LOCAL ASYMPTOTIC MINIMAX ESTIMATION OF NONREGULAR PARAMETERS WITH TRANSLATION-SCALE EQUIVARIANT MAPS

KYUNGCHUL SONG

ABSTRACT. When a parameter of interest is defined to be a nondifferentiable transform of a regular parameter, the parameter does not have an influence function, rendering the existing theory of semiparametric efficient estimation inapplicable. However, when the nondifferentiable transform is a known composite map of a continuous piecewise linear map with a single kink point and a translation-scale equivariant map, this paper demonstrates that it is possible to define a notion of asymptotic optimality of an estimator as an extension of the classical local asymptotic minimax estimation. This paper establishes a local asymptotic risk bound and proposes a general method to construct a local asymptotic minimax decision.

KEY WORDS. Nonregular Parameters; Translation-Scale Equivariant Transforms; Semiparametric Efficiency; Local Asymptotic Minimax Estimation.

AMS CLASSIFICATION. 62C05, 62C20.

1. INTRODUCTION

This paper investigates the problem of optimal estimation of a parameter $\theta \in \mathbf{R}$ which takes the following form:

$$\theta = (f \circ g)(\boldsymbol{\beta}), \tag{1.1}$$

where $\boldsymbol{\beta} \in \mathbf{R}^d$ is a regular parameter for which a semiparametric efficiency bound is well defined, g is a translation-scale equivariant map, and f is a continuous piecewise linear map with a single kink point.

Examples abound, including $\max\{\beta_1, \beta_2, \beta_3\}$, $\max\{\beta_1, 0\}$, $|\beta_1|$, $|\max\{\beta_1, \beta_2\}|$, etc., where $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3)$ is a regular parameter, i.e., a parameter which is differentiable in the underlying probability. For example, one might be interested in the absolute difference between two conditional means $\theta = |\mathbf{E}[Y|X = x_1] - \mathbf{E}[Y|X = x_2]|$, or the maximum between two different treatment effects $\theta = \max_{1 \leq j \leq J} \beta_j$ with $\beta_j = \mathbf{E}[Y|X = x, D = j] - \mathbf{E}[Y|X = x, D = 0]$, where Y is an outcome variable, $D \in \{0, 1, \cdot \cdot \cdot, J\}$ treatment type (with 0 representing no treatment), and X is a discrete covariate vector. Another example involves the object of interest bounded by unknown quantities β_1 and β_2, forming a common bound $\min\{\beta_1, \beta_2\}$. Such a bound frequently arises in economics literature (e.g. Manski and Pepper (2000) for returns to education and Haile and Tamer (2003) for bidders' valuations in English auctions.)

In contrast to the ease with which such parameters arise in applied researches, a formal analysis of the optimal estimation problem has remained a challenging task. One

Date: December 29, 2012.

might consistently estimate θ by using plug-in estimator $\hat{\theta} = f(g(\hat{\boldsymbol{\beta}}))$, where $\hat{\boldsymbol{\beta}}$ is a $\sqrt{n}$-consistent estimator of $\boldsymbol{\beta}$. However, there have been concerns about the asymptotic bias that such an estimator carries, and some researchers have proposed ways to reduce the bias (Manski and Pepper (2000), Haile and Tamer (2003), Chernozhukov, Lee, and Rosen (2010)). However, Doss and Sethuraman (1989) showed that a sequence of estimators of a parameter for which there is no unbiased estimator must have variance diverging to infinity if the bias decreases to zero. (See also Hirano and Porter (2010) for a recent result for nondifferentiable parameters.) Given that one cannot eliminate the bias entirely without its variance exploding, the bias reduction may do the estimator either harm or good.

Many early researches on estimation of a nonregular parameter considered a parametric model and focused on finite sample optimality properties. For example, estimation of a normal mean under bound restrictions or order restrictions has been studied, among many others, by Lovell and Prescott (1970), Casella and Strawderman (1981), Bickel (1981), Moors (1981), and more recently van Eeden and Zidek (2004). Closer to this paper are researches by Blumenthal and Cohen (1968a,b) who studied estimation of $\max\{\beta_1, \beta_2\}$, when i.i.d. observations from a location family of symmetric distributions or normal distributions are available. On the other hand, the notion of asymptotic efficient estimation through the convolution theorem and the local asymptotic minimax theorem initiated by Hajék (1972) and Le Cam (1979) has mostly focused on regular parameters, and in many cases, resulted in regular estimators as optimal estimators. Hence the classical theory of semiparametric estimation widely known and well summarized in monographs such as Bickel, Klassen, Ritov, and Wellner (1993) and in later chapters of van der Vaart and Wellner (1996) does not directly apply to the problem of estimation of $\theta = (f \circ g)(\boldsymbol{\beta})$. This paper attempts to fill this gap from the perspective of local asymptotic minimax estimation.

This paper finds that for the class of nonregular parameters of the form (1.1), we can extend the existing theory of local asymptotic minimax estimation and construct a reasonable class of optimal estimators that are nonregular in general and asymptotically biased. The class of optimal estimators take the form of a plug-in estimator with semiparametrically efficient estimator of $\boldsymbol{\beta}$ except that it involves an additive bias-adjustment term which can be computed using simulations.

To deal with nondifferentiability, this paper first focuses on the special case where f is an identity, and utilizes the approach of generalized convolution theorem in van der Vaart (1989) to establish the local asymptotic minimax risk bound for the parameter θ. However, such a risk bound is hard to use in our set-up where f or g is potentially asymmetric, because the risk bound involves minimization of the risk over the distributions of "noise" in the convolution theorem. This paper uses the result of Dvoretsky, Wald, and Wolfowitz (1951) to reduce the risk bound to one involving minimization over a real line. And then this paper proposes a local asymptotic minimax decision of a simple form:

$$g(\hat{\boldsymbol{\beta}}) + \hat{c}/\sqrt{n},$$

where $\hat{\boldsymbol{\beta}}$ is a semiparametrically efficient estimator of $\boldsymbol{\beta}$ and $\hat{c}$ is a bias adjustment term that can be computed through simulations.

Next, extension to the case where f is continuous piecewise linear with a single kink point is done by making use of the insights of Blumenthal and Cohen (1968a) and applying them to local asymptotic minimax estimation. Thus, an estimator of the form

$$\hat{\theta}_{mx} \equiv f\left(g(\hat{\boldsymbol{\beta}}) + \frac{\hat{c}}{\sqrt{n}}\right), \tag{1.2}$$

with appropriate bias adjustment term $\hat{c}$, is shown to be local asymptotic minimax. In several situations, the bias adjustment term $\hat{c}$ can be set to zero. In particular, when $\theta = \mathbf{s}'\boldsymbol{\beta}$, for some known vector $\mathbf{s} \in \mathbf{R}^d$, so that θ is a regular parameter, the bias adjustment term can be set to be zero, and an optimal estimator in (1.2) is reduced to $\mathbf{s}'\hat{\boldsymbol{\beta}}$ which is a semiparametric efficient estimator of $\theta = \mathbf{s}'\boldsymbol{\beta}$. This confirms the continuity of this paper's approach with the standard method of semiparametric efficiency.

This paper offers results from a small sample simulation study for the cases of $\theta = \max\{\beta_1, \beta_2\}$ and $\theta = \max\{0, \beta_2\}$, where the bias adjustment suggested by the local asymptotic minimax estimation is either not necessary or only very minimal. This paper compares the method with two alternative bias reduction methods: fixed bias reduction method and a selective bias reduction method. The method of local asymptotic minimax estimation shows relatively robust performance in terms of the finite sample risk.

The next section defines the scope of the paper by introducing nondifferentiable transforms that this paper focuses on. The section also introduces regularity conditions for probabilities that identify $\boldsymbol{\beta}$. Section 3 investigates optimal decisions based on the local asymptotic maximal risks. Section 4 presents and discusses Monte Carlo simulation results. All the mathematical proofs are relegated to the Appendix.

We introduce some notation. Let $\mathbb{N}$ be the collection of natural numbers. Let $\mathbf{1}_d$ be a $d \times 1$ vector of ones with $d \geq 2$. For a vector $\mathbf{x} \in \mathbf{R}^d$ and a scalar c, we simply write $\mathbf{x} + c = \mathbf{x} + c\mathbf{1}_d$, or write $\mathbf{x} = c$ instead of $\mathbf{x} = c\mathbf{1}_d$. We define $S_1 \equiv \{\mathbf{x} \in \mathbf{R}^d : \mathbf{x}'\mathbf{1}_d = 1\}$, where the notation $\equiv$ indicates definition. For $\mathbf{x} \in \mathbf{R}^d$, the notation $\max(\mathbf{x})$ (or $\min(\mathbf{x})$) means the maximum (or the minimum) over the entries of the vector $\mathbf{x}$. 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 meanings are obvious. We let $\bar{\mathbf{R}} = [-\infty, \infty]$ and view it as a two-point compactification of $\mathbf{R}$, and let $\bar{\mathbf{R}}^d$ be the product of its d copies.

2. Nondifferentiable Transforms of a Regular Parameter

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

$$\theta = (f \circ g)(\boldsymbol{\beta}), \tag{2.1}$$

where $\boldsymbol{\beta} \in \mathbf{R}^d$ is a regular parameter (the meaning of regularity for $\boldsymbol{\beta}$ is clarified in Assumption 3 below), and $g : \bar{\mathbf{R}}^d \to \bar{\mathbf{R}}$ and $f : \bar{\mathbf{R}} \to \bar{\mathbf{R}}$ satisfy the following assumptions.

Assumption 1: (i) The map $g : \bar{\mathbf{R}}^d \to \bar{\mathbf{R}}$ is Lipschitz continuous, $g(\mathbf{R}^d) \subset \mathbf{R}$, and satisfies the following.
(a) (Translation Equivariance) For each $c \in \bar{\mathbf{R}}$ and $\mathbf{x} \in \mathbf{R}^d$, $g(\mathbf{x} + c) = g(\mathbf{x}) + c$.
(b) (Scale Equivariance) For each $u \geq 0$ and $\mathbf{x} \in \mathbf{R}^d$, $g(u\mathbf{x}) = ug(\mathbf{x})$.
(ii) The map $f : \bar{\mathbf{R}} \to \bar{\mathbf{R}}$ is continuous, piecewise linear with one kink at a point in $\mathbf{R}$.

Assumption 1 essentially defines the scope of this paper. Some examples of g are as follows.

EXAMPLES 1: (a) $g(\mathbf{x}) = \mathbf{s}'\mathbf{x}$, where $\mathbf{s} \in S_1$.
(b) $g(\mathbf{x}) = \max(\mathbf{x})$ or $g(\mathbf{x}) = \min(\mathbf{x})$.
(c) $g(\mathbf{x}) = \max\{\min(\mathbf{x}_1), \mathbf{x}_2\}$, $g(\mathbf{x}) = \max(\mathbf{x}_1) + \max(\mathbf{x}_2)$, $g(\mathbf{x}) = \min(\mathbf{x}_1) + \min(\mathbf{x}_2)$, $g(\mathbf{x}) = \max(\mathbf{x}_1) + \min(\mathbf{x}_2)$, or $g(\mathbf{x}) = \max(\mathbf{x}_1) + \mathbf{s}'\mathbf{x}$ with $\mathbf{s} \in S_1$, where $\mathbf{x}_1$ and $\mathbf{x}_2$ are subvectors of $\mathbf{x}$.

One might ask whether the representation of parameter θ as a composition map $f \circ g$ of $\boldsymbol{\beta}$ in (2.1) is unique. The following lemma gives an affirmative answer.

LEMMA 1: *Suppose that f_1 and f_2 are $\mathbf{R}$-valued maps on $\mathbf{R}$ that are non-constant on $\mathbf{R}$, and g_1 and g_2 satisfy* Assumption 1(i). *If $f_1 \circ g_1 = f_2 \circ g_2$, we have*

$$f_1 = f_2 \text{ and } g_1 = g_2.$$

As we shall see later, the local asymptotic minimax risk bound involves g and the optimal estimators involve the maps f and g. The uniqueness result of Lemma 1 removes ambiguity that could potentially arise when θ had multiple equivalent representations with different maps f and g.

We introduce briefly conditions for probabilities that identify $\boldsymbol{\beta}$, in a manner adapted from van der Vaart (1991) and van der Vaart and Wellner (1996). Let $\mathcal{P} \equiv \{P_\alpha : \alpha \in \mathcal{A}\}$ be a family of distributions on a measurable space $(\mathcal{X}, \mathcal{G})$ indexed by $\alpha \in \mathcal{A}$, where the set $\mathcal{A}$ is a subset of a Euclidean space or an infinite dimensional space. We assume that we have i.i.d. draws $Y_1, \cdots, Y_n$ from $P_{\alpha_0} \in \mathcal{P}$ so that $\mathbf{X}_n \equiv (Y_1, \cdots, Y_n)$ is distributed as $P_{\alpha_0}^n$. Let $\mathcal{P}(P_{\alpha_0})$ be the collection of maps $t \to P_{\alpha_t}$ such that for some $h \in L_2(P_{\alpha_0})$,

$$\int \left\{ \frac{1}{t}\left(dP_{\alpha_t}^{1/2} - dP_{\alpha_0}^{1/2}\right) - \frac{1}{2} h dP_{\alpha_0}^{1/2} \right\}^2 \to 0, \text{ as } n \to \infty.$$

When this convergence holds, we say that P_{α_t} *converges in quadratic mean* to P_{α_0}, call $h \in L_2(P_{\alpha_0})$ a *score function* associated with this convergence, and call the set of all such h's a *tangent set*, denoting it by $T(P_{\alpha_0})$. We assume that the tangent set is a linear subspace of $L_2(P_{\alpha_0})$. Taking $\langle \cdot, \cdot \rangle$ to be the usual inner product in $L_2(P_{\alpha_0})$, we write $H \equiv T(P_{\alpha_0})$ and view $(H, \langle \cdot, \cdot \rangle)$ as a subspace of a separable Hilbert space, with $\bar{H}$ denoting its completion. For each $h \in H$, $n \in \mathbb{N}$, and $\lambda_h \in \mathcal{A}$, let $P_{\alpha_0 + \lambda_h/\sqrt{n}}$ be probabilities converging in quadratic mean to P_{α_0} as $n \to \infty$ having h as its associated score. We simply write $P_{n,h} = P_{\alpha_0 + \lambda_h/\sqrt{n}}^n$ and consider sequences of such probabilities $\{P_{n,h}\}_{n \geq 1}$ indexed by $h \in H$. (See van der Vaart (1991) and van der Vaart and Wellner (1996), Section 3.11 for details.) The collection $\mathcal{E}_n \equiv (\mathcal{X}_n, \mathcal{G}_n, P_{n,h}; h \in H)$ constitutes a sequence of statistical experiments for $\boldsymbol{\beta}$ (Blackwell (1951)). As for $\mathcal{E}_n$, we assume local asymptotic normality as follows.

ASSUMPTION 2: (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 $\mathcal{E} = (\mathcal{X}, \mathcal{G}, P_h; h \in H)$, 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 $\boldsymbol{\beta}$ associated with the space of probability sequences $\{\{P_{n,h}\}_{n \geq 1} : h \in H\}$ (van der Vaart (1991).) Taking $\boldsymbol{\beta}$ as an $\mathbf{R}^d$-valued map on $\{P_{n,h} : h \in H\}$, we can regard the map as a sequence of $\mathbf{R}^d$-valued maps on H and write it as $\boldsymbol{\beta}_n(h)$.

Assumption 3: (Regular Parameter) There exists a continuous linear $\mathbf{R}^d$-valued map, $\dot{\boldsymbol{\beta}}$, on H such that

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

as $n \to \infty$.

Assumption 3 requires that $\boldsymbol{\beta}_n(h)$ is *regular* in the sense of van der Vaart and Wellner (1996, Section 3.11). The map $\dot{\boldsymbol{\beta}}$ in Assumption 3 is associated with the semiparametric efficiency bound of $\boldsymbol{\beta}$. For each $\mathbf{b} \in \mathbf{R}^d$, $\mathbf{b}'\dot{\boldsymbol{\beta}}(\cdot)$ defines a continuous linear functional on H, and hence there exists $\dot{\beta}_{\mathbf{b}}^* \in \bar{H}$ such that $\mathbf{b}'\dot{\boldsymbol{\beta}}(h) = \langle \dot{\beta}_{\mathbf{b}}^*, h \rangle$, $h \in H$. Then for any $\mathbf{b} \in \mathbf{R}^d$, $||\dot{\beta}_{\mathbf{b}}^*||^2$ represents the asymptotic variance bound of the parameter $\mathbf{b}'\boldsymbol{\beta}$. The map $\dot{\beta}_{\mathbf{b}}^*$ is called an *efficient influence function* for $\mathbf{b}'\boldsymbol{\beta}$ in the literature (e.g. van der Vaart (1991)). Let $\mathbf{e}_m$ be a $d \times 1$ vector whose m-th entry is one and the other entries are zero, and let Σ be a $d \times d$ matrix whose (m,k)-th entry is given by $\langle \dot{\beta}_{\mathbf{e}_m}^*, \dot{\beta}_{\mathbf{e}_k}^* \rangle$. As for Σ, we assume the following:

Assumption 4: Σ is invertible.

The inverse of matrix Σ is called the semiparametric efficiency bound for $\boldsymbol{\beta}$. In particular, Assumption 4 requires that there is no redundancy among the entries of $\boldsymbol{\beta}$, i.e., one entry of $\boldsymbol{\beta}$ is not defined as a linear combination of the other entries.

3. Local Asymptotic Minimax Estimators

3.1. **Loss Functions.**

For a decision $d \in \mathbf{R}$ and the object of interest $\theta \in \mathbf{R}$, we consider the following form of a loss function:

$$L(d, \theta) = \tau(|d - \theta|), \tag{3.1}$$

where $\tau : \mathbf{R} \to \mathbf{R}$ is a map that satisfies the following assumption.

Assumption 5: $\tau(\cdot)$ is nonnegative, strictly increasing, $\tau(y) \to \infty$ as $y \to \infty$, $\tau(0) = 0$, and for each $M > 0$, there exists $c_M > 0$ such that for all $x, y \in \mathbf{R}$,

$$|\tau_M(x) - \tau_M(y)| \leq c_M |x - y|,$$

where $\tau_M(\cdot) = \min\{\tau(\cdot), M\}$.

While Assumption 5 is satisfied by many loss functions, it excludes the hypothesis testing type loss function $\tau(|d-\theta|) = 1\{|d-\theta| > c\}$, $c \in \mathbf{R}$. The following lemma establishes a lower bound for the local asymptotic minimax risk when f is an identity. Let $Z \in \mathbf{R}^d$ be a random vector having distribution equal to $N(0, \Sigma)$.

Lemma 2: *Suppose that Assumptions 1-5 hold. Then for any sequence of estimators* $\hat{\theta}$,

$$\begin{aligned} &\liminf_{n\to\infty} \sup_{h\in H} \mathbf{E}_h \left[\tau(|\sqrt{n}\{\hat{\theta} - g(\boldsymbol{\beta}_n(h))\}|)\right] \\ \geq\ &\inf_{F\in\mathcal{F}} \sup_{\mathbf{r}\in\mathbf{R}^d} \int \mathbf{E}\left[\tau(|g(Z+\mathbf{r}) - g(\mathbf{r}) + w|)\right] dF(w), \end{aligned}$$

where $\mathcal{F}$ *denotes the collection of probability measures on the Borel* σ*-field of* $\mathbf{R}$.

The lemma establishes a lower bound for the risk. The result is obtained by using a version of a generalized convolution theorem in van der Vaart (1989) which is adapted to the current set-up. The main difficulty with Lemma 2 is that the supremum over $\mathbf{r} \in \mathbf{R}^d$ and the infimum over $F \in \mathcal{F}$ do not have an explicit solution in general. Hence this paper considers simulating the lower bound in Lemma 2 by using random draws from a distribution approximating that of Z. The main obstacle in this approach is that the risk lower bound involves infimum over an infinite dimensional space $\mathcal{F}$.

We obtain a much simpler formulation by using the classical purification result of Dvoretsky, Wald, and Wolfowitz (1951) for zero sum games, where it is shown that the risk of a randomized decision can be replaced by that of a nonrandomized decision when the distributions of observations are atomless. This result has had an impact on the literature of purifications in incomplete information games (e.g. Milgrom and Weber (1985)). In our set-up, the observations are not drawn from an atomless distribution, but in the limiting experiment, we can regard them as drawn from a shifted normal distribution. This enables us to use their result to obtain the following theorem.

Theorem 1: *Suppose that Assumptions 1-5 hold. Then for any sequence of estimators* $\hat{\theta}$,

$$\liminf_{n\to\infty} \sup_{h\in H} \mathbf{E}_h \left[\tau(|\sqrt{n}\{\hat{\theta} - g(\boldsymbol{\beta}_n(h))\}|)\right] \geq \inf_{c\in\mathbf{R}} B(c;1),$$

where for $c \in \mathbf{R}$, *and* $a \geq 0$,

$$B(c;a) \equiv \sup_{\mathbf{r}\in\mathbf{R}^d} \mathbf{E}\left[\tau(a|g(Z+\mathbf{r}) - g(\mathbf{r}) + c|)\right].$$

The main feature of the lower bound in Theorem 1 is that it involves infimum over a *single-dimensional* space $\mathbf{R}$ in its risk bound. This simpler form now makes it feasible to simulate the lower bound for the risk.

This paper proposes a method of constructing a local asymptotic minimax estimator as follows. Suppose that we are given a consistent estimator $\hat{\Sigma}$ of Σ and a semiparametrically efficient estimator $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ which satisfy the following assumptions. (See Bickel, Klaasen, Ritov, and Wellner (1993) for semiparametric efficient estimators from various models.)

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

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

(ii) For each $t \in \mathbf{R}^d$, $\text{sup}_{h\in H}\left|P_{n,h}\{\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_n(h)) \leq t\} - P\{Z \leq t\}\right| \to 0$ as $n \to \infty$.

Assumption 6 imposes $\sqrt{n}$-consistency of $\hat{\Sigma}$ and convergence in distribution of $\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_n(h))$, both uniform over $h \in H$. The uniform convergence can be proved through the central limit theorem uniform in $h \in H$. Under regularity conditions, the uniform central limit theorem of a sum of i.i.d. random variables follows from a Berry-Esseen bound, as long as the third moment of the random variable is bounded uniformly in $h \in H$.

For technical facility, we follow a suggestion by Strasser (1985) (p.440) and consider a truncated loss: $\tau_M(\cdot) = \min\{\tau(\cdot), M\}$ for large M. To simulate the risk lower bound in Theorem 1, we first draw $\{\boldsymbol{\xi}_i\}_{i=1}^L$ i.i.d. from $N(0, I_d)$. For a fixed large $M_1 > 0$, let

$$\hat{B}_{M_1}(c;1) \equiv \sup_{\mathbf{r}\in[-M_1,M_1]^d} \frac{1}{L}\sum_{i=1}^{L} \tau_{M_1}\left(\left|g(\hat{\Sigma}^{1/2}\boldsymbol{\xi}_i + \mathbf{r}) - g(\mathbf{r}) + c\right|\right).$$

Then we obtain

$$\hat{c}_{M_1} \equiv \frac{1}{2}\left\{\sup \hat{E}_{M_1} + \inf \hat{E}_{M_1}\right\}, \tag{3.2}$$

where, with $\eta_{n,L} \to 0$ as $n, L \to \infty$, $\eta_{n,L}\sqrt{n} \to \infty$ as $n \to \infty$ and $\eta_{n,L}\sqrt{L} \to \infty$ as $L \to \infty$,

$$\hat{E}_{M_1} \equiv \left\{c \in [-M_1, M_1] : \hat{B}_{M_1}(c;1) \leq \inf_{c_1\in[-M_1,M_1]} \hat{B}_{M_1}(c_1;1) + \eta_{n,L}\right\}.$$

Let us construct

$$\hat{\theta}_{mx} \equiv g(\hat{\boldsymbol{\beta}}) + \frac{\hat{c}_{M_1}}{\sqrt{n}}. \tag{3.3}$$

The following theorem affirms that $\hat{\theta}_{mx}$ is local asymptotic minimax for $\theta = g(\boldsymbol{\beta})$.

THEOREM 2: *Suppose that the conditions of Theorem 1 and Assumption 6 hold. Then,*

$$\lim_{M_1\geq M:M\uparrow\infty}\ \underset{n\to\infty}{\text{limsup}}\ \sup_{h\in H} \mathbf{E}_h\left[\tau_M(|\sqrt{n}\{\hat{\theta}_{mx} - g(\boldsymbol{\beta}_n(h))\}|)\right] \leq \inf_{c\in\mathbf{R}} B(c;1).$$

Recall that the candidate estimators considered in Theorem 1 were not restricted to plug-in estimators with an additive bias adjustment term. As standard in the literature of local asymptotic minimax estimation, the candidate estimators are any sequences of measurable functions of observations including both regular and nonregular estimators. The main thrust of Theorem 2 is the finding that it is sufficient for local asymptotic minimax estimation to consider a plug-in estimator using a semiparametrically efficient estimator of $\boldsymbol{\beta}$ with an additive bias adjustment term as in (3.3). It remains to find optimal bias adjustment, which can be done using the simulation method proposed earlier.

We now extend the result to the case where f is not an identity map, but a continuous piecewise linear map with a single kink point. The main idea is taken from the proof of Theorem 3.1 of Blumenthal and Cohen (1968a).

THEOREM 3: *Suppose that the conditions of Theorem 1 hold, and let s be the maximum absolute slope from the two linear components of f. Then for any sequence of estimators $\hat{\theta}$,*

$$\liminf_{n\to\infty} \sup_{h\in H} \mathbf{E}_h \left[\tau(|\sqrt{n}\{\hat{\theta} - (f\circ g)(\boldsymbol{\beta}_n(h))\}|)\right] \geq \inf_{c\in\mathbf{R}} B(c;s).$$

The bounds in Theorems 1 and 3 involve a bias adjustment term c^* that minimizes $B(c;1)$. A similar bias adjustment term appears in Takagi (1994)'s local asymptotic minimax estimation result. While the bias adjustment term arises here due to asymmetric nondifferentiable map $f \circ g$ of a regular parameter, it arises in his paper due to an asymmetric loss function, and the decision problem in this paper cannot be reduced to his set-up, even if we assume a parametric family of distributions indexed by an open interval as he does in his paper.

Now let us search for a class of local asymptotic minimax estimators that achieve the lower bound in Theorem 3. It turns out that an estimator of the form:

$$\tilde{\theta}_{mx} \equiv f\left(g(\hat{\boldsymbol{\beta}}) + \frac{\hat{c}_{M_1}}{\sqrt{n}}\right), \tag{3.4}$$

where $\hat{c}_{M_1}$ is the bias-adjustment term introduced previously, is local asymptotic minimax. To see this intuitively, first observe that we lose no generality by considering $f_s(\cdot) \equiv f(\cdot)/s$ instead of $f(\cdot)$. Hence we assume $s = 1$ and note that $f(\cdot)$ is then a contraction mapping so that

$$|\tilde{\theta}_{mx} - (f\circ g)(\boldsymbol{\beta}_n(h))| \leq |\hat{\theta}_{mx} - g(\boldsymbol{\beta}_n(h))|.$$

It follows from Theorem 2 that the decision $\tilde{\theta}_{mx}$ achieves the bound $\inf_{c\in\mathbf{R}} B(c;1)$. We state this result as follows and a formal proof is given in the appendix.

THEOREM 4: *Suppose that the conditions of Theorem 2 and Assumption 6 hold. Then,*

$$\lim_{M_1\geq M:M\uparrow\infty} \limsup_{n\to\infty} \sup_{h\in H} \mathbf{E}_h \left[\tau_M(|\sqrt{n}\{\tilde{\theta}_{mx} - (f\circ g)(\boldsymbol{\beta}_n(h))\}|)\right] \leq \inf_{c\in\mathbf{R}} B(c;s).$$

The estimator $\tilde{\theta}_{mx}$ is in general a nonregular estimator that is asymptotically biased. When $g(\boldsymbol{\beta}) = \mathbf{s}'\boldsymbol{\beta}$ with $\mathbf{s} \in S_1$, the risk bound (with $s = 1$) in Theorem 4 becomes

$$\inf_{c\in\mathbf{R}} \mathbf{E}\left[\tau\left(|g(Z)+c|\right)\right] = \mathbf{E}\left[\tau\left(|\mathbf{s}'Z|\right)\right],$$

where the equality follows by Anderson's Lemma. In this case, it suffices to set $\hat{c}_{M_1} = 0$, because the infimum over $c \in \mathbf{R}$ is achieved at $c = 0$. The minimax decision thus becomes simply

$$\tilde{\theta}_{mx} = f(\hat{\boldsymbol{\beta}}'\mathbf{s}). \tag{3.5}$$

This has the following consequences.

EXAMPLES 2: (a) When $\theta = \boldsymbol{\beta}'\mathbf{s}$ for a known vector $\mathbf{s} \in S_1$, $\tilde{\theta}_{mx} = \hat{\boldsymbol{\beta}}'\mathbf{s}$. Therefore, the decision in (3.5) reduces to a semiparametric efficient estimator of $\boldsymbol{\beta}'\mathbf{s}$.
(b) When $\theta = \max\{a\boldsymbol{\beta}'\mathbf{s}+b, 0\}$ for a known vector $\mathbf{s} \in S_1$ and known constants $a, b \in \mathbf{R}$, $\tilde{\theta}_{mx} = \max\{a\hat{\boldsymbol{\beta}}'\mathbf{s} + b, 0\}$.
(c) When $\theta = |\beta|$ for a scalar parameter β, $\tilde{\theta}_{mx} = |\hat{\beta}|$. ■

The examples of (b)-(c) involve nondifferentiable transform f, and hence $\tilde{\theta}_{mx} = f(\hat{\boldsymbol{\beta}}'\mathbf{s})$ as an estimator of θ is asymptotically biased in these examples. Nevertheless, the plug-in estimator $\tilde{\theta}_{mx}$ that does not require any bias adjustment is local asymptotic minimax. We provide another example that has the bias adjustment term equal to zero. This example is motivated by Blumenthal and Cohen (1968a).

Examples 3: Suppose that $\theta = \max\{\beta_1, \beta_2\}$, where $\boldsymbol{\beta} = (\beta_1, \beta_2) \in \mathbf{R}^2$ is a regular parameter, and the 2×2 matrix Σ is a diagonal matrix with identical diagonal entries σ^2. We take $\tau(x) = x^2$, i.e., the squared error loss. Then one can show that the local asymptotic minimax risk bound is achieved by $\hat{\theta}_{mx} = \max\{\hat{\beta}_1, \hat{\beta}_2\}$, where $\hat{\boldsymbol{\beta}} = (\hat{\beta}_1, \hat{\beta}_2)$ is a semiparametrically efficient estimator of $\boldsymbol{\beta}$. To see this, first note that the local asymptotic minimax risk bound in Theorem 1 becomes

$$\inf_{c \in \mathbf{R}} \sup_{r \geq 0} \mathbf{E} \left(\max\{Z_1 - r, Z_2\} - c \right)^2 .$$

For each $c \in \mathbf{R}$, $\mathbf{E} \left(\max\{Z_1 - r, Z_2\} - c \right)^2$ is quasiconvex in $r \geq 0$ so that the supremum over $r \geq 0$ is achieved at $r = 0$ or $r \to \infty$. When $r = 0$, the bound becomes $Var(\max\{Z_1, Z_2\})$ and when $r \to \infty$, the bound becomes $Var(Z_1)$. By (5.10) of Moriguti (1951), we have $Var(\max\{Z_1, Z_2\}) \leq Var(Z_1)$, so that the local asymptotic risk bound becomes $Var(Z_1) = \sigma^2$ with $r = \infty$ and $c = 0$. On the other hand, it is not hard to see from (A.3) of Blumenthal and Cohen (1968b) that the local asymptotic maximal risk of $\tilde{\theta}_{mx} = \max\{\hat{\beta}_1, \hat{\beta}_2\}$ is equal to σ^2, confirming that it is indeed local asymptotic minimax. This result parallels the finding by Blumenthal and Cohen (1968a) that for squared error loss and with observations of two independent random variables X_1 and X_2 from a location family of symmetric distributions, $\max\{X_1, X_2\}$ is a minimax decision, and the risk bound is σ^2. ■

4. Monte Carlo Simulations

4.1. **Simulation Designs.** As mentioned in the introduction, various methods of bias reduction for nondifferentiable parameters have been proposed in the literature. In the simulation study, this paper compares the finite sample risk performances of the local asymptotic minimax estimator proposed in this paper with estimators that perform bias reductions in two methods: fixed bias reduction and selective bias reduction.

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 $X_1 \sim N(\boldsymbol{\beta}, \Sigma)$, where

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} 0 \\ \delta_0/\sqrt{n} \end{bmatrix} \text{ and } \Sigma = \begin{bmatrix} 2 & 1/2 \\ 1/2 & 4 \end{bmatrix}, \tag{4.1}$$

where δ_0 is chosen from grid points in $[-10, 10]$. The parameters of interest are as follows:

$$\theta_1 \equiv \max\{\beta_1, \beta_2\} \text{ and } \theta_2 \equiv \max\{\beta_2, 0\}.$$

When δ_0 is close to zero, parameters θ_1 and θ_2 have $\boldsymbol{\beta}$ close to the kink point of the nondifferentiable map. However, when δ_0 is away from zero, the parameters become

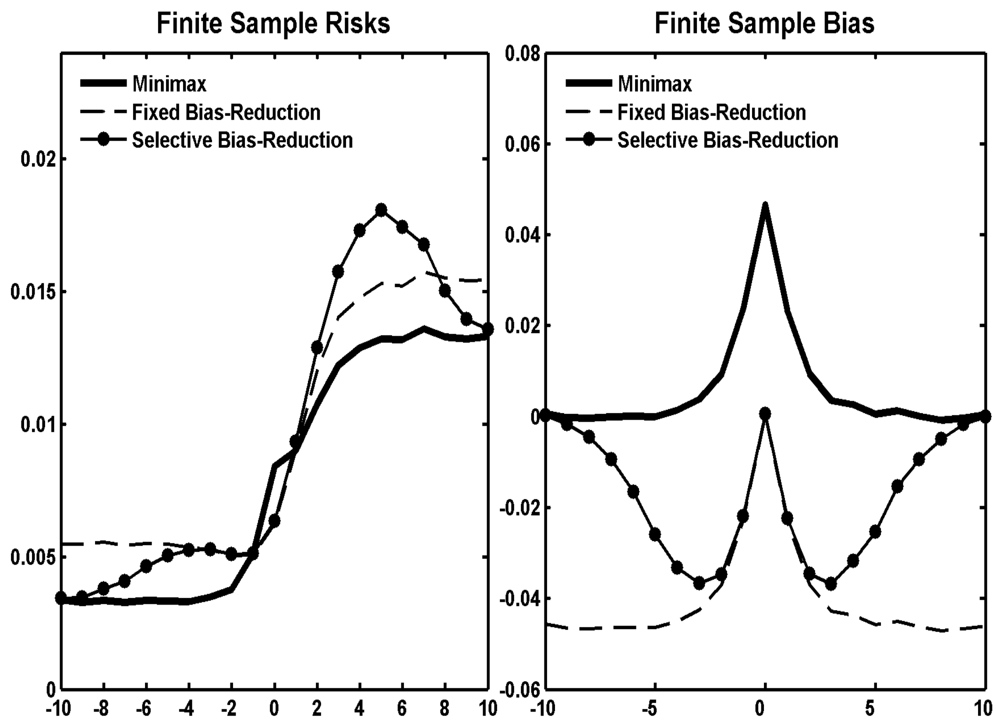

FIGURE 1. Comparison of the Local Asymptotic Minimax Estimators with Estimators Obtained through Other Bias-Reduction Methods: $\theta_1 = \max\{\beta_1, \beta_2\}$.

more like a regular parameter themselves. We take $\hat{\boldsymbol{\beta}} = \frac{1}{n}\sum_{i=1}^{n} X_i$ as the estimator of $\boldsymbol{\beta}$. As for the finite sample risk, we adopt the mean squared error:

$$\mathbf{E}\left[(\hat{\theta}_j - \theta_j)^2\right], \; j = 1, 2,$$

where $\hat{\theta}_j$ is a candidate estimator for θ_j. We evaluated the risk using Monte Carlo simulations. The sample size was 300. The Monte Carlo simulation number was set to be 20,000. In the simulation study, we investigate the finite sample risk profile of decisions by varying δ_0.

4.2. **Minimax Decision and Bias Reduction for** $\max\{\beta_1, \beta_2\}$**.** In the case of $\theta_1 \equiv \max\{\beta_1, \beta_2\}$, $b_F \equiv \mathbf{E}[\max\{X_{11} - \beta_1, X_{12} - \beta_2\}]$ becomes the asymptotic bias of the estimator $\hat{\theta}_1 \equiv \max\{\hat{\beta}_1, \hat{\beta}_2\}$ when $\beta_1 = \beta_2$. One may consider the following estimator of b_F :

$$\hat{b}_F \equiv \frac{1}{L}\sum_{i=1}^{L} \max\left(\hat{\Sigma}^{1/2}\boldsymbol{\xi}_i\right),$$

where $\boldsymbol{\xi}_i$ is drawn i.i.d. from $N(0, I_2)$. This adjustment term $\hat{b}_F$ is fixed over different values of $\beta_2 - \beta_1$ (in large samples). Since the bias of $\max\{\hat{\beta}_1, \hat{\beta}_2\}$ becomes prominent only when β_1 is close to β_2, one may instead consider performing bias adjustment only when the estimated difference $|\beta_2 - \beta_1|$ is close to zero. Thus we also consider the following estimated adjustment term:

$$\hat{b}_S \equiv \left(\frac{1}{L} \sum_{i=1}^{L} \max \left(\hat{\Sigma}^{1/2} \boldsymbol{\xi}_i \right) \right) 1 \left\{ |\hat{\beta}_2 - \hat{\beta}_1| < 1.7/n^{1/3} \right\}.$$

We compare the following two estimators with the minimax decision $\tilde{\theta}_{mx}$:

$$\hat{\theta}_F \equiv \max\{\hat{\beta}_1, \hat{\beta}_2\} - \hat{b}_F/\sqrt{n} \text{ and } \hat{\theta}_S \equiv \max\{\hat{\beta}_1, \hat{\beta}_2\} - \hat{b}_S/\sqrt{n}.$$

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 1.

The finite sample risks of $\hat{\theta}_F$ are better than the minimax decision $\hat{\theta}_{mx}$ only locally around $\delta_0 = 0$. The bias reduction using $\hat{b}_F$ improves the estimator's performance in this case. However, for other values of δ_0, the bias reduction 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 1 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 $\tilde{\theta}_{mx}$. 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 $\max\{\hat{\beta}_1, \hat{\beta}_2\}$. 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 as shown on the left panel in Figure 1. 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 a 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 $\hat{c}_{M_1}$ of the minimax decision is zero. This means that the estimator $\hat{\theta}_{mx}$ requires 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 1 shows, it performs better than the other two estimators except when δ_0 is locally around zero, or when $\beta_2 - \beta_1$ is around roughly between -0.057 and 0.041.

4.3. **Minimax Decision and Bias Reduction for** $\max\{0, \beta_2\}$**.** We consider $\theta_2 \equiv \max\{\beta_2, 0\}$. The bias of the plug-in estimator $\hat{\theta}_2 \equiv \max\{0, \hat{\beta}_2\}$, due to its asymmetric nature, might cause a concern at first glance. The bias at the kink point $\beta_2 = 0$ is equal

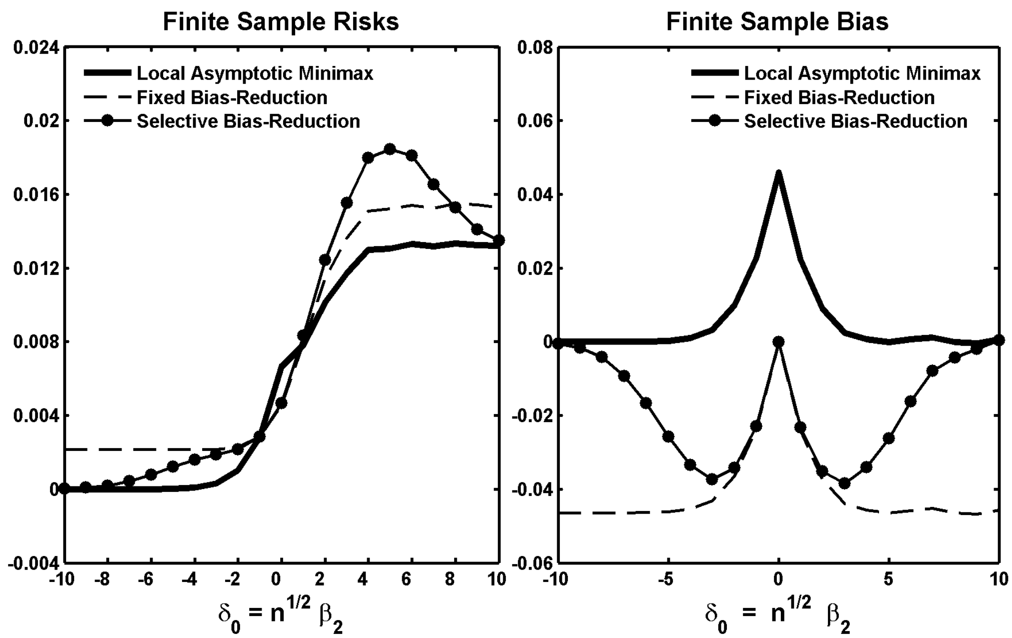

FIGURE 2. Comparison of the Local Asymptotic Minimax Estimators with Estimators Obtained through Other Bias-Reduction Methods: $\theta_2 = \max\{0, \beta_2\}$.

to $b_F \equiv \mathbf{E}[\max\{0, X_{12} - \beta_2\}]$ and its estimator is taken to be

$$\hat{b}_F \equiv \frac{1}{L}\sum_{i=1}^{L} \max\{0, \hat{\sigma}_2 \xi_i\},$$

where ξ_i is drawn i.i.d. from $N(0,1)$ and $\hat{\sigma}_2^2$ is the sample variance of $\{X_{i2}\}_{i=1}^n$. The fixed bias reduction method uses this estimator to perform bias reduction. As for the selective bias reduction method, we also consider the following estimated adjustment term:

$$\hat{b}_S \equiv \left(\frac{1}{L}\sum_{i=1}^{L} \max\{0, \hat{\sigma}_2 \xi_i\}\right) 1\left\{|\hat{\beta}_2| < 1.7/n^{1/3}\right\}.$$

As before, we compare the following two estimators with the minimax decision $\tilde{\theta}_{mx}$:

$$\hat{\theta}_F \equiv \max\{0, \hat{\beta}_2\} - \hat{b}_F/\sqrt{n} \text{ and } \hat{\theta}_S \equiv \max\{0, \hat{\beta}_2\} - \hat{b}_S/\sqrt{n}.$$

The results are shown in Figure 2. The results are similar to the case of $\max\{\beta_1, \beta_2\}$. Except for the case where β_2 is local around zero, roughly between 0.057 and -0.057, the local asymptotic minimax estimator performs better than the other methods. This result shows the overall robustness properties of the local asymptotic minimax approach.

5. Conclusion

The paper proposes local asymptotic minimax estimators for a class of nonregular parameters that are constructed by applying translation-scale equivariant transform to a regular parameter. The results are extended to the case where the nonregular parameters are transformed further by a piecewise linear map with a single kink. The local asymptotic minimax estimators take the form of a plug-in estimator with an additive bias adjustment term. The bias adjustment term can be computed by a simulation method. A small scale Monte Carlo simulation study demonstrates the robust finite sample risk properties of the local asymptotic minimax estimators, as compared to estimators based on alternative bias correction methods.

6. Appendix: Mathematical Proofs

Proof of Lemma 1: First, suppose to the contrary that $f_1(y) \neq f_2(y)$ for some $y \in \mathbf{R}$. Then since $f_1 \circ g_1 = f_2 \circ g_2$, it is necessary that $g_1(\boldsymbol{\beta}) \neq g_2(\boldsymbol{\beta})$ for some $\boldsymbol{\beta} \in \mathbf{R}^d$ such that $g_1(\boldsymbol{\beta}) = y$. Hence

$$(f_1 \circ g_1)(\boldsymbol{\beta}) \neq (f_2 \circ g_1)(\boldsymbol{\beta}). \tag{6.1}$$

Now observe that $f_2(g_1(\boldsymbol{\beta})) = f_2(g_2(\boldsymbol{\beta}) + g_1(\boldsymbol{\beta}) - g_2(\boldsymbol{\beta})) = f_2(g_2(\boldsymbol{\beta} + g_1(\boldsymbol{\beta}) - g_2(\boldsymbol{\beta})))$. Since $f_1 \circ g_1 = f_2 \circ g_2$, the last term is equal to

$$\begin{aligned} f_1(g_1(\boldsymbol{\beta} + g_1(\boldsymbol{\beta}) - g_2(\boldsymbol{\beta}))) &= f_1(2g_1(\boldsymbol{\beta}) - g_2(\boldsymbol{\beta})) = f_1(g_1(2\boldsymbol{\beta} - g_2(\boldsymbol{\beta}))) \\ &= f_2(g_2(2\boldsymbol{\beta} - g_2(\boldsymbol{\beta}))) = f_2(g_2(\boldsymbol{\beta})) = f_1(g_1(\boldsymbol{\beta})). \end{aligned}$$

Therefore, we conclude that $f_2(g_1(\boldsymbol{\beta})) = f_1(g_1(\boldsymbol{\beta}))$ contradicting (6.1).

Second, suppose to the contrary that $g_1(\boldsymbol{\beta}) \neq g_2(\boldsymbol{\beta})$ for some $\boldsymbol{\beta} \in \mathbf{R}^d$ and $f_1 = f_2$. First suppose that $g_1(\boldsymbol{\beta}) > g_2(\boldsymbol{\beta})$. Fix arbitrary $a \in \mathbf{R}$ and $c \geq 0$ and let $c_\Delta = c/\Delta_{1,2}(\boldsymbol{\beta})$ and $\Delta_{1,2}(\boldsymbol{\beta}) = g_1(\boldsymbol{\beta}) - g_2(\boldsymbol{\beta})$. Then

$$\begin{aligned} f_1(a + c) &= f_1(a + \Delta_{1,2}(c_\Delta \boldsymbol{\beta})) = f_1(a + g_2(c_\Delta \boldsymbol{\beta}) + \Delta_{1,2}(c_\Delta \boldsymbol{\beta}) - g_2(c_\Delta \boldsymbol{\beta})) \\ &= f_1(g_2(a + c_\Delta \boldsymbol{\beta} + \Delta_{1,2}(c_\Delta \boldsymbol{\beta}) - g_2(c_\Delta \boldsymbol{\beta}))) \\ &= f_2(g_2(a + c_\Delta \boldsymbol{\beta} + \Delta_{1,2}(c_\Delta \boldsymbol{\beta}) - g_2(c_\Delta \boldsymbol{\beta}))) \\ &= f_1(g_1(a + c_\Delta \boldsymbol{\beta} + \Delta_{1,2}(c_\Delta \boldsymbol{\beta}) - g_2(c_\Delta \boldsymbol{\beta}))) \\ &= f_1(a + g_1(c_\Delta \boldsymbol{\beta} - g_2(c_\Delta \boldsymbol{\beta}) + c_\Delta \Delta_{1,2}(\boldsymbol{\beta}))) = f_1(a + 2c). \end{aligned}$$

The choice of $a \in \mathbf{R}$ and $c \geq 0$ are arbitrary, and hence $f_1(\cdot)$ is constant on $\mathbf{R}$, contradicting the nonconstancy condition for f_1.

Second, suppose that $g_1(\boldsymbol{\beta}) < g_2(\boldsymbol{\beta})$. Then, fix arbitrary $a \in \mathbf{R}$ and $c \leq 0$ and let $c_\Delta = c/\Delta_{1,2}(\boldsymbol{\beta})$. Then similarly as before, we have

$$\begin{aligned} f_1(a + c) &= f_1(a + \Delta_{1,2}(c_\Delta \boldsymbol{\beta})) \\ &= f_1(a + g_1(c_\Delta \boldsymbol{\beta} - g_2(c_\Delta \boldsymbol{\beta}) + c_\Delta \Delta_{1,2}(\boldsymbol{\beta}))) = f_1(a + 2c), \end{aligned}$$

because $\Delta_{1,2}(c_\Delta \boldsymbol{\beta}) = c$. Therefore, again, $f_1(\cdot)$ is constant on $\mathbf{R}$, contradicting the nonconstancy condition for f_1. ■

We view convergence in distribution $\xrightarrow{D}$ in the proofs as convergence in $\bar{\mathbf{R}}^d$, so that the limit distribution is allowed to be deficient in general. Choose $\{h_i\}_{i=1}^m$ from an orthonormal basis $\{h_i\}_{i=1}^\infty$ of H. For $\mathbf{p} \in \mathbf{R}^m$, we consider $h(\mathbf{p}) \equiv \Sigma_{i=1}^m p_i h_i$ so that $\dot{\beta}_j(h(\mathbf{p})) = \sum_{i=1}^m \dot{\beta}_j(h_i)p_i$, where $\dot{\beta}_j$ is the j-th element of $\dot{\boldsymbol{\beta}}$. Let $\mathbf{B}$ be $m \times d$ matrix such that

$$\mathbf{B} \equiv \begin{bmatrix} \dot{\beta}_1(h_1) & \dot{\beta}_2(h_1) & \cdots & \dot{\beta}_d(h_1) \\ \dot{\beta}_1(h_2) & \dot{\beta}_2(h_2) & \cdots & \dot{\beta}_d(h_2) \\ \vdots & \vdots & & \vdots \\ \dot{\beta}_1(h_m) & \dot{\beta}_2(h_m) & \cdots & \dot{\beta}_d(h_m) \end{bmatrix}. \tag{6.2}$$

We assume that $m \geq d$ and $\mathbf{B}$ is a full column rank matrix. We also define $\boldsymbol{\zeta} \equiv (\zeta(h_1), \cdots, \zeta(h_m))'$, where ζ is the Gaussian process that appears in Assumption 3, and with $\lambda > 0$, let $F_{\lambda,m}(\cdot)$ be the cdf of $N(0, I_m/\lambda)$, and let $Z_{\lambda,m} \in \mathbf{R}^d$ be a random vector following $N(0, \mathbf{B}'\mathbf{B}/(\lambda+1))$.

Let $g : \bar{\mathbf{R}}^d \to \bar{\mathbf{R}}$ be a translation equivariant map, i.e., a map that satisfies Assumptions 1(i)(a) and (b). Suppose that $\hat{\theta} \in \mathbf{R}$ is a sequence of estimators such that along $\{P_{n,0}\}$,

$$\begin{bmatrix} \sqrt{n}\{\hat{\theta} - g(\boldsymbol{\beta}_n(h))\} \\ \log dP_{n,h}/dP_{n,0} \end{bmatrix} \xrightarrow{D} \begin{bmatrix} V - g(\dot{\boldsymbol{\beta}}(h) + \boldsymbol{\beta}_g) \\ \zeta(h) - \frac{1}{2}\langle h, h\rangle \end{bmatrix},$$

for some nonstochastic vector $\boldsymbol{\beta}_g \in \bar{\mathbf{R}}^d$ such that $g(\boldsymbol{\beta}_g)$ whenever $\boldsymbol{\beta}_g \in \mathbf{R}^d$, and $V \in \bar{\mathbf{R}}^d$ is a random vector having a potentially deficient distribution. Let $\mathcal{L}_g^h$ be the limiting distribution of $\sqrt{n}\{\hat{\theta} - g(\boldsymbol{\beta}_n(h))\}$ in $\bar{\mathbf{R}}^d$ along $\{P_{n,h}\}$ for each $h \in H$. The following lemma is an adaptation of the generalized convolution theorem in van der Vaart (1989).

Lemma A1: *Suppose that Assumptions 1(i) and 2-4 hold. For any $\lambda > 0$, the distribution $\int \mathcal{L}_g^{h(\mathbf{p})} dF_\lambda(\mathbf{p})$ is equal to that of $-g(Z_{\lambda,m} + W_{\lambda,m} + \boldsymbol{\beta}_g)$, where $W_{\lambda,m} \in \bar{\mathbf{R}}$ is a random variable having a potentially deficient distribution independent of $Z_{\lambda,m}$.*

Proof: Using Assumption 1(i) and applying Le Cam's third lemma, we find that for all $C \in \mathcal{B}(\bar{\mathbf{R}})$, the Borel σ-field of $\bar{\mathbf{R}}$,

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

where $(-g)^{-1}(C) \equiv \{\mathbf{x} \in \bar{\mathbf{R}}^d : -g(\mathbf{x}) \in C\}$. The second equality uses translation equivariance of g. Let N_λ be the distribution of $N(0, I_m/(\lambda+1))$. We can write

$$\begin{aligned} &\int \mathcal{L}_g^{h(\mathbf{p})}(C) dF_{\lambda,m}(\mathbf{p}) \\ &= \int \mathbf{E}\left[1_{(-g)^{-1}(C)}\left(-v + \mathbf{B}'\mathbf{p} + \boldsymbol{\beta}_g\right) e^{\mathbf{p}'\boldsymbol{\zeta} - \frac{\lambda+1}{2}||\mathbf{p}||^2}\right] \left(\frac{\lambda}{2\pi}\right)^{m/2} d\mathcal{L}_g^0(v)d\mathbf{p} \\ &= \int \mathbf{E}\left[1_{(-g)^{-1}(C)}\left(-v + \mathbf{B}'\left(\mathbf{p} + \frac{\boldsymbol{\zeta}}{1+\lambda}\right) + \boldsymbol{\beta}_g\right) c_\lambda(\boldsymbol{\zeta})\right] d\mathcal{L}_g^0(v)dN_\lambda(\mathbf{p}), \end{aligned}$$

where $c_\lambda(\boldsymbol{\zeta}) \equiv e^{\frac{1}{2(\lambda+1)}\boldsymbol{\zeta}'\boldsymbol{\zeta}} \cdot (\lambda/(\lambda+1))^{m/2}$. When we let $W_{\lambda,m}$ be a random variable having the distribution $\mathcal{W}_{\lambda,m}$ defined by

$$\mathcal{W}_{\lambda,m}(C) \equiv \int \mathbf{E}\left[1_{(-g)^{-1}(C)}\left(v - \frac{\mathbf{B}'\boldsymbol{\zeta}}{1+\lambda}\right) c_\lambda(\boldsymbol{\zeta})\right] d\mathcal{L}_g^0(v),\ C \in \mathcal{B}(\bar{\mathbf{R}}),$$

the distribution $\int \mathcal{L}_g^{h(\mathbf{p})} dF_\lambda(\mathbf{p})$ is equal to that of $-g(Z_{\lambda,m} + W_{\lambda,m} + \boldsymbol{\beta}_g)$. ■

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

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

For any two probability measures P and Q on $\mathcal{B}(\mathbf{R}^d)$, define

$$d_{\mathcal{P}}(P, Q) \equiv \sup\left\{\left|\int f dP - \int f dQ\right| : ||f||_{BL} \leq 1\right\}. \tag{6.3}$$

For the proof of Theorem 1, we employ two lemmas. The first lemma is Lemma 3 of Chamberlain (1987), which is used to write the risk using a distribution that has a finite set support, and the second lemma is Theorem 3.2 of Dvoretsky, Wald, and Wolfowitz (1951).

Lemma A2 (Chamberlain (1987)): *Let $h : \mathbf{R}^m \to \mathbf{R}^d$ be a Borel measurable function and let P be a probability measure on $(\mathbf{R}^m, \mathcal{B}(\mathbf{R}^m))$ with a support $A_P \subset \mathbf{R}^m$. If $\int ||h|| dP < \infty$, then there exists a probability measure Q whose support is a finite subset of A_P and*

$$\int h dP = \int h dQ.$$

Lemma A3 (Dvoretsky, Wald and Wolfowitz (1951)): *Let $\mathcal{P}$ be a finite set of distributions on $(\mathbf{R}^m, \mathcal{B}(\mathbf{R}^m))$, where each distribution is atomless, and for each $P \in \mathcal{P}$, let $W_P : \mathbf{R}^m \times \mathcal{T} \to [0, M]$ be a bounded measurable map with some $M > 0$, where $\mathcal{T} \equiv \{d_1, \cdots, d_J\}$ is a finite subset of $\mathbf{R}$.*

Then, for each randomized decision $\delta : \mathbf{R}^m \to \Delta_{\mathcal{T}}$, with $\Delta_{\mathcal{T}}$ denoting the simplex on $\mathbf{R}^J$, there exists a measurable map $v : \mathbf{R}^m \to \mathcal{T}$ such that for each $P \in \mathcal{P}$,

$$\sum_{j=1}^J \int_{\mathbf{R}^m} W_P(\mathbf{x}, d_j)\delta_j(\mathbf{x}) dP(\mathbf{x}) = \int_{\mathbf{R}^m} W_P(\mathbf{x}, v(\mathbf{x})) dP(\mathbf{x}),$$

where $\delta_j(\mathbf{x})$ denotes the j-th entry of $\delta(\mathbf{x})$.

Proof of Lemma 2: We write

$$\begin{aligned} & \sqrt{n}\{\hat{\theta} - g(\boldsymbol{\beta}_n(h))\} \\ = \ & \sqrt{n}\{\hat{\theta} - g(\boldsymbol{\beta}_n(0))\} - g(\sqrt{n}(\boldsymbol{\beta}_n(h) - \boldsymbol{\beta}_n(0)) + \boldsymbol{\beta}_{n,g}), \end{aligned}$$

where $\boldsymbol{\beta}_{n,g} = \sqrt{n}\{\boldsymbol{\beta}_n(0) - g(\boldsymbol{\beta}_n(0))\}$. Applying Prohorov's Theorem and invoking Assumption 2, we note that for any subsequence of $\{P_{n,0}\}$, there exists a further subsequence along which $\boldsymbol{\beta}_{n',g} \to \boldsymbol{\beta}_g$, and

$$\left[\begin{array}{c} \sqrt{n}\{\hat{\theta} - g(\boldsymbol{\beta}_{n'}(h))\} \\ \log dP_{n',h}/dP_{n',0} \end{array}\right] \xrightarrow{D} \left[\begin{array}{c} V - g(\dot{\boldsymbol{\beta}}(h) + \boldsymbol{\beta}_g) \\ \zeta(h) - \frac{1}{2}\langle h, h\rangle \end{array}\right],$$

where $V \in \bar{\mathbf{R}}^d$ has a potentially deficient distribution and $\boldsymbol{\beta}_g$ is a nonstochastic vector in $\bar{\mathbf{R}}^d$. From now on, it suffices to focus only on these subsequences. Since $g(\boldsymbol{\beta}_{n,g}) = 0$ for all $n \geq 1$ by definition and by Assumption 1(i), we have $g(\boldsymbol{\beta}_g) = 0$, whenever $\boldsymbol{\beta}_g \in \mathbf{R}^d$.

As in the proof of Theorem 3.11.5 of van der Vaart and Wellner (1996), choose an orthonormal basis $\{h_i\}_{i=1}^{\infty}$ from $\bar{H}$. We fix m and take $\{h_i\}_{i=1}^{m} \subset H$ and consider $h(\mathbf{p}) = \sum p_i h_i$ for some $\mathbf{p} = (p_i)_{i=1}^m \in \mathbf{R}^m$ such that $h(\mathbf{p}) \in H$. Fix $\lambda > 0$ and let $F_{\lambda,m}(\mathbf{p})$ be as defined prior to Lemma A1. Hence note that for fixed $M > 0$,

$$\begin{aligned} \liminf_{n\to\infty} \mathcal{R}_n(\hat{\theta}) &= \liminf_{n\to\infty} \sup_{h\in H} \mathbf{E}_h\left[\tau_M\left(|V_{n,h}|\right)\right] \\ &\geq \liminf_{n\to\infty} \int \mathbf{E}_{h(\mathbf{p})}\left[\tau_M\left(\left|V_{n,h(\mathbf{p})}\right|\right)\right] dF_{\lambda,m}(\mathbf{p}), \end{aligned}$$

where $V_{n,h} \equiv \sqrt{n}\{\hat{\theta} - g(\boldsymbol{\beta}_n(h))\}$. By Lemma A1, the last liminf is equal to $\mathbf{E}[\tau_M(|g(Z_{\lambda,m} + W_{\lambda,m} + \boldsymbol{\beta}_g)|)]$, where $Z_{\lambda,m}$ is as defined prior to Lemma A1 and $W_{\lambda,m} \in \bar{\mathbf{R}}$ is a random variable having a potentially deficient distribution and independent of $Z_{\lambda,m}$. It is not hard to see that for any sequence $\lambda_m \to 0$ as $m \to \infty$, $Z_{\lambda_m,m}$ converges in distribution to Z. Since $\{[Z'_{\lambda,m}, W_{\lambda,m}]' : (\lambda, m) \in (0,\infty) \times \{1, 2, \cdots\}\}$ is uniformly tight in $\bar{\mathbf{R}}^{d+1}$, by Prohorov's Theorem, for any subsequence of $\{\lambda_m\}_{m=1}^{\infty}$, there exists a further subsequence such that $[Z'_{\lambda_m,m}, W_{\lambda_m,m}]'$ converges in distribution to $[Z', W]'$ for some random variable W having a potentially deficient distribution. Since $g(\boldsymbol{\beta}_g) = 0$ whenever $\boldsymbol{\beta}_g \in \mathbf{R}^d$, we bound $\liminf_{n\to\infty}\mathcal{R}_n(\hat{\theta})$ from below by

$$\begin{aligned} \sup_{\mathbf{r}\in\mathbf{R}^d : g(\mathbf{r})=0} \mathbf{E}\left[\tau_M\left(|g(Z + W + \mathbf{r})|\right)\right] &= \sup_{\mathbf{r}\in\mathbf{R}^d} \mathbf{E}\left[\tau_M\left(|g(Z + W + \mathbf{r} - g(\mathbf{r}))|\right)\right] \\ &= \sup_{\mathbf{r}\in\mathbf{R}^d} \int \mathbf{E}\left[\tau_M\left(|g(Z + \mathbf{r}) - g(\mathbf{r}) + w|\right)\right] dF(w), \end{aligned}$$

where F is the (potentially deficient) distribution of W. The first equality above follows because $\{\mathbf{r} \in \mathbf{R}^d : g(\mathbf{r}) = 0\} = \{\mathbf{r} - g(\mathbf{r}) : \mathbf{r} \in \mathbf{R}^d\}$ by the translation equivariance of g. Thus we conclude that

$$\liminf_{n\to\infty} \mathcal{R}_n(\hat{\theta}) \geq \lim_{M\uparrow\infty} \inf_{F\in\mathcal{F}^*} \sup_{\mathbf{r}\in\mathbf{R}^d} \int \mathbf{E}\left[\tau_M(|g(Z + \mathbf{r}) - g(\mathbf{r}) + w|)\right] dF(w), \tag{6.4}$$

where $\mathcal{F}^*$ is the collection of potentially deficient distributions on $\mathcal{B}(\mathbf{R})$.

Fix $F \in \mathcal{F}^*$ and let $W \in \bar{\mathbf{R}}$ have distribution F. Suppose that $P\{W \in \bar{\mathbf{R}} \backslash \mathbf{R}\} > 0$. For all $\mathbf{r} \in \mathbf{R}^d$,

$$\mathbf{E}\left[\tau_M\left(|g(Z + W + \mathbf{r} - g(\mathbf{r}))|\right)\right] \geq \mathbf{E}\left[\tau_M\left(|g(Z + W + \mathbf{r} - g(\mathbf{r}))|\right) 1\{W \in \bar{\mathbf{R}} \backslash \mathbf{R}\}\right].$$

For all $u \in \bar{\mathbf{R}} \backslash \mathbf{R}$, $Z + u + \mathbf{r} - g(\mathbf{r}) \in \bar{\mathbf{R}} \backslash \mathbf{R}$, because $Z + \mathbf{r} - g(\mathbf{r}) \in \mathbf{R}$. Since u is a scalar and $Z + \mathbf{r} - g(\mathbf{r}) \in \mathbf{R}$, we have by Assumption 1(i)(a),

$$g(Z + u + \mathbf{r} - g(\mathbf{r})) = g(Z + \mathbf{r} - g(\mathbf{r})) + u \in \{-\infty, \infty\}, \text{ almost everywhere.}$$

Hence for $u \in \bar{\mathbf{R}} \backslash \mathbf{R}$, $\tau_M (|g(Z + u + \mathbf{r} - g(\mathbf{r}))|) = M$, a.e., so that

$$\begin{aligned} & \mathbf{E}\left[\tau_M (|g(Z + W + \mathbf{r} - g(\mathbf{r}))|) 1\{W \in \bar{\mathbf{R}} \backslash \mathbf{R}\}\right] \\ = \; & M \cdot P\{W \in \bar{\mathbf{R}} \backslash \mathbf{R}\} \uparrow \infty, \end{aligned}$$

as $M \uparrow \infty$. Therefore, the lower bound in (6.4) remains the same if we replace $\mathcal{F}^*$ by $\mathcal{F}$. Since τ_M increases in M, we obtain the desired bound by sending $M \uparrow \infty$. ■

Proof of Theorem 1: In view of Lemma 1, it suffices to show that for each $M > 0$,

$$\begin{aligned} & \inf_{F \in \mathcal{F}} \sup_{\mathbf{r} \in \mathbf{R}^d} \int \mathbf{E}\left[\tau_M(|g(Z + \mathbf{r}) - g(\mathbf{r}) + w|)\right] dF(w) \\ \geq \; & \inf_{c \in \mathbf{R}} \sup_{\mathbf{r} \in \mathbf{R}^d} \mathbf{E}\left[\tau_M(|g(Z + \mathbf{r}) - g(\mathbf{r}) + c|)\right], \end{aligned} \tag{6.5}$$

because $\mathcal{F}$ includes point masses at points in $\mathbf{R}$. The proof is complete then by sending $M \uparrow \infty$, because the last infimum is increasing in $M > 0$. Let $W \in \mathbf{R}$ be a random variable having distribution $F_W \in \mathcal{F}$, and choose arbitrary $M_1 > 0$ which may depend on the choice of $F_W \in \mathcal{F}$,

$$\begin{aligned} & \sup_{\mathbf{r} \in [-M_1, M_1]^d} \mathbf{E}\left[\tau_M (|g(Z + W + \mathbf{r} - g(\mathbf{r}))|) 1\{W \in [-M_1, M_1]\}\right] \\ \geq \; & \inf_{u \in \mathbf{R}} \sup_{\mathbf{r} \in [-M_1, M_1]^d} \mathbf{E}\left[\tau_M (|g(Z + u + \mathbf{r} - g(\mathbf{r}))|)\right] P\{W \in [-M_1, M_1]\}. \end{aligned} \tag{6.6}$$

Once this inequality is established, we send $M_1 \uparrow \infty$ on both sides to obtain the following inequality:

$$\begin{aligned} & \sup_{\mathbf{r} \in \mathbf{R}^d} \mathbf{E}\left[\tau_M (|g(Z + W + \mathbf{r} - g(\mathbf{r}))|)\right] \\ \geq \; & \inf_{u \in \mathbf{R}} \sup_{\mathbf{r} \in \mathbf{R}^d} \mathbf{E}\left[\tau_M (|g(Z + u + \mathbf{r} - g(\mathbf{r}))|)\right]. \end{aligned}$$

(Note that by the definition of $\mathcal{F}$, the distribution of W is tight in $\mathbf{R}$.) And since the lower bound does not depend on the choice of F_W, we take infimum over $F_W \in \mathcal{F}$ of the left hand side of the above inequality to deduce (6.5).

We fix large enough $M_1 > 0$ so that $P\{W \in [-M_1, M_1]\} > 0$. Then

$$\begin{aligned} & \sup_{\mathbf{r} \in [-M_1, M_1]^d} \mathbf{E}\left[\tau_M (|g(Z + W + \mathbf{r} - g(\mathbf{r}))|) 1\{W \in [-M_1, M_1]\}\right] \\ \geq \; & P\{W \in [-M_1, M_1]\} \times \sup_{\mathbf{r} \in [-M_1, M_1]^d} \int_{[-M_1, M_1]} \bar{\kappa}(u, \mathbf{r}) dF_{M_1}(u) \end{aligned}$$

where $\bar{\kappa}(u, \mathbf{r}) \equiv \mathbf{E}\left[\kappa(Z + u + \mathbf{r} - g(\mathbf{r}))\right]$ and $\kappa(\mathbf{x}) \equiv \tau_M(|g(\mathbf{x})|)$, and

$$\int_{A \cap [-M_1, M_1]} dF_{M_1}(u) = \int_{A \cap [-M_1, M_1]} dF_W(u) / P\{W \in [-M_1, M_1]\},$$

for all $A \in \mathcal{B}(\mathbf{R})$. Take $K > 0$ and let $\mathcal{R}_K \equiv \{\mathbf{r}_1, \cdots, \mathbf{r}_K\} \subset [-M_1, M_1]^d$ be a finite set such that $\mathcal{R}_K$ become dense in $[-M_1, M_1]^d$ as $K \to \infty$. Since $\int \bar{\kappa}(u, \mathbf{r}) dF_{M_1}(u)$ is Lipschitz in $\mathbf{r}$ (due to Assumption 5), for any fixed $\eta > 0$, we can take $\mathcal{R}_K$ such that

$$\max_{\mathbf{r} \in \mathcal{R}_K} \int \bar{\kappa}(u, \mathbf{r}) dF_{M_1}(u) \geq \sup_{\mathbf{r} \in [-M_1, M_1]^d} \int \bar{\kappa}(u, \mathbf{r}) dF_{M_1}(u) - \eta \tag{6.7}$$

Let $\mathcal{F}_{M_1}$ be the collection of probabilities with support confined to $[-M_1, M_1]$, so that we deduce that

$$\begin{aligned} &\sup_{\mathbf{r} \in [-M_1, M_1]^d} \mathbf{E}\left[\tau_M\left(|g(Z + W + \mathbf{r} - g(\mathbf{r}))|\right) 1\left\{W \in [-M_1, M_1]\right\}\right] \\ &\geq \; P\{W \in [-M_1, M_1]\} \left(\inf_{F \in \mathcal{F}_{M_1}} \max_{\mathbf{r} \in \mathcal{R}_K} \int_{[-M_1, M_1]} \bar{\kappa}(u, \mathbf{r}) dF(u) - \eta \right). \end{aligned} \tag{6.8}$$

Since $\mathcal{F}_{M_1}$ is uniformly tight, $\mathcal{F}_{M_1}$ is totally bounded for $d_{\mathcal{P}}$ defined in (6.3) (e.g. Theorems 11.5.4 of Dudley (2002)). Hence we fix $\varepsilon > 0$ and choose $F_1, \cdots, F_N$ such that for any $F \in \mathcal{F}_{M_1}$, there exists $j \in \{1, \cdots, N\}$ such that $d_{\mathcal{P}}(F_j, F) < \varepsilon$. Hence for $F \in \mathcal{F}_{M_1}$, we take F_j such that $d_{\mathcal{P}}(F_j, F) < \varepsilon$, so that

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

Since $\bar{\kappa}(\cdot, \mathbf{r})$ is Lipschitz continuous and bounded on $[-M_1, M_1]$, $C_K \equiv \max_{\mathbf{r} \in \mathcal{R}_K} ||\bar{\kappa}(\cdot, \mathbf{r})||_{BL} < \infty$. Therefore,

$$\inf_{F \in \mathcal{F}_{M_1}} \max_{\mathbf{r} \in \mathcal{R}_K} \int \bar{\kappa}(u, \mathbf{r}) dF(u) \geq \min_{1 \leq j \leq N} \max_{\mathbf{r} \in \mathcal{R}_K} \int \bar{\kappa}(u, \mathbf{r}) dF_j(u) - C_K \varepsilon. \tag{6.9}$$

By Lemma A2, we can select for each F_j and for each $\mathbf{r}_k \in \mathcal{R}_K$ a distribution $G_{j,k}$ with a finite set support such that

$$\int \bar{\kappa}(u, \mathbf{r}_k) dF_j(u) = \int \bar{\kappa}(u, \mathbf{r}_k) dG_{j,k}(u). \tag{6.10}$$

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

$$\begin{aligned} \min_{1 \leq j \leq N} \max_{1 \leq k \leq K} \int \bar{\kappa}(u, \mathbf{r}_k) dG_{j,k}(u) \;&\geq\; \inf_{G \in \mathcal{F}_{K,N}} \max_{\mathbf{r} \in \mathcal{R}_K} \int \bar{\kappa}(u, \mathbf{r}) dG(u) \\ &=\; \inf_{G \in \mathcal{F}_{K,N}} \max_{\mathbf{r} \in \mathcal{R}_K} \int \int \kappa(z + u) d\Lambda_{\mathbf{r}}(z) dG(u), \end{aligned}$$

where $\Lambda_{\mathbf{r}}$ is the distribution of $Z + \mathbf{r} - g(\mathbf{r})$.

For the last $\inf_{G \in \mathcal{F}_{K,N}} \max_{\mathbf{r} \in \mathcal{R}_K}$, we regard $Z + \mathbf{r} - g(\mathbf{r})$ as a state variable distributed by $\Lambda_{\mathbf{r}}$ with $\Lambda_{\mathbf{r}}$ parametrized by $\mathbf{r}$ in a finite set $\mathcal{R}_K$. We view the conditional distribution of W given $Z + \mathbf{r} - g(\mathbf{r})$ (which is $G \in \mathcal{F}_{K,N}$) as a randomized decision. Each randomized decision has a finite set support contained in $\mathcal{T}_{K,N}$, and for each $\mathbf{r} \in \mathcal{R}_K$, $\Lambda_{\mathbf{r}}$ is atomless. Finally κ is bounded. We apply Lemma A3 to find that the last $\inf_{G \in \mathcal{F}_{K,N}} \max_{\mathbf{r} \in \mathcal{R}_K}$ is equal to that with randomized decisions replaced by nonrandomized decisions (the collection $\mathcal{P}$ and the finite set $\{d_1, \cdots, d_J\}$ in the lemma correspond to $\{\Lambda_{\mathbf{r}} : \mathbf{r} \in \mathcal{R}_K\}$ and $\mathcal{T}_{K,N}$ respectively here), whereby we can now write it as

$$\min_{u \in \mathcal{T}_{K,N}} \max_{\mathbf{r} \in \mathcal{R}_K} \int \kappa(z+u) d\Lambda_{\mathbf{r}}(z) = \min_{u \in \mathcal{T}_{K,N}} \max_{\mathbf{r} \in \mathcal{R}_K} \mathbf{E}\left[\tau_M(|g(Z+u+\mathbf{r}) - g(\mathbf{r})|)\right].$$

Since $\mathbf{E}\left[\tau_M(|g(Z+u+\mathbf{r}) - g(\mathbf{r})|)\right]$ is Lipschitz continuous in u and $\mathbf{r}$, we send $\varepsilon \downarrow 0$ and then $\eta \downarrow 0$ (along with $K \uparrow \infty$) to conclude from (6.7), (6.9), and (6.10) that

$$\begin{aligned} &\inf_{F \in \mathcal{F}_{M_1}} \sup_{\mathbf{r} \in [-M_1, M_1]^d} \int \mathbf{E}\left[\kappa(Z+u+\mathbf{r} - g(\mathbf{r}))\right] dF(u) \\ &\geq \inf_{u \in \mathbf{R}} \sup_{\mathbf{r} \in [-M_1, M_1]^d} \mathbf{E}\left[\tau_M(|g(Z+\mathbf{r}) - g(\mathbf{r}) + u|)\right]. \end{aligned}$$

Therefore, combining this with (6.8), we obtain (6.6). ■

For given $M_1, a > 0$ and $c \in \mathbf{R}$, define

$$B_{M_1}(c; a) \equiv \sup_{\mathbf{r} \in [-M_1, M_1]^d} \mathbf{E}\left[\tau_{M_1}(a|g(Z+\mathbf{r}) - g(\mathbf{r}) + c|)\right], \tag{6.11}$$

and

$$E_{M_1}(a) \equiv \left\{ c \in [-M_1, M_1] : B_{M_1}(c; a) \leq \inf_{c_1 \in [-M_1, M_1]} B_{M_1}(c_1; a) \right\}.$$

Let $c^*_{M_1}(a) \equiv 0.5\{\sup E_{M_1}(a) + \inf E_{M_1}(a)\}$. We simply write $E_{M_1} = E_{M_1}(1)$, $B_{M_1}(c) = B_{M_1}(c; 1)$, and $c^*_{M_1} = c^*_{M_1}(1)$.

Lemma A4: *Suppose that Assumptions* 1-2 *hold.*
(i) *There exists* $M_0 > 0$ *such that for all* $M_1 > M_0$ *and for all* $a > 0$,

$$c^*_{M_1} = c^*_{M_1}(a).$$

(ii) *There exists* M_0 *such that for any* $M_1 > M_0$ *and* $\varepsilon > 0$,

$$\sup_{h \in H} P_{n,h}\left\{\left|\hat{c}_{M_1} - c^*_{M_1}\right| > \varepsilon\right\} \to 0,$$

as $n, L \to \infty$ *jointly.*

Proof: (i) Define $B(c; a)$ to be $B_{M_1}(c; a)$ with $M_1 = \infty$. For any $a > 0$, we have $B(c; a) \uparrow \infty$, as $|c| \uparrow \infty$. Therefore, the set

$$S \equiv \text{argmin}_{c \in \mathbf{R}} B(c; a) \tag{6.12}$$

is bounded in $\mathbf{R}$. Note that the set S does not depend on a because τ is a strictly increasing function. Increase M_1 large enough so that $S \subset [-M_1, M_1]$. Then certainly for any $a > 0$,

$$c^*_{M_1}(a) = \frac{1}{2}\{\sup S + \inf S\},$$

delivering the desired result.

(ii) Let S be the bounded set defined in (6.12). Take large M_1 such that for some $\bar{\varepsilon} > 0$, $S \subset [M_1 + \bar{\varepsilon}, M_1 - \bar{\varepsilon}]$. Let the Hausdorff distance between the two subsets E_1 and E_2 of $\mathbf{R}$ be denoted by $d_H(E_1, E_2)$. First we show that

$$d_H(E_{M_1}, \hat{E}_{M_1}) \to_P 0, \tag{6.13}$$

as $n \to \infty$ and $L \to \infty$ uniformly over $h \in H$. For this, we use arguments in the proof of Theorem 3.1 of Chernozhukov, Hong and Tamer (2007). Fix $\varepsilon \in (0, \bar{\varepsilon})$ and let $E_{M_1}^{\varepsilon} \equiv \{x \in [-M_1, M_1] : \sup_{y \in E_{M_1}} |x - y| \leq \varepsilon\}$. It suffices for (6.13) to show that for any $\varepsilon > 0$,

$$\begin{aligned}&\text{(a) } \inf_{h \in H} P_{n,h} \left\{ \sup_{c \in E_{M_1}} \hat{B}_{M_1}(c) \leq \inf_{c \in [-M_1, M_1]} \hat{B}_{M_1}(c) + \eta_{n,L} \right\} \to 1, \\ &\text{(b) } \inf_{h \in H} P_{n,h} \left\{ \sup_{c \in \hat{E}_{M_1}} B_{M_1}(c) < \inf_{c \in [-M_1, M_1] \backslash E_{M_1}^{\varepsilon}} B_{M_1}(c) \right\} \to 1,\end{aligned}$$

as $n, L \to \infty$ jointly, where the last term $o_P(1)$ is uniform over $h \in H$. This is because (a) implies $\inf_{h \in H} P_{n,h}\{E_{M_1} \subset \hat{E}_{M_1}\} \to 1$ and (b) implies that $\inf_{h \in H} P_{n,h}\{\hat{E}_{M_1} \cap ([-M_1, M_1] \backslash E_{M_1}^{\varepsilon}) = \varnothing\} \to 1$ so that $\inf_{h \in H} P_{n,h}\{\hat{E}_{M_1} \subset E_{M_1}^{\varepsilon}\} \to 1$, and hence for any $\varepsilon > 0$,

$$\sup_{h \in H} P_{n,h} \left\{ d_H(E_{M_1}, \hat{E}_{M_1}) > \varepsilon \right\} \to 0, \text{ as } n, L \to \infty \text{ jointly.}$$

We focus on (a). First, define $f(\xi; c, \mathbf{r}) \equiv \tau_{M_1}\left(|g(\xi + \mathbf{r}) - g(\mathbf{r}) + c|\right)$ and $\mathcal{J} \equiv \{f(\cdot; c, \mathbf{r}) : (c, \mathbf{r}) \in [-M_1, M_1] \times [-M_1, M_1]^d\}$. The class $\mathcal{J}$ is uniformly bounded, and $f(\xi; c, \mathbf{r})$ is Lipschitz continuous in $(c, \mathbf{r}) \in [-M_1, M_1] \times [-M_1, M_1]^d$. Using the maximal inequality (e.g. Theorems 2.14.2 and 2.7.11 of van der Vaart and Wellner (1996)) and Assumptions 1, 2, and 6(i), we find that for some $C_{M_1} > 0$ that depends only on $M_1 > 0$,

$$\mathbf{E}\left[\sup_{c \in [-M_1, M_1]} \left| \hat{B}_{M_1}(c) - B_{M_1}(c) \right| \right] \leq C_{M_1} \left\{ L^{-1/2} + n^{-1/2} \right\}. \tag{6.14}$$

The last bound does not depend on $h \in H$. 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 $L \to \infty$.

Now let us turn to (b). Fix $\varepsilon > 0$. By (6.14), with probability approaching 1 uniformly over $h \in H$,

$$\begin{aligned}\sup_{c \in \hat{E}_{M_1}} B_{M_1}(c) &\leq \sup_{c \in \hat{E}_{M_1}} \hat{B}_{M_1}(c) + o_P(1) \leq \sup_{c \in E_{M_1}^{\varepsilon/2}} \hat{B}_{M_1}(c) + o_P(1) \\ &\leq \sup_{c \in E_{M_1}^{\varepsilon/2}} B_{M_1}(c) + o_P(1),\end{aligned}$$

where the second inequality follows due to $\eta_{n,L} \to 0$ as $n, L \to \infty$ and (6.14). Since $\tau(\cdot)$ is strictly increasing, and Z has full support on $\mathbf{R}$ by Assumption 4, we have $\inf_{c \in [-M_1, M_1] \backslash E_{M_1}^{\varepsilon}} B_{M_1}(c) > \sup_{c \in E_{M_1}} B_{M_1}(c) \geq 0$. Note that this last supremum does not depend on $h \in H$. Hence we obtain (b).

For the main conclusion of the lemma, observe that $\left|\hat{c}_{M_1} - c_{M_1}^*\right|$ is equal to

$$\frac{1}{2}\left| \sup \hat{E}_{M_1} + \inf \hat{E}_{M_1} - \sup E_{M_1} - \inf E_{M_1} \right|$$

which we can write as

$$\begin{aligned}&\frac{1}{2}\left| \inf_{y \in \hat{E}_{M_1}} \{y - \sup E_{M_1}\} - \sup_{x \in E_{M_1}} \left\{ x - \inf \hat{E}_{M_1} \right\} \right| \\ =\ &\frac{1}{2}\left| \sup_{y \in \hat{E}_{M_1}} \inf(y - E_{M_1}) - \inf_{x \in E_{M_1}} \sup \left\{ x - \hat{E}_{M_1} \right\} \right|.\end{aligned}$$

We write the last term as

$$\frac{1}{2}\left|\sup_{y\in\hat{E}_{M_1}}\inf(y-E_{M_1})+\sup_{x\in E_{M_1}}\inf\left\{\hat{E}_{M_1}-x\right\}\right|\leq\frac{1}{2}\left\{\sup_{y\in\hat{E}_{M_1}}d(y,E_{M_1})+\sup_{x\in E_{M_1}}d(\hat{E}_{M_1},x)\right\},$$

where $d(y,A)=\inf_{x\in A}|y-x|$. The last term is bounded by $d_H(E_{M_1},\hat{E}_{M_1})$. The desired result follows from (6.13). ■

Proof of Theorem 2: Fix $M>0$ and $\varepsilon>0$, and take $M_1\geq M_2>M$ such that

$$\begin{aligned}\text{(6.15)}\qquad &\sup_{\mathbf{r}\in[-M_2,M_2]^d}\mathbf{E}\left[\tau_M(|g(Z+\mathbf{r})-g(\mathbf{r})+c^*_{M_1}|)\right]\\ &\geq\ \sup_{\mathbf{r}\in\mathbf{R}^d}\mathbf{E}\left[\tau_M(|g(Z+\mathbf{r})-g(\mathbf{r})+c^*_{M_1}|)\right]-\varepsilon.\end{aligned}$$

This is possible for any choice of $\varepsilon>0$ because $\tau_M(\cdot)$ is continuous and bounded by M. Note that

$$\begin{aligned}&\sup_{h\in H}\mathbf{E}_h\left[\tau_M(\sqrt{n}|\hat{\theta}-\theta_n(h)|)\right]\\ =\ &\sup_{h\in H}\mathbf{E}_h\left[\tau_M(\sqrt{n}|g(\hat{\boldsymbol{\beta}})+\hat{c}_{M_1}/\sqrt{n}-g(\boldsymbol{\beta}_n(h))|)\right]\\ \leq\ &\sup_{\mathbf{r}\in\mathbf{R}^d}\mathbf{E}_h\left[\tau_M(|g(\sqrt{n}\{\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_n(h)\}+\mathbf{r})-g(\mathbf{r})+\hat{c}_{M_1}|)\right].\end{aligned}$$

Using Lemma A4(ii) and Assumption 6, we observe that for all $t\in\mathbf{R}^d$,

$$P_{n,h}\{\sqrt{n}\{\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_n(h)\}+\mathbf{r}-g(\mathbf{r})+\hat{c}_{M_1}\leq t\}=P\left\{Z+\mathbf{r}-g(\mathbf{r})+c^*_{M_1}\leq t\right\}+o(1),$$

uniformly over $h\in H$. Since Z is a continuous random vector, the convergence is uniform over $(\mathbf{r},t)\in\mathbf{R}^d\times\mathbf{R}$. Therefore,

$$\begin{aligned}\limsup_{n\to\infty}\sup_{h\in H}\mathbf{E}_h\left[\tau_M(\sqrt{n}|\hat{\theta}-\theta_n(h)|)\right]\ &=\ \sup_{\mathbf{r}\in\mathbf{R}^d}\mathbf{E}\left[\tau_M(|g(Z+\mathbf{r})-g(\mathbf{r})+c^*_{M_1}|)\right]\\ &\leq\ \sup_{\mathbf{r}\in[-M_2,M_2]^d}\mathbf{E}\left[\tau_M(|g(Z+\mathbf{r})-g(\mathbf{r})+c^*_{M_1}|)\right]+\varepsilon,\end{aligned}$$

by (6.15). Since $M_1\geq M_2>M$, the last supremum is bounded by

$$\begin{aligned}&\sup_{\mathbf{r}\in[-M_1,M_1]^d}\mathbf{E}\left[\tau_{M_1}(|g(Z+\mathbf{r})-g(\mathbf{r})+c^*_{M_1}|)\right]\\ =\ &\inf_{-M_1\leq c\leq M_1}\sup_{\mathbf{r}\in[-M_1,M_1]^d}\mathbf{E}\left[\tau_{M_1}(|g(Z+\mathbf{r})-g(\mathbf{r})+c|)\right],\end{aligned}$$

where the equality follows by the definition of $c^*_{M_1}$. We conclude that

$$\limsup_{n\to\infty}\sup_{h\in H}\mathbf{E}_h\left[\tau_M(\sqrt{n}|\hat{\theta}-\theta_n(h)|)\right]\leq\inf_{-M_1\leq c\leq M_1}\sup_{\mathbf{r}\in\mathbf{R}^d}\mathbf{E}\left[\tau(|g(Z+\mathbf{r})-g(\mathbf{r})+c|)\right]+\varepsilon$$

Since the choice of ε and M_1 were arbitrary, sending $M_1\uparrow\infty$ and $M_2\uparrow\infty$ (along with $\varepsilon\downarrow 0$), and then sending $M\uparrow\infty$, we obtain the desired result. ■

Proof of Theorem 3: Suppose that $f(x)$ has a kink point at $x = m$. Then write

$$\begin{aligned} f(g(\boldsymbol{\beta}_n(h))) &= f(g(\boldsymbol{\beta}_n(h) - m) + m) - f(m) + f(m) \\ &= \tilde{f}(g(\tilde{\boldsymbol{\beta}}_n(h)) + f(m), \end{aligned}$$

where $\tilde{f}(x) = f(x+m) - f(m)$ and $\tilde{\boldsymbol{\beta}}_n(h) = \boldsymbol{\beta}_n(h) - m$. Certainly $\tilde{\boldsymbol{\beta}}_n(h)$ satisfies Assumption 3 for $\boldsymbol{\beta}_n(h)$ and $\tilde{f}$ satisfies Assumption 1(ii) for f, only with its kink point now at the origin. Therefore, we lose no generality by assuming that f has a kink point at the origin, i.e.,

$$f(x) = a_1 x 1\{x \geq 0\} + a_2 x 1\{x < 0\},$$

for some constants a_1 and a_2 in $\mathbf{R}$, and $s = \max\{|a_1|, |a_2|\}$. Let

$$\begin{aligned} H_{n,1}(b) &\equiv \{h \in H : g(\boldsymbol{\beta}_n(h)) \geq b\}, \text{ and} \\ H_{n,2}(b) &\equiv \{h \in H : g(\boldsymbol{\beta}_n(h)) \leq b\}. \end{aligned}$$

First, note that

$$\begin{aligned} \sup_{h\in H} \mathbf{E}_h\left[\tau(|\sqrt{n}\{\hat{\theta} - \theta_n(h)\}|)\right] &= \sup_{h\in H} \mathbf{E}_h\left[\tau(|\sqrt{n}\{\hat{\theta} - f(g(\boldsymbol{\beta}_n(h)))\}|)\right] \\ &\geq \max_{k=1,2} \sup_{h\in H_{n,k}(0)} \mathbf{E}_h\left[\tau_M\left(|\sqrt{n}\{\hat{\theta} - a_k g(\boldsymbol{\beta}_n(h))\}|\right)\right]. \end{aligned}$$

Now we employ an argument similar to one in the proof of Theorem 3.1 of Blumenthal and Cohen (1968a). We fix arbitrary $\varepsilon > 0$, and choose any large number $b > 0$. Note that from sufficiently large n on,

$$\begin{aligned} &\sup_{h\in H_{n,1}(0)} \mathbf{E}_h\left[\tau_M\left(|\sqrt{n}\{\hat{\theta} - a_1 g(\boldsymbol{\beta}_n(h))\}|\right)\right] \\ &= \sup_{h\in H_{n,1}(0)} \mathbf{E}_h\left[\tau_M\left(|\sqrt{n}\{\hat{\theta} - a_1 b - a_1 g(\boldsymbol{\beta}_n(h) - b)\}|\right)\right] \\ &\geq \sup_{h\in H_{n,1}(-b)} \mathbf{E}_h\left[\tau_M\left(|\sqrt{n}\{\tilde{\theta} - a_1 g(\boldsymbol{\beta}_n(h))\}|\right)\right] - \varepsilon, \end{aligned}$$

where $\tilde{\theta} \equiv \hat{\theta} - a_1 b$. Let $\tilde{V}_{n,h} \equiv \sqrt{n}\{\tilde{\theta} - g(\boldsymbol{\beta}_n(h))\}$, $h(\mathbf{p})$, $\mathbf{p} = (p_i)_{i=1}^m \in \mathbf{R}^m$, and $F_{\lambda,m}(\mathbf{p})$ be as in the proof of Lemma 2, so that we have

$$\begin{aligned} &\liminf_{n\to\infty} \sup_{h\in H_{n,1}(-b)} \mathbf{E}_h\left[\tau_M\left(|\sqrt{n}\{\tilde{\theta} - a_1 g(\boldsymbol{\beta}_n(h))|\right)\right] \\ &\geq \liminf_{n\to\infty} \int_{\{\mathbf{p}\in\mathbf{R}^m : h(\mathbf{p})\in H_{n,1}(-b)\}} \mathbf{E}_h\left[\tau_M\left(|\tilde{V}_{n,h(\mathbf{p})}|\right)\right] dF_{\lambda,m}(\mathbf{p}). \end{aligned}$$

Since τ_M is bounded by M, we take b large enough so that the last liminf is bounded from below by

$$\liminf_{n\to\infty} \int \mathbf{E}_h\left[\tau_M\left(|\tilde{V}_{n,h(\mathbf{p})}|\right)\right] dF_{\lambda,m}(\mathbf{p}) - \varepsilon.$$

By following the same arguments as in the proofs of Lemma 2 and Theorem 1, we deduce that the last liminf is bounded from below by

$$\inf_{c\in\mathbf{R}} \sup_{\mathbf{r}\in\mathbf{R}^d} \mathbf{E}\left[\tau_M(|a_1||g(Z+\mathbf{r}) - g(\mathbf{r}) + c|)\right] - \varepsilon.$$

We proceed similarly with $\sup_{h\in H_{n,2}(0)} \mathbf{E}_h[\tau_M(|\sqrt{n}\{\hat{\theta}-a_2 g(\boldsymbol{\beta}_n(h))\}|)]$ to conclude that

$$\begin{aligned} & \liminf_{n\to\infty} \sup_{h\in H} \mathbf{E}_h[\tau(|\sqrt{n}\{\hat{\theta}-\theta_n(h)\}|)] \\ \geq\ & \max_{k=1,2} \inf_{c\in\mathbf{R}} \sup_{\mathbf{r}\in\mathbf{R}^d} \mathbf{E}\left[\tau_M(|a_k||g(Z+\mathbf{r})-g(\mathbf{r})+c|)\right]-3\varepsilon \\ =\ & \inf_{c\in\mathbf{R}} \sup_{\mathbf{r}\in\mathbf{R}^d} \mathbf{E}\left[\tau_M(s|g(Z+\mathbf{r})-g(\mathbf{r})+c|)\right]-3\varepsilon, \end{aligned}$$

where the last equality follows because τ_M is an increasing function. By sending $M\uparrow\infty$ and $\varepsilon\downarrow 0$, we obtain the desired result. ■

PROOF OF THEOREM 4: First, observe that a real valued map that assigns $y\in\mathbf{R}$ to $f(y)/s$ is a contraction mapping, because the maximum absolute slope of the line segments of f is equal to s. Hence for $M>0$,

$$\begin{aligned} & \sup_{h\in H} \mathbf{E}_h\left[\tau_M(|\sqrt{n}\{\tilde{\theta}_{mx}-\theta_n(h)\}|)\right] \\ \leq\ & \sup_{h\in H} \mathbf{E}_h\left[\tau_M(s\sqrt{n}|g(\hat{\boldsymbol{\beta}})+\hat{c}_{M_1}/\sqrt{n}-g(\boldsymbol{\beta}_n(h))|)\right]. \end{aligned}$$

Fix $\varepsilon>0$, choose $M_1\geq M$, and follow the proof of Theorem 2 to find that the $\limsup_{n\to\infty}$ of the last supremum is bounded by

$$\sup_{\mathbf{r}\in[-M_1,M_1]^d} \mathbf{E}\left[\tau_{M_1}(s|g(Z+\mathbf{r})-g(\mathbf{r})+c^*_{M_1}|)\right]+\varepsilon.$$

By Lemma A4(i), from some large M_1 on, the last supremum is equal to

$$\begin{aligned} & \sup_{\mathbf{r}\in[-M_1,M_1]^d} \mathbf{E}\left[\tau_{M_1}(s|g(Z+\mathbf{r})-g(\mathbf{r})+c^*_{M_1}(s)|)\right] \\ =\ & \inf_{c\in[-M_1,M_1]} \sup_{\mathbf{r}\in[-M_1,M_1]^d} \mathbf{E}\left[\tau_{M_1}(s|g(Z+\mathbf{r})-g(\mathbf{r})+c^*_{M_1}(s)|)\right] \\ \leq\ & \inf_{c\in[-M_1,M_1]} \sup_{\mathbf{r}\in\mathbf{R}^d} \mathbf{E}\left[\tau(s|g(Z+\mathbf{r})-g(\mathbf{r})+c|)\right]. \end{aligned}$$

Sending $M_1\uparrow\infty$, we obtain the desired result. ■

7. ACKNOWLEDGEMENT

I thank Keisuke Hirano, Marcelo Moreira, Ulrich Müller and Frank Schorfheide for valuable comments. This research was supported by the Social Sciences and Humanities Research Council in Canada.

REFERENCES

[1] 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.
[2] 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.
[3] BICKEL, P. J. , A.J. KLAASSEN, Y. RIKOV, AND J. A. WELLNER (1993): *Efficient and Adaptive Estimation for Semiparametric Models*, Springer Verlag, New York.
[4] BLUMENTHAL, S. AND A. COHEN (1968a): "Estimation of the larger translation parameter," *Annals of Mathematical Statistics*, 39, 502-516.

[5] BLUMENTHAL, S. AND A. COHEN (1968b): "Estimation of the larger of two normal means," *Journal of the American Statistical Association*, 63, 861-876.
[6] BOUBARKI, N. (1970): *Theory of Sets*, Springer-Verlag.
[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] CHERNOZHUKOV, V., S. LEE AND A. ROSEN (2009): "Intersection bounds: estimation and inference," Cemmap Working Paper, CWP 19/09.
[10] DOSS, H. AND J. SETHURAMAN (1989): "The price of bias reduction when there is no unbiased estimate," *Annals of Statistics,* 17, 440-442.
[11] DUDLEY, R. M. (2002): *Real Analysis and Probability*, Cambridge University Press, New York.
[12] 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.
[13] HAILE, P. A. AND E. TAMER (2003): "Inference with an incomplete model of English auctions," *Journal of Political Economy*, 111, 1-51.
[14] HÁJEK, J. (1972): "Local asymptotic minimax and admissibility in estimation," in L. Le Cam, J. Neyman and E. L. Scott, eds, *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*, Vol 1, University of California Press, Berkeley, p.175-194.
[15] HIRANO, K. AND J. PORTER (2012): "Impossibility results for nondifferentiable functionals," Forthcoming in *Econometrica*.
[16] LE CAM, L. (1979): "On a theorem of J. Hájek," in J. Jurečková, ed. *Contributions to Statistics - Hájek Memorial Volume*, Akademian, Prague, p.119-135.
[17] LEHMANN E. L. (1986): *Testing Statistical Hypotheses*, 2nd Ed. Chapman&Hall, New York.
[18] 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.
[19] MANSKI C. F. AND J. PEPPER (2000): "Monotone instrumental variables: with an application to the returns to schooling," *Econometrica* 68, 997–1010.
[20] MILGROM, P. J. AND R. J. WEBER (1985): "Distributional strategies for games with incomplete information," *Mathematics of Operations Research*, 10, 619-632.
[21] MOORS, J. J. A. (1981): "Inadmissibility of linearly invariant estimators in truncated parameter spaces," *Journal of the American Statistical Association*, 76, 910-915.
[22] MORIGUTI, S. (1951): "Extremal properties of extreme value distribution," *Annals of Mathematical Statistics*, 22, 523-536.
[23] STRASSER, H. (1985): *Mathematical Theory of Statistics*, Walter de Gruyter, New York.
[24] TAKAGI, Y. (1994): "Local asymptotic minimax risk bounds for asymmetric loss functions," *Annals of Statistics* 22, 39–48.
[25] VAN DER VAART, A. W. (1989): "On the asymptotic information bound," *Annals of Statistics* 17, 1487-1500.
[26] VAN DER VAART, A. W. (1991): "On differentiable functionals," *Annals of Statistics* 19, 178-204.
[27] VAN DER VAART, A. W. AND J. A. WELLNER (1996): *Weak Convergence and Empirical Processes*, Springer-Verlag, New York.
[28] VAN EEDEN, C., AND J. V. ZIDEK (2004): "Combining the data from two normal populations to estimate the mean of one when their means difference is bounded," *Journal of Multivariate Analysis* 88, 19-46.

DEPARTMENT OF ECONOMICS, UNIVERSITY OF BRITISH COLUMBIA, 997 - 1873 EAST MALL, VANCOUVER, BC, V6T 1Z1, CANADA
E-mail address: kysong@mail.ubc.ca