Point Decisions for Interval-Identified Parameters

Kyungchul Song[1]
Department of Economics, University of British Columbia

January 31, 2013

Abstract

This paper considers a decision-maker who prefers to make a point decision when the object of interest is interval-identified with regular bounds. When the bounds are just identified along with known interval length, the local asymptotic minimax decision with respect to a symmetric convex loss function takes an obvious form: an efficient lower bound estimator plus the half of the known interval length. However, when the interval length or any nontrivial upper bound for the length is not known, the minimax approach suffers from triviality because the maximal risk is associated with infinitely long identified intervals. In this case, this paper proposes a local asymptotic minimax regret approach and shows that the mid-point between semiparametrically efficient bound estimators is optimal.

Key words and Phrases: Interval Identification, Semiparametric Efficiency, Local Asymptotic Minimax Regret Decisions.

JEL Classifications: C01, C13, C14, C44.

1 Introduction

Many objects of inference in economics are 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, treatment decisions are based on the estimated treatment effects in various program evaluations or medical studies. In such an environment, a point decision (or estimate) about the object

[1] I thank Don Andrews, Xu Cheng, Jinyong Hahn, Ulrich Müller, Aureo de Paula, Frank Schorfheide, Xun Tang, and Yoon-Jae Whang for valuable comments. I thank Co-Editor and two referees for their detailed comments that have led to a substantial improvement of the paper. All errors are mine. Corresponding address: Kyungchul Song, Department of Economics, University of British Columbia, 997-1873 East Mall, Vancouver, BC, Canada, V6T 1Z1. Email address: kysong@mail.ubc.edu.

of interest is preferred to a set estimate 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 (or estimate). This paper attempts to address this dilemma by searching for an optimal point decision when the object of interest is interval-identified.

Searching for a point estimator of an interval-identified parameter may sound odd at first. The problem of estimation is fundamentally a decision problem, for it is the problem of producing an "appropriate" mapping from the observed data into the parameter space. A formal definition of "appropriateness" of an estimator presumes a decision-theoretic description of the problem. As far as the nature of the problem is decision-theoretic, the candidate class of decisions - whether it be a point or an interval - is a matter of the decision maker's choice. This decision-theoretic perspective where an optimal decision is pursued regardless of the informational content of the observations makes contrast with the perspective of inference where one is allowed to declare "non-discovery" when the informational content of the observations is thin. This paper demonstrates how the problem of point estimation can be formulated in decision-theoretic terms even when the object of interest is interval identified.

Let X_n be a random vector of observations with distribution P_n. We are interested in an object $\theta_0 \in [\theta_L, \theta_U]$, where $\theta_B = (\theta_U, \theta_L)^\top$, $\theta_L \leq \theta_U$, is identified by P_n. See Manski (2003) for an overview of the bound approach. (See also Imbens and Manski (2004), Stoye (2009a) and Fan and Park (2009) for examples and for inference procedures.) The boundary parameter θ_B is not known, but has $\sqrt{n}$-consistent and asymptotically efficient estimators constructed using X_n. This paper focuses on the problem of deciding on (or estimating) θ_0 after observing X_n.

When the interval length has a known bound along which the interval bounds are just identified, one can show that the local asymptotic minimax approach of Hájek (1972) and Le Cam (1979) with respect to a symmetric convex loss gives an intuitive solution: a semiparametrically efficient lower bound estimator plus the half of the upper bound for the interval length. When there is no known bound for the interval length that is nontrivial (i.e., there exist sampling variations in which the bound can be potentially violated when the bound is not externally forced on the variations), the local asymptotic minimax approach does not provide a meaningful solution, because the worst possible scenario that the minimax approach focuses on arises when the interval is infinitely long.

This paper introduces the approach of local asymptotic minimax regret, and shows that the mid-point between efficient estimators of upper and lower bounds is a minimax regret decision. We call this estimator a *mid-point decision*. First, the paper establishes a lower bound for the local asymptotic minimax regret bound. Defining a minimax regret bound encounters a difficulty that pertains to the analysis of a set-identified parameter: the asymptotics that sustains a fixed interval length with the growing sample size does not provide a good description of the finite sample environment for decisions. To address this issue, this paper adopts what this paper calls *asymptotics of near identification* where the interval length shrinks along with the sample size as the bound estimators become more accurate.

The minimax regret approach has a long history in statistical decision theory. (See Berger (1985) for example.) In econometrics, the approach has been recently employed by Manski (2004) and a few others such as Tetenov (2007), Hirano and Porter (2009) and Stoye (2009b) who applied the approach in search of good statistical treatment rules.

A recent contribution by Kitagawa (2012) explores robust Bayes decisions for set-identified parameters. The main advantage of Kitagawa (2012) is three-fold. First, it offers a finite sample inference method as compared to this paper's approach that relies on asymptotic theory. Second, the class of identified sets are more general than the class of identified sets here - which are intervals. Third, in contrast with this paper, Kitagawa (2012) allows the loss functions to be asymmetric. On the other hand, the robustness of decisions that this paper pursues is the robustness against any local perturbation of the true probability in any direction, and hence various semiparametric or nonparametric models for observations are accommodated. This makes contrast with Kitagawa (2012)'s approach which focuses on a parametric model for observations, and pursues robustness only against various priors for the object of interest, not against (local) likelihood misspecifications, while assuming a single prior distribution for the identified parameters.

2 Boundary Parameters, Loss, and Risks

Let us introduce boundary parameter θ_B formally. Let $\mathbb{N}$ be the collection of natural numbers. Suppose that $\mathcal{P} = \{P_\alpha : \alpha \in \mathcal{A}\}$ is 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. Suppose that $Y_1, \cdots, Y_n$ are i.i.d. draws from $P_{\alpha_0} \in \mathcal{P}$ so that $X_n \equiv (Y_1, \cdots, Y_n)$ is distributed as $P^n_{\alpha_0}$. We focus on local asymptotic analysis centered around α_0. For this, we consider $\alpha_{n,h} = \alpha_0 + \lambda_h/\sqrt{n}$, $\lambda_h \in \mathcal{A}$, and the direction λ_h is indexed by $h \in H$, where $(H, \langle \cdot, \cdot \rangle)$ is a subspace of a separable Hilbert space called a *tangent space*. Hence we denote $P_{n,h} = P^n_{\alpha_{n,h}}$ for simplicity, and consider sequences $\{P_{n,h}\}_{n \geq 1}$ indexed by $h \in H$. From here on, one can view H as an infinite dimensional parameter space indexing the sequences of distributions for X_n. (For a formal background, see Appendix A.)

The boundary parameter θ_B is identified under $P_{n,h}$ for each $n \in \mathbb{N}$ and is viewed here as a sequence of $\mathbf{R}^2$-valued maps on H, i.e., $\theta_{B,n}(h) = (\theta_{U,n}(h), \theta_{L,n}(h))^\top$, $h \in H$. *Identification* here means that the map $\theta_{B,n}(\cdot)$ is point-valued in $\mathbf{R}^2$, not set-valued. We keep the notation $\theta_B = (\theta_U, \theta_L)^\top$ when we do not need to make explicit its dependence on n.

As for $\theta_B = (\theta_U, \theta_L)^\top$, we assume two standard conditions: local asymptotic normality of probabilities that identify θ_B (Assumption A1) and a smooth behavior of θ_B at the local perturbation of the probabilities (Assumption A2). These two conditions are well-known and well studied in the literature of semiparametrically efficient estimation (e.g. van der Vaart (1991), and Bickel, Klaassen, Ritov, and Wellner (1993)). Their formal statements are found in Appendix A.

The loss for each decision $d \in \mathbf{R}$ is given by

$$L\left(d - \theta_0\right) = |d - \theta_0|^\alpha, \ \alpha \in [1, \infty). \tag{1}$$

The loss function is symmetric. While assuming symmetric losses can serve as a benchmark case, this assumption excludes interesting situations where asymmetric losses are appropriate as in Manski (2004) and Hirano and Porter (2009).

For each $n \in \mathbb{N}$ and given an observed random vector X_n taking values in a set $\mathcal{X}_n$, define $\mathcal{D}_n$ to be the collection of random variables of the form $\hat{d} = \Psi_n(X_n)$, where for each $n \in \mathbb{N}$, Ψ_n is a measurable function: $\mathcal{X}_n \to \mathbf{R}$. Each member $\hat{d} \in \mathcal{D}_n$ is a candidate estimator of θ_0 that is constructed using X_n. Each $\hat{d} \in \mathcal{D}_n$ is associated with its normalized risk $n^{\alpha/2}\mathbf{E}_h[L(\hat{d} - \theta_0)]$, where $\mathbf{E}_h$ denotes expectation under $P_{n,h}$.

Since θ_0 is not point-identified, the risk is not identified either. This paper introduces what this paper calls the *identifiable maximal risk*:

$$\rho_h(\hat{d}) = \sup_{\theta \in [\theta_{L,n}(h), \theta_{U,n}(h)]} n^{\alpha/2}\mathbf{E}_h\left[L(\hat{d} - \theta)\right], \ \hat{d} \in \mathcal{D}_n. \tag{2}$$

(Throughout the paper, the supremum of a nonnegative map over an empty set is set to be zero.) The identifiable maximal risk is the largest risk possible by any potential location of θ_0 in the identified interval $[\theta_{L,n}(h), \theta_{U,n}(h)]$. (The form of the identifiable maximal risk in (2) is due to comments from Don Andrews. A related notion is also employed by Tetenov (2009).)

3 Local Asymptotic Minimax Regret Decisions

3.1 Overview and Examples

This paper focuses on a minimax regret approach.[2] The unknown state in the decision problem is the parameter $h \in H$. For each n, if the decision-maker knew $h \in H$ which indexes the underlying probabilities, she would solve the following problem:

$$\inf_{\hat{d} \in \mathcal{D}_n} \rho_h(\hat{d}) = \inf_{\hat{d} \in \mathcal{D}_n} \sup_{\theta_{L,n}(h) \leq \theta \leq \theta_{U,n}(h)} n^{\alpha/2} \mathbf{E}_h \left[L(\hat{d} - \theta) \right].$$

The maximal regret is written as:

$$\sup_{h \in H : \Delta_n(h) \geq 0} \left\{ \rho_h(\hat{d}) - \inf_{\hat{d} \in \mathcal{D}_n} \rho_h(\hat{d}) \right\},$$

where for $h \in H$,

$$\Delta_n(h) \equiv \theta_{U,n}(h) - \theta_{L,n}(h).$$

Once $h \in H$ is known, so are $\theta_{U,n}(h)$ and $\theta_{L,n}(h)$, and hence a minimizer of $\rho_h(\hat{d})$ over $\hat{d} \in \mathcal{D}_n$ is $\{\theta_{U,n}(h) + \theta_{L,n}(h)\}/2$, so that we have

$$\inf_{\hat{d} \in \mathcal{D}_n} \rho_h(\hat{d}) = n^{\alpha/2} L \left(\frac{\Delta_n(h)}{2} \right).$$

Therefore, the maximal regret is equal to

$$\sup_{h \in H : \Delta_n(h) \geq 0} \left\{ \rho_h(\hat{d}) - n^{\alpha/2} L \left(\frac{\Delta_n(h)}{2} \right) \right\}.$$

A *local asymptotic minimax regret decision* is defined to be one that minimizes an asymptotic version of this maximal regret. In Theorem 2 below, it is shown that if $(\tilde{\theta}_U, \tilde{\theta}_L)$ is a semiparametrically efficient estimator of (θ_U, θ_L), the mid-point decision

$$\tilde{d}_{1/2} \equiv \frac{\tilde{\theta}_U + \tilde{\theta}_L}{2}$$

is a local asymptotic minimax regret decision.

To see this result heuristically, assume that $L(x) = x^2$ and consider for simplicity candi-

[2] As compared to a previous version of this paper, the definition of a minimax regret decision is now reformulated here thanks to a referee's comment and Co-Editor's suggestion.

date decisions of the following form:

$$\hat{d}(\tau, b) = \tau\hat{\theta}_U + (1-\tau)\hat{\theta}_L + b, \ \ \tau \in \mathbf{R} \text{ and } b \in \mathbf{R}, \tag{3}$$

where $\hat{\theta}_B = (\hat{\theta}_U, \hat{\theta}_L)^\top \in \mathbf{R}^2$ is such that for all sample sizes n,

$$\sqrt{n}(\hat{\theta}_B - \theta_{B,n}(h)) \sim N(0, S), \text{ with a positive definite matrix } S. \tag{4}$$

(The choice of candidate decisions $\hat{d}(\tau, b)$ made here is only for a heuristic purpose.) Then we can write for $\theta \in [\theta_{L,n}(h), \theta_{U,n}(h)]$,

$$\sqrt{n}\{\hat{d}(\tau, b) - \theta\} = \tilde{Z}_L + \tau\tilde{Z}^\Delta + (\tau - s_n(h))\sqrt{n}\Delta_n(h) + \sqrt{n}b, \tag{5}$$

where $\tilde{Z}_U \equiv \sqrt{n}(\hat{\theta}_U - \theta_{U,n}(h))$, $\tilde{Z}_L \equiv \sqrt{n}(\hat{\theta}_L - \theta_{L,n}(h))$, $\tilde{Z}^\Delta \equiv \tilde{Z}_U - \tilde{Z}_L$, and $s_n(h) \equiv (\theta - \theta_{L,n}(h))/\Delta_n(h)$. From (5), the maximal regret (after maximizing over $\theta \in [\theta_{L,n}(h), \theta_{U,n}(h)]$) becomes:

$$\sup_{s\in[0,1]} \mathbf{E}\left[\left(\tilde{Z}_L + \tau\tilde{Z}^\Delta + (\tau - s)\sqrt{n}\Delta_n(h) + \sqrt{n}b\right)^2\right] - \left(\frac{\sqrt{n}\Delta_n(h)}{2}\right)^2.$$

When $\tau > 1/2$, the maximal regret can be made to diverge to infinity by setting $s = 0$ and $\sqrt{n}\Delta_n(h) \uparrow \infty$. When $\tau < 1/2$, the regret can be made to diverge to infinity by setting $s = 1$ and $\sqrt{n}\Delta_n(h) \uparrow \infty$. When $\tau = 1/2$, the maximal regret becomes

$$\sup_{s\in[0,1]} \mathbf{E}\left[\left(\frac{\tilde{Z}_U + \tilde{Z}_L}{2}\right)^2\right] + n\left(b + (1-s)\Delta_n(h)\right)\left(b - s\Delta_n(h)\right). \tag{6}$$

One can easily check that unless $b = 0$, the maximal regret can be made to be infinity, e.g., by taking $s = 1\{b < 0\}$ and $\sqrt{n}\Delta_n(h) \uparrow \infty$. Hence the minimax regret is achieved by taking $\tau = 1/2$ and $b = 0$, i.e., by a mid point decision.

The heuristic derivation of optimal decisions so far in large part relies on the choice of the candidate decisions of the form (3). It is the main result of this paper that such an optimality result, once we replace $\hat{\theta}_B$ by a semiparametrically efficient estimator $\tilde{\theta}_B$ of θ_B, continues to hold even when we expand the candidate decisions widely so that any decision (as any measurable function of observations) are now taken to be a candidate decision.

When one adopts the seemingly natural asymptotics of a *fixed* positive interval length $\Delta_n(0) = c > 0$ for all n, we have $\sqrt{n}\Delta_n(h) \to \infty$ as $n \to \infty$. Since $\tilde{Z}_U + \tilde{Z}_L$ is stochastically bounded for all n, it is like focusing only on the case where the interval length is infinity, resulting in the maximal regret for any decision $\hat{d}(\tau, b)$ taking the value of either $\mathbf{E}[(\tilde{Z}_U + \tilde{Z}_L)^2]/4$ (with $\tau = 1/2$ and $b = 0$) or ∞. Hence to reflect properly the finite sample situation with finite interval length, this paper introduces alternative asymptotics which this paper calls *asymptotics of near identification*. The asymptotic scheme is expressed in the following assumption.

Assumption 1: $\Delta_0 \equiv \lim_{n\to\infty} \sqrt{n}\Delta_n(0)$ exists in $[0, \infty)$.

Assumption 1 says that the interval length $\Delta_n(h)$ at $h = 0$ (i.e. at the true data generating process) converges to a constant Δ_0 at the rate of $\sqrt{n}$. Along with the differentiability condition for the bound parameters (Assumption A2 in the appendix), this assumption implies that for all $h \in H$,

$$\Delta_h \equiv \lim_{n\to\infty} \sqrt{n}\Delta_n(h)$$

exists in $[0, \infty)$.

One major criticism regarding asymptotics of near identification that the author has received in numerous occasions is that the paper's analysis begins with a partially identified parameter but reduces the analysis to the easy case of point identification through asymptotics of near identification. This criticism stems perhaps from misunderstanding the basic motivation for the asymptotic device here. The asymptotics of near identification is analogous to the local power analysis in hypothesis test where one chooses a sequence of alternatives that are local around the boundary of the null hypothesis. The local analysis is motivated by the fact that for a consistent test, under a fixed alternative, its power converges to one, preventing one from comparing the power properties of different consistent tests. Similarly, the asymptotic near identification approach chooses a sequence of probabilities that are local around point identification. Point-identification is not assumed under near identification, just as the null hypothesis is not assumed under local alternatives.

Now let us consider examples.

Example 1 (Missing Outcomes with Identified Treatment Probability): Let (Y, W) be a pair of random variables such that $W \in \{0, 1\}$, $\mathbf{E}W = p$ and $\mathbf{E}[Y|W = 1] = \mu$. The econometrician observes (YW, W), and is interested in the parameter $\theta = \mathbf{E}[Y]$. As shown in Imbens and Manski (2004), the identified interval for θ is $[\theta_L, \theta_U]$, where $\theta_L = \mu p$ and $\theta_U = \mu p + 1 - p$.

Example 2 (Bounds on Conditional Treatment Effects): This example is taken from Manski (1990). Suppose that $Y_d \in [0, 1]$, $d \in \{0, 1\}$, is an outcome for the treated $(d = 1)$ or not treated $(d = 0)$. A random variable $D \in \{0, 1\}$ represents the treatment incidence, with $D = 1$ representing treatment, and a discrete random vector X represents a vector of covariates. The object of interest is the conditional treatment effect:

$$T(x) = \mathbf{E}[Y_1|X = x] - \mathbf{E}[Y_0|X = x].$$

Without further assumptions about the underlying data generating process, Manski (1990) showed that $T(x)$ is in the interval $[\theta_L, \theta_U]$, where

$$\begin{aligned} \theta_L &= y_1 p_1 - y_0 p_0 - p_1, \\ \theta_U &= y_1 p_1 - y_0 p_0 - p_1 + 1, \end{aligned}$$

$p_d = P\{D = d|X = x\}$ and $y_d = \mathbf{E}[Y_d|X = x, D = d]$, $d \in \{0, 1\}$.

Example 3 (Censored Outcomes and Covariates due to Survey Nonresponses): Horowitz and Manski (1998) established bounds for a conditional outcome given covariates and its asymptotic bias, when outcomes and covariates are censored. Let $Y \in [0, 1]$ be an outcome variable and $D \in \{0, 1\}$ denotes the censoring indicator, with $D = 1$ representing the noncensoring event. The object of interest is $\mathbf{E}[Y|X \in A]$, where X is a covariate vector and A is a designated set. Horowitz and Manski (1998) showed that $\mathbf{E}[Y|X \in A]$ is identified in the interval $[\theta_L, \theta_U]$, where

$$\begin{aligned} \theta_L &= \mathbf{E}[Y|X \in A, D = 1] \cdot q_1 \text{ and} \\ \theta_U &= \mathbf{E}[Y|X \in A, D = 1] \cdot q_1 + q_0, \end{aligned}$$

with

$$q_1 = \frac{P\{X \in A|D = 1\}P\{D = 1\}}{P\{X \in A|D = 1\}P\{D = 1\} + P\{D = 0\}}$$

and $q_0 = 1 - q_1$.

Example 4 (Missing Treatments): Molinari (2010) offers identification analysis for status quo treatment effects and average treatment effects, when some treatment decisions are not observed due to survey nonresponse. One of her examples is the following. Suppose that $Y_d \in [0, 1]$ denotes the potential outcome for the treated $(d = 1)$ and the untreated $(d = 0)$, $Z \in \{0, 1\}$ indicates the treatment decision with $Z = 1$ meaning the decision to be treated, and $D \in \{0, 1\}$ the selection indicator, where $D = 1$ means that the treatment decision is observed. Then Molinari showed that the average treatment effect $\mathbf{E}[Y_1] - \mathbf{E}[Y_0]$

is interval identified with the identified interval $[\theta_L, \theta_U]$, where

$$\begin{aligned}\theta_L &= B_L - P\{D=0\} \text{ and}\\ \theta_U &= B_U + P\{D=0\},\end{aligned}$$

and, with $y_{s,d} \equiv \mathbf{E}\left[Y1\{D=s, Z=d\}\right]$, $d, s \in \{0,1\}$,

$$\begin{aligned}B_L &= y_{1,1} - y_{1,0} - P\left\{D=1, Z=1\right\} \text{ and}\\ B_U &= y_{1,1} - y_{1,0} + P\left\{D=1, Z=0\right\}.\end{aligned}$$

3.2 Local Asymptotic Minimax Regret Decisions

In this section, we formally present the main results of this paper. Let $Z = (Z_U, Z_L)^\top \in \mathbf{R}^2$ be a normal random vector that has the same distribution as the asymptotic distribution of $\sqrt{n}(\tilde{\theta}_B - \theta_B)$, where $\tilde{\theta}_B \equiv (\tilde{\theta}_U, \tilde{\theta}_L)^\top$ is a semiparametrically efficient estimator of $\theta_{B,n}$ (without imposing the inequality restriction $\theta_{L,n} \leq \theta_{U,n}$ and without imposing any interval length restriction such that $\theta_U = \theta_L + c$). Let Σ be a 2×2 matrix such that for each $b \in \mathbf{R}^2$,

$$b^\top Z \sim N(0, b^\top \Sigma b). \tag{7}$$

Here we take $N(0, b^\top \Sigma b)$ to be a point mass at zero when $b^\top \Sigma b = 0$. Note that we do not require that Σ be invertible. A formal definition of Σ is given in Appendix A. We write

$$\Sigma = \left[\begin{array}{cc} \sigma_U^2 & \sigma_{L,U} \\ \sigma_{L,U} & \sigma_L^2 \end{array}\right].$$

In many cases, the matrix Σ can be found using the method of projection in the L_2 space (e.g. Bickel, Klaassen, Ritov and Wellner (1993)). We also define

$$\sigma_\Delta^2 \equiv \sigma_U^2 - 2\sigma_{L,U} + \sigma_L^2 \text{ and } Z^\Delta \equiv Z_U - Z_L.$$

Thus σ_Δ^2 is the variance of Z^Δ.

Theorem 1 below establishes a lower bound for the local asymptotic minimax regret, and Theorem 2, a result that the mid-point decision $\tilde{d}_{1/2} \equiv (\tilde{\theta}_U + \tilde{\theta}_L)/2$ is local asymptotic minimax regret among the sequences of decisions in $\mathcal{D}_n$. For a given decision $\hat{d}$, and $\varepsilon > 0$, let

$$\tilde{\mathcal{R}}_n^\varepsilon(\hat{d}) \equiv \sup_{h \in H_n^\varepsilon} \left\{\rho_h(\hat{d}) - \inf_{\hat{d} \in \mathcal{D}_n} \rho_h(\hat{d})\right\},$$

where $H_n^\varepsilon \equiv \{h \in H : \Delta_n(h) \geq -\varepsilon/\sqrt{n}\}$. The restriction of the supremum to H_n^ε represents our focus on the restriction $\theta_U \geq \theta_L$ (up to a negligible error).

THEOREM 1: *Suppose that Assumptions 1 and A1-A2 (in the appendix) hold. Suppose further that $\sigma_\Delta^2 > 0$ and $\alpha \in [1,2]$. Then for any sequence of estimators $\hat{d}$ in $\mathcal{D}_n$,*

$$\lim_{\varepsilon \downarrow 0} \liminf_{n \to \infty} \tilde{\mathcal{R}}_n^\varepsilon(\hat{d}) \geq \mathbf{E}\left[L\left(\frac{Z_U + Z_L}{2}\right)\right].$$

The condition $\sigma_\Delta^2 > 0$ in Theorem 1 is plausible for the case where the interval length is not a known constant. When $\alpha > 2$, it is not hard to see that the minimax regret becomes trivially infinity.

In Theorem 2 below, we ascertain that the bound in Theorem 1 is sharp. For technical facility, we follow a suggestion by Strasser (1985) (p.440) and consider instead

$$\tilde{\mathcal{R}}_n^\varepsilon(\hat{d}) \equiv \sup_{h \in H_n^\varepsilon} \left\{ \rho_{h,M}(\hat{d}) - \inf_{\hat{d} \in \mathcal{D}_n} \rho_{h,M}(\hat{d}) \right\}, \tag{8}$$

where $L_M(\cdot) \equiv \min\left\{L(\cdot), Mn^{-\alpha/2}\right\},\ M > 0$, and

$$\rho_{h,M}(\hat{d}) \equiv \sup_{\theta \in [\theta_{L,n}(h), \theta_{U,n}(h)]} n^{\alpha/2} \mathbf{E}_h\left[L_M(\hat{d} - \theta)\right].$$

As for $\tilde{\theta}_B = (\tilde{\theta}_U, \tilde{\theta}_L)^\top$, we make the following assumption.

ASSUMPTION 2: $\sup_{h \in H} |P_{n,h}\{\sqrt{n}(\tilde{\theta}_B - \theta_{B,n}(h)) \leq t\} - P\{Z \leq t\}| \to 0$ for each $t \in \mathbf{R}^2$.

The uniform convergence of distributions can often be verified using the uniform central limit theorem. 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 $h \in H$. (See e.g. Theorem 3 of Chow and Teicher (2003), p. 322.)

THEOREM 2: *Suppose that the conditions of Theorem 1 hold. Furthermore, suppose that Assumption 2 holds. Then,*

$$\lim_{M \uparrow \infty} \lim_{\varepsilon \downarrow 0} \limsup_{n \to \infty} \tilde{\mathcal{R}}_{n,M}^\varepsilon(\tilde{d}_{1/2}) \leq \mathbf{E}\left[L\left(\frac{Z_U + Z_L}{2}\right)\right].$$

REMARKS 1: The results of Theorems 1 and 2 continue to hold even when $Var(Z_U + Z_L) = 0$. In this case, the minimax regret bound becomes zero. This case arises when one knows

$\theta_U + \theta_L$ although θ_U and θ_L are not known separately. In other words, one knows precisely the minimax decision $(\theta_U + \theta_L)/2$. Hence there is no regret for this decision.

2: Interestingly, the mid-point decision does not require using a boundary estimator $\hat{\theta}_B = (\hat{\theta}_U, \hat{\theta}_L)^\top$ that satisfies $\hat{\theta}_L \leq \hat{\theta}_U$ despite the known restriction that $\theta_L \leq \theta_U$.

3: Suppose that one mistakenly assumes point identification (i.e., $\theta_U = \theta_L = \theta_0$) and uses its efficient estimator $\tilde{d}_0 \equiv \hat{\tau}^* \tilde{\theta}_U + (1 - \hat{\tau}^*)\tilde{\theta}_L$ under the overidentifying restrictions $\theta_U = \theta_L$, where $\hat{\tau}^*$ is a consistent estimator of $\tau^* \equiv -\sigma_{L,\Delta}/\sigma_\Delta^2$ and $\sigma_{L,\Delta} = \sigma_{L,U} - \sigma_L^2$. Unless $\sigma_U^2 = \sigma_L^2$, this estimator is not local asymptotic minimax regret in general. In fact, when $\sigma_U^2 \neq \sigma_L^2$, one can show that the local asymptotic maximal regret of $\tilde{d}_0$ is infinity!

Example 1 (Cont'd) : Let (Y, W) be as in Example 1, and define p, μ and θ similarly, so that the identified interval for θ is $[\theta_L, \theta_U]$, where $\theta_L = \mu p$ and $\theta_U = \mu p + 1 - p$. Now, suppose that the econometrician observes a random sample $\{(Y_i W_i, W_i)\}_{i=1}^n$ from the distribution of $(Y \cdot W, W)$, and p is unknown, and there is no known *nontrivial* upper bound for $1-p$. (Note that the obvious bound 1 is a *trivial* bound for $1 - p$, because the bound holds vacuously, i.e., the bound does not induce non-vacuous restrictions on the space H.) When one takes $\tilde{\theta}_U = \bar{\mu}\hat{p} + 1 - \hat{p}$ and $\tilde{\theta}_L = \bar{\mu}\hat{p}$ as efficient estimators of bounds, where $\bar{\mu} = \sum_{i=1}^n Y_i W_i / \sum_{i=1}^n W_i$, $\hat{p} = \frac{1}{n}\sum_{i=1}^n W_i$, the mid-point decision given by

$$\tilde{d}_{1/2} = \frac{\tilde{\theta}_U + \tilde{\theta}_L}{2} = \frac{1}{n}\sum_{i=1}^n \left(Y_i - \frac{1}{2}\right) W_i + \frac{1}{2}$$

is local asymptotic minimax regret.

4 Simulations

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 object 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:

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

where $Y_{L,i} \sim N(0,1)$, $Y_{U,i} \sim N(0,1)$, and $Z_i \sim \text{Uniform}[-1/2, 1/2]$. The scale parameter (a_U, a_L) was chosen from $\{(4,1), (3,2), (2,3)\}$, and $w = 0.8$. The mean vectors θ_L and θ_U were chosen to be $-\Delta/2$ and $\Delta/2$, where Δ denotes the interval length. The sample sizes were taken to be from $\{300, 1000\}$.

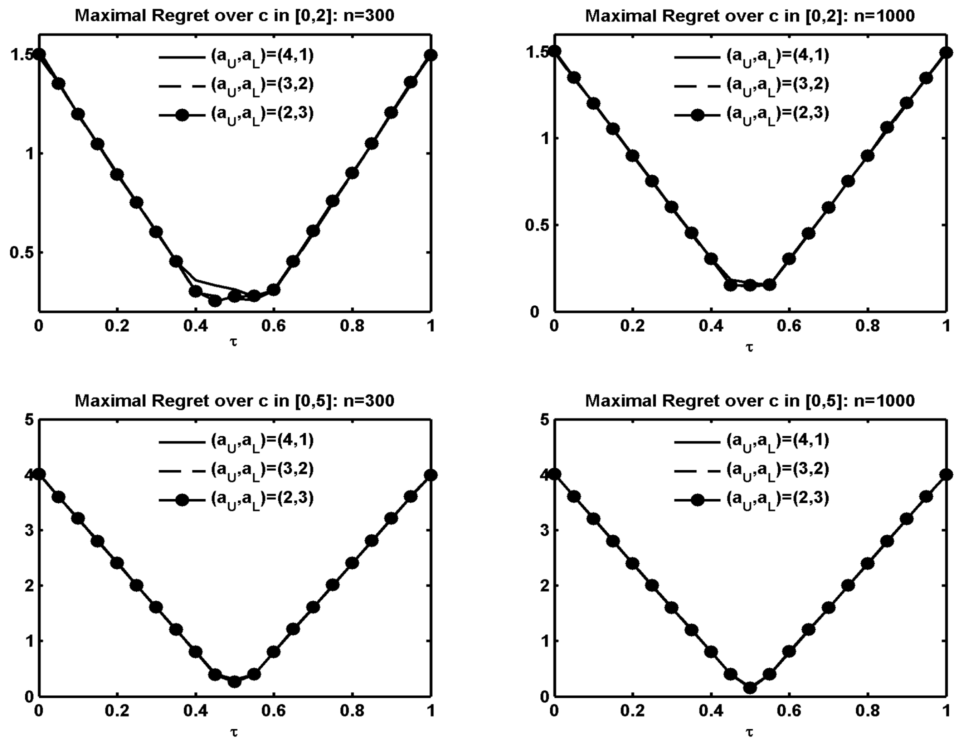

Figure 1: The four panels plot maximal regret of a weighted average decision $\tilde{\delta}_\tau = \tau\tilde{\theta}_U + (1-\tau)\tilde{\theta}_L$ against $\tau \in [0,1]$. The maximal regret over c in [0,2] takes the supremum of the regret over the interval lengths from 0 to 2. Hence when c is small (relative to the sample size), the situation is closer to point-identification. In this case, the mid-point decision is not necessarily an optimal decision and one is better off by taking into account the variance discrepancies of the bound estimators. However, when the interval lengths can be potentially large (as in the case of $c = 5$ for example), the maximal regret is minimized conspicuously at $\tau = 1/2$, showing that the mid-point decision is optimal.

We investigate the finite sample optimality of the mid-point decision $\tilde{d}_{1/2}$. The finite sample maximal regret of a decision $\hat{d}$ is taken to be the maximum of

$$\sup_{s\in[0,1]} \mathbf{E}\left[\left|\hat{d} - (\theta_L + s\Delta)\right|\right] - \left|\frac{\Delta}{2}\right|$$

over $\Delta \in [0, c]$, where $\Delta = \theta_U - \theta_L$. We considered $c \in \{2, 5\}$. When c is large, the domain of the maximum in the maximal regret becomes large, increasing the maximal regret. When c is small, the situation is closer to the case where the parameter is point-identified. We allow for discrepancies in the variances of the upper and lower bound estimators $\tilde{\theta}_U$ and $\tilde{\theta}_L$. For this, we have considered three pairs of $(a_U, a_L) \in \{(4,1), (3,2), (2,3)\}$. The decisions under consideration are of the form:

$$\tilde{d}_\tau = \tau\tilde{\theta}_U + (1-\tau)\tilde{\theta}_L$$

with the weight τ running in the interval $[0,1]$.

In Figure 1, the maximal regret for decisions $\tilde{d}_\tau$ is plotted against $\tau \in [0,1]$. When c is small, the mid-point decision is not necessarily an optimal decision, as shown in the figure where the maximal regret is not necessarily lowest at $\tau = 1/2$. In this case, the variance discrepancies play a role. However, as c becomes larger (so that the model is farther from point-identification), the mid-point decision emerges as the unique optimal decision in terms of the maximal regret, regardless of the variance discrepancies between the two bound estimators.

5 Conclusion

This paper investigates the problem of making a point-decision for an interval-identified object, when the interval length is not known. This paper demonstrates that the mid-point of efficient upper and lower bound estimators is a reasonable point decision according to the minimax regret principle.

Various extensions from the results may be of interest. One extension is to accommodate the situation where the decision is binary and loss functions are asymmetric and the object of interest is interval-identified. Such a question is relevant in the context of treatment decisions. (Manski (2004), Tetenov (2007), and Hirano and Porter (2009).) Although this extension is very interesting, the extension does not appear obvious to the author, and seems to require a substantial development that warrants a separate paper.

References

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

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

[3] Blackwell, D. (1951): "Comparison of experiments," *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, Berkeley, p.93-102.

[4] Chow, Y. S. and H. Teicher (2003): *Probability Theory: Independence, Interchangeability, Martingales.* Third Edition, Springer-Verlag, New York.

[5] Fan, Y. and S. Park (2009): "Partial Identification of the Distribution of Treatment Effects and its Confidence Sets," (with Sang Soo Park), in Thomas B. Fomby and R. Carter Hill, eds, *Nonparametric Econometric Methods* (*Advances in Econometrics, Volume 25*), Emerald Group Publishing Limited, p.3-70.

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

[7] 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.

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

[9] Horowitz, J. and C. F. Manski (1998): "Censoring of outcomes and regressors due to survey nonresponse: identification and estimation using weights and imputations," *Journal of Econometrics*, 84, 37-58.

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

[11] Kitagawa, T. (2012): "Estimation and inference for set-identified parameters using posterior lower probability," Working Paper, University College London.

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

[13] 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.

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

[15] Manski, C. F. (2003): *Partial Identification of Probability Distributions*, Springer, New York.

[16] Manski, C. F. (2004): "Statistical treatment rules for heterogeneous populations," *Econometrica* 72, 1221-1246.

[17] Molinari, F. (2010): "Missing treatments," *Journal of Business and Economic Statistics*, 28, 82-95.

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

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

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

[21] Tetenov, A. (2007): "Statistical treatment choice based on asymmetric minmax regret criteria," Working Paper.

[22] Tetenov, A. (2009): "Measuring precision of statistical inference on partially identified parameters," Working Paper.

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

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

[25] van der Vaart, A. W. (1998): *Asymptotic Statistics*, Cambridge University Press, New York.

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

[27] Winter, B. B. (1975): "A portmanteau theorem for vague convergence," *Studia Scientiarum Mathematicarum Hungarica* 10, 247-253.

6 Appendix

6.1 Appendix A: Conditions for the Boundary Parameter

To save space, we follow the general formulation in van der Vaart and Wellner (1996: Section 3.11) and refer the reader to it for further details. Suppose that $\mathcal{P} = \{P_\alpha : \alpha \in \mathcal{A}\}$ is 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. Suppose that $Y_1, \cdots, Y_n$ are i.i.d. draws from $P_{\alpha_0} \in \mathcal{P}$ so that $X_n \equiv (Y_1, \cdots, Y_n)$ is a measurable map from a measurable space $(\Omega_n, \mathcal{F}_n)$ into another measurable space $(\mathcal{X}_n, \mathcal{G}_n)$, distributed as $P^n_{\alpha_0}$. Now let $\mathcal{P}(P_{\alpha_0})$ be the collection of maps $t \to P_{\alpha_t}$ such that for some $h \in L_2(P_{\alpha_0})$,

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

When this convergence holds, we say that P_{α_t} *converges in quadratic mean* to P_{α_0} and call $h \in L_2(P_{\alpha_0})$ a *score function* associated with this convergence. The set of all such h's is called a *tangent set* and denoted by $T(P_{\alpha_0})$. We assume that $T(P_{\alpha_0})$ 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 $h \in H$ and $n \in \mathbb{N}$, we consider a path of the form: $\alpha_{t(n,h)} = \alpha_0 + t(n,h)$, where $t(n,h) = \lambda_h / \sqrt{n}$, $\lambda_h \in \mathcal{A}$, and $P_{\alpha_{t(n,h)}}$ converges in quadratic mean to P_{α_0} as $n \to \infty$ having h as its associated score. We simply write $P_{n,h} = P^n_{\alpha_{t(n,h)}}$ 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 $\mathcal{E}_n = (\mathcal{X}_n, \mathcal{G}_n, P_{n,h}; h \in H)$ constitutes a sequence of statistical experiments for the boundary parameter (Blackwell (1951)). As for $\mathcal{E}_n$, we assume local asymptotic normality as follows.

Assumption A1: 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 $h \in H$, $\zeta_n(h) \rightsquigarrow \zeta(h)$ (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 weak convergence of measures. Local asymptotic normality of experiments 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 $\mathcal{E} = (\mathcal{X}, \mathcal{G}, P_h; h \in H)$, where P_h, $h \in H$, is a probability measure on a measurable space $(\mathcal{X}, \mathcal{G})$ such that $\log dP_h / dP_0 = \zeta(h) - \frac{1}{2} \langle h, h \rangle$. (Note that the asymptotic Gaussianity is concerned with the log-likelihood process of the potential distributions that identify θ_B. This does not mean that candidate decisions are constructed only based on asymptotically normal estimators of the boundary parameter θ_B.)

Assumption A2: There exists a continuous linear $\mathbf{R}^2$-valued map on H, $\dot{\theta}_B = (\dot{\theta}_U, \dot{\theta}_L)^\top$, such that for each $h \in H$,

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

as $n \to \infty$.

Assumption A2 says that $\theta_{B,n}(h)$ is *regular* in the sense of van der Vaart and Wellner (1996, Section 3.11). The map $\dot{\theta}_B$ is associated with the semiparametric efficiency bound of $\theta_{B,n}$ in the following way. For each $b \in \mathbf{R}^2$, $b^\top \dot{\theta}_B(\cdot)$ defines a continuous linear functional on H, and hence there exists $\dot{\theta}^*_{B,b} \in \bar{H}$ such that $b^\top \dot{\theta}_B(h) = \langle \dot{\theta}^*_{B,b}, h\rangle$, $h \in H$. Then for any $b \in \mathbf{R}^2$, $||\dot{\theta}^*_{B,b}||^2$ 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 (1991), p.180.) The map $\dot{\theta}^*_{B,b}$ is called an *efficient influence function* for $b^\top \theta_{B,n}$ in the literature. Let $\mathbf{e}_1 = [1,0]^\top$ and $\mathbf{e}_2 = [0,1]^\top$, and define

$$\dot{\theta}^*_U = \dot{\theta}^*_{B,\mathbf{e}_1} \text{ and } \dot{\theta}^*_L = \dot{\theta}^*_{B,\mathbf{e}_2}.$$

We also define

$$\sigma^2_U \equiv \langle \dot{\theta}^*_U, \dot{\theta}^*_U\rangle,\ \sigma^2_L \equiv \langle \dot{\theta}^*_L, \dot{\theta}^*_L\rangle,\ \sigma_{L,U} \equiv \langle \dot{\theta}^*_L, \dot{\theta}^*_U\rangle \tag{9}$$

and

$$\Sigma \equiv \begin{bmatrix} \sigma^2_U & \sigma_{L,U} \\ \sigma_{L,U} & \sigma^2_L \end{bmatrix}.$$

This Σ is the matrix that appears in (7).

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

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

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

$$H_R = \left\{ h \in H : (\dot{\theta}_U - \dot{\theta}_L)(h) \geq -\Delta_0 \right\}. \tag{10}$$

The tangent set H_R is a convex affine cone. When $\Delta_0 = \infty$, $H_R = H$, and hence in this case, the inequality restriction leaves the tangent set intact. When $\Delta_0 < \infty$ and we consider slices of the tangent space such that $(\dot{\theta}_U - \dot{\theta}_L)(h) = r - \Delta_0$, $r \in [0, \infty)$, the analysis also remains the same regardless of whether we use H_R or H. This technique is used in the proof of Theorem 1. (See also Hirano and Porter (2009) for their use of this method.)

6.2 Appendix B: Mathematical Proofs

For any $A \subset \mathbf{R}^d$, let $vol_d(A)$ denote its Lebesgue measure on $\mathbf{R}^d$. A set $A \subset \mathbf{R}^d$ is called *centrally symmetric* if $A = -A$. 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.) When a set A is a centrally symmetric convex body, we say A is a *cscb*.

LEMMA A1: (i) *For any convex bodies A and B in $\mathbf{R}^d$, $d \geq 1$, $vol_d(A \cap (B+a))^{1/d}$ is concave in $a \in \{b \in \mathbf{R}^d : A \cap (B+b) \neq \varnothing\}$.*
(ii) *Let $a \in \mathbf{R}^d$ and $c \in \mathbf{R}$. Then the following two statements are equivalent.*
(a) *For some cscb's $A, B \subset \mathbf{R}^d$, $vol_d(A \cap (B+a)) \leq vol_d(A \cap (B+ca))$.*
(b) *For any cscb's $C, D \subset \mathbf{R}^d$, $vol_d(C \cap (D+a)) \leq vol_d(C \cap (D+ca))$.*

PROOF: (ii) For $a = 0$, the inequality becomes trivially an equality. Assume that $a \in \mathbf{R}^d \backslash \{0\}$. For $c \in \mathbf{R}$, let $H(c) = \{x \in \mathbf{R}^{d+1} : x_{d+1} = c\}$, where x_{d+1} denotes the $(d+1)$-th entry of $x \in \mathbf{R}^{d+1}$. For any convex body $A_1 \subset \mathbf{R}^{d+1}$, let

$$\begin{aligned} S(c; A_1) &\equiv \{(x_1, \cdots, x_d) \in \mathbf{R}^d : (x_1, \cdots, x_d, c) \in A_1 \cap H(c)\} \text{ and} \\ f(c; A_1) &\equiv vol_d(S(c; A_1)). \end{aligned}$$

Then $f(c; A_1)$ 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 $A_1 \in \mathbf{R}^{d+1}$ and $c \in \mathbf{R}$, $f(c; A_1) \leq f(0; A_1)$ by Lemma 38.20 of Strasser (1985) and $f(c; A_1) = f(-c; A_1)$. Therefore, for any cscb $A_1 \in \mathbf{R}^{d+1}$, $f(c_1; A_1) \leq f(c_2; A_1)$ if and only if $|c_1| \geq |c_2|$. Since the latter inequality does not involve A_1, this statement implies that

$$\begin{aligned} \exists \text{ cscb } A_1 &\subset \mathbf{R}^{d+1} \text{ s.t. } f(c_1; A_1) \leq f(c_2; A_1) \\ &\quad \text{if and only if} \\ \forall \text{ cscb } C_1 &\subset \mathbf{R}^{d+1} \text{ s.t. } f(c_1; C_1) \leq f(c_2; C_1). \end{aligned} \tag{11}$$

Choose cscb's A and B in $\mathbf{R}^d$ as in the lemma and let

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

where $m(c) = \max\{|c|, 1\}$. Then, $\bar{P}_{A,B} \equiv \bar{A} \cap \bar{B}$ is a cscb in $\mathbf{R}^{d+1}$. The statement in (a) is equivalent to $f(1; \bar{P}_{A,B}) \leq f(c; \bar{P}_{A,B})$, because

$$f(c; \bar{P}_{A,B}) = vol_d(A \cap (B + ca)).$$

The inequality holds if and only if $f(1; \bar{P}_{C,D}) \leq f(c; \bar{P}_{C,D})$ for all cscb's $C, D \subset \mathbf{R}^d$ by (11), which is equivalent to (b). ■

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

Then $g(\cdot)$ is quasiconcave if $\varphi_t(\cdot)$ is quasiconcave for all $t \in \mathbf{T}$, and $g(\cdot)$ is quasiconvex if $\varphi_t(\cdot)$ is quasiconvex for all $t \in \mathbf{T}$.

Proof: Straightforward. ■

Lemma A3: *Let $V \in \mathbf{R}$ be a continuous random variable with a quasiconcave density function that is symmetric around $b \in \mathbf{R}$, and let $\Lambda : \mathbf{R} \to [0, \infty)$ be symmetric around zero, and let $K \equiv \{\delta \in \mathbf{R} : \mathbf{E}[\Lambda(V + \delta)] < \infty\}$.*

Then $\mathbf{E}[\Lambda(V + \delta)]$ is quasiconvex in $\delta \in K$ if Λ is quasiconvex, and $\mathbf{E}[\Lambda(V + \delta)]$ is quasiconcave in $\delta \in K$ if Λ is quasiconcave.

Proof: We focus on the case where Λ is quasiconvex. The case with quasiconcave Λ can be dealt with similarly by considering $-\Lambda$. Write

$$\mathbf{E}[\Lambda(V + \delta)] = \int_0^\infty P\{V + \delta \in \mathbf{R} \backslash A(t)\} dt,$$

where $A(t) \equiv \{z \in \mathbf{R} : \Lambda(z) \leq t\}$. Note that

$$P\{V + \delta \in A(t)\} = \int_0^\infty \eta(\delta; t, e) de,$$

where $\eta(\delta; t, e) \equiv vol_1(A(t) \cap cl(\{z \in \mathbf{R} : f(z) > e\} + b + \delta))$, f is the density of $V - b$, and for any set B, $cl(B)$ is the closure of B. Since f is quasiconcave and Λ is quasiconvex, $A(t)$ and $cl\{z \in \mathbf{R} : f(z) > e\}$ are closed intervals with their centers at zero, for all e and t in $\mathbf{R}$. By Lemma A1(i), $\eta(\cdot; t, e)$ is quasiconcave over $J(t, e)$ for all e and t in $\mathbf{R}$, where

$$J(t, e) \equiv \{\delta \in \mathbf{R} : A(t) \cap cl(\{z \in \mathbf{R} : f(z) > e\} + b + \delta) \neq \varnothing\}.$$

Since $A(t)$ and $cl(\{z \in \mathbf{R} : f(z) > e\} + b)$ are intervals, $J(t, e)$ is also an interval. Since

$\eta(\cdot;t,e) \geq 0$ and $\eta(\delta;t,e) = 0$ for all $\delta \in \mathbf{R} \backslash J(t,e)$, we conclude that $\eta(\cdot;t,e)$ is quasiconcave over $\mathbf{R}$ for all e and t in $\mathbf{R}$. Since $A(t)$ and $cl\{z \in \mathbf{R} : f(z) > e\}$ are closed intervals centered at zero for all t and e, we apply Lemma A1(ii) to find that when $\delta_1 \in \mathbf{R}$ and $\delta_2 \in \mathbf{R}$ are given, $\eta(\delta_1;t,e) \leq \eta(\delta_2;t,e)$ for some t and e if and only if $\eta(\delta_1;t,e) \leq \eta(\delta_2;t,e)$ for all t and e. That is, the ordering of $\{\eta(\delta;t,e) : \delta \in \mathbf{R}\}$ remains the same as we shift t and e. Hence by Lemma A2, $P\{V + \delta \in A(t)\}$ is quasiconcave in $\delta \in \mathbf{R}$ for each $t \in \mathbf{R}$.

Take any $\delta_1, \delta_2 \in K$ such that $P\{V + \delta_1 \in A(t)\} \geq P\{V + \delta_2 \in A(t)\}$ for some $t \in \mathbf{R}$. Then, since $A(t)$ is a closed interval centered at zero for all $t \in \mathbf{R}$ and δ_1 and δ_2 are real numbers, the preceding inequality holds if and only if for all $t' \in (0,\infty)$, $P\{V + \delta_1 \in A(t')\} \geq P\{V + \delta_2 \in A(t')\}$, by Lemma A1(ii) again. Therefore,

$$\mathbf{E}[\Lambda(V+\delta)] = \int_0^\infty (1 - P\{V + \delta \in A(t)\})\, dt$$

is quasiconvex in $\delta \in K$ by Lemma A2. ■

Lemma A4: *Let $V \in \mathbf{R}$ be a continuous random variable with a quasiconcave density function that is symmetric around b and let $\Lambda : \mathbf{R} \to [0,\infty]$ be convex and symmetric around zero.*

Then, for $s \geq 0$,

$$\mathbf{E}[\Lambda(V - b - s/2)] \leq \inf_{c \in \mathbf{R}} \sup_{\delta \in [0,s]} \mathbf{E}[\Lambda(V + c + \delta)]. \tag{12}$$

Proof: Since Λ is quasiconvex and symmetric around zero, by Lemma A3,

$$\inf_{c \in \mathbf{R}} \sup_{\delta \in [0,s]} \mathbf{E}[\Lambda(V + c + \delta)] = \inf_{c \in \mathbf{R}} \max_{\delta \in \{0,s\}} \mathbf{E}[\Lambda(V + c + \delta)].$$

Let $\tilde{\Lambda}(d) \equiv \{\Lambda(d+s) + \Lambda(d)\}/2$. It suffices to show that

$$\mathbf{E}\left[\tilde{\Lambda}\left(V - b - \frac{s}{2}\right)\right] \leq \mathbf{E}\left[\tilde{\Lambda}(V + c)\right], \text{ for all } c \in \mathbf{R}. \tag{13}$$

This is because the left hand side of (13) is equal to the left hand side of (12) (due to the fact that the distribution of $\Lambda(V - b - s/2)$ and that of $\Lambda(V - b + s/2)$ are identical) and because

$$\begin{aligned} \mathbf{E}\left[\tilde{\Lambda}(V + c)\right] &= \frac{1}{2}\{\mathbf{E}[\Lambda(V + c)] + \mathbf{E}[\Lambda(V + c + s)]\} \\ &\leq \max_{\delta \in \{0,s\}} \mathbf{E}[\Lambda(V + c + \delta)], \end{aligned}$$

for all $c \in \mathbf{R}$.

We now show (13). Let $A(t) \equiv \{z \in \mathbf{R} : \tilde{\Lambda}(z) \leq t\}$ and observe that

$$\mathbf{E}[\tilde{\Lambda}(V+c)] = \int_0^\infty P\left\{V + c \in \mathbf{R} \backslash A(t)\right\} dt. \tag{14}$$

Let f be the density function of $V - b$ and write $P\left\{V + c \in A(t)\right\}$ as

$$\begin{aligned} & \int_0^\infty vol_1\left(A(t) \cap cl(\{z \in \mathbf{R} : f(z) > e\} + b + c)\right) de \\ = & \int_0^\infty vol_1((A(t) + s/2) \cap cl(\{z \in \mathbf{R} : f(z) > e\} + b + c + s/2))de. \end{aligned}$$

Since $\tilde{\Lambda}$ is symmetric around $-s/2$ and convex, $A(t) + s/2$ is convex and symmetric around zero. By Lemma 38.20 of Strasser (1985),

$$\begin{aligned} & vol_1\left((A(t) + s/2) \cap cl(\{z : f(z) > e\} + b + c + s/2)\right) \\ \leq & \ vol_1((A(t) + s/2) \cap cl\{z : f(z) > e\}). \end{aligned}$$

The inequality becomes equality when $b + c + s/2 = 0$ or $c = -b - s/2$. This implies that

$$P\left\{V + c \in A(t)\right\} \leq P\left\{V - b - s/2 \in A(t)\right\}. \tag{15}$$

In view of (14), this proves the result in (13). ■

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

$$\bar{\theta} \equiv \begin{bmatrix} \dot{\theta}_U(h_1) & \dot{\theta}_L(h_1) \\ \vdots & \vdots \\ \dot{\theta}_U(h_m) & \dot{\theta}_L(h_m) \end{bmatrix} \text{ and } \bar{\Delta}_\theta \equiv \begin{bmatrix} (\dot{\theta}_U - \dot{\theta}_L)(h_1) \\ \vdots \\ (\dot{\theta}_U - \dot{\theta}_L)(h_m) \end{bmatrix}, \tag{16}$$

and $\bar{\zeta} \equiv (\zeta(h_1), \cdots, \zeta(h_m))^\top$, where ζ is the Gaussian process that appears in Assumption A1. We assume that $m \geq 2$ and $\bar{\theta}$ is full column rank. We fix $\lambda > 0$, $q \in \mathbf{R}$, and let $A_\lambda \in \mathbf{R}^m \sim N(0, I/\lambda)$ and let $F_{\lambda,q}(a)$ be the cdf of $\bar{\Sigma} A_\lambda + \bar{\mu}_q$, where

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

Then, it is easy to check that for all realizations of A_λ,

$$(\dot{\theta}_U - \dot{\theta}_L)((\bar{\Sigma}A_\lambda + \bar{\mu}_q)^\top h_B) = q, \tag{17}$$

where $h_B = (h_1, \cdots, h_m)^\top$. Suppose that $\hat{\theta}_B$ is a sequence of estimators such that for each $h \in H$ and $V_{n,h} \equiv \sqrt{n}\{\hat{\theta}_B - \theta_{B,n}(h)\}$,

$$\begin{bmatrix} V_{n,h} \\ \log dP_{n,h}/dP_{n,0} \end{bmatrix} \to_V \begin{bmatrix} \mathcal{L}^h \\ \zeta(h) - \frac{1}{2}\langle h, h\rangle \end{bmatrix}, \tag{18}$$

where $\to_V$ denotes vague convergence and $\mathcal{L}^h$ is a potentially deficient distribution. Finally let $\bar{\Sigma}_\lambda \equiv \bar{\Sigma} + \lambda I$ and let $Z_{\lambda,q,m} \in \mathbf{R}^2$ and $Z_{0,q,m} \in \mathbf{R}^2$ be normal random vectors such that

$$\begin{aligned} Z_{\lambda,q,m} &\sim N(\bar{\theta}^\top(I - \bar{\Sigma}\bar{\Sigma}_\lambda^{-1}\bar{\Sigma})\bar{\mu}_q, \bar{\theta}^\top\bar{\Sigma}\bar{\Sigma}_\lambda^{-1}\bar{\Sigma}\bar{\theta}) \text{ and} \\ Z_{0,q,m} &\sim N(\bar{\theta}^\top\bar{\mu}_q, \bar{\theta}^\top\bar{\Sigma}\bar{\theta}). \end{aligned}$$

Let $\mathcal{Z}_{\lambda,q,m}$ denote the distribution of $Z_{\lambda,q,m}$. The following result is a conditional version of the convolution theorem in Theorem 2.2 of van der Vaart (1989). For any positive integer d, $\mathcal{B}(\mathbf{R}^d)$ denotes the Borel σ-field of $\mathbf{R}^d$ and for two measures F and G on $\mathcal{B}(\mathbf{R}^d)$, $F * G$ denotes the convolution of F and G.

Lemma A5: *Suppose that $\sigma_\Delta^2 > 0$. Then for any $\lambda > 0$ and $q \in \mathbf{R}$,*

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

where $\mathcal{M}_{\lambda,q,m}$ denotes a potentially deficient distribution on $\mathcal{B}(\mathbf{R}^2)$.

Furthermore, as first $\lambda \to 0$ and then $m \to \infty$, $Z_{\lambda,q,m}$ weakly converges to the conditional distribution of Z given $Z_U - Z_L = q$.[3]

Proof: Let $\bar{\mathbf{R}} = [-\infty, \infty]$ be the usual two-point compactification of $\mathbf{R}$ and $\bar{\mathbf{R}}^2$ the product of its two copies. By (18), along $\{P_{n,0}\}$,

$$\left(V_{n,h(a)},\ \log\frac{dP_{n,h(a)}}{dP_{n,0}}\right) \to_V \left(Z_0 - \bar{\theta}^\top a,\ \bar{\zeta}^\top a - \frac{1}{2}||a||^2\right),$$

where Z_0 is a random vector distributed as $\mathcal{L}^0$, and $||a||^2 = a^\top a$. By Le Cam's third lemma,

[3]The conditional distribution of Z^L given $Z^\Delta = 0$ is taken to be a point mass at zero when its variance is zero. This case arises when $\sigma_L^2 - \sigma_{L,\Delta}^2/\sigma_\Delta^2 = 0$.

we find that for all $B \in \mathcal{B}(\bar{\mathbf{R}}^2)$,

$$\mathcal{L}^{h(a)}(B) = \int \mathbf{E}[1_B\{v - \bar{\theta}^\top a\}e^{\bar{\zeta}^\top a - \frac{1}{2}||a||^2}]d\mathcal{L}^0(v).$$

Hence $\int \mathcal{L}^{h(a)}dF_{\lambda,q}(a)(B)$ is equal to

$$\int\int \mathbf{E}\left[1_B\left\{v - \bar{\theta}^\top \bar{\Sigma}_q(a)\right\} e^{\bar{\Sigma}_q(a)^\top \bar{\zeta} - \frac{1}{2}\bar{\Sigma}_q(a)^\top \bar{\Sigma}_q(a) - \frac{\lambda||a||^2}{2}}\right]\left(\frac{\lambda}{2\pi}\right)^{\frac{m}{2}} d\mathcal{L}^0(v)da,$$

where $\bar{\Sigma}_q(a) \equiv \bar{\Sigma}a + \bar{\mu}_q$. Using the fact that $\bar{\Sigma}$ is idempotent and going through some tedious calculations, we write $\int \mathcal{L}^{h(a)}dF_{\lambda,q}(a)(B)$ as

$$\int\int \mathbf{E}\left[1_B\left\{v - J_{\lambda,q}(a) - \bar{\theta}^\top \bar{\Sigma}\bar{\Sigma}_\lambda^{-1}\bar{\Sigma}\bar{\zeta}\right\} c_\lambda(\bar{\zeta})\right] d\mathcal{L}^0(v)dN(a),$$

where $J_{\lambda,q}(a) \equiv \bar{\theta}^\top \bar{\Sigma}\bar{\Sigma}_\lambda^{-1/2}a + \bar{\theta}^\top(I - \bar{\Sigma}\bar{\Sigma}_\lambda^{-1}\bar{\Sigma})\bar{\mu}_q$, $N(\cdot)$ is the cdf of $N(0, I)$, and

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

For any $B \in \mathcal{B}(\mathbf{R}^2)$, $\int_B J_{\lambda,q}(a)dN(a) = P\{Z_{\lambda,q,m} \in B\}$. Letting $\mathcal{M}_{\lambda,q,m}$ be a measure such that

$$\mathcal{M}_{\lambda,q,m}(B) = \int \mathbf{E}\left[1_B\left\{v - \bar{\theta}^\top \bar{\Sigma}\bar{\Sigma}_\lambda^{-1}\bar{\Sigma}\bar{\zeta}\right\} c_\lambda(\bar{\zeta})\right] d\mathcal{L}^0(v), \text{ for all } B \in \mathcal{B}(\bar{\mathbf{R}}^2),$$

we obtain the desired result.

As for the second statement, note that as $\lambda \to 0$, $Z_{\lambda,q,m} \to_d N(\bar{\theta}^\top(I - \bar{\Sigma})\bar{\mu}_q, \bar{\theta}^\top\bar{\Sigma}\bar{\theta})$. Note that $\bar{\Sigma}\bar{\mu}_q = 0$. Hence as $\lambda \to 0$, $Z_{\lambda,q,m} \to_d Z_{0,q,m}$. Since $\{h_i\}_{i=1}^\infty$ is an orthonormal basis for a complete Hilbert space, as $m \to \infty$, the Euclidean distance between $\bar{\theta}^\top\bar{\mu}_q$ and $[(\sigma_{U,\Delta}/\sigma_\Delta^2)q, (\sigma_{L,\Delta}/\sigma_\Delta^2)q]^\top$ and that between

$$\bar{\theta}^\top\bar{\Sigma}\bar{\theta} \text{ and } \begin{bmatrix} \sigma_U^2 - \sigma_{U,\Delta}^2/\sigma_\Delta^2 & \sigma_{U,L} - \sigma_{U,\Delta}\sigma_{L,\Delta}/\sigma_\Delta^2 \\ \sigma_{U,L} - \sigma_{U,\Delta}\sigma_{L,\Delta}/\sigma_\Delta^2 & \sigma_L^2 - \sigma_{L,\Delta}^2/\sigma_\Delta^2 \end{bmatrix}$$

become zero, where $\sigma_{U,\Delta} \equiv \sigma_U^2 - \sigma_{L,U}$ and $\sigma_{L,\Delta} \equiv \sigma_{L,U} - \sigma_L^2$. Hence as $m \to \infty$, the distribution of $Z_{0,q,m}$ converges to the conditional distribution of Z given $Z_U - Z_L = q$. ■

Lemma A6: *Let $V \in \mathbf{R}$ be a continuous random variable that has a density function symmetric around zero. Let $\varphi_\alpha(t) \equiv \mathbf{E}L_\alpha(t + V) - L_\alpha(t)$, for $L_\alpha(t) = |t|^\alpha$, $\alpha \in [1, 2]$. Then $\varphi_\alpha(\cdot)$ is quasiconcave and symmetric around zero.*

PROOF: The symmetry around zero is obvious. Let L'_α be the first order derivative of L_α when $\alpha \in (1,2]$, and define $L'_\alpha(t) = 1\{t > 0\} - 1\{t < 0\}$ when $\alpha = 1$. Note that for all $v \geq 0$,

$$\begin{aligned} L'_\alpha(t+v) - L'_\alpha(t) &\leq L'_\alpha(t) - L'_\alpha(t-v) \text{ if } t \geq 0 \text{ and} \\ L'_\alpha(t+v) - L'_\alpha(t) &\geq L'_\alpha(t) - L'_\alpha(t-v) \text{ if } t \leq 0. \end{aligned} \tag{19}$$

Let f be the density of V. Splitting the absolute values and using Leibnitz's rule and symmetry of f, we find that

$$\varphi'_\alpha(t) = \frac{1}{2}\int_{-\infty}^{\infty} \{L'_\alpha(t+v) + L'_\alpha(t-v) - 2L'_\alpha(t)\} f(v)dv.$$

From (19), if $t \geq 0$, $\varphi'_\alpha(t) \leq 0$ and if $t < 0$, $\varphi'_\alpha(t) \geq 0$, completing the proof. ■

PROOF OF THEOREM 1: Suppose that $\sigma^2_\Delta > 0$. Since $\hat{d}$ can be viewed as an arbitrary measurable map from Ω_n into $\mathbf{R}$, we lose no generality by writing

$$\hat{d} = \tau^* \bar{\theta}_{U,n} + (1-\tau^*)\bar{\theta}_{L,n}, \tag{20}$$

where $\bar{\theta}_{U,n}$ and $\bar{\theta}_{L,n}$ are any measurable maps: $\Omega_n \to \mathbf{R}$ and $\tau^* \equiv -\sigma_{L,\Delta}/\sigma^2_\Delta$. Let

$$\begin{aligned} V_{n,h} &\equiv \sqrt{n}\{\bar{\theta}_{B,n} - \theta_{B,n}(h)\} \text{ and} \\ S_\tau(v) &\equiv \tau v_U + (1-\tau)v_L, \text{ for } v = (v_U, v_L)^\top \in \mathbf{R}^2 \text{ and } \tau \in \mathbf{R}, \end{aligned} \tag{21}$$

where $\bar{\theta}_{B,n} \equiv (\bar{\theta}_{U,n}, \bar{\theta}_{L,n})^\top$. Let $\tilde{L}_M(\cdot) \equiv \min\{L(\cdot), M\}$, $M > 0$, so that $n^{\alpha/2}L_M(v) = \tilde{L}_M(\sqrt{n}v)$ for all $v \in \mathbf{R}^2$.

Take $0 < J_M \to \infty$ as $M \uparrow \infty$, and let

$$L_M^{(J)}(v) \equiv \frac{1}{2^{J_M}} \sum_{j=0}^{2^{J_M}M} 1\left\{\tilde{L}_M(v) > j2^{-J_M}\right\}$$

and $C_j \equiv \{v \in \mathbf{R} : \tilde{L}_M(v) \in ((j-1)2^{-J_M}, j2^{-J_M}]\}$. For all $j = 0, \cdots, 2^{J_M}M$ and all $v \in C_j$, $L_M^{(J)}(v) = (j-1)2^{-J_M}$. The set $\cup_{j=2^{J_M}M+1}^{\infty} C_j$ is empty because when $v \in \cup_{j=2^{J_M}M+1}^{\infty} C_j$, it means that $\tilde{L}_M(v) > M$, which contradicts the fact that $\tilde{L}_M(\cdot) \leq M$. Therefore,

$$\sup_{v\in\mathbf{R}} \left| L_M^{(J)}(v) - \tilde{L}_M(v) \right| = \max_{0\leq j\leq 2^{J_M}M} \sup_{v\in C_j} \left| L_M^{(J)}(v) - \tilde{L}_M(v) \right| \leq 2^{-J_M}. \tag{22}$$

Also note that $L_M^{(J)}(\cdot) \leq \tilde{L}_M(\cdot)$ and as $M \uparrow \infty$, $L_M^{(J)}(\cdot)$ increases.

For each $r \in [0, \infty)$ and for each $\varepsilon > 0$, let

$$H_n^\varepsilon(r) \equiv \{h \in H : r - \varepsilon \leq \sqrt{n}\Delta_n(h) \leq r + \varepsilon\}. \tag{23}$$

For each $r \in [0, \infty)$, the set $\{P_{n,h} : h \in H_n^\varepsilon(r)\}$ collects probabilities $P_{n,h}$ such that the length of the identified interval is approximately $r/\sqrt{n}$. Also, $H(r) \equiv \{h \in H : \Delta_h = r\}$, where we recall $\Delta_h \equiv \dot{\Delta}(h) + \Delta_0$. Now we write $\tilde{\mathcal{R}}_n^\varepsilon(\hat{d})$ as

$$\begin{aligned}
&\sup_{h \in H_n^\varepsilon} \sup_{s \in [0,1]} \mathbf{E}\left(L(S_{\tau^*}(V_{n,h}) + (\tau^* - s)\sqrt{n}\Delta_n(h)) - L\left(\frac{\sqrt{n}\Delta_n(h)}{2}\right)\right) \qquad (24)\\
\geq\ &\sup_{h \in H_n^\varepsilon} \sup_{s \in [0,1]} \mathbf{E}\left(L_M^{(J)}(S_{\tau^*}(V_{n,h}) + (\tau^* - s)\sqrt{n}\Delta_n(h)) - L\left(\frac{\sqrt{n}\Delta_n(h)}{2}\right)\right)\\
\geq\ &\sup_{r \in [0,\infty)} \sup_{h \in H(r)} \sup_{s \in [0,1]} \mathbf{E}\left(L_M^{(J)}(S_{\tau^*}(V_{n,h}) + (\tau^* - s)r) - L\left(\frac{r + \varepsilon}{2}\right)\right) - \varepsilon M - 2^{-J_M},
\end{aligned}$$

from some large n on.

By Helly's Lemma (e.g. Lemma 2.5 of van der Vaart (1998) on page 9), every subsequence of $V_{n,h}$ has a further subsequence that vaguely converges to $\mathcal{L}^h$ and $\sqrt{n}\Delta_n(h) \to r$ as $n \to \infty$ for each $h \in H(r)$. By Assumption A1, without loss of generality, we pick any subsequence $V_{n',h}$ such that along $\{P_{n,0}\}$,

$$\begin{bmatrix} V_{n',h} \\ \log dP_{n',h}/dP_{n',0} \end{bmatrix} \to_V \begin{bmatrix} \mathcal{L}^h \\ \zeta(h) - \frac{1}{2}\langle h, h\rangle \end{bmatrix},$$

(the vague limit $\mathcal{L}^h$ may depend on the choice of this subsequence), so that we bound the $\liminf_{n' \to \infty}$ of the last infimum in (24) from below by

$$\begin{aligned}
&\sup_{r \in [0,\infty)} \sup_{h \in H(r)} \sup_{s \in [0,1]} \liminf_{n' \to \infty} \mathbf{E}_h\left[L_M^{(J)}(S_{\tau^*}(V_{n',h}) + (\tau^* - s)r) - L\left(\frac{r + \varepsilon}{2}\right)\right]\\
\geq\ &\sup_{r \in [0,\infty)} \sup_{h \in H(r)} \sup_{s \in [0,1]} \int \left[L_M^{(J)}(S_{\tau^*}(v) + (\tau^* - s)r) - L\left(\frac{r + \varepsilon}{2}\right)\right] d\mathcal{L}^h(v),
\end{aligned}$$

by Theorem 3 of Winter (1975).

Since $\Delta_h = (\dot{\theta}_U - \dot{\theta}_L)(h) + \Delta_0$, it follows that $(\dot{\theta}_U - \dot{\theta}_L)(h) = r - \Delta_0$ if and only if $h \in H(r)$. 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 $h(a) \in H(r)$. Fix $\lambda > 0$ and let $F_{\lambda,q}(a)$ be as defined prior to Lemma A5 above with $q = r - \Delta_0$. Then the support of $F_{\lambda,q}(\cdot)$ is confined to the set of a's such that $h(a) \in H(r)$ from (17), so that

$$\sup_{h\in H(r)} \sup_{s\in[0,1]} \int \left[L_M^{(J)}(S_{\tau^*}(v) + (\tau^* - s)r) - L\left(\frac{r+\varepsilon}{2}\right)\right] d\mathcal{L}^h(v)$$
$$\geq \sup_{s\in[0,1]} \int\int \left[L_M^{(J)}(S_{\tau^*}(v) + (\tau^* - s)r) - L\left(\frac{r+\varepsilon}{2}\right)\right] d\mathcal{L}^{h(a)}(v) dF_{\lambda,q}(a).$$

By Lemma A5, the above double integral is equal to

$$\int \left[L_M^{(J)}(S_{\tau^*}(v) + (\tau^* - s)r) - L\left(\frac{r+\varepsilon}{2}\right)\right] d(\mathcal{Z}_{\lambda,q,m} * \mathcal{M}_{\lambda,q,m})(v)$$
$$= \int \mathbf{E}\left[L_M^{(J)}(S_{\tau^*}(Z_{\lambda,q,m} + w) + (\tau^* - s)r) - L\left(\frac{r+\varepsilon}{2}\right)\right] d\mathcal{M}_{\lambda,q,m}(w)$$
$$\geq \int \mathbf{E}\left[L_M(S_{\tau^*}(Z_{\lambda,q,m} + w) + (\tau^* - s)r) - L\left(\frac{r+\varepsilon}{2}\right)\right] d\mathcal{M}_{\lambda,q,m}(w) - 2^{-J_M},$$

where $\mathcal{M}_{\lambda,q,m}$ is a potentially deficient distribution as in Lemma A5.

We conclude that by sending $\varepsilon \downarrow 0$, the $\liminf_{n\to\infty}$ of $\tilde{\mathcal{R}}_n^{\varepsilon}(\hat{d})$ is bounded from below by

$$\sup_{r\in[0,\infty)} \sup_{h\in H(r)} \sup_{s\in[0,1]} \int \mathbf{E}\left[L_M(S_{\tau^*}(Z_{\lambda,q,m} + w) + (\tau^* - s)r) - L\left(\frac{r}{2}\right)\right] d\mathcal{M}_{\lambda,q,m}(w) - 2^{-J_M}.$$

By Helly's Lemma, we can find subsequences $\lambda_k \to 0$ and $m_j \to \infty$ such that the measure $\mathcal{M}_{\lambda_k,q,m_j}$ has a vague limit, say, $\mathcal{M}_q$. Along this subsequence, by therefore, the $\liminf_{m_j\to\infty}$ $\liminf_{\lambda_k\to\infty}$ of the above supremum is bounded from below by

$$\sup_{r\in[0,\infty)} \sup_{s\in[0,1]} \liminf_{m_j\to\infty} \liminf_{\lambda_k\to 0} \int \mathbf{E}\left[L_M\left(S_{\tau^*}(Z_{\lambda_k,q,m_j} + w) + (\tau^* - s)\, r\right) - L\left(\frac{r}{2}\right)\right] d\mathcal{M}_{\lambda_k,q,m_j}(w)$$
$$\geq \sup_{r\in[0,\infty)} \sup_{s\in[0,1]} \int \mathbf{E}\left[L_M\left(S_{\tau^*}(Z + w) + (\tau^* - s)\, r\right) - L\left(\frac{r}{2}\right) | Z^{\Delta} = r - \Delta_0\right] d\mathcal{M}_q(w).$$

We bound the $\lim_{M\uparrow\infty}$ of above supremum from below by

$$\sup_{r\in[0,\infty)} \sup_{s\in[0,1]} \lim_{M\uparrow\infty} \int \mathbf{E}\left[L_M\left(S_{\tau^*}(Z + w) + (\tau^* - s)\, r\right) - L\left(\frac{r}{2}\right) | Z^{\Delta} = q\right] d\mathcal{M}_q(w)$$
$$= \sup_{r\in[0,\infty)} \sup_{s\in[0,1]} \int \mathbf{E}\left[L\left(S_{\tau^*}(Z + w) + (\tau^* - s) r\right) - L\left(\frac{r}{2}\right) | Z^{\Delta} = q\right] d\mathcal{M}_q(w)$$
$$= \sup_{r\in[0,\infty)} \sup_{s\in[0,1]} \int \mathbf{E}\left[L\left(S_{\tau^*}(Z + w) + (\tau^* - s) r\right) - L\left(\frac{r}{2}\right)\right] d\mathcal{M}_q(w),$$

where the first equality uses the Monotone Convergence Theorem and the second equality uses the fact that $S_{\tau^*}(Z+w) = S_{\tau^*}(Z) + S_{\tau^*}(w)$ and $S_{\tau^*}(Z)$ is independent of Z^{Δ}. As noted in Section 3.1, when $\tau^* > 1/2$, the above supremum is infinity because we can take $s = 0$ and r as large as possible, and when $\tau^* < 1/2$, the above supremum is infinity because we can take $s = 1$ and r as large as possible. Therefore, the above supremum is bounded from below by

$$\begin{aligned}&\sup_{r\in[0,\infty)}\sup_{s\in[0,1]}\int \mathbf{E}\left[L\left(S_{1/2}(Z+w)+\left(\frac{1}{2}-s\right)r\right)-L\left(\frac{r}{2}\right)\right]d\mathcal{M}_q(w)\\ \geq\ &\sup_{r\in[0,\infty)}\mathbf{E}\left[L\left(S_{1/2}(Z)+\frac{r}{2}\right)-L\left(\frac{r}{2}\right)\right],\end{aligned}$$

where the last inequality uses the argument in the proof of (13). By Lemma A6, the above expectation is quasiconcave in r and symmetric around zero, and hence the supremum over $r \in [0,\infty)$ is achieved when $r = 0$, delivering the desired lower bound of $\mathbf{E}\left[L\left(S_{1/2}(Z)\right)\right]$. ■

Proof of Theorem 2: First suppose that

$$\sigma_U^2 + 2\sigma_{U,L} + \sigma_L^2 > 0. \tag{25}$$

We write $\tilde{\mathcal{R}}^{\varepsilon}_{n,M}(\tilde{d}_{1/2})$ as

$$\sup_{h\in H_n^{\varepsilon}}\sup_{s\in[0,1]}\mathbf{E}\left(L_M\left(S_{1/2}(Z_{n,h})+\left(\frac{1}{2}-s\right)\sqrt{n}\Delta_n(h)\right)-L\left(\frac{\sqrt{n}\Delta_n(h)}{2}\right)\right),$$

where $Z_{n,h} \equiv \sqrt{n}(\tilde{\theta}_B - \theta_{B,n}(h))$. For any $\Delta \in [-\varepsilon,\infty)$, $s \in [0,1]$, and $t \in \mathbf{R}$,

$$P\left\{S_{1/2}(Z_{n,h})+\left(\frac{1}{2}-s\right)\Delta\leq t\right\}\to P\left\{S_{1/2}(Z)+\left(\frac{1}{2}-s\right)\Delta\leq t\right\},\ \text{as } n\to\infty, \tag{26}$$

by Assumption 2. The convergence is in fact uniform over $(s,\Delta) \in [0,1]\times[-\varepsilon,\infty)$ because $\mathbf{E}[S_{1/2}^2(Z)] = \frac{1}{4}\mathbf{E}[(Z_U+Z_L)^2] > 0$ from (25), meaning that $S_{1/2}(Z)$ is a continuous random variable. Therefore, we conclude that

$$\begin{aligned}&\limsup_{n\to\infty}\sup_{h\in H_n^{\varepsilon}}\sup_{s\in[0,1]}\mathbf{E}\left(L_M\left(S_{1/2}(Z_{n,h})+\left(\frac{1}{2}-s\right)\sqrt{n}\Delta_n(h)\right)-L\left(\frac{\sqrt{n}\Delta_n(h)}{2}\right)\right)\\ \leq\ &\sup_{\Delta\in[-\varepsilon,\infty)}\sup_{s\in[0,1]}\mathbf{E}\left(L_M\left(S_{1/2}(Z)+\left(\frac{1}{2}-s\right)\Delta\right)-L\left(\frac{\Delta}{2}\right)\right).\end{aligned}$$

The above supremum is increasing in $\varepsilon > 0$ and increasing in M. By sending $\varepsilon \downarrow 0$ and then

$M \uparrow \infty$, the above supremum becomes $\sup_{\Delta \in [0,\infty)} \tilde{\beta}\left(-\Delta/2\right)$, where

$$\tilde{\beta}(z) = \mathbf{E}\left[L\left(\frac{Z_L + Z_U}{2} - z\right)\right] - L(z).$$

Certainly $\tilde{\beta}(z)$ is symmetric around zero, and quasiconcave by Lemma A6. Hence the supremum of $\tilde{\beta}(-\Delta/2)$ over $\Delta \in [0,\infty)$ achieved at $\Delta = 0$.

Now suppose that $\sigma_U^2 + 2\sigma_{U,L} + \sigma_L^2 = 0$. This means that $S_{1/2}(Z_{n,h}) \to_P 0$ uniformly in $h \in H$. Define

$$\zeta_n(s,\Delta) \equiv S_{1/2}(Z_{n,h}) + \left(\frac{1}{2} - s\right)\Delta \text{ and } \zeta(s,\Delta) \equiv \left(\frac{1}{2} - s\right)\Delta.$$

Certainly, $\sup_{(s,\Delta)\in[0,1]\times[-\varepsilon,\infty)} |\zeta_n(s,\Delta) - \zeta(s,\Delta)| \to_P 0$. Therefore, following the steps after (26), we obtain the desired bound. In fact, this bound is zero.[4] ■

[4]When $\sigma_U^2 + 2\sigma_{U,L} + \sigma_L^2 = 0$, we have $\dot{\theta}_U(h) + \dot{\theta}_L(h) = 0$ for all $h \in H$. Therefore, $\theta_{U,n}(h) + \theta_{L,n}(h)$ is locally constant at $h = 0$. This is the case where we know $\theta_{U,n}(h) + \theta_{L,n}(h)$ although we do not know separately $\theta_{U,n}(h)$ and $\theta_{L,n}(h)$. Hence we still know the minimax decision $\{\theta_{U,n}(h) + \theta_{L,n}(h)\}/2$, causing no regret.