} = O_P(1/\sqrt{L} + 1/\sqrt{n}) \text{ as } n\to\infty \text{ and } L\to\infty \tag{24}$$

uniformly over $h \in H$. From this (a) follows because $\eta_{n,L}\sqrt{n} \to \infty$ as $n \to \infty$ and $\eta_{n,L}\sqrt{L} \to \infty$ as $L \to \infty$.

We turn to (b). By (24), with probability approaching 1 uniformly over $h \in H$,

$$\begin{aligned}
\sup_{c\in\tilde{E}_\pi} Q_\pi(c) &\leq \sup_{c\in\tilde{E}_\pi}\tilde{Q}_\pi(c) \leq \sup_{c\in\tilde{E}_\pi} Q_\pi(c) + o_P(1) \\
&\leq \sup_{c\in E_\pi} Q_\pi(c) + o_P(1) \\
&< \inf_{c\in[-M,M]\backslash E_\pi^\varepsilon} Q_\pi(c) + o_P(1).
\end{aligned}$$

This completes the proof of (b). Since (a) implies $\inf_{h\in H} P_{n,h}\{E_\pi \subset \tilde{E}_\pi\} \to 1$ and (b) implies

$\inf_{h\in H} P_{n,h}\{\tilde{E}_\pi \subset E_\pi^\varepsilon\} \to 1$, we conclude that for any $\varepsilon > 0$,

$$\sup_{h\in H} P_{n,h}\left\{d_H(E_\pi, \tilde{E}_\pi) > \varepsilon\right\} \to 0, \text{ as } n \to \infty \text{ and } L \to \infty.$$

Observe that

$$\begin{aligned}
|\tilde{c}_\pi^* - c_\pi^*| &= \frac{1}{2}\left|\max \tilde{E}_\pi + \min \tilde{E}_\pi - \max E_\pi - \min E_\pi\right| \\
&= \frac{1}{2}\left|\max_{y\in\tilde{E}_\pi}\{y - \max E_\pi\} - \min_{x\in E_\pi}\left\{x - \min \tilde{E}_\pi\right\}\right| \\
&= \frac{1}{2}\left|\max_{y\in\tilde{E}_\pi}\min(y - E_\pi) - \min_{x\in E_\pi}\max\left\{x - \tilde{E}_\pi\right\}\right|.
\end{aligned}$$

We write the last term as

$$\frac{1}{2}\left|\max_{y\in\tilde{E}_\pi}\min(y - E_\pi) + \max_{x\in E_\pi}\min\left\{\tilde{E}_\pi - x\right\}\right| \le \frac{1}{2}\left\{\max_{y\in\tilde{E}_\pi} d(y, E_\pi) + \max_{x\in E_\pi} d(\tilde{E}_\pi, y)\right\},$$

where $d(y, A) = \inf_{x\in A}|y - x|$. The last term is bounded by $d_H(E_\pi, \tilde{E}_\pi)$. Hence we obtain the wanted result. ■

Proof of Theorem 2: Define $\bar{\theta}_\pi = \hat{a}'\tilde{\beta} + \tilde{c}_\pi^*/\sqrt{n}$ and let $\Delta_{n,a}(h) = \sqrt{n}\{\beta_n(h) - a'\beta_n(h)\}$ and $Z_n(h) = \sqrt{n}(\tilde{\beta} - \beta_n(h))$. Observe that

$$\begin{aligned}
&\sqrt{n}\left\{\bar{\theta}_\pi - \psi(\beta_n(h))\right\} \\
=\ &\sqrt{n}\left\{\bar{\theta}_\pi - a'\beta_n(h)\right\} - \psi(r_n(h)) \\
=\ &\sqrt{n}\hat{a}'\{\tilde{\beta} - \beta_n(h)\} + \tilde{c}_\pi^* - \psi(r_n(h)) + \sqrt{n}\,(\hat{a} - a)'\,\beta_n(h) \\
=\ &\hat{a}'Z_n(h) + \tilde{c}_\pi^* - \psi(\Delta_{n,a}(h)) + \sqrt{n}\,(\hat{a} - a)'\,\Delta_{n,a}(h).
\end{aligned}$$

The last equality uses the fact that

$$\begin{aligned}
\sqrt{n}\,(\hat{a} - a)'\,\beta_n(h) &= (\hat{a} - a)'\sqrt{n}\{\beta_n(h) - a'\beta_n(h)\} \\
&= (\hat{a} - a)'\,\Delta_{n,a}(h),
\end{aligned}$$

where the first equality follows because $\hat{a}, a \in S_1$. Since φ is a contraction map, for each

$r \in \mathbf{R}^d$, from some large n on,

$$
\begin{aligned}
& \mathcal{R}^{\varepsilon}_{n,M}(\tilde{\theta}_\pi; r) \\
= & \sup_{h \in H^{\varepsilon}_n(r)} \mathbf{E}_h \left[\tau_M(|\sqrt{n}\{\varphi(\bar{\theta}_\pi) - (\varphi \circ \psi)(\beta_n(h))\}|) \right] \\
\leq & \sup_{h \in H^{\varepsilon}_n(r)} \mathbf{E}_h \left[\tau_M(|\sqrt{n}\{\bar{\theta}_\pi - \psi(\beta_n(h))\}|) \right] \\
\leq & \sup_{h \in H^{\varepsilon}_n(r)} \sup_{r-\varepsilon \leq s \leq r+\varepsilon} \mathbf{E}_h \left[\tau_M(|\hat{a}'Z + \tilde{c}^*_\pi - \psi(s) + (\hat{a} - a)'\, s|) \right].
\end{aligned}
$$

By Assumption 6 and Lemma A3, for each $t \in \mathbf{R}$,

$$
P_{n,h}\left\{\hat{a}'Z_n(h) + \tilde{c}^*_\pi \leq t\right\} \to P\left\{a'Z + c^*_\pi \leq t\right\}
$$

uniformly over $h \in H$. Since $a'Z + c^*_\pi$ is continuous, the above convergence is uniform over $t \in \mathbf{R}$. Using this uniform convergence and the fact that τ_M is continuous and bounded,

$$
\operatorname*{limsup}_{n\to\infty} \mathcal{R}^{\varepsilon}_{n,M}(\tilde{\theta}_\pi; r) \leq \sup_{r-\varepsilon \leq s \leq r+\varepsilon} \mathbf{E}\left[\tau_M(|a'Z + c^*_\pi - \psi(s)|)\right].
$$

Using Fatou's Lemma,

$$
\begin{aligned}
& \operatorname*{limsup}_{n\to\infty} \int \mathcal{R}^{\varepsilon}_{n,M}(\tilde{\theta}_\pi; r)\pi(r)dr \\
\leq & \int \operatorname*{limsup}_{n\to\infty} \mathcal{R}^{\varepsilon}_{n,M}(\tilde{\theta}_\pi; r)\pi(r)dr \leq \int \sup_{r-\varepsilon \leq s \leq r+\varepsilon} \mathbf{E}_h\left[\tau_M(|a'Z + c^*_\pi - \psi(s)|)\right]\pi(r)dr.
\end{aligned}
$$

By sending $\varepsilon \to 0$ and using continuity of $\mathbf{E}\left[\tau_M(|a'Z + c^*_\pi - \psi(s)|)\right]$ in s,

$$
\sup_{r-\varepsilon \leq s \leq r+\varepsilon} \mathbf{E}\left[\tau_M(|a'Z + c^*_\pi - \psi(s)|)\right] \to \mathbf{E}\left[\tau_M(|a'Z + c^*_\pi - \psi(r)|)\right].
$$

We use the bounded convergence theorem to conclude that

$$
\begin{aligned}
\operatorname*{limsup}_{n\to\infty} \int \mathcal{R}^{\varepsilon}_{n,M}(\tilde{\theta}_\pi; r)\pi(r)dr & \leq \int \mathbf{E}\left[\tau_M(|a'Z + c^*_\pi - \psi(r)|)\right]\pi(r)dr \\
& = \inf_{c \in [-M,M]} \int \mathbf{E}\left[\tau_M(|a'Z + c - \psi(r)|)\right]\pi(r)dr,
\end{aligned}
$$

by the definition of c^*_π. By sending $M \to \infty$, we obtain the wanted result. ■

We introduce some notations. Define $||\cdot||_{BL}$ on the space of functions on $\mathbf{R}^d$:

$$||f||_{BL} = \sup_{x\neq y} |f(x)-f(y)|/||x-y|| + \sup_x |f(x)|.$$

For any two probability measures P and Q on $\mathcal{B}$, define

$$d_{\mathcal{P}}(P,Q) = \sup\left\{\left|\int f dP - \int f dQ\right| : ||f||_{BL}\leq 1\right\}. \tag{25}$$

Proof of Theorem 3: As in the proof of Theorem 1, we first consider the case of φ being an identity. We write

$$\begin{aligned}
& \sqrt{n}\{\hat{\theta}-\psi(\beta_n(h))\} \\
= \ & \sqrt{n}\{\hat{\theta}-\psi(\beta_n(0))\} - \sqrt{n}\psi\left(\beta_n(h)-\beta_n(0)+\beta_n(0)-\psi(\beta_n(0))\right) \\
= \ & \sqrt{n}\{\hat{\theta}-\psi(\beta_n(0))\} - \psi(\sqrt{n}(\beta_n(h)-\beta_n(0))+\beta_{n,\psi})
\end{aligned}$$

where $\beta_{n,\psi} = \sqrt{n}\{\beta_n(0)-\psi(\beta_n(0))\}$. Applying Prohorov's Theorem, we note that for any subsequence of $\{P_{n,0}\}$, there exists a further subsequence along which

$$\begin{aligned}
\sqrt{n}\{\hat{\theta}-\psi(\beta_n(0))\} &\to_d V \text{ and} \\
\sqrt{n'}\{\hat{\theta}-\psi(\beta_{n'}(h))\} &\to_d V-\psi(\dot{\beta}(h)+\beta_\psi),
\end{aligned}$$

under $h\in H$, where V is a random vector and β_ψ is a nonstochastic vector in $[-\infty,\infty]^d$. From now on, it suffices to focus only on these subsequences.

Using Lemma A2 and following the arguments of the proof of Theorem 1 of Song (2009), we obtain that

$$\begin{aligned}
\sup_{\delta\in\Psi}\lim_{\varepsilon\to 0}\liminf_{n\in N(\delta,\varepsilon)} \mathcal{R}_n(\hat{\theta}) &\geq \sup_{\delta\in\Psi}\mathbf{E}\left[\tau_M\left(|\psi(Z+W+\delta)|\right)\right] \\
&\geq \sup_{r\in[-M,M]^d}\mathbf{E}\left[\tau_M\left(|\psi(Z+W+r-\psi(r))|\right)1\left\{W\in[-M,M]^d\right\}\right].
\end{aligned}$$

The main part of the proof is the proof of the inequality : for any $M>0$,

$$\begin{aligned}
& \sup_{r\in[-M,M]^d}\mathbf{E}\left[\tau_M\left(|\psi(Z+W+r-\psi(r))|\right)1\left\{W\in[-M,M]^d\right\}\right] \\
\geq \ & \inf_{w\in\mathbf{R}^d}\sup_{r\in[-M,M]^d}\mathbf{E}\left[\tau_M\left(|\psi(Z+w+r-\psi(r))|\right)\right].
\end{aligned} \tag{26}$$

Once this inequality is established, we send $M\to\infty$ to complete the proof.

For brevity, we write

$$\begin{aligned}
&\sup_{r\in[-M,M]^d} \mathbf{E}\left[\tau_M\left(|\psi(Z+W+r-\psi(r))|\right)1\left\{W\in[-M,M]^d\right\}\right] \\
\geq\ &\sup_{r\in[-M,M]^d}\int \bar{g}(w,r)dF(w),
\end{aligned}$$

where $\bar{g}(w,r)=\mathbf{E}\left[g(Z+w+r-\psi(r))1\left\{w\in[-M,M]^d\right\}\right]$ and $g(x)=\tau_M(|\psi(x)|)$, and F denotes the cdf of W.

Take $K>0$ and let $\mathcal{R}_K=\{r_1,\cdots,r_K\}\subset[-M,M]^d$ be a finite set such that $\mathcal{R}_K$ become dense in $[-M,M]^d$ as $K\to\infty$. Define $\mathcal{F}_M$ to be the collection of distributions whose support is restricted to $[-M,M]^d$. Then $\mathcal{F}_M$ is uniformly tight. Note that

$$\int \bar{g}(w,r)dF(w)$$

is Lipschitz in r uniformly over $F\in\mathcal{F}_M$. Therefore, for any fixed $\eta>0$, we can take $\mathcal{R}_K$ independently of $F\in\mathcal{F}_M$ such that

$$\max_{r\in\mathcal{R}_K}\int\bar{g}(w,r)dF(w)\geq\sup_{r\in[-M,M]^d}\int\bar{g}(w,r)dF(w)-\eta \tag{27}$$

Since $\mathcal{F}_M$ is uniformly tight, by Theorems 11.5.4 of Dudley (2002), $\mathcal{F}_M$ is totally bounded for $d_{\mathcal{P}}$ defined in (25). Hence we fix $\varepsilon>0$ and choose $F_1,\cdots,F_N$ such that for any $F\in\mathcal{F}_M$, there exists $j\in\{1,\cdots,N\}$ such that $d_{\mathcal{P}}(F_j,F)<\varepsilon$. For F_j and F such that $d_{\mathcal{P}}(F_j,F)<\varepsilon$,

$$\left|\max_{r\in\mathcal{R}_K}\int\bar{g}(w,r)dF(w)-\max_{r\in\mathcal{R}_K}\int\bar{g}(w,r)dF_j(w)\right|\leq\max_{r\in\mathcal{R}_K}||\bar{g}(\cdot,r)||_{BL}\varepsilon.$$

Since $\bar{g}(\cdot,r)$ is Lipschitz continuous and bounded, $\max_{r\in\mathcal{R}_K}||\bar{g}(\cdot,r)||_{BL}<\infty$. We let $C_K=\max_{r\in\mathcal{R}_K}||\bar{g}(\cdot,r)||_{BL}$. Therefore,

$$\inf_{F\in\mathcal{F}_M}\max_{r\in\mathcal{R}_K}\int\bar{g}(w,r)dF(w)\geq\inf_{1\leq j\leq N}\max_{r\in\mathcal{R}_K}\int\bar{g}(w,r)dF_j(w)-C_K\varepsilon. \tag{28}$$

By Lemma 3 of Chamberlain (1987), we can select for each F_j and for each $r_k\in\mathcal{R}_K$ a multinomial distribution $G_{j,k}$ such that

$$\int\bar{g}(w,r_k)dF_j(w)\geq\int\bar{g}(w,r_k)dG_{j,k}(w)-\nu_{k,j}$$

where $\nu_{k,j}$ can be taken to be arbitrarily small. Hence

$$\inf_{F\in\mathcal{F}}\max_{r\in\mathcal{R}_K}\int\bar{g}(w,r)dF(w)\geq\inf_{1\leq j\leq N}\max_{1\leq k\leq K}\int\bar{g}(w,r_k)dG_{j,k}(w)-C_K\varepsilon-\nu_{k,j}. \quad (29)$$

Then let $\mathcal{W}_{K,N}$ be the union of the supports of $G_{j,k}$, $j=1,\cdots,N$ and $k=1,\cdots,K$. The set $\mathcal{W}_{K,N}$ is finite. Let $\mathcal{F}_{K,N}$ be the space of discrete probability measures with a support in $\mathcal{W}_{K,N}$. Then,

$$\begin{aligned}
&\inf_{1\leq j\leq N}\max_{1\leq k\leq K}\int\bar{g}(w,r_k)dG_{j,k}(w)\\
\geq\ &\inf_{G\in\mathcal{F}_{K,N}}\max_{r\in\mathcal{R}_K}\int\bar{g}(w,r)dG(w)\\
=\ &\inf_{G\in\mathcal{F}_{K,N}}\max_{r\in\mathcal{R}_K}\int\int g(z+w-\psi(r))d\Lambda_r(z)dG(w),
\end{aligned}$$

where Λ_r is the distribution of $Z+r$. For the last $\inf_{G\in\mathcal{F}_{K,N}}\max_{r\in\mathcal{R}_K}$, we regard $g(Z_1+w-\psi(r))$ as a loss function with Z_1 representing a state variable distributed by Λ_r with Λ_r parametrized by r in a finite set $\mathcal{R}_K$. We view r as the parameter of interest and the conditional distribution of Z_1+W given Z_1 as a randomized decision. The conditional distribution of Z_1+W given $Z_1=z$ is equal to the distribution of $z+W$, because W and Z_1 are independent. The distribution $G\in\mathcal{F}_{K,N}$ has a common finite support $\mathcal{W}_{K,N}$ and the distribution associated with Λ_r is atomless. Hence, by Theorem 3.1 of Dvoretzky, Wald, and Wolfowitz (1951), the last $\inf_{G\in\mathcal{F}_{K,N}}\max_{r\in\mathcal{R}_K}$ is equal to that with randomized decisions replaced by nonrandominzed decisions, enabling us to write it as

$$\begin{aligned}
&\inf_{w\in\mathcal{W}_{K,N}}\sup_{r\in\mathcal{R}_K}\int g(z+w-\psi(r))d\Lambda_r(z)\\
=\ &\inf_{w\in\mathcal{W}_{K,N}}\sup_{r\in\mathcal{R}_K}\mathbf{E}\left[\tau_M(|\psi(Z+w+r)-\psi(r)|)\right].
\end{aligned}$$

Since $\mathbf{E}\left[\tau_M(|\psi(Z+w+r)-\psi(r)|)\right]$ is continuous in w and r, we send $\nu_{k,j}\to 0$, $\varepsilon\to 0$ and then $\eta\to 0$ to conclude from (27), (28), and (29) that

$$\begin{aligned}
&\inf_{F\in\mathcal{F}_M}\sup_{r\in[-M,M]^d}\int\mathbf{E}\left[g(Z+w+r-\psi(r))\right]dF(w)\\
\geq\ &\inf_{w\in\mathcal{W}_{K,N}}\sup_{r\in[-M,M]^d}\mathbf{E}\left[\tau_M(|\psi(Z+w+r)-\psi(r)|)\right]\\
\geq\ &\inf_{w\in\mathbf{R}^d}\sup_{r\in[-M,M]^d}\mathbf{E}\left[\tau_M(|\psi(Z+w+r)-\psi(r)|)\right].
\end{aligned}$$

Therefore, we obtain (26). ■

PROOF OF THEOREM 4: Let $Q_{mx}(w) = \sup_{r\in\mathbf{R}^d} \mathbf{E}\left[\tau_M(|\psi(Z+w+r) - \psi(r)|)\right]$ and, taking a large number $M > 0$, define

$$w^*_{mx} = \frac{1}{2}\{\max E_{mx} + \min E_{mx}\}$$

where $E_{mx} = \{w \in [-M,M]^d : Q_{mx}(w) \leq \inf_{w\in[-M,M]^d} Q_{mx}(w)\}$. Since φ is a contraction map, it suffices to define

$$\bar{\theta}_{mx} = \psi(\tilde{\beta}) + \frac{\tilde{w}^*_{mx}}{\sqrt{n}}$$

and establish that

$$\lim_{M\to\infty} \sup_{\delta\in\Psi} \lim_{\varepsilon\to 0} \limsup_{n\in N(\delta;\varepsilon)} \sup_{h\in H} \mathbf{E}_h\left[\tau_M\left(\sqrt{n}\left|\bar{\theta}_{mx} - \psi(\beta_n(h))\right|\right)\right] \tag{30}$$

achieves the bound. (See the proof of Theorem 2.) First, observe that

$$\begin{aligned}
& \sup_{h\in H} \mathbf{E}_h\left[\tau_M\left(\sqrt{n}\left|\bar{\theta}_{mx} - \psi(\beta_n(h))\right|\right)\right] \\
= & \sup_{h\in H} \mathbf{E}_h\left[\tau_M\left(\sqrt{n}\left|\psi(\tilde{\beta}) + \tilde{w}^*_{mx} - \psi(\beta_n(h))\right|\right)\right] \\
= & \sup_{h\in H} \mathbf{E}_h\left[\tau_M\left(\left|\psi(\sqrt{n}\{\tilde{\beta} - \beta_n(h)\} + \tilde{w}^*_{mx} + \sqrt{n}\{\beta_n(h) - \psi(\beta_n(h))\})\right|\right)\right] \\
\leq & \sup_{h\in H} \sup_{r\in\mathbf{R}^d} \mathbf{E}_h\left[\tau_M\left(\left|\psi(\sqrt{n}\{\tilde{\beta} - \beta_n(h)\} + \tilde{w}^*_{mx} + r - \psi(r))\right|\right)\right].
\end{aligned}$$

Consistency of $\tilde{w}^*_{mx}$ for w^*_{mx} can be shown as in the proof of Lemma A3. Since $\sqrt{n}\{\tilde{\beta} - \beta_n(h)\} + \tilde{w}^*_{mx} \to_d Z + w^*_{mx}$ uniformly over h and τ_M is uniformly continuous, we deduce that the limit in (30) is bounded by

$$\begin{aligned}
& \sup_{r\in\mathbf{R}^d} \mathbf{E}\left[\tau_M\left(|\psi\left(Z + w^*_{mx} + r - \psi(r)\right)|\right)\right] \\
= & \inf_{w\in[-M,M]^d} \sup_{r\in\mathbf{R}^d} \mathbf{E}\left[\tau_M\left(|\psi\left(Z + w + r\right) - \psi(r)|\right)\right].
\end{aligned}$$

Hence the proof is complete. ■

Recall that a function $f : \mathbf{R}^d \to \mathbf{R}$ is *positive homogeneous of degree* k if for all $x \in \mathbf{R}^d$ and for all $u \geq 0$, $f(ux) = u^k f(x)$.

DEFINITION 2: (i) Define $\mathcal{A}$ to be the collection of maps $\delta : \mathbf{R}^d \to \mathbf{R}$ such that for some positive integer m,

$$\delta(x) = \sum_{j=1}^{m} c_j \delta_j(x) + c_0,$$

where c_0 and c_j's are real numbers and δ_j's are positive homogeneous functions of degree $1 \leq k_j \leq m$.
(ii) Define $\mathcal{A}_{EQ}$ to be the collection of maps $\delta : \mathbf{R}^d \to \mathbf{R}$ such that $\delta \in \mathcal{A}$ and for any $x \in \mathbf{R}^d$ and $c \in \mathbf{R}$, $\delta(x+c) = \delta(x) + c$.

Note that $\mathcal{A}$ is the affine space of positive homogeneous functions, and $\mathcal{A}_{EQ}$ is a subset of $\mathcal{A}$ that includes only *location equivariant* members of $\mathcal{A}$.

LEMMA A4: *Every function in $\mathcal{A}_{EQ}$ takes the form of $\delta_1(\cdot) + c$, where c is a constant and δ_1 is positive homogeneous of degree 1.*

PROOF : Let $\mathcal{A}_{EQ}(m)$ be the collection of maps $\delta : \mathbf{R}^d \to \mathbf{R}$ such that

$$\delta(x) = \sum_{j=1}^{m} c_j \delta_j(x) + c_0, \tag{31}$$

where c_0 and c_j's are constants and δ_j's are positive homogeneous functions of degree $1 \leq j \leq m$, and for any $x \in \mathbf{R}^d$ and $c \in \mathbf{R}$, $\delta(x+c) = \delta(x) + c$. Without loss of generality, let us assume that δ_j is homogeneous of degree j.

It suffices to show that for every positive integer m, $\mathcal{A}_{EQ}(m) = \mathcal{A}_{EQ}(1)$. We define the following property for any map δ in the form (31).

(Property A): We say δ in the form (31) satisfies Property A if for each $j = 1, \cdots, m$, there exists a constant κ_j such that for any $c \in \mathbf{R}$,

$$\delta_j(x+c) - \delta_j(x) = \kappa_j c \tag{32}$$

for all $x \in \mathbf{R}^d$, and $\sum_{j=1}^{m} \kappa_j c_j = 1$, $\kappa_j \neq 0$.

Consider the situation with $m = 1$. Then, δ_1 is positive homogeneous of degree 1. By taking $\kappa_1 = 1$, we find that δ_1 satisfies Property A due to location equivariance. Hence all the members of $\mathcal{A}_{EQ}(1)$ satisfy Property A.

Fix arbitrary positive integer m_1 such that all the δ's in $\mathcal{A}_{EQ}(m_1)$ satisfy Property A. Take $\delta \in \mathcal{A}_{EQ}(m_1 + 1)$ such that

$$\delta(x) = \sum_{j=1}^{m_1+1} c_j \delta_j(x) + c_0.$$

Choose $c \neq 0$ and write

$$\begin{aligned} c &= \delta(x+c) - \delta(x) = \sum_{j=1}^{m_1+1} c_j \left\{\delta_j(x+c) - \delta_j(x)\right\} \qquad (33) \\ &= \sum_{j=1}^{m_1} c_j \left\{\delta_j(x+c) - \delta_j(x)\right\} + c_{m_1+1}\left\{\delta_{m_1+1}(x+c) - \delta_{m_1+1}(x)\right\}. \end{aligned}$$

The first equality follows from location equivariance of δ. Since a decision of the form $\sum_{j=1}^{m_1} c_j\delta_j(x)$ belongs to $\mathcal{A}_{EQ}(m_1)$ and all the members of $\mathcal{A}_{EQ}(m_1)$ satisfy Property A,

$$\sum_{j=1}^{m_1} c_j \left\{\delta_j(x+c) - \delta_j(x)\right\} = c\sum_{j=1}^{m_1} c_j\kappa_j = c \qquad (34)$$

using the condition that $\sum_{j=1}^{m_1} c_j\kappa_j = 1$. Hence we have (from (33))

$$c_{m_1+1}\left\{\delta_{m_1+1}(x+c) - \delta_{m_1+1}(x)\right\} = 0.$$

If $c_{m_1+1} \neq 0$, δ_{m_1+1} is location invariant. However, location invariance is not possible when δ_{m_1+1} is positive homogeneous of degree $k \geq 1$. Therefore $c_{m_1+1} = 0$. We conclude that $\mathcal{A}_{EQ}(m_1) = \mathcal{A}_{EQ}(m_1+1)$. By rolling back the arguments for $\mathcal{A}_{EQ}(m_1-1)$ and $\mathcal{A}_{EQ}(m_1)$ and so on, we deduce that $\mathcal{A}_{EQ}(m) = \mathcal{A}_Q(1)$ for any $m \geq 1$. ■

LEMMA A5: *Suppose that V is a random vector in $\mathbf{R}^d$ (with a tight distribution) and $\delta \in \mathcal{A}_{EQ}$. If there exists no $c \in \mathbf{R}$ such that $\delta(x) = \psi(x) + c$ for all $x \in \mathbf{R}^d$, then for each $M > 0$,*

$$\sup_{r\in\mathbf{R}^d} \mathbf{E}\left[\tau_M(|\delta(V+r) - \psi(r)|)\right] = M.$$

PROOF: Fix $m > 0$ and choose $\delta \in \mathcal{A}_{EQ}(m)$. By Lemma A4, we can write

$$\delta(x) = \delta_1(x) + c_1$$

for some $c_1 \in \mathbf{R}$, where $\delta_1(x)$ is positive homogeneous of degree 1. Take any $M > 0$. Under the assumption of the lemma, for every $c \in \mathbf{R}$, there exists $\bar{r} \in \mathbf{R}^d$ such that $\delta(\bar{r}) \neq \psi(\bar{r}) + c$. We take $c = c_1$ and $\bar{r} \in \mathbf{R}^d$ such that $\delta_1(\bar{r}) \neq \psi(\bar{r})$. Let $\eta = \delta_1(\bar{r}) - \psi(\bar{r})$. Using positive homogeneity of δ_1 with degree 1, we can write the supremum in Lemma A5 as

$$\begin{aligned} &\sup_{r\in\mathbf{R}^d} \mathbf{E}\left[\tau_M\left(|\delta_1(V+ur) - \psi(ur)|\right)\right] \\ = &\sup_{r\in\mathbf{R}^d} \mathbf{E}\left[\tau_M\left(u\left|\delta_1(V/u+r) - \psi(r)\right|\right)\right] \text{ for any } u > 0. \end{aligned}$$

Then choosing $r = \bar{r}$, we bound the last term from below by

$$
\begin{aligned}
& \mathbf{E}\left[\tau_M(u\left|\psi(V/u+\bar{r}) - \psi(\bar{r}) + \eta\right|)\right] \\
\geq\ & \mathbf{E}\left[\tau_M(\max\{u\left|\eta\right| - u\left|\psi(V/u+\bar{r}) - \psi(\bar{r})\right|, 0\})\right] \\
\geq\ & \tau_M\left(\max\{u\left|\eta\right| - u\mathrm{sup}_{z\in[-K,K]}\left|\psi(z/u+\bar{r}) - \psi(\bar{r})\right|, 0\}\right) P\left\{||V|| \leq K\right\} \\
& -\mathbf{E}\left[\tau_M(\max\{u\left|\eta\right| - u\left|\psi(V/u+\bar{r}) - \psi(\bar{r})\right|, 0\})1\{||V|| > K\}\right],
\end{aligned}
$$

for any $K > 0$. As for the last term, as $K \to \infty$,

$$
\begin{aligned}
& \mathbf{E}\left[\tau_M(\max\{u\left|\eta\right| - u\left|\psi(V/u+r) - \psi(r)\right|, 0\})1\{||V|| > K\}\right] \\
\leq\ & MP\{||V|| > K\} \to 0.
\end{aligned}
$$

Since ψ is uniformly continuous, we have as $u \to \infty$,

$$
u\mathrm{sup}_{z\in[-K,K]}\left|\psi(z/u+\bar{r}) - \psi(\bar{r})\right| = o(u) \text{ and } u\left|\eta\right| \to \infty.
$$

Hence from some large u on,

$$
\tau_M\left(\max\{u\left|\eta\right| - o(u), 0\}\right) P\left\{||V|| \leq K\right\} = MP\left\{||V|| \leq K\right\},
$$

because $\tau(y) \to \infty$ as $y \to \infty$. By sending $K \to \infty$, we obtain the wanted result. ■

Proof of Theorem 5: Choose $\hat{\theta}_n \in \mathcal{D}_{n,M}^{AV}$ such that $\hat{\theta}_n$ is asymptotically equivalent to $\varphi(a'\tilde{\beta} + c_\pi^*/\sqrt{n})$ for some c_π^*. From the proofs of Theorems 1 and 3, we can show that

$$
\sup_{\delta\in\Psi} \lim_{\varepsilon\to 0} \liminf_{n\in N(\delta;\varepsilon)} \mathcal{R}_n(\hat{\theta}) \geq \sup_{r\in\mathbf{R}^d} \mathbf{E}\left[\tau_M\left(\left|a'Z + c_\pi^* + r - \psi(r)\right)\right|\right)\right].
$$

Since $a'Z + c_\pi^*$ is not of the form $\psi(Z) + c$ when ψ is not differentiable, Lemma A5 implies that the last supremum is equal to M. Hence the local asymptotic minimax risk bound is infinity by sending $M \to \infty$ in the lemma, and the decision $\hat{\theta}_n$ cannot achieve the minimax risk bound that is achieved by $\tilde{\theta}_{mx}$. ■

References

[1] Andrews, D. W. K. (1999): "Estimation when a parameter is on the boundary," *Econometrica*, 67, 1341-1383.

[2] Andrews, D. W. K. (2001): “Testing when a parameter is on the boundary of the maintained hypothesis," *Econometrica*, 69, 683-734.

[3] Begun, J. M., W. J. Hall, W-M., Huang, and J. A. Wellner (1983): “Information and asymptotic efficiency in parametric-nonparametric models,” *Annals of Statistics*, 11, 432-452.

[4] Bickel, P. J. (1981): "Minimax estimation of the mean of a normal distribution when the parameter space is restricted," *Annals of Statistics*, 9, 1301-1309.

[5] Blumenthal, S. and A. Cohen (1968a): "Estimation of the larger translation parameter," *Annals of Mathematical Statistics*, 39, 502-516.

[6] Blumenthal, S. and A. Cohen (1968b): "Estimation of two ordered translation parameters," *Annals of Mathematical Statistics*, 39, 517-530.

[7] Casella G. and W. E. Strawderman (1981): "Estimating a bounded normal mean," *Annals of Statistics,* 9, 870-878.

[8] Charras, A. and C. van Eeden (1991): "Bayes and admissibility properties of estimators in truncated parameter spaces," *Canadian Journal of Statistics,* 19, 121-134.

[9] Chamberlain, G. (1987): "Asymptotic efficiency in estimation with conditional moment restrictions," *Journal of Econometrics* 34, 305-334.

[10] Chernozhukov, V. and H. Hong. (2004): "Likelihood estimation and inference in a class of nonregular econometric models," *Econometrica,* 72, 1445-1480.

[11] Chernozhukov, V., H. Hong and E. Tamer (2007): "Estimation and confidence regions for parameter sets in econometric models," *Econometrica,* 75, 1243-1284.

[12] Chernozhukov, V., S. Lee and A. Rosen (2009): "Intersection bounds: estimation and inference," Cemmap Working Paper, CWP 19/09.

[13] Doss, H. and J. Sethuraman (1989): "The price of bias reduction when there is no unbiased estimate," *Annals of Statistics,* 17, 440-442.

[14] Dudley, R. M. (2002): *Real Analysis and Probability*, Cambridge University Press, New York.

[15] Dunford, N. and J. T. Schwartz (1988): *Linear Operators. Part I: General Theory*, Interscience, New York.

[16] Dvoretsky, A., A. Wald. and J. Wolfowitz (1951): "Elimination of randomization in certain statistical decision procedures and zero-sum two-person games," *Annals of Mathematical Statistics* 22, 1-21.

[17] Fan, Y. and J. Wu (2009), "Partial identification of the distribution of treatment effects in switching regime models and its confidence sets," forthcoming in *Review of Economic Studies*.

[18] Ghosh, J. K., S. Ghosal, and T. Samanta (1994): "Stability and convergence of posterior in nonregular problems," *Statistical Decision Theory and Related Topics V*, eds. by S. S. Gupta and J. O. Berger, 183-199, Springer, New York.

[19] Giné, E. and J. Zinn (1991), "Gaussian characterization of uniform Donsker classes of functions," *Annals of Probability*, 19, 758-782.

[20] Haile, P. A. and E. Tamer (2003): "Inference with an incomplete model of English auctions," *Journal of Political Economy*, 111, 1-51.

[21] Hirano, K. and J. Porter (2003): "Asymptotic efficiency in parametric structural models with parameter-dependent support," *Econometrica*, 71, 1307–1338.

[22] Hirano, K. and J. Porter (2009a): "Asymptotics for statistical treatment rules," *Econometrica*, 77, 1683–1701.

[23] Hirano, K. and J. Porter (2009b): "Impossibility results for nondifferentiable functionals," Working Paper.

[24] Hoderlein, S. and J. Stoye (2009): "Revealed preferences in a heterogeneous population," Working Paper.

[25] Lehmann E. L. (1986): *Testing Statistical Hypotheses*, 2nd Ed. Chapman&Hall, New York.

[26] Lehmann E. L. and G. Casella (1998): *Theory of Point Estimation*, Springer, New York.

[27] Lovell, M. C. and E. Prescott (1970): "Multiple regression with inequality constraints: pretesting bias, hypothesis testing, and efficiency," *Journal of the American Statistical Association*, 65, 913-915.

[28] Manski, C. F. (1989): "Anatomy of the selection problem," *Journal of Human Resources*, 24, 343-360.

[29] MANSKI, C. F. (1990): "Nonparametric bounds on treatment effects," *American Economic Review*, 80, 319-323.

[30] MANSKI, C. F. (1997): "Monotone treatment response," *Econometrica*, 65, 1311-1334.

[31] MANSKI, C. F., AND J. PEPPER (2000): "Monotone instrumental variables: with an application to the returns to schooling," *Econometrica*, 68, 997-1010.

[32] MANSKI, C. F. (2008): *Identification for Prediction and Decision*, Harvard University Press.

[33] MOORS, J. J. A. (1981): "Inadmissibility of linearly invariant estimators in truncated parameter spaces," *Journal of the American Statistical Association*, 76, 910-915.

[34] NOLAN, D. AND D. POLLARD (1987): "U-processes: rates of convergence," *Annals of Statistics*, 15, 780-799.

[35] PFLUG, G. CH. (1983): "The limiting log-likelihood process for discontinuous density families," *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 64, 15-35.

[36] PLOBERGER, W. AND P. C. B. PHILLIPS (2010): "Optimal estimation under nonstandard conditions," Cowles Foundation Discussion Paper 1748.

[37] SONG, K. (2009): "Point decisions for interval-identified parameters," Working Paper.

[38] STRASSER, H. (1985): *Mathematical Theory of Statistics*, Walter de Gruyter, New York.

[39] VAN DER VAART, A. W. (1988): *Statistical Estimation in Large Sample Spaces*, CWI Tracts 44, Amsterdam.

[40] VAN DER VAART, A. W. (1989): "On the asymptotic information bound," *Annals of Statistics* 17, 1487-1500.

[41] VAN DER VAART, A. W. (1991): "On differentiable functionals," *Annals of Statistics* 19, 178-204.

[42] VAN DER VAART, A. W. AND J. A. WELLNER (1996): *Weak Convergence and Empirical Processes*, Springer-Verlag, New York.