Inference on sets in finance

Victor Chernozhukov
MIT

Emre Kocatulum
DRW Holdings, LLC

Konrad Menzel
New York University

We consider the problem of inference on a class of sets describing a collection of admissible models as solutions to a single smooth inequality. Classical and recent examples include the Hansen–Jagannathan sets of admissible stochastic discount factors, Markowitz–Fama mean–variance sets for asset portfolio returns, and the set of structural elasticities in Chetty's (2012) analysis of demand with optimization frictions. The econometric structure of the problem allows us to construct convenient and powerful confidence regions based on the weighted likelihood ratio and weighted Wald statistics. Our statistics differ from existing statistics in that they enforce either exact or first-order equivariance to transformations of parameters, making them especially appealing in the target applications. We show that the resulting inference procedures are more powerful than the structured projection methods. Last, our framework is also useful for analyzing intersection bounds, namely sets defined as solutions to multiple smooth inequalities, since multiple inequalities can be conservatively approximated by a single smooth inequality. We present two empirical examples showing how the new econometric methods are able to generate sharp economic conclusions.

Keywords. Hansen–Jagannathan bound, Markowitz–Fama bounds, Chetty bounds, mean–variance sets, optimization frictions, inference, confidence set.

JEL classification. C10, C50.

1. Introduction

In this paper, we consider the problem of inference on a class of sets describing a collection of admissible models as solutions to a single smooth inequality,

$$\Theta_0 = \big\{ \theta \in \Theta : m(\theta) \le 0 \big\}, \tag{1.1}$$

where $\Theta \subset \mathbb{R}^k$ is a compact parameter space and $m : \mathbb{R}^k \mapsto \mathbb{R}$ is the inequality-generating smooth function, which is estimable from the data. This structure arises in a number of important applications in financial economics and public finance, including the Hansen–Jagannathan (HJ) (1991) set for the stochastic discount factor (SDF), Markowitz–Fama (MF) sets for portfolio returns, and bounds on structural demand elasticities in the presence of optimization frictions developed by Chetty (2012). We describe these examples in some detail below in this section.

We show that this structure leads to highly tractable and powerful inference procedures, and demonstrate their usefulness in substantive empirical examples. Furthermore, the structure could be used to conservatively approximate more complicated problems, where the sets of interest are given as intersections of solutions to multiple smooth inequalities. Indeed, for the latter case, we propose a procedure for approximating the multiple inequalities conservatively by a single smooth inequality. The benefit from doing so is the highly tractable "regular" inference.

The basic problem of this paper is to develop inference methods on sets defined by nonlinear inequality restrictions while accounting for uncertainty in the estimation of parameters of the inequality-generating functions. Specifically, the problem is to construct a confidence region R such that

$$\lim_{n \to \infty} \mathrm{P}\{\Theta_0 \subseteq R\} = 1 - \alpha, \qquad d_{\mathrm{H}}(\Theta_0, R) = O_p(1/\sqrt{n}),$$

where d_{H} is the Hausdorff distance, defined below in (A.4). We construct confidence regions for the set Θ_0 using likelihood ratio (LR) and Wald-type statistics, building on and simultaneously enriching the approaches suggested in Chernozhukov, Hong, and Tamer (2007) and in Beresteanu and Molinari (2008) and Molchanov (1998), respectively. We also would like to ensure that confidence regions R are as small as possible and converge to Θ_0 at the fastest attainable speed.

Once R is constructed, we can test any composite hypotheses involving the parameter θ without compromising the significance level. For example, a typical application of the HJ bounds determines which combinations of Θ_e the first two moments of the SDF generated by a given family of economic models fall in the HJ set. Indeed, in that case $\Theta_e = \{\mu(\varrho), \sigma(\varrho), \varrho \in [0, \infty)\}$, where ϱ is the elasticity of power utility function, and we can check which values of ϱ give us overlap of Θ_e with the confidence set R for the HJ set.[1] Similar comments about applicability of our approach go through for the Markowitz and MF sets as well. We should also mention that our confidence regions R

---

[1]In cases where curve $\varrho \mapsto \mu(\varrho)$, $\varrho \mapsto \sigma(\varrho)$ are estimated, we need to construct a confidence set R_e for Θ_e and then examine whether (and where) R_e and R overlap to determine the plausible values of ϱ.

are also more powerful than the structured projection methods, which rely on building confidence sets for the frontier-determining sufficient parameters (e.g., frontier-spanning portfolios), and then projecting them to obtain confidence sets for HJ sets or MF sets. We demonstrate this in the empirical section, where we compute the confidence sets for HJ sets based on our approach and based on the structured projection approach, and we show that our confidence region is much smaller and lies strictly inside the confidence region based on structured projections. We provide a theoretical explanation for this phenomenon in Section 2.5.

1.1 *Examples*

EXAMPLE 1 (Mean–Variance Set for the Stochastic Discount Factor). So as to describe the problem, we recall Cochrane's (2005) assertion that the science of asset pricing could be effectively summarized using the two equations

$$P_t = E_t[M_{t+1}X_{t+1}],$$

$$M_{t+1} = f(Z_{t+1}, \text{parameters}),$$

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, i.e., a set that is not a singleton. SDFs are not unique as long as 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 the 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. Note that the investor's optimal consumption plan, and therefore the marginal rate of substitution, generally depends on a set of additional state variables Z_t, including lifetime wealth and noncapital income. Hence, when markets are incomplete, asset prices alone do not, in general, pin down a unique stochastic discount factor across different aggregate states of the economy or across investors with different values of Z_t.

An important empirical problem in the context of the Capital Asset Pricing Model (CAPM) is to check which families of SDFs price the assets correctly and which do not. This reasoning forms the basis for many approaches to estimation and tests for particular specifications of an asset pricing model, most prominently Hansen and Singleton's (1982) seminal analysis of the consumption-based CAPM; see also Ludvigson (2012) for a survey of recent developments on the subject. One leading approach to performing the check is to see whether the mean and standard deviation of SDFs (μ, σ) of the stochas-

tic discount factor M_t are admissible. Let K be a compact convex body in $\mathbb{R}^2$. The set of admissible means and standard deviations in K,

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

is introduced by Hansen and Jagannathan (1991) and is known as the Hansen–Jagannathan (HJ) set. The boundary of that set $\partial\Theta_0$ is known as the HJ bound. To describe the admissible pairs in a canonical setting, let v and Σ denote the vector of mean returns and covariance matrix to assets $1, \ldots, N$, which are assumed not to vary with information sets at each period t. In the following discussion, we consider the "minimal sufficient parameters"

$$S_{vv} = v'\Sigma^{-1}v, \qquad S_{v1} = v'\Sigma^{-1}1_N, \qquad S_{11} = 1_N'\Sigma^{-1}1_N, \tag{1.2}$$

where 1_N is a column vector of 1s. Then the minimum variance $\sigma_{\text{HJ}}^2(\mu)$ achievable by a SDF given mean μ of the SDF is

$$\sigma_{\text{HJ}}^2(\mu) = (1 - \mu v)'\Sigma^{-1}(1 - \mu v) = S_{vv}\mu^2 - 2S_{v1}\mu + S_{11}.$$

Therefore, the HJ set is

$$\Theta_{\text{HJ}} = \big\{\underbrace{(\mu, \sigma)}_{\theta} \in \underbrace{(\mathbb{R} \times \mathbb{R}_+) \cap K}_{\Theta} : \underbrace{\sigma_{\text{HJ}}(\mu) - \sigma \leq 0}_{m(\theta)}\big\};$$

that is,

$$\Theta_0 = \{\theta \in \Theta : m_{\text{HJ}}(\theta) \leq 0\}.$$

Note that the inequality-generating function $m_{\text{HJ}}(\theta)$ depends on the unknown means μ and the covariance of returns Σ via the minimal sufficient parameters, $\gamma = (S_{vv}, S_{v1}, S_{11})'$, so that we can write $m_{\text{HJ}}(\theta) = m_{\text{HJ}}(\theta, \gamma)$. The reasoning behind the derivation can be extended to conditional HJ-type bounds given any observed variables in the investors' information set. These bounds are, in general, sharper than the unconditional bounds described in this example and could be treated similarly to the original HJ bounds, forming intersections of conditional admissible sets.

EXAMPLE 2 (Mean–Variance Analysis of Asset Portfolios). Let us now describe the second problem. The classical Markowitz (1952) problem is to minimize the variance of a portfolio given some attainable level of return

$$\min_{w} \mathrm{E}_t\big[r_{p,t+1} - \mathrm{E}_t[r_{p,t+1}]\big]^2 \quad \text{such that} \quad \mathrm{E}_t[r_{p,t+1}] = \mu,$$

where $r_{p,t+1}$ is the return of the portfolio, determined as $r_{p,t+1} = w'r_{t+1}$, and 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 the covariance of returns Σ do not vary with time period t, so that the problem becomes

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

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

$$\sigma_M^2(\mu) = \frac{S_{11}\mu^2 - 2S_{v1}\mu + S_{vv}}{S_{vv}S_{11} - S_{v1}^2},$$

where the minimal sufficient parameters $\gamma = (S_{vv}, S_{v1}, S_{11})'$ are the same as in (1.2).

Therefore, given a compact convex subset $K \subset \mathbb{R}^2$, the Markowitz (M) set of admissible standard deviations and means in K is given by

$$\Theta_M = \{(\mu, \sigma) \in (\mathbb{R} \times \mathbb{R}_+) \cap K : \sigma_M(\mu) - \sigma \leq 0\} = \{\theta \in \Theta : m_M(\theta) \leq 0\},$$

where $m_M(\theta) := \sigma_M(\mu) - \sigma$. The boundary of the set Θ_M is known as the efficient frontier. Note that as in the HJ example, the inequality-generating function $m_M(\theta) = m_M(\theta, \gamma)$ depends on the unknown parameters—the means and covariance of returns, $\gamma = (S_{vv}, S_{v1}, S_{11})'$.

On the other hand, a test for efficiency of a given feasible mean–variance combination of portfolio returns can be derived from the complementary Markowitz set

$$\Theta_M^c = \{\theta \in \Theta : -m_M(\theta) \leq 0\}.$$

Noting that any feasible portfolio is in the Markowitz set by construction, a test whether the first two moments of returns for a portfolio belong to Θ_M^c is a one-sided analog of a (two-sided) test whether the portfolio attains the efficient mean–variance frontier and, therefore, is more powerful than the latter. Inference for this problem is, therefore, also covered by our framework in (1.1).

EXAMPLE 3 (Multifactor Efficient Frontiers). A third example due to Fama (1996) extends the unconditional portfolio choice problem in Markowitz (1952) to multiple sources of priced risk. More specifically, suppose that there exist k state variables $Z_t = (z_{1t}, \ldots, z_{kt})'$ ("factors") such that $(r_t', Z_t')'$ are jointly normal (or, more generally, jointly spherical). We assume that Z_t follows a k-variate standard normal (or other spherical) distribution and that

$$E_t[r_{t+1}|Z_{t+1}] = v + B'Z_{t+1},$$

where v is an N-dimensional vector of average returns, and the $k \times N$ covariance matrix $B = [\beta_1, \ldots, \beta_N] = \mathrm{Cov}_t(Z_{t+1}, r_{t+1})$ gives the factor loadings for the vector of asset returns. As in the previous examples, we denote the (unconditional) mean of asset returns by $v \in \mathbb{R}^N$ and denote their covariance matrix with Σ. Note that for the purposes of this example, we take r_{it} to denote the excess return of asset i relative to a risk-free asset with return r_f.

This analysis is motivated by the investor's lifetime consumption–investment problem, where the state variables Z_t denote other factors influencing prices for consumption goods and the household's income from wage labor or entrepreneurial activity in period t. In particular, the household not only faces a trade-off between mean and variance of portfolio returns, but may also wish to use their portfolio to insure against price or income shocks.

Specifically, we consider the problem of constructing a minimum variance portfolio that targets a mean return μ^* and a vector β^* of loadings with respect to the factors Z_t in the absence of borrowing constraints. Denoting $D := [v, B']$ and $\delta^* := (\mu^*, \beta^{*'})'$, the efficient portfolio weights w can be found by solving the program

$$\min_{w} w'\Sigma w \quad \text{s.t.} \quad D'w = \delta^*. \tag{1.3}$$

Note that since the investor has access to the risk-free asset and faces no short-sale constraints, the portfolio weights w are not required to be nonnegative or to sum to 1. Solving this program yields efficient portfolio weights $w^* = \Sigma^{-1}D(D'\Sigma D)^{-1}\delta^*$ so that the lower bound on the variance of a portfolio with beta equal to β and mean μ is given by

$$\sigma^2_{\text{MF}}(\mu, \beta) := (w^*)'\Sigma w^* = \delta^{*'}(D'\Sigma D)^{-1}\delta^*.$$

The resulting multifactor efficient frontier bounds the cone

$$\Theta_{\text{MF}} := \left\{ (\mu^*, \beta^{*'}, \sigma)' \in (\mathbb{R}^{k+1} \times \mathbb{R}_+) \cap K : \sigma_{\text{MF}}(\mu^*, \beta^*) - \sigma \le 0 \right\}$$

of feasible mean–variance–beta combinations, where K is some compact convex body in $\mathbb{R}^{k+2}$. Again, this cone can also be characterized by an inequality condition on the moment function

$$m_{\text{MF}}(\theta, \gamma) := \sqrt{(\mu^*, \beta^{*'})(D'\Sigma D)^{-1}(\mu^*, \beta^{*'})'} - \sigma,$$

where $\theta = (\mu^*, \beta^{*'}, \sigma)'$ and $\gamma = \text{vec}(D'\Sigma D)$.

Extending the two-fund theorem, Fama (1996) shows that we can always construct $k + 2$ different funds from the N assets that span the entire multifactor efficient frontier. In the presence of heterogeneity in preferences—wealth and consumption possibilities—different investors may typically choose different points on the efficient cone, so that only the entire set constitutes an adequate representation of the opportunities to insure against different factor risks in the market.

EXAMPLE 4 (Bounds on Demand Elasticities). To illustrate the wider applicability of our methods, we also consider inference on demand elasticities in the presence of an optimization friction. Chetty (2012) considers a household's dynamic labor supply and consumption problem, and proposes bounds on price elasticities that allow for optimization errors of unknown form but that are bounded in magnitude. His analysis assumes that preferences can be represented by quasilinear flow utilities and lead to Marshallian demand functions of the form

$$\log x^*_{it}(p) = \alpha - \varepsilon \log p + \nu_{it},$$

where p is the price of the good of interest and ν_{it} is a preference shock for household i in period t. Here ε is the (constant) "structural" demand elasticity corresponding to the first-best solution to the household's planning problem.

Let U^*_{it} be the discounted lifetime utility from following the optimal consumption plan and let $e_{it}(x)$ denote the minimal expenditure needed to finance a utility level U^*_{it}

under the constraint that x_{it} is held fixed at x. The analysis then considers "small" optimization errors in the households' observed choice of x_{it} such that for every period t, the average utility loss measured in terms of the difference in expenditure as a percentage of the optimal budget is bounded by some $\delta > 0$, i.e.,

$$\frac{1}{N} \sum_{i=1}^{N} \frac{e_{it}(x_{it}) - e_{it}(x_{it}^*(p_t))}{p_t x_{it}^*} \le \delta,$$

where the average is taken over a sample of households $i = 1, \ldots, N$.

Chetty (2012) derives the lower and upper bounds ε_L and ε_U for the structural elasticities that are compatible with an observed elasticity ε^o given a price change $\Delta \log p$ and in the presence of an optimization error $\delta > 0$. This bound is given by the two equalities

$$\varepsilon^o = \varepsilon_U - 2\frac{(2\varepsilon_U \delta)^{1/2}}{\Delta \log p} \quad \text{and} \quad \varepsilon^o = \varepsilon_L + 2\frac{(2\varepsilon_L \delta)^{1/2}}{\Delta \log p},$$

where δ is assumed to be small.[2] With a few basic manipulations, we can obtain the following equivalent characterization of the (δ, ε) set as a lower bound for the distortion needed to reconcile ε^o with a structural elasticity ε:

$$\delta \ge \frac{(\varepsilon - \varepsilon^o)^2 (\Delta \log p)^2}{8\varepsilon} =: \delta_{OF}(\varepsilon, \varepsilon^o, \Delta \log p) \quad \text{if and only if} \quad \varepsilon \in [\varepsilon_L, \varepsilon_U]. \quad (1.4)$$

Chetty (2012) gives several interpretations for this optimization friction, which could result, e.g., from adjustment costs or misperception of prices. In the context of the consumption–CAPM model, Lettau and Ludvigson (2009) conducted a similar sensitivity analysis, allowing for optimization errors in terms of the Euler equations characterizing the household's optimal investment decision.

The resulting set of structural elasticity/optimization friction pairs (ε, δ) compatible with the observed elasticity ε^o is given by

$$\Theta_{OF} := \{(\varepsilon, \delta) \in \mathbb{R}_+^2 : \delta_{OF}(\varepsilon, \varepsilon^o, \Delta \log p) - \delta \le 0\}.$$

The bounds in (1.4) can, therefore, also be expressed as an inequality condition on the moment function

$$m_{OF}(\theta, \gamma) := \frac{(\varepsilon - \varepsilon^o)^2 (\Delta \log p)^2}{8\varepsilon} - \delta,$$

where $\theta = (\varepsilon, \delta)$ and $\gamma = \varepsilon^o$, and $\Delta \log p$ serves as a fixed design parameter. In applications, one might be interested in intersection bounds from several studies so as to obtain tighter bounds, as was done in Chetty (2012). As we discuss below and illustrate in the empirical example, it is possible to perform this intersection in a smooth fashion, retaining the general single-equation structure as given in (1.1).

---

[2]See Chetty (2012, pp. 983–985, eqs. (12) and (13)).

1.2 *Related literature*

Our results extend and complement other existing procedures for set inference. Chernozhukov, Hong, and Tamer (2007) developed criterion-based methods for inference under fairly general conditions. Our approach based on weighted directed Hausdorff distance exploits more closely the structure of the econometric problem at hand and is substantially easier to implement in practice, in particular if the researcher is mainly interested in inference on a subvector of the parameter θ.

Set inference based on the directed Hausdorff distance was previously suggested by Molchanov (1998) and Beresteanu and Molinari (2008). By using weighting in the construction of the Wald-type statistics, we make this approach asymptotically invariant to parameter transformations, which results in noticeably sharper confidence sets, at least in the canonical empirical example given below. These invariance properties should be seen as complementary to questions of efficient estimation of the identified set as studied by Kaido and Santos (2014); indeed, many types of confidence sets are centered around the same efficient set estimate, so we use invariance and precision considerations to select among various potential constructions of sets. Furthermore, our results on formal validity of the bootstrap for the weighted LR-type and Wald-type statistics are also of independent interest (for the class of problems with structures that are similar to those studied here, see Beresteanu and Molinari (2008), and see Kaido and Santos (2014) for related bootstrap results for Hausdorff statistics).

Critchley, Marriott, and Salmon (1996) analyzed the geometric structure of Wald-type tests with nonlinear restrictions for point-identified problems and pointed out the beneficial role of invariance of the distribution of the statistic with respect to reparameterizations. In regular problems, the effect of nonlinearities on inference is typically asymptotically negligible due to the delta method and its extensions. However, our results show that for inference on sets with a nontrivial diameter, the effect of nonlinearities is of the same order as that of sampling variation. In particular, we propose inference procedures that rely on quantities that are asymptotically pivotal and (asymptotically) invariant to parameter transformation, and we illustrate that procedures failing to meet these requirements lead to overly conservative inference. In particular, these conditions are not met by the classical Hausdorff distance except in very special cases.

Regarding the applications, mean–variance analysis for asset returns and the stochastic discount factor provides powerful tools to summarize the opportunities for risk diversification in a given asset market, and can serve as a basis for spanning and efficiency tests. Cochrane and Hansen (1992) illustrate the usefulness of the HJ bounds for analyzing consumption and asset market data under various modeling assumptions, including various extensions of standard asset pricing models. The importance of sampling error and the role of statistical inference in portfolio analysis has long been recognized in the literature: pointwise inference for the HJ bounds was first developed by Cecchetti, Lam, and Mark (1994) and Hansen, Heaton, and Luttmer (1995), who consider specification tests based on the Euclidean distance of a given SDF to the HJ set; Hansen, Heaton, and Luttmer (1995) and Luttmer (1996) extend the model to allow for

short-sale constraints and transaction costs; Hansen, Heaton, and Luttmer (1995) also establish consistency of the estimated HJ bound and Luttmer (1996) proposes methods for pointwise inference on the bound; Burnside (1994) develops methods for pointwise inference using the vertical distance of a SDF to the HJ frontier and shows how to invert that test to obtain confidence regions for structural parameters (e.g., the discount rate and the elasticity of intertemporal substitution). Our paper considers inference on the bound that is uniform in that the proposed confidence regions cover the entire admissible set with the prespecified probability, which amounts to controlling a familywise error rate for all null hypotheses concerning points in the HJ set. Our formal results also apply to inference on preference parameters (as developed by Burnside (1994)) and various extensions of the HJ framework. However, we use the original HJ bounds as our leading empirical example to illustrate the technical ideas in the paper and the potential for applications.

For the Markowitz set, Gibbons, Ross, and Shanken (1989) develop regression-based efficiency tests, and Britten-Jones (1999) considers inference on the tangency portfolio and tests for portfolio efficiency. Recent work by Peñaranda and Sentana (2010) combines restrictions derived from the two dual approaches to mean–variance analysis so as to derive more powerful spanning tests for (sub-) markets. A summary of the recent literature on mean–variance efficiency tests is provided by Sentana (2009). We show that the problem of constructing confidence sets, and, therefore, the econometric contribution of this paper, is quite distinct from and hence complementary to these previous efforts. Specifically, we find that statistical properties of confidence sets are more sensitive to the choice of the underlying test statistic than are pointwise hypothesis tests, and we show how to construct confidence sets satisfying certain optimality criteria.

NOTATION. Our formal arguments generalize in a straightforward manner to asymptotics under triangular arrays, where the data generating process $P_n \in \mathbf{P}$, a common set of probability measures, which translates into statements that are uniform in $\mathbf{P}$. However, to keep the focus on the main issues, we shall suppress the dependence of P_n on n in what follows. Moreover, the set inference methods (as opposed to doing inference on points) does not suffer from the same problems.[3]

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. Results on inferential validity and consistency for the inference procedure are relegated to Appendix A. In the remainder of the Appendix, we collect the proofs of the main results. Additional data are available in a supplementary file on the journal website, http://qeconomics.org/supp/387/code_and_data.zip.

---

[3]See, e.g., Romano and Shaikh (2010), Andrews and Guggenberger (2009), and Andrews and Soares (2010) for uniformity results in pointwise inference.

2. Estimation and inference results

2.1 *Basic constructions*

We first introduce our basic framework. We have a real-valued inequality-generating function $m(\theta)$, and the set of interest is the solution of the inequalities generated by the function $m(\theta)$ over a compact parameter space $\Theta \subset \mathbb{R}^k$:

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

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 an estimate of the inequality-generating function. For example, in the HJ and M examples, the estimate takes the form

$$\widehat{m}(\theta) = m(\theta, \widehat{\gamma}), \qquad \widehat{\gamma} = (\widehat{S}_{vv}, \widehat{S}_{v1}, \widehat{S}_{11})'.$$

Note that although this formulation only allows for one single inequality restrictions, we provide a method of incorporating smoothed intersection bounds into our framework in Section 2.5.

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

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

where $\widehat{s}(\theta)$ is a weighting function, ideally, the standard deviation of $\widehat{m}(\theta)$, and $\widehat{k}_{\mathrm{L}}(1-\alpha)$ is some suitable estimate of $k_{\mathrm{L}}(1-\alpha)$, the $(1-\alpha)$ quantile of the statistic

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

Note that $\mathcal{L}_n$ is a LR-type statistic, as in Chernozhukov, Hong, and Tamer (2007).

Next we shall consider confidence regions based on inverting a Wald-type statistic. So as to eliminate boundary effects on the distribution of this statistic, we need to assume that the moment conditions, and, therefore, the set estimator, are well defined in a neighborhood of the parameter space. To this end, we define the distance of a point θ to a set $A \subset \mathbb{R}^k$ as

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

We let the set

$$\Theta^\delta := \{\theta \in \mathbb{R}^k : d(\theta, \Theta) \leq \delta\}$$

denote the δ-expansion of Θ in $\mathbb{R}^k$ for any $\delta > 0$. We denote the natural set estimator for Θ_0 in the expansion of the parameter space by

$$\widehat{\Theta}_{0,\delta} := \{\theta \in \Theta^\delta : \widehat{m}(\theta) \leq 0\}.$$

We show below that the behavior of the moment function near the boundary of the parameter space is, in general, relevant for the statistical behavior of our procedure. However, for the practical applications discussed in this paper, the restriction of the parameter space Θ to a compact set is entirely for technical reasons, so that an extension of the definitions to Θ^δ is unproblematic.

Given these definitions, our Wald-based confidence region is

$$R_W := \left\{ \theta \in \Theta : \left[\sqrt{n} d(\theta, \widehat{\Theta}_{0,\delta}) / \widehat{w}(\theta) \right]^2 \leq \widehat{k}_W(1 - \alpha) \right\}, \tag{2.3}$$

where $\widehat{w}(\theta)$ is a weighting function, particular forms of which we will suggest later, and $\widehat{k}_W(1 - \alpha)$ is a suitable estimate of $k_W(1 - \alpha)$, the $(1 - \alpha)$ quantile of $\mathcal{W}_n$, where $\mathcal{W}_n$ is the weighted W-statistic

$$\mathcal{W}_n := \sup_{\theta \in \Theta_0} \left[\sqrt{n} d(\theta, \widehat{\Theta}_{0,\delta}) / \widehat{w}(\theta) \right]^2. \tag{2.4}$$

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

$$\sqrt{\mathcal{W}_n} \propto d^+(\Theta_0, \widehat{\Theta}_{0,\delta}) := \sup_{\theta \in \Theta_0} \inf_{\theta' \in \widehat{\Theta}_{0,\delta}} \|\theta - \theta'\|.$$

REMARK 1 (Parameter Restrictions and Subvector Inference). As already pointed out by Beresteanu and Molinari (2008), inference based on the directed Hausdorff distance can be easily adapted to problems concerning a proper subvector $\theta_1 \in \Theta_1$ of $\theta = (\theta_1', \theta_2')'$. In our setup this can be done by replacing the distance $d(\cdot, \cdot)$ in the definition of the weighted Wald statistic $\mathcal{W}_n$ by a pseudometric $\tilde{d}(\cdot, \cdot)$ that puts no weight on deviations in subspaces of Θ corresponding to the components θ_2, i.e., for some metric $\tilde{d}_1(\cdot, \cdot)$ on Θ_1 we have

$$\tilde{d}(\theta, \tilde{\theta}) \equiv \tilde{d}\left((\theta_1', \theta_2')', (\tilde{\theta}_1', \tilde{\theta}_2')' \right) = \tilde{d}_1(\theta_1, \tilde{\theta}_1)$$

for all θ, $\tilde{\theta}$. Inference for other affine subspaces of Θ can be implemented in an analogous manner. Restrictions on the parameter space or the metric $d(\cdot, \cdot)$ imply that the supremum in the definition of $\mathcal{W}_n$ is taken over a smaller, or lower dimensional subset, resulting in less conservative critical values for tests and construction of confidence sets.

REMARK 2 (Invariance Motivation for Weighted W-Statistics). The weighted W-statistic (2.4) is generally not a distance, but we argue in Section 2.3 that it provides a very useful extension/generalization of the canonical directed Hausdorff distance. Note that, like any Euclidean norm, the Hausdorff distance is not invariant with respect to changes in the scale of the different components of θ, which results in confidence regions that are not equivariant to such transformations. Our empirical results below illustrate that as a result, the shape of confidence sets alters dramatically as we change the weight on different dimensions of the parameter space. In sharp contrast, the introduction of

certain types of weights makes the W-statistic (first-order) invariant to such transformations, as shown in Section 2.3, which makes the resulting confidence regions (first-order) equivariant. As a result, in all of our empirical examples, such weighting dramatically improves the confidence regions (e.g., compare Figures 5 and 6 in the example concerning inference on HJ sets).

REMARK 3 (Inference Based on the Support Function). As shown in Beresteanu and Molinari (2008), there is a close correspondence between inference based on the support function for the (convex hull of the) identified set and inference based on the directed Hausdorff distance. An inferential theory based on a weighted support function was developed in a more recent paper by Chandrasekhar, Chernozhukov, Molinari, and Schrimpf (2012), and under technical regularity conditions our approach based on the weighted directed Hausdorff distance can be shown to be asymptotically equivalent.

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 maintain the following general regularity conditions:

CONDITIONS R. For some $\delta > 0$, the inequality-generating function $m : \Theta^\delta \to \mathbb{R}$ and its estimator $\widehat{m} : \Theta^\delta \to \mathbb{R}$ are well defined.

R.1. The estimator $\widehat{m}$ is $n^{-1/2}$-consistent for m and asymptotically Gaussian, namely, in the metric space of bounded functions $\ell^\infty(\Theta^\delta)$,

$$\sqrt{n}(\widehat{m} - m) \rightsquigarrow \mathbb{G},$$

where $\mathbb{G}$ is a zero-mean Gaussian process with continuous paths a.s., and a nondegenerate covariance function. Specifically, we let $s^2(\theta) := \mathrm{E}[\mathbb{G}^2(\theta)]$ and assume that $\inf_{\theta \in \Theta^\delta} s^2(\theta) > 0$.

R.2. Functions $\widehat{m}$ and m admit continuous gradients $\nabla_\theta \widehat{m} : \Theta^\delta \mapsto \mathbb{R}^k$ and $\nabla_\theta m : \Theta^\delta \mapsto \mathbb{R}^k$ with probability 1, where

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

uniformly in $\theta \in \Theta^\delta$. The gradient $\theta \mapsto \nabla_\theta m$ is uniformly Lipschitz and

$$\inf_{\theta \in \Theta^\delta} \left\| \nabla_\theta m(\theta) \right\| > 0.$$

R.3. Weighting functions $\widehat{s} : \Theta^\delta \mapsto \mathbb{R}$ and $\widehat{w} : \Theta^\delta \mapsto \mathbb{R}$ satisfy, uniformly in $\theta \in \Theta^\delta$,

$$\widehat{s}(\theta) = s(\theta) + o_p(1), \qquad \widehat{w}(\theta) = w(\theta) + o_p(1),$$

where $s : \Theta^\delta \mapsto \mathbb{R}_+$ and $w : \Theta^\delta \mapsto \mathbb{R}_+$ are continuous functions with values bounded away from zero.

In Condition R.1, we require the estimates of the inequality-generating functions to satisfy a uniform central limit theorem. Many sufficient conditions for this are provided

by the theory of empirical processes; see, e.g., van der Vaart and Wellner (1996). In our finance examples, this condition follows from asymptotic normality of the estimates of the mean returns and covariance of returns. In Condition R.2, we require that the gradient of the estimate of the inequality-generating function be consistent for the gradient of the inequality-generating function. Moreover, we require that the norm $\|\nabla_\theta m(\theta)\|$ be bounded away from zero, which is an identification condition and allows us to estimate, at a parametric rate, 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 be consistent, and the weight functions be well behaved.

   Under these conditions, we can state the following general result.

THEOREM 1 (Limit Laws of LR and W Statistics). *Under Conditions* R,

$$\mathcal{L}_n \rightsquigarrow \mathcal{L}, \quad \mathcal{L} = \sup_{\theta \in \partial\Theta_0} \left[\frac{\mathbb{G}(\theta)}{s(\theta)}\right]_+^2, \tag{2.5}$$

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

where both $\mathcal{W}$ *and* $\mathcal{L}$ *have distribution functions that are continuous at their* $(1-\alpha)$ *quantiles for* $\alpha < 1/2$. *Furthermore, if*

$$\widehat{w}(\theta) = \frac{s(\theta)}{\|\nabla_\theta m(\theta)\|} + o_p(1)$$

uniformly in $\theta \in \Theta^\delta$, *then the two statistics are asymptotically equivalent:*

$$\mathcal{W}_n = \mathcal{L}_n + o_p(1).$$

   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 $\mathbb{G}$. Moreover, we see that under an appropriate choice of the weighting functions, the two statistics are asymptotically equivalent.

   For our application to HJ and MF sets, the following conditions are sufficient.

CONDITIONS C.

   C.1. We have that $m(\theta) = m(\theta, \gamma_0)$, where $\gamma_0 \in \Gamma \subset \mathbb{R}^d$ for all $\theta \in \Theta^\delta$ and some $\delta > 0$. The value γ_0 is in the interior of Γ, and there is an estimator $\widehat{\gamma}$ of γ_0 that obeys

$$\sqrt{n}(\widehat{\gamma} - \gamma_0) \rightsquigarrow \Omega^{1/2}Z, \quad Z = N(0, I_d)$$

for some positive-definite Hermitian matrix Ω. Moreover, there is $\widehat{\Omega} \to_p \Omega$.

   C.2. The gradient map $(\theta, \gamma) \mapsto \nabla_\theta m(\theta, \gamma)$, mapping $\Theta^\delta \times \Gamma \mapsto \mathbb{R}^k$, exists and is uniformly Lipschitz-continuous. Moreover, $\inf_{(\theta,\gamma)\in\Theta^\delta\times\Gamma} \|\nabla_\theta m(\theta, \gamma)\| > 0$.

C.3. The gradient map $(\theta, \gamma) \mapsto \nabla_\gamma m(\theta, \gamma)$, mapping $\Theta^\delta \times \Gamma \mapsto \mathbb{R}^k$, exists and is uniformly Lipschitz-continuous.

We show in Proposition 3 in Appendix B that these conditions hold for the canonical versions of the HJ and MF problems, and Chetty's (2012) bounds for demand elasticities. Under these conditions, we immediately conclude that in the metric space of bounded functions $\ell^\infty(\Theta^\delta)$,

$$\sqrt{n}\big(\widehat{m}(\cdot) - m(\cdot)\big) = \nabla_\gamma m(\cdot, \bar\gamma)' \sqrt{n}(\widehat{\gamma} - \gamma_0) + o_p(1)$$
$$\rightsquigarrow \nabla_\gamma m(\cdot, \gamma_0)' \Omega^{1/2} Z,$$

where $\nabla_\gamma m(\theta, \bar\gamma)$ denotes the gradient with each of its rows evaluated at a value $\bar\gamma$ on the line connecting $\widehat{\gamma}$ 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

$$\mathbb{G}(\theta) = \nabla_\gamma m(\theta, \gamma_0)' \Omega^{1/2} Z. \tag{2.7}$$

More formally, we show below that conclusions of Theorem 1 hold with

$$\mathcal{L} = \sup_{\theta \in \partial\Theta_0} \left[\frac{\nabla_\gamma m(\theta, \gamma_0)' \Omega^{1/2}}{s(\theta)} Z \right]_+^2, \tag{2.8}$$

$$\mathcal{W} = \sup_{\theta \in \partial\Theta_0} \left[\frac{\nabla_\gamma m(\theta, \gamma_0)' \Omega^{1/2}}{\|\nabla_\theta m(\theta, \gamma_0)\| \cdot w(\theta)} Z \right]_+^2. \tag{2.9}$$

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

$$s(\theta) = \big\|\nabla_\gamma m(\theta, \gamma_0)' \Omega^{1/2}\big\|, \tag{2.10}$$

$$w(\theta) = \frac{\|\nabla_\gamma m(\theta, \gamma_0)' \Omega^{1/2}\|}{\|\nabla_\theta m(\theta, \gamma_0)\|}. \tag{2.11}$$

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

$$\widehat{s}(\theta) := \big\|\nabla_\gamma m(\theta, \widehat{\gamma})' \widehat{\Omega}^{1/2}\big\|, \tag{2.12}$$

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

We formalize the preceding discussion as the following corollary.

COROLLARY 1 (Limit Laws of LR and W Statistics Under Conditions C). *Under Conditions C, Conditions R hold with the limit Gaussian process stated in equation* (2.7). *The plug-in estimates of the weighting functions* (2.12) *and* (2.13) *are uniformly consistent for the weighting functions* (2.10) *and* (2.11). *Therefore, the conclusions of Theorem 1 hold with the limit laws for our statistics given by the laws of random variables stated in* (2.8) *and* (2.9).

Given this asymptotic characterization of the laws of $\mathcal{L}$ and $\mathcal{W}$, we can obtain consistent estimates for the respective $1 - \alpha$ percentiles, $\hat{k}_W(1 - \alpha)$ and $\hat{k}_L(1 - \alpha)$. It is then relatively standard to establish asymptotic coverage properties and root-n consistency for confidence regions of the form (2.1) and (2.3). Procedures for obtaining critical values $\hat{k}(1 - \alpha)$ via bootstrap or simulation, and formal results establishing large sample properties of the confidence regions are given in Appendix A.

2.3 *Invariance and similarity properties of confidence regions based on LR and Wald statistics*

We next proceed to state the invariance properties of the proposed inference procedures with respect to parameter transformations. We distinguish between an exact invariance and an asymptotic invariance. A parameter transformation is a one-to-one mapping $\eta : \Theta \to Y$, where $Y = \eta(\Theta)$, and we denote the population and sample moment conditions for the transformed problem by $m_\eta(\cdot) := m(\eta^{-1}(\cdot))$ and $\widehat{m}_\eta(\cdot) := \widehat{m}(\eta^{-1}(\cdot))$, respectively, which are mappings from $Y \to \mathbb{R}$. We define H^* as the set of parameter transformations η that are continuously differentiable in $\theta \in \Theta^\delta$ and such that Conditions R.1–R.3 hold for the transformed moments $m_\eta(\cdot)$.

In this section, we discuss invariance properties of inference based on the LR and W statistics. Let

$$\mathcal{L}_n(\theta) := \left[\sqrt{n}\widehat{m}_n(\theta)/\widehat{s}(\theta)\right]^2_+ \quad \text{and} \quad \mathcal{W}_n(\theta) := \left(\sqrt{n}d(\theta, \widehat{\Theta}_{0,\delta})/\widehat{w}(\theta)\right)^2,$$

and let the critical values $\hat{k}_L(1 - \alpha)$ and $\hat{k}_W(1 - \alpha)$ be consistent estimators of the asymptotic $1 - \alpha$ quantiles of $\mathcal{L}_n = \sup_{\theta \in \partial\Theta_0} \mathcal{L}_n(\theta)$ and $\mathcal{W}_n = \sup_{\theta \in \partial\Theta_0} \mathcal{W}_n(\theta)$, respectively. Consider the decision functions for including a value of θ in the confidence regions based on the LR and Wald-type statistics,

$$\phi_{Ln}(\theta) := \mathbf{1}\{\mathcal{L}_n(\theta) \leq \hat{k}_L(1 - \alpha)\} \quad \text{and} \quad \phi_{Wn}(\theta) := \mathbf{1}\{\mathcal{W}_n(\theta) \leq \hat{k}_W(1 - \alpha)\},$$

respectively. Similarly, given the parameter transformation η, we consider the statistics

$$\mathcal{L}_n(\theta; \eta) := \left[\sqrt{n}\widehat{m}_\eta(\eta(\theta))/\widehat{s}(\theta)\right]^2_+ \quad \text{and}$$

$$\mathcal{W}_n(\theta; \eta) := \left(\sqrt{n}d(\eta(\theta), \eta(\widehat{\Theta}_{0,\delta}))/\widehat{w}(\theta; \eta)\right)^2$$

with a weight function $\widehat{w}(\theta; \eta)$ possibly depending on η, and the resulting decision functions

$$\phi_{Ln}(\theta; \eta) := \mathbf{1}\{\mathcal{L}_n(\theta; \eta) \leq \hat{k}_{L\eta}(1 - \alpha)\} \quad \text{and}$$

$$\phi_{Wn}(\theta; \eta) := \mathbf{1}\{\mathcal{W}_n(\theta; \eta) \leq \hat{k}_{W\eta}(1 - \alpha)\},$$

where the critical values $\hat{k}_{L\eta}(1 - \alpha)$ and $\hat{k}_{W\eta}(1 - \alpha)$ are estimates of the asymptotic $1 - \alpha$ quantiles of $\mathcal{L}_n(\eta) = \sup_{\theta \in \partial\Theta_0} \mathcal{L}_n(\theta; \eta)$ and $\mathcal{W}_n(\eta) = \sup_{\theta \in \partial\Theta_0} \mathcal{W}_n(\theta; \eta)$, respectively.

We say that the decision function $\phi_n(\theta; \eta)$ has one of the following properties:

- It is *invariant* if for any $\theta \in \Theta$, we have $\phi_n(\theta) = \phi_n(\theta; \eta)$ for any parameter transformation $\eta \in H^*$.

- It is *asymptotically invariant* to first order if for any parameter transformation $\eta \in H^*$, any $\theta \in \Theta$, and any sequence $\theta_n \to \theta$, we have $P(\phi_n(\theta_n) \neq \phi_n(\theta_n; \eta)) \to 0$ as $n \to \infty$.

These invariance properties describe whether the parameter transformations affect the inclusion of any sequence of points θ_n in the confidence sets. Note in particular that (exact) invariance is a property of a given realization of $\phi_n(\theta; \eta)$ and implies asymptotic invariance. Furthermore, it is easy to verify that invariance of a decision function $\phi_n(\theta; \eta)$ implies analogous equivariance properties for the corresponding confidence sets

$$R_{\mathrm{LR}} \equiv \{\theta \in \Theta : \phi_{\mathrm{L}n}(\theta; \eta) = 1\} \quad \text{and} \quad R_{\mathrm{W}} \equiv \{\theta \in \Theta : \phi_{\mathrm{W}n}(\theta; \eta) = 1\}.$$

We also would like to mention another property, which characterizes the precision of the confidence sets. We say that a test based on $\phi_n(\theta)$ is *asymptotically similar* on the boundary of Θ_0 if $\lim_n \mathrm{E}[\phi_n(\theta)] = c$ for some constant $c \in (0, 1)$ and any $\theta \in \partial\Theta_0$. Similarity here means that any point on the boundary of the identified set can be expected to be included with asymptotic probability c, which does not vary with the location of the point.

Given these definitions, we can now characterize the invariance and similarity properties of inference procedures based on the weighted LR and Wald statistics.

PROPOSITION 1 (Invariance Properties of Decision Functions Based on $\mathcal{L}_n(\theta; \eta)$ and $\mathcal{W}_n(\theta; \eta)$). *Suppose Conditions R.1–R.3 hold. Then* (i) *the decision function $\phi_{\mathrm{L}n}(\theta; \eta)$ based on $\mathcal{L}_n(\theta; \eta)$ is invariant and asymptotically similar on the boundary of Θ_0, whereas* (ii) *the decision function $\phi_{\mathrm{W}n}(\theta; \eta)$ based on $\mathcal{W}_n(\theta; \eta)$ is asymptotically invariant to first order if the weighting function is of the form $\widehat{w}(\theta; \eta) = \frac{b(\theta)}{\|\nabla_\eta m_n(\eta(\theta))\|} + o_p(1)$ for some function $b(\theta) > 0$ that does not depend on η. Furthermore,* (iii) *if $\widehat{w}(\theta; \eta) = b\frac{\widehat{s}(\theta)}{\|\nabla_\eta m_n(\eta(\theta))\|} + o_p(1)$ for some constant $b > 0$, then $\phi_{\mathrm{W}n}(\theta)$ is asymptotically similar on the boundary of Θ_0.*

Notice in particular the different roles the norm of the gradient of $m(\theta)$ and the standard deviation $s(\theta)$ play for the properties of the weighted Wald statistic: Choosing weights that are inversely proportional to $\|\nabla_\theta m(\theta)\|$ in the limit corrects for the dependence of the Hausdorff distance on the parameterization of the problem, and accounting for $s(\theta)$ also gives similarity on the boundary. In particular, only the weights in part (iii) of Proposition 1 yield results for confidence sets based on the Wald statistic that compare to the performance of LR-based inference. Our empirical results below illustrate that the difference is important, since the lack of invariance or precision can lead to overturning the main economic conclusions in the empirical analysis.

2.4 *Structured projection approach*

We also compare our procedure to alternative confidence regions from a structured projection approach that is based on a confidence set for the point-identified parameter γ that characterizes the bound $m(\theta) = m(\theta, \gamma)$ under Conditions C. This confidence region is then "projected" to obtain a confidence region for Θ_0. We call the projection approach "structured" when γ is a minimal sufficient parameter for the bounds on θ, and $\widehat{\gamma}$ is the minimal sufficient statistics. For example, the HJ bounds from Example 1 can be characterized by the three-dimensional parameter

$$\gamma = (S_{vv}, S_{v1}, S_{11})',$$

whereas the Markowitz–Fama mean–variance frontiers are described in terms of the first two moments of the spanning tangency portfolios. Note that the structured approach avoids creating some confidence regions for high-dimensional mean and variance parameters v and Σ, and then projecting them to obtain a confidence region for Θ_0. Such an approach would be extremely conservative, and working with the minimal parameter γ instead reduces the conservativeness dramatically. Since inference for γ is standard, the structured projection approach seems much more natural and "economically appealing." However, despite its dimension-reducing and intuitive appeal, this approach remains very conservative and is much less powerful than the approach based on the optimally weighted LR-type and W statistics. We illustrate the superior performance of LR-based confidence sets compared to the structured projection approach empirically for the Hansen–Jagannathan set (see Figure 7).

For the construction of projection confidence sets, we assume that we can construct a $(1 - \alpha)$ confidence set R_γ for γ_0. This confidence region can then be projected onto the parameter space Θ to form the confidence set for Θ_0:

$$R_{\text{Proj}} := \bigcup_{\gamma \in R_\gamma} \left\{ \theta \in \Theta : m(\theta, \gamma) \le 0 \right\}. \tag{2.14}$$

Under Conditions C, $\sqrt{n}(\widehat{\gamma} - \gamma_0)$ satisfies a central limit theorem (CLT), so we can construct an elliptical joint $1 - \alpha$ confidence region for the quantity γ as

$$R_\gamma = \left\{ \gamma \in \Gamma : n(\widehat{\gamma} - \gamma)' \widehat{\Omega}^{-1} (\widehat{\gamma} - \gamma) \le \widehat{c}(1 - \alpha) \right\}, \tag{2.15}$$

where $\widehat{c}(1 - \alpha)$ is either a $1 - \alpha$ quantile of the $\chi^2(k)$ variable or any consistent estimate of such a quantile. Under asymptotic normality of $\widehat{\gamma}$, this construction approximates an upper contour set of the density of the estimator and, therefore, gives an (approximate) smallest volume confidence set for the parameter γ.

PROPOSITION 2 (Basic Validity of the Structured Projection Approach). *Let R_γ be a $1 - \alpha$ confidence set for the parameter γ. Then $\mathrm{P}[\Theta_0 \subset R_{\text{proj}}] \ge \mathrm{P}[\gamma_0 \in R_\gamma] = 1 - \alpha$. In particular, under Conditions C, the region R_γ in (2.15) obeys*

$$\mathrm{P}[\gamma_0 \in R_\gamma] = 1 - \alpha + o(1),$$

so that the confidence region (2.14) *based on structured projection obeys*

$$P[\Theta_0 \subset R_{\text{proj}}] \geq 1 - \alpha + o(1).$$

Projected confidence sets were first proposed by Scheffé (1953) for the problem of joint confidence bounds for all linear combinations of the form $c'\gamma$, where $\gamma \in \mathbb{R}^d$ is a parameter vector and $c \in \mathbb{S}^{d-1}$ is the $d-1$ dimensional unit sphere in $\mathbb{R}^d$. For the problem of confidence bands for the linear regression function, it has been shown that optimality of Scheffé's (1953) method depends crucially on equivariance with respect to translations and orthogonal transformations of the original parameter; see Sections 9.4 and 9.5 in Lehmann and Romano (2005) for a discussion. Bohrer (1973) showed the optimality of projection bounds for a generalization of Scheffé's (1953) original problem, but they were shown to be suboptimal if coverage was only required for a restricted set of functionals corresponding to vectors c for a proper subset of $\mathbb{S}^{d-1}$; see Casella and Strawderman (1980) and Naiman (1984).

The main reason for the poor statistical properties of the projection approach is that it is based on a confidence set R_γ for the sufficient parameter γ, which is designed to ensure simultaneous coverage of all components of γ_0 under any distribution corresponding to the value γ_0. However, only certain deviations from γ_0 would result in a violation of the moment inequality at a given value of θ. As a result, there are at least three instances in which equivariance with respect to the reduced-form parameters γ fails: in all our examples, (i) the dimension of θ is strictly lower than that of γ, (ii) we consider inference problems that are one-sided rather than symmetric, where the parameter space is only a compact subset of $\mathbb{R}^k$, and (iii) the inequality generating function $m(\theta, \gamma)$ is nonlinear in γ. Even after approximate linearization, our inference problem reduces to inference on $c'\gamma$ with c restricted to a small subset of the sphere $\mathbb{S}^{d-1}$, which is exactly the case where the structured projection approach becomes suboptimal. In the empirical application to the HJ problem, the inferior performance of the structured projection-based set relative to the LR-based confidence set can be seen very clearly in Figure 7. Specifically, we obtain an LR-based set that is much smaller and lies strictly inside the projection-based set.

2.5 *Approximating multiple inequalities by a single smooth inequality*

Throughout the paper, we only consider the case of a single inequality restriction $m : \Theta \to \mathbb{R}$ satisfying certain smoothness conditions with respect to θ. However, this framework approximately encompasses the case of multiple inequalities, since it is possible to conservatively approximate intersection bounds of the form $g(\theta) \leq 0$, where $g(\theta) := (g_1(\theta), \ldots, g_J(\theta))'$ represents a vector of constraints, by a single smooth bound $m(\theta) \leq 0$. To be specific, we can form an inequality-generating function

$$m_\lambda(\theta) := H\big(g_1(\theta), \ldots, g_J(\theta); \lambda\big) := \sum_{j=1}^{J} \frac{\exp(\lambda g_j(\theta))}{\sum_{l=1}^{J} \exp(\lambda g_l(\theta))} g_j(\theta),$$

where $\lambda \in \mathbb{R}$ is a fixed, positive scalar. This function m_λ is clearly smooth with respect to the underlying function g and is a conservative approximation to $\max_{1 \leq j \leq J} g_j(\theta)$. This approximation is related to the smoothed intersection bounds for valuation cumulative distribution functions (c.d.f.s) in Haile and Tamer (2003). We can now give an explicit error bound for the approximation as a function of the smoothing parameter λ, as shown by the following lemma.

LEMMA 1 (Properties of Smooth Max). (i) *We have that*

$$H(g_1, \ldots, g_J, \lambda) \leq 0 \quad \textit{if } \max_{1 \leq j \leq J} g_j(\theta) \leq 0.$$

(ii) *For any $\lambda > 0$, the approximation error obeys*

$$\sup_{(g_1,\ldots,g_J)' \in \mathbb{R}^J} \left| \max(g_1, \ldots, g_J) - H(g_1, \ldots, g_J, \lambda) \right| \leq \lambda^{-1} W\left(\frac{J-1}{e} \right),$$

where e is Euler's number, and the function $W : \mathbb{R} \mapsto \mathbb{R}$ is Lambert's product logarithm function, i.e., $W(z)$ is defined as the solution w of the equation $w \exp\{w\} = z$. In particular, $W(z) \geq 0$ if $z \geq 0$ and $W(z) \leq \log(z)$ if $z \geq e$.

This lemma implies that the approximation error relative to the max function (with respect to the sup-norm) is inversely proportional to the smoothing parameter λ, where the constant of proportionality grows only very slowly in J. In the paper, we state our results and assumptions on estimation and inference directly in terms of the scalar moment $m(\theta) = m_\lambda(\theta)$ and its empirical analog $\widehat{m}(\theta) := \widehat{m}_\lambda(\theta) := H(\widehat{g}_1(\theta), \ldots, \widehat{g}_J(\theta); \lambda)$. We also consider the quantity $1/\lambda$ fixed at some small value. We also note that the approximation error is largest when several elements in $\{g_1, \ldots, g_J\}$ are close to the maximal value, whereas it is negligible when the maximum of $\{g_1, \ldots, g_J\}$ is well separated.

We now state formally the sense in which the statistical properties of the procedures for a single moment inequality are inherited by their analogs for a smoothed intersection bound based on the moment $m_\lambda(\theta)$. To this end, we define

$$\Theta_0(\lambda) := \{\theta \in \Theta : m_\lambda(\theta) \leq 0\},$$

the identification region satisfying the weighted average of the J inequality restrictions. Clearly, $\Theta_0 \subset \Theta_0(\lambda)$, so that inference based on the smoothed maximum is generally conservative compared to a "hard" intersection bound. Note that Lemma 1 also implies a bound on the approximation error bound for the set $\Theta_0(\lambda)$ relative to the "hard" intersection bound: Suppose that the parameter space is convex and gradients $\|\nabla_\theta g_j(\theta)\|$ are bounded away from zero, that is, $\kappa_0 := \min_j \inf_{\theta \in \Theta} \|g_j(\theta)\| > 0$. Then Lemma 1 implies that the Hausdorff distance $d_H(\Theta_0(\lambda), \Theta_0)$ can be bounded by $W(\frac{J-1}{e})/(\lambda \kappa)$, where the conclusion follows from the mean-value theorem in combination with the Cauchy–Schwarz inequality.

Analogous to the previous sections, define

$$\widehat{\Theta}_{0,\delta}(\lambda) := \{\theta \in \Theta^\delta : \widehat{m}_\lambda(\theta) \leq 0\},$$

and let the likelihood ratio and Wald statistics for the smoothed maximum be written

$$\mathcal{L}_n(\theta) := \left[\sqrt{n}\widehat{m}_\lambda(\theta)/\widehat{s}(\theta)\right]_+^2 \quad \text{and} \quad \mathcal{W}_n(\theta) := \left(\sqrt{n}d\bigl(\theta, \widehat{\Theta}_{0,\delta}(\lambda)\bigr)/\widehat{w}(\theta)\right)^2.$$

We also assume that the critical values $\widehat{k}_L(1-\alpha)$ and $\widehat{k}_W(1-\alpha)$ are consistent estimators of the asymptotic $1-\alpha$ quantiles of $\mathcal{L}_n = \sup_{\theta\in\Theta_0}\mathcal{L}_n(\theta)$ and $\mathcal{W}_n = \sup_{\theta\in\Theta_0}\mathcal{W}_n(\theta)$, respectively.

It is immediate that for any fixed value of $\lambda > 0$, the population moment $m_\lambda(\theta)$ and its sample analog $\widehat{m}_\lambda(\theta)$ satisfy Conditions R and C, respectively, if these conditions are met by the moment functions $g_1(\theta), \ldots, g_J(\theta)$ and $\widehat{g}_1(\theta), \ldots, \widehat{g}_J(\theta)$ individually. Hence the basic results on inferential validity, consistency of the confidence regions relative to the smoothed bound $\Theta_0(\lambda)$, and consistency of the bootstrap directly carry over from the single inequality case.

So as to characterize the invariance and similarity properties for the procedure based on the smoothed maximum, let the decision functions for including a value of θ in the confidence regions based on the LR and Wald-type statistic be

$$\phi_{Ln}(\theta) := \mathbf{1}\bigl\{\mathcal{L}_n(\theta) \le \widehat{k}_L(1-\alpha)\bigr\} \quad \text{and} \quad \phi_{Wn}(\theta) := \mathbf{1}\bigl\{\mathcal{W}_n(\theta) \le \widehat{k}_W(1-\alpha)\bigr\},$$

respectively. We can now consider the same set of parameter transformations $\eta \in H^*$, where we denote the smoothed moment condition after the parameter transformation with $m_{\eta,\lambda}(\theta) := m_\lambda(\eta^{-1}(\theta))$, and denote its sample analog with $\widehat{m}_{\eta,\lambda}(\theta) := \widehat{m}_\lambda(\eta^{-1}(\theta))$. We can then also define the decision functions $\phi_{Ln}(\theta; \eta)$ and $\phi_{Wn}(\theta; \eta)$ in analogy with the case of a single inequality restriction.

With these definitions, the following corollary follows directly from Proposition 1.

COROLLARY 2 (Invariance Properties for Smoothed Maximum). *Suppose Conditions R.1–R.3 hold. Then* (i) *the decision function $\phi_{Ln}(\theta; \eta)$ based on $\mathcal{L}_n(\theta; \eta)$ is invariant and asymptotically similar on the boundary of $\Theta_0(\lambda)$, whereas* (ii) *the decision function $\phi_{Wn}(\theta; \eta)$ based on $\mathcal{W}_n(\theta; \eta)$ is asymptotically invariant to first order if the weighting function is of the form $\widehat{w}(\theta; \eta) = \frac{b(\theta)}{\|\nabla_\eta m_\eta(\eta(\theta))\|} + o_p(1)$ for some function $b(\theta) > 0$ that does not depend on η. Furthermore,* (iii) *if $\widehat{w}(\theta; \eta) = b\frac{\widehat{s}(\theta)}{\|\nabla_\eta m_\eta(\eta(\theta))\|} + o_p(1)$ for some constant $b > 0$, then $\phi_{Wn}(\theta)$ is asymptotically similar on the boundary of $\Theta_0(\lambda)$.*

Hence, the invariance and similarity properties are inherited by the corresponding procedures based on the smoothed maximum $\widehat{m}_\lambda(\theta)$ for any fixed value of $\lambda > 0$. However, it is important to note that these properties are given with respect to the smoothed intersection bound $\Theta_0(\lambda)$, a superset of the identification region Θ_0.

3. EMPIRICAL APPLICATIONS

3.1 *Hansen–Jagannathan mean–variance sets for the SDF*

To illustrate the performance of our procedure, we estimate confidence sets for the Hansen–Jagannathan sets of means and variances of stochastic discount factors. To keep
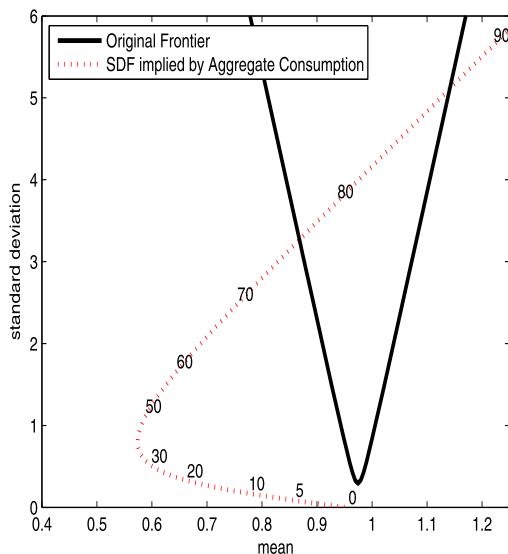
FIGURE 1. Estimated HJ bounds with stochastic discount factors implied by constant relative risk aversion (CRRA) preferences. Data labels represent values for the intertemporal elasticity of substitution (IES) ϱ.

results comparable with Hansen and Jagannathan (1991), we construct the sample for the empirical exercise following the data description in Hansen and Jagannathan (1991). The two asset series used are annual treasury bond returns and annual New York Stock Exchange (NYSE) value-weighted dividend included returns. These nominal returns are converted to real returns by using the implicit price deflator based on personal consumption expenditures used by Hansen and Jagannathan (1991). Asset returns are from the Center for Research in Security Prices (CRSP); the implicit price deflator is available from the St. Louis Federal Reserve and is based on the National Income and Product Accounts of the United States. We use data for the years 1960–2012.

Figure 1 reports the estimated bound consisting of the mean-standard deviation pairs that satisfy

$$m(\theta, \widehat{\gamma}) = 0,$$

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

We can compare the estimated HJ bounds with mean–variance combinations implied by the consumption CAPM model. In the model, the economy is equivalent to a representative agent with constant elasticity of intertemporal substitution preferences[4]

$$u(C_t) = \frac{C_t^{1-\varrho}}{1-\varrho},$$

[4]For clarity of exposition, we focus on the classical empirical formulation of the CAPM model as, e.g., in Hansen and Singleton (1982). Most aspects of this analysis can, in principle, be extended to more recent developments in empirical models for asset prices, e.g., Epstein and Zin (1989) preferences as in Bansal and Yaron (2004) or habit formation as in Campbell and Cochrane (1999).

where C_t is the aggregate consumption. Then the stochastic discount factor implied by consumption growth is given by $M_t(\varrho) = \beta(\frac{C_t}{C_{t+1}})^\varrho$, so that we can estimate the first two moments of $M_t(\varrho)$. Specifically, let $\mu_C(\varrho) := \mathbb{E}[M_t(\varrho)]$ and $\sigma_C(\varrho) := \sqrt{\text{Var}(M_t(\varrho))}$ be the mean and standard deviation, respectively, of the stochastic discount factor given an isoelastic utility function with an elasticity of intertemporal substitution equal to $1/\varrho$. We can characterize the feasible set Θ_{EIS} of mean–variance pairs by the moment restriction

$$0 = m_{\text{EIS}}\big((\mu, \sigma)\big) := \mu_C\big(\sigma_C^{-1}(\sigma)\big) - \mu,$$

where $\sigma_C^{-1}(\cdot)$ denotes the inverse function of $\sigma_C(\varrho)$.[5] Given a time series of observations for the growth rate of consumption $\frac{C_{t+1}}{C_t}$, we define the empirical analog $\widehat{m}_{\text{EIS}}((\mu, \sigma))$ analogously, where the expectations in the definition of $\mu_C(\cdot)$ and $\sigma_C(\cdot)$ are replaced with averages. The mean–variance pairs reported in Figure 1 were obtained using data on per capita expenditures for consumption and services in the United States from 1959 to 2012, assuming a discount factor $\beta = 0.95$.

It is well known that it is difficult to reconcile asset prices and aggregate consumption empirically in a representative framework. For our data, the values of ϱ corresponding to mean–variance pairs for the SDF that fall inside the estimated HJ set range from about 75 to 88, suggesting a very low elasticity of intertemporal substitution and implying a high variance for the SDF. These values for ϱ are unrealistically large, but are in line with other findings in the empirical literature on the consumption-based CAPM.[6]

To allow for serial dependence in the time series for consumption and asset returns, we computed confidence sets using the block bootstrap using block length equal to 5 years. As a robustness check, we repeated the analysis using block lengths equal to 1, 3, and 7 years, and obtained almost identical results. In particular, treating observations as independent across periods does not appear to affect our empirical conclusions. This may in part be due to the fact that the time series for consumption and asset returns used for estimation are only at an annual frequency, so that dependence may not be a major concern. It is also important to note that our formal results regarding validity and invariance of confidence sets do not require that observations be independent or from a stationary process. Rather, it is possible to combine our method of construction with any simulation or resampling scheme that estimates the asymptotic distributions of $\mathcal{L}_n$ and $\mathcal{W}_n$ consistently.

To represent the sampling uncertainty in estimating γ, we plot 100 bootstrap draws of the HJ frontier in Figure 2, where observations were drawn with replacement from the bivariate time series of stock and bond returns. To represent the sampling uncertainty in estimating the mean-standard deviation pairs $(\mu_C(\varrho), \sigma_C(\varrho))$ of the consumption-based SDF, we also plot 100 bootstrap draws of $\varrho \mapsto (\widehat{\mu}_C(\varrho), \widehat{\sigma}_C(\varrho))$. We see that the sampling uncertainty is quite considerable for both HJ frontier and for mean-standard

---

[5]It can be verified that $\sigma_C(\varrho)$ is strictly increasing in ϱ, so that this inverse is well defined.

[6]For example, in Lettau and Ludvigson (2009), the values for the elasticity of intertemporal substitution minimizing the mean squared error in the Euler equations characterizing the household's investment problem are comparable in magnitude.
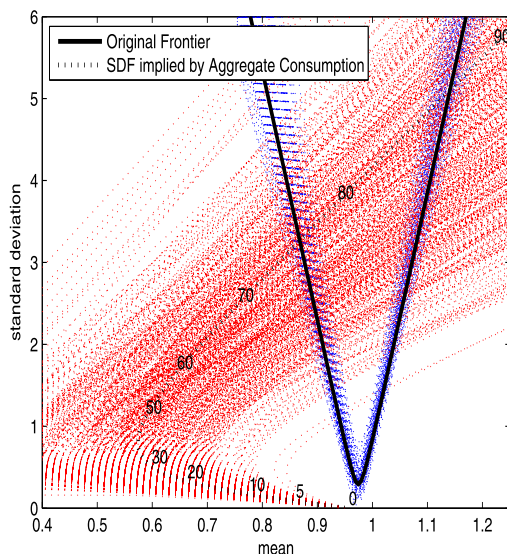
FIGURE 2. Estimated HJ bounds, mean–standard deviation pairs based on consumption SDF, and bootstrap draws.

deviation pairs implied by the consumption-based SDF. In fact, an intriguing feature of this graph is that near the apex of the HJ frontier, the pair $(\hat{\mu}_C(\varrho), \hat{\sigma}_C(\varrho))$ and its bootstrap draws are close to the HJ frontier at low values of ϱ. The low values of ϱ are considered to be "reasonable," since they correspond to a relatively high intertemporal elasticity of substitution and appear to be well microfounded. The inability of reasonable values of ϱ to reconcile aggregate consumption data with asset prices has been a major theme of the empirical literature on the consumption-based CAPM model starting with Hansen and Singleton (1982). However, from Figures 1 and 2 it is not obvious that an empirical test would reject the canonical/baseline model underlying that literature, and we report confidence sets based on the various approaches discussed in earlier sections based on which we can make inferential statements about the benchmark model.

Figure 3 shows the 95% confidence region based on the LR statistic. By construction, the LR confidence region covers most of the bootstrap draws below the HJ bounds. However, it should also be noted that the confidence bound based on the LR statistic is fairly tight relative to the boostrapped frontiers, and does not include any unnecessary areas of the parameter space. Noting that the set Θ_e for the consumption-based SDF is defined by a moment equality, we can also form a confidence band R_{EIS} for Θ_{EIS} based on the statistic $\mathcal{L}_n(\theta) := [\frac{\hat{m}_{\text{EIS}}(\theta)}{\hat{s}(\theta)}]^2$, noting that the asymptotic arguments in the derivations for the (one-sided) LR statistic can be easily extended to the two-sided case if we replace squared positive parts $[\cdot]_+^2$ with the usual square $(\cdot)^2$. The lower and upper bounds in Figures 3–7 were constructed using separate estimates of the local standard deviation $s((\mu, \sigma))$ based on the negative and positive deviations of $\hat{m}_{\text{EIS}}((\mu, \sigma))$, respectively, to improve the approximation. Critical values were obtained using the nonparametric bootstrap.
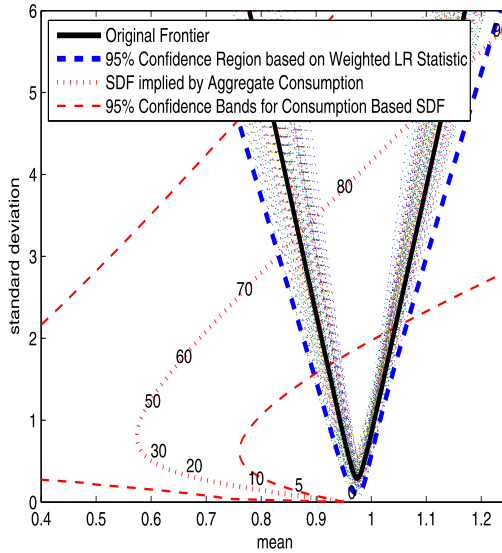
FIGURE 3. The 95% confidence set for the HJ set using the LR statistic. The 95% confidence set for mean and standard deviation of the consumption-based SDF was also constructed using the LR statistic for the moment equality defining the corresponding mean–variance pairs.

Most importantly, the LR-based confidence region for the HJ set does not overlap with the confidence set for the consumption-based SDF for "small" values of ϱ (in fact, for any $\varrho \in [0, 58]$). The absence of overlap for the 95% regions implies the rejection of any $\varrho \in [0, 58]$ at the 10% significance level, where we can form Bonferroni bounds from the false rejection rates of the respective tests underlying the two confidence regions. This is clear evidence against the benchmark formulation of the consumption-based CAPM and, therefore, is an important empirical conclusion. In what follows, we show that the same empirical conclusion cannot be reached for this example using less precise or noninvariant methods. Specifically, we will show that if we use confidence regions based on either LR statistics without precision weighting, or Wald statistics without invariance/precision weighting, or regions based on structural projection, we will not be able to reach the same empirical conclusion. So invariance and precision considerations in construction of the confidence regions turn out to be quite important for reaching sharp economic conclusions.

Figure 4 plots the 95% confidence region based on an unweighted LR statistic. Comparing Figure 3 and Figure 4, it can be seen that precision weighting plays a very important role in delivering good confidence sets. Without precision weighting, the unweighted LR statistic delivers a confidence region that includes implausible regions in the parameter space where the standard deviation of the discount factor is zero. Moreover, the confidence region becomes too imprecise to reject the canonical model of the stochastic discount factor with realistic (small) values of ϱ.

Figure 5 plots the confidence region based on the Wald statistic with no invariance/precision weighting, which is equivalent to a confidence region based on the directed Hausdorff distance. Similar to Figure 4, the confidence set covers a large area of
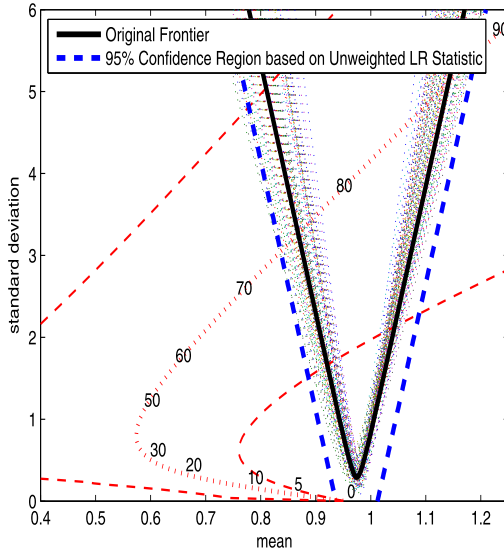
FIGURE 4. The 95% confidence set for the HJ set using an unweighted LR statistic.
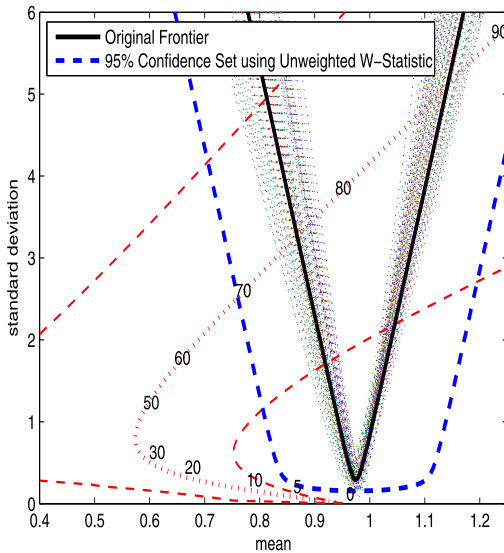
FIGURE 5. The 95% confidence set for the HJ set using an unweighted W-statistic (directed Hausdorff distance).

the parameter space that is excluded from any bootstrap realization of the HJ set. The shape of the confidence region based on the Wald statistic in Figure 5 seem counterintuitive because at first sight, as the confidence bounds do not appear to be a uniform enlargement of the estimated frontier $\partial\widehat{\Theta}$. However, this visual impression is only due to the fact that the plot shows units of μ and σ at different scales. The observation that the weighting and scaling of the different components of θ seem "unnatural" in this par-
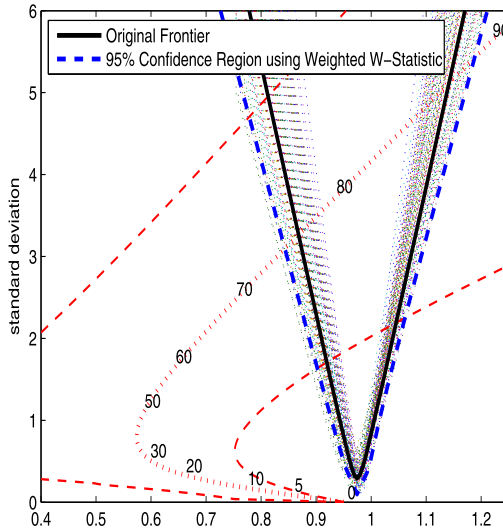
FIGURE 6. The 95% confidence set for the HJ set using a weighted W-statistic.

ticular graph emphasizes the potential problems associated with the noninvariance of inference based on the unweighted Wald statistic.

Figure 6 plots the confidence region based on the weighted Wald statistic, where weights induce first-order invariance and similarity via precision weighting. This weighting fixes the problem and generates a statistic that is (first-order) invariant to parameter transformations. As a result, the confidence set looks very similar to the weighted LR-based confidence set in Figure 3 in that it covers most of the bootstrap draws below the HJ bounds and its shape reflects local sampling uncertainty in an adequate manner. This practical evidence therefore emphasizes the importance of introducing invariance and precision inducing weights in the Wald-based approach, which we argued for theoretically in the previous sections.

Finally, in Figure 7, we compare our results to confidence regions from the structured projection approach that is based on a confidence set for the point-identified parameters. As described in Section 2.5, we construct an elliptical joint $1 - \alpha$ confidence region for the quantity γ defined in (1.2) based on the quadratic form for the estimator $\widehat{\gamma}$, $\widehat{T}(\gamma) := (\widehat{\gamma} - \gamma)' \operatorname{var}(\widehat{\gamma})^{-1} (\widehat{\gamma} - \gamma)$. For the diameter of this confidence ellipsoid, we used both a bootstrap and a chi-squared approximation to the distribution of $\widehat{T}(\gamma)$, which both yield qualitatively similar results.

The $1 - \alpha$ confidence set for $\theta = (\mu, \sigma)$ is obtained by projecting the $1 - \alpha$ confidence region for γ onto Θ using the condition $m(\theta, \gamma) \leq 0$. We report the resulting confidence set from the structured projection approach in Figure 7 together with the LR-based confidence set proposed in this paper. The structured projection confidence set performs quite poorly relative to the LR-based confidence set: in particular, the latter is much smaller and lies strictly inside the former. In fact, the precision of the confidence set based on structured projection is poor enough to overturn the major empirical
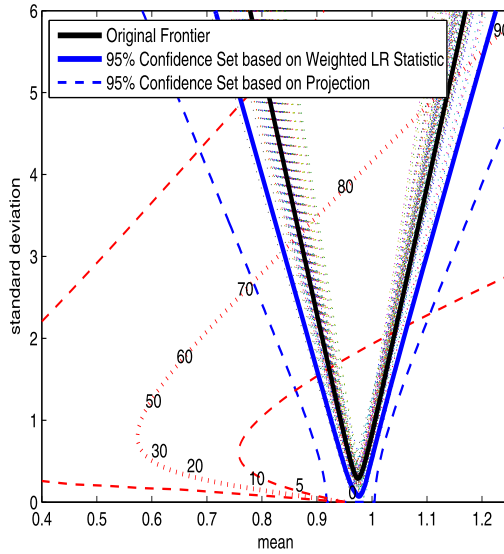
FIGURE 7. The 95% confidence sets for the HJ set using a weighted LR statistic and a projection approach.

conclusion that the consumption-based CAPM cannot be reconciled with small values of ϱ.

This should be expected since the projection confidence bounds are based on a confidence set for the point-identified parameter that does not account for the specific shape of the bounds as a function of $\widehat{\gamma}$. More specifically, the elliptical joint confidence set for γ (which minimizes volume under joint normality of $\widehat{\gamma}$) guards us against deviations from the true value in any direction in $\mathbb{R}^3$, but most of these deviations are irrelevant for the bounds for θ, since these are only one-sided and the parameter space for θ is only two-dimensional. The fact that the standard confidence set for γ treats all directions in the parameter space Γ symmetrically may be far from ideal for inference on the (μ, σ) frontier, since the bound on the standard deviation is a nonlinear function whose derivative with respect to γ varies widely across different values of $(\mu, \sigma(\mu))$. Note that for confidence sets for a point-identified parameter, by the delta method, the effect of nonlinearities is asymptotically negligible to first order. However, when the object of interest is a set with a nontrivial diameter, the resulting effect is of first order even for large samples.

3.2 *Bounds on the elasticity of labor supply*

In his metaanalysis, Chetty (2012) reports pointwise confidence bounds for the structural Hicksian elasticity ε of labor supply at the intensive margin for given values of the optimization friction δ. The reported bounds result from the intersection of bounds of the form (1.4) from estimates $\widehat{\varepsilon}_j$ obtained from J empirical studies exploiting different natural experiments that vary the effective income tax $\Delta_j \log p$, $j = 1, \ldots, J$.

We apply the bootstrap procedure proposed in this paper to obtain joint confidence sets for (δ, ε) based on the LR and Wald statistics. More specifically, we consider the moments obtained from individual empirical elasticities $\widehat{\varepsilon}_j$:

$$\widehat{g}_{\mathrm{OF},j}\big((\varepsilon, \delta)'\big) = g_{\mathrm{OF},j}\big((\varepsilon, \delta)', \widehat{\varepsilon}_j\big) := \frac{(\varepsilon - \widehat{\varepsilon}_j)^2 (\Delta_j \log p)^2}{8\varepsilon} - \delta.$$

These "raw" moments are then aggregated by a smooth function

$$\widehat{m}_{\mathrm{OF}}^*(\varepsilon, \delta; \lambda) = \sum_{j=1}^{J} \frac{\exp(\lambda \widehat{g}_{\mathrm{OF},j}(\varepsilon, \delta))}{\displaystyle\sum_{l=1}^{J} \exp(\lambda \widehat{g}_{\mathrm{OF},l}(\varepsilon, \delta))} \widehat{g}_{\mathrm{OF},j}(\varepsilon, \delta),$$

where $\lambda \in \mathbb{R}$ is a fixed, positive scalar. Note that as discussed in Section 2, this transformation approximates the maximum of $\widehat{g}_{\mathrm{OF},1}(\varepsilon, \delta), \ldots, \widehat{g}_{\mathrm{OF},J}(\varepsilon, \delta)$ as $\lambda \to \infty$, but satisfies the smoothness conditions for our procedure for any finite value of $\lambda > 0$. For our empirical results, we set $\lambda = 3$, where we compute the smoothing weights based on the studentized moment functions.

We use a parametric bootstrap to obtain the critical value $k(1 - \alpha)$, where we approximate the sampling distribution of the estimators for the respective elasticities by a joint normal distribution centered around the estimates reported in panel A of Table 1 in Chetty (2012) with standard deviations equal to the respective standard errors and zero covariances. This approach can be justified by an assumption that the studies were based on mutually independent random samples from possibly different populations.

Figure 8 shows that the estimated bounds coincide with the set reported in Figure 8 of Chetty (2012) except for the use of the smoothed maximum function instead of the in-
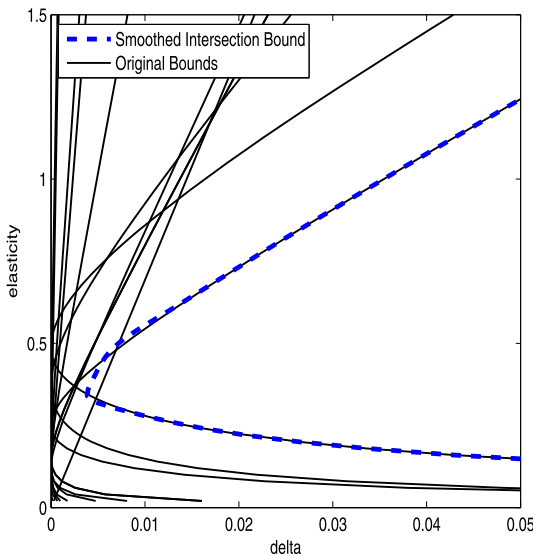

FIGURE 8. Bounds on structural labor supply elasticity based on estimates from different studies, and their smoothed intersection bound.
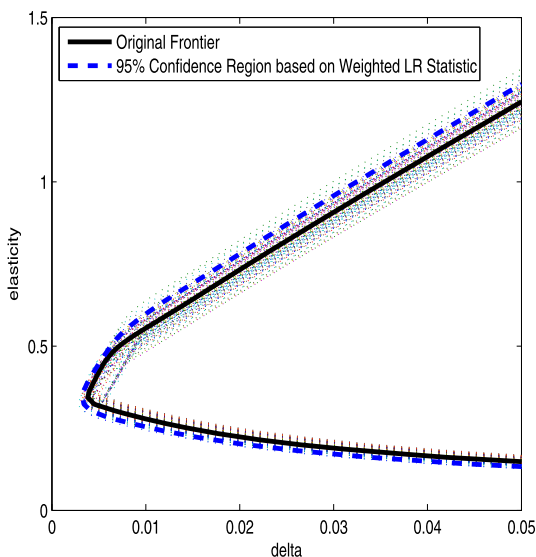
Figure 9. Smoothed bound with 95% confidence set for the plausible pairs of the structural elasticity ε and friction δ using the LR statistic.

tersection of (ε, δ) sets, which leads to a slightly wider set. The 95% confidence set based on the LR statistic[7] reported in Figure 9 is fairly narrow around the estimated bound, and does not appear to differ very much from the collections of confidence intervals in Chetty (2012). Chetty presents confidence intervals that are pointwise with respect to δ; that is, for each fixed value of friction δ, the interval covers structural elasticity ε with a prescribed probability. In contrast, our set estimator covers all plausible values of (δ, ε) with a prescribed probability. Thus it simultaneously performs inference on both structural elasticity ε and the friction amount δ. The LR confidence region is a valid joint confidence set for (ε, δ) and not only pointwise in δ. Furthermore, it is not conservative in that we assume a joint sampling distribution for the elasticity estimates instead of constructing Fréchet–Hoeffding bounds. The LR confidence region excludes all points with $\delta \leq 0.3\%$, so that an optimization friction of at least that size would be needed to reconcile the different elasticities found in the studies considered in this metanalysis.

Finally, we also report a 95% confidence region based on the Wald statistic *without* optimal reweighting.[8] As in the case of HJ bounds, the shape of the resulting confidence set does not reflect the sampling variation in the estimated bounds, and the critical value

[7]Note that for the LR statistic we used, the standard error of the negative part of $\widehat{m}^*(\theta)$ is a weighting function that improves the local approximation due to the asymmetry of the distribution for small values of ϱ. Note that for the $N(0, s(\theta)^2)$ distribution, the standard deviation of the negative part is proportional to $s(\theta)$, so that this weighting scheme is asymptotically equivalent to weighting by the (inverse of the) local standard deviation of $\widehat{m}^*(\theta)$.

[8]To adjust for the differences in order of magnitude, we constructed the Hausdorff distance based on the norm $\|(\varepsilon, \delta)\| = \sqrt{\varepsilon^2 + 100 \cdot \delta^2}$. Note that in the graph, the confidence region looks poorly centered around the estimated bound, but this optical impression is, in fact, due to the different scaling of the two axes and the difference in the slope of the frontier above and below its apex.
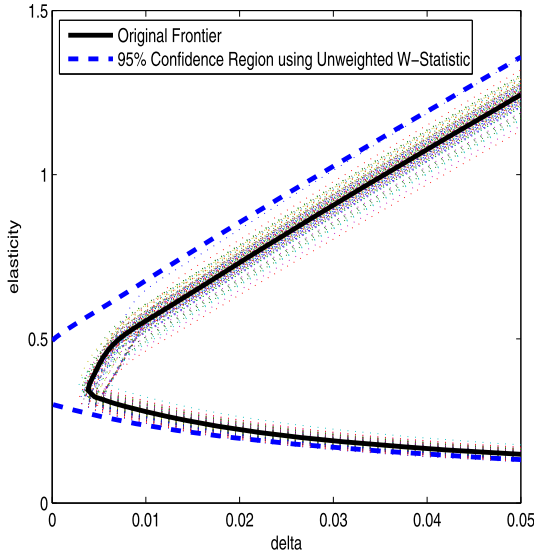
FIGURE 10. Smoothed bound with 95% confidence set for the plausible pairs of the structural elasticity ε and friction δ using an unweighted W-statistic.

for the Wald statistic is determined by perturbations of the frontier at very low values for ε. More importantly, in contrast to the LR-based region, the confidence set based on the Wald statistic includes points with $\delta = 0$, failing to reject that the empirical elasticities can be reconciled in a model with no optimization frictions and changing one of the main conclusions of the analysis. Using weighted W statistics instead fixes this problem and gives a confidence set that is very similar to the LR-based confidence set; we do not report this confidence set for brevity.

4. Conclusion

In this paper, we provide new methods for inference on parameter sets and frontiers that can be characterized by a smooth nonlinear inequality. The proposed procedures are straightforward to implement computationally and have favorable statistical properties. By analyzing the geometric and statistical properties of different statistics, we illustrate the importance of equivariance and similarity considerations for achieving tight confidence regions. In particular, while local weighting is irrelevant for the statistical properties of the estimated frontier, it matters greatly for the size and shape of confidence sets. We also consider smoothed intersection bounds from multiple inequality restrictions, where we give an exact upper bound for the approximation error that depends only on the smoothing parameter.

    We illustrate the practical usefulness of these procedures in financial econometrics with various classical examples from mean–variance analysis, including inference on Hansen–Jagannathan mean–variance sets of admissible stochastic discount factors, Markowitz–Fama mean–variance sets of admissible portfolios, and factor-based asset

pricing. As a second application, we consider Chetty's (2012) joint bounds for the elasticity of labor supply and an optimization friction. This example suggests a broader range of uses for set inference in the context of possibly misspecified or incomplete economic models.

In both examples, using invariant or precision weighted statistics is important for maintaining major empirical conclusions that have been reached informally in prior empirical work, e.g. the inability of large values of the elasticity of intertemporal substitution to generate plausible distributions of stochastic discount factors, or the need for nontrivial optimization frictions to reconcile estimated demand elasticities from different settings. Therefore, the empirical examples illustrate our formal points about the advantages of inference based on a precision weighted metric that is invariant to parameter transformations.

Appendix A: Basic Validity and Convergence Rates for the Confidence Regions

This appendix formalizes basic validity of the confidence region and consistency of the bootstrap for inference on the identified set. We begin by giving a basic validity result for confidence regions assuming that we have suitable estimates of the quantiles of LR and W statistics, and we verify basic validity of our confidence regions. A basic procedure for constructing suitable estimates of these quantiles via bootstrap or simulation 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*

$$\widehat{k}_j(1-\alpha) = k_j(1-\alpha) + o_p(1), \tag{A.1}$$

for $j = L, W$, where $k_W(1-\alpha)$ and $k_L(1-\alpha)$ are the $(1-\alpha)$ quantiles of $\mathcal{W}$ or $\mathcal{L}$, respectively. Then as the sample size n grows to infinity, confidence regions R_{LR} and R_W cover Θ_0 with probability approaching $1 - \alpha$:

$$P[\Theta_0 \subseteq R_{LR}] = P[\mathcal{L}_n \leq \widehat{k}_L(1-\alpha)] \to P[\mathcal{L} \leq k_L(1-\alpha)] = (1-\alpha), \tag{A.2}$$

$$P[\Theta_0 \subseteq R_W] = P[\mathcal{W}_n \leq \widehat{k}_W(1-\alpha)] \to P[\mathcal{W} \leq k_W(1-\alpha)] = (1-\alpha). \tag{A.3}$$

We next recall that given a Euclidian metric $d(\cdot, \cdot)$ on $\mathbb{R}^K$, the (symmetric) Hausdorff distance between two nonempty sets $A, B \subset \mathbb{R}^k$ is defined as

$$d_H(A, B) := \max\left\{\sup_{a \in A} d(a, B), \sup_{b \in B} d(b, A)\right\}. \tag{A.4}$$

Our next result shows that the confidence regions based on the LR and the Wald statistic are also root-n consistent estimators for the set Θ_0 with respect to the Hausdorff distance. This result also establishes that the Pitman rates for R_W and R_{LR} are also root-n, namely any alternative set Θ_A such that $\sqrt{n}d_H(\Theta_0, \Theta_A) \to \infty$ will not be covered by R_W and R_{LR} for n large enough, and hence will be rejected with probability tending to 1.

THEOREM 3 (Confidence Regions Are $\sqrt{n}$ Consistent Estimator of Θ_0). *Under Conditions* R, *the confidence regions based on the LR and Wald statistics are consistent with respect to the Hausdorff distance at a root-n rate, that is,*

$$d_{\mathrm{H}}(R_{\mathrm{LR}}, \Theta_0) = O_p(n^{-1/2}) \quad and \quad d_{\mathrm{H}}(R_{\mathrm{W}}, \Theta_0) = O_p(n^{-1/2}).$$

As a consequence, if $\sqrt{n}d_{\mathrm{H}}(\Theta_0, \Theta_A) \to \infty$, then

$$\mathrm{P}[\Theta_A \subseteq R_{\mathrm{LR}}] \to 0 \quad and \quad \mathrm{P}[\Theta_A \subseteq R_{\mathrm{W}}] \to 0.$$

While the results in Theorems 1–3 were stated in terms of the high-level Conditions R, the next corollary shows that these results can be applied to any problem where Conditions C holds. These conditions hold in HJ, MF, and other problems listed in Section 1, as we verify below.

COROLLARY 3 (Limit Laws of LR and W Statistics Under Conditions C). *Suppose that Conditions C holds and that consistent estimates of quantiles of statistics* (2.8) *and* (2.9) *are available. Then the conclusions of Theorem 2 apply.*

We now turn to estimation of quantiles for the 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 $\mathbb{G}$ or a weighted version of this process appearing in the limit expressions. Next, we can use the estimated law to obtain consistent estimates of the laws of the 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, which can also be adapted to allow for various forms of cross section and time series dependence.

In what follows, we simplify the notation by writing our limit statistics as a special case of the statistic

$$\mathcal{S} := \sup_{\theta \in \partial \Theta_0} [V(\theta)]_+^2, \quad V(\theta) := \tau(\theta)\mathbb{G}(\theta). \tag{A.5}$$

Thus, $\mathcal{S} = \mathcal{L}$ for $\tau(\theta) = 1/s(\theta)$ and $\mathcal{S} = \mathcal{W}$ for $\tau(\theta) = 1/[\|\nabla_\theta m(\theta)\| \cdot w(\theta)]$. We take τ to be a continuous function bounded away from zero on the parameter space. We also need to introduce the following notation 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 $\mathcal{Q}_V$.

In the following, we assume that 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 that can be 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 cover such methods of producing realizations of V^* as the bootstrap, sub-sampling, or other simulation approaches.

We require that the estimate Q_{V^*} be consistent for Q_V in any metric ρ_K metrizing weak convergence, where we can take the metric to be the Kantorovich–Rubinstein metric. Note that there are many results that verify this basic consistency condition for different processes V and various bootstrap, simulation, and subsampling schemes, as we discuss in more detail below.

To define the Kantorovich–Rubinstein metric, let $\theta \mapsto v(\theta)$ be an element of a metric space (M, d_M), and let $\mathrm{BL}_1(M)$ be a class of Lipschitz functions $\varphi : M \to \mathbb{R}$ that satisfy

$$\left|\varphi(v) - \varphi(v')\right| \leq d_M(v, v'), \qquad \sup_{v \in M} \left|\varphi(v)\right| \leq 1.$$

The Kantorovich–Rubinstein distance between probability laws Q_V and Q'_V of random elements V and V' taking values in M is defined as

$$\rho_K(Q, Q'; M) := \sup_{\varphi \in \mathrm{BL}_1(M)} \left|\mathrm{E}_Q \varphi(V) - \mathrm{E}_{Q'} \varphi(V)\right|.$$

As stated earlier, we require the estimate Q_{V^*} to be consistent for Q_V in the metric ρ_K, that is,

$$\rho_K(Q_{V^*}, Q_V; C(\Theta)) = o_p(1). \tag{A.6}$$

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

$$\widehat{\partial\Theta_0} = \left\{\theta \in \Theta : \widehat{m}(\theta) = 0\right\}. \tag{A.7}$$

This estimate turns out to be consistent at the usual root-n rate by an argument similar to that given in Chernozhukov, Hong, and Tamer (2007). Next, define $Q_{\mathcal{S}^*}$ as the law of the random variable

$$\mathcal{S}^* = \sup_{\theta \in \widehat{\partial\Theta_0}} \left[V^*(\theta)\right]_+^2. \tag{A.8}$$

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

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

$$\rho_K(Q_{\mathcal{S}^*}, Q_{\mathcal{S}}; \mathbb{R}) = o_p(1). \tag{A.9}$$

Consistency in the Kantorovich–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 4 (Consistent Estimation of Critical Values $k_L(1-\alpha)$ and $k_W(1-\alpha)$). *Suppose Conditions* R.1–R.3 *hold and that we have a consistent estimate of the law of our limit Gaussian processes V. Then the estimates of the laws of the limit statistics $\mathcal{S} = \mathcal{W}$ or $\mathcal{L}$ defined above are consistent. As a consequence, we have that the estimates of the quantiles are consistent in the sense of* (A.1).

It is useful to give a similar result under more primitive conditions, Conditions C.1 and C.2. Recall that in this case, our estimator satisfies

$$\sqrt{n}(\widehat{\gamma} - \gamma) \rightsquigarrow \Omega^{1/2} Z,$$

so that our limit statistics take the form

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

where $t(\theta)$ is a vector-valued weight function; in particular, we have

$$t(\theta) = \left(\nabla_\gamma m(\theta, \gamma)' \Omega^{1/2}\right)/s(\theta) \quad \text{for } \mathcal{S} = \mathcal{L},$$

$$t(\theta) = \left(\nabla_\gamma m(\theta, \gamma)' \Omega^{1/2}\right)/\left(\left\|\nabla_\theta m(\theta, \gamma)\right\| \cdot w(\theta)\right) \quad \text{for } \mathcal{S} = \mathcal{W}.$$

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

$$\rho_K(\mathcal{Q}_{Z^*}, \mathcal{Q}_Z) = o_p(1). \tag{A.10}$$

For instance, we can simulate the distribution using draws $Z^* \sim N(0, I)$ or applying any valid bootstrap method to $\widehat{\Omega}^{-1/2}\sqrt{n}(\widehat{\gamma} - \gamma)$. Alternatively, it is possible to use subsampling (Politis and Romano (1994)). Then the estimate $\mathcal{Q}_{V^*}$ of the law $\mathcal{Q}_{V^*}$ is defined as

$$V^*(\theta) = \widehat{t}(\theta)' Z^*, \tag{A.11}$$

where $\widehat{t}(\theta)$ is a vector-valued weighting function that is uniformly consistent for the weighting function $t(\theta)$. In particular, we can use

$$\widehat{t}(\theta) = \left(\nabla_\gamma m(\theta, \widehat{\gamma})' \widehat{\Omega}^{1/2}\right)/\widehat{s}(\theta) \quad \text{for } \mathcal{S} = \mathcal{L},$$

$$\widehat{t}(\theta) = \left(\nabla_\gamma m(\theta, \widehat{\gamma})' \widehat{\Omega}^{1/2}\right)/\left(\left\|\nabla_\theta m(\theta, \widehat{\gamma})\right\| \cdot w(\theta)\right) \quad \text{for } \mathcal{S} = \mathcal{W}.$$

In this definition, we hold the hatted quantity fixed and the only random element being resampled or simulated is Z^*, with the law denoted as $\mathcal{Q}_{Z^*}$. Then we define the random variable

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

and use its law $\mathcal{Q}_{\mathcal{S}^*}$ to estimate the law $\mathcal{Q}_{\mathcal{S}}$. In the definition of the law $\mathcal{Q}_{\mathcal{S}^*}$ we fix, i.e., condition on, the estimated quantities $\widehat{t}(\cdot)$ and $\widehat{\partial\Theta_0}$.

We can now state the following corollary, which is proven in Appendix C.

COROLLARY 4 (Consistent Estimation of Critical Values Under Conditions C). *Suppose that Conditions C hold and that we have a consistent estimate of the law of Z, so* (A.10) *holds. Then this provides us with a consistent estimate of the law of our limit Gaussian process* $\mathbb{G}$*; hence, all of the conclusions of Theorem 3 hold.*

APPENDIX B: VERIFICATION OF CONDITIONS C FOR MF AND HJ BOUNDS

This appendix establishes the validity of Conditions C for the HJ and MF problems, so that validity of inference and invariance properties for the procedure follow from the general results in Section 2.2, Section 2.3, and Appendix A. As before, let $v := \mathbb{E}[r_{t+1}]$ and $\Sigma := \mathbb{E}[(r_{t+1} - v)(r_{t+1} - v)']$, and let the parameters $\theta = (\mu, \sigma)'$ and $\gamma_{\mathrm{HJ}} = (S_{vv}, S_{v1}, S_{11})'$ for the HJ bounds, where

$$S_{vv} = v'\Sigma^{-1}v, \qquad S_{v1} = 1'_N \Sigma^{-1} v, \quad \text{and} \quad S_{11} = 1'_N \Sigma^{-1} 1_N.$$

Also let $\widehat{\gamma}_{\mathrm{HJ}}$ denote the sample analog obtained by replacing v and Σ with the mean $\widehat{v}$ and variance $\widehat{\Sigma}$ of the sample $r_1, \ldots, r_T$. For the MF bounds, let $\gamma_{\mathrm{MF}} = \mathrm{vec}(D'\Sigma D)$, where

$$B := \mathrm{Cov}_t(Z_{t+1}, r_{t+1}) \quad \text{and} \quad D := [v, B'],$$

let $\widehat{\gamma}_{\mathrm{MF}}$ be the sample analog, and let μ^* and β^* be the target mean portfolio return and factor loadings.

Recall that the Hansen–Jagannathan mean–variance bound for the stochastic discount factor is defined by the moment condition

$$m_{\mathrm{HJ}}(\theta, \gamma_{\mathrm{HJ}}) = \sqrt{S_{vv}\mu^2 - 2S_{v1}\mu + S_{11}} - \sigma,$$

whereas the multifactor efficient (MF) mean–variance set for portfolio returns is characterized by

$$m_{\mathrm{MF}}(\theta, \gamma_{\mathrm{MF}}) := \sqrt{(\mu^*, \beta^{*\prime})(D'\Sigma D)^{-1}(\mu^*, \beta^{*\prime})'} - \sigma.$$

We now give regularity conditions for the validity of inference and confidence intervals based on the LR and Wald statistics derived from the moment functions $\widehat{m}_{\mathrm{HJ}}(\theta)$ and $\widehat{m}_{\mathrm{MF}}(\theta)$, respectively.

PROPOSITION 3 (Verification of Conditions C for HJ and MF Problems). *Suppose that $\widehat{v}, \widehat{\Sigma}$, and $\widehat{B}$ satisfy a CLT. Furthermore, let Θ be a rectangle in $\mathbb{R} \times \mathbb{R}_+$, assume that the absolute values of the elements of v and eigenvalues of Σ are bounded between finite strictly positive constants, and assume that $N\sum_{i=1}^N v_i^2 - (\sum_{i=1}^N v_i)^2$ is bounded away from zero. Then Conditions C holds for the HJ and MF bounds.*

For the last condition, note that, for the purposes of this paper, we treat the number of assets N as finite and that the Cauchy–Schwarz inequality implies that the difference $N\sum_{i=1}^N v_i^2 - (\sum_{i=1}^N v_i)^2 \geq 0$. The difference is strictly positive only if the mean return v_i is not constant across assets.

PROOF OF PROPOSITION 3. First note that γ_{HJ} and γ_{MF} are continuous, differentiable functions of the elements in v, Σ, and B, so that Condition C.1 follows from a CLT for $\widehat{v}$, $\mathrm{vec}(\widehat{\Sigma})$, $\mathrm{vec}(\widehat{B})$, and the delta rule.

Next, we check Condition C.2 for the HJ bounds. Recall that the moment function defining the HJ bound was given by

$$m_{HJ}\big((\mu, \sigma)', \gamma_{HJ}\big) = \sqrt{S_{vv}\mu^2 - 2S_{v1}\mu + S_{11}} - \sigma$$

with S_{vv}, S_{v1}, S_{11} as defined in Section 1.

Since the derivative of $m_{HJ}((\mu, \sigma)', \gamma_{HJ})$ with respect to σ is equal to -1, the lower bound on the norm of the gradient holds for all $\theta \in \Theta$. Also, since the eigenvalues of Σ are bounded away from zero and the elements of $|v|$ are bounded, $|S_{vv}|$ and $|S_{v1}|$ are also bounded. Furthermore, we can easily verify that

$$S_{vv}\mu^2 - 2S_{v1}\mu + S_{11} \geq \frac{S_{vv}S_{11} - S_{v1}^2}{S_{vv}}$$

for all values of μ. Now, since the eigenvalues of Σ^{-1} are bounded away from zero and $v'v1_N'1_N - (1_N'v)^2$ is bounded from below by a positive constant by assumption, we can also bound $S_{vv}\mu^2 - 2S_{v1}\mu + S_{11}$ away from zero.

Noting that the parameter space $\Theta^\delta \times \Gamma$ is compact, it follows that the second derivatives of $m^{(HJ)}(\theta, \gamma_{HJ})$ with respect to γ_{HJ} and θ, respectively, are bounded. In particular, this implies Lipschitz continuity of the gradients $\nabla_\gamma m^{(HJ)}(\theta, \gamma_{HJ})$ and $\nabla_\theta m^{(HJ)}(\theta, \gamma_{HJ})$.

The arguments for the moment function $m_{MF}(\theta, \gamma_{MF})$ for the MF mean–variance set for asset portfolios are completely analogous. □

APPENDIX C: PROOFS

This appendix collects the remaining proofs for the results from the main text and Appendix A.

PROOF OF LEMMA 1. Without loss of generality (w.l.o.g.), let $\max_j g_j = g_1$ and rewrite

$$\lambda\Big(\max_j g_j - H(g_1, \ldots, g_J; \lambda)\Big) = \frac{\displaystyle\sum_{j=1}^{J} \exp\{\lambda(g_j - g_1)\}\lambda(g_1 - g_j)}{\displaystyle\sum_{j=1}^{J} \exp\{\lambda(g_j - g_1)\}}$$

$$= \frac{\displaystyle\sum_{j=1}^{J} \exp\{-h_j\}h_j}{\displaystyle\sum_{j=1}^{J} \exp\{-h_j\}},$$

where $h_j := \lambda(g_1 - g_j) \geq 0$. Clearly, this expression is nonnegative, and since $h_1 = 0$, the denominator is bounded from below by 1. Next note that the function

$$F(h_2, \ldots, h_J) := \frac{\displaystyle\sum_{j=2}^{J} \exp\{-h_j\}h_j}{1 + \displaystyle\sum_{j=2}^{J} \exp\{-h_j\}}$$

is strictly quasi-concave on $\mathbb{R}_+^{J-1}$, so that the usual first-order conditions for a local extremum are sufficient for a global maximum. We can now verify that the first-order conditions for maximization of $F(h_2, \ldots, h_J)$ have the symmetric solution $h_2 = \cdots = h_J = h^* := 1 + W^*$, where $W^* := W(\frac{J-1}{e})$. Note that by definition of the product logarithm, $W^* = \frac{J-1}{e}\exp\{-W^*\} = (J-1)\exp\{-1 - W^*\}$, so that

$$\max_{h_2,\ldots,h_J \geq 0} F(h_1, \ldots, h_J) \equiv \frac{\displaystyle\sum_{j=2}^{J} \exp\{-h^*\}h^*}{1 + \displaystyle\sum_{j=2}^{J} \exp\{-h^*\}} = \frac{(J-1)\exp\{-1 - W^*\}(1 + W^*)}{1 + (J-1)\exp\{-1 - W^*\}}$$

$$= \frac{W^* + (J-1)\exp\{-1 - W^*\}W^*}{1 + (J-1)\exp\{-1 - W^*\}} = W^* = W\left(\frac{J-1}{e}\right).$$

Since $\lambda(\max_j g_j - H(g_1, \ldots, g_J; \lambda)) = F(\lambda(g_1 - g_2), \ldots, \lambda(g_1 - g_J))$, we therefore have that

$$\sup_{g_1,\ldots,g_J} \lambda\left|\max_j g_j - H(g_1, \ldots, g_J; \lambda)\right| = \sup_{h_2,\ldots,h_J \geq 0} |F(h_2, \ldots, h_J)| = W\left(\frac{J-1}{e}\right),$$

which establishes the conclusion. □

Next, we prove four lemmas that will be used to justify the local approximation for the Wald statistic. We consider a (stochastic or deterministic) sequence of moment functions $m_n(\theta) := m(\theta) - q_n$, where $q_n = o_p(1)$, and $m(\theta)$ satisfies Conditions R.1 and R.2 from the main text, and the corresponding sequence of parameter sets $\Theta_n := \{\theta \in \Theta : m_n(\theta) \leq 0\}$.

LEMMA 2. *Suppose the parameter space Θ is compact. Suppose that the gradient $\nabla_\theta m_n(\theta)$ is bounded away from zero uniformly in θ and $n = 1, 2, \ldots$, and is Lipschitz continuous in θ with Lipschitz constant $L < \infty$. Also let θ_n be any sequence such that θ_n approaches the boundary of Θ_n, i.e., $d(\theta_n, \partial\Theta_n) \to 0$. Then there exists $\delta > 0$ such that the projection of θ_n on Θ_n is unique for all such sequences whenever $d(\theta_n, \Theta_n) < \delta$.*

PROOF. Suppose the statement was not true. Then for some sequence θ_n, we could construct a subsequence $\theta_{k(n)}$ such that there are (at least) two distinct projections of $\theta_{s(n)}$ onto $\partial\Theta_n$ for each n. By compactness of Θ, $\theta_{k(n)}$ has a convergent sub-subsequence $\theta_{b(n)}$

with $\lim_n \theta_{b(n)} = \theta_0$, say. Since by construction every member of $\theta_{b(n)}$ has two distinct projections onto $\partial \Theta_n$, we can inscribe a ball of radius $r_n := d(\theta_{b(n)}, \partial \Theta_n)$ centered at $\theta_{b(n)}$ into Θ / Θ_n such that this ball has at least two distinct points $(\theta^*_{1,b(n)}, \theta^*_{2,b(n)})$ in common with $\partial \Theta_n$.

By properties of the projection, the radii of these balls corresponding to the projection points, $N(\theta^*_{j,b(n)}) := r_n^{-1}(\theta_{b(n)} - \theta^*_{j,b(n)})$ for $j = 1, 2$, are also normal vectors to the surface $\partial \Theta_n$ at $\theta^*_{1,b(n)}$ and $\theta^*_{2,b(n)}$, respectively. Note that since the gradient $\nabla_\theta m_n(\theta)$ is bounded away from zero, we can w.l.o.g. normalize the length of the normal vectors of the surface $\partial \Theta_n$ to 1.

Note that the two points $\theta^*_{1,b(n)}$ and $\theta^*_{2,b(n)}$ are equidistant to $\theta_{b(n)}$, and therefore lie on a one-dimensional sphere $S^1(r_n, \theta_{b(n)}) \subset \Theta$ with center $\theta_{b(n)}$ and radius r_n. The curve $\eta(s)$ that corresponds to a sphere of radius r, where s is the arc length, has constant curvature $\kappa(s) \equiv \frac{1}{r_n}$, so that the normal vectors (with length normalized to 1) $N(\theta^*_{j,b(n)}) := r_n^{-1}(\theta_{b(n)} - \theta^*_{j,b(n)})$ for $j = 1, 2$ satisfy

$$\left\| N(\theta^*_{2,b(n)}) - N(\theta^*_{1,b(n)}) \right\| \gtrsim r_n^{-1} \left\| \theta^*_{2,b(n)} - \theta^*_{1,b(n)} \right\|.$$

Since $r_n \to 0$, there is no upper bound on r_n^{-1}, so that $N(\theta)$ is not Lipschitz-continuous.

However, the normal vector of $\partial \Theta_0$ at θ standardized to length 1 is given by

$$N(\theta) := \left\| \nabla_\theta m(\theta) \right\|^{-1} \nabla_\theta m(\theta).$$

Condition R.2 implies that $N(\theta)$ is Lipschitz-continuous in θ with a constant $1/\delta < \infty$, which leads to a contradiction. Finally, note that since that Lipschitz constant L was assumed to be uniform in γ, there exists $\delta := \delta(L) > 0$ such that projection is unique whenever $r_n := d(\theta_n, \Theta_n) < \delta$ along all sequences θ_n and $m_n(\cdot)$ □

LEMMA 3. *Suppose Conditions R.1 and R.2 hold, and let θ_n be a sequence such that $d(\theta_n, \Theta_0) \to 0$. Define $\bar{\theta}_n \equiv \arg\min_{\widehat{m}(\theta) \leq 0} \|\theta_n - \theta\|^2$. Then $\bar{\theta}_n$ is uniquely defined with probability approaching 1 (w.p.a.1) and it satisfies $d(\bar{\theta}_n, \Theta_0) \xrightarrow{p} 0$.*

PROOF. For a given value of $\delta > 0$, let $\eta := \min\{m(\theta) : d(\theta, \Theta_0) \geq \delta\}$. By continuity of $m(\theta)$ and compactness of Θ, $\eta > 0$. By Condition R.1, $\sup_{\theta \in \Theta_0} \widehat{m}(\theta) < \eta$ w.p.a.1 and, therefore, $d(\bar{\theta}_n, \Theta_0) < \delta$ w.p.a.1. Uniqueness follows from Lemma 2, noting that the conclusion holds regardless of whether the sequence $m_n(\theta)$ is deterministic or stochastic.   □

LEMMA 4. *Suppose Conditions R.1 and R.2 hold, and consider a sequence of parameter values $\theta_n \in \Theta$ such that $\theta_n \to \partial \Theta_0$. For a sequence $q_n = o_p(1)$ and $\widehat{\Theta}_n := \{\theta \in \Theta : \widehat{m}(\theta) \leq q_n\}$, we have*

$$\sqrt{n} d(\theta_n, \widehat{\Theta}_{0,\delta}) = \left[\left\| \nabla_\theta m(\theta_n) \right\|^{-1} \left(\mathbb{G}_n(\theta_n) + \sqrt{n}(m(\theta_n) - q_n) \right) + o_p(1) \right]_+.$$

In particular, for $q_n \equiv 0$, we have

$$\sqrt{n} d(\theta_n, \widehat{\Theta}_{0,\delta}) = \left[\left\| \nabla_\theta m(\theta_n) \right\|^{-1} \left(\mathbb{G}_n(\theta_n) + \sqrt{n} m(\theta_n) \right) + o_p(1) \right]_+.$$

PROOF. Without loss of generality, we only consider the case $q_n \equiv 0$. Note that for the general case, we can apply the same argument to the functions $\widehat{m}(\theta) - q_n$ and $m(\theta) - q_n$, which inherit the properties of Conditions R.1 and R.2.

The projection of θ_n onto the set $\widehat{\Theta}_{0,\delta} := \{\theta \in \Theta^\delta : \widehat{m}(\theta) \le 0\}$ is given by

$$\bar{\theta}_n := \arg \min_{\theta' \in \Theta^\delta : \widehat{m}(\theta') \le 0} \left\| \theta_n - \theta' \right\|^2.$$

The Lagrangian for this constrained minimization problem is $\|\theta_n - \theta'\|^2 + 2\lambda \widehat{m}(\theta')$, where the Kuhn–Tucker multiplier $\lambda \ge 0$. Since $\delta > 0$ is fixed and $d(\theta_n, \Theta_0) \to 0$, it follows from Lemma 3 that $\bar{\theta}_n$ is contained in the interior of Θ^δ w.p.a.1. Hence for n large enough, the quantity $\bar{\theta}_n$ can be taken to be an interior solution of the saddle-point problem

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

We can now use a mean-value expansion to obtain

$$(\bar{\theta}_n - \theta_n) + \nabla_\theta \widehat{m}(\bar{\theta}_n)\lambda = 0,$$
$$\widehat{m}(\theta_n) + \nabla_\theta m(\check{\theta}_n)'(\bar{\theta}_n - \theta_n) = 0$$

for an intermediate value $\check{\theta}_n$. By Lemma 3, $\nabla_\theta \widehat{m}(\check{\theta}_n) = \nabla_\theta m(\theta_n) + o_p(1)$ and $\nabla_\theta m(\bar{\theta}_n) = \nabla_\theta m(\theta_n) + o_p(1)$. Hence, solving for $(\bar{\theta}_n - \theta_n)$ and applying the extended continuous mapping theorem, we obtain

$$\bar{\theta}_n - \theta_n = -\left[\nabla_\theta m(\theta_n)\left(\nabla_\theta m(\theta_n)'\nabla_\theta m(\theta_n)\right)^{-1} + o_p(1)\right]\widehat{m}(\theta_n).$$

Now, since by Condition R.2, $\sqrt{n}(\widehat{m} - m) = \mathbb{G}_n = O_p(1)$ in $\ell^\infty(\Theta^\delta)$, we obtain

$$\sqrt{n}(\bar{\theta}_n - \theta_n) = -\nabla_\theta m(\theta_n)\left(\nabla_\theta m(\theta_n)'\nabla_\theta m(\theta_n)\right)^{-1}\left(\mathbb{G}_n(\theta_n) + \sqrt{n}m(\theta_n)\right)$$
$$+ o_p(1).$$

Thus we have that

$$\sqrt{n}d(\theta_n, \widehat{\Theta}_n^\delta) = \sqrt{n}\|\bar{\theta}_n - \theta_n\|\mathbb{1}\{\widehat{m}(\theta_n) > 0\}$$
$$= \left|\left(\nabla_\theta m(\theta_n)'\nabla_\theta m(\theta_n)\right)^{-1/2}\left(\mathbb{G}_n(\theta_n) + \sqrt{n}m(\theta_n)\right) + o_p(1)\right|$$
$$\times \mathbb{1}\{\mathbb{G}_n(\theta_n) + \sqrt{n}m(\theta_n) > 0 + o_p(1)\}$$
$$= \left[\left\|\nabla_\theta m(\theta_n)\right\|^{-1}\left(\mathbb{G}_n(\theta_n) + \sqrt{n}m(\theta_n)\right) + o_p(1)\right]_+,$$

which concludes the proof. □

C.1 *Proof of Theorem* 1

Part 1: Limit law of $\mathcal{L}_n$. Let $\mathbb{G}_n = \sqrt{n}(\widehat{m} - m)$. We first show that

$$
\begin{aligned}
\mathcal{L}_n &= \sup_{\theta \in \Theta_0} \left[\sqrt{n}\widehat{m}(\theta)/\widehat{s}(\theta)\right]_+^2 = \sup_{\theta \in \Theta_0} \left[(\mathbb{G}_n(\theta) + \sqrt{n}m(\theta))/\widehat{s}(\theta)\right]_+^2 \\
&=_{(1)} \sup_{\theta \in \partial\Theta_0} \left[(\mathbb{G}_n(\theta) + \sqrt{n}m(\theta))/\widehat{s}(\theta) + o_p(1)\right]_+^2 \\
&=_{(2)} \sup_{\theta \in \partial\Theta_0} \left[(\mathbb{G}_n(\theta) + \sqrt{n}m(\theta))/s(\theta) + o_p(1)\right]_+^2 \\
&\rightsquigarrow_{(3)} \sup_{\theta \in \partial\Theta_0} \left[(\mathbb{G}(\theta) + \sqrt{n}m(\theta))/s(\theta)\right]_+^2,
\end{aligned}
$$

where the steps follows from the following arguments.

To show step (1), note that

$$
\sup_{\theta \in \Theta_0} \left[(\mathbb{G}_n(\theta) + \sqrt{n}m(\theta))/\widehat{s}(\theta)\right]_+^2 \geq \sup_{\theta \in \partial\Theta_0} \left[(\mathbb{G}_n(\theta) + \sqrt{n}m(\theta))/\widehat{s}(\theta)\right]_+^2.
$$

So we need to show that w.p.a.1,

$$
\sup_{\theta \in \Theta_0} \left[(\mathbb{G}_n(\theta) + \sqrt{n}m(\theta))/\widehat{s}(\theta)\right]_+^2 \leq \sup_{\theta \in \partial\Theta_0} \left[(\mathbb{G}_n(\theta) + \sqrt{n}m(\theta))/\widehat{s}(\theta) + o_p(1)\right]_+^2.
$$

To show this, take a sequence $\theta_n \in \Theta_0$ such that

$$
\sup_{\theta \in \Theta_0} \left[(\mathbb{G}_n(\theta) + \sqrt{n}m(\theta))/\widehat{s}(\theta)\right]_+^2 = \left[(\mathbb{G}_n(\theta_n) + \sqrt{n}m(\theta_n))/\widehat{s}(\theta_n)\right]_+^2,
$$

where the supremum is attained by compactness of Θ_0, continuity hypotheses on $\widehat{m}$, m, and $\widehat{s}$, and uniform positivity of $\widehat{s}$. Since $\sup_{\theta \in \Theta} |\mathbb{G}_n| = O_p(1)$, θ_n can be selected to obey

$$
\sqrt{n}m(\theta_n)/\widehat{s}(\theta_n) = O_p(1).
$$

We now show that θ_n approaches $\partial\Theta_0$ as n increases: fix $\kappa > 0$, and let

$$
\zeta(\kappa) := \sup\{m(\theta)/s(\theta) : \theta \in \Theta_0, d(\theta, \partial\Theta_0) \geq \kappa\}.
$$

Next, note that Condition R.2 implies that $m(\theta) < 0$ for each $\theta \in \text{int}\,\Theta_0$: suppose this was not true, but that $m(\theta_0) = 0$ for some $\theta_0 \in \text{int}\,\Theta_0$. Since for $\theta \in \Theta_0$, $m(\theta) \leq 0$, then θ_0 represents a local maximum of $m(\cdot)$ in the interior of Θ_0. Since $m(\theta)$ is differentiable, this implies that $\nabla_\theta m(\theta_0) = 0$. However, Condition R.2 implies that $\|\nabla_\theta m(\theta)\| > 0$, a contradiction. Therefore, $m(\theta) < 0$ for each $\theta \in \text{int}\,\Theta_0$, so that by continuity and compactness, $\zeta(\kappa) < 0$.

By Conditions R.1 and R.3, we have $\sup_{\theta \in \Theta_0} |\widehat{m}(\theta)/\widehat{s}(\theta) - m(\theta)/s(\theta)| < -\zeta$ w.p.a.1 for any strictly negative value of ζ. Hence, the supremum of $\sqrt{n}[\widehat{m}(\theta)/\widehat{s}(\theta)]_+^2$ for values $\theta \in \Theta_0$ such that $d(\theta, \partial\Theta_0) \geq \kappa$ is equal to zero w.p.a.1, so that the supremum over all of Θ_0 is attained at a value of θ such that $d(\theta, \partial\Theta_0) < \kappa$. Since κ can be chosen arbitrarily

close to zero, we can construct a sequence of positive random variables $\kappa_n = o_p(1)$ such that

$$d(\theta_n, \partial\Theta_0) \leq \kappa_n$$

w.p.a.1. We conclude that

$$\sup_{\theta\in\Theta_0} \left[(\mathbb{G}_n(\theta) + \sqrt{n}m(\theta))/\widehat{s}(\theta) \right]_+^2$$

$$\leq \sup_{\theta\in\partial\Theta_0, \theta+\eta\in\Theta_0, \|\eta\|\leq\kappa_n} \left[(\mathbb{G}_n(\theta+\eta) + \sqrt{n}m(\theta+\eta))/\widehat{s}(\theta+\eta) \right]_+^2.$$

Using stochastic equicontinuity of $\mathbb{G}_n$ implied by Condition R.1, the last quantity is equal to

$$\sup_{\theta\in\partial\Theta_0, \theta+\eta\in\Theta_0, \|\eta\|\leq\kappa_n} \left[(\mathbb{G}_n(\theta) + \sqrt{n}m(\theta+\eta))/\widehat{s}(\theta+\eta) + o_p(1) \right]_+^2$$

for some $o_p(1)$ term. Because $\sqrt{n}m(\theta+\eta) \leq 0$ for $\theta+\eta \in \Theta_0$ and $m(\theta) = 0$ for $\theta \in \partial\Theta_0$, we conclude that the last quantity is equal to

$$\sup_{\theta\in\partial\Theta_0} \left[(\mathbb{G}_n(\theta)/\widehat{s}(\theta+\eta) + o_p(1) \right]_+^2$$

for the same $o_p(1)$ term. This verifies equality (1).

Equality (2) follows from using Condition R.3 and the fact that $\sup_{\theta\in\Theta} |\mathbb{G}_n(\theta)| = O_p(1)$ is implied by Condition R.1. Equality (3) follows from the application of Condition R.1 and the continuous mapping theorem.

Part 2: Limit Law of $\mathcal{W}_n$. Recall that we define the set estimator $\widehat{\Theta}_{0,\delta}$ as

$$\widehat{\Theta}_{0,\delta} := \left\{ \theta \in \Theta^\delta : \widehat{m}(\theta) \leq 0 \right\}.$$

Analogous to part 1, we establish the conclusion by the steps

$$\sqrt{\mathcal{W}_n} = \sup_{\theta\in\Theta_0} \sqrt{n}d(\theta, \widehat{\Theta}_{0,\delta})/\widehat{w}(\theta)$$

$$=_{(1)} \sup_{\theta\in\partial\Theta_0} \left[(\mathbb{G}_n(\theta) + \sqrt{n}m(\theta))/\widehat{w}(\theta) + o_p(1) \right]_+$$

$$=_{(2)} \sup_{\theta\in\partial\Theta_0} \left[(\mathbb{G}_n(\theta) + \sqrt{n}m(\theta))/w(\theta) + o_p(1) \right]_+$$

$$\rightsquigarrow_{(3)} \sup_{\theta\in\partial\Theta_0} \left[(\mathbb{G}(\theta) + \sqrt{n}m(\theta))/w(\theta) \right]_+,$$

where steps (1)–(3) are proven as follows.

To establish step (1), we first show that

$$\sup_{\theta\in\Theta_0} \sqrt{n}d(\theta, \widehat{\Theta}_{0,\delta})/\widehat{w}(\theta) = \sup_{\Theta_n} \sqrt{n}d(\theta, \widehat{\Theta}_{0,\delta})/\widehat{w}(\theta) \tag{C.1}$$

holds w.p.a.1, where

$$\Theta_n = \big\{\theta \in \Theta_0 : d(\theta, \partial\Theta_0)/w(\theta) \le \kappa_n\big\}$$

and κ_n is some sequence of positive random variables converging to zero in probability, $\kappa_n = o_p(1)$. Note that the right-hand side of (C.1) is less than or equal to the left-hand side of (C.1) by construction, so we only need to show that w.p.a.1, the right-hand side cannot be less. To this end, fix some $\kappa > 0$ and note that using the same line of reasoning as for part 1, we have

$$\zeta(\kappa) := \sup\big\{m(\theta) : \theta \in \Theta_0, d(\theta, \partial\Theta_0) \ge \kappa\big\} < 0.$$

Furthermore, for every $\zeta < 0$, we have that w.p.a.1, $\sup_{\theta \in \Theta} |\widehat{m}(\theta) - m(\theta)| < -\zeta$ and, therefore, each $\theta \in \Theta_0$ with $d(\theta, \partial\Theta_0) \ge \kappa$ is included in $\widehat{\Theta}_{0,\delta}$. In that event, we can only have $d(\theta, \widehat{\Theta}_{0,\delta}) > 0$ for values of θ within a distance κ of the boundary $\partial\Theta_0$. Since κ can be chosen arbitrarily small, we can choose a sequence $\kappa_n = o(1)$ such that the right-hand side in (C.1) holds with probability converging to 1.

Next, let θ_n be a sequence such that

$$\sup_{\theta \in \Theta_0} \sqrt{n}d(\theta, \widehat{\Theta}_{0,\delta})/\widehat{w}(\theta) = \sqrt{n}d(\theta_n, \widehat{\Theta}_{0,\delta})/\widehat{w}(\theta_n).$$

Note that since $d(\theta, \widehat{\Theta}_{0,\delta})$ and $\widehat{w}(\theta)$ are continuous functions of θ, $\widehat{w}(\theta)$ is bounded away from zero w.p.a.1, so that since Θ_0 is compact, the supremum is attained w.p.a.1. Since $\delta > 0$ is fixed, by Lemma 3 w.p.a.1 we have $d(\bar{\theta}_n, \theta_n)/\widehat{w}(\theta) < \delta$, so that $\bar{\theta}_n$ is contained in the interior of Θ^δ w.p.a.1.

It follows from Lemma 4 that

$$\sqrt{n}d(\theta_n, \widehat{\Theta}_{0,\delta}) = \big[\big\|\nabla_\theta m(\theta_n)\big\|^{-1}\big(\mathbb{G}_n(\theta_n) + \sqrt{n}m(\theta_n)\big) + o_p(1)\big]_+. \tag{C.2}$$

Hence, dividing by $\widehat{w}(\theta)$ and using stochastic equicontinuity of $\mathbb{G}_n(\theta)$, we can follow the same line of reasoning as in part 1 to obtain that w.p.a.1,

$$\sup_{\theta \in \Theta_0} \sqrt{n}d(\theta, \widehat{\Theta}_{0,\delta})/\widehat{w}(\theta) \le \sup_{\theta \in \partial\Theta_0, \theta+\eta \in \Theta_0, \|\eta\| \le \kappa_n}\left[\frac{\mathbb{G}_n(\theta) + \sqrt{n}m(\theta + \eta)}{\|\nabla_\theta m(\theta)\|\widehat{w}(\theta + \eta)} + o_p(1)\right]_+.$$

Noting again that $\sqrt{n}m(\theta + \eta) \le 0$ for $\theta + \eta \in \Theta_0$, we can bound the expression by

$$\sup_{\theta \in \Theta_0} \sqrt{n}d(\theta, \widehat{\Theta}_{0,\delta})/\widehat{w}(\theta) \le \sup_{\theta \in \partial\Theta_0, \theta+\eta \in \Theta_0, \|\eta\| \le \kappa_n}\left[\frac{\mathbb{G}_n(\theta)}{\|\nabla_\theta m(\theta)\|\widehat{w}(\theta + \eta)} + o_p(1)\right]_+,$$

where again the $o_p(1)$ term is the same as in the first inequality.

Therefore, we can use uniform convergence of the weighting function $\widehat{w}$ to w from Condition R.3 and the continuous mapping theorem to obtain

$$\mathcal{W}_n \rightsquigarrow \sup_{\theta \in \partial\Theta_0} \big[\big(\big\|\nabla_\theta m(\theta)\big\|w(\theta)\big)^{-1}\mathbb{G}(\theta)\big]_+^2, \tag{C.3}$$

where the intermediate steps are analogous to equalities (2) and (3) in the proof for $\mathcal{L}_n$.

Part 3: Continuity of the Limit Distributions. The continuity of the distribution function $\mathcal{L}$ on $(0, \infty)$ follows from Davydov, Lifshits, and Smorodina (1998) and from the assumption that the covariance function of $\mathbb{G}$ is nondegenerate, i.e., $\inf_{\theta \in \Theta} \text{Var}(\mathbb{G}(\theta)) > 0$. The probability that $\mathcal{L}$ is greater than zero is equal to the probability that $\max_j \sup_{\theta \in \Theta} \mathbb{G}_j(\theta) > 0$, which is greater than the probability that $\mathbb{G}_{j'}(\theta') > 0$ 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. □

C.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 the corollary. □

C.3 Proof of Theorem 2

We have that $P[\Theta_0 \subseteq R_{\text{LR}}] = P[\mathcal{L}_n \leq \widehat{k}_{\text{L}}(1 - \alpha)]$ 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,

$$P[\mathcal{L}_n \leq \widehat{k}_{\text{L}}(1 - \alpha)] \leq P[\mathcal{L}_n \leq k_{\text{L}}(1 - \alpha) + \epsilon] + o(1) \to P[\mathcal{L} \leq k_{\text{L}}(1 - \alpha) + \epsilon],$$
$$P[\mathcal{L}_n \leq \widehat{k}_{\text{L}}(1 - \alpha)] \geq P[\mathcal{L}_n \leq k_{\text{L}}(1 - \alpha) - \epsilon] - o(1) \to P[\mathcal{L} \leq k_{\text{L}}(1 - \alpha) - \epsilon].$$

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

$$P[\mathcal{L}_n \leq \widehat{k}_{\text{L}}(1 - \alpha)] \to P[\mathcal{L} \leq k_{\text{L}}(1 - \alpha)] = (1 - \alpha).$$

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

C.4 Proof of Theorem 3

We give the proof only for confidence regions based on the LR statistic. The arguments for the Wald-type confidence sets are completely analogous. Let δ_n be a null sequence where $\delta_n > 0$ for all n and $\sqrt{n}\delta_n \to \infty$. To show convergence with respect to Hausdorff distance, we establish that w.p.a.1, (a) $R_{\text{LR}} \subset \Theta_0^{\delta_n}$ and (b) $\Theta_0 \subset R_{\text{LR}}^{\delta_n}$, where for $\delta > 0$, $A^\delta := \{x \in \mathbb{R}^k : d(x, A) < \delta\}$ denotes the δ-expansion of a set A in $\mathbb{R}^k$.

To prove statement (a), consider a sequence $\theta_n \in \Theta \setminus \Theta_0^{\delta_n}$. We have to show that $\theta_n \notin R_{\text{LR}}$ w.p.a.1 as n increases: For any fixed $\bar{\delta} > 0$, let $\eta := \inf_{\theta : d(\theta, \Theta_0) \geq \bar{\delta}} m(\theta)$. Since Θ is compact and $m(\theta)$ is continuous by Condition R.2, it follows from the definition of Θ_0 that $\eta > 0$. Hence, along any sequence θ_n such that $d(\theta_n, \Theta_0) \geq \bar{\delta}$ for all n, we have $\sqrt{n}m(\theta_n) \to \infty$. Since $\theta \in R_{\text{LR}}$ if and only if for the critical value $\widehat{k}_{\text{L}}$, $\mathcal{L}\mathcal{R}_n \leq \widehat{k}_{\text{L}}$, where $\widehat{k}_{\text{L}}$ is tight for any n large enough, it follows that $\theta_n \notin R_{\text{LR}}$ w.p.a.1. Hence we can restrict our attention to (sub-) sequences for which $d(\theta_n, \Theta_0) \to 0$.

Now suppose that $d(\theta_n, \Theta_0) \to 0$ and $\sqrt{n}d(\theta_n, \Theta_0) \to \infty$. Let θ_n^* be the projection of θ_n onto $\partial\Theta_0$, where θ_n^* need not converge to a particular point. Then using the same

steps as in the proof of Lemma 4, it follows from a mean-value expansion of $m(\theta)$ around θ_n and continuity of the gradient $\nabla_\theta m(\theta)$ that $\sqrt{n}m(\theta_n) \to \infty$. Hence the LR statistic diverges to infinity and $\theta_n \notin R_{\mathrm{LR}}$ w.p.a.1.

It remains to check claim (b), namely that $\Theta_0 \subset R_{\mathrm{LR}}^{\delta_n}$. To this end, consider a sequence $\theta_n \in \Theta_0$ and let $q_n = \widehat{s}(\theta)\sqrt{\widehat{k}_{\mathrm{L}}(1-\alpha)/n}$. Also define $\bar{\theta}_n \equiv \arg\min_{\widehat{m}(\theta) \leq q_n} \|\theta_n - \theta\|^2$, so that $d(\theta_n, \widehat{\Theta}_{0,\delta}) = \|\theta_n - \bar{\theta}_n\|$. Now note that Lemma 3 implies that $d(\bar{\theta}_n, \Theta_0) \overset{p}{\to} 0$, so that we can apply Lemma 4 to obtain

$$\sqrt{n}d\|\bar{\theta}_n - \theta_n\| = \left[\left\|\nabla_\theta m(\theta_n)\right\|^{-1}\left(\mathbb{G}_n(\theta_n) + \sqrt{n}\left(m(\theta_n) - q_n\right)\right) + o_p(1)\right]_+.$$

By the assumption of Condition R.2, the norm of $\nabla_\theta m(\theta_n)$ is bounded away from zero, and $\mathbb{G}_n(\theta_n)$ is stochastically bounded by Condition R.1. Since $\widehat{s}(\theta)$ and $\widehat{k}$ are stochastically bounded by Condition R.3, $\sqrt{n}q_n$ is also stochastically bounded.

Noting that $m(\theta_n) \leq 0$, it follows that $\sqrt{n}(\bar{\theta}_n - \theta_n) = O_p(1)$. Therefore,

$$\sup_{\theta \in \Theta_0} \sqrt{n}d(\theta, R_{\mathrm{LR}}) = O_p(1),$$

so that for any sequence δ_n such that $\sqrt{n}\delta_n \to \infty$, we have $\sup_{\theta \in \Theta_0} d(\theta, R_{\mathrm{LR}})/\delta_n \to 0$. Hence $\Theta_0 \subset R_{\mathrm{LR}}^{\delta_n}$ w.p.a.1, which concludes the proof.   □

C.5 *Proof of Corollary 3*

This corollary immediately follows from the assumed conditions and Corollary 1.   □

C.6 *Proof of Theorem 4*

In what follows, let A denote an absolute positive constant. We have by definition of the Kantarovich–Rubinstein metric that

$$\mathrm{E}_{\mathcal{Q}_{V*}}\left[\varphi(V^*)\right] - \mathrm{E}_{\mathcal{Q}_V}\left[\varphi(V)\right] = o_p(1) \quad \text{uniformly in } \varphi \in \mathrm{BL}_1\big(C(\Theta)\big).$$

This implies that

$$\mathrm{E}_{\mathcal{Q}_{V*}}\left[\varphi\big([V^*]_+^2\big)\right] - \mathrm{E}_{\mathcal{Q}_V}\left[\varphi\big([V]_+^2\big)\right] = o_p(1) \quad \text{uniformly in } \varphi \in \mathrm{BL}_1\big(C(\Theta)\big),$$

since the composition $\varphi \circ [\cdot]_+^2 \in A \cdot \mathrm{BL}_1(C(\Theta))$ for $\varphi \in \mathrm{BL}_1(C(\Theta))$. This further implies that

$$\mathrm{E}_{\mathcal{Q}_{V*}}\left[\varphi'\Big(\sup_{R_n}[V^*]_+^2\Big)\right] - \mathrm{E}_{\mathcal{Q}_V}\left[\varphi'\Big(\sup_{R_n}[V]_+^2\Big)\right] = o_p(1) \quad \text{uniformly in } \varphi' \in \mathrm{BL}_1(\mathbb{R}),$$

since the composition $\varphi'(\sup_{R_n}[\cdot]_+^2) \in A \cdot \mathrm{BL}_1(C(\Theta))$ for $\varphi' \in \mathrm{BL}_1(\mathbb{R})$ and R_n denoting any sequence of closed nonempty subsets in Θ. Therefore, by the extended continuous mapping theorem (see, e.g., Theorem 18.11 in van der Vaart (1998)),

$$\mathrm{E}_{\mathcal{Q}_{V*}}\left[\varphi'\Big(\sup_{\partial\Theta_0}[V^*]_+^2\Big)\right] - \mathrm{E}_{\mathcal{Q}_V}\left[\varphi'\Big(\sup_{\partial\Theta_0}[V]_+^2\Big)\right] = o_p(1) \quad \text{uniformly in } \varphi' \in \mathrm{BL}_1(\mathbb{R}).$$

(Note that here we compute expectations over V^* and V taking $\widehat{\partial\Theta}_0$ as given; note that our bootstrap method treats $\widehat{\partial\Theta}_0$ as fixed.) Also note that for any sequence of sets R_n converging to a set R, we have that

$$\left|\mathrm{E}_{\mathcal{Q}_V}\left[\varphi'\left(\sup_{R_n}[V]_+^2\right) - \varphi'\left(\sup_{R}[V]_+^2\right)\right]\right|$$

$$\leq \mathrm{E}_{\mathcal{Q}_V}\left[\left|\sup_{R_n}[V]_+^2 - \sup_{R}[V]_+^2\right| \wedge 1\right] = o_p(1) \quad \text{uniformly in } \varphi' \in \mathrm{BL}_1(\mathbb{R}),$$

since

$$\sup_{R_n}[V]_+^2 - \sup_{R}[V]_+^2 = o_p(1)$$

by stochastic equicontinuity of the process V. Since by Condition R.1–R.3 and Theorem 3, $\widehat{\partial\Theta}_0$ converges to $\partial\Theta_0$ in the Hausdorff distance, we have by the extended continuous mapping theorem

$$\left|\mathrm{E}_{\mathcal{Q}_V}\left[\varphi'\left(\sup_{\widehat{\partial\Theta}_0}[V]_+^2\right) - \varphi'\left(\sup_{\partial\Theta_0}[V]_+^2\right)\right]\right| = o_p(1) \quad \text{uniformly in } \varphi' \in \mathrm{BL}_1(\mathbb{R}),$$

where $\mathcal{Q}_V$ computes the expectation over V, treating $\widehat{\partial\Theta}_0$ as fixed.

Combining the steps above, we conclude by the triangle inequality that

$$\left|\mathrm{E}_{\mathcal{Q}_{V^*}}\left[\varphi'\left(\sup_{\widehat{\partial\Theta}_0}[V^*]_+^2\right) - \varphi'\left(\sup_{\partial\Theta_0}[V]_+^2\right)\right]\right| = o_p(1) \quad \text{uniformly in } \varphi' \in \mathrm{BL}_1(\mathbb{R}),$$

which is the same as

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

(Note that the bootstrap random variable $\mathcal{S}^*$ is computed having fixed $\widehat{\partial\Theta}_0$.)

It is known 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 $\mathcal{S}_n$ defined on probability space $(\Omega', \mathcal{F}', P_n)$, implies the convergence of the distribution function

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

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

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

at each continuity point p of the mapping $s \mapsto \inf\{s : \mathrm{Pr}_{\mathcal{Q}_{\mathcal{S}}}[\mathcal{S} \leq s] \geq p\}$. 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 $\rho_K(\mathcal{Q}_{\mathcal{S}^*}, \mathcal{Q}_{\mathcal{S}}) = o_p(1)$, we obtain the convergence in probability of the distribution function

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

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 $s \mapsto \inf\{s : \Pr_{\mathcal{Q}_{\mathcal{S}}}[\mathcal{S} \leq s] \geq p\}$.   □

C.7 *Proof of Corollary* 4

To prove this corollary, it suffices to show that

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

Without loss of generality, we can take $\sup_\theta \|\widehat{t}(\theta)\| \leq 1$ and $\sup_\theta \|t(\theta)\| \leq 1$. The claim will then follow from

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

That $\rho_K(\mathcal{Q}_{\widehat{t}'Z^*}, \mathcal{Q}_{\widehat{t}'Z}; C(\Theta)) = o_p(1)$ follows immediately from $\rho_K(\mathcal{Q}_{Z^*}, \mathcal{Q}_Z) = o_p(1)$ and $\varphi(\widehat{t}' \cdot) \in \mathrm{BL}_1(\mathbb{R}^k)$. Indeed,

$$\left| \varphi\big(\widehat{t}(\theta)' z(\theta)\big) - \varphi\big(\widehat{t}(\theta)' \bar{z}(\theta)\big) \right| \leq \sup_\theta \left| \widehat{t}(\theta)'\big(z(\theta) - \bar{z}(\theta)\big) \right| \wedge 2$$

$$\leq \left[\left(\sup_\theta \|\widehat{t}(\theta)\| \sup_\theta \|z(\theta) - \bar{z}(\theta)\| \right) \wedge 2 \right]$$

$$\leq \left[\sup_\theta \|z(\theta) - \bar{z}(\theta)\| \wedge 2 \right].$$

That $\rho_K(\mathcal{Q}_{\widehat{t}'Z}, \mathcal{Q}_{t'Z}; C(\Theta)) = o_p(1)$ follows because, uniformly in $\varphi \in \mathrm{BL}_1(C(\Theta))$,

$$\left| \mathrm{E}_{\mathcal{Q}_Z}[\varphi(\widehat{t}'Z)] - \varphi(t'Z) \right| \leq \mathrm{E}_{\mathcal{Q}_Z}\left[\sup_\theta \left| \big(\widehat{t}(\theta) - t(\theta)\big)' Z(\theta) \right| \wedge 2 \right]$$

$$\leq \mathrm{E}_{\mathcal{Q}_Z}\left[\sup_\theta \|\widehat{t}(\theta) - t(\theta)\| \sup_\theta \|Z(\theta)\| \wedge 2 \right] = o_p(1),$$

where $\mathrm{E}_{\mathcal{Q}_Z}$ computes the expectation over Z, treating $\widehat{t}$ as fixed.   □

C.8 *Proof of Proposition* 2

The proof for this result is obvious and is, therefore, omitted from the paper.

C.9 *Proof of Proposition* 1

For the LR-type statistic, it is sufficient to notice that $\mathcal{L}_n(\theta; \eta) = \mathcal{L}_n(\eta^{-1}(\eta(\theta))) = \frac{\sqrt{n}\widehat{m}_n(\theta)}{\widehat{s}(\theta)}$ and, therefore, only depends on quantities evaluated at θ. Hence,

$$\mathcal{L}_n = \sup_{\theta \in \Theta_0} \mathcal{L}_n(\theta) = \sup_{\eta \in \eta(\Theta_0)} \mathcal{L}_n\big(\eta^{-1}(\eta)\big)$$

and $\phi_{Ln}(\theta; \eta)$ is invariant with respect to parameter transformations. Asymptotic similarity follows from Condition R.1 and the continuous mapping theorem.

Next, define $\widehat{H}_{0,\delta} = \{\eta \in \eta(\Theta^\delta) : \widehat{m}(\eta) \leq 0\}$, noting that $\widehat{H}_{0,\delta} = \eta(\widehat{\Theta}_{0,\delta})$. We now distinguish three cases regarding the limit point of the sequence θ_n: (i) $\theta_0 \in \operatorname{int} \Theta_0$, the interior of Θ_0, (ii) $\theta \in \Theta/\Theta_0$, and (iii) $\theta \in \partial\Theta_0$.

Since the weighting functions $\widehat{w}(\theta; \eta)$ were assumed to satisfy Condition R.3, Theorem 3 implies that in case (i), $\phi_{Wn}(\theta; \eta) \to 0$ for all values of η, including the identity transformation, so that by continuity of $\mathcal{W}_n(\theta; \eta)$ in θ, we have

$$\lim_{n\to\infty} P\big(\phi_{Wn}(\theta_n; \eta) \neq \phi_{Wn}(\theta_n)\big)$$

$$\leq \max\Big\{ \lim_{n\to\infty} P\big(\phi_{Wn}(\theta_0; \eta) = 1\big), \lim_{n\to\infty} P\big(\phi_{Wn}(\theta_0) = 1\big) \Big\} = 0.$$

Similarly, in case (ii), we have

$$\lim_{n\to\infty} P\big(\phi_{Wn}(\theta_n; \eta) \neq \phi_{Wn}(\theta_n)\big)$$

$$\leq \max\Big\{ \lim_{n\to\infty} P\big(\phi_{Wn}(\theta_0; \eta) = 0\big), \lim_{n\to\infty} P\big(\phi_{Wn}(\theta_0) = 0\big) \Big\} = 0.$$

Finally, consider the third case in which $\theta_0 \in \partial\Theta_0$. Using the expansion in Lemma 4, we obtain that

$$\mathcal{W}_n(\theta_n; \eta) = \left[\frac{d(\eta(\theta_n), \widehat{H}_{0,\delta})}{w(\theta_n; \eta)}\right]^2 = \left[\frac{d(\eta(\theta_n), \eta(\widehat{\Theta}_{0,\delta}))}{w(\theta_n; \eta)}\right]^2$$

$$= \left[\frac{\mathbb{G}_n(\theta_n) + \sqrt{n}m(\theta_n)}{\|\nabla_\eta m_\eta(\theta_n; \eta)\| w(\theta_n; \eta)} + o_P(1)\right]_+^2. \tag{C.4}$$

Now suppose we split the sequence θ_n into three (possibly trivial) subsequences θ_{q_n}, θ_{r_n}, and θ_{s_n}, respectively, such that $\sqrt{n}m(\theta_{q_n}) = O(1)$, $\sqrt{n}m(\theta_{r_n}) \to \infty$, and $\sqrt{n}m(\theta_{s_n}) \to -\infty$. We can now analyze the behavior of $\phi_{Wn}(\cdot)$ separately along each of these subsequences. By Theorem 3,

$$\lim_{n\to\infty} P\big(\phi_{Wr_n}(\theta_{r_n}; \eta) = 0\big) = \lim_{n\to\infty} P\big(\phi_{Ws_n}(\theta_{s_n}; \eta) = 1\big) = 0$$

for all η, so that by the same arguments as before, the test is asymptotically invariant along these subsequences.

Finally, consider the sequence θ_{q_n}: if $\widehat{w}(\theta; \eta) = \frac{b(\theta)}{\|\nabla_\eta m(\eta(\theta))\|} + o_p(1)$, the expansion in (C.4) together with the continuous mapping theorem implies that along the subsequence θ_{q_n},

$$\mathcal{W}_{q_n}(\theta_{q_n}; \eta) = \left[\frac{\mathbb{G}_n(\theta_{q_n}) + \sqrt{q_n}m(\theta_{q_n})}{b(\theta_{q_n})} + o_p\big(1 + \sqrt{q_n}m(\theta_{q_n})\big)\right]_+^2$$

$$= \left[\frac{\mathbb{G}_n(\theta_{q_n}) + \sqrt{q_n}m(\theta_{q_n})}{b(\theta_{q_n})} + o_p(1)\right]_+^2,$$

so that the leading term of this expression is a function of θ that does not depend on η.

Now note that for $\theta \in \partial\Theta_0$, we have $m(\theta) = 0$, so that by this expansion and stochastic equicontinuity of $\mathbb{G}(\theta)$, we have

$$\mathcal{W}_n(\eta) := \sup_{\theta \in \partial\Theta_0} \mathcal{W}_n(\theta; \eta) \rightsquigarrow \mathcal{W} := \sup_{\theta \in \partial\Theta_0} \left[\frac{\mathbb{G}(\theta)}{b(\theta)}\right]_+^2,$$

where the limit does not depend on η. It follows that $\widehat{k}_{\mathcal{W}\eta}(1-\alpha) = \widehat{k}_{\mathcal{W}}(1-\alpha)$ for all η, so that $\phi_{\mathcal{W}n}(\theta; \eta)$ is asymptotically invariant along θ_{q_n} with respect to transformations $\eta \in H^*$, which establishes the second conclusion.

To establish the last claim, consider the Wald statistic in (2.4) with weighting function $\widehat{w}(\theta) = \frac{s(\theta)}{\|\nabla_\theta m(\theta)\|} + o_p(1)$. By Condition R.1 and the approximation in (C.2), $\mathcal{W}_n(\theta)$ converges in distribution to $\max\{0, Z\}^2$, where $Z \sim N(0, 1)$ for all values of $\theta \in \partial\Theta_0$ and is, therefore, asymptotically pivotal on the boundary of Θ_0. Hence for any fixed critical level k, the limit $\lim_{n\to\infty} P(\mathcal{W}_n(\theta) > k)$ is constant across all values of $\theta \in \partial\Theta_0$ for any value k.

REFERENCES

Andrews, D. and P. Guggenberger (2009), "Validity of subsampling and 'plug-in asymptotic' inference for parameters defined by moment inequalities." *Econometric Theory*, 25 (3), 669–709. [317]

Andrews, D. and G. Soares (2010), "Inference for parameters defined by moment inequalities using generalized moment selection." *Econometrica*, 78 (1), 119–157. [317]

Bansal, R. and A. Yaron (2004), "Risks for the long run: A potential resolution of asset pricing puzzles." *Journal of Finance*, 59 (4), 1481–1509. [329]

Beresteanu, A. and F. Molinari (2008), "Asymptotic properties for a class of partially identified models." *Econometrica*, 76 (4), 763–814. [310, 316, 319, 320]

Bohrer, R. (1973), "An optimality property of Scheffé bounds." *The Annals of Statistics*, 1 (4), 766–772. [326]

Britten-Jones, M. (1999), "The sampling error in estimates of mean–variance efficient portfolio weights." *Journal of Finance*, 54 (2), 655–671. [317]

Burnside, C. (1994), "Hansen–Jagannathan bounds as classical tests of asset-pricing models." *Journal of Business & Economic Statistics*, 12 (1), 57–79. [317]

Campbell, J. and J. Cochrane (1999), "By force of habit: A consumption-based explanation of aggregate stock market behavior." *Journal of Political Economy*, 107 (2), 205–251. [329]

Casella, G. and W. Strawderman (1980), "Confidence bands for linear regression with restricted predictor variables." *Journal of the American Statistical Association*, 75 (372), 862–868. [326]

Cecchetti, S., P.-S. Lam, and N. Mark (1994), "Testing volatility restrictions on intertemporal marginal rates of substitution implied by Euler equations and asset returns." *Journal of Finance*, 49 (1), 123–152. [316]

Chandrasekhar, A., V. Chernozhukov, F. Molinari, and P. Schrimpf (2012), "Inference for best linear approximations to set identified functions." Working paper, Stanford, MIT, Cornell, and UBC. [320]

Chernozhukov, V., H. Hong, and E. Tamer (2007), "Estimation and confidence regions for parameter sets in econometric models." *Econometrica*, 75, 1243–1284. [310, 316, 318, 341]

Chetty, R. (2012), "Bounds on elasticities with optimization frictions: A synthesis of micro and macro evidence on labor supply." *Econometrica*, 80 (3), 969–1018. [309, 310, 314, 315, 322, 335, 336, 337, 339]

Cochrane, J. and L. Hansen (1992), "Asset pricing explorations for macroeconomics." *NBER Macroeconomics Annual*, 7, 115–182. [316]

Cochrane, J. H. (2005), *Asset Pricing*. Princeton University Press, Princeton, NJ. [311]

Critchley, F., P. Marriott, and M. Salmon (1996), "On the differential geometry of the Wald test with nonlinear restrictions." *Econometrica*, 64 (5), 1213–1222. [316]

Davydov, Y., M. Lifshits, and N. Smorodina (1998), *Local Properties of Distributions of Stochastic Functionals*. American Mathematical Society, Providence, RI. [351]

Epstein, L. and S. Zin (1989), "Substitution, risk aversion and the temporal behavior of consumption and asset returns: A theoretical framework." *Econometrica*, 57 (4), 937–969. [329]

Fama, E. (1996), "Multifactor portfolio efficiency and multifactor asset pricing." *Journal of Financial and Quantitative Analysis*, 31 (4), 441–465. [313, 314]

Gibbons, M., S. Ross, and J. Shanken (1989), "A test on the efficiency of a given portfolio." *Econometrica*, 57 (5), 1121–1152. [317]

Haile, P. and E. Tamer (2003), "Inference with an incomplete model of English auctions." *Journal of Political Economy*, 111 (1), 1–51. [327]

Hansen, L. P., J. Heaton, and E. Luttmer (1995), "Econometric evaluation of asset pricing models." *Review of Financial Studies*, 8, 237–274. [316, 317]

Hansen, L. P. and R. Jagannathan (1991), "Implications of security market data for models of dynamic economies." *Journal of Political Economy*, 99, 225–262. [310, 312, 329]

Hansen, L. P. and K. Singleton (1982), "Generalized instrumental variables estimation of nonlinear rational expectations models." *Econometrica*, 50 (5), 1269–1286. [311, 329, 331]

Kaido, H. and A. Santos (2014), "Asymptotically efficient estimation of models defined by convex moment inequalities." *Econometrica*, 82 (1), 387–413. [316]

Lehmann, E. and J. Romano (2005), *Testing Statistical Hypotheses*. Springer, New York. [326]

Lettau, M. and S. Ludvigson (2009), "Euler equation errors." *Review of Economic Dynamics*, 12, 255–283. [315, 330]

Ludvigson, S. (2012), "Advances in consumption-based asset pricing: Empirical tests." In *Handbook of the Economics of Finance*, Vol. 2 (G. M. Constantinides, M. Harris, and R. M. Stulz, eds.), Elsevier, Amsterdam. [311]

Luttmer, E. (1996), "Asset pricing in economies with frictions." *Econometrica*, 64 (6), 1439–1467. [316, 317]

Markowitz, H. (1952), "Portfolio selection." *Journal of Finance*, 7 (1), 77–91. [312, 313]

Molchanov, I. S. (1998), "A limit theorem for solutions of inequalities." *Scandinavian Journal of Statistics*, 25, 235–242. [310, 316, 319]

Naiman, D. (1984), "Optimal simultaneous confidence bounds." *The Annals of Statistics*, 12 (2), 702–715. [326]

Peñaranda, F. and E. Sentana (2010), "Spanning tests in return and stochastic discount factor mean–variance frontiers: A unifying approach." Working paper, UPF and CEMFI. [317]

Politis, D. N. and J. P. Romano (1994), "Large sample confidence regions based on subsamples under minimal assumptions." *The Annals of Statistics*, 22 (4), 2031–2050. [342]

Romano, J. and A. Shaikh (2010), "Inference for the identified set in partially identified econometric models." *Econometrica*, 78, 169–211. [317]

Scheffé, H. (1953), "A method for judging all contrasts in the analysis of variance." *Biometrika*, 40 (1/2), 87–104. [326]

Sentana, E. (2009), "The econometrics of mean–variance efficiency tests: A survey." *Econometrics Journal*, 12, C65–C101. [317]

van der Vaart, A. W. (1998), *Asymptotic Statistics*. Cambridge University Press, Cambridge. [352]

van der Vaart, A. W. and J. A. Wellner (1996), *Weak Convergence and Empirical Processes*. Springer-Verlag, New York. [321]