Point Decisions for Interval-Identified Parameters

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

October 5, 2009

Abstract

This paper focuses on a situation where the decision-maker prefers to make a point-decision when the object of interest is interval-identified. Such a situation frequently arises when the interval-identified parameter is closely related to an optimal policy decision. To obtain a reasonable decision, this paper slices asymptotic normal experiments into subclasses corresponding to localized interval lengths, and finds a local asymptotic minimax decision for each subclass. Then, this paper suggests a decision that is based on the subclass minimax decisions, and explains the sense in which the decision is reasonable. One remarkable aspect of this solution is that the optimality of the solution remains intact even when the order of the interval bounds is misspecified. A small sample simulation study illustrates the solution's usefulness.

Key words and Phrases: Partial Identification, Inequality Restrictions, Local Asymptotic Minimax Estimation, Semiparametric Efficiency

JEL Classifications: C01, C13, C14, C44

[1]This research has benefited from conversations with Xu Cheng, Aureo de Paula, Frank Schorfheide, and Xun Tang who kindly offered valuable comments. All errors are mine. Corresponding address: Kyungchul Song, Department of Economics, University of Pennsylvania, 528 McNeil Building, 3718 Locust Walk, Philadelphia, Pennsylvania 19104-6297. Email address: kysong@sas.upenn.edu.

"Ranges are for cattle. Give me a number."

A remark ascribed to Lyndon B. Johnson, requoted from Manski (2007), p.8.

1 Introduction

Many objects of inference in economics are directly related to decisions that are to be implemented in practice. For example, estimation of willingness-to-pay in discrete choice models is closely related to transportation and environmental policies or marketing strategies. Also, optimal treatment decisions are based on the estimated treatment effects in various program evaluations or medical studies. In such an environment, a point decision about the object of interest is preferred to a set decision for practical reasons. A natural way to proceed in this case would be to introduce identifying restrictions for the object of interest and obtain its reasonable estimator using data. However, the decision-maker faces a dilemma when the empirical results are sensitive to the identifying restrictions that have no *a priori* justification other than that of practical convenience. Relying on the unjustifiable restrictions will erode the validity of the decision, while shedding them will yield no guidance as to a reasonable point decision. This paper attempts to address this dilemma by offering an optimal point decision when the object of interest is interval-identified.

Interval-identification frequently arises when the ultimate object of interest is a reduced form parameter that is partially identified. Examples abound, including predicted outcomes with sample-selected observations (Manski (1989)) or with covariates observed to lie in intervals (Manski and Tamer (2002)), local features of a structural function from nonseparable models (Chesher (2005)), marginal effects from nonlinear panel models (Chernozhukov, Fernandez-Val, Hahn, and Newey (2009)), and various treatment effects (e.g. Manski (1990), Manski and Pepper (2002), Battacharya, Shaikh and Vytlacil (2009)). Numerous other examples are found in the monographs of Manski (2004, 2007).

Providing a point decision when the parameter is interval-identified may sound odd. Indeed, when we fully know the bounds, choosing a particular value in the identified interval as a decision cannot be empirically motivated. However, when we do not know the bounds, and have to rely on their estimators that are subject to certain sampling variations, the question of how to obtain a reasonable point decision is still empirically relevant, because all the point decisions in the estimated interval may not be equally supported by the observations.

As far as the author is concerned, optimal inference under partial identification has not appeared in the literature except for a few researches. Canay (2009) considered inference under moment inequality models through generalized empirical likelihoods and established large deviation optimality. Andrews and Jia (2008) recently suggested a moment inequality test that has reasonable properties in terms of average asymptotic powers. This paper's approach is distinctive in many aspects. Most of all, this paper's main concern is to find a reasonable point decision when the object of interest is partially identified.

This paper takes candidate point decisions to be in the form of weighted averages of bound estimators that are potentially irregular and asymptotically biased, and considers a loss function that is a convex increasing function of the distance between the point decision and the interval-identified parameter. The allowance for irregular estimators is important in our context because most bound estimators obtained under the inequality constraint are irregular inherently.

Asymptotic theory with fixed positive interval length does not capture the finite sample situation in which the ratio of the interval length to the finite sample variance of its estimator is finite. Hence this paper considers asymptotic theory under *near-identification* where the length of the identified interval in each sequence of experiments shrinks to zero at the rate of $\sqrt{n}$. This approach of near-identification has precedents in Imbens and Manski (2004) and Stoye (2009a). A similar local parameter approach in the presence of the parameter on the boundary was employed by Andrews (1999).

To obtain a reasonable point decision, this paper first slices asymptotically normal experiments into subclasses according to the localized lengths of the identified interval. For each subclass, this paper establishes a local asymptotic minimax risk bound, and finds a local asymptotic minimax decision. This minimax decision is reduced to a semiparametrically efficient estimator of the parameter of interest when the parameter is known to be point-identified. Therefore, the decision can be viewed as a generalization of a semiparametric efficient estimator under point identification.

When the length of the interval is known, the local asymptotic minimax solution immediately offers a feasible solution. When this is not the case, the local asymptotic minimax solution depends on the index of the subclass experiment, rendering the solution impractical. In this situation, this paper proposes that we replace the subclass index by an appropriate shrinkage estimator. When the loss function is the absolute deviation loss, we prove that the resulting estimator is a hybrid decision between two extreme subclass minimax decisions. Remarkably, the hybrid decision remains intact even when one mistakes the upper bound for the lower bound. Hence knowledge of which of the bound parameters constitutes the upper or lower bound is only ancillary to the decision.

The main proposal of this paper is the following. When the interval length is small, one uses a semiparametrically efficient estimator of the object of interest that is computed under the assumption that the object is point-identified. However, when the interval length is large, it is optimal to take the average of the efficient upper and lower bound estimators as the point-decision.

The inference problem that arises when the parameter of interest is partially identified has long been noted and addressed in the literature. (e.g. Gilstein and Leamer (1983), Phillips (1989), and Hansen, Heaton, and Luttmer (1995), to name but a few.) One of the major approaches in such a situation is to shed the restrictions, and develop inference procedures for set-identified parameters or the identified set themselves. This approach has been most notably advanced by numerous researches of Charles Manski (see his monographs (1995, 2004), and references therein). Recent methodological developments include, among many researches, inference of interval-identified parameters (Horowitz and Manski (2000), Imbens and Manski (2004) and Stoye (2009a)) and parameters partially identified by moment inequalities (Rosen (2008), Andrews and Soares (2007), Andrews and Jia (2008), Pakes, Porter, Ho and Ishii (2006), Andrews and Guggenberger (2009a,b), Bugni (2007), and Fan and Park (2007)). A general inference procedure under partial identification has also been investigated by Chernozhukov, Hong, and Tamer (2007), Romano and Shaikh (2009), and Beresteanu and Molinari (2008). Recently Moon and Schorfheide (2009) contrasted Bayesian and frequentist approaches to inference under partial identification.

The remainder of this paper is organized as follows. In the next section, we introduce the inferential environment for the interval-identified parameter and discuss examples. The third section defines a decision-theoretic framework, introducing the decision space, loss functions, and asymptotically normal experiments. The fourth section proposes subclass local asymptotic minimax optimality and presents a method to construct optimal decisions. Then in the fifth section, the finite sample risk properties of this decision are illustrated through a Monte Carlo simulation study. The sixth section concludes. All the technical proofs are relegated to the appendix.

2 Point Decisions for an Interval-Identified Parameter

2.1 Decision Making for Interval-Identified Parameters

Suppose that we are given an *observable* random vector X from $P \in \mathcal{P}$, where $\mathcal{P}$ is a model, i.e., the collection of probabilities potentially serving as a data generating process for X. We are interested in knowing about a parameter $\theta_0 \in \mathbf{R}$ which is known to lie in an identified interval, $[\theta_L(P), \theta_U(P)]$, where $\theta_B(P) = (\theta_U(P), \theta_L(P))^\top$ is an $\mathbf{R}^2$-valued map on $\mathcal{P}$. Throughout this paper, $(\theta_U, \theta_L)^\top$ denotes a vector in $\mathbf{R}^2$, whereas $[\theta_L, \theta_U]$ denotes an interval in $\mathbf{R}$.

Interval-identification frequently arises in numerous contexts as mentioned in the introduction. We discuss three examples where the interest in a point decision of an interval-identified parameter naturally arises.

EXAMPLE 1: (OPTIMAL CONTINUOUS TREATMENT DECISION) Suppose that $\theta(t)$ denotes the treatment effect parameter with continuous treatment $t \in \mathbf{T}$. For example $\theta(t)$ is the expected earnings differential of a worker due to treatment t that is the duration of a person in the job training program. Or in medical studies, $\theta(t)$ indicates the expected reduction in health risks and t denotes the dosage of a new medicine. Suppose further that for each $t \in \mathbf{T}$,

$$\theta_L(t,P) \leq \theta(t) \leq \theta_U(t,P),$$

where $\theta_L(t,P)$ and $\theta_U(t,P)$ are identified upper and lower bounds. Then, the optimal treatment effect θ^* is identified as an interval $[\theta_L(P), \theta_U(P)]$, where $\theta_L(P) = \sup_{t\in\mathbf{T}} \theta_L(t,P)$ and $\theta_U(P) = \sup_{t\in\mathbf{T}} \theta_U(t,P)$. Suppose that an optimal policy rule for φ_0 which is, for example, the level of treatment or the entailed budget, etc. is given by $\varphi(\theta^*, P)$ where $\varphi(\cdot, P)$ is point-identified. Then, from the decision-maker's point of view, a reasonable point decision about θ^* can be useful. ■

EXAMPLE 2: (OPTIMAL RESERVE PRICE FROM AUCTION MODELS) Suppose that a decision-maker designs an English auction by setting an optimal reserve price with n bidders. Haile and Tamer (2003) considered a simple assumption that bidders neither bid more than their private valuations nor let an opponent win at a price that they would be willing to beat. Using this assumption, they derived bounds for the bidders' valuation distributions. Then, the optimal reserve price p^* is shown to lie in an interval $[p_L, p_U]$, where p_L and p_U are identified from the distributions of observed bids. In this situation, the auction designer may want to prefer a point decision about p^* to an interval decision even when it is interval-identified. ■

EXAMPLE 3: (ESTIMATION OF WILLINGNESS-TO-PAY) Many environmental or transportation analyses require accurate estimates of the willingness-to-pay of consumers who face different policy outcomes. In particular, McFadden (1999) explained how one can analyze willingness-to-pay from a random utility model. One example he considered involves two policy outcomes; the injury state and the restored state. Then the willingness-to-pay for the restoration policy is defined to be the amount of income that is to be taken from a consumer in the restored state to make her utility the same as that of a consumer in the injury state. Using the assumption that the indirect utility is linear in the unobserved heterogeneity, McFadden (1999) showed that the expected willingness to pay w is interval-identified in $[w_L, w_U]$, where w_L and w_U are upper and lower bounds involving the choice probabilities under the injury and the restored states. ■

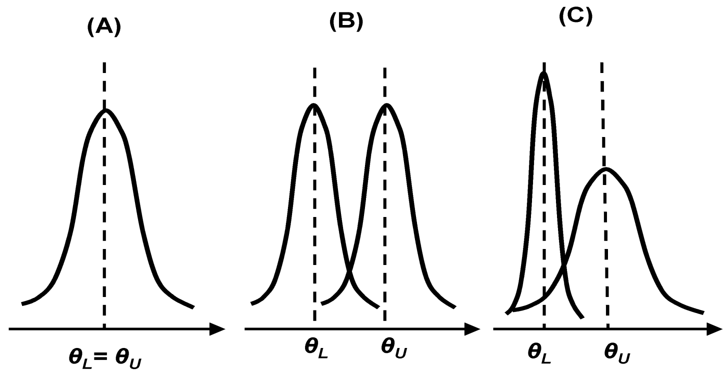

Figure 1: Illustrative Distributions of Bound Estimators

As mentioned in the introduction, the problem of making a point decision for an interval-identified parameter is an empirically relevant question in a finite-sample situation. We illustrate this point by a simple example. Suppose that we observe $X_L \sim N(\theta_L, \sigma_L^2)$ and $X_U \sim N(\theta_U, \sigma_U^2)$ and the parameter of interest θ_0 lies in the interval $[\theta_L, \theta_U]$. We assume that X_L and X_U are independent for simplicity. Figure 1 illustrates three possible cases. Certainly, when $\sigma_L^2 = \sigma_U^2$ and θ_0 is point-identified (as in (A) of Figure 1) so that $\theta_0 = \theta_L = \theta_U$, an efficient estimator of θ_0 is $(X_L + X_U)/2$. When $\theta_L < \theta_U$ and $\sigma_L^2 = \sigma_U^2$ as in (B), this paper demonstrates that a reasonable point decision continues to be $(X_L + X_U)/2$. Intuitively, this will cease to be the case when $\sigma_L^2 \neq \sigma_U^2$, or, say, $\sigma_L^2 < \sigma_U^2$ as in (C) of Figure 1. In this case, the informative content of X_U regarding θ_U is relatively small, making it intuitive to rely less on the estimate X_U of θ_U than on X_L. One might conjecture that an optimal decision in this situation will be a weighted average of X_L and X_U where the weights are given by $\sigma_U^2/(\sigma_L^2 + \sigma_U^2)$ for X_L and $\sigma_L^2/(\sigma_L^2 + \sigma_U^2)$ for X_U. Later, we will show that this is indeed the case only when the length of the interval is small. Even in this simple situation, the answer is less immediate when X_L and X_U are correlated.

2.2 Inequality Restrictions on Boundary Parameters

When a parameter is identified in an interval with bounds, say, $\theta_1(P)$ and $\theta_2(P)$, we often know which of the bounds is an upper bound and which a lower bound. When this is not the case, one might consider using $\theta_L(P) = \min\{\theta_1(P), \theta_2(P)\}$ and $\theta_U(P) = \max\{\theta_1(P), \theta_2(P)\}$. The main difficulty with this choice is, as pointed out by Hirano and Porter (2009) recently, that the bound parameters are not differentiable in P, preventing us from relying on the established framework of local asymptotic minimax estimation. Nevertheless, as long as one is interested in a point decision about θ_0, this paper's proposal does not require knowledge of the order of the bounds. This is demonstrated in the following way. First, the paper focuses on the situation where one knows the order of the bounds, and establishes a reasonable solution. Then it is shown that the decision that is derived remains invariant even when one interchanges the roles of the upper and the lower bounds.[2]

3 The Environment for Decision-Making

3.1 Boundary Parameters

We introduce a sequence of experiments that satisfy local asymptotic normality in the sense of Le Cam. For details, the reader is referred to Strasser (1985) or van der Vaart (1988). Let $\mathcal{B}$ be the Borel σ-field of $\mathbf{R}^2$ 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 $(\mathbf{R}^2, \mathcal{B})$ indexed by $h \in H$, so that $\mathcal{E}_n = (\mathbf{R}^2, \mathcal{B}, P_{n,h}; h \in H)$ constitutes a sequence of experiments. As for $\mathcal{E}_n$, we assume local asymptotic normality as follows.

Assumption 1: For each $h \in H$,

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

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

The notation $\rightsquigarrow$ denotes the weak convergence of measures. The local asymptotic normality of $\mathcal{E}_n$ was introduced by Le Cam (1960). The condition essentially reduces the decision problem to one in which an optimal decision is sought under a single Gaussian shift experiment $\mathcal{E} = (\mathbf{R}^2, \mathcal{B}, P_h; h \in H)$, 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 associated with the space of probability sequences $\{\{P_{n,h}\}_{n\geq 1} : h \in H\}$ (van der Vaart (1991).)

[2] This procedure is analogous to the usual argument used to show that a likelihood ratio test is a uniformly most powerful one-sided test of a normal mean. One first shows that the test is a most powerful test of simple null and alternative hypotheses, using the Neyman-Pearson Lemma. Then, it is shown that the test remains invariant when the alternative hypothesis is altered in one direction, proving that the test is uniformly most powerful against one-sided alternatives.

We introduce a sequence of bound parameters $\theta_B(P_{n,h}) = (\theta_L(P_{n,h}), \theta_U(P_{n,h}))^\top$ which are identified under $P_{n,h}$ for each $n \geq 1$ and each $h \in H$. The identification here means that the map $\theta_B(\cdot)$ is point-valued, and not set-valued for each $n \geq 1$. We can view the sequence as a sequence of $\mathbf{R}^2$-valued maps on H and write $\theta_{B,n}(h) = (\theta_{L,n}(h), \theta_{U,n}(h))^\top$. We let

$$\Delta_n(h) \equiv \sqrt{n}(\theta_{U,n}(h) - \theta_{L,n}(h)) \text{ and } \Delta_h \equiv \lim_{n\to\infty} \Delta_n(h)$$

and assume that Δ_h takes values in $\mathbf{R}$ for each $h \in H$. We also define

$$\Delta_0 \equiv \lim_{n\to\infty} \Delta_n(0) = \lim_{n\to\infty} \sqrt{n}\left\{\theta_{U,n}(0) - \theta_{L,n}(0)\right\}. \tag{1}$$

The asymptotic device here sends the interval length to zero at the rate of $\sqrt{n}$, so that the asymptotic theory maintains the finite sample feature that the ratio of the interval length to the variance of its estimator is finite in the limit.

The sequence of the boundary parameters $\theta_{B,n}(h)$ are assumed to be *regular* in the sense of van der Vaart (1991):[3]

Assumption 2: There exists a continuous linear $\mathbf{R}^2$-valued map on H, $\dot{\theta}_B = (\dot{\theta}_U, \dot{\theta}_U)^\top$, such that

$$\sqrt{n}(\theta_{U,n}(h) - \theta_{U,n}(0), \theta_{L,n}(h) - \theta_{L,n}(0)) \to (\dot{\theta}_U(h), \dot{\theta}_L(h))$$

as $n \to \infty$.

The continuous linear map $\dot{\theta}_B$ is associated with the semiparametric efficiency bound of the boundary parameter in the following way. Let $\dot{\theta}_B^* = (\dot{\theta}_U^*, \dot{\theta}_L^*)^\top \in \bar{H} \times \bar{H}$ be such that for each $b \in \mathbf{R}^2$and each $h \in H$, $b^\top \dot{\theta}_B(h) = \langle b^\top \dot{\theta}_B^*, h\rangle$. Then for any $b \in \mathbf{R}^2$, $||b^\top \dot{\theta}_B^*||^2$ represents the asymptotic variance bound of the parameter $b^\top \theta_{B,n}$ (without imposing the inequality restriction $\theta_{L,n}(h) \leq \theta_{U,n}(h)$) (e.g. van der Vaart and Wellner (1996), Section 3.11.) For example, suppose that $P_{n,h} = P_h^n$ for some $\{P_h : h \in H\}$ and $\langle h_1, h_2\rangle = \int h_1(x)h_2(x)dP_0$. Then, by the Riesz representation theorem, there exist $\dot{\theta}_U^* \in \bar{H}$ and $\dot{\theta}_L^* \in \bar{H}$ such that

$$\dot{\theta}_U(h) = \int \dot{\theta}_U^*(x)h(x)dP_0 \text{ and } \dot{\theta}_L(h) = \int \dot{\theta}_L^*(x)h(x)dP_0 \text{ for all } h \in H.$$

The maps $\dot{\theta}_U^*$ and $\dot{\theta}_L^*$ are called the efficient influence functions of $\theta_{U,n}$ and $\theta_{L,n}$ in the

[3] As pointed out by Hirano and Porter (2009), the bounds in Example 2 do not satisfy Assumption 2 when the number of bidders is different across the auctions, because the bounds there are nondifferentiable. However, when the number of bidders in the auctions in the sample are large, the bounds can be approximated by differentiable functionals, and this paper's framework still applies as an approximate benchmark solution. A general treatment of the problem accommodating nondifferentiable bounds deserves a separate research.

literature (e.g. van der Vaart (1991)). For future references, we define also

$$\sigma_U^2 \equiv \langle \dot{\theta}_U^*, \dot{\theta}_U^* \rangle, \ \sigma_L^2 \equiv \langle \dot{\theta}_L^*, \dot{\theta}_L^* \rangle, \text{ and } \sigma_{L,U} \equiv \langle \dot{\theta}_L^*, \dot{\theta}_U^* \rangle. \tag{2}$$

The inequality restriction $\theta_{L,n}(h) \leq \theta_{U,n}(h)$ suggests focusing on a proper subset of H. For all $n \geq 1$,

$$\begin{aligned} \sqrt{n}\left(\theta_{U,n}(h) - \theta_{U,n}(0)\right) &\geq \sqrt{n}\left(\theta_{L,n}(h) - \theta_{U,n}(0)\right) \\ &= \sqrt{n}\left(\theta_{L,n}(h) - \theta_{L,n}(0)\right) - \Delta_n(0). \end{aligned}$$

In the limit with $n \to \infty$, we have $\dot{\Delta}_\theta(h) \geq -\Delta_0$ where $\dot{\Delta}_\theta = \dot{\theta}_U - \dot{\theta}_L$ by Assumption 2. Hence the tangent set under the inequality restriction is given by

$$H_R = \left\{ h \in H : \dot{\Delta}_\theta(h) \geq -\Delta_0 \right\}. \tag{3}$$

It is important to note that the tangent set H_R is a convex affine cone, contains a convex cone, namely, $\{h \in H : \dot{\Delta}_\theta(h) \geq 0\}$, but is not a linear space. Suppose that θ_0 is point-identified and has an efficient influence function $\tilde{\kappa}$ associated with H_R in the sense of van der Vaart (1989, 1991). When $\tilde{\kappa}$ is contained in the linear span of $\{h \in H : \dot{\Delta}_\theta(h) \geq 0\}$, a best regular estimator which is a regular estimator with asymptotic distribution $N(0, ||\tilde{\kappa}||^2)$ (after scale and location normalization) is local asymptotic minimax by Theorem 2.4 of van der Vaart (1989). This approach does not apply in our case because due to partial identification, we do not have a well-defined influence function for θ_0.

3.2 Loss Functions and Decisions

The main interest lies in a decision that is as close to the object of interest θ_0 as possible. Let us introduce the decision space and loss functions $L(d, \theta_0)$.

Assumption 3: (i) The decision space D is given by $D = \mathbf{R}$.
(ii) For each decision $d \in D$,

$$L(d, \theta_0) = \psi\left(|d - \theta_0|\right), \tag{4}$$

where $\psi : [0, \infty) \to [0, \infty)$ is continuous, convex, and increasing and for each $M > 0$, $\min\{\psi(\cdot), M\}$ is uniformly continuous.

The decision space D is taken to be $\mathbf{R}$. Although this might appear innocuous at first, this excludes some important problems such as finding optimal discrete treatment rules (Manski (2004), Stoye (2009b), and Hirano and Porter (2008)). The loss function requires that ψ is continuous and convex which is stronger than the standard condition that ψ is lower semicontinuous. This additional condition is introduced to deal with partial-identification by using the so called identifiable maximal risk in Section 4.1.

A reasonable decision will utilize the fact that θ_0 lies in the identifiable interval. Candidate decisions $\hat{d}$ for θ_0 are assumed to be linked with boundary estimators in the following way. Given a boundary estimator $\hat{\theta}_B = (\hat{\theta}_U, \hat{\theta}_L)^\top \in \mathbf{R}^2$ of $\theta_B = (\theta_U, \theta_L)^\top$, and a (possibly stochastic) number $\hat{\tau}$, we define

$$T(\hat{\theta}_B; \hat{\tau}) \equiv \hat{\tau}\hat{\theta}_U + (1 - \hat{\tau})\hat{\theta}_L. \tag{5}$$

We need to specify the nature of the stochastic number $\hat{\tau}$. Let $\mathcal{T}_n$ be the collection of random variables $\hat{\tau}$ such that for each $h \in H$,

$$\hat{\tau} \to_P \tau \text{ along } \{P_{n,h}\}$$

and τ is a real number that does not depend on h. For example, the class $\mathcal{T}_n$ includes all the estimators $\hat{\tau}$ such that for each $h \in H$,

$$P_{n,h}\left\{|\hat{\tau} - \tau_n(h)| > \varepsilon\right\} \to 0,$$

where the sequence $\tau_n(h)$ satisfies that $\tau_n(0) \to \tau$ for some $\tau \in \mathbf{R}$ and $\sqrt{n}\{\tau_n(h) - \tau_n(0)\} \to \dot{\tau}(h)$ for a continuous linear real map $\dot{\tau} : H \to \mathbf{R}$. In the definition (5), we do not require that $\hat{\tau}$ take values in $[0, 1]$. Hence the combination $\hat{\tau}\hat{\theta}_U + (1 - \hat{\tau})\hat{\theta}_L$ is not necessarily a convex combination of $\hat{\theta}_U$ and $\hat{\theta}_L$. Moreover, the boundary estimator $\hat{\theta}_B$ (after location and scale normalization) is allowed to be asymptotically biased for θ_B.

Definition 1: Let $\mathcal{D}_n$ be the collection of random vectors $\hat{d} \in \mathbf{R}$ in the form $\hat{d} = T(\hat{\theta}_B; \hat{\tau})$, for some $\hat{\tau} \in \mathcal{T}_n$ and some random vector $\hat{\theta}_B \in \mathbf{R}^2$ such that for each $h \in H$, there exists a random vector $V^h = (V_U^h, V_L^h) \in \mathbf{R}^2$ satisfying that

$$\sqrt{n}(\hat{\theta}_B - \theta_{B,n}(h)) \to_d V^h, \text{ along } \{P_{n,h}\}_{n \geq 1},$$

and $\sup_{h \in H} \mathbf{E}L(|V_U^h| + |V_L^h|) < \infty$.

The class $\mathcal{D}_n$ includes as a special case $\hat{d} = T(\hat{\theta}_B; \hat{\tau})$ where $\sqrt{n}(\hat{\theta}_B - \theta_{B,n}(h))$ is asymptotically normal. More importantly, the distribution of V^h may not have mean zero and may vary with the shift h. Therefore we allow the boundary estimator $\hat{\theta}_B$ to be *irregular*. The irregularity of the boundary estimator $\hat{\theta}_B$ arises inherently when $\hat{\theta}_B$ is obtained under the constraint : $\hat{\theta}_L \leq \hat{\theta}_U$, for each $n \geq 1$. We illustrate two methods of obtaining a boundary estimator $\hat{\theta}_B$ satisfying the inequality constraint $\hat{\theta}_U \geq \hat{\theta}_L$ in finite samples. Let $\bar{\theta}_B = (\bar{\theta}_U, \bar{\theta}_L)^\top$ be an estimator of θ_B obtained without imposing the inequality constraint $\bar{\theta}_U \geq \bar{\theta}_L$. Since the main point of exposition here is to illuminate the nature of irregularity, we focus on the case where

$$\sqrt{n}(\bar{\theta}_B - \theta_{B,n}(h)) \to_d V$$

where the distribution of $V = (V_U, V_L)$ is identical for all $h \in H$.

Example 4 (Reflecting Transform) : Consider $\hat{\theta}_B = (\max(\bar{\theta}_U, \bar{\theta}_L), \min(\bar{\theta}_U, \bar{\theta}_L))^\top$. Then, along the sequence of $\{P_{n,h}\}_{n\geq 1}$,

$$\sqrt{n}\begin{bmatrix} \hat{\theta}_U - \theta_{U,n}(h) \\ \hat{\theta}_L - \theta_{L,n}(h) \end{bmatrix} \to_d \begin{bmatrix} \max\{V_U, V_L - \Delta_h\} \\ \min\{V_U + \Delta_h, V_L\} \end{bmatrix}.$$

The distribution is reflected on the line $\{(v_1, v_2) : v_1 - v_2 + \Delta_h = 0\}$, and hence absolutely continuous with respect to the Lebesgue measure. This estimator $\hat{\theta}_B$ satisfies the conditions in Definition 1. ■

Example 5 (Censoring Transform) : For any $\gamma \in \mathbf{R}$, define $\hat{\theta}_B \equiv (T_U^\gamma(\bar{\theta}_B), T_L^\gamma(\bar{\theta}_B))^\top$, where, letting $(a)_+ \equiv \max(a, 0)$ as usual,

$$\begin{aligned} T_U^\gamma(\bar{\theta}_B) &\equiv \gamma\bar{\theta}_U + (1-\gamma)\bar{\theta}_L + (1-\gamma)(\bar{\theta}_U - \bar{\theta}_L)_+ \text{ and} \\ T_L^\gamma(\bar{\theta}_B) &\equiv \gamma\bar{\theta}_U + (1-\gamma)\bar{\theta}_L - \gamma(\bar{\theta}_U - \bar{\theta}_L)_+. \end{aligned}$$

It is easy to see that along $\{P_{n,h}\}_{n\geq 1}$,

$$\sqrt{n}\begin{bmatrix} \hat{\theta}_U - \theta_{U,n}(h) \\ \hat{\theta}_L - \theta_{L,n}(h) \end{bmatrix} \to_d \begin{bmatrix} \tilde{V}_U \\ \tilde{V}_L \end{bmatrix},$$

where

$$\begin{aligned} \tilde{V}_U &= \gamma V_U + (1-\gamma)V_L + (1-\gamma)[(V_U - V_L) + \Delta_h]_+ - (1-\gamma)\Delta_h \text{ and} \\ \tilde{V}_L &= \gamma V_U + (1-\gamma)V_L - \gamma[\{V_U - V_L\} + \Delta_h]_+ + \gamma\Delta_h. \end{aligned}$$

Note that the distribution of $(\tilde{V}_U, \tilde{V}_L)^\top$ is not regular, and not absolutely continuous in general. The joint distribution of $(\tilde{V}_U, \tilde{V}_L)^\top$ is censored on the half space $\{(v_1, v_2) : v_1 - v_2 + \Delta_h < 0\}$. ■

4 Asymptotic Optimal Decisions

4.1 Subclass Experiments and Minimax Optimality

The standard approach of local asymptotic minimax estimation faces difficulty due to partial identification of θ_0. Since θ_0 is not point-identified, the risk is not identified either. This may create an inconvenient situation in which an optimal solution depends on θ_0. This paper circumvents this difficulty by introducing what we call the *identifiable maximal risk*:

$$\rho_h(\hat{d}) = \sup_{F \in (\theta_{B,n}(h))} \int \mathbf{E}_h \left[L(\sqrt{n}\{\hat{d} - \theta\}) \right] dF(\theta), \tag{6}$$

where $\mathcal{F}(\theta_{B,n}(h))$ is the collection of probability measures with support in $[\theta_{L,n}(h), \theta_{U,n}(h)]$. (Throughout the paper, the supremum of a nonnegative quantity over an empty set is set to be zero.) The identifiable maximal risk is the largest risk possible by any arbitrary randomization of the unidentified parameter θ_0. The support of the randomization is restricted to the identified interval $[\theta_{L,n}(h), \theta_{U,n}(h)]$. The identifiable maximal risk is point-identified, and taken in this paper as the criterion of comparison among different decisions.

One might consider alternatively the Bayesian approach, by considering instead

$$\rho_{h,G}(\hat{d}) = \int \mathbf{E}_h \left[L(\sqrt{n}\{\hat{d} - \theta\}) \right] dG(\theta)$$

where G is a probability measure that is absolutely continuous with respect to the Lebesgue measure on $\mathbf{R}$ and has a support in $[\theta_{L,n}(h), \theta_{U,n}(h)]$. This choice has the disadvantage that when the parameter θ_0 is point-identified so that $\theta_{L,n}(h) = \theta_{U,n}(h)$, the risk is reduced to zero. Therefore, this choice does not ascertain continuity from point-identification to interval-identification. To avoid this inconvenience, one might consider G such that its probability measure over the identified interval is normalized to one regardless of the length of the interval. In this case, the resulting risk may fail to be increasing in the interval-length. This is odd because naturally the decision should become riskier as the identified interval becomes longer. As we demonstrate later, the use of the identifiable maximal risk (6) does not suffer from these difficulties.

The adoption of the identifiable maximal risk does not resolve the matter entirely. Recall that the tangent set H_R is not a linear space. By deliberately focusing on convex subcones in H_R, one may find an estimator that is local asymptotic minimax over each convex subcone as in Example of van der Vaart (1989). This latter approach poses a two-fold difficulty in our situation. First, as mentioned by van der Vaart (1989), p.1492, it is still an unresolved matter how one obtains such an estimator in semiparametric models.[4] Furthermore, the choice of convex cones should be done in a way that avoids rendering the minimax problem trivial. Indeed, when $h \in H_R$ exists such that $\Delta_h = \infty$, the maximal risk becomes infinity regardless of the decision involved. The constraint arising from our need to avoid triviality makes it even harder to find a generic solution such as one that appeared in Example of van der Vaart (1989), p.1492. In the light of the difficulty in employing the standard minimax approach, this paper first slices the asymptotic normal experiments into subclasses according to the associated localized interval length Δ_h. Corresponding to each subclass, we find a local asymptotic minimax decision. Then, in the second stage, we replace the subclass index by a shrinkage type estimator. It is demonstrated that this type of decision has a hybrid character that performs well both when $r \to 0$ and when $r \to \infty$.

For each $r \in [0, \infty)$ and each small $\varepsilon > 0$, let $H_{n,R}(r;\varepsilon) \equiv \{h \in H : r - \varepsilon \leq \Delta_n(h) \leq r + \varepsilon\}$. The set $H_{n,R}(r;\varepsilon)$ approximates the hyperplane $H_R(r) \equiv \{h \in H : \Delta_h = r\}$. The focus is on the supremum of the identifiable maximal risk over $h \in H_{n,R}(r;\varepsilon)$ for each $r \in [0, \infty)$ and each $\varepsilon > 0$,

$$R_n^{\varepsilon}(\hat{d}; r) \equiv \sup_{h \in H_{n,R}(r;\varepsilon)} \rho_h(\hat{d}). \tag{7}$$

The theorem below establishes the local asymptotic minimax risk bound for each r. Let $Z = (Z_U, Z_L)^\top \in \mathbf{R}^2$ be a normal random vector such that

$$Z \sim N(0, \Sigma),$$

where, with $\sigma_U^2, \sigma_{L,U}$ and σ_L^2 defined in (2),

$$\Sigma \equiv \begin{bmatrix} \sigma_U^2 & \sigma_{L,U} \\ \sigma_{L,U} & \sigma_L^2 \end{bmatrix}.$$

It is worth noting that Σ^{-1} is the semiparametric efficiency bound for the boundary vector $\theta_{B,n}$ without imposing the inequality restriction $\theta_{L,n} \leq \theta_{U,n}$. Hence the matrix Σ can be found using the usual method of projection in the L_2 space (e.g. Begun, Hall, Huang, and Wellner (1983), Newey (1990) and Bickel, Klaassen, Ritov and Wellner (1993)).

[4] Due to this difficulty, Tripathi (2000) confined attention only to regular estimators when he investigated efficient estimation of a finite dimensional parameter in the presence of an infinite dimensional nuisance parameter under shape restrictions.

THEOREM 1: *Suppose that* Assumptions 1-3 *hold. Then for any* $\hat{d} \in \mathcal{D}_n$ *and* $r \in [0,\infty)$,

$$\lim_{\varepsilon \to 0} \liminf_{n \to \infty} R_n^\varepsilon(\hat{d}; r) \ \geq \begin{cases} \mathbf{E}\left[L\left(Z_L - \frac{r}{2}\right)|Z_U - Z_L = 0\right] \text{ if } \sigma_\Delta^2 > 0 \text{ or } r = \Delta_0 \\ \\ 0 \quad \text{ if } \sigma_\Delta^2 = 0 \text{ and } r \neq \Delta_0. \end{cases}.$$

The minimax risk bound increases in r. This reflects that deciding on an interval-identified parameter is riskier than that on a point-identified parameter. As the identified interval becomes longer, the lowest possible maximal risk also increases. When $\sigma_\Delta^2 = 0$ and $r \neq \Delta_0$, the slice $H_R(r)$ of the normal experiment becomes empty, and hence the bound is trivially zero.

A local asymptotic minimax decision is one that achieves the bound in Theorem 1. The verification of this achievement often requires to prove the uniform integrability condition for the sequence of decisions. To dispense with such a requirement, we follow the suggestion by Strasser (1985) (p.480) and consider instead

$$R_{n,M}^\varepsilon(\hat{d}; r) \equiv \sup_{h \in H_{n,R}(r;\varepsilon)} \rho_{h,M}(\hat{d}), \tag{8}$$

where $L_M(\cdot) \equiv \min\{L(\cdot), M\}$, $M > 0$, and

$$\rho_{h,M}(\hat{d}) \equiv \sup_{F \in (\theta_{B,n}(h))} \int \mathbf{E}_h\left[L_M(\sqrt{n}\{\hat{d} - \theta\})\right] dF(\theta).$$

Then we adopt the definition of local asymptotic minimaxity as follows.

DEFINITION 2: (i) A decision $\tilde{d}(r)$ is said to be *subclass local asymptotic minimax* (*S-LAM*) *at* $r \in [0,\infty)$, if for all $\hat{d} \in \mathcal{D}_n$,

$$\lim_{M \uparrow \infty} \lim_{\varepsilon \to 0} \limsup_{n \to \infty} R_{n,M}^\varepsilon(\tilde{d}(r); r) \ \leq \ \mathbf{E}\left[L\left(Z_L - \frac{r}{2}\right)|Z_U - Z_L = 0\right]. \tag{9}$$

(ii) When $\tilde{d}(r)$ does not depend on r and the above inequality holds for all $r \in [0,\infty)$, the decision is said to be *uniformly S-LAM.*

Since the quantity $R_n^\varepsilon(\hat{d}; r)$ involves supremum over $H_{n,R}(r;\varepsilon)$, the lower bound in Theorem 1 holds locally uniformly over r. In other words, for any $r \in \mathbf{R}$ and any $r_n \to r$, the lower bound in Theorem 1 remains invariant when we replace $R_n^\varepsilon(\hat{d}; r)$ in the theorem by $R_n^\varepsilon(\hat{d}; r_n)$. Similarly, when a decision $\tilde{d}(r)$ is S-LAM in the sense of Definition 2, the decision

$\tilde{d}(r)$ is also locally uniformly S-LAM at r, i.e., for all $r_n \to r$,

$$\lim_{M \uparrow \infty} \lim_{\varepsilon \to 0} \limsup_{n \to \infty} R^{\varepsilon}_{n,M}(\tilde{d}(r); r_n) \ \leq \mathbf{E}\left[L\left(Z_L - \frac{r}{2} \right) | Z_U - Z_L = 0 \right]. \tag{10}$$

Theorem 1 shows that the minimax risk bound involves the conditional distribution of $Z_L - r/2$ given $Z_U - Z_L = 0$. This form guides our search for a S-LAM decision. Let $\tilde{\theta}_B = (\tilde{\theta}_U, \tilde{\theta}_L)^\top$ be a semiparametrically efficient estimator of θ_B such that

$$\sqrt{n}\left(\tilde{\theta}_B - \theta_{B,n}(h) \right) \to_d Z. \tag{11}$$

This estimator $\tilde{\theta}_B$ is a *best regular* estimator of $\theta_{B,n}(h)$ (with respect to the tangent set H) in the sense of van der Vaart (1988). Hence $\tilde{\theta}_B$ is an estimator of $\theta_{B,n}(h)$ obtained without imposing the inequality constraint $\theta_L \leq \theta_U$.

Let $\sigma^2_\Delta \equiv \sigma^2_U - 2\sigma_{L,U} + \sigma^2_L$, and $\sigma_{L,\Delta} \equiv \sigma_{L,U} - \sigma^2_L$. Let τ^* be such that

$$\tau^* \equiv \begin{cases} -\sigma_{L,\Delta}/\sigma^2_\Delta \text{ if } \sigma^2_\Delta > 0 \\ 0 \text{ if } \sigma^2_\Delta = 0, \end{cases}$$

and let $\hat{\tau}^*$ be a consistent estimator of τ^*. We require the convergence in (11) to be uniform in $h \in H$:

Assumption 4: (i) $\sup_{h \in H} |P_{n,h}\{\sqrt{n}(\tilde{\theta}_B - \theta_{B,n}(h)) \leq t\} - P\{Z \leq t\}| \to 0$ for each $t \in \mathbf{R}$. (ii) For all $\varepsilon > 0$, $\sup_{h \in H} P_{n,h}\{|\hat{\tau}^* - \tau^*| > \varepsilon\} \to 0$.

The uniform convergence of distributions and the uniform consistency can often be verified using the uniform central limit theorem (Giné and Zinn (1991)). We consider the following decision:

$$\tilde{d}(r) \equiv \hat{\tau}^* \tilde{\theta}_U + (1 - \hat{\tau}^*)\tilde{\theta}_L + \left(\frac{1}{2} - \hat{\tau}^* \right) \frac{r}{\sqrt{n}}. \tag{12}$$

This random decision $\tilde{d}(r)$ is shown to be S-LAM at r in the following theorem.

Theorem 2: *Suppose that* Assumptions 1-4 *are satisfied. Then the following holds.*
(i) *For each* $r \in [0, \infty)$, $\tilde{d}(r)$ *is S-LAM at* r.
(ii) *When* $\tau^* = 1/2$, $\tilde{d}_{1/2} \equiv (\tilde{\theta}_U + \tilde{\theta}_L)/2$ *is uniformly S-LAM.*

The result of Theorem 2 immediately applies to the case where the parameter θ_0 is known to be point-identified. When the parameter is point-identified so that $r = 0$, the S-LAM decision is reduced to $\tilde{d}(0)$. One can check that $\tilde{d}(0)$ is a semiparametric efficient estimator of θ_0 under the restriction that $\theta_L(P) = \theta_U(P)$.[5] Hence the estimator in (12) can be viewed as a generalization of an efficient estimator of θ_0 to a situation where θ_0 is interval-identified.

Result (ii) says that when $\tau^* = 1/2$, $\tilde{d}_{1/2}$ is uniformly S-LAM. For example, when the semiparametric efficiency bounds for θ_U and θ_L are identical, it follows that $\tau^* = 1/2$. The estimator $\tilde{d}_{1/2}$ does not depend on r and it is S-LAM uniformly over $r \in [0, \infty)$.

When $\sigma_{L,U} = 0$, we have $\tau^* = \sigma_L^2/(\sigma_U^2 + \sigma_L^2)$ so that we take $\hat{\tau}^* = \hat{\sigma}_L^2/(\hat{\sigma}_U^2 + \hat{\sigma}_L^2)$, where $\hat{\sigma}_L^2$ and $\hat{\sigma}_U^2$ are consistent estimators of σ_L^2 and σ_U^2. The S-LAM decision becomes

$$\tilde{d}(r) = \tilde{d}(0) + \frac{\hat{\sigma}_U^2 - \hat{\sigma}_L^2}{2(\hat{\sigma}_U^2 + \hat{\sigma}_L^2)} \frac{r}{\sqrt{n}}, \text{ where } \tilde{d}(0) = \frac{\tilde{\theta}_U \hat{\sigma}_L^2}{\hat{\sigma}_U^2 + \hat{\sigma}_L^2} + \frac{\tilde{\theta}_L \hat{\sigma}_U^2}{\hat{\sigma}_U^2 + \hat{\sigma}_L^2}.$$

Therefore, when $\sigma_L^2 < \sigma_U^2$, the first component $\tilde{d}(0)$ of the S-LAM decision leans toward $\tilde{\theta}_L$ and is close to the S-LAM decision when r is small. However, this leaning effect is offset by the presence of the term $r(\hat{\sigma}_U^2 - \hat{\sigma}_L^2)/(2\sqrt{n}(\hat{\sigma}_U^2 + \hat{\sigma}_L^2))$. Hence when the length of the identified interval is large, the significance of considering the variance components σ_U and σ_L is reduced.

Certainly, $\tilde{d}(0)$ is a reasonable decision when r is close to zero, as it is S-LAM when $r = 0$. The following result shows that under the absolute deviation loss, the decision $\tilde{d}_{1/2}$ is approximately S-LAM for large r.

THEOREM 3: *Suppose that* Assumptions 1-4 *hold. Furthermore assume that* $L(x) = |x|$. *Then for any* $r_n \to \infty$,

$$\lim_{M \uparrow \infty} \lim_{\varepsilon \to 0} \limsup_{n \to \infty} \left\{ R_{n,M}^{\varepsilon}(\tilde{d}_{1/2}; r_n) - \mathbf{E}\left[L\left(Z_L - \frac{r_n}{2}\right) | Z_U - Z_L = 0 \right] \right\} = 0.$$

The above result applies only to the absolute deviation loss. For a general loss function L, the result depends on the limit behavior of $L(x+y) - L(x)$ as $x \to \infty$.

When the researcher does not have reasonable values of r in mind, S-LAM decisions are not of much practical value. This paper considers

$$\hat{r} \equiv \sqrt{n}(\tilde{\theta}_U - \tilde{\theta}_L) 1\{|\tilde{\theta}_U - \tilde{\theta}_L| > b_n\}, \tag{13}$$

where $b_n \to 0$ is a sequence such that $\sqrt{n} b_n \to \infty$. Such a shrinkage estimator $\hat{r}$ was used by Stoye (2009a) to construct a uniformly valid confidence interval for an interval-identified

[5] In this case, the tangent space is given by $H_0 = \{h \in \bar{H} : \langle h, \dot{\Delta}_\theta \rangle = 0\}$. (e.g. Example 3.2.3 of Bickel, Klaassen, Ritov, and Wellner (1993).) The efficient influence function of θ_0 is given by $\dot{\theta}_L^* - \langle \dot{\theta}_L^*, \dot{\Delta}_\theta^* \rangle \langle \dot{\Delta}_\theta^*, \dot{\Delta}_\theta^* \rangle^{-1} \dot{\Delta}_\theta^*$ (or equivalently $\dot{\theta}_U^* - \langle \dot{\theta}_U^*, \dot{\Delta}_\theta^* \rangle \langle \dot{\Delta}_\theta^*, \dot{\Delta}_\theta^* \rangle^{-1} \dot{\Delta}_\theta^*$) and the variance bound is given by $\sigma_L^2 - \sigma_{L,\Delta}\sigma_\Delta^{-2}\sigma_{\Delta,L}$ (or equivalently $\sigma_U^2 - \sigma_{U,\Delta}\sigma_\Delta^{-2}\sigma_{\Delta,U}$). This variance bound is achieved by $\tilde{d}(0)$.

parameter. Then this paper proposes the following estimator as a reasonable point-decision:

$$\tilde{d}(\hat{r}) = \hat{\tau}^* \tilde{\theta}_U + (1 - \hat{\tau}^*) \tilde{\theta}_L + \left(\frac{1}{2} - \hat{\tau}^* \right) \frac{\hat{r}}{\sqrt{n}}. \tag{14}$$

It is interesting to observe that $\tilde{d}(\hat{r})$ is a hybrid of the two decisions $\tilde{d}_{1/2}$ and $\tilde{d}(0)$. To see this, note that when $|\tilde{\theta}_U - \tilde{\theta}_L| > b_n$,

$$\tilde{d}(\hat{r}) = \tilde{d}_{1/2} = (\tilde{\theta}_U + \tilde{\theta}_L)/2.$$

This means that when the length of the interval is large, we take an average of efficient estimators of their upper and lower bounds. On the other hand, when $|\tilde{\theta}_U - \tilde{\theta}_L| \leq b_n$,

$$\tilde{d}(\hat{r}) = \tilde{d}(0) = \hat{\tau}^* \tilde{\theta}_U + (1 - \hat{\tau}^*) \tilde{\theta}_L.$$

When the length of the interval is small, we take as the decision the efficient estimator $\tilde{d}(0)$ of θ_0 under point-identification. When the loss is the absolute deviation loss, the use of $\tilde{d}(\hat{r})$ is reasonable because it approximates a S-LAM decision both when $r \to 0$ and $r \to \infty$.

Corollary 1: *Suppose that* Assumptions 1-4 *hold. Then the following holds.*
(i) *For any* $r_n \to 0$,

$$\lim_{M \uparrow \infty} \lim_{\varepsilon \to 0} \limsup_{n \to \infty} \left\{ R^{\varepsilon}_{n,M}(\tilde{d}(\hat{r}); r_n) - \mathbf{E}\left[L\left(Z_L - \frac{r_n}{2} \right) | Z_U - Z_L = 0 \right] \right\} = 0.$$

(ii) *Furthermore assume that* $L(x) = |x|$. *Then for any* $r_n \to \infty$ *such that* $r_n/(\sqrt{n} b_n) \to \infty$,

$$\lim_{M \uparrow \infty} \lim_{\varepsilon \to 0} \limsup_{n \to \infty} \left\{ R^{\varepsilon}_{n,M}(\tilde{d}(\hat{r}); r_n) - \mathbf{E}\left[L\left(Z_L - \frac{r_n}{2} \right) | Z_U - Z_L = 0 \right] \right\} = 0.$$

A remarkable aspect of the solution $\tilde{d}(\hat{r})$ is that the solution remains numerically invariant even if one mistakenly believes that $\theta_U \leq \theta_L$ and $\theta_0 \in [\theta_U, \theta_L]$ and implements the solution accordingly. Indeed, the decisions $\tilde{d}(0)$ and $\tilde{d}_{1/2}$ and indicator $1\{|\tilde{\theta}_U - \tilde{\theta}_L| > b_n\}$ remain unchanged even when one mistakes θ_U for θ_L. Therefore, the optimality property of the solution $\tilde{d}(\hat{r})$ remains intact regardless of whether one knows the order between θ_L and θ_U or not.

5 Simulations

The decision $\tilde{d}(\hat{r})$ in this paper is justified based on the asymptotic theory under near-identification. Hence it is important to investigate whether the concept of asymptotic optimality approximates its finite sample counterpart reasonably well. In this section, we present and discuss some simulation results that illuminate the theoretical findings of the paper.

The data generating proccess was as follows. Suppose that the econometrician observes the i.i.d. data set $\{(X_{L,i}, X_{U,i})\}_{i=1}^n$ with unknown means $\mathbf{E}X_{L,i} = \theta_L$ and $\mathbf{E}X_{U,i} = \theta_U$. The parameter of interest θ_0 is known to lie in $[\theta_L, \theta_U]$. In the simulation study, we generated $X_{L,i}$ and $X_{U,i}$ as follows:

$$\begin{aligned} X_{L,i} &= 4a_L \times (wY_{L,i} + (1-w)Z_i)/2 + \theta_L \text{ and} \\ X_{L,i} &= 4a_U \times (wY_{U,i} + s \times (1-w)Z_i)/2 + \theta_U, \end{aligned}$$

where $Y_{L,i} \sim N(0,1)$ and $Y_{U,i} \sim N(0,1)$. The scale parameters a_L and a_U determine the scale of observations and were set to be $(a_L, a_U) = (3,1)$. The parameters w and s are employed to vary the degree and the sign of correlation between X_L and X_U. When $s = -1$, both random variables are negatively correlated. The mean vectors θ_L and θ_U were chosen to be $-b/2$ and $b/2$, where b denotes the interval length ranging from 0 to 0.5.

The main object of interest is the finite sample identifiable maximal risk for four types of decisions: hybrid decision $\tilde{d}(\hat{r})$, $\tilde{d}_{1/2}$, $\tilde{d}(0)$, and S-LAM decision $\tilde{d}(r)$. As for $\tilde{d}(\hat{r})$, we considered

$$\hat{r} = \sqrt{n}(\tilde{\theta}_U - \tilde{\theta}_L)1\{|\tilde{\theta}_U - \tilde{\theta}_L|/\hat{\sigma}_\Delta > c/(n^{1/3}\log(n))\},$$

where $\hat{\sigma}_\Delta$ denotes the standard deviation of $\{X_{U,i} - X_{L,i}\}_{i=1}^n$ and c is chosen from $\{2,4\}$. The estimators $\tilde{\theta}_U$ and $\tilde{\theta}_L$ are taken to be the sample means of $\{X_{L,i}\}_{i=1}^n$ and $\{X_{U,i}\}_{i=1}^n$. The estimated weight $\hat{\tau}^*$ was taken to be

$$\hat{\tau}^* = -\frac{\text{the sample covariance of } \{X_{L,i}\}_{i=1}^n \text{ and } \{X_{U,i} - X_{L,i}\}_{i=1}^n}{\text{the sample variance of } \{X_{U,i} - X_{L,i}\}_{i=1}^n}.$$

For computational simplicity, the finite sample identifiable maximal risk was computed to be the simulated quantity of

$$\sup_{p \in [0,1]} p\mathbf{E}\left[|\hat{d} - \theta_U|\right] + (1-p)\mathbf{E}\left[|\hat{d} - \theta_L|\right]. \tag{15}$$

When the finite sample density of $\hat{d}$ is quasiconcave, the above quantity coincides with the identifiable maximal risk defined in (6). When this is not the case, the quantity in (6) can

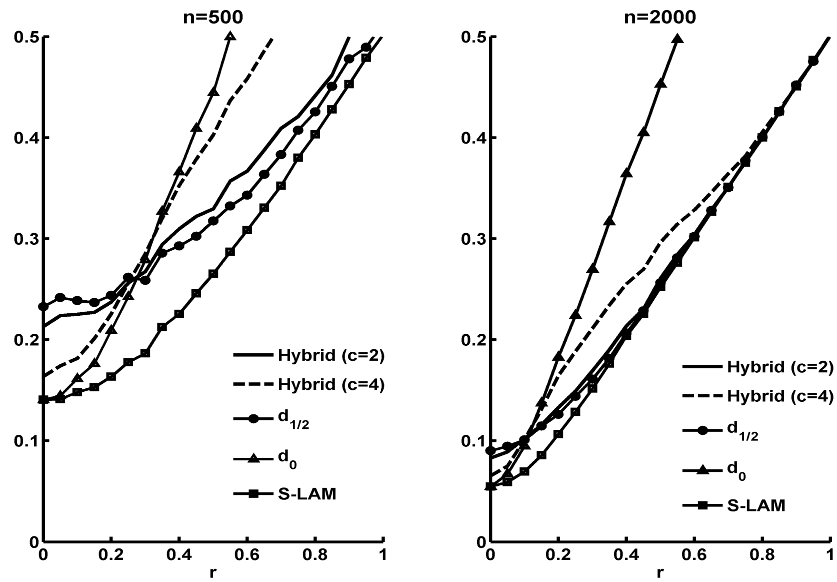

Figure 2: Finite Sample Identifiable Maximal Risks for Various Decisions

be viewed as approximating (6). The Monte Carlo simulation number was 2000.

As shown in Figure 3, the identifiable maximal risk of the S-LAM decision envelopes those of the other decisions from below. Hence the S-LAM decision shows its ideal character in finite samples. The decision $\tilde{d}(0)$ (denoted by 'd_0' in the figure) performs as well as the S-LAM, when the length of the interval is close to zero, but performs poorly when the length of the interval is large. On the other hand, the decision $\tilde{d}_{1/2}$ (denoted by '$d_{1/2}$' in the figure) performs well when the length of the interval is large, but performs poorly relative to $\tilde{d}(0)$ when the interval length is close to zero. These results show the complementary nature of the two decisions $\tilde{d}(0)$ and $\tilde{d}_{1/2}$. When the length of the interval is close to zero, the decision $\tilde{d}(\hat{r})$ mitigates the relatively poor performance of the decision $\tilde{d}_{1/2}$ as it tends to identify with $\hat{d}(0)$. On the other hand, when the length of the interval is large, $\tilde{d}(\hat{r})$ moves toward $\hat{d}_{1/2}$, retaining its good performance in this situation. When the sample size is large, $\tilde{d}(\hat{r})$ performs better than $\hat{d}(0)$ for large r and performs better than $\hat{d}_{1/2}$ when r is close to zero. The dominance of $\tilde{d}(\hat{r})$ over $\hat{d}_{1/2}$ under r close to zero entails cost: the finite sample risk is higher than that of $\hat{d}_{1/2}$ in the middle ranges of r.

The performance of the decisions varies depending on the correlation between X_L and X_U. This fact is illustrated in Figure 3 where the results from three different correlations between X_L and X_U are considered. We generated the observations with $(w, s) = (0.2, 1)$, $(w, s) = (0.8, 1)$ and $(w, s) = (0.2, -1)$, resulting in the average sample correlation coefficients between X_L and X_U around 0.59, 0.01, and -0.59 respectively. It is interesting to note that

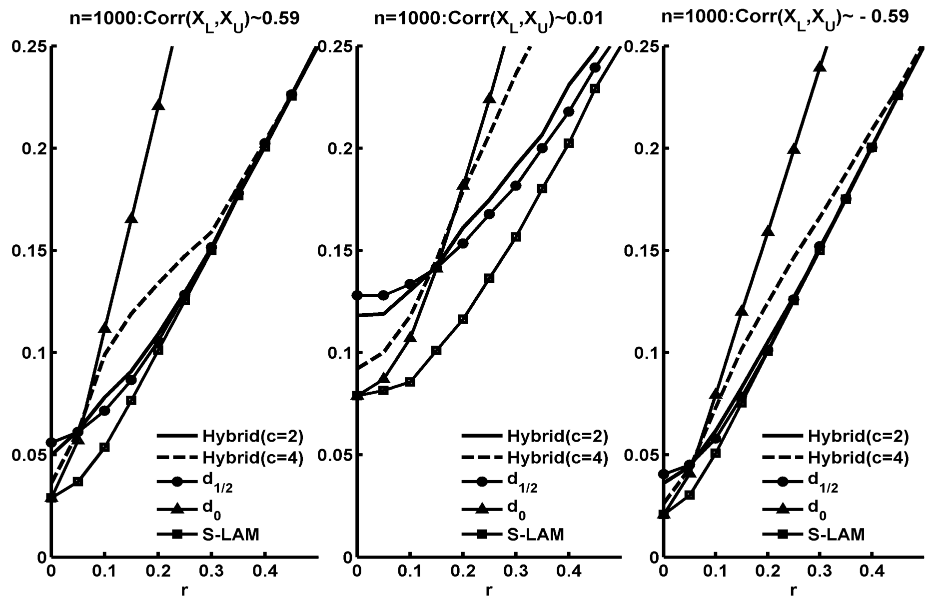

Figure 3: Finite Sample Risks over Various Correlations between X_L and X_U

when the observations X_L and X_U are close to be uncorrelated, the discrepancy between the finite sample risk of S-LAM estimator and the other decisions becomes large. However, when X_L and X_U are positively or negatively correlated to some degree, all the decisions of $\tilde{d}(\hat{r})$ and $\tilde{d}_{1/2}$ and $\tilde{d}(0)$ become closer to S-LAM. It is interesting to observe that the finite sample risk in the case of negative correlation is slightly lower than that in the case of positive correlation. This contrast becomes starker when we set the correlation coefficients to closer to 1 and -1 respectively. This is anticipated because the lower bound in Theorem 1, $\mathbf{E}\left[L\left(Z_L - r/2\right)|Z_U - Z_L = 0\right]$ becomes lower when Z_U and Z_L are negatively correlated than when they are positively correlated. This phenomenon arises intuitively because the conditioning on $Z_U = Z_L$ becomes more binding when Z_U and Z_L are negatively correlated than when they are positively correlated, resulting in a lower conditional variance of Z_L given $Z_U = Z_L$.

6 Conclusion

This paper has developed a notion of an optimal decision for an interval-identified parameter by considering increasing subclasses of experiments. This paper finds that a subclass local asymptotic minimax decision is given by an estimator that involves a weighted average of semiparametrically efficient bound estimators. The development of this paper suggests various ramifications. First, an investigation into ways to construct confidence sets for the interval-identified parameter using the local asymptotic minimax decision will be useful. Second, an extension to a situation with an identified box with many sides can be of practical interest. One interesting question in this situation is what class of functionals of the local asymptotic minimax decisions for the box-identified parameter will inherit the local asymptotic minimaxity of the latter. When this class of functionals is usefully characterized, one can easily generate local asymptotic minimax decisions for these functionals of the box-identified parameter from the local asymptotic minimax decision for this parameter.

7 Proofs

7.1 Proofs of the Auxiliary Results

The first part of the following lemma is Theorem 3.1 in Fukuda and Uno (2007) which results from Brunn's Theorem on convex bodies. (A convex body is a convex compact set with nonempty interior.) In the following lemma, we define *vol* to be the Lebesgue measure on $\mathbf{R}^d$. We say that a set $A \subset \mathbf{R}^d$ is *centrally symmetric* if $A = -A$.

Lemma A1: (i) *For any convex bodies* A *and* B *in* $\mathbf{R}^d$, $vol(A \cap (B+a))^{1/d}$ *is concave in* $a \in \{a : A \cap (B+a) \neq \varnothing\}$.
(ii) *Suppose further that* A *and* B *are centrally symmetric convex bodies in* $\mathbf{R}^d$, *and that for some* $a \in \mathbf{R}^d$ *and* $c \in \mathbf{R}\backslash\{0\}$, $vol(A \cap (B+a)) \leq vol(A \cap (B+ca))$. *Then, for all centrally convex* $C \in \mathbf{R}^d$, $vol(C \cap (B+a)) \leq vol(C \cap (B+ca))$.

Proof: (ii) For $a = 0$, the inequality becomes trivially an equality. For the remainder of the proof, we assume that $a \in \mathbf{R}\backslash\{0\}$. For $c \in \mathbf{R}$, let $H(c) = \{x \in \mathbf{R}^{d+1} : x_{d+1} = c\}$. Then, for any convex A, $f(c; A) \equiv vol(A \cap H(c))$ is quasiconcave in c on its support by Brunn's Theorem. (e.g. Theorem 5.1 of Ball (1997).) Furthermore, for any centrally symmetric convex set A and $c \in \mathbf{R}$, $f(c; A) \leq f(0; A)$ by Lemma 38.20 of Strasser (1985) and $f(c; A) = f(-c; A)$. Therefore, for any convex and centrally symmetric A, $f(c_1; A) \leq f(c_2; A)$ if and only if $|c_1| \geq |c_2|$. Since the latter inequality does not involve A, this statement implies that

$$
\begin{aligned}
f(c_1; A) &\leq f(c_2; A) \text{ for some centrally symmetric convex } A \text{ if and only if} \qquad (16)\\
f(c_1; C) &\leq f(c_2; C) \text{ for all centrally symmetric convex } C.
\end{aligned}
$$

We let

$$\begin{aligned}
\bar{A} &= \{(x_1, \cdots, x_d, w) : (x_1, \cdots, x_d) \in A \text{ and } w \in [-\max\{|c|, 1\}, \max\{|c|, 1\}]\}, \text{ and} \\
\bar{B} &= \{(x_1, \cdots, x_d, w) : (x_1, \cdots, x_d) \in B + wa \text{ and } w \in [-\max\{|c|, 1\}, \max\{|c|, 1\}]\}.
\end{aligned}$$

Then, observe that $\bar{P} = \bar{A} \cap \bar{B}$ is a convex body and centrally symmetric. The assumption of the lemma is tantamount to $f(1; \bar{P}) \leq f(c; \bar{P})$. The wanted result immediately follows from (16). ■

LEMMA A2: *Suppose that* $\{\varphi_t : t \in \mathbf{T}\}$, $\mathbf{T} \subset \mathbf{R}$, *is a class of quasiconcave functions such that for any* $t, t' \in \mathbf{T}$ *and* x_1, x_2, $\varphi_t(x_1) \leq \varphi_t(x_2)$ *if and only if* $\varphi_{t'}(x_1) \leq \varphi_{t'}(x_2)$. *Then, for any measure* μ *on the Borel* σ*-field of* $\mathbf{T}$, *the function* g *defined by* $g(\cdot) \equiv \int \varphi_t(\cdot) d\mu(t)$ *is quasiconcave.*

PROOF: For any $\alpha \in [0, 1]$, $g(\alpha x_1 + (1 - \alpha) x_2) \geq \int \min\{\varphi_t(x_1), \varphi_t(x_2)\} d\mu(t)$. Without loss of generality, suppose that $\varphi_t(x_1) \leq \varphi_t(x_2)$. Then, $\varphi_t(x_1) \leq \varphi_t(x_2)$ for all $t \in \mathbf{T}$. Hence

$$\int \min\{\varphi_t(x_1), \varphi_t(x_2)\} d\mu(t) = \min\left\{\int \varphi_t(x_1) d\mu(t), \int \varphi_t(x_2) d\mu(t)\right\} = \min\{g(x_1), g(x_2)\},$$

yielding the wanted result. ■

LEMMA A3: *Let* $V \in \mathbf{R}$ *be any continuous random variable with a quasiconcave density function that is symmetric around* b. *Then, for any quasiconvex* $L : [0, \infty) \to [0, \infty)$ *that is symmetric around zero,* $\mathbf{E}[L(V + \delta)]$ *is quasiconvex in* δ.

PROOF: Write $\mathbf{E}[L(V + \delta)] = \int_0^\infty P\{L(V + \delta) > t\} dt = \int_0^\infty P\{V + \delta \in \mathbf{R} \backslash A(t)\} dt$, where $A(t) \equiv \{z \in \mathbf{R} : L(z) \leq t\}$. Observe that

$$P\{V + \delta \in A(t)\} = \int_0^\infty vol(A(t) \cap (\{z : f(z) > e\} + b + \delta)) de,$$

where f is the joint density of $V - b$. Let

$$\eta(\delta; t, e) \equiv vol(A(t) \cap (\{z : f(z) > e\} + b + \delta)).$$

Since $\{z : f(z) > e\}$ is convex, by Lemma A1(i), η is quasiconcave in δ for all e and t. Since $A(t)$ and $\{z : f(z) > e\}$ are centrally symmetric and convex for all t and e, we apply Lemma A1(ii) to find that the order of $\eta(\delta; t, e)$ in δ is preserved through any shifts in e or t. Therefore, by Lemma A2, $\mathbf{E}[L(V + \delta)]$ is quasiconvex in δ. ■

Lemma A4: *Let $V \in \mathbf{R}$ be any continuous random variable with a quasiconcave density function that is symmetric around b and for a given $s \in [0, \infty)$, let $\mathcal{G}(s)$ be the collection of all the probability measures with a support in the interval $[0, s]$. Then, for any $L : \mathbf{R} \to \mathbf{R}$ that is convex and symmetric around zero,*

$$\sup_{G \in \mathcal{G}(s)} \inf_{c \in \mathbf{R}} \int \mathbf{E}\left[L(V + c + \delta)\right] dG(\delta) \geq \mathbf{E}\left[L\left(V - b - \frac{s}{2}\right)\right].$$

Proof: Since L is quasiconvex, we deduce that by Lemma A3, for any $G \in \mathcal{G}(s)$,

$$\inf_{c \in \mathbf{R}} \int \mathbf{E}\left[L(V + c + \delta)\right] dG(\delta) \leq \sup_{p \in [0,1]} \inf_{c \in \mathbf{R}} J(c, p),$$

where $J(c, p) \equiv \mathbf{E}\left[L(V + c + s)\right] p + \mathbf{E}\left[L(V + c)\right](1 - p)$. Since $\mathcal{G}(s)$ includes two-point distributions with support in $\{0, s\}$, we conclude that

$$\sup_{G \in \mathcal{G}(s)} \int \mathbf{E}\left[L(V + c + \delta)\right] dG(\delta) = \sup_{p \in [0,1]} J(c, p)$$

Hence for the statement of the lemma, it suffices to show that

$$J(c^*, p^*) \leq J(c, p^*), \text{ for all } c \in \mathbf{R}, \tag{17}$$

where $c^* \equiv -s/2 - b$ and $p^* \equiv 1/2$, because $J(c^*, p^*) \leq \inf_{c \in \mathbf{R}} J(c, p^*) \leq \sup_{p \in [0,1]} \inf_{c \in \mathbf{R}} J(c, p)$.

To show (17), let $\tilde{L}(d) = \{L(d + s) + L(d)\}/2$. Then, $\tilde{L}(d)$ is symmetric around $-s/2$. Since L is convex, so is $\tilde{L}(d)$. Observe that

$$\frac{1}{2}\left\{\mathbf{E}\left[L(V + c + s)\right] + \mathbf{E}\left[L(V + c)\right]\right\} = \int_0^\infty P\left\{\tilde{L}(V + c) > t\right\} dt = \int_0^\infty P\left\{V + c \in \mathbf{R} \backslash A(t)\right\} dt,$$

where $\tilde{A}(t) = \{z \in \mathbf{R} : \tilde{L}(z) \leq t\}$. Note that $\tilde{A}(t)$ is convex and symmetric around $-s/2$. Observe that

$$\begin{aligned} P\left\{V + \delta \in \tilde{A}(t)\right\} &= \int_0^\infty vol(\tilde{A}(t) \cap (\{z : f(z) > e\} + b + c))de \\ &= \int_0^\infty vol((\tilde{A}(t) + s/2) \cap (\{z : f(z) > e\} + b + c + s/2))de, \end{aligned}$$

where f is the joint density of $V - b$. By Lemma 38.20 of Strasser (1985),

$$\begin{aligned} vol((\tilde{A}(t)+s/2)\cap(\{z \quad &: \quad f(z)>e\}+b+c+s/2)) \\ &\leq \quad vol((\tilde{A}(t)+s/2)\cap(\{z: f(z)>e\}). \end{aligned}$$

The equality follows when $b + c + s/2 = 0$ or $c = -s/2 - b = c^*$. This proves the result in (17). ■

We assume the environment of Theorem 1. Furthermore, we assume that $\sigma_\Delta^2 > 0$. Choose $\{h_i\}_{i=1}^m$ from an orthonormal basis $\{h_i\}_{i=1}^\infty$ of H. For $a \in \mathbf{R}^k$, we consider $h = \Sigma_{i=1}^m a_i h_i$ so that $\dot{\theta}_U(h) = \sum_{i=1}^m a_i \dot{\theta}_U(h_i) = a^\top \dot{\theta}_U$ and $\dot{\theta}_L(h) = a^\top \dot{\theta}_L$, where $\dot{\theta}_U = (\dot{\theta}_U(h_1), \cdots, \dot{\theta}_U(h_m))^\top$ and $\dot{\theta}_L = (\dot{\theta}_L(h_1), \cdots, \dot{\theta}_L(h_m))^\top$. Let $\bar{\theta}_B$ and $\bar{\Delta}_\theta$ be $m \times 2$ and $m \times 1$ matrices such that

$$\bar{\theta}_B \equiv \begin{bmatrix} \dot{\theta}_U(h_1) & \dot{\theta}_L(h_1) \\ \dot{\theta}_U(h_2) & \dot{\theta}_L(h_2) \\ \vdots & \vdots \\ \dot{\theta}_U(h_m) & \dot{\theta}_L(h_m) \end{bmatrix} \text{ and } \bar{\Delta}_\theta \equiv \begin{bmatrix} \dot{\Delta}_\theta(h_1) \\ \dot{\Delta}_\theta(h_2) \\ \vdots \\ \dot{\Delta}_\theta(h_m) \end{bmatrix}, \tag{18}$$

and $\bar{\zeta} \equiv (\zeta(h_1), \cdots, \zeta(h_m))^\top$, where ζ is the Gaussian process that appears in Assumption 1. We assume that $m \geq 2$ and $\bar{\theta}_B$ is full column rank. We fix $\lambda > 0$ and let $A_\lambda \in \mathbf{R}^m \sim N(0, I/\lambda)$ and let $F_\lambda(a|q)$ be the cdf of $\bar{\Sigma} A_\lambda + \bar{\mu}_q$ where

$$\bar{\mu}_q = \bar{\Delta}_\theta(\bar{\Delta}_\theta^\top \bar{\Delta}_\theta)^{-1} q \text{ and } \bar{\Sigma} \equiv I - \bar{\Delta}_\theta(\bar{\Delta}_\theta^\top \bar{\Delta}_\theta)^{-1} \bar{\Delta}_\theta^\top.$$

Finally let $\mathcal{L}(V^{h(a)})$ denote the distribution of $V^{h(a)}$ where V^h is the random vector that appears in Definition 2 and let $Z_\lambda(q) = (Z_{U,\lambda}(q), Z_{L,\lambda}(q))^\top \in \mathbf{R}^2$ be a random vector distributed as $N(\bar{\theta}_B^\top (I - \bar{\Sigma}\bar{\Sigma}_\lambda^{-1}\bar{\Sigma})\bar{\mu}_q, \bar{\theta}_B^\top \bar{\Sigma}\bar{\Sigma}_\lambda^{-1}\bar{\Sigma}\bar{\theta}_B)$, where $\bar{\Sigma}_\lambda = \bar{\Sigma} + \lambda I$. The following result is a variant of the convolution theorem that appears in Theorem 2.2 of van der Vaart (1989).

LEMMA A5: (i) *For any* $\lambda > 0$, $B \in \mathcal{B}$, *and* $q \in \mathbf{R}$,

$$P\left\{\int \mathcal{L}(V^{h(a)}) dF_\lambda(a|q) \in B\right\} = P\left\{Z_\lambda^{(m)}(q) + W_\lambda^{(m)}(q) \in B\right\},$$

where $W_\lambda^{(m)}(q) \in \mathbf{R}^2$ *is a random vector independent of* $Z_\lambda(q)$.
(ii) *Furthermore, as* $\lambda \to \infty$ *and* $m \to \infty$, $Z_\lambda^{(m)}(q)$ *weakly converges to the conditional distribution of* Z *given* $Z^\Delta = q$.

PROOF: (i) Under $P_{n,h}$, the local asymptotic normal experiments yield that

$$V_n^0 \rightsquigarrow V^0 \text{ along } \{P_{n,0}\} \text{ and } V_n^{h(a)} \rightsquigarrow V^0 - \bar{\theta}_B^\top a, \text{ along } \{P_{n,0}\}.$$

By Le Cam's third lemma, $P\left\{V^{h(a)} \in B\right\} = \mathbf{E}[1_B(V_0 - \bar{\theta}_B^\top a)e^{a^\top \bar{\zeta} - \frac{1}{2}||a||^2}]$. Note that

$$\begin{aligned}
& P\left\{\int \mathcal{L}(V^{h(a)}) dF_\lambda(a|q) \in B\right\} \\
= & \int \mathbf{E}\left[1_B\left\{V_0 - \bar{\theta}_B^\top(\bar{\Sigma}a + \bar{\mu}_q)e^{(\bar{\Sigma}a+\bar{\mu}_q)^\top \bar{\zeta} - \frac{1}{2}(\bar{\Sigma}a+\bar{\mu}_q)^\top(\bar{\Sigma}a+\bar{\mu}_q) - \frac{\lambda}{2}a^\top a}\right\}\right] \times (\lambda/2\pi)^{-m/2} e^{-\frac{\lambda}{2}a^\top a} da.
\end{aligned}$$

After some tedious calculations, we can write the integrand as

$$\begin{aligned}
& \int \mathbf{E}\left[1_B\left\{V_0 - \bar{\theta}_B^\top\left(\bar{\Sigma}a + \bar{\mu}_q + \bar{\Sigma}\bar{\Sigma}_\lambda^{-1}\bar{\Sigma}(\bar{\zeta} - \bar{\mu}_q)\right)\right\} \cdot c_q(\lambda)\right] dN(a|0, \bar{\Sigma}_\lambda^{-1}) \\
= & \int \mathbf{E}\left[1_B\left\{V_0 - \bar{\theta}_B^\top(\bar{\Sigma}a + (I - \bar{\Sigma}\bar{\Sigma}_\lambda^{-1}\bar{\Sigma})\bar{\mu}_q + \bar{\Sigma}\bar{\Sigma}_\lambda^{-1}\bar{\Sigma}\bar{\zeta})\right\} \cdot c_q(\lambda)\right] dN(a|0, \bar{\Sigma}_\lambda^{-1})
\end{aligned}$$

where $N(\cdot|0, \bar{\Sigma}_\lambda^{-1})$ is the cdf of $N(0, \bar{\Sigma}_\lambda^{-1})$ and

$$c_q(\lambda) = e^{-\frac{1}{2}(\bar{\zeta}-\bar{\mu}_q)^\top \bar{\Sigma}\bar{\Sigma}_\lambda^{-1}\bar{\Sigma}(\bar{\zeta}-\bar{\mu}_q) + \bar{\mu}_q^\top \bar{\zeta}} \times \left\{\det(\bar{\Sigma}_\lambda^{-1})\right\} \lambda^{-m/2} e^{-\frac{1}{2}\bar{\mu}_q^\top \bar{\mu}_q}.$$

Letting $W_\lambda(q)$ be a random vector with the distribution:

$$P\left\{W_\lambda(q) \in B\right\} = \mathbf{E}\left[1_B\left\{V_0 - \bar{\theta}_B^\top \bar{\Sigma}\bar{\Sigma}_\lambda^{-1}\bar{\Sigma}\bar{\zeta}\right\} \cdot c_q(\lambda)\right],$$

we obtain the wanted result.

(ii) As $\lambda \to 0$, $Z_\lambda^{(m)}(q) \to_d Z_0^{(m)}(q)$ where $Z_0^{(m)}(q) \in \mathbf{R}^2$ is a normal random vector with mean $\bar{\theta}_B^\top(I - \bar{\Sigma})\bar{\mu}_q = \bar{\theta}_B^\top \bar{\mu}_q$ and variance $\bar{\theta}_B^\top \bar{\Sigma}\bar{\theta}_B$. We can choose $\{h_i\}_{i=1}^m$ such that as $m \to \infty$, the Euclidean distances between $\bar{\theta}_B^\top \bar{\mu}_q$ and $[(\sigma_{U,\Delta}/\sigma_\Delta^{-2})q, (\sigma_{L,\Delta}/\sigma_\Delta^{-2})q]^\top$ and between $\bar{\theta}_B^\top \bar{\Sigma}\bar{\theta}_B$ and the matrix

$$\left[\begin{array}{ll}
\sigma_U^2 - \sigma_{U,\Delta}^2/\sigma_\Delta^2 & \sigma_{U,L} - \sigma_{U,\Delta}\sigma_{L,\Delta}/\sigma_\Delta^2 \\
\sigma_{U,L} - \sigma_{U,\Delta}\sigma_{L,\Delta}/\sigma_\Delta^2 & \sigma_L^2 - \sigma_{L,\Delta}^2/\sigma_\Delta^2
\end{array}\right]$$

become zero. Hence as $m \to \infty$, the distribution of $Z_0^{(m)}(q)$ converges to the conditional distribution of Z given $Z^\Delta = q$. ■

7.2 Proofs of the Main Results

Proof of Theorem 1: First, we assume that $\sigma_\Delta^2 > 0$. Fix $\varepsilon > 0$ and $r \in [\varepsilon, \infty)$. Let $\mathcal{G}(s)$ be the collection of probability measures that have a support in $[0, s]$, and let $S(V; \tau) \equiv \tau V_U + (1-\tau)V_L$ for $V = (V_U, V_L)^\top \in \mathbf{R}^2$. We write $V_n^h \equiv \sqrt{n}(\hat{\theta}_B - \theta_{B,n}(h))$. Fix $M > 0$ and a decision $\hat{d} = T(\hat{\theta}_B; \hat{\tau})$. We can write, from some large n on,

$$\begin{aligned}
\rho_h(\hat{d}) &\geq \sup_{G \in \mathcal{G}(\Delta_n(h))} \int \mathbf{E}_h \left[L_M(S(V_n^h; \hat{\tau}) - (1-\hat{\tau})\Delta_n(h) + \delta) \right] dG(\delta) \qquad (19) \\
&\geq \sup_{G \in \mathcal{G}(r-\varepsilon)} \int \mathbf{E}_h \left[L_M(S(V_n^h; \hat{\tau}) - (1-\hat{\tau})\Delta_n(h) + \delta) \right] dG(\delta),
\end{aligned}$$

for all $h \in H_{n,R}(r; \varepsilon)$. The second inequality follows because $\mathcal{G}(s)$ is increasing in s. Since $\Delta_n(h) \to \Delta_h$ as $n \to \infty$, we deduce that for each $\varepsilon > 0$, from some large n on, $H_R(r) \subset H_{n,R}(r, \varepsilon)$. Using this and (19), we find that

$$\liminf_{n\to\infty} \sup_{H_{n,R}(r;\varepsilon)} \rho_h(\hat{d}) \geq \liminf_{n\to\infty} \sup_{H_R(r)} \sup_{G \in \mathcal{G}(r-\varepsilon)} \int \mathbf{E}_h \left[L_M(S(V_n^h; \hat{\tau}) - (1-\hat{\tau})\Delta_n(h) + \delta) \right] dG(\delta).$$

Since L_M is bounded and continuous, and for each $h \in H_R(r)$,

$$S(V_n^h; \hat{\tau}) - (1-\hat{\tau})\Delta_n(h) \to_d S(V^h; \tau) - (1-\tau)r, \text{ along } \{P_{n,h}\}_{n\geq 1},$$

we deduce that for each $h \in H_R(r)$,

$$\begin{aligned}
& \liminf_{n\to\infty} \sup_{G \in \mathcal{G}(r-\varepsilon)} \int \mathbf{E}_h \left[L_M(S(V_n^h; \hat{\tau}) - (1-\hat{\tau})\Delta_n(h) + \delta) \right] dG(\delta) \\
\geq & \sup_{G \in \mathcal{G}(r-\varepsilon)} \int \liminf_{n\to\infty} \mathbf{E}_h \left[L_M(S(V_n^h; \hat{\tau}) - (1-\hat{\tau})\Delta_n(h) + \delta) \right] dG(\delta) \\
= & \sup_{G \in \mathcal{G}(r-\varepsilon)} \int \mathbf{E}_h \left[L_M(S(V^h; \tau) - (1-\tau)r + \delta) \right] dG(\delta).
\end{aligned}$$

The first inequality uses Fatou's Lemma. Since $\Delta_h = \dot{\Delta}_\theta(h) + \Delta_0$, it follows that $\dot{\Delta}_\theta(h) = r - \Delta_0$ if and only if $h \in H_R(r)$.

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 $\bar{H}$. Fix m and take $\{h_i\}_{i=1}^m \subset H$ and consider $h(a) = \sum a_i h_i$ for some $a = (a_i)_{i=1}^m \in \mathbf{R}^m$ such that $h(a) \in H_R(r)$. Fix $\lambda > 0$ and let $F_\lambda(a|q)$ be as defined prior to Lemma A5 above. Then the support of $F_\lambda(\cdot|r - \Delta_0)$ is confined to the set of a's such that $h(a) \in H_R(r)$. Utilizing the previous result and Fubini's theorem,

$$\begin{aligned}
& \liminf_{n\to\infty} \sup_{h\in H_R(r)} \sup_{G\in\mathcal{G}(r-\varepsilon)} \int \mathbf{E}_h \left[L_M(S(V_n^h;\hat{\tau}) - (1-\hat{\tau})\Delta_n(h) + \delta)\right] dG(\delta) \\
\geq & \int \sup_{G\in\mathcal{G}(r-\varepsilon)} \int \mathbf{E}_{h(a)} \left[L_M(S(V^{h(a)};\tau) - (1-\tau)r + \delta)\right] dG(\delta) dF_\lambda(a|r-\Delta_0) \\
\geq & \sup_{G\in\mathcal{G}(r-\varepsilon)} \int\int \mathbf{E}_{h(a)} \left[L_M(S(V^{h(a)};\tau) - (1-\tau)r + \delta)\right] dF_\lambda(a|r-\Delta_0) dG(\delta) \\
\geq & \sup_{G\in\mathcal{G}(r-\varepsilon)} \int\int \mathbf{E}_{h(a)} \left[L(S(V^{h(a)};\tau) - (1-\tau)r + \delta)\right] dF_\lambda(a|r-\Delta_0) dG(\delta) - q_M,
\end{aligned}$$

where $q_M > 0$ is a sequence such that $q_M \to 0$ as $M \to \infty$. The last inequality uses the condition that $\sup_{h\in H} \mathbf{E}[L(|V_U^h| + |V_L^h|)] < \infty$ in Definition 2. Using Lemma A5 and letting $q = r - \Delta_0$, we find that the last double integral is equal to

$$\begin{aligned}
& \int \mathbf{E}\left[L(S(Z_\lambda(q) + W_\lambda(q);\tau) - (1-\tau)r + \delta)\right] dG(\delta) \\
= & \int \mathbf{E}\left[L(Z_{L,\lambda}(q) + \tau Z_\lambda^\Delta(q) + S(W_\lambda(q);\tau) - (1-\tau)r + \delta)\right] dG(\delta) \\
= & \int\int \mathbf{E}\left[L(Z_{L,\lambda}(q) + \tau Z_\lambda^\Delta(q) + w - (1-\tau)r + \delta)\right] dG(\delta) dQ_{\lambda,q}(w)
\end{aligned}$$

where $Z_\lambda(q)$ and $W_\lambda(q)$ are as in Lemma A5 and $Z_\lambda^\Delta(q) \equiv Z_{U,\lambda}(q) - Z_{L,\lambda}(q)$ and $Q_{\lambda,q}$ is the cdf of $S(W_\lambda(q);\tau)$. The last equality follows because $Z_\lambda(q)$ and $W_\lambda(q)$ are independent. The supremum of the last double integral over $G \in \mathcal{G}(r-\varepsilon)$ is bounded from below by

$$\int\int \mathbf{E}\left[L(Z_{L,\lambda}(q) + \tau Z_\lambda^\Delta(q) + w - (1-\tau)r + \delta)\right] dG_{1/2}(\delta) dQ_{\lambda,q}(w),$$

where $G_{1/2}$ denotes the two-point distribution with equal masses at $\{0\}$ and $\{r-\varepsilon\}$. From the proof of (17) in the proof of Lemma A4, the last term is bounded from below by

$$\begin{aligned}
& \int \mathbf{E}\left[L\left(Z_{L,\lambda}(q) + \tau Z_\lambda^\Delta(q) - \mu_q - \frac{r-\varepsilon}{2} + \delta\right)\right] dG_{1/2}(\delta) \\
= & \mathbf{E}\left[L\left(Z_{L,\lambda}(0) + \tau Z_\lambda^\Delta(0) - \frac{r-\varepsilon}{2}\right)\right],
\end{aligned}$$

where $\mu_q = \mathbf{E}\left[Z_{L,\lambda}(q) + \tau Z_\lambda^\Delta(q)\right]$. The equality above follows because

$$\begin{aligned}
& \mathbf{E}\left[L(Z_{L,\lambda}(0) + \tau Z_\lambda^\Delta(0) - (r-\varepsilon)/2)\right] \\
= & \mathbf{E}\left[L(Z_{L,\lambda}(0) + \tau Z_\lambda^\Delta(0) + (r-\varepsilon)/2)\right].
\end{aligned}$$

We conclude that

$$\liminf_{n\to\infty} \sup_{H_{n,R}(r;\varepsilon)} \rho_h(\hat{d}) \geq \mathbf{E}\left[L(Z_{L,\lambda}(0) + \tau Z_\lambda^\Delta(0) - (r-\varepsilon)/2)\right] - q_M. \quad (20)$$

The left-hand side does not depend on λ and hence we send $\lambda \to 0$ on the right-hand side. As $\lambda \to 0$ and $m \to \infty$, $Z_\lambda(0)$ weakly converges to the conditional distribution of Z given $Z^\Delta = 0$, where $Z^\Delta \equiv Z_U - Z_L$. By sending $\lambda \to 0$ and $m \to \infty$ on the right-hand side of (20), we conclude that

$$\liminf_{n\to\infty} \sup_{H_{n,R}(r;\varepsilon)} \rho_h(\hat{d}) \geq \mathbf{E}\left[L(Z_L - (r-\varepsilon)/2)|Z^\Delta = 0\right] - q_M.$$

When we send $\varepsilon \to 0$, the conditional expectation becomes $\mathbf{E}\left[L(Z_L - r/2)|Z^\Delta = 0\right]$ by continuity of L. The wanted result then follows by sending $M \to \infty$.

Let us consider the case where $\sigma_\Delta^2 = 0$. In this case, if $r \neq \Delta_0$, $H_R(r) = \varnothing$, so that the lower bound becomes trivially zero. However if $r = \Delta_0$, we proceed as before by replacing $F_\lambda(a|q)$ with $\tilde{F}_\lambda(a)$ where $\tilde{F}_\lambda(\cdot)$ denotes the cdf of a normal random vector in $\mathbf{R}^2$ with mean zero and variance I/λ. The conditional distribution of Z given $Z^\Delta = 0$ is equal to the unconditional distribution of Z because $\mathbf{E}Z^\Delta = 0$ and $Var(Z^\Delta) = \sigma_\Delta^2 = 0$. Therefore, we obtain the wanted result. ■

Proof of Theorem 2: If $\sigma_\Delta^2 = 0$ and $r \neq \Delta_0$, the set $H_{n,R}(r;\varepsilon)$ converges to an empty set as $\varepsilon \to 0$. Hence $R_{n,M}^\varepsilon(\hat{d};r)$ converges to zero trivially. In the following we assume that either $\sigma_\Delta^2 > 0$ or $r = \Delta_0$. The case with $\sigma_\Delta^2 = 0$ and $r = \Delta_0$ is simpler as in the proof of Theorem 1. Hence we focus only on the case where $\sigma_\Delta^2 > 0$.

(i) Let $\hat{b}(r) \equiv (1/2 - \hat{\tau}^*)r$, $b(r) \equiv (1/2 - \tau^*)r$, and $Z_n^h \equiv \sqrt{n}\{\tilde{\theta} - \theta_n(h)\}$. First, observe that from some large n on, $\sup_{h\in H_{n,R}(r;\varepsilon)} \rho_{h,M}(\tilde{d}(r))$ is bounded by

$$\begin{aligned} & \sup_{h\in H_{n,R}(r;\varepsilon)} \sup_{G\in\mathcal{G}(r+\varepsilon)} \int \mathbf{E}_h\left[L_M\left(\sqrt{n}\left(\tilde{d}(r) - \theta_{U,n}(h)\right) + \delta\right)\right] dG(\delta) \quad (21) \\ = & \sup_{h\in H_{n,R}(r;\varepsilon)} \sup_{G\in\mathcal{G}(r+\varepsilon)} \int \mathbf{E}_h\left[L_M\left(S(Z_n^h;\hat{\tau}^*) - (1-\hat{\tau}^*)\Delta_n(h) + \delta + \hat{b}(r)\right)\right] dG(\delta) \\ \leq & \sup_{h\in H_{n,R}(r;\varepsilon)} \sup_{G\in\mathcal{G}(r+\varepsilon)} \sup_{r-\varepsilon\leq s\leq r+\varepsilon} \int \mathbf{E}_h\left[L_M\left(S(Z_n^h;\hat{\tau}^*) - (1-\hat{\tau}^*)s + \delta + \hat{b}(r)\right)\right] dG(\delta). \end{aligned}$$

Observe that by Assumption 4, for each $t \in \mathbf{R}$,

$$\sup_{h\in H}\left|P\left\{S(Z_n^h;\hat{\tau}^*) + \hat{b}(r) \leq t\right\} - P\left\{S(Z;\tau^*) + b(r) \leq t\right\}\right| \to 0.$$

Since the distribution of $S(Z;\tau^*)+b(r)$ is continuous, the above convergence is uniform in t (e.g. van der Vaart (1998)). Thus, using the fact that $(1-\hat{\tau})s=(1-\tau^*)s+o_P(1)$ uniformly in $s\in[r-\varepsilon,r+\varepsilon]$ and in $h\in H$, we deduce that

$$\sup_{h\in H}\left|P\left\{S(Z_n^h;\hat{\tau}^*)+\hat{b}(r)-(1-\hat{\tau}^*)s\le t\right\}-P\left\{S(Z;\tau^*)+b(r)-(1-\tau^*)s\le t\right\}\right|\to 0$$

uniformly in t and s. Therefore, the $\limsup_{n\to\infty}$ of the last term in (21) becomes the $\limsup_{n\to\infty}$ of

$$\sup_{h\in H_{n,R}(r;\varepsilon)}\sup_{G\in\mathcal{G}(r+\varepsilon)}\sup_{r-\varepsilon\le s\le r+\varepsilon}\int \mathbf{E}_h\left[L_M\left(S(Z;\tau^*)-(1-\tau^*)s+\delta+b(r)\right)\right]dG(\delta)$$

because L_M is bounded and continuous. Since the distribution of Z does not depend on h, this limsup becomes

$$\sup_{G\in\mathcal{G}(r+\varepsilon)}\sup_{r-\varepsilon\le s\le r+\varepsilon}\int \mathbf{E}\left[L_M\left(S(Z;\tau^*)-(1-\tau^*)s+\delta+b(r)\right)\right]dG(\delta).$$

The conditional distribution of Z_L given $Z^\Delta=0$ is a normal distribution. The mean of $Z_L+\tau^*Z^\Delta$ is equal to zero and its variance is equal to

$$\begin{aligned}\sigma_L^2+\tau^{*2}\sigma_\Delta^2+2\tau^*\sigma_{L,\Delta} &= \sigma_L^2+\tau^{*2}\sigma_\Delta^2+2\tau^*\sigma_{L,\Delta}\\ &= \sigma_L^2-\sigma_{L,\Delta}^2/\sigma_\Delta^2.\end{aligned}$$

The last quantity is the conditional variance of Z_L given $Z^\Delta=0$. Since $S(Z;\tau^*)=Z_L+\tau^*Z^\Delta$ is normal, the distribution of $Z_L+\tau^*Z^\Delta$ is equal to the conditional distribution of Z_L given $Z^\Delta=0$. Therefore,

$$\begin{aligned}&\sup_{G\in\mathcal{G}(r+\varepsilon)}\sup_{r-\varepsilon\le s\le r+\varepsilon}\int \mathbf{E}\left[L_M\left(S(Z;\tau^*)-(1-\tau^*)s+\delta+b(r)\right)\right]dG(\delta)\\ =&\sup_{G\in\mathcal{G}(r+\varepsilon)}\sup_{r-\varepsilon\le s\le r+\varepsilon}\int \mathbf{E}\left[L_M\left(Z_L-s/2-(1/2-\tau^*)s+\delta+b(r)\right)|Z^\Delta=0\right]dG(\delta)\\ =&\sup_{G\in\mathcal{G}(r+\varepsilon)}\sup_{r-\varepsilon\le s\le r+\varepsilon}\int \mathbf{E}\left[L_M\left(Z_L-s/2-(1/2-\tau^*)(s-r)+\delta\right)|Z^\Delta=0\right]dG(\delta)\\ \le&\sup_{G\in\mathcal{G}(r+\varepsilon)}\int \mathbf{E}\left[L_M\left(Z_L-(r+\varepsilon)/2+\delta\right)|Z^\Delta=0\right]dG(\delta)+\varphi_M(\varepsilon),\end{aligned}$$

for some function $\varphi_M(\varepsilon)\ge 0$ such that $\lim_{\varepsilon\to 0}\varphi_M(\varepsilon)=0$. The last inequality is due to the assumed uniform continuity of $L_M(\cdot)$. (See Assumption 3(ii).) By Lemma A3,

$\mathbf{E}\left[L_M\left(Z_L-(r+\varepsilon)/2+\delta\right)|Z^{\Delta}=0\right]$ is quasiconvex in δ. Hence the supremum over $G \in \mathcal{G}(r+\varepsilon)$ is equal to the supremum over two-point distributions with support in $\{0, r+\varepsilon\}$. However, the value of $\mathbf{E}\left[L_M\left(Z_L-(r+\varepsilon)/2+\delta\right)|Z^{\Delta}=0\right]$ is identical regardless of whether $\delta=0$ or $\delta=r+\varepsilon$ by the central symmetry of L_M and that of the conditional distribution of Z_L given $Z^{\Delta}=0$. Hence we bound the last term in the previous display by

$$\mathbf{E}\left[L_M(Z_L-(r+\varepsilon)/2)|Z^{\Delta}=0\right]+\varphi_M(\varepsilon).$$

By sending $\varepsilon \to 0$ and then $M \to \infty$, we obtain the wanted result.

(ii) The proof is the same as that of (i) only with τ^* replaced by 1/2. Since $\tilde{d}_{1/2}$ does not depend on r, the wanted result immediately follows. ■

Proof of Theorem 3: Following the similar steps as in the proof of Theorem 2 and using Lemma A3, we find that as $n \to \infty$,

$$\begin{aligned}
R_{n,M}(\tilde{d}_{1/2};r_n) &\leq \sup_{G\in\mathcal{G}(r_n+\varepsilon)}\int \mathbf{E}\left[L_M\left(Z_L+\frac{Z^{\Delta}}{2}-\frac{r_n}{2}+\delta\right)\right]dG(\delta)\\
&= \mathbf{E}\left[L_M\left(Z_L+\frac{Z^{\Delta}}{2}-\frac{r_n}{2}\right)\right]+\varphi_M(\varepsilon)+o(1)\\
&\leq \mathbf{E}\left[L\left(Z_L+\frac{Z^{\Delta}}{2}-\frac{r_n}{2}\right)\right]+\varphi_M(\varepsilon)+o(1).
\end{aligned}$$

Now, the leading expectation is rewritten as

$$\begin{aligned}
&\mathbf{E}\left[L\left(Z_L+\tau^*Z^{\Delta}+\left(\frac{1}{2}-\tau^*\right)Z^{\Delta}-\frac{r_n}{2}\right)\right]\\
=&\int \mathbf{E}\left[L\left(Z_L+\left(\frac{1}{2}-\tau^*\right)s\sigma_{\Delta}-\frac{r_n}{2}\right)|Z^{\Delta}=0\right]d\Phi(s).
\end{aligned}$$

The above equality follows because the distribution of $Z_L+\tau^*Z^{\Delta}$ is equal to the conditional distribution of Z_L given $Z^{\Delta}=0$ and $Z_L+\tau^*Z^{\Delta}$ and Z^{Δ} are uncorrelated and hence independent due to their joint normality. Therefore,

$$\begin{aligned}
&R_{n,M}(\tilde{d}_{1/2};r_n)-\mathbf{E}\left[L\left(Z_L-\frac{r_n}{2}\right)|Z^{\Delta}=0\right]\\
\leq&\int \mathbf{E}\left[L\left(Z_L+\left(\frac{1}{2}-\tau^*\right)s\sigma_{\Delta}-\frac{r_n}{2}\right)-L\left(Z_L-\frac{r_n}{2}\right)|Z^{\Delta}=0\right]d\Phi(s)+\varphi_M(\varepsilon)+o(1).
\end{aligned}$$

The leading integral is written as

$$\begin{aligned}&\int \mathbf{E}\left[\left|Z_L+\left(\frac{1}{2}-\tau^*\right)s\sigma_\Delta-\frac{r_n}{2}\right|-\left|Z_L-\frac{r_n}{2}\right||Z^\Delta=0\right]d\Phi(s)\\ =\ & -\int\left(\frac{1}{2}-\tau^*\right)s\sigma_\Delta d\Phi(s)+o(1)\text{ as }r_n\to\infty,\end{aligned}$$

because as $r_n \to \infty$,

$$\begin{aligned}\int P\left\{Z_L+\left(\frac{1}{2}-\tau^*\right)s\sigma_\Delta-\frac{r_n}{2}>0|Z^\Delta=0\right\}d\Phi(s) &\to 0\text{ and}\\ P\left\{Z_L-\frac{r_n}{2}>0|Z^\Delta=0\right\} &\to 0.\end{aligned}$$

However, $\int\left(\frac{1}{2}-\tau^*\right)s\sigma_\Delta d\Phi(s)=\left(\frac{1}{2}-\tau^*\right)\sigma_\Delta\int sd\Phi(s)=0$, yielding the wanted result. ■.

Proof of Corollary 1: (i) When $r_n \to 0$, $|\tilde{d}(\hat{r}_n)-\tilde{d}(0)|=0$ with probability approaching one. Since $R^\varepsilon_{n,M}(\tilde{d}(0);r_n)$ involves the supremum over $h\in H_{n,R}(r;\varepsilon)$ and from some large n on,

$$H_{n,R}(r_n;\varepsilon)\subset H_{n,R}(r;2\varepsilon),$$

we deduce that

$$\lim_{M\uparrow\infty}\lim_{\varepsilon\to 0}\limsup_{n\to\infty}R^\varepsilon_{n,M}(\tilde{d}(0);r_n)\leq\lim_{M\uparrow\infty}\lim_{\varepsilon\to 0}\limsup_{n\to\infty}R^\varepsilon_{n,M}(\tilde{d}(0);0)$$

as $r_n \to 0$. By Theorem 2(ii) with $r=0$, the wanted result follows.

(ii) When $r_n/(\sqrt{n}b_n)\to\infty$, $|\tilde{d}(\hat{r}_n)-\tilde{d}_{1/2}|=0$ with probability approaching one. The wanted result follows from Theorem 3. ■

References

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

[2] Andrews, D. W. K. and P. Guggenberger. (2009a): "Hybrid and size-corrected subsampling methods," forthcoming in *Econometrica*.

[3] Andrews, D. W. K. and P. Guggenberger. (2009b): "Validity of subsampling and "plug-in asymptotic" inference for parameters defined by moment inequalities," forthcoming in *Econometric Theory*.

[4] Andrews, D. W. K. and P. Jia. (2008): "Inference for parameters defined by moment inequalities: a recommended moment selection procedure," Working Paper.

[5] Andrews, D. W. K. and G. Soares. (2007): "Inference for parameters defined by moment inequalities using generalized moment selection," Cowles Foundation Discussion Paper 1631.

[6] Ball, K. (1997): "An elementary introduction to modern convex geometry," in *Flavors of Geometry*, MSRI Publications 31, Cambridge University Press, Cambridge, 1997.

[7] Battacharya, J., A. Shaikh, and E. J. Vytlacil. (2008): "Treatment effect bounds under monotonicity assumptions: an application to Swan-Ganz catheterization," *American Economic Review Papers and Proceedings*, 98, 351-356.

[8] 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.

[9] Beresteanu, A. and F. Molinari (2008): "Asymptotic properties for a class of partially identified models," *Econometrica*, 76, 763-814.

[10] Bickel, P. J., C. A. J. Klassen, Y. Ritov, J. A. Wellner (1993): *Efficient and Adaptive Estimation for Semiparametric Models*, Johns Hopkins University Press, Baltimore.

[11] Bugni, F.A. (2007): "Bootstrap inference in partially identified models," Working Paper.

[12] Canay, I. (2009): "EL inference for partially identified models: large deviations optimality and bootstrap validity," Working Paper.

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

[14] Chernozhukov, V., I. Fernández-Val, J. Hahn, and W. Newey (2009): "Identification and estimation of marginal effects in nonlinear panel models," CEMMP, CWP05/09, University of College London.

[15] Chesher, A. (2005): "Nonparametric identification under discrete variation," *Econometrica*, 73, 1525-1550.

[16] Fan, Y. and S. Park (2007): "Confidence sets for some partially identified parameters," Working paper, Vanderbilt University.

[17] Fukuda, K. and T. Uno (2007): "Polynomial time algorithms for maximizing the intersection volume of polytopes," *Pacific Journal of Optimization*, 3, 37-52.

[18] Gilstein, C. Z. and E. Leamer (1983): "Robust sets of regression estimates," *Econometrica*, 51, 555–582.

[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] Hajék, J. (1970): "A characterization of limiting distributions of regular estimators," *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* 14, 323-330.

[22] Hajék, 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.

[23] Hajék, J. and Z. Šidák (1967): *Theory of Rank Tests*, Academic Press, New York.

[24] Hansen, L., J. Heaton, and E. G. Luttmer (1995): "Econometric evaluation of asset pricing models," *Review of Financial Studies*, 8, 237–274.

[25] Hirano, K. and J. Porter (2008): "Asymptotics for statistical treatment rules," Forthcoming in *Econometrica*.

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

[27] Horowitz, J. and C. F. Manski (2000): "Nonparametric analysis of randomized experiments with missing covariate and outcome data," *Journal of the American Statistical Association*, 95, 77-84.

[28] Imbens, G. and C. Manski (2004): "Confidence intervals for partially identified parameters," *Econometrica*, 72, 1845-1857.

[29] Le Cam, L. (1960): "Locally asymptotically normal families of distributions," *University of California Publications in Statistics,* 3, 3-32.

[30] Le Cam, L. (1972): "Limits of experiments," 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. 245-261.

[31] Le Cam, L. (1979): "On a theorem of J. Hajék," in J. Jurečková, ed. *Contributions to Statistics - Hajék Memorial Volume*, Akademian, Prague, p.119-135.

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

[33] Manski, C. F. (1990): "Nonparametric bounds on treatment effects," *American Economic Review, Papers and Proccedings*, 80, 319–323.

[34] Manski, C. F. (2004): *Partial Identification of Probability Distributions*, Springer, New York.

[35] Manski, C. F. (2008): *Identification for Prediction and Decision*, Havard University Press.

[36] Manski, C. F. and J. V. Pepper (2000): "Monotone instrumental variables: with an application to the returns to schooling," *Econometrica*, 68, 997-1010.

[37] Manski, C. F. and E. Tamer (2002): "Inference on regressions with interval data on a regressor or outcome," *Econometrica*, 70, 519–547.

[38] McFadden, D. (1999): "Computing willingness-to-pay in random utility models," In R. Melvin, C. Moore and R. Riezman, eds. *Trade Theory and Econometrics*, New York: Routledge, 253-274.

[39] Moon, H. R. and F. Schorfheide (2009): "Bayesian and frequentist inference in partially identified models," NBER Working Paper 14882.

[40] Newey, W. K. (1990): "Semiparametric efficiency bounds," *Journal of Applied Econometrics*, 5, 99-135.

[41] Pakes, A. J. R. Porter, K. Ho, and J. Ishii (2006): "Moment inequalities and their application," Working Paper.

[42] Phillips, P. C. B. (1989): "Partially identified econometric models," *Econometric Theory*, 5, 181–240.

[43] Romano P. and A. Shaikh (2009): "Inference for the identified set in partially identified econometric models," Working Paper.

[44] Rosen, A. (2008): "Confidence sets for partially identified parameters that satisfy a finite number of moment Inequalities," *Journal of Econometrics*, 146, 107-117.

[45] Stoye, J. (2009a): "More on confidence intervals for partially identified models," *Econometrica*, 77, 1299-1315.

[46] Stoye, J. (2009b): "Minimax regret treatment choice with finite samples," *Journal of Econometrics*, 151, 70-81.

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

[48] Tripathi, G. (2000): "Local semiparametric efficiency bounds under shape restrictions," *Econometric Theory* 16, 729-739.

[49] Uhrin, B. (1984): "Some remarks about the convolution of unimodal functions," *Annals of Probability*, 12, 640-645.

[50] van der Vaart, A. W. (1988): *Statistical Estimation in Large Sample Spaces*, CWI Tracts 44, Amsterdam.

[51] van der Vaart, A. W. (1989): "On the asymptotic information bound," *Annals of Statistics* 17, 1487-1500.

[52] van der Vaart, A. W. (1991): "On differentiable functionals," *Annals of Statistics* 19, 178-204.

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

[54] Wintner, A. (1938): *Asymptotic Distributions and Infinite Convolutions,* Edwards Brothers, Ann Arbor.