cemmap
centre for microdata methods and practice

Inference on sets in finance

Victor Chernozhukov
Emre Kocatulum
Konrad Menzel

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP04/12

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL
An ESRC Research Centre

INFERENCE ON SETS IN FINANCE

VICTOR CHERNOZHUKOV[†] EMRE KOCATULUM[§] KONRAD MENZEL[‡]

Abstract. In this paper we introduce various set inference problems as they appear in finance and propose practical and powerful inferential tools. Our tools will be applicable to any problem where the set of interest solves a system of smooth estimable inequalities, though we will particularly focus on the following two problems: the admissible mean-variance sets of stochastic discount factors and the admissible mean-variance sets of asset portfolios. We propose to make inference on such sets using weighted likelihood-ratio and Wald type statistics, building upon and substantially enriching the available methods for inference on sets.

Keywords: Hansen-Jagannathan set, Marokowitz set, Inference

JEL classification: C10, C50

Date: August 5, 2008.

1. Introduction

In this paper we introduce various set inference problems as they appear in finance and propose practical and powerful inferential tools. Our tools will be applicable to any problem where the set of interest solves a system of estimable inequalities, though we will particularly focus on the following two problems: The first problem will deal with mean-variance sets of stochastic discount factors and the second with mean-variance sets of admissible portfolios.

Let us now introduce the problem. We begin by recalling two equations used by [Cochrane, 2005] to effectively summarize the science of asset pricing:

$$\begin{aligned} P_t &= E_t[M_{t+1}X_{t+1}] \\ M_{t+1} &= f(Z_{t+1}, parameters), \end{aligned}$$

where P_t is an asset price, X_{t+1} is the asset payoff, M_{t+1} is the stochastic discount factor (SDF) or pricing kernel (PK), which is a function f of some data Z_{t+1} and parameters, and E_t is the conditional expectation given information at time t. The set of SDFs M_t that can price existing assets generally form a proper set, that is, a set that is not a singleton. SDFs are not unique, because the existing payoffs to assets do not span the entire universe of possible random payoffs. Dynamic asset pricing models provide families of potential SDFs, for example, the standard consumption model predicts that an appropriate SDF can be stated in terms of intertemporal marginal rate of substitution:

$$M_t = \beta \frac{u'(C_{t+1})}{u'(C_t)},$$

where u denotes a utility function parameterized by some parameters, C_t denotes consumption at time t, and β denotes the subjective discount factor.

The basic econometric problem is to check which families of SDFs price the assets correctly and which do not. In other words, we want to check whether given families or subfamilies of SDFs are valid or not. One leading approach for performing the check is

to see whether mean and standard deviation of SDFs

$$\{\mu_M, \sigma_M\}$$

are admissible. The set of admissible means and standard deviations

$$\Theta_0 := \{ \text{ admissible pairs } (\mu, \sigma^2) \in R^2 \cap K\},$$

which is introduced by [Hansen and Jagannathan, 1991] is known as the Hansen-Jagannathan set and the boundary of the set Θ_0 is known as the Hansen-Jagannathan bound. In order to give a very specific, canonical example, let v and Σ denote the vector of mean returns and covariance matrix to assets $1, ..., N$ which are assumed not to vary with information sets at each period t. Let us denote

$$A = v'\Sigma^{-1}v, B = v'\Sigma^{-1}1_N, C = 1_N'\Sigma^{-1}1_N \tag{1.1}$$

where 1_N is a column vector of ones. Then the minimum variance $\sigma^2(\mu)$ achievable by a SDF given mean μ of the SDF is equal to

$$\sigma^2(\mu) = (1 - \mu v)' \Sigma^{-1} (1 - \mu v) = A\mu^2 - 2B\mu + C$$

Therefore, the HJ set is equal to

$$\Theta_0 = \{\underbrace{(\mu, \sigma)}_{\theta} \in \underbrace{\mathbb{R}^2 \cap K}_{\Theta} : \underbrace{\sigma(\mu) - \sigma}_{m(\theta)} \leq 0\},$$

where K is any compact set. That is,

$$\Theta_0 = \{\theta \in \Theta : m(\theta) \leq 0\}.$$

Note that the inequality-generating function $m(\theta)$ depends on the unknown parameters, the means and covariance of returns, $m(\theta) = m(\theta, \gamma)$ and $\gamma =$ vec (v, Σ).

Let us now describe the second problem. The classical [Markowitz, 1952] problem is to minimize the risk of a portfolio given some attainable level of return:

$$\min_w E_t[r_{p,t+1} - E_t[r_{p,t+1}]]^2 \text{ such that } E_t[r_{p,t+1}] = \mu,$$

where $r_{p,t+1}$ is portfolios return, determined as $r_{p,t+1} = wr_{t+1}$, where w is a vector of portfolio "weights" and r_{t+1} is a vector of returns on available assets. In a canonical version of the problem, we have that the vector of mean returns v and covariance of returns Σ do not vary with time period t, so that the problem becomes:

$$\sigma(\mu) = \min_{w} w'\Sigma w \text{ such that } w'v = \mu.$$

An explicit solution for $\sigma(\mu)$ takes the form,

$$\sigma^2(\mu) = \frac{C\mu^2 - 2B\mu + A}{AC - B^2}$$

where A,B and C are as in equation 1.1.

Therefore, the Markowitz (M) set of admissible standard deviations and means is given by

$$\Theta_0 = \{\underbrace{(\mu, \sigma)}_{\theta} \in \underbrace{\mathbb{R}^2 \cap K}_{\Theta} : \underbrace{\sigma(\mu) - \sigma}_{m(\theta)} \leq 0\},$$

that is,

$$\Theta_0 = \{\theta \in \Theta : m(\theta) \leq 0\}.$$

The boundary of the set Θ_0 is known as the efficient frontier. Note that as in HJ example, the inequality-generating function $m(\theta)$ depends on the unknown parameters, the means and covariance of returns, $m(\theta) = m(\theta, \gamma)$, where $\gamma =$ vec (v, Σ).

The basic problem of this paper is to develop inference methods on HJ and M sets, accounting for uncertainty in the estimation of parameters of the inequality-generating functions. The problem is to construct a confidence region R such that

$$\lim_{n\to\infty} P\{\Theta_0 \subseteq R\} = 1 - \alpha.$$

We will construct confidence regions for HJ sets using LR and Wald-type Statistics, building on and simultaneously enriching the approaches suggested in [Chernozhukov et al., 2007], [Beresteanu and Molinari, 2008], and [Molchanov, 1998]. We also would like to ensure that confidence regions R are as small as possible and converge to Θ_0 at the most rapid

attainable speed. We need the confidence region R for entire set Θ_0 in order to test validity of sets of SDFs. Once R is constructed, we can test infinite number of composite hypotheses, current and future, without compromising the significance level. Indeed, a typical application of HJ sets determines which sets of (μ, σ)'s within a given family fall in the HJ set and which do not. Similar comments about applicability of our approach go through for the M sets as well.

Our approach to inference using weighted Wald-type statistics complements and enriches the approach based on the directed Hausdorff distance suggested in [Beresteanu and Molinari, 2008] and [Molchanov, 1998]. By using weighting in the construction of the Wald-type statistics, we endow this approach with better invariance properties to parameter transformations, which results in noticeably sharper confidence sets, at least in the canonical emprical example that we will show. Thus, our construction is of independent interest for this type of inference, and is a useful complement to the work of [Beresteanu and Molinari, 2008] and [Molchanov, 1998]. Furthermore, our results on formal validity of the bootstrap for LR-type and W-type statistics are also of independent interest.

The rest of the paper is organized as follows. In Section 2 we present our estimation and inference results. In Section 3 we present an empirical example, illustrating the constructions of confidence sets for HJ sets. In Section 4 we draw conclusions and provide direction for further research. In the Appendix, we collect the proofs of the main results.

2. Estimation and Inference Results

2.1. **Basic Constructions.** We first introduce our basic framework. We have an inequality-generating function:

$$m : \Theta \mapsto \mathbb{R}.$$

The set of interest is the solution of the inequalities generated by the function $m(\theta)$ over a compact parameter space Θ:

$$\Theta_0 = \{\theta \in \Theta : m(\theta) \leq 0\}.$$

A natural estimator of Θ_0 is its empirical analog

$$\widehat{\Theta}_0 = \{\theta \in \Theta : \widehat{m}(\theta) \leq 0\},$$

where $\widehat{m}(\theta)$ is the estimate of the inequality-generating function. For example, in HJ and M examples, the estimate takes the form

$$\hat{m}(\theta) = m(\theta, \hat{\gamma}), \quad \hat{\gamma} = \text{ vec } (\widehat{v}, \widehat{\Sigma}).$$

Our proposals for confidence regions are based on (1) LR-type statistic and (2) Wald-type statistic. The LR-based confidence region is

$$R_{LR} = \left\{\theta \in \Theta : \left[\sqrt{n}\hat{m}(\theta)/s(\theta)\right]_+^2 \leq \hat{k}(1-\alpha)\right\}, \tag{2.1}$$

where $s(\theta)$ is the weighting function; ideally, the standard error of $\hat{m}(\theta)$; and $\hat{k}(1-\alpha)$ is a suitable estimate of

$$k(1-\alpha) = (1-\alpha) - \text{ quantile of } \mathcal{L}_n,$$

where

$$\mathcal{L}_n = \sup_{\theta \in \Theta_0} \left[\sqrt{n}\hat{m}(\theta)/s(\theta)\right]_+^2 \tag{2.2}$$

is the LR-type statistic, as in [Chernozhukov et al., 2007].

Our Wald-based confidence region is

$$R_W = \{\theta \in \Theta : [\sqrt{n}d(\theta, \widehat{\Theta}_0)/w(\theta)]^2 \leq \hat{k}(1-\alpha)\}, \tag{2.3}$$

where $w(\theta)$ is the weighting function, particular forms of which we will suggest later; and $\hat{k}$ is a suitable estimate of

$$k(1-\alpha) = (1-\alpha) - \text{ quantile of } \mathcal{W}_n,$$

where $\mathcal{W}_n$ is the weighted W-statistic

$$\mathcal{W}_n = \sup_{\theta \in \Theta_0} [\sqrt{n}d(\theta, \widehat{\Theta}_0)/w(\theta)]^2. \tag{2.4}$$

Recall that quantity $d(\theta, \widehat{\Theta}_0)$ is the distance of a point θ to a set $\widehat{\Theta}_0$, that is,

$$d(\theta, \widehat{\Theta}_0) := \inf_{\theta' \in \widehat{\Theta}_0} \|\theta - \theta'\|.$$

In the special case, where the weight function is flat, namely $w(\theta) = w$ for all θ, the W-statistic $\mathcal{W}_n$ becomes the canonical directed Hausdorff distance ([Molchanov, 1998], [Beresteanu and Molinari, 2008]):

$$\sqrt{\mathcal{W}_n} \propto d(\Theta_0, \widehat{\Theta}_0) = \sup_{\theta \in \Theta_0} \inf_{\theta' \in \widehat{\Theta}_0} \|\theta - \theta'\|.$$

The weighted statistic (2.4) is generally *not* a distance, but we argue that it provides a very useful extension of the canonical directed Hausdorff distance. In fact, in our empirical example precision weighting dramatically improves the confidence regions.

2.2. **A Basic Limit Theorem for LR and W statistics.** In this subsection, we develop a basic result on the limit laws of the LR and W statistics. We will develop this result under the following general regularity conditions:

R.1 *The estimates $\theta \mapsto \widehat{m}(\theta)$ of the inequality-generating function $\theta \mapsto m(\theta)$ are asymptotically Gaussian, namely, we have that in the metric space of bounded functions $\ell^\infty(\Theta)$*

$$\sqrt{n}(\widehat{m}(\theta) - m(\theta)) =_d G(\theta) + o_P(1),$$

where $G(\theta)$ is a Gaussian process with zero mean and a non-degenerate covariance function.

R.2 *Functions $\theta \mapsto \widehat{m}(\theta)$ and $\theta \mapsto m(\theta)$ admit continuous gradients $\nabla_\theta \widehat{m}(\theta)$ and $\nabla_\theta m(\theta)$ over the domain Θ, with probability one, where the former is a uniformly consistent estimate of the latter, namely uniformly in $\theta \in \Theta$*

$$\nabla_\theta \widehat{m}(\theta) = \nabla_\theta m(\theta) + o_P(1).$$

Moreover, the norm of the gradient $\|\nabla_\theta m(\theta)|$ is bounded away from zero.

R.3 *Weighting functions satisfy uniformly in* $\theta \in \Theta$

$$s(\theta) = \sigma(\theta) + o_p(1), \quad w(\theta) = \omega(\theta) + o_p(1),$$

where $\sigma(\cdot) \geq 0$ *and* $\omega(\cdot) \geq 0$ *are continuous functions bounded away from zero.*

In Condition R.1, we require the estimates of the inequality-generating functions to satisfy a uniform central limit theorem. There are plenty of sufficient conditions for this to hold provided by the theory of empirical processes. In our example, this condition will follow from asymptotic normality of the estimates of the mean returns and covariance of returns. In Condition R.2, we require that gradient of the estimate of the inequality-generating function is consistent for the gradient of the inequality-generating function. Moreover, we require that the minimal eigenvalue of $\nabla_\theta m(\theta)\nabla m(\theta)'$ is bounded away from zero, which is an identification condition that allows us to estimate, at a usual speed, the boundary of the set Θ_0, which we define as

$$\partial\Theta_0 := \{\theta \in \Theta : m(\theta) = 0\}.$$

In Condition R.3, we require that the estimates of the weight functions are consistent for the weight functions, which are well-behaved.

Under these conditions we can state the following general result.

THEOREM 1 *(Limit Laws of LR and W Statistics). Under R.1-R.3*

$$\mathcal{L}_n \;=_d\; \mathcal{L} + o_p(1), \quad \mathcal{L} = \sup_{\theta\in\partial\Theta_0} \left[\frac{G(\theta)}{\sigma(\theta)}\right]_+^2, \tag{2.5}$$

$$\mathcal{W}_n \;=_d\; \mathcal{W} + o_p(1), \quad \mathcal{W} = \sup_{\theta\in\partial\Theta_0} \left[\frac{G(\theta)}{\|\nabla_\theta m(\theta)\| \cdot \omega(\theta)}\right]_+^2, \tag{2.6}$$

where both $\mathcal{W}$ *and* $\mathcal{L}$ *have distribution functions that are continuous at their* $(1-\alpha)$*-quantiles for* $\alpha < 1/2$*. The two statistics are asymptotically equivalent under the following condition:*

$$\mathcal{W}_n =_d \mathcal{L}_n + o_p(1) \quad \textit{if} \quad w(\theta) = \frac{\|\nabla_\theta m(\theta)\|}{\sigma(\theta)} \quad \textit{for each} \quad \theta \in \Theta.$$

We see from this theorem that the LR and W statistics converge in law to well-behaved random variables that are continuous transformations of the limit Gaussian process $G(\theta)$. Moreover, we see that under an appropriate choice of the weighting functions, the two statistics are asymptotically equivalent.

For our application to HJ and M sets, the following conditions will be sufficient

C.1 *Estimator of the true parameter value γ_0 characterizing the inequality generating function $m(\theta) = m(\theta, \gamma_0)$, where γ_0 denotes the true parameter value, is such that $\sqrt{n}(\hat{\gamma} - \gamma_0) \to_d \Omega^{1/2} Z$, $Z = N(0, I_d)$.*

C.2 *Gradients $\nabla_\theta m(\theta, \gamma)$ and $\nabla_\gamma m(\theta, \gamma)$ are continuous over the compact parameter space $(\theta, \gamma) \in \Theta \times \Gamma$, where Γ is some set that includes an open neigborhood of γ_0. Moreover, the minimal eigenvalue of $\nabla_\theta m(\theta, \gamma) \nabla_\theta m(\theta, \gamma)'$ is bounded away from zero over $(\theta, \gamma) \in \Theta \times \Gamma$.*

It is straightforward to verify that these conditions hold for the canonical versions of the HJ and M problems.

Under these conditions we immediately conclude that the following approximation is true uniformly in θ, that is, in the metric space of bounded functions $\ell^\infty(\Theta)$:

$$\begin{aligned} \sqrt{n}(\widehat{m}(\theta) - m(\theta)) &= \nabla_\gamma m(\theta, \bar{\gamma})' \sqrt{n}(\hat{\gamma} - \gamma_0) + o_p(1) && (2.7) \\ &= {}_d \nabla_\gamma m(\theta, \gamma_0)' \Omega^{1/2} Z + o_p(1), && (2.8) \end{aligned}$$

where $\nabla m(\theta, \bar{\gamma})$ denotes the gradient with each of its rows evaluated at a value $\bar{\gamma}$ on the line connecting $\hat{\gamma}$ and γ_0, where value $\bar{\gamma}$ may vary from row to row of the matrix. Therefore, the limit process in HJ and M examples takes the form:

$$G(\theta) = \nabla_\gamma m(\theta, \gamma_0)' \Omega^{1/2} Z. \tag{2.9}$$

This will lead us to conclude formally below that conclusions of Theorem 1 hold with

$$\mathcal{L} = \sup_{\theta \in \partial\Theta_0} \left[\frac{\nabla_\gamma m(\theta,\gamma)'\Omega^{1/2}}{\sigma(\theta)} Z \right]_+^2, \tag{2.10}$$

$$\mathcal{W} = \sup_{\theta \in \partial\Theta_0} \left[\frac{\nabla_\gamma m(\theta,\gamma)'\Omega^{1/2}}{\|\nabla_\theta m(\theta,\gamma)\| \cdot \omega(\theta)} Z \right]_+^2. \tag{2.11}$$

A good strategy for choosing the weighting function for LR and W is to choose the studentizing Anderson-Darling weights

$$\sigma(\theta) = \|\nabla_\gamma m(\theta,\gamma_0)'\Omega^{1/2}\|, \tag{2.12}$$

$$\omega(\theta) = \frac{\|\nabla_\gamma m(\theta,\gamma_0)'\Omega^{1/2}\|}{\|\nabla_\theta m(\theta,\gamma_0)\|}. \tag{2.13}$$

The natural estimates of these weighting functions are given by the following plug-in estimators:

$$s(\theta) := \|\nabla_\gamma m(\theta,\widehat{\gamma})'\widehat{\Omega}^{1/2}\|, \tag{2.14}$$

$$w(\theta) := \frac{\|\nabla_\gamma m(\theta,\widehat{\gamma})'\widehat{\Omega}^{1/2}\|}{\|\nabla_\theta m(\theta,\widehat{\gamma})\|}. \tag{2.15}$$

We formalize the preceding discussion as the following corollary.

Corollary 1 *(Limit Laws of LR and W statistics in HJ and M problems). Suppose that Conditions C.1-C.2 hold. Then conditions R.1 and R.2 hold with the limit Gaussian process stated in equation (2.9). Furthermore, the plug-in estimates of the weighting functions (2.14) and (2.15) are uniformly consistent for the weighting functions (2.12) and (2.13), so that Condition R.3 holds. Therefore, conclusions of Theorem 1 hold with the limit laws for our statistics given by the laws of random variables stated in equations (2.10) and (2.11).*

2.3. **Basic Validity of the Confidence Regions.** In this section we shall suppose that we have suitable estimates of the quantiles of LR and W statistics and will verify basic validity of our confidence regions. In the next section we will provide a construction of such suitable estimates by the means of bootstrap and simulation.

Our result is as follows.

THEOREM 2 *(Basic Inferential Validity of Confidence Regions). Suppose that for $\alpha < 1/2$ we have consistent estimates of quantiles of limit statistics $\mathcal{W}$ and $\mathcal{L}$, namely,*

$$\hat{k}(1-\alpha) = k(1-\alpha) + o_p(1), \tag{2.16}$$

where $k(1-\alpha)$ is $(1-\alpha)$-quantile of either $\mathcal{W}$ or $\mathcal{L}$. Then as the sample size n grows to infinity, confidence regions R_{LR} and R_W cover Θ_0 with probability approaching $1-\alpha$:

$$Pr_P[\Theta_0 \subseteq R_{LR}] = Pr_P[\mathcal{L}_n \leq \hat{k}(1-\alpha)] \to Pr_P[\mathcal{L} \leq k(1-\alpha)] = (1-\alpha), \tag{2.17}$$

$$Pr_P[\Theta_0 \subseteq R_W] = Pr_P[\mathcal{W}_n \leq \hat{k}(1-\alpha)] \to Pr_P[\mathcal{W} \leq k(1-\alpha)] = (1-\alpha). \tag{2.18}$$

The result further applies to HJ and M problems.

COROLLARY 2*(Limit Laws of LR and W statistics in HJ and M problems). Suppose that Conditions C.1-C.2 hold and that consistent estimates of quantiles of statistics (2.10) and (2.11) are available. Then conclusions of Theorem 2 apply.*

2.4. **Estimation of Quantiles of LR and W Statistics by Bootstrap and Other Methods.** In this section we show how to estimate quantiles of LR and W statistics using bootstrap, simulation, and other resampling schemes under general conditions. The basic idea is as follows: First, let us take any procedure that consistently estimates the law of our basic Gaussian process G or a weighted version of this process appearing in the limit expressions. Second, then we can show with some work that we can get consistent estimates of the laws of LR and W statistics, and thus also obtain consistent estimates of their quantiles. It is well-known that there are many procedures for accomplishing the first step, including such common schemes as the bootstrap, simulation, and subsampling, including both cross-section and time series versions.

In what follows, we will ease the notation by writing our limit statistics as a special case of the following statistic:

$$\mathcal{S} = \sup_{\theta \in \partial\Theta_0} [V(\theta)]_+, \quad V(\theta) = \tau(\theta)G(\theta). \tag{2.19}$$

Thus, $\mathcal{S} = \mathcal{L}$ for $\tau(\theta) = 1/s(\theta)$ and $\mathcal{S} = \mathcal{W}$ for $\tau(\theta) = 1/[\|\nabla_\theta m(\theta)\| \cdot \omega(\theta)]$. We take τ to be a continuous function bounded away from zero on the parameter space. We also need to introduce the following notations and concepts. Our process V is a random element that takes values in the metric space of continuous functions $C(\Theta)$ equipped with the uniform metric. The underlying measure space is $(\Omega, \mathcal{F})$ and we denote the law of V under the probability measure P by the symbol $\mathcal{Q}_V$.

Suppose we have an estimate $\mathcal{Q}_{V^*}$ of the law $\mathcal{Q}_V$ of the Gaussian process V. This estimate $\mathcal{Q}_{V^*}$ is a probability measure generated as follows. Let us fix another measure space $(\Omega', \mathcal{F}')$ and a probability measure P^* on this space, then given a random element V^* on this space taking values in $C(\Theta)$, we denote its law under P^* by $\mathcal{Q}_{V^*}$. We thus identify the probability measure P^* with a data-generating process by which we generate draws or realizations of V^*. This identification allows us to encompass such methods of producing realizations of V^* as the bootstrap, subsampling, or other simulation methods. We require that the estimate $\mathcal{Q}_{V^*}$ is consistent for $\mathcal{Q}_V$ in any metric ρ_K metrizing weak convergence, where we can take the metric to be the Kantarovich-Rubinstein metric. Let us mention right away that there are many results that verify this basic consistency condition for various rich forms of processes V and various bootstrap, simulation, and subsampling schemes for estimating the laws of these processes, as we will discuss in more detail below.

In order to recall the definition of the Kantarovich-Rubinstein metric, let $\theta \mapsto v(\theta)$ be an element of a metric space (M, d), and $Lip(M)$ be a class of Lipschitz functions $\varphi : M \to \mathbb{R}$ that satisfy:

$$|\varphi(v) - \varphi(v')| \leq d(v, v') \wedge 1, \quad |\varphi(v)| \leq 1,$$

The Kantarovich-Rubinstein distance between probability laws $\mathcal{Q}$ and $\mathcal{Q}'$ is

$$\rho_K(\mathcal{Q}, \mathcal{Q}'; M) := \sup_{\varphi \in Lip(M)} |E_{\mathcal{Q}}\varphi - E_{\mathcal{Q}'}\varphi|.$$

As stated earlier, we require that the estimate $\mathcal{Q}_{V^*}$ is consistent for $\mathcal{Q}_V$ in the metric ρ_K, that is

$$\rho_K(\mathcal{Q}_{V^*}, \mathcal{Q}_V; C(\Theta)) = o_p(1). \tag{2.20}$$

Let $\mathcal{Q}_{\mathcal{S}}$ denote the probability law of $\mathcal{S} = \mathcal{W}$ or $\mathcal{L}$, which is in turn induced by the law $\mathcal{Q}_{\mathcal{V}}$ of the Gaussian process V. We need to define the estimate $\mathcal{Q}_{\mathcal{S}^*}$ of this law. First, we define the following plug-in estimate of the boundary set $\partial\Theta_0$, which we need to state here:

$$\widehat{\partial\Theta_0} = \{\theta \in \Theta : \widehat{m}(\theta) = 0\}. \tag{2.21}$$

This estimate turns out to be consistent at the usual root-n rate, by the argument like the one given in [Chernozhukov et al., 2007]. Then define $\mathcal{Q}_{\mathcal{S}^*}$ as the law of the following random variable

$$\mathcal{S}^* = \sup_{\theta \in \widehat{\partial\Theta_0}} [V^*(\theta)]_+ \tag{2.22}$$

In this definition, we hold the hatted quantities fixed, and the only random element is V^* that is drawn according to the law $\mathcal{Q}_{V^*}$.

We will show that the estimated law $\mathcal{Q}_{\mathcal{S}^*}$ is consistent for $\mathcal{Q}_{\mathcal{S}}$ in the sense that

$$\rho_K(\mathcal{Q}_{\mathcal{S}^*}, \mathcal{Q}_{\mathcal{S}}; \mathbb{R}) = o_p(1). \tag{2.23}$$

Consistency in the Kantarovich-Rubinstein metric in turn implies consistency of the estimates of the distribution function at continuity points, which in turn implies consistency of the estimates of the quantile function.

Equipped with the notations introduced above we can now state our result.

THEOREM 3 *(Consistent Estimation of Quantiles) Suppose Conditions R.1-R.3 hold, and any mechanism, such as bootstrap or other method, is available, which provides a*

consistent estimate of the law of our limit Gaussian processes V , namely equation (2.20) holds. Then, the estimates of the laws of the limit statistics $S = \mathcal{W}$ or $\mathcal{L}$ defined above are consistent in the sense of equation (2.23). As a consequence, we have that the estimates of the quantiles are consistent in the sense of equation (2.16).

We now specialize this result to the HJ and M problems. We begin by recalling that our estimator satisfies

$$\sqrt{n}(\hat{\gamma} - \gamma) =_d \Omega^{1/2} Z + o_p(1).$$

Then our limit statistics take the form:

$$\mathcal{S} = \sup_{\theta \in \partial\Theta_0} [V(\theta)]_+^2, \quad V(\theta) = t(\theta)' Z,$$

where $t(\theta)$ is a vector valued weight function, in particular, for $\mathcal{S} = \mathcal{L}$ we have $t(\theta) = (\nabla_\gamma m(\theta, \gamma)' \Omega^{1/2})/\sigma(\theta)$ and for $\mathcal{S} = \mathcal{W}$ we have $t(\theta) = (\nabla_\gamma m(\theta, \gamma)' \Omega^{1/2})/(\|\nabla_\theta m(\theta, \gamma)\| \cdot \omega(\theta))$. Here we shall assume that we have a consistent estimate $\mathcal{Q}_{Z^*}$ of the law $\mathcal{Q}_Z$ of Z, in the sense that,

$$\rho_K(\mathcal{Q}_{Z^*}, \mathcal{Q}_Z) = o_p(1). \tag{2.24}$$

There are many methods that provide such consistent estimates of the laws. Bootstrap is known to be valid for various estimation methods ([van der Vaart and Wellner, 1996]); simulation method that simply draws $Z \sim N(0, I)$ is another valid method; and subsampling is another rather general method ([Politis and Romano, 1994]). Next, the estimate $\mathcal{Q}_{V^*}$ of the law $\mathcal{Q}_{V^*}$ is then defined as:

$$V^*(\theta) = \widehat{t}(\theta)' Z^*, \tag{2.25}$$

where $\widehat{t}(\theta)$ is a vector valued weighting function that is uniformly consistent for the weighting function $t(\theta)$. In this definition we hold the hatted quantity fixed, and the only random element is Z^* that is drawn according to the law $\mathcal{Q}_{Z^*}$. Then, we define the random variable

$$\mathcal{S}^* = \sup_{\theta \in \widehat{\partial\Theta_0}} [V^*(\theta)]_+^2,$$

and use its law $\mathcal{Q}_{\mathcal{S}^*}$ to estimate the law $\mathcal{Q}_{\mathcal{S}}$.

We can now state the following corollary.

Corollary 3 *(Consistent Estimation of Quantiles in HJ and M problems) Suppose Conditions C.1-C.2 hold, and any mechanism, such as bootstrap or other method, that provides a consistent estimate of the law of Z is available, namely equation (2.24) holds. Then, this provides us with a consistent estimate of the law of our limit Gaussian process G, namely equation (2.20) holds. Then, all of the conclusions of Theorem 3 hold.*

3. Empirical Example

For empirical example we use HJ bounds which are widely used in testing asset pricing models. For comparison purposes, the data used in this section is very similar to data used in [Hansen and Jagannathan, 1991]. The two asset series used are annual treasury bond returns and annual NYSE value-weighted dividend included returns. These nominal returns are converted to real returns by using implicit price deflator based on personal consumption expenditures as in [Hansen and Jagannathan, 1991]. Returns data is from CRSP. Implicit price deflator is available from St. Louis Fed and based on National Income and Product Accounts of United States. The time period is 1959-2006 (inclusive).

Figure 1 simply traces out the mean-standard deviation pairs which satisfy

$$m\left(\theta, \hat{\gamma}\right) = 0$$

where $\hat{\gamma}$ is estimated using sample moments.

Figure 2 represents the uncertainity caused by the estimation of γ. To estimate the distribution of $\hat{\gamma}$ bootstrap method is used. Observations are drawn with replacement from the bivariate time series of stock and bond returns. 100 bootstraps result in 100 $\hat{\gamma}$. The resulting HJ bounds are included in the figure.

In Figure 3 in addition to the bootstrapped curves 90% confidence region based on LR statistic is presented. LR based confidence region covers most of the bootstrap draws below the HJ bounds as expected. An attractive outcome of using this method is that the resulting region does not include any unnecessary areas that is not covered by bootstrap draws.

Figure 4 plots 90% confidence region based on unweighted LR statistic. Comparison of Figure 3 and Figure 4 reveals that precision weighting plays a very important role in delivering good confidence sets. Without precision weighting LR statistic delivers a confidence region that includes unlikely regions in the parameter space where standard deviation of the discount factor is zero. On the other hand precision weighted LR based confidence region is invariant to parameter transformations, for example, changes in units of measurement. This invariance to parameter transformations is the key property of a statistic to deliver desirable confidence regions that does not cover unnecessary areas.

Figure 5 plots confidence region based on Wald-based statistic with no precision weighting. This is identical to the confidence region based on Hausdorff distance. Similar to Figure 4 this region covers a large area of the parameter space where no bootstrap draws appear. This picture reveals a key weakness of using an unweighted Wald-based statistic or Hausdorff distance to construct confidence regions. These methods are not invariant to parameter transformations which results in confidence regions with undesirable qualities that cover unnecessary areas in the parameter space. The problem in Figure 4 and Figure 5 are of similar nature. In both of these cases the statistics underlying the confidence regions are not invariant to parameter transformations therefore when drawing confidence regions uncertainity in one part of the plot is assumed to be identical to uncertainity in other parts of the plot. However a quick look at the Figure 2 reveals that uncertainity regarding the location of the HJ bound varies for a given mean or standard deviation of the stochastic discount factor.

Figure 6 plots the confidence region based on weighted Wald statistic. Weighting fixes the problem and generates a statistic that is invariant to parameter transformations. The resulting confidence set looks very similar to weighted LR based confidence set in Figure 3 as it covers most of the bootstrap draws below the HJ bounds and does not include unnecessary regions in the parameter space.

4. Conclusion

In this paper we provided various inferential procedures for inference on sets that solve a system of inequalities. These procedures are useful for inference on Hansen-Jagannathan mean-variance sets of admissible stochastic discount factors and Markowitz mean-variance sets of admissible portfolios.

Appendix A. Proofs

A.1. **Proof of Theorem 1.** Part 1. (*Limit law of* $\mathcal{L}_n$.) Let $G_n = \sqrt{n}(\widehat{m} - m)$. Then

$$\begin{aligned}\mathcal{L}_n = \sup_{\theta\in\Theta_0}\left[\sqrt{n}\hat{m}(\theta)/s(\theta)\right]_+^2 &= \sup_{\theta\in\Theta_0}\left[(G_n(\theta)+\sqrt{n}m(\theta))/s(\theta)\right]_+^2 \\ &=_d \sup_{\theta\in\Theta_0}\left[(G(\theta)+\sqrt{n}m(\theta))/\sigma(\theta)+o_p(1)\right]_+^2 \\ &= \sup_{\theta\in\partial\Theta_0}\left[(G(\theta)+\sqrt{n}m(\theta))/\sigma(\theta)+o_p(1)\right]_+^2\end{aligned}$$

The steps, apart from the last, immediately follow from Conditions R.1 and R.3. The last step follows from the argument given below. Indeed, take any sequence $\theta_n \in \Theta_0$ such that

$$\sup_{\theta\in\Theta_0}\left[(G(\theta)+\sqrt{n}m(\theta))/\sigma(\theta)+o_p(1)\right]_+^2 = \left[(G(\theta_n)+\sqrt{n}m(\theta_n))/\sigma(\theta_n)+o_p(1)\right]_+^2.$$

In order for this to occur we need to have that

$$\sqrt{n}m(\theta_n)/\sigma(\theta_n) = O_p(1),$$

which is only possible in view of condition R.2 if, for some stochastically bounded sequence of positive random variables $C_n = O_p(1)$,

$$\sqrt{n}d(\theta_n, \partial\Theta_0) \le C_n.$$

Therefore we conclude that

$$\begin{aligned}&\sup_{\theta\in\Theta_0}\left[(G(\theta)+\sqrt{n}m(\theta))/\sigma(\theta)+o_p(1)\right]_+^2 \\ &= \sup_{\theta\in\partial\Theta_0,\theta+\lambda/\sqrt{n}\in\Theta_0,\|\lambda\|\le C_n}\left[(G(\theta+\lambda/\sqrt{n})+\sqrt{n}m(\theta+\lambda/\sqrt{n}))/\sigma(\theta+\lambda/\sqrt{n})+o_p(1)\right]_+^2\end{aligned}$$

Using stochastic equicontinuity of G and continuity of σ, the last quantity is further approximated by

$$\sup_{\theta\in\partial\Theta_0,\theta+\lambda/\sqrt{n}\in\Theta_0,\|\lambda\|\le C_n}\left[(G(\theta)+\sqrt{n}m(\theta+\lambda/\sqrt{n}))/\sigma(\theta)+o_p(1)\right]_+^2.$$

Because $\sqrt{n}m(\theta + \lambda/\sqrt{n}) \leq 0$ and $m(\theta) = 0$ for $\theta \in \Theta_0$ and $\theta + \lambda/\sqrt{n} \in \Theta_0$, we conclude that the last quantity is necessarily equal to $\sup_{\theta \in \partial\Theta_0} [G(\theta)/\sigma(\theta)]_+^2$, yielding the conclusion we needed.

Part 2. (*Limit Law of* $\mathcal{W}_n$). We will begin by justifying the approximation holding with probability going to one

$$\sup_{\theta \in \Theta_0} \sqrt{n} d(\theta, \widehat{\Theta}_0) = \sup_{\Theta_n} \sqrt{n} d(\theta, \widehat{\Theta}_0). \tag{A.1}$$

where

$$\Theta_n = \{\theta \in \Theta_0 : \sqrt{n} d(\theta, \partial\Theta_0) \leq C_n\}$$

where C_n is some stochastically bounded sequence of positive random variables, $C_n = O_p(1)$. Note that right hand side is less than or equal to the left hand side in general, so we only need to show that the right hand side can not be less. Indeed, let θ_n be any sequence such that

$$\sup_{\theta \in \Theta_0} \sqrt{n} d(\theta, \widehat{\Theta}_0) = \sqrt{n} d(\theta_n, \widehat{\Theta}_0).$$

If $\widehat{m}(\theta_n) \leq 0$, then $d(\theta_n, \widehat{\Theta}_0) = 0$, and the claim follows trivially since the right hand side of (A.1) is non-negative and is less than or equal to the left hand side of (A.1). If $\widehat{m}(\theta_n) > 0$, then $d(\theta_n, \widehat{\Theta}_0) > 0$, but for this and for $\theta_n \in \Theta_0$ to take place we must have that $0 < \widehat{m}(\theta_n) = O_p(1/\sqrt{n})$, which by Condition R.2 implies that $d(\theta_n, \widehat{\Theta}_0) = O_p(1/\sqrt{n})$.

In the discussion the quantity $\theta^*(\theta)$ as follows

$$\theta^*(\theta) \in \arg\min_{\theta' \in \partial\Theta_0} \|\theta - \theta'\|^2.$$

The argmin set $\theta^*(\theta)$ is a singleton simultaneously for all $\theta \in \Theta_n$, provided n is sufficiently large. This follows from condition R.2 imposed on the gradient $\nabla_\theta m$. Moreover, by examining the optimality condition we can conclude that we must have that for $\theta \in \Theta_n$

$$(I - \nabla_\theta m(\theta)(\nabla_\theta m(\theta)' \nabla_\theta m(\theta))^{-1} \nabla_\theta m(\theta)')(\theta - \theta^*) = o_p(1) \tag{A.2}$$

The projection of $\theta \in \Theta$ onto the set $\widehat{\Theta} := \{\theta \in \Theta : \widehat{m}(\theta) \leq 0\}$ is given by

$$\tilde{\theta}(\theta) = \arg \min_{\theta : \widehat{m}(\theta') \leq 0} \|\theta - \theta'\|^2.$$

If $\widehat{m}(\theta) \leq 0$, then $\tilde{\theta}(\theta) = \theta$. If $\widehat{m}(\theta) > 0$, then $\tilde{\theta}(\theta) = \bar{\theta}(\theta)$, where

$$\bar{\theta}(\theta) = \arg \min_{\theta : \widehat{m}(\theta') = 0} \|\theta - \theta'\|^2.$$

In what follows we will suppress the indexing by θ in order to ease the notation, but it should be understood that we will make all the claims uniformly in $\theta \in \Theta_n$. For each θ, the Lagrangian for this problem is $\|\theta - \theta'\|^2 + 2\widehat{m}(\theta')'\lambda$. Therefore, the quantity $\bar{\theta}(\theta)$ can be take to be an interior solution of the saddle-point problem

$$(\bar{\theta} - \theta) + \nabla_\theta \widehat{m}(\bar{\theta})\lambda = 0$$

$$\widehat{m}(\bar{\theta}) = 0$$

The corner solutions do not contribute to the asymptotic behavior of $\mathcal{W}_n$, and thus can be ignored. A formal justification for this will be presented in future versions of this work. Using mean-value expansion we obtain

$$(\bar{\theta} - \theta) + \nabla_\theta \widehat{m}(\bar{\theta})\lambda = 0$$

$$m(\theta^*) + \nabla_\theta m(\check{\theta})(\bar{\theta} - \theta^*) + \widehat{m}(\bar{\theta}) - m(\bar{\theta}) = 0$$

Since $\nabla_\theta \widehat{m}(\check{\theta}) = \nabla_\theta m(\theta) + o_p(1)$ and $\nabla_\theta m(\bar{\theta}) = \nabla_\theta m(\theta) + o_p(1)$ uniformly in $\theta \in \Theta$, solving for $(\bar{\theta} - \theta)$ we obtain

$$\begin{aligned} \bar{\theta} - \theta^* &= [\nabla_\theta m(\theta)(\nabla_\theta m(\theta)'\nabla_\theta m(\theta))^{-1} + o_p(1)](\widehat{m}(\bar{\theta}) - m(\bar{\theta})) \\ &+ (I - \nabla_\theta m(\theta)(\nabla_\theta m(\theta)'\nabla_\theta m(\theta))^{-1}\nabla_\theta m(\theta)' + o_p(1))(\theta - \theta^*) \end{aligned}$$

Using that $\sqrt{n}(\widehat{m}(\theta) - m(\theta)) =_d G(\theta) + o_p(1)$, we obtain

$$\begin{aligned} \sqrt{n}(\bar{\theta} - \theta^*) &=_d \nabla_\theta m(\theta)(\nabla_\theta m(\theta)'\nabla_\theta m(\theta))^{-1} G(\theta) \\ &+ (I - \nabla_\theta m(\theta)(\nabla_\theta m(\theta)'\nabla_\theta m(\theta))^{-1}\nabla_\theta m(\theta)')(\theta - \theta^*) \end{aligned}$$

Furthermore, by $\theta \in \Theta_n$ and by the approximate orthgonality condition (A.2) we further have that $(I - \nabla_\theta m(\theta)(\nabla_\theta m(\theta)'\nabla_\theta m(\theta))^{-1}\nabla_\theta m(\theta)')(\theta - \theta^*) = o_p(1)$, so that

$$\sqrt{n}(\bar{\theta} - \theta^*) =_d \nabla_\theta m(\theta)(\nabla_\theta m(\theta)'\nabla_\theta m(\theta))^{-1}G(\theta) + o_p(1).$$

We next approximate $1(\widehat{m}(\theta) > 0)$ using that

$$\begin{aligned}
\sqrt{n}\widehat{m}(\theta) &= \sqrt{n}\widehat{m}(\bar{\theta}) + \check{M}_\theta\sqrt{n}(\theta - \bar{\theta}) \\
&= \nabla m(\theta)'\sqrt{n}(\theta - \bar{\theta}) + o_p(1), \\
&= G(\theta) + o_p(1)
\end{aligned}$$

where we used that $\widehat{m}(\bar{\theta}) = 0$.

Thus, uniformly in $\theta \in \Theta_n$ we have that

$$\begin{aligned}
\sqrt{n}d(\theta, \widehat{\Theta}_n) &= \|\bar{\theta} - \theta\|^2 1\{\nabla m(\theta)\sqrt{n}(\theta - \bar{\theta}) > 0 + o_p(1)\} \\
&= |\nabla_\theta m(\theta)'\nabla_\theta m(\theta))^{-1/2}G(\theta)|1\{G(\theta) > 0 + o_p(1)\} \\
&= [\|\nabla_\theta m(\theta)\|^{-1}G(\theta) + o_p(1)]_+
\end{aligned}$$

Therefore, given the initial approximation (A.1) we obtain that

$$\mathcal{W}_n =_d \sup_{\theta \in \partial\Theta_0} [\|\nabla_\theta m(\theta)\|^{-1}G(\theta)]_+ + o_p(1). \tag{A.3}$$

Part 3. (*Continuity of the Limit Distributions*). The continuity of the distribution function $\mathcal{L}$ on $(0, \infty)$ follows from the Davydov et al (1998) and from the assumption that the covariance function of G is non-degenerate. Probability that $\mathcal{L}$ is greater than zero is equal to the probability that $\max_j \sup_{\theta \in \Theta} G_j(\theta) > 0$ which is greater than the probability that $G_{j'}(\theta') > 0$ for some fixed j' and θ', but the latter is equal to $1/2$. Therefore the claim follows. The claim of continuity of the distribution function of $\mathcal{W}$ on $(0, \infty)$ follows similarly. □

A.2. **Proof of Corollary 1.** This corollary immediately follows from the assumed conditions and from the comments given in the main text preceding the statement of Corollary 1. □

A.3. **Proof of Theorem 2.** We have that $Pr_P[\Theta_0 \subseteq R_{LR}] = Pr_P[\mathcal{L}_n \leq \hat{k}(1-\alpha)]$ by the construction of the confidence region. We then have that for any $\alpha < 1/2$ that $k(1-\alpha)$ is a continuity point of the distribution function of $\mathcal{L}$, so that for any sufficiently small ϵ

$$Pr_P[\mathcal{L}_n \leq \hat{k}(1-\alpha)] \leq Pr_P[\mathcal{L}_n \leq k(1-\alpha)+\epsilon] \to Pr_P[\mathcal{L} \leq k(1-\alpha)+\epsilon]$$
$$Pr_P[\mathcal{L}_n \leq \hat{k}(1-\alpha)] \geq Pr_P[\mathcal{L}_n \leq k(1-\alpha)-\epsilon] \to Pr_P[\mathcal{L} \leq k(1-\alpha)-\epsilon]$$

Since we can set ϵ as small as we like and $k(1-\alpha)$ is a continuity point of the distribution function of $\mathcal{L}$, we have that

$$Pr_P[\mathcal{L}_n \leq \hat{k}(1-\alpha)] \to Pr_P[\mathcal{L} \leq k(1-\alpha)] = (1-\alpha).$$

We can conclude similarly for the W-statistic $\mathcal{W}_n$. □.

A.4. **Proof of Corollary 2.** This corollary immediately follows from the assumed conditions and Corollary 1. □

A.5. **Proof of Theorem 3.** We have that

$$E_{P^*}[\varphi(V^*)] - E_P[\varphi(V)] = o_p(1) \text{ uniformly in } \varphi \in Lip(C(\Theta)).$$

This implies that

$$E_{P^*}[\varphi([V^*]_+)] - E_P[\varphi([V]_+)] = o_p(1) \text{ uniformly in } \varphi \in Lip(C(\Theta)),$$

since the composition $\varphi \circ [\cdot]_+ \in Lip(C(\Theta))$ for $\varphi \in Lip(C(\Theta))$. This further implies that

$$E_{P^*}[\varphi'(\sup_{R_n}[V^*]_+)] - E_P[\varphi'(\sup_{R_n}[V]_+)] = o_p(1) \text{ uniformly in } \varphi' \in Lip(\mathbb{R}),$$

since the composition $\varphi'(\sup_{R_n}[\cdot]_+) \in Lip(C(\Theta))$ for $\varphi' \in Lip(\mathbb{R})$ and R_n denoting any sequence of closed non-empty subsets in Θ. We have that $\widehat{\partial\Theta_0}$ converges to $\partial\Theta_0$ in the

Hausdorff distance, so that

$$|E_P[\varphi'(\sup_{\widehat{\partial\Theta_0}}[V]_+) - \varphi'(\sup_{\partial\Theta_0}[V]_+)]|$$
$$\leq E[|\sup_{\widehat{\partial\Theta_0}}[V]_+ - \sup_{\partial\Theta_0}[V]_+| \wedge 1] = o_p(1) \text{ uniformly in } \varphi' \in Lip(\mathbb{R}),$$

since $\sup_{\widehat{\partial\Theta_0}}[V]_+ - \sup_{\partial\Theta_0}[V] = o_p(1)$ by stochastic equicontinuity of the process V. Since metric ρ_K is a proper metric that satisfies the triangle inequality, we have shown that

$$\rho_K(\mathcal{Q}_{\mathcal{S}^*}, \mathcal{Q}_{\mathcal{S}}) = o_p(1).$$

Next, we note that the convergence $\rho_K(\mathcal{Q}_{\mathcal{S}_n}, \mathcal{Q}_{\mathcal{S}}) = o(1)$, for any sequence of laws $\mathcal{Q}_{\mathcal{S}_n}$ of a sequence of random variables S_n defined on probability space $(\Omega', \mathcal{F}', P_n)$ implies the convergence of the distribution function

$$Pr_{\mathcal{Q}_{\mathcal{S}_n}}[\mathcal{S}_n \leq s] = Pr_{\mathcal{Q}_{\mathcal{S}}}[\mathcal{S} \leq s] + o(1)$$

at each continuity point $(0, \infty)$ of the mapping $s \mapsto Pr[\mathcal{S} \leq s]$ and also convergence of quantile functions

$$\inf\{s : Pr_{\mathcal{Q}_{\mathcal{S}_n}}[\mathcal{S}_n \leq s] \geq p\} = \inf\{s : Pr_{\mathcal{Q}_{\mathcal{S}}}[\mathcal{S} \leq s] \geq p\} + o(1)$$

at each continuity point p of the mapping $s \mapsto \inf\{s : Pr_{\mathcal{Q}_{\mathcal{S}}}[\mathcal{S} \leq s] \geq p\}$. Recall from Theorem 1 that the set of continuity points necessarily includes the region $(0, 1/2)$.

By the Extended Continuous Mapping Theorem we conclude that since $\rho_K(\mathcal{Q}_{\mathcal{S}^*}, \mathcal{Q}_{\mathcal{S}}) = o_p(1)$, for any sequence of laws $\mathcal{Q}_{\mathcal{S}^*}$ of random variable $\mathcal{S}^*$ defined on probability space $(\Omega', \mathcal{F}', P^*)$, we obtain the convergence in probability of the distribution function

$$Pr_{\mathcal{Q}_{\mathcal{S}^*}}[\mathcal{S}^* \leq s] = Pr_{\mathcal{Q}_{\mathcal{S}}}[\mathcal{S} \leq s] + o_p(1)$$

at each continuity point $(0, \infty)$ of the mapping $s \mapsto Pr[\mathcal{S} \leq s]$ and also convergence in probability of the quantile functions

$$\inf\{s : Pr_{\mathcal{Q}_{\mathcal{S}^*}}[\mathcal{S}^* \leq s] \geq p\} = \inf\{s : Pr_{\mathcal{Q}_{\mathcal{S}}}[\mathcal{S} \leq s] \geq p\} + o_p(1)$$

at each continuity point p of the mapping $s \mapsto \inf\{s : Pr_{\mathcal{Q}_{\mathcal{S}}}[\mathcal{S} \leq s] \geq p\}$. □

A.6. **Proof of Corollary 3.** In order to prove this corollary it suffices to show that

$$\rho_K(\mathcal{Q}_{\hat{t}'Z^*}, \mathcal{Q}_{t'Z}; C(\Theta)) = o_p(1).$$

Without loss of generality we can take $\sup \|\hat{t}\| \leq 1$ and $\sup \|t\| \leq 1$. The claim will follow from

$$\rho_K(\mathcal{Q}_{\hat{t}'Z^*}, \mathcal{Q}_{t'Z}; C(\Theta)) \leq \rho_K(\mathcal{Q}_{\hat{t}'Z^*}, \mathcal{Q}_{\hat{t}'Z}; C(\Theta)) + \rho_K(\mathcal{Q}_{\hat{t}'Z}, \mathcal{Q}_{\hat{t}'Z}; C(\Theta)) = o_p(1).$$

That $\rho_K(\mathcal{Q}_{\hat{t}'Z^*}, \mathcal{Q}_{\hat{t}'Z}; C(\Theta)) = o_p(1)$ follows immediately from $\rho_K(\mathcal{Q}_{Z^*}, \mathcal{Q}_Z) = o_p(1)$ and from the mapping $\varphi(\hat{t}'\cdot) \in Lip(\mathbb{R}^k)$ (indeed, $|\varphi(\hat{t}'z) - \varphi(t'z)| \leq \sup |\hat{t}'(z - z')| \wedge 1 \leq [(\sup \|t\| \sup \|z - z'\|) \wedge 1] \leq [\sup \|z - z'\| \wedge 1]$. That $\rho_K(\mathcal{Q}_{\hat{t}'Z}, \mathcal{Q}_{\hat{t}'Z}; C(\Theta)) = o_p(1)$ follows because uniformly in $\varphi \in Lip(C(\Theta)$

$$|E[\varphi(\hat{t}'Z)] - \varphi(t'Z)| \leq E[\sup |(\hat{t} - t)'Z| \wedge 1] \leq E[\sup \|\hat{t} - t\| \|Z\| \wedge 1] = o_p(1). \quad \square$$

References

[Beresteanu and Molinari, 2008] Beresteanu, A. and Molinari, F. (2008). Asymptotic properties for a class of partially identified models. *Econometrica*.

[Chernozhukov et al., 2007] Chernozhukov, V., Hong, H., and Tamer, E. (2007). Estimation and confidence regions for parameter sets in econometric models. *Econometrica*.

[Cochrane, 2005] Cochrane, J. H. (2005). *Asset Pricing*. Princeton University Press.

[Hansen and Jagannathan, 1991] Hansen, L. P. and Jagannathan, R. (1991). Implications of security market data for models of dynamic economies. *The Journal of Political Economy*.

[Markowitz, 1952] Markowitz, H. (1952). Portfolio selection. *The Journal of Finance*.

[Molchanov, 1998] Molchanov, I. S. (1998). A limit theorem for solutions of inequalities. *Scandinavian Journal of Statistics*.

[Politis and Romano, 1994] Politis, D. N. and Romano, J. P. (1994). Large sample confidence regions based on subsamples under minimal assumptions. *The Annals of Statistics*.

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

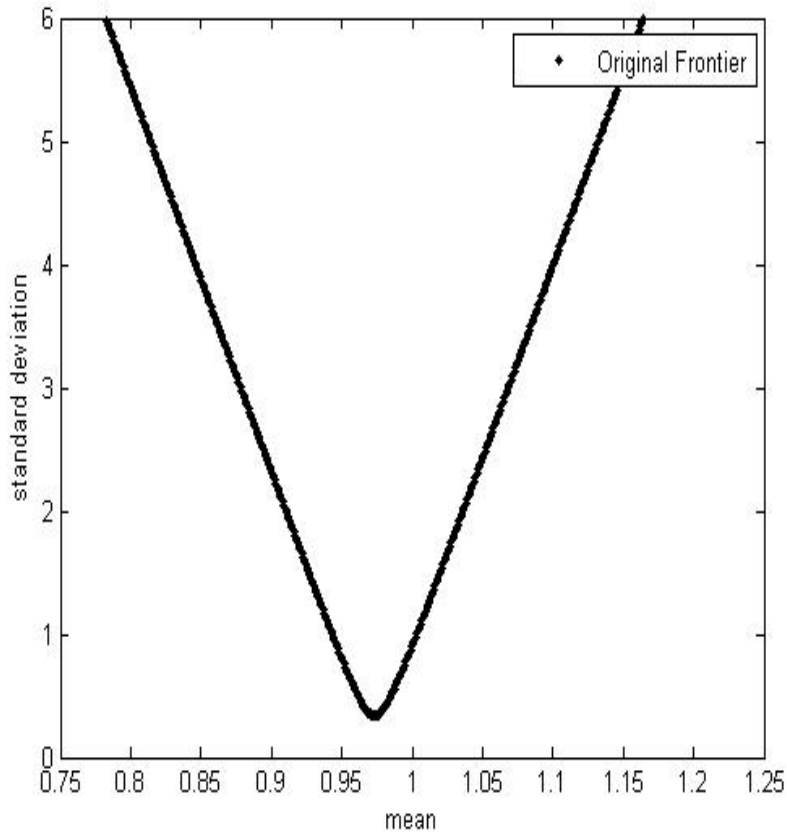

FIGURE 1. Estimated HJ Bounds

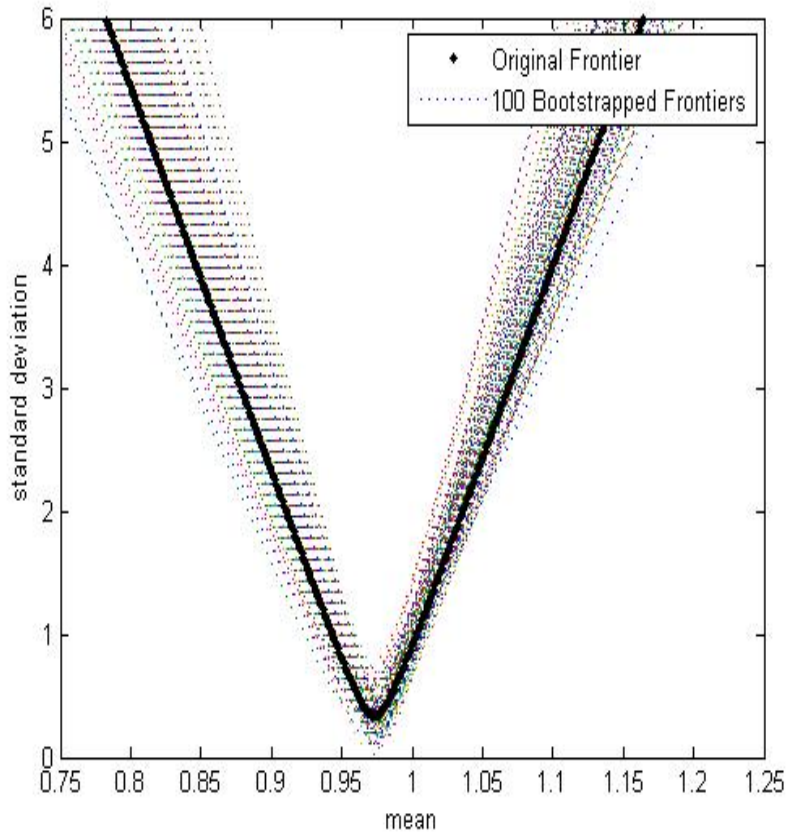

FIGURE 2. Estimated HJ Bounds and Bootstrap Draws

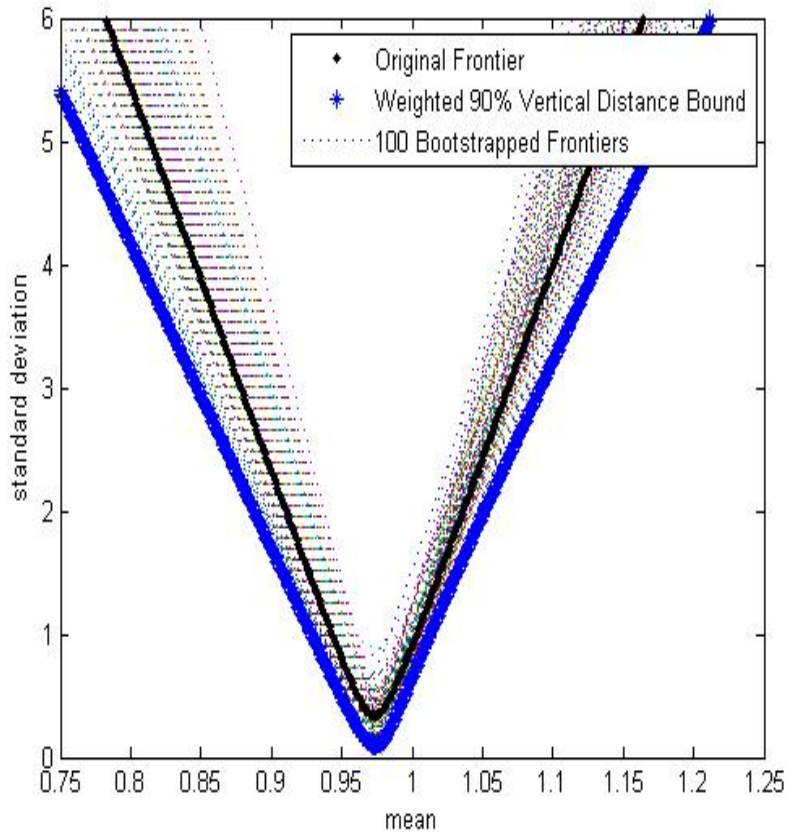

FIGURE 3. 90% Confidence Region using LR Statistic

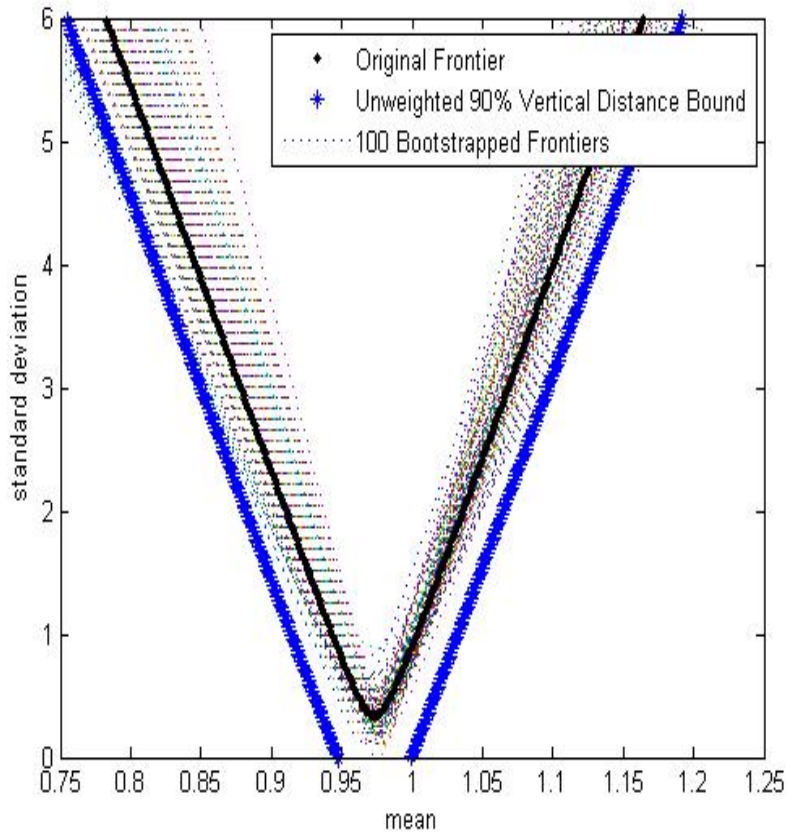

FIGURE 4. 90% Confidence Region using Unweighted LR Statistic

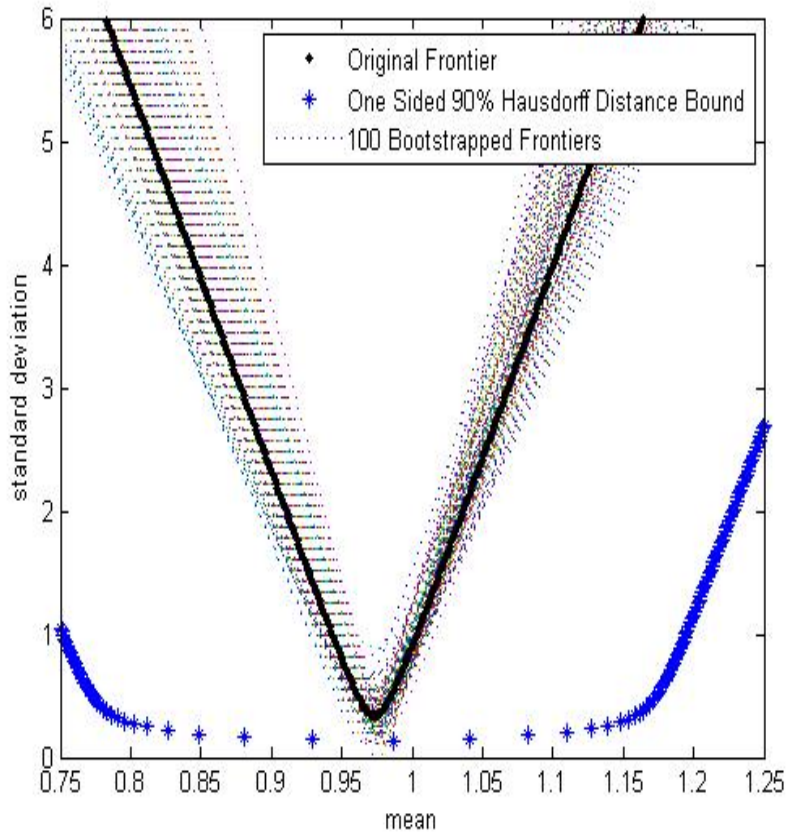

FIGURE 5. 90% Confidence Region using Unweighted W Statistic (H-Distance)

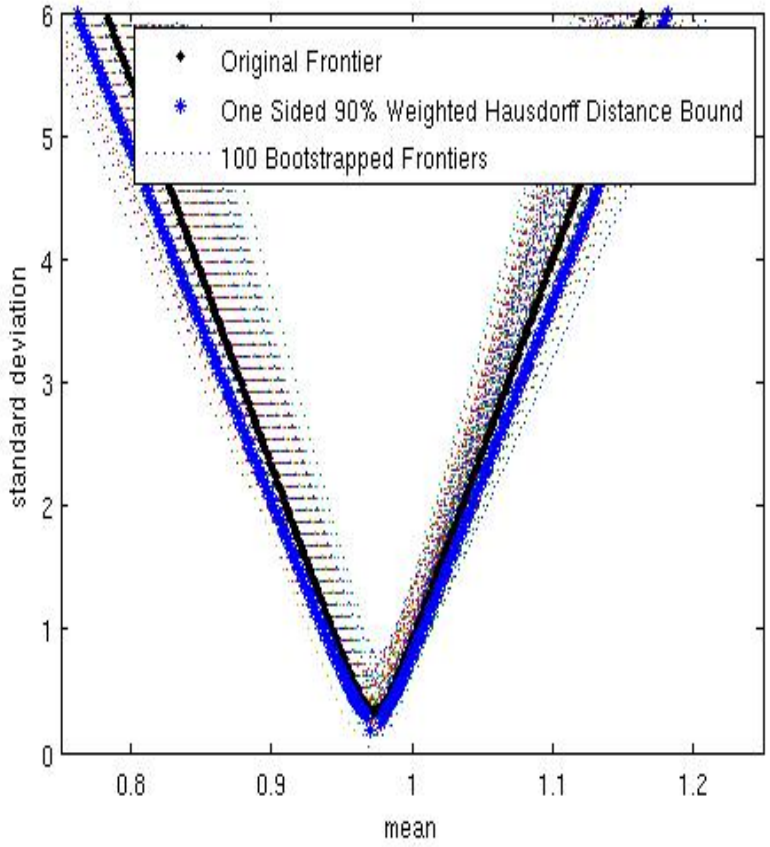

FIGURE 6. 90% Confidence Region using Weighted W Statistic