cemmap
centre for microdata methods and practice

Inference for functions of partially identified parameters in moment inequality models

Federico Bugni
Ivan Canay
Xiaoxia Shi

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP22/14

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

An ESRC Research Centre

Inference for Functions of Partially Identified Parameters in Moment Inequality Models*

Federico A. Bugni
Department of Economics
Duke University
federico.bugni@duke.edu

Ivan A. Canay
Department of Economics
Northwestern University
iacanay@northwestern.edu

Xiaoxia Shi
Department of Economics
University of Wisconsin, Madison
xshi@ssc.wisc.edu

Ver. 10.2: May 4, 2014

Abstract

This paper introduces a bootstrap-based inference method for functions of the parameter vector in a moment (in)equality model. As a special case, our method yields marginal confidence sets for individual coordinates of this parameter vector. Our inference method controls asymptotic size uniformly over a large class of data distributions. The current literature describes only two other procedures that deliver uniform size control for this type of problem: projection-based and subsampling inference. Relative to projection-based procedures, our method presents three advantages: (i) it weakly dominates in terms of finite sample power, (ii) it strictly dominates in terms of asymptotic power, and (iii) it is typically less computationally demanding. Relative to subsampling, our method presents two advantages: (i) it strictly dominates in terms of asymptotic power (for reasonable choices of subsample size), and (ii) it appears to be less sensitive to the choice of its tuning parameter than subsampling is to the choice of subsample size.

Keywords: Partial Identification, Moment Inequalities, Subvector Inference, Hypothesis Testing.

JEL Classification: C01, C12, C15.

*We thank Francesca Molinari, Elie Tamer, the participants at the Workshop on Inference in Nonstandard Problems at Cornell University, the Workshop on Mathematical Statistics of Partially Identified at Oberwolfach, the conference on Partial Identification, Weak Identification, and Related Econometric Problems at Cowles Fundation, the First Conference in Econometric Theory at Universidad de San Andrés, Argentina, the workshop on Recent Developments in Partial Identification: Theory and Practice at Toulouse, the UCLA mini conference in Econometrics, and seminar participants at Brown, Northwestern, University of Maryland and Korea University for helpful comments. Bugni and Canay thank the National Science Foundation for research support via grants SES-1123771 and SES-1123586, respectively. Takuya Ura provided excellent research assistance. Any and all errors are our own.

1 Introduction

In recent years, substantial interest has been drawn to partially identified models defined by moment (in)equalities of the following generic form,

$$
\begin{aligned}
&E_F[m_j(W_i,\theta)] \geq 0 \text{ for } j = 1,\dots,p \ , \\
&E_F[m_j(W_i,\theta)] = 0 \text{ for } j = p+1,\dots,k \ ,
\end{aligned} \tag{1.1}
$$

where $\{W_i\}_{i=1}^n$ is an i.i.d. sequence of random variables with distribution F and $m = (m_1, \cdots, m_k)' : \mathbb{R}^d \times \Theta \to \mathbb{R}^k$ is a known measurable function of the finite dimensional parameter vector $\theta \in \Theta \subseteq \mathbb{R}^{d_\theta}$. Methods to conduct inference on θ have been proposed, for example, by Chernozhukov et al. (2007), Romano and Shaikh (2008), Andrews and Guggenberger (2009), and Andrews and Soares (2010).[1] As a common feature, these papers construct *joint* confidence sets (CS's) for the vector θ by inverting hypothesis tests for $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$. However, in empirical work, researchers often report *marginal* confidence intervals for each coordinate of θ, either to follow the tradition of standard t-test-based inference or because only few individual coordinates of θ are of interest. The current practice appears to be reporting projections of the joint CS's for the vector θ, e.g., Ciliberto and Tamer (2010) and Grieco (2013).

Although convenient, projecting joint CS's suffers from three problems. First, when interest lies in individual components of θ, projection methods are typically conservative (even asymptotically). This may lead to confidence intervals that are unnecessarily wide, a problem that gets exacerbated when the dimension of θ becomes reasonably large. Second, the projected confidence intervals do not necessarily inherit the good asymptotic power properties of the joint CS's. Yet, the available results in the literature are mostly limited to asymptotic properties of joint CS's. Finally, computing the projections of a joint CS is typically unnecessarily burdensome if the researcher is only interested in individual components. This is because one needs to compute the joint CS first, which itself requires searching over a potentially large dimensional space Θ for all the points not rejected by a hypothesis test.

In this paper, we address the practical need for marginal CS's by proposing a test to conduct inference directly on individual coordinates, or more generally, on a function $f : \Theta \to \mathbb{R}^{d_\gamma}$ of the parameter vector θ. The hypothesis testing problem is

$$
H_0 : f(\theta) = \gamma_0 \quad \text{vs.} \quad H_1 : f(\theta) \neq \gamma_0 \ , \tag{1.2}
$$

for a hypothetical value $\gamma_0 \in \mathbb{R}^{d_\gamma}$. We then construct a CS for $f(\theta)$ by exploiting the well-known duality between tests and CS's. Our test controls asymptotic size uniformly over a large class of data distributions (see Theorem 2.1) and has several attractive properties for practitioners: (i) it has finite sample power that weakly dominates that of projection-based tests for all alternative hypothesis (see Theorem 3.1), (ii) it has an asymptotic power that strictly dominates the one from projection-based tests under reasonable assumptions (see Remark 3.3), and (iii) it is less computationally demanding that projection-based tests whenever the function $f(\cdot)$ introduces dimension reduction, i.e., $d_\gamma << d_\theta$.[2] In addition, one corollary of our analysis is that our marginal CS's are always a subset of those constructed by projecting joint CS's (see Remark 3.1).

Leaving projection-based inference aside, there is another inference procedure that also addresses the hypothesis testing problem in (1.2): the subsampling method proposed by Romano and Shaikh (2008,

[1]Additional references include Imbens and Manski (2004), Beresteanu and Molinari (2008), Rosen (2008), Stoye (2009), Bugni (2010), Canay (2010), Romano and Shaikh (2010), Galichon and Henry (2011), Pakes et al. (2011), Bontemps et al. (2012), Bugni et al. (2012), and Romano et al. (2013) among others.

[2]This is commonly the case when d_θ is large and $d_\gamma = 1$, see Examples 2.1 and 2.2.

Section 3.4). In their paper, the authors propose comparing a profiled criterion function similar to ours with a subsampling critical value. However, subsampling inference presents well-known challenges that make some applied researchers reluctant to use it when other alternatives are available. To be more specific, subsampling inference is known to be very sensitive to the choice of subsample size and, even when the subsample size is chosen to minimize error in the coverage probability, it is more imprecise than its bootstrap alternatives, see Politis and Romano (1994); Bugni (2010, 2014). In this paper, we provide additional results that support our bootstrap-based inference over the subsampling inference in Romano and Shaikh (2008). First, we show that our test is no less asymptotically powerful than the subsampling test under reasonable assumptions (see Theorem 3.2). Second, we formalize the conditions under which our test has strictly higher asymptotic power (see Remark 3.5). Finally, we note that our test appears to be less sensitive to the choice of its tuning parameter κ_n than subsampling is to the choice of subsample size (see Remark 3.6).

As previously mentioned, the asymptotic results in this paper hold uniformly over a large class of nuisance parameters. In particular, the test we propose controls asymptotic size over a large class of distributions F and can be inverted to construct uniformly valid CS's (see Remark 2.4). This represents an important difference with other methods that could also be used for inference on components of θ, such as Pakes et al. (2011), Chen et al. (2011), Kline and Tamer (2013), and Wan (2013). The test proposed by Pakes et al. (2011) is, by construction, a test for each coordinate of the parameter θ. However, such test controls size over a much smaller class of distributions than the one we consider in this paper (c.f. Andrews and Han, 2009). The approach recently introduced by Chen et al. (2011) is especially useful for parametric models with unknown functions, which do not correspond exactly with the model in (1.1). In addition, the asymptotic results in that paper hold pointwise and so it is unclear whether it controls asymptotic size over the same class of distributions we consider. The method in Kline and Tamer (2013) is Bayesian in nature, requires either the function $m(W_i, \theta)$ to be separable (in W_i and θ) or the data to be discretely-supported, and focuses on inference about the identified set as opposed to identifiable parameters. Finally, Wan (2013) introduces a computationally attractive inference method based on MCMC, but derives pointwise asymptotic results. Due to these reasons, we do not devote special attention to these papers.

We view our test as an attractive alternative to applied researchers and so we have included a step by step algorithm to implement our test in Appendix A.1. We use this algorithm in the Monte Carlo simulations of Section 4. Our numerical results support all the theoretical findings about asymptotic size control (Section 2) and asymptotic power advantages (Section 3).

2 New test: the minimum resampling test

2.1 Motivating examples

The hypothesis test in (1.2) involves a function $f : \Theta \to \Gamma \subseteq \mathbb{R}^{d_\gamma}$ of the partially identified parameter θ in the moment (in)equality model in (1.1). As the next two examples illustrate, applied researchers often are only interested in a few individual coordinates of the vector θ and resort to projection-based inference. The leading application of our inference method is therefore the construction of marginal CS's for coordinates of θ, which is done by setting $f(\theta) = \theta_s$ for a $s \in \{1, \ldots, d_\theta\}$ in (1.2) and collecting all values of $\gamma_0 \in \Gamma$ for which H_0 is not rejected.

Example 2.1. Ciliberto and Tamer (2010) investigate the empirical importance of firm heterogeneity as a determinant of market structure in the U.S. airline industry. They show that the competitive effects of large airlines (American, Delta, United) are different from those of low cost carriers and Southwest. The parameter θ entering the profit functions in Ciliberto and Tamer (2010) is close to 30 dimensional in some of the specifications they use. However, interest is centered in the competitive effect of American Airlines, in whether two airlines have the same coefficients, or in some other restriction that involves a small number of components of θ. The authors report a table with the smallest cube that contains a 95% confidence region for $\Theta_I(F)$. This is what we call "the standard practice" in this paper. □

Example 2.2. Grieco (2013) introduces an entry model that includes both publicly observed and privately known structural errors for each firm and studies the impact of supercenters - large stores such as Wal-Mart - on the profitability of rural grocery stores. The parameter θ in his application is multi-dimensional with 11 components. However, interest centers on the coefficient that measures the presence of a supercenter on the value of a grocery store. In his application, Grieco (2013) also reports projections of the confidence set for θ onto parameter axes and clarifies that such table "exaggerates the size of the confidence sets of the full model" (see Grieco, 2013, footnote 54). □

While marginal CS's for the vector θ is our leading application, our inference framework can accommodate other hypotheses testing problems that can be of empirical interest. For example, one could test whether two coordinates of θ are equal (or not) by setting $f(\theta) = \theta_s - \theta_{s'}$ for $s \neq s'$ in (1.2) and $\gamma_0 = 0$. In addition, one could test hypotheses about counterfactuals, provided the counterfactual can be represented as a known function of θ, which is typically the case.

2.2 Framework and test statistic

In order to describe our new test for the hypotheses in (1.2), we need to introduce some basic notation. We assume throughout the paper that F, the distribution of the observed data, belongs to a *baseline distribution space* that we define below. We then introduce an appropriate baseline and null parameter space for (γ, F), which is the tuple composed of the parameter γ, i.e., the image of the function f, and the distribution of the data. In order to keep the exposition as reader friendly as possible, we summarize the most important notation in Table 2, Appendix A.

Definition 2.1 (Baseline Distribution Space)**.** The baseline space of probability distributions, denoted by $\mathcal{P} \equiv \mathcal{P}(a, M, \Psi)$, is the set of distributions F such that, when paired with some $\theta \in \Theta$, the following conditions hold:

(i) $\{W_i\}_{i=1}^n$ are i.i.d. under F,

(ii) $\sigma^2_{F,j}(\theta) = Var_F(m_j(W_i, \theta)) \in (0, \infty)$, for $j = 1, \dots, k$,

(iii) $Corr_F(m(W_i, \theta)) \in \Psi$,

(iv) $E_F[|m_j(W_i, \theta)/\sigma_{F,j}(\theta)|^{2+a}] \leq M$,

where Ψ is a specified closed set of $k \times k$ correlation matrices, and M and a are fixed positive constants.

Definition 2.2 (Identified Set)**.** For any $F \in \mathcal{P}$, the identified set $\Theta_I(F)$ is the set of parameters $\theta \in \Theta$ that satisfy the moments restrictions in (1.1).

Definition 2.3 (Null Set and Null Identified Set)**.** For any $F \in \mathcal{P}$ and $\gamma \in \Gamma$, the null set $\Theta(\gamma) \equiv \{\theta \in \Theta : f(\theta) = \gamma\}$ is the set of parameters satisfying the null hypothesis, and the null identified set $\Theta_I(F, \gamma) \equiv \{\theta \in \Theta_I(F) : f(\theta) = \gamma\}$ is the set of parameters in the identified set satisfying the null hypothesis.

Definition 2.4 (Parameter Space for (γ, F))**.** The parameter space for (γ, F) is given by $\mathcal{L} \equiv \{(\gamma, F) : F \in \mathcal{P}, \gamma \in \Gamma\}$. The null parameter space is $\mathcal{L}_0 \equiv \{(\gamma, F) : F \in \mathcal{P}, \gamma \in \Gamma, \Theta_I(F, \gamma) \neq \emptyset\}$.

Our test is based on a non-negative function $Q_F : \Theta \to \mathbb{R}_+$, referred to as *population criterion function*, with the property that $Q_F(\theta) = 0$ if and only if $\theta \in \Theta_I(F)$. In the context of the moment (in)equality model in (1.1), it is convenient to consider criterion functions that are specified as follows (see, e.g., Andrews and Guggenberger, 2009; Andrews and Soares, 2010; Bugni et al., 2012),

$$Q_F(\theta) = S(E_F[m(W, \theta)], \Sigma_F(\theta)) \ , \tag{2.1}$$

where $\Sigma_F(\theta) \equiv Var_F(m(W, \theta))$ and $S : \mathbb{R}^p_{[+\infty]} \times \mathbb{R}^{k-p} \times \Psi \to \mathbb{R}_+$ is the test function specified by the econometrician that needs to satisfy several regularity assumptions.[3] The (properly scaled) sample analogue criterion function is

$$Q_n(\theta) = S(\sqrt{n}\bar{m}_n(\theta), \hat{\Sigma}_n(\theta)) \ , \tag{2.2}$$

where $\bar{m}_n(\theta) \equiv (\bar{m}_{n,1}(\theta), \ldots, \bar{m}_{n,k}(\theta))$, $\bar{m}_{n,j}(\theta) \equiv n^{-1}\sum_{i=1}^{n} m_j(W_i, \theta)$ for $j = 1, \ldots, k$, and $\hat{\Sigma}_n(\theta)$ is a consistent estimator of $\Sigma_F(\theta)$. Sometimes it is more convenient to work with correlation matrices, in which case we use $\Omega_F(\theta) \equiv D_F^{-1/2}(\theta)\Sigma_F(\theta)D_F^{-1/2}(\theta)$, $\hat{\Omega}_n(\theta) \equiv \hat{D}_n^{-1/2}(\theta)\hat{\Sigma}_n(\theta)\hat{D}_n^{-1/2}(\theta)$, $D_F(\theta) = Diag(\Sigma_F(\theta))$, and $\hat{D}_n(\theta) = Diag(\hat{\Sigma}_n(\theta))$. Finally, for a given $\gamma_0 \in \Gamma$, the test statistic we use for testing (1.2) is the *profiled* version of $Q_n(\theta)$,

$$T_n(\gamma_0) \equiv \inf_{\theta \in \Theta(\gamma_0)} Q_n(\theta) \ . \tag{2.3}$$

Theorem C.4 in the Appendix adapts results from Bugni et al. (2013) to show that, along relevant subsequences of parameters $(\gamma_n, F_n) \in \mathcal{L}_0$,

$$\inf_{\theta \in \Theta(\gamma_n)} Q_n(\theta) \xrightarrow{d} J(\Lambda, \Omega) \equiv \inf_{(\theta,\ell) \in \Lambda} S(v_\Omega(\theta) + \ell, \Omega(\theta, \theta)) \ , \tag{2.4}$$

where $v_\Omega : \Theta \to \mathbb{R}^k$ is a $\mathbb{R}^k$-valued tight Gaussian process with covariance (correlation) kernel $\Omega \in \mathcal{C}(\Theta^2)$, and Λ is the limit (in the Hausdorff metric) of the set

$$\Lambda_{n,F_n}(\gamma_n) \equiv \left\{(\theta, \ell) \in \Theta(\gamma_n) \times \mathbb{R}^k : \ell = \sqrt{n}D_{F_n}^{-1/2}(\theta)E_{F_n}[m(W, \theta)]\right\} \ . \tag{2.5}$$

The limit distribution $J(\Lambda, \Omega)$ in (2.4) depends on the set Λ and the function Ω, and so does its $1-\alpha$ quantile, which we denote by $c_{(1-\alpha)}(\Lambda, \Omega)$. The non-random set $\Lambda_{n,F_n}(\gamma_n)$ and its limit Λ are introduced for technical convenience and do not have an intuitive interpretation. However, the form of $J(\Lambda, \Omega)$ is quite natural in this context as it resembles a "profiled" version of the usual limit distribution $S(v_\Omega(\theta) + \ell, \Omega(\theta))$.

Remark 2.1. Theorem C.4 gives the asymptotic distribution of our test statistic under a (sub)sequence of parameters (γ_n, F_n) that satisfies certain properties. It turns out that these types of (sub)sequences are the relevant ones to determine the asymptotic coverage of confidence sets that are derived by test inversion. In other words, controlling the asymptotic coverage of a confidence set for γ involves a infimum over (γ, F) - see (2.15) - and thus we present the asymptotic derivations along sequences of parameters (γ_n, F_n) rather than (γ_0, F_n) to accommodate this case. If the goal were to simply control the asymptotic size of the test for $H_0 : f(\theta) = \gamma_0$, then deriving results for sequences (γ_0, F_n) would have been sufficient.

Having an expression for $J(\Lambda, \Omega)$, our goal is to construct feasible critical values that approximate $c_{(1-\alpha)}(\Lambda, \Omega)$ asymptotically. This requires approximating the limiting set Λ and the limiting correlation

[3] See Assumptions M.1-M.9 in the Appendix for these regularity conditions and (3.14) for an example.

function Ω. The limiting correlation function can be estimated using standard methods. On the other hand, the approximation of Λ is non-standard and presents certain difficulties that we describe in the next section.

2.3 Test MR: minimum resampling

The main challenge in approximating the quantiles of $J(\Lambda, \Omega)$ lies in the approximation of the set Λ. Part of the difficulty relates to the approximation of ℓ, although this can be addressed using the GMS approach in Andrews and Soares (2010) that consists in replacing ℓ with $\varphi = (\varphi_1, \ldots, \varphi_k)$, where

$$\varphi_j = \varphi_j(\kappa_n^{-1}\sqrt{n}\hat{\sigma}_{n,j}^{-1}(\theta)\bar{m}_{n,j}(\theta)) \text{ for } j = 1, \ldots, p \text{ and } \varphi_j = 0 \text{ for } j = p+1, \ldots, k \ , \tag{2.6}$$

is the GMS function satisfying the assumptions in Andrews and Soares (2010). The thresholding sequence $\{\kappa_n\}_{n\geq 1}$ satisfies $\kappa_n \to \infty$ and $\kappa_n/\sqrt{n} \to 0$.[4] However, the real challenge in our context is due to the fact that the relevant points within the set Λ are the cluster points of the sequence

$$\{(\theta_n, \sqrt{n}D_{F_n}^{-1/2}(\theta_n)E_{F_n}[m(W, \theta_n)])\}_{n\geq 1} \ , \tag{2.7}$$

where θ_n is the infimum of $Q_n(\theta)$ over $\Theta(\gamma_n)$ and, hence, *random*. This has two immediate technical consequences. First, we cannot borrow results from Andrews and Soares (2010) as those hold for non-random sequences of parameters $\{(\theta_n, F_n)\}_{n\geq 1}$ with $\theta_n \in \Theta_I(F_n)$ for all $n \in \mathbb{N}$ and cannot be extended to random sequences. Second, the random sequence $\{\theta_n\}_{n\geq 1}$ in (2.7) could be such that $\theta_n \notin \Theta_I(F_n)$ for all $n \in \mathbb{N}$ (especially in models where $\Theta_I(F_n)$ has empty interior). To see why, note that in our setup the null hypothesis implies that there is $(\gamma_n, F_n) \in \mathcal{L}_0$ for all $n \in \mathbb{N}$, meaning that there exists $\theta_n^\star \in \Theta_I(F_n)$ such that $\gamma_n = f(\theta_n^\star)$ for all $n \in \mathbb{N}$ (see Definition 2.4). There is, however, no guarantee that the random minimizing sequence $\{\theta_n\}_{n\geq 1}$ in (2.7) satisfies $\theta_n \in \Theta_I(F_n)$. This is problematic because it implies that the set Λ contains tuples (θ, ℓ) such that $\ell_j < 0$ for $j = 1, \ldots, p$, or $\ell_j \neq 0$ for $j = p+1, \ldots, k$ and so, if an infimum is attained, it could be attained at a value of θ that is *not* associated with $\ell_j \geq 0$ for $j = 1, \ldots, p$ and $\ell_j = 0$ for $j = p+1, \ldots, k$. Thus, along sequences that converge to such tuples, the GMS function $\varphi(\cdot)$ is *not* a conservative estimator of $\sqrt{n}D_{F_n}^{-1/2}(\theta_n)E_{F_n}[m(W, \theta_n)]$.

In this paper we circumvent the aforementioned difficulties by combining two approximations to $J(\Lambda, \Omega)$ that share common elements. They both use the same estimate of Ω,

$$\hat{\Omega}_n(\theta) \equiv \hat{D}_n^{-1/2}(\theta)\hat{\Sigma}_n(\theta)\hat{D}_n^{-1/2}(\theta) \quad \text{where} \quad \hat{\Sigma}_n(\theta) \equiv n^{-1}\sum_{i=1}^{n}(m(W_i, \theta) - \bar{m}_n(\theta))(m(W_i, \theta) - \bar{m}_n(\theta))' \ .$$

They also use the same asymptotic approximation to the stochastic process $v_\Omega(\theta)$,

$$v_n^*(\theta) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\hat{D}_n^{-1/2}(\theta)(m(W_i, \theta) - \bar{m}_n(\theta))\zeta_i \quad \text{and} \quad \{\zeta_i \sim N(0,1)\}_{i=1}^{n} \text{ is i.i.d.}^{5} \tag{2.8}$$

[4] The GMS function $\varphi(\cdot)$ in Andrews and Soares (2010) might also depend on $\hat{\Sigma}_n(\theta)$. For simplicity we consider those that only depend on $\kappa_n^{-1}\sqrt{n}\hat{\sigma}_{n,j}^{-1}(\theta)\bar{m}_{n,j}(\theta)$, which represents all but one of the φ-functions in Andrews and Soares (2010).

[5] We note that one could alternatively use a bootstrap approximation, $n^{-1/2}\sum_{i=1}^{n}\hat{D}_n^{-1/2}(\theta)(m(W_i^*, \theta) - \bar{m}_n(\theta))$, where $\{W_i^*\}_{i=1}^{n}$ is an i.i.d. sample drawn with replacement from original sample $\{W_i\}_{i=1}^{n}$. In our simulations, the asymptotic approximation is computationally faster.

The first resampling test statistic we use to approximate $J(\Lambda,\Omega)$ is

$$T_n^{R1}(\gamma_0) \equiv \inf_{\theta\in\hat{\Theta}_I(\gamma_0)} S(v_n^*(\theta) + \varphi(\kappa_n^{-1}\sqrt{n}\hat{D}_n^{-1}(\theta)\bar{m}_n(\theta)), \hat{\Omega}_n(\theta)) \ , \tag{2.9}$$

where

$$\hat{\Theta}_I(\gamma_0) \equiv \{\theta \in \Theta(\gamma_0) : Q_n(\theta) \leq T_n(\gamma_0)\} \tag{2.10}$$

is the set of minimizers of $Q_n(\theta)$, $\varphi = (\varphi_1, \ldots, \varphi_k)$ is as in (2.6), and $T_n(\gamma_0)$ is as in (2.3). Using $T_n^{R1}(\gamma_0)$ to simulate the quantiles of $J(\Lambda,\Omega)$ is based on an approximation to the set Λ that replaces ℓ with $\varphi(\cdot)$ and enforces θ to be *close* to $\Theta_I(F)$ by using the approximation $\hat{\Theta}_I(\gamma_0)$ to $\Theta_I(F,\gamma_0)$. Even though $\hat{\Theta}_I(\gamma_0)$ is generally not a consistent estimator of $\Theta_I(F,\gamma_0)$, it follows from Bugni et al. (2013, Lemma D.13) that

$$\lim_{n\to\infty} \inf_{F\in\mathcal{P}_0} P_F(\hat{\Theta}_I(\gamma_0) \subseteq \Theta_I^{\ln\kappa_n}(\gamma_0)) = 1 \ , \tag{2.11}$$

where $\Theta_I^{\ln\kappa_n}(\gamma_0)$ is the non-random expansion of $\Theta_I(F,\gamma_0)$ defined in Table 2. The result in (2.11) is precisely what we need for our approach to work. Since $\varphi_j(\cdot) \geq 0$ for $j = 1, \ldots, p$ and $\varphi_j(\cdot) = 0$ for $j = p+1, \ldots, k$, using this approximation in the definition of $T_n^{R1}(\gamma_0)$ guarantees that the (in)equality restrictions are not violated by much when evaluated at the θ that approximates the infimum in (2.9). This makes the GMS function $\varphi(\cdot)$ a valid replacement for ℓ and plays an important role in establishing our results.

Remark 2.2. As in other M-estimation problems, it is not necessary to impose that $\hat{\Theta}_I(\gamma_0)$ is the set of exact minimizers of $Q_n(\theta)$. This set could be replaced with an "approximate" set of minimizers, i.e., $\hat{\Theta}_I(\gamma_0) \equiv \{\theta \in \Theta(\gamma_0) : Q_n(\theta) \leq T_n(\gamma_0) + o_p(1)\}$, without affecting any of our results. This is relevant for situations in which the optimization algorithm is only guaranteed to approximate the exact minimizers.

The second resampling test statistic we use to approximate $J(\Lambda,\Omega)$ is

$$T_n^{R2}(\gamma_0) \equiv \inf_{\theta\in\Theta(\gamma_0)} S(v_n^*(\theta) + \kappa_n^{-1}\sqrt{n}\hat{D}_n^{-1}(\theta)\bar{m}_n(\theta), \hat{\Omega}_n(\theta)) \ . \tag{2.12}$$

Using $T_n^{R2}(\gamma_0)$ to simulate the quantiles of $J(\Lambda,\Omega)$ is based on an approximation to the set Λ that replaces ℓ with $\kappa_n^{-1}\sqrt{n}\hat{D}_n^{-1}(\theta)\bar{m}_n(\theta)$. This is not equivalent to the GMS approach: (a) it could be the case that $\kappa_n^{-1}\sqrt{n}\hat{\sigma}_{n,j}^{-1}(\theta)\bar{m}_{n,j}(\theta) < 0$ for some $j = 1, \ldots, p$, and (b) it also includes the term $\kappa_n^{-1}\sqrt{n}\hat{\sigma}_{n,j}^{-1}(\theta)\bar{m}_{n,j}(\theta)$ for $j = p+1, \ldots, k$ (i.e. equality restrictions).[6] As explained before, the set Λ contains tuples (θ,ℓ) such that $\ell_j < 0$ for $j = 1, \ldots, p$, or $\ell_j \neq 0$ for $j = p+1, \ldots, k$. This second approximation directly contemplates this possibility and therefore avoids the need of an estimator of $\Theta_I(F,\gamma_0)$ satisfying (2.11). This also plays an important role in establishing the consistency in level of our test.

Remark 2.3. Replacing $\hat{\Theta}_I(\gamma_0)$ with $\Theta(\gamma_0)$ while keeping the function $\varphi(\cdot)$ in (2.9) would not result in a valid approximation to $J(\Lambda,\Omega)$ and, subsequently, would not result in a valid test for the null hypothesis of interest. Therefore, it is important for $T_n^{R1}(\gamma_0)$ to use $\hat{\Theta}_I(\gamma_0)$ and for $T_n^{R2}(\gamma_0)$ to use $\kappa_n^{-1}\sqrt{n}\hat{D}_n^{-1}(\theta)\bar{m}_n(\theta)$ rather than $\varphi(\cdot)$.

We now have all the elements to define the new critical value and the minimum resampling test we propose in this paper.

[6] The GMS approach requires both that $\varphi_j(\cdot) \geq 0$ for $j = 1, \ldots, p$ and $\varphi_j(\cdot) = 0$ for $j = p+1, \ldots, k$ in order for the approach to have good power properties, see Andrews and Soares (2010, Assumption GMS6 and Theorem 3).

Definition 2.5 (Minimum Resampling Critical Value)**.** Let $T_n^{R1}(\gamma_0)$ and $T_n^{R2}(\gamma_0)$ be defined as in (2.9) and (2.12) respectively, where $v_n^*(\theta)$ is defined as in (2.8) and is common to both test statistics. The Minimum Resampling critical value $\hat{c}_n^{MR}(\gamma_0, 1-\alpha)$ is defined as the $1-\alpha$ quantile of

$$T_n^{MR}(\gamma_0) \equiv \min\left\{T_n^{R1}(\gamma_0), T_n^{R2}(\gamma_0)\right\} \ . \tag{2.13}$$

Definition 2.6 (Minimum Resampling Test)**.** Let $\Theta(\gamma_0)$ be defined as in Definition 2.3 and $\hat{c}_n^{MR}(\gamma_0, 1-\alpha)$ be defined as in Definition (2.5). The Minimum Resampling test (or Test MR) is

$$\phi_n^{MR}(\gamma_0) \equiv 1\left\{\inf_{\theta\in\Theta(\gamma_0)} Q_n(\theta) > \hat{c}_n^{MR}(\gamma_0, 1-\alpha)\right\} \ . \tag{2.14}$$

The profiled test statistic $\inf_{\theta\in\Theta(\gamma_0)} Q_n(\theta)$ is standard in point identified models and has been proposed in the context of partially identified models for a subsampling test by Romano and Shaikh (2008). The novelty in our Test MR lies in the critical value $\hat{c}_n^{MR}(\gamma_0, 1-\alpha)$. This is because each of the two basic resampling approximations we combine - embedded in $T_n^{R1}(\gamma_0)$ and $T_n^{R2}(\gamma_0)$ - has good power properties in particular directions and neither of them dominate each other in terms of asymptotic power - see Example 3.1. By combining these two approximations into the test statistic $T_n^{MR}(\gamma_0)$, the Minimum Resampling Test $\phi_n^{MR}(\gamma_0)$ dominates each of these basic approximations and has two important additional properties. The first property is summarized in the next theorem.

Theorem 2.1. *Let Assumptions A.1-A.7 hold. Then, for* $\alpha \in (0, 1/2)$*,*

$$\limsup_{n\to\infty} \sup_{(\gamma,F)\in\mathcal{L}_0} E_F[\phi_n^{MR}(\gamma)] \leq \alpha \ .$$

Remark 2.4. Let $CS_n^\star(1-\alpha) \equiv \{\gamma \in \Gamma : T_n^{MR}(\gamma) \leq \hat{c}_n^{MR}(\gamma, 1-\alpha)\}$ be a $1-\alpha$ CS for γ. It follows from Theorem 2.1 that

$$\liminf_{n\to\infty} \inf_{(\gamma,F)\in\mathcal{L}_0} P_F\left(\gamma \in CS_n^\star(1-\alpha)\right) \geq 1-\alpha \ , \tag{2.15}$$

meaning that the Minimum Resampling test can be inverted to construct valid CS's for γ. In particular, this allows us to construct confidence intervals for individual components of θ when $f(\theta) = \theta_s$ for some $s = 1, \dots, d_\theta$.

Remark 2.5. All the assumptions we use throughout the paper can be found in Appendix B. The first four assumptions in Theorem 2.1 are regularity conditions that allow us to use uniform Donsker theorems, see Remark B.1. Assumptions A.5 and A.6 are rather technical conditions that are discussed in Remarks B.2 and B.3. Finally, Assumption A.7 is a key sufficient condition for the asymptotic validity of our test that requires the criterion function to satisfy a minorant-type condition as in Chernozhukov et al. (2007) and the normalized population moments to be sufficiently smooth. See Remark B.4 for a detailed discussion.

The second property concerns the asymptotic power properties of our test relative to a subsampling test applied to $\inf_{\theta\in\Theta(\gamma_0)} Q_n(\theta)$ (as in Romano and Shaikh, 2008) or a test based on checking whether the image under $f(\cdot)$ of a Generalized Moment Selection (GMS) confidence set for θ (as in Andrews and Soares, 2010) intersects $\Theta(\gamma_0)$. The power analysis is involved, so we devote the entire Section 3 to this task.

3 Minimum resampling versus existing alternatives

The critical value of the Minimum Resampling test from Definition 2.6 is the $1-\alpha$ quantile of $T_n^{MR}(\gamma_0) \equiv \min\left\{T_n^{R1}(\gamma_0), T_n^{R2}(\gamma_0)\right\}$, where $T_n^{R1}(\gamma_0)$ and $T_n^{R2}(\gamma_0)$ are defined in (2.9) and (2.12), respectively. If we let $\hat{c}_n^{R1}(\gamma_0, 1-\alpha)$ and $\hat{c}_n^{R2}(\gamma_0, 1-\alpha)$ be the $1-\alpha$ quantiles of $T_n^{R1}(\gamma_0)$ and $T_n^{R2}(\gamma_0)$, respectively, then two "basic" resampling tests, denoted by Test R1 and Test R2, could be defined as follows,

$$\phi_n^{R1}(\gamma_0) \equiv 1\left\{\inf_{\theta\in\Theta(\gamma_0)} Q_n(\theta) > \hat{c}_n^{R1}(\gamma_0, 1-\alpha)\right\} \ , \tag{3.1}$$

$$\phi_n^{R2}(\gamma_0) \equiv 1\left\{\inf_{\theta\in\Theta(\gamma_0)} Q_n(\theta) > \hat{c}_n^{R2}(\gamma_0, 1-\alpha)\right\} \ . \tag{3.2}$$

By construction $\hat{c}_n^{MR}(\gamma, 1-\alpha) \leq \min\{\hat{c}_n^{R1}(\gamma, 1-\alpha), \hat{c}_n^{R2}(\gamma, 1-\alpha)\}$, and thus it follows immediately that for any $(\gamma, F) \in \mathcal{L}$ and all $n \in \mathbb{N}$,

$$\phi_n^{MR}(\gamma) \geq \phi_n^{R1}(\gamma) \quad \text{and} \quad \phi_n^{MR}(\gamma) \geq \phi_n^{R2}(\gamma) \ . \tag{3.3}$$

In this section we study the properties of each of these basic resampling tests. This is interesting for the following reasons. First, we show that Test R1 dominates (in terms of finite sample power) the standard practice of computing the image under $f(\cdot)$ of a CS for θ and checking whether it includes γ_0. Second, we show that Test R2 dominates (in terms of asymptotic power) a subsampling test applied to $\inf_{\theta\in\Theta(\gamma_0)} Q_n(\theta)$ under certain conditions. By the inequalities in (3.3) these results imply that Test MR weakly dominates both of these tests. We formalize these statements in the next subsections, and also present two examples (Examples 3.1 and 3.2) that illustrate cases in which Test MR has strictly better asymptotic power and size control than the two existing tests.

3.1 Power advantages over Test BP

Examples 2.1 and 2.2 illustrate that the standard practice in applied work to test the hypotheses in (1.2) involves computing a CS for the parameter θ first, and then rejecting the null hypothesis whenever the image of this CS under $f(\cdot)$ does not equal γ_0. We refer to this test as Test BP, to emphasize the fact that this test comes as a By-Product of constructing a CS for the entire parameter θ, and was not specifically designed to test the hypotheses in (1.2). Using the notation introduced in the previous section, we define a generic $1-\alpha$ CS for θ as

$$CS_n(1-\alpha) = \{\theta \in \Theta : Q_n(\theta) \leq \hat{c}_n(\theta, 1-\alpha)\} \ , \tag{3.4}$$

where $\hat{c}_n(\theta, 1-\alpha)$ is such that $CS_n(1-\alpha)$ has the correct asymptotic coverage. CS's that have the structure in (3.4) and control asymptotic coverage have been proposed by Romano and Shaikh (2008); Andrews and Guggenberger (2009); Andrews and Soares (2010); Canay (2010); and Bugni (2009), among others.

Definition 3.1 (Test BP)**.** Let $CS_n(1-\alpha)$ be a CS for θ that controls asymptotic size. Test BP rejects the null hypothesis in (1.2) according to the following rejection rule

$$\phi_n^{BP}(\gamma_0) \equiv 1\left\{CS_n(1-\alpha) \cap \Theta(\gamma_0) = \emptyset\right\} \ . \tag{3.5}$$

Definition 3.1 shows that Test BP depends on the confidence set $CS_n(1-\alpha)$. It follows that Test BP inherits its size and power properties from the properties of $CS_n(1-\alpha)$, and these properties in turn depend on the particular choice of test statistic and critical value used in the construction of $CS_n(1-\alpha)$. All the tests we consider in this paper are functions of the sample criterion function defined in (2.2) and therefore their relative power properties do not depend on the choice of the particular function $S(\cdot)$. However, the relative performance of Test BP with respect to our test does depend on the choice of critical value used in $CS_n(1-\alpha)$. Bugni (2010) shows that GMS tests have more accurate asymptotic size than subsampling tests. Andrews and Soares (2010) show that GMS tests are more powerful than Plug-in asymptotics or subsampling tests. This means that, asymptotically, Test BP implemented with a GMS CS will be less conservative and more powerful than the analogous test implemented with plug-in asymptotics or subsampling. We therefore adopt the GMS version of the specification test in Definition 3.1 as the "benchmark version" of Test BP. This is summarized in the maintained Assumption M.4, see Appendix B.

By appropriately modifying the arguments in Bugni et al. (2013), we show that

$$\phi_n^{R1}(\gamma) = 1\{\inf_{\theta\in\Theta(\gamma)} Q_n(\theta) > \hat{c}_n^{R1}(\gamma, 1-\alpha)\} \geq 1\{\nexists\theta \in CS_n(1-\alpha) : f(\theta) = \gamma\} = \phi_n^{BP}(\gamma) \ , \qquad (3.6)$$

whenever Tests BP and Test R1 are implemented with the same sequences $\{\kappa_n\}_{n\geq 1}$ and the same function $\varphi(\cdot)$. By (3.3), this means that Test MR weakly dominates Test BP in terms of finite sample power. We summarize this in the next theorem.

Theorem 3.1. *For any $(\gamma, F) \in \mathcal{L}$ it follows that $\phi_n^{R1}(\gamma) \geq \phi_n^{BP}(\gamma)$ for all $n \in \mathbb{N}$.*

Corollary 3.1. *For any sequence $\{(\gamma_n, F_n) \in \mathcal{L}\}_{n\geq 1}$, $\liminf_{n\to\infty}(E_{F_n}[\phi_n^{R1}(\gamma_n)] - E_{F_n}[\phi_n^{BP}(\gamma_n)]) \geq 0$.*

Remark 3.1. Theorem 3.1 is a statement that holds for all $n \in \mathbb{N}$ and $(\gamma, F) \in \mathcal{L}$. In particular, since it holds for parameters $(\gamma, F) \in \mathcal{L}_0$, this is a result about finite sample power and size. This theorem and (3.3) imply that Test MR cannot be more conservative nor have lower power than Test BP for all $n \in \mathbb{N}$ and $(\gamma, F) \in \mathcal{L}$. In fact, Theorem 3.1 and (3.3) imply an even stronger conclusion: The CS for γ defined in Remark 2.4 is always a subset of the one produced by projecting the joint CS in (3.5).

Remark 3.2. When the dimension of Θ is big relative to that of Γ - e.g., the function $f(\cdot)$ selects one of several elements of Θ - the implementation of Test MR is computationally attractive as it involves inverting the test over a smaller dimension. In other words, in cases where $dim(\Gamma)$ is much smaller than $dim(\Theta)$, Test MR has power *and* computational advantages over Test BP.

Remark 3.3. Under a condition similar to Bugni et al. (2013, Assumption A.9), we can show that Test R1 has asymptotic power that is *strictly* higher than that of Test BP for certain local alternative hypotheses. The proof is similar to that in Bugni et al. (2013, Theorem 6.2) and so we do not include it here. We do illustrate a situation in which our test has strictly better asymptotic power in Example 3.1.

Test R1 corresponds to the Resampling Test introduced by Bugni et al. (2013) to test the correct specification of the model in (1.1). Using this test for the hypotheses we consider in this paper would result in a test with correct asymptotic size by the inequality in (3.3). Unfortunately, Test R1 presents two disadvantages relative to Test MR. First, there is no guarantee that Test R1 has better asymptotic power than the subsampling test proposed by Romano and Shaikh (2008). Second, there are cases in which Test MR has strictly higher asymptotic power than Test R1 for the hypotheses in (1.2) - see Example 3.1 for an illustration.

3.2 Power advantages over Test SS

We have compared Test MR with Test BP using the connection between Test MR and the first basic resampling test, Test R1. In this section we show that Test MR dominates a subsampling test by using its connection to the second basic resampling test, Test R2, which is not discussed in Bugni et al. (2013) but has recently been used for a different testing problem in Gandhi et al. (2013).

Romano and Shaikh (2008, Section 3.4) propose to test the hypothesis in (1.2) using the test statistic in (2.3) with a subsampling critical value. Concretely, the test they propose, which we denote by Test SS, is

$$\phi_n^{SS}(\gamma_0) \equiv 1\left\{\inf_{\theta\in\Theta(\gamma_0)} Q_n(\theta) > \hat{c}_n^{SS}(\gamma_0, 1-\alpha)\right\}, \tag{3.7}$$

where $\hat{c}_n^{SS}(\gamma_0, 1-\alpha)$ is the $(1-\alpha)$ quantile of the distribution of $Q_{b_n}^{SS}(\theta)$, which is identical to $Q_n(\theta)$ but computed using a random sample of size b_n without replacement from $\{W_i\}_{i=1}^n$. We assume the subsample size satisfies $b_n \to \infty$ and $b_n/n \to 0$, and show in Theorem C.3 in the Appendix that, conditional on the data,

$$T_{b_n}^{SS}(\gamma_n) \equiv \inf_{\theta\in\Theta(\gamma_n)} Q_{b_n}^{SS}(\theta) \xrightarrow{d} J(\Lambda^{SS}, \Omega) \equiv \inf_{(\theta,\ell)\in\Lambda^{SS}} S(v_\Omega(\theta)+\ell, \Omega(\theta,\theta)),\ a.s.\ , \tag{3.8}$$

where $v_\Omega : \Theta \to \mathbb{R}^k$ is a $\mathbb{R}^k$-valued tight Gaussian process with covariance (correlation) kernel $\Omega \in \mathcal{C}(\Theta^2)$, and Λ^{SS} is the limit (in the Hausdorff metric) of the set

$$\Lambda_{b_n,F_n}^{SS}(\gamma_n) \equiv \left\{(\theta,\ell) \in \Theta(\gamma_n)\times\mathbb{R}^k : \ell = \sqrt{b_n} D_{F_n}^{-1/2}(\theta) E_{F_n}[m(W,\theta)]\right\}\ . \tag{3.9}$$

Romano and Shaikh (2008, Remark 3.11) note that constructing a test for the hypotheses in (1.2) using Test BP would typically result in a conservative test, and use this as a motivation for introducing Test SS. However, they do not provide a formal comparison between their test and Test BP.

To compare Test SS with Test R2 (and Test MR), we define a class of distributions in the alternative hypotheses that are local to the null hypothesis. Notice that the null hypothesis in (1.2) can be written as $\Theta(\gamma_0)\cap\Theta_I(F) \neq \emptyset$, so we do this by defining sequences of distributions F_n for which $\Theta(\gamma_0)\cap\Theta_I(F_n) = \emptyset$ for all $n \in \mathbb{N}$, but where $\Theta(\gamma_n)\cap\Theta_I(F_n) \neq \emptyset$ for a sequence $\{\gamma_n\}_{n\geq 1}$ that approaches γ_0. These alternatives are conceptually similar to those in Andrews and Soares (2010), but the proof of our result involves additional challenges that are specific to the infimum present in the definition of our test statistic. The following definition formalizes the class of local alternative distributions we consider.

Definition 3.2 (Local Alternatives)**.** Let $\gamma_0 \in \Gamma$. The sequence $\{F_n\}_{n\geq 1}$ is a sequence of local alternatives if there is $\{\gamma_n \in \Gamma\}_{n\geq 1}$ such that $\{(\gamma_n, F_n) \in \mathcal{L}_0\}_{n\geq 1}$ and

(a) For all $n \in \mathbb{N}$, $\Theta_I(F_n)\cap\Theta(\gamma_0) = \emptyset$.

(b) $d_H(\Theta(\gamma_n), \Theta(\gamma_0)) = O(n^{-1/2})$.

(c) For any $\theta \in \Theta$, $\kappa_n^{-1} G_{F_n}(\theta) = o(1)$, where $G_F(\theta) \equiv \partial D_F^{-1/2}(\theta) E_F[m(W,\theta)]/\partial\theta'$.

Under the assumption that F_n is a local alternative (see Assumption A.9) and some smoothness conditions (see Assumptions A.7 and A.10) we show that Test R2 has weakly higher asymptotic power than Test SS. This is the content of the next theorem.

Theorem 3.2. *Let Assumptions A.1-A.10 hold. Then,*

$$\liminf_{n\to\infty}\ \left(E_{F_n}[\phi_n^{R2}(\gamma_0)] - E_{F_n}[\phi_n^{SS}(\gamma_0)]\right)\ \geq\ 0\ . \tag{3.10}$$

Remark 3.4. To show that the asymptotic power of Test MR weakly dominates that of Test SS, Theorem 3.2 relies on Assumption A.8, i.e., $\limsup_{n\to\infty} \kappa_n \sqrt{b_n/n} \leq 1$. For the problem of inference on the entire parameter θ, Andrews and Soares (2010) show the analogous result that the asymptotic power of the GMS test weakly dominates that of subsampling tests, based on the stronger condition that $\lim_{n\to\infty}\kappa_n\sqrt{b_n/n}=0$. Given that Theorem 3.2 allows for $\limsup_{n\to\infty}\kappa_n\sqrt{b_n/n}=K\in(0,1]$, we view our result as relatively more robust to the choice of κ_n and b_n. We notice that Theorem 3.2 is consistent with the possibility of a failure of (3.10) whenever Assumption A.8 is violated, i.e., when $\limsup_{n\to\infty}\kappa_n\sqrt{b_n/n}>1$. Remark 3.11 provides a concrete example in which this possibility occurs. In any case, for the recommended choice of $\kappa_n=\sqrt{\ln n}$ in Andrews and Soares (2010, Page 131), a violation of this assumption implies a b_n larger than $O(n^c)$ for all $c\in(0,1)$, which can result in Test SS having poor finite sample power properties, as discussed in Andrews and Soares (2010, Page 137).

Remark 3.5. The inequality in (3.10) can be strict for certain sequences of local alternatives. Lemma C.10 proves this result under the conditions in Assumption A.11. Intuitively, we require a sequence of alternative hypotheses in which one or more moment (in)equality is slack by magnitude that is $o(b_n^{-1/2})$ and larger than $O(\kappa_n n^{-1/2})$. We provide an illustration of Assumption A.11 in Example 3.2. We also empirically illustrate the role of this assumption in our Monte Carlo simulations.

Remark 3.6. There are reasons to prefer Test MR over Test SS beyond the asymptotic power result in Theorem 3.2. First, we find in all our simulations that Test SS is significantly more sensitive to the choice of b_n than Test MR is to the choice of κ_n. Second, in the context of inference on the entire parameter θ, subsampling tests have been shown to have an error in rejection probability (ERP) of order $O_p(b_n/n+b_n^{-1/2})\geq O_p(n^{-1/3})$, while GMS-type tests have an ERP of order $O_p(n^{-1/2})$ (c.f. Bugni, 2014). We expect an analogous result to hold for the problem of inference on $f(\theta)$, but a formal proof is well beyond the scope of this paper.

3.3 Understanding the new test and its power advantages

The previous sections derived two important results. On the one hand, Test R1 weakly dominates Test BP in terms of finite sample power and, under certain conditions and for some alternatives, strictly dominates Test BP in terms of asymptotic power. On the other hand, Test R2 weakly dominates Test SS in terms of asymptotic power for all the alternatives in Definition 3.2, and strictly dominates Test SS for certain local alternatives. These power properties are inherited by Test MR by virtue of the inequalities in (3.3). We summarize these lessons in the following corollary.

Corollary 3.2. *Let Assumptions A.1-A.10 hold. Then*

$$\liminf_{n\to\infty}(E_{F_n}[\phi_n^{MR}(\gamma_n)]-E_{F_n}[\phi_n^{SS}(\gamma_n)])\geq 0 \quad \text{and} \quad \liminf_{n\to\infty}(E_{F_n}[\phi_n^{MR}(\gamma_n)]-E_{F_n}[\phi_n^{BP}(\gamma_n)])\geq 0\ . \tag{3.11}$$

The result in Corollary 3.2 is useful as it shows that Test MR cannot be asymptotically dominated by any of the other tests. The next natural step is to understand the type of alternatives for which both of the inequalities in (3.11) become strict. There are two ways of obtaining such result. First, it could be the case that either Test R1 strictly dominates Test BP (as in Remark 3.3), or that Test R2 strictly dominates Test SS (as in Remark 3.5). We illustrate a situation in which Test R2 has strictly better asymptotic power than Test SS in Example 3.2. It could also be possible that Test MR has strictly better asymptotic power than both Test R1 and Test R2, which results in Test MR strictly dominating (asymptotically) Test BP and Test SS. We illustrate this situation in Example 3.1.

Example 3.1 (Test MR vs. Tests R1 and R2)**.** Let $W=(W_1,W_2,W_3)\in\mathbb{R}^3$ be a random vector with distribution F_n, $V_{F_n}[W]=I_3$, $E_{F_n}[W_1]=\mu_1\kappa_n/\sqrt{n}$, $E_{F_n}[W_2]=\mu_2\kappa_n/\sqrt{n}$, and $E_{F_n}[W_3]=\mu_3/\sqrt{n}$ for some

$\mu_1 > 1$, $\mu_2 \in (0,1)$, and $\mu_3 \in \mathbb{R}$. Consider the following model with $\Theta = [-C, C]^3$ for some $C > 0$,

$$\begin{aligned} E_{F_n}[m_1(W_i,\theta)] &= E_{F_n}[W_{i,1} - \theta_1] \geq 0 \ , \\ E_{F_n}[m_2(W_i,\theta)] &= E_{F_n}[W_{i,2} - \theta_2] \geq 0 \ , \\ E_{F_n}[m_3(W_i,\theta)] &= E_{F_n}[W_{i,3} - \theta_3] = 0 \ . \end{aligned} \tag{3.12}$$

We are interested in testing the hypotheses

$$H_0 : \theta = (0,0,0) \text{ vs. } H_1 : \theta \neq (0,0,0) \ ,$$

which implies that $f(\theta) = \theta$, $\Theta(\gamma_0) = \{(0,0,0)\}$, and $\hat{\Theta}_I(\gamma_0) = \{(0,0,0)\}$.[7] Note that H_0 is true if and only if $\mu_3 = 0$. The model in (3.12) is linear in θ, and so many relevant parameters and estimators do not depend on θ. These include $\hat{\sigma}_j(\theta) = \hat{\sigma}_j$ for $j = 1,2,3$, so $\hat{D}_n^{-1/2}(\theta) = \hat{D}_n^{-1/2}$, $\tilde{v}_{n,j}(\theta) = \tilde{v}_{n,j} = \sqrt{n}\hat{\sigma}_j^{-1}(\bar{W}_{n,j} - E_{F_n}[W_j])$, and

$$v_n^*(\theta) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \hat{D}_n^{-1/2}(\theta)(m(W_i,\theta) - \bar{m}_n(\theta))\zeta_i = \frac{1}{\sqrt{n}} \sum_{i=1}^n \hat{D}_n^{-1/2}(W_i - \bar{W}_n)\zeta_i = v_n^* \ , \tag{3.13}$$

where $\{\zeta_i\}_{i=1}^n$ is i.i.d. $N(0,1)$. It follows that $\{v_n^* | \{W_i\}_{i=1}^n\} \sim N(0,1)$ a.s. For simplicity here, we use the Modified Method of Moments criterion function given by

$$S(m,\Sigma) = \sum_{j=1}^p [m_j/\sigma_j]_-^2 + \sum_{j=p+1}^k (m_j/\sigma_j)^2 \ , \tag{3.14}$$

where $[x]_- \equiv \min\{x,0\}$, and the simplest function $\varphi(\cdot)$ proposed by Andrews and Soares (2010),

$$\varphi_j(x) = \infty * 1\{x > 1\} \text{ for } j = 1,\ldots,p \ , \text{ and } \varphi_j(x) = 0 \text{ for } j = p+1,\ldots,k \ . \tag{3.15}$$

The sample criterion function is given by

$$Q_n(\theta) = \left[\sqrt{n}\hat{\sigma}_1^{-1}(\bar{W}_1 - \theta_1)\right]_-^2 + \left[\sqrt{n}\hat{\sigma}_2^{-1}(\bar{W}_2 - \theta_2)\right]_-^2 + \left[\sqrt{n}\hat{\sigma}_3^{-1}(\bar{W}_3 - \theta_3)\right]^2 \ ,$$

and so the test statistic satisfies

$$\begin{aligned} \inf_{\theta \in \Theta(\gamma_0)} Q_n(\theta) &= \left[\sqrt{n}\hat{\sigma}_1^{-1}\bar{W}_1\right]_-^2 + \left[\sqrt{n}\hat{\sigma}_2^{-1}\bar{W}_2\right]_-^2 + \left[\sqrt{n}\hat{\sigma}_3^{-1}\bar{W}_3\right]^2 \ , \\ &= \left[\tilde{v}_{n,1} + \mu_1\hat{\sigma}_1^{-1}\kappa_n\right]_-^2 + \left[\tilde{v}_{n,2} + \mu_2\hat{\sigma}_2^{-1}\kappa_n\right]_-^2 + \left[\tilde{v}_{n,3} + \hat{\sigma}_3^{-1}\mu_3\right]^2 \ , \\ &\xrightarrow{d} [Z_3 + \mu_3]^2, \quad Z_3 \sim N(0,1). \end{aligned}$$

To study the behavior of the Test R1, Test R2, and Test MR, we derive convergence statements that occur conditionally on $\{W_i\}_{i=1}^n$, and exploit that $\kappa_n^{-1}\tilde{v}_{n,j} \xrightarrow{p} 0$ and $\hat{\sigma}_j^{-1} \xrightarrow{p} \sigma_j^{-1} = 1$ for $j = 1,2,3$. Below we also use the notation $Z = (Z_1, Z_2, Z_3) \sim N(\mathbf{0}_3, I_3)$.

Test R1: This test uses the (conditional) $(1-\alpha)$ quantile of the following random variable,

$$T_n^{R1}(\gamma_0) = \inf_{\theta \in \hat{\Theta}_I(\gamma_0)} \left\{ \left[v_{n,1}^* + \varphi_1(\kappa_n^{-1}\sqrt{n}\hat{\sigma}_1^{-1}(\bar{W}_1 - \theta_1))\right]_-^2 + \left[v_{n,2}^* + \varphi_2(\kappa_n^{-1}\sqrt{n}\hat{\sigma}_2^{-1}(\bar{W}_2 - \theta_2))\right]_-^2 + \left[v_{n,3}^*\right]^2 \right\} \ ,$$

[7] In this example we use $f(\theta) = \theta$ for simplicity, as it makes the infimum over $Q_n(\theta)$ trivial. We could generate the same conclusions using a different function by adding some complexity to the structure of the example.

$$= \left[v^*_{n,1} + \infty * 1\{\kappa_n^{-1}\tilde{v}_{n,1} + \mu_1\hat{\sigma}_1^{-1} > 1\}\right]_-^2 + \left[v^*_{n,2} + \infty * 1\{\kappa_n^{-1}\tilde{v}_{n,2} + \mu_2\hat{\sigma}_2^{-1} > 1\}\right]_-^2 + \left[v^*_{n,3}\right]^2 ,$$
$$\stackrel{d}{\to} [Z_2]_-^2 + [Z_3]^2 \text{ w.p.a.1 },$$

since $\mu_1 > 1$ and $\mu_2 < 1$.

Test R2: This test uses the (conditional) $(1-\alpha)$ quantile of the following random variable,

$$T_n^{R2}(\gamma_0) = \inf_{\theta \in \Theta(\gamma_0)} \Big\{ \left[v^*_{n,1} + \kappa_n^{-1}\sqrt{n}\hat{\sigma}_1^{-1}(\bar{W}_1 - \theta_1)\right]_-^2 + \left[v^*_{n,2} + \kappa_n^{-1}\sqrt{n}\hat{\sigma}_2^{-1}(\bar{W}_2 - \theta_2)\right]_-^2$$
$$+ [v^*_{n,3} + \kappa_n^{-1}\sqrt{n}\hat{\sigma}_3^{-1}(\bar{W}_3 - \theta_3)]^2 \Big\} ,$$
$$= \left[v^*_{n,1} + \kappa_n^{-1}\tilde{v}_{n,1} + \mu_1\hat{\sigma}_1^{-1}\right]_-^2 + \left[v^*_{n,2} + \kappa_n^{-1}\tilde{v}_{n,2} + \mu_2\hat{\sigma}_2^{-1}\right]_-^2 + \left[v^*_{n,3} + \kappa_n^{-1}\tilde{v}_{n,3} + \hat{\sigma}_3^{-1}\kappa_n^{-1}\mu_3\right]^2 ,$$
$$\stackrel{d}{\to} [Z_1 + \mu_1]_-^2 + [Z_2 + \mu_2]_-^2 + [Z_3]^2 \text{ w.p.a.1 } .$$

Test MR: This test uses the (conditional) $(1-\alpha)$ quantile of the following random variable,

$$T_n^{MR}(\gamma_0) = \min\{T_n^{R1}(\gamma_0), T_n^{R2}(\gamma_0)\} \stackrel{d}{\to} \min\{[Z_1 + \mu_1]_-^2 + [Z_2 + \mu_2]_-^2 , [Z_2]_-^2\} + [Z_3]^2 \text{ w.p.a.1 } .$$

□

The example provides important lessons about the relative power of all these tests, as well as illustrating a case in which Test MR has strict better power than both Test BP and Test SS. We summarize these lessons in the following remarks.

Remark 3.7. Since $\min\{[Z_1 + \mu_1]_-^2 + [Z_2 + \mu_2]_-^2, [Z_2]_-^2\} \geq 0$, it follows that the null rejection probability of Test MR along this sequence will not exceed α under H_0. More importantly, note that

$$P([Z_1 + \mu_1]_-^2 + [Z_2 + \mu_2]_-^2 < [Z_2]_-^2) \geq P(Z_1 + \mu_1 \geq 0)P(Z_2 < 0) > 0 ,$$
$$P([Z_1 + \mu_1]_-^2 + [Z_2 + \mu_2]_-^2 > [Z_2]_-^2) \geq P(Z_1 + \mu_1 < 0)P(Z_2 \geq 0) > 0 , \quad (3.16)$$

which implies that whether $T_n^{MR}(\gamma_0)$ equals $T_n^{R1}(\gamma_0)$ or $T_n^{R2}(\gamma_0)$ is random, conditionally on $\{W_i\}_{i=1}^n$. This means that using Test MR is not equivalent to using either Test R1 or Test R2.

Remark 3.8. Example 3.1 and (3.16) show that the conditional distribution of $T_n^{MR}(\gamma_0)$ is (asymptotically) strictly dominated by the conditional distributions of $T_n^{R1}(\gamma_0)$ or $T_n^{R2}(\gamma_0)$. Given that all these tests use the same test statistic, what determines their relative asymptotic power is the limit of their respective critical values. In the example above, we can numerically compute the $1-\alpha$ quantiles of the limit distributions of $T_n^{R1}(\gamma_0)$, $T_n^{R2}(\gamma_0)$, and $T_n^{MR}(\gamma_0)$ after fixing some values for μ_1 and μ_2. For example, setting both of these parameters close to 1 results in asymptotic 95% quantiles of $T_n^{R1}(\gamma_0)$, $T_n^{R2}(\gamma_0)$, and $T_n^{MR}(\gamma_0)$ equal to 5.15, 4.18, and 4.04, respectively.

Remark 3.9. Example 3.1 illustrates that the basic resampling tests, Test R1 and R2, do not dominate each other in terms of asymptotic power. For example, if we consider the model in (3.12) with the second inequality removed, it follows that

$$T_n^{R1}(\gamma_0) \stackrel{d}{\to} [Z_3]^2 \quad \text{and} \quad T_n^{R2}(\gamma_0) \stackrel{d}{\to} [Z_1 + \mu_1]_-^2 + [Z_3]^2 . \quad (3.17)$$

In this case Test R1 has strictly better asymptotic power than Test R2. For this example, taking μ_1 close to 1 gives asymptotic 95% quantiles of Tests R1 and R2 equal to 3.84 and 4.00, respectively. On the other hand, if we consider the model in (3.12) with the first inequality removed, it follows that

$$T_n^{R1}(\gamma_0) \xrightarrow{d} [Z_2]_-^2 + [Z_3]^2 \quad \text{and} \quad T_n^{R2}(\gamma_0) \xrightarrow{d} [Z_2 + \mu_2]_-^2 + [Z_3]^2 \ . \tag{3.18}$$

Since $[Z_2 + \mu_2]_-^2 \leq [Z_2]_-^2$ (with strict inequality when $Z_2 < 0$), this case represents a situation where Test R1 has strictly worse asymptotic power than Test R2. For this example, taking μ_2 close to 1 results in asymptotic 95% quantiles of Tests R1 and R2 equal to 5.13 and 4.00, respectively.

Example 3.2 (Test R2 versus Test SS)**.** Let $W = (W_1, W_2, W_3) \in \mathbb{R}^3$ be a random vector with distribution F_n, $V_{F_n}[W] = I_3$, $E_{F_n}[W_1] = \mu_1 \kappa_n/\sqrt{n}$, $E_{F_n}[W_2] = \mu_2/\sqrt{n}$, and $E_{F_n}[W_3] = 0$ for some $\mu_1 \geq 0$ and $\mu_2 \leq 0$. Consider the model in (3.12) with $\Theta = [-C, C]^3$ for some $C > 0$, and the hypotheses

$$H_0 : f(\theta) = (\theta_1, \theta_2) = (0, 0) \text{ vs. } H_1 : f(\theta) = (\theta_1, \theta_2) \neq (0, 0) \ .$$

In this case $\Theta(\gamma_0) = \{(0, 0, \theta_3) : \theta_3 \in [-C, C]\}$ and H_0 is true if and only if $\mu_2 = 0$. The model in (3.12) is linear in θ, and so many relevant parameters and estimators do not depend on θ. These include $\hat{\sigma}_j(\theta) = \hat{\sigma}_j$ for $j = 1, 2, 3$, so $\hat{D}_n^{-1/2}(\theta) = \hat{D}_n^{-1/2}$, $\tilde{v}_{n,j} = \sqrt{n}\hat{\sigma}_j^{-1}(\bar{W}_j - E_{F_n}[W_j])$, and $v_n^*(\theta) = v_n^*$ as defined in (3.13). For simplicity, we assume that $\kappa_n\sqrt{b_n/n}$ converges to a constant K.

To study the behavior of the tests we derive convergence statements that occur conditionally on $\{W_i\}_{i=1}^n$, and exploit that $\kappa_n^{-1}\tilde{v}_{n,j} \xrightarrow{p} 0$ and $\hat{\sigma}_j^{-1} \xrightarrow{p} \sigma_j^{-1} = 1$ for $j = 1, 2, 3$. We also use $Z = (Z_1, Z_2, Z_3) \sim N(\mathbf{0}_3, I_3)$ and assume Andrews and Soares (2010, Assumption GMS5). As in Example 3.1, $S(\cdot)$ is as in (3.14) and $\varphi(\cdot)$ as in (3.15), which results in a sample criterion function given by

$$Q_n(\theta) = \left[\sqrt{n}\hat{\sigma}_1^{-1}(\bar{W}_1 - \theta_1)\right]_-^2 + \left[\sqrt{n}\hat{\sigma}_2^{-1}(\bar{W}_2 - \theta_2)\right]_-^2 + \left[\sqrt{n}\hat{\sigma}_3^{-1}(\bar{W}_3 - \theta_3)\right]^2 \ .$$

The test statistic satisfies

$$\begin{aligned}\inf_{\theta \in \Theta(\gamma_0)} Q_n(\theta) &= \inf_{\theta_3 \in [-C, C]} \left[\sqrt{n}\hat{\sigma}_1^{-1}\bar{W}_1\right]_-^2 + \left[\sqrt{n}\hat{\sigma}_2^{-1}\bar{W}_2\right]_-^2 + \left[\sqrt{n}\hat{\sigma}_3^{-1}(\bar{W}_3 - \theta_3)\right]^2 \ , \\ &= \left[\tilde{v}_{n,1} + \mu_1\hat{\sigma}_1^{-1}\kappa_n\right]_-^2 + \left[\tilde{v}_{n,2} + \mu_2\hat{\sigma}_2^{-1}\right]_-^2 \xrightarrow{d} [Z_1]^2 1\{\mu_1 = 0\} + [Z_2 + \mu_2]^2 \ .\end{aligned}$$

Test R2: This test uses the (conditional) $(1 - \alpha)$ quantile of the following random variable,

$$\begin{aligned}T_n^{R2}(\gamma_0) &= \inf_{\theta \in \Theta(\gamma_0)} \left\{\left[v_{n,1}^* + \kappa_n^{-1}\sqrt{n}\hat{\sigma}_1^{-1}(\bar{W}_1 - \theta_1)\right]_-^2 + \left[v_{n,2}^* + \kappa_n^{-1}\sqrt{n}\hat{\sigma}_2^{-1}(\bar{W}_2 - \theta_2)\right]_-^2 + \left[\sqrt{n}\hat{\sigma}_3^{-1}(\bar{W}_3 - \theta_3)\right]^2\right\} \ , \\ &= \left[v_{n,1}^* + \kappa_n^{-1}\tilde{v}_{n,1} + \mu_1\hat{\sigma}_1^{-1}\right]_-^2 + \left[v_{n,2}^* + \kappa_n^{-1}\tilde{v}_{n,2} + \kappa_n^{-1}\mu_2\hat{\sigma}_2^{-1}\right]_-^2 \xrightarrow{d} [Z_1 + \mu_1]_-^2 + [Z_2]_-^2 \text{ w.p.a.1} \ .\end{aligned}$$

Test SS: This test draws $\{W_i^\star\}_{i=1}^{b_n}$ i.i.d. with replacement from $\{W_i\}_{i=1}^n$ and computes $v_{b_n}^\star(\theta) = \frac{1}{\sqrt{b_n}}\sum_{i=1}^{b_n} \hat{D}_{b_n}^{\star,-1/2}(\theta)m(W_i^\star, \theta)$. Now define

$$\tilde{v}_{b_n}^\star(\theta) = \frac{1}{\sqrt{b_n}}\sum_{i=1}^{b_n} \hat{D}_{b_n}^{\star,-1/2}(\theta)\{m(W_i^\star, \theta) - E_{F_n}[m(W_i, \theta)]\} \ , \tag{3.19}$$

and since $\tilde{v}_{b_n}^\star(\theta) = \tilde{v}_{b_n}^\star$, Politis et al. (Theorem 2.2.1, 1999) implies that $\{\tilde{v}_{b_n}^\star | \{W_i\}_{i=1}^n\} \xrightarrow{d} N(0, 1)$ a.s.

Test SS uses the conditional $(1-\alpha)$ quantile of the following random variable

$$\begin{aligned} T^{SS}_{b_n}(\gamma_0) &= \inf_{\theta \in \Theta(\gamma_0)} \left\{ [\sqrt{b_n}\hat{\sigma}^{\star,-1}_{b_n,1}(\bar{W}^{\star}_{b_n,1} - \theta_1)]^2_- + [\sqrt{b_n}\hat{\sigma}^{\star,-1}_{b_n,2}(\bar{W}^{\star}_{b_n,2} - \theta_2)]^2_- + [\sqrt{b_n}\hat{\sigma}^{\star,-1}_{b_n,3}(\bar{W}^{\star}_{b_n,3} - \theta_3)]^2 \right\} , \\ &= [\tilde{v}^{\star}_{b_n,1} + \hat{\sigma}^{\star,-1}_{b_n,1}\sqrt{b_n}\mu_1\kappa_n/\sqrt{n}]^2_- + [\tilde{v}^{\star}_{n,b_n,2} + \hat{\sigma}^{\star,-1}_{b_n,2}\sqrt{b_n}\mu_2/\sqrt{n}]^2_- , \\ &\xrightarrow{d} [Z_1 + K\mu_1]^2_- + [Z_2]^2_- \text{ w.p.a.1} , \end{aligned}$$

where we used $\kappa_n\sqrt{b_n/n} \to K$. □

Remark 3.10. In Example 3.2, $T^{R2}_n(\gamma_0)$ and $T^{SS}_{b_n}(\gamma_0)$ have the same asymptotic distribution, conditionally on $\{W_i\}^n_{i=1}$, when $\mu_1 = 0$ or $K = 1$. However, if $\mu_1 > 0$ and $K < 1$, it follows that $T^{SS}_{b_n}(\gamma_0)$ asymptotically strictly dominates $T^{R2}_n(\gamma_0)$ in the first order stochastic sense, conditionally on $\{W_i\}^n_{i=1}$. Specifically,

$$P\left([Z_2 + \mu_2]^2_- > q_{1-\alpha}([Z_1 + \mu_1]^2_- + [Z_2]^2_-)\right) > P\left([Z_2 + \mu_2]^2_- > q_{1-\alpha}([Z_1 + K\mu_1]^2_- + [Z_2]^2_-)\right) , \tag{3.20}$$

where $q_{1-\alpha}(X)$ denotes the $1-\alpha$ quantile of X. Thus, Test R2 is *strictly* less conservative under H_0 (i.e. when $\mu_2 = 0$) and *strictly* more powerful under H_1 (i.e. when $\mu_2 < 0$).

Remark 3.11. Example 3.2 shows that Test SS could deliver higher power than Test R2 if $\mu_1 > 0$ and $K > 1$, i.e., if Assumption A.8 is violated. However, for the recommended choice of $\kappa_n = \sqrt{\ln n}$ in Andrews and Soares (2010, Page 131), a violation of this assumption can result in Test SS having poor finite sample power properties, as already discussed in Remark 3.4.

4 Monte Carlo simulations

In this section we consider an entry game model similar to that in Canay (2010). Suppose that firm $j \in \{1, 2\}$ decides whether to enter ($z_{j,i} = 1$) a market $i \in \{1, \dots, n\}$ or not ($z_{j,i} = 0$) based on its profit function $\pi_{j,i} = (\varepsilon_{j,i} - \theta_j z_{-j,i})1\{z_{j,i} = 1\}$, where $\varepsilon_{j,i}$ is firm j's benefit of entry in market i and $z_{-j,i}$ denotes the decision of the other firm. Let $\varepsilon_{j,i} \sim \text{Uniform}(0, 1)$ and $\theta_0 = (\theta_1, \theta_2) \in (0, 1)^2$. There are four outcomes in this game: (i) $W_i \equiv (z_{1,i}, z_{2,i}) = (1, 1)$ is the unique Nash equilibrium (NE) if $\varepsilon_{j,i} > \theta_j$ for all j; (ii) $W_i = (1, 0)$; is the unique NE if $\varepsilon_{1,i} > \theta_1$ and $\varepsilon_{2,i} < \theta_2$; (iii) $W_i = (0, 1)$ is the unique NE if $\varepsilon_{1,i} < \theta_1$ and $\varepsilon_{2,i} > \theta_2$ and; (iv) there are multiple equilibria if $\varepsilon_{j,i} < \theta_j$ for all j as both $W_i = (1, 0)$ and $W_i = (0, 1)$ are NE. Without further assumptions this model implies

$$\begin{aligned} E_F[m_1(W_i, \theta)] &= E_F[z_{1,i}z_{2,i} - (1-\theta_1)(1-\theta_2)] = 0 \\ E_F[m_2(W_i, \theta)] &= E_F[z_{1,i}(1 - z_{2,i}) - \theta_2(1-\theta_1)] \geq 0 \\ E_F[m_3(W_i, \theta)] &= E_F[\theta_2 - z_{1,i}(1 - z_{2,i})] \geq 0 . \end{aligned} \tag{4.1}$$

The identified set $\Theta_I(F)$ in this model is a curve in $\mathbb{R}^2$. We generate data using $\theta_0 = (0.3, 0.5)$ as the true parameter and $p = 0.7$ as the true probability of selecting $W_i = (1, 0)$ in the region of multiple equilibria. This gives an identified set having a first coordinate ranging from 0.19 to 0.36, and a second coordinate ranging from 0.45 to 0.56 (see Canay, 2010, Figure 1). Our setup differs from that in Canay (2010) in that we add the following redundant inequality,

$$E_F[m_4(\zeta_i, \theta)] = E_F[\zeta_i + \theta_1 - \theta_2] \geq 0 , \tag{4.2}$$

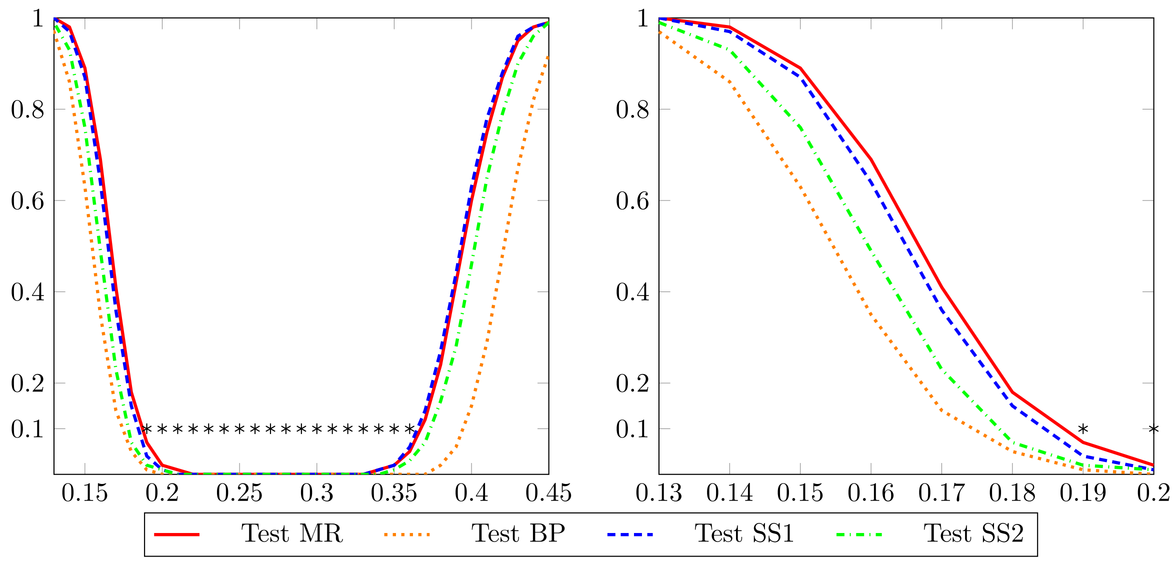

Figure 1: Rejection probabilities under the null and alternative hypotheses when $f(\theta) = \theta_1$. Tests considered are: Test MR (solid red line), Test BP (dotted orange line), Test SS1 (dashed blue line), and Test SS2 (dashed-dotted green line). Black asterisks indicate values of θ_1 in the identified set at the nominal level. Left panel shows rejection rates to the left and right of the identified set. Right panel zooms-in the power differences to the left. In all cases $n = 1,000$, $\alpha = 0.10$, and $MC = 2,000$.

where $\zeta_i \sim N(\mu_\zeta, 1)$ and $\mu_\zeta = 0.38$. There are three aspects worth noting about this inequality. First, the value of μ_ζ was chosen to place the line determining the inequality at $\kappa_n/\sqrt{n}$ distance from the identified set. As a consequence, the identified set $\Theta_I(F)$ remains unchanged. Second, the line determining the inequality is close enough to $\Theta_I(F)$ so that it can occasionally appear to be binding in finite samples. More specifically, the inequality is used to generate the situation required in Assumption A.11 and hence illustrate a case where Test MR delivers higher power than Test SS.[8] Finally, the inequality only affects power for alternatives approaching the identified set from the left (right) in the case of θ_1 (θ_2), and does not affect power in the other direction.

We set $n = 1,000$ and $\alpha = 0.10$, and simulate the data by taking independent draws of $\varepsilon_{j,i} \sim \text{Uniform}(0, 1)$ for $j = \{1, 2\}$ and computing the equilibrium according to the region in which $\varepsilon_i \equiv (\varepsilon_{1,i}, \varepsilon_{2,i})$ falls. We consider subvector inference for this model, with

$$H_0 : f(\theta) = \theta_s = \gamma_0 \quad \text{vs} \quad H_1 : f(\theta) = \theta_s \neq \gamma_0 \quad \text{for } s = 1, 2 \,, \tag{4.3}$$

and perform $MC = 2,000$ Monte Carlo replications. We report results for Test MR (with $\kappa_n = \sqrt{\ln n} = 2.63$)[9], Test BP, Test SS1 (with $b_n = n^{2/3} = 100$ as considered in Bugni, 2014), and Test SS2 (with $b_n = n/4 = 250$ as considered in Ciliberto and Tamer, 2010).[10]

Figure 1 shows the rejection probabilities under the null and alternative hypotheses for the first coordinate, i.e., $f(\theta) = \theta_1$. The results show that Test MR has null rejection probabilities (at the boundary of the identified set) closer to the nominal size than those of Test BP which is highly conservative in this

[8] We also run simulations of the model without (4.2) and these simulations are available upon request. For that model Test MR and Test SS show similar rejection rates for appropriate choices of b_n, which is consistent with the theory as the local asymptotic power of Test MR and Test SS coincide in the model determined by (4.1) alone.

[9] We computed Test MR with $\kappa_n = 0.80\sqrt{\ln n}$ and $\kappa_n = 0.90\sqrt{\ln n}$, as in Bugni et al. (2013), and obtained numerically identical results to those reported here.

[10] We tried different choices of subsampling sizes and found that $b_n = n^{2/3}$ works best and that $n/4$ works worst. The choice $b_n = n^{2/3}$ corresponds to the optimal rate for the subsample size to minimize ERP; see Bugni (2010). The choice $b_n = n/4$ is the subsample size rule used by Ciliberto and Tamer (2010).

case. Test SS can be close to Test MR or Test BP depending on the subsample size b_n. The right panel illustrates the power differences more clearly, and it shows that the differences in the power of Test MR with respect to that of Test BP, Test SS1, and Test SS2 could be as high as 0.30, 0.05, and 0.20, respectively. Given that the simulation standard error with 2,000 replications is between 0.007 and 0.01, depending on the alternative, these differences are significant. The power differences between Test MR and Test SS1 disappear for alternatives approaching the identified set from the right, as in that case the inequality in (4.3) does not play any role and both tests have the same asymptotic local power.

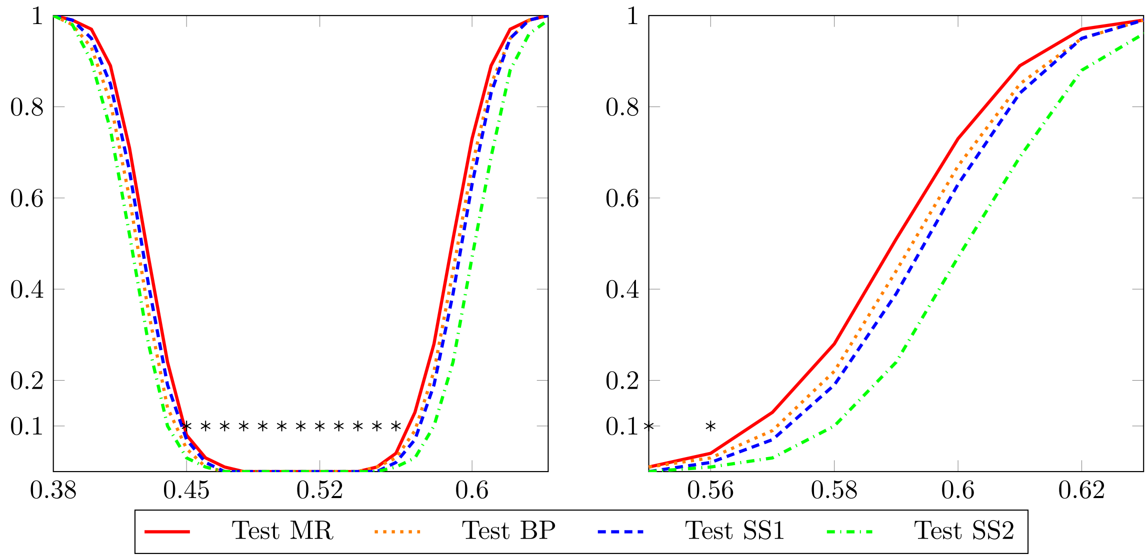

Figure 2: Rejection probabilities under the null and alternative hypotheses when $f(\theta) = \theta_2$. Tests considered are: Test MR (solid red line), Test BP (dotted orange line), Test SS1 (dashed blue line), and Test SS2 (dashed-dotted green line). Black asterisks indicate values of θ_2 in the identified set at the nominal level. Left panel shows rejection rates to the left and right of the identified set. Right panel zooms-in the power differences to the right. In all cases $n = 1,000$, $\alpha = 0.10$, and $MC = 2,000$.

Figure 2 shows the rejection probabilities under the null and alternative hypotheses for the second coordinate, i.e., $f(\theta) = \theta_2$. The results show again that Test MR has null rejection probabilities (at the boundary of the identified set) closer to the nominal size than those of Test BP, although Test BP performs better than in the previous case. Similarly, the performance of Test SS highly depends on the choice of subsampling size and it performs particularly poorly for the subsampling size rule used by Ciliberto and Tamer (2010). The differences in the power of Test MR with respect to that of Test BP, Test SS1, and Test SS2 could be as high as 0.07, 0.12, and 0.27, respectively. As before, given that the simulation standard error is between 0.007 and 0.01, these differences are significant.

Table 1 reports rejection probabilities for Test MR and Test SS after "size-correcting" the critical values to make the null rejection probabilities at the boundary (i.e. at $\theta_1 = 0.19$ or $\theta_2 = 0.56$) equal to the nominal level $\alpha = 0.10$. In the case of the first coordinate, the size-adjusted power of Test MR and Test SS are identical, suggesting that the power differences in the right panel of Figure 1 are due to better size control. On the contrary, inference for the second coordinate exhibits a size-adjusted power of Test MR consistently above that of Test SS by about 0.04-0.05. We note that this model is not symmetric in its coordinates, so it would be expected to observe differences in performance across coordinates.

The results of the simulations are consistent with the theoretical results in Theorems 2.1 and 3.2, Corollary 3.2, and Remark 3.5. They illustrate a concrete case where Test MR provides strictly higher power than Test

	H_0	H_1				
	$\theta_1 = 0.19$	$\theta_1 = 0.18$	$\theta_1 = 0.17$	$\theta_1 = 0.16$	$\theta_1 = 0.15$	$\theta_1 = 0.14$
Rejection Rate						
Test MR	0.10	0.24	0.50	0.77	0.93	0.99
Test SS1	0.10	0.25	0.50	0.77	0.93	0.99
Test SS2	0.10	0.16	0.36	0.64	0.87	0.97
	H_0	H_1				
	$\theta_2 = 0.56$	$\theta_2 = 0.57$	$\theta_2 = 0.58$	$\theta_2 = 0.59$	$\theta_2 = 0.60$	$\theta_2 = 0.61$
Rejection Rate						
Test MR	0.10	0.17	0.33	0.57	0.78	0.92
Test SS1	0.10	0.14	0.28	0.51	0.73	0.90
Test SS2	0.10	0.06	0.13	0.28	0.51	0.73

Table 1: Size Adjusted Rejection probabilities for Test MR and Test SS. In all cases $n = 1,000$, $\alpha = 0.10$ and $MC = 2,000$.

BP and Test SS. But the simulations reveal additional features. For example, in all our simulations the finite sample behavior of Test MR is insensitive to the choice of κ_n while the finite sample behavior of Test SS is *highly sensitive* to the choice of b_n. The simulations also illustrate that Test BP can be highly conservative and suffer from low power. Finally, note that Test SS can perform similarly to Test MR whenever b_n is appropriately chosen and the conditions in Assumption A.11 are not met - to the right (left) of $\Theta_I(F)$ for θ_1 (θ_2).

5 Concluding remarks

In this paper, we introduce a test, denoted Test MR, for the null hypothesis $H_0 : f(\theta) = \gamma_0$, where $f(\cdot)$ is a known function, γ_0 is a known constant, and θ is a parameter that is partially identified by a moment (in)equality model. Our test can be used to construct CS's for $f(\theta)$ by exploiting the well-known duality between tests and CS's. The leading application of our inference method is the construction of marginal CS's for individual coordinates of a parameter vector θ, which is implemented by setting $f(\theta) = \theta_s$ for $s = 1, \ldots, d_\theta$ and collecting all values of $\gamma_0 \in \Gamma$ for which H_0 is not rejected.

We show that our inference method controls asymptotic size uniformly over a large class of distributions of the data. The current literature describes only two other procedures that deliver uniform size control for these types of problems: projection-based and subsampling inference. Relative to projection-based procedure, our method presents three advantages: (i) it weakly dominates in terms of finite sample power, (ii) it strictly dominates in terms of asymptotic power, and (iii) it is typically less computationally demanding. Relative to a subsampling, our method presents two advantages: (i) it strictly dominates in terms of asymptotic power (for reasonable choices of subsample size), (ii) it appears to be less sensitive to the choice of its tuning parameter than subsampling is to the choice of subsample size.

There are two interesting extensions of the test we propose that are worth mentioning. First, our paper does not consider conditional moment restrictions, c.f. Andrews and Shi (2013), Chernozhukov et al. (2013), Armstrong (2011), and Chetverikov (2013). Second, our asymptotic framework is one where the limit distributions do not depend on tuning parameters used at the moment selection stage, as opposed to Andrews and Barwick (2012) and Romano et al. (2013). These two extensions are well beyond the scope of this paper and so we leave them for future research.

Appendix A Notation and computational algorithm

Throughout the Appendix we employ the following notation, not necessarily introduced in the text.

$\mathcal{P}_0$	$\{F \in \mathcal{P} : \Theta_I(F) \neq \emptyset\}$
$\mathcal{L}$	$\{(\gamma, F) : F \in \mathcal{P}, \gamma \in \Gamma\}$
$\Theta(\gamma)$	$\{\theta \in \Theta : f(\theta) = \gamma\}$
$\Theta_I(F,\gamma)$	$\{\theta \in \Theta_I(F) : f(\theta) = \gamma\}$
$\mathcal{L}_0$	$\{(\gamma, F) : F \in \mathcal{P}, \gamma \in \Gamma, \Theta_I(F,\gamma) \neq \emptyset\}$
$\hat{\Theta}_I(\gamma)$	$\{\theta \in \Theta(\gamma) : Q_n(\theta) \leq T_n(\gamma)\}$
$\Theta_I^{\ln \kappa_n}(\gamma)$	$\{\theta \in \Theta(\gamma) : S(\sqrt{n} E_{F_n}[m(W,\theta)], \Sigma_{F_n}(\theta)) \leq \ln \kappa_n\}$
$\Lambda_{n,F}(\gamma)$	$\{(\theta,\ell) \in \Theta(\gamma) \times \mathbb{R}^k : \ell = \sqrt{n} D_F^{-1/2}(\theta) E_F[m(W_i,\theta)]\}$
$\Lambda_{b_n,F}^{SS}(\gamma)$	$\{(\theta,\ell) \in \Theta(\gamma) \times \mathbb{R}^k : \ell = \sqrt{b_n} D_F^{-1/2}(\theta) E_F[m(W,\theta)]\}$
$\Lambda_{n,F}^{R2}(\gamma)$	$\{(\theta,\ell) \in \Theta(\gamma) \times \mathbb{R}^k : \ell = \kappa_n^{-1}\sqrt{n} D_F^{-1/2}(\theta) E_F[m(W_i,\theta)]\}$
$\Lambda_{n,F}^{R1}(\gamma)$	$\{(\theta,\ell) \in \Theta_I^{\delta_n}(\gamma) \times \mathbb{R}^k : \ell = \kappa_n^{-1}\sqrt{n} D_F^{-1/2}(\theta) E_F[m(W,\theta)]\}$

Table 2: Important Notation

For any $u \in \mathbb{N}$, $\mathbf{0}_u$ is a column vector of zeros of size u, $\mathbf{1}_u$ is a column vector of ones of size u, and I_u is the $u \times u$ identity matrix. We use $\mathbb{R}_{++} = \{x \in \mathbb{R} : x > 0\}$, $\mathbb{R}_+ = \mathbb{R}_{++} \cup \{0\}$, $\mathbb{R}_{+,\infty} = \mathbb{R}_+ \cup \{+\infty\}$, $\mathbb{R}_{[+\infty]} = \mathbb{R} \cup \{+\infty\}$, and $\mathbb{R}_{[\pm\infty]} = \mathbb{R} \cup \{\pm\infty\}$. We equip $\mathbb{R}^u_{[\pm\infty]}$ with the following metric d. For any $x_1, x_2 \in \mathbb{R}^u_{[\pm\infty]}$, $d(x_1,x_2) = \left(\sum_{i=1}^u (G(x_{1,i}) - G(x_{2,i}))^2\right)^{1/2}$, where $G : \mathbb{R}_{[\pm\infty]} \to [0,1]$ is such that $G(-\infty) = 0$, $G(\infty) = 1$, and $G(y) = \Phi(y)$ for $y \in \mathbb{R}$, where Φ is the standard normal CDF. Also, $1\{\cdot\}$ denotes the indicator function.

Let $\mathcal{C}(\Theta^2)$ denote the space of continuous functions that map Θ^2 to Ψ and $\mathcal{S}(\Theta \times \mathbb{R}^k_{[\pm\infty]})$ denote the space of compact subsets of the metric space $(\Theta \times \mathbb{R}^k_{[\pm\infty]}, d)$. In addition, let d_H denote the Hausdorff metric associated with d. We use "$\overset{H}{\to}$" to denote convergence in the Hausdorff metric, i.e., $A_n \overset{H}{\to} B \iff d_H(A_n, B) \to 0$. Finally, for non-stochastic functions of $\theta \in \Theta$, we use "$\overset{u}{\to}$" to denote uniform in θ convergence, e.g., $\Omega_{F_n} \overset{u}{\to} \Omega \iff \sup_{\theta,\theta' \in \Theta} d(\Omega_{F_n}(\theta,\theta'), \Omega(\theta,\theta')) \to 0$. Also, we use $\Omega(\theta)$ and $\Omega(\theta,\theta)$ equivalently.

We denote by $l^\infty(\Theta)$ the set of all uniformly bounded functions that map $\Theta \to \mathbb{R}^u$, equipped with the supremum norm. The sequence of distributions $\{F_n \in \mathcal{P}\}_{n\geq 1}$ determine a sequence of probability spaces $\{(\mathcal{W}, \mathcal{A}, F_n)\}_{n\geq 1}$. Stochastic processes are then random maps $X : \mathcal{W} \to l^\infty(\Theta)$. In this context, we use "$\overset{d}{\to}$", "$\overset{p}{\to}$", and "$\overset{a.s.}{\to}$" to denote weak convergence, convergence in probability, and convergence almost surely in the $l^\infty(\Theta)$ metric, respectively, in the sense of van der Vaart and Wellner (1996). In addition, for every $F \in \mathcal{P}$, we use $\mathcal{M}(F) \equiv \{D_F^{-1/2}(\theta) m(\cdot,\theta) : \mathcal{W} \to \mathbb{R}^k\}$ and denote by ρ_F the coordinate-wise version of the "intrinsic" variance semimetric, i.e.,

$$\rho_F(\theta,\theta') \equiv \left\| \left\{ V_F\left[\sigma_{F,j}^{-1}(\theta) m_j(W,\theta) - \sigma_{F,j}^{-1}(\theta') m_j(W,\theta')\right]^{1/2} \right\}_{j=1}^k \right\| . \tag{A-1}$$

A.1 Algorithm for Test MR

Algorithm A.1 below summarizes the steps required to implement Test MR as defined in Section 2. A few aspects are worth enphasizing. Note that in line 3 a matrix of $n \times B$ of independent $N(0,1)$ is simulated and the same matrix is used to compute $T_n^{R1}(\gamma)$ and $T_n^{R2}(\gamma)$ (lines 25 and 26). The algorithm involves $2B+1$ optimization problems (lines 22, 25, and 26), however solving this problem is typically significantly faster than computing Test BP, which requires a way to compute a test statistic and a quantile for each $\theta \in \Theta$. Relative to subsampling, Test MR does not need to resample from the original data at each $b = 1, \ldots, B$ (line 24), which speeds up computation in our simulations. These computational advantages are even more noticeable when computing a confidence set (as in Remark 2.4).

Algorithm A.1 Algorithm to Implement the Minimum Resampling Test

Inputs: γ, Θ, κ_n, B, $f(\cdot)$, $\varphi(\cdot)$, $m(\cdot)$, $S(\cdot)$, α ▷ We set: $\kappa_n = \sqrt{\ln n}$.
$\Theta(\gamma) \leftarrow \{\theta \in \Theta : f(\theta) = \gamma\}$ ▷ Restriction set
$\zeta \leftarrow n \times B$ matrix of independent $N(0,1)$ ▷ Normal Draws needed for Test MR

function QSTAT(type, θ, $\{W_i\}_{i=1}^n$, $\{\zeta_i\}_{i=1}^n$) ▷ Computes Criterion Function for a given θ
  $\bar{m}_n(\theta) \leftarrow n^{-1}\sum_{i=1}^n m(W_i,\theta)$. ▷ Moments for a given θ
  $\hat{D}_n(\theta) \leftarrow \mathrm{Diag}(\mathrm{var}(m(W_i,\theta)))$. ▷ Variance matrix for a given θ
  $\hat{\Omega}_n(\theta) \leftarrow \mathrm{cor}(m(W_i,\theta))$ ▷ Correlation matrix for a given θ
  $\hat{\sigma}^2_{n,j}(\theta) \leftarrow \hat{D}_n(\theta)[j,j]$ ▷ Variance of the j-th moment for a given θ
  if type=0 **then** ▷ Type 0 is for Test Statistic
    $v(\theta) \leftarrow \sqrt{n}\bar{m}_n(\theta)$ ▷ Scaled average
    $\ell(\theta) \leftarrow \mathbf{0}_{k\times 1}$ ▷ Test Statistic does not involve ℓ
  else if type=1 **then** ▷ Type 1 is for Test R1
    $v(\theta) \leftarrow n^{-1/2}\sum_{i=1}^n (m(W_i,\theta) - \bar{m}_n(\theta))\zeta_i$ ▷ Define Stoch. process
    $\ell(\theta) \leftarrow \varphi(\kappa_n^{-1}\sqrt{n}\hat{D}_n^{-1/2}(\theta)\bar{m}_n(\theta))$
  else if type=2 **then** ▷ Type 2 is for Test R2
    $v(\theta) \leftarrow n^{-1/2}\sum_{i=1}^n (m(W_i,\theta) - \bar{m}_n(\theta))\zeta_i$ ▷ Define Stoch. process
    $\ell(\theta) \leftarrow \kappa_n^{-1}\sqrt{n}\hat{D}_n^{-1/2}(\theta)\bar{m}_n(\theta)$
  end if
  return $Q(\theta) \leftarrow S(v(\theta) + \ell(\theta), \hat{\Omega}_n(\theta))$
end function

function TESTMR(B, $\{W_i\}_{i=1}^n$, ζ, $\Theta(\gamma)$, α) ▷ Test MR
  $T_n \leftarrow \min_{\theta\in\Theta(\gamma)} \mathrm{QSTAT}(0,\theta,\{W_i\}_{i=1}^n)$ ▷ Compute Test Statistic
  $\hat{\Theta}_I(\gamma) \leftarrow \{\theta \in \Theta(\gamma) : \mathrm{QSTAT}(0,\theta,\{W_i\}_{i=1}^n) \leq T_n\}$ ▷ Estimated null identified set
  for b=1,...,B **do**
    $T^{R1}[b] \leftarrow \min_{\theta\in\hat{\Theta}_I(\gamma)} \mathrm{QSTAT}(1,\theta,\{W_i\}_{i=1}^n, \zeta[,b])$ ▷ type=1. Uses bth column of ζ
    $T^{R2}[b] \leftarrow \min_{\theta\in\Theta(\gamma)} \mathrm{QSTAT}(2,\theta,\{W_i\}_{i=1}^n, \zeta[,b])$ ▷ type=2. Uses bth column of ζ
    $T^{MR}[b] \leftarrow \min\{T^{R1}[b], T^{R2}[b]\}$
  end for
  $\hat{c}_n^{MR} \leftarrow \mathrm{QUANTILE}(T^{MR}, 1-\alpha)$ ▷ T^{MR} is $B \times 1$. Gets $1-\alpha$ quantile
  return $\phi^{MR} \leftarrow 1\{T_n > \hat{c}_n^{MR}\}$ ▷ Reject if Test statistic above MR critical value
end function

Appendix B Assumptions

Assumption A.1. For every $F \in \mathcal{P}$ and $j = 1,\ldots,k$, $\{\sigma^{-1}_{F,j}(\theta)m_j(\cdot,\theta) : \mathcal{W} \to \mathbb{R}\}$ is a *measurable* class of functions indexed by $\theta \in \Theta$.

Assumption A.2. The empirical process $v_n(\cdot)$ with j-component

$$v_{n,j}(\theta) = n^{-1/2}\sigma^{-1}_{F,j}(\theta)\sum_{i=1}^n (m_j(W_i,\theta) - \bar{m}_{n,j}(\theta)), \quad j = 1,\ldots,k\ , \tag{B-1}$$

is asymptotically ρ_F-equicontinuous uniformly in $F \in \mathcal{P}$ in the sense of van der Vaart and Wellner (1996, page 169). This is, for any $\varepsilon > 0$,

$$\lim_{\delta\downarrow 0}\limsup_{n\to\infty}\sup_{F\in\mathcal{P}} P_F^*\left(\sup_{\rho_F(\theta,\theta')<\delta} ||v_n(\theta) - v_n(\theta')|| > \varepsilon\right) = 0\ ,$$

where P_F^* denotes outer probability and ρ_F is the coordinate-wise intrinsic variance semimetric in (A-1).

Assumption A.3. For some constant $a > 0$ and all $j = 1, \ldots, k$,

$$\sup_{F \in \mathcal{P}} E_F \left[\sup_{\theta \in \Theta} \left| \frac{m_j(W,\theta)}{\sigma_{F,j}(\theta)} \right|^{2+a} \right] < \infty \ .$$

Assumption A.4. For any $F \in \mathcal{P}$ and $\theta, \theta' \in \Theta$, let $\Omega_F(\theta, \theta')$ be a $k \times k$ correlation matrix with typical $[j_1, j_2]$-component

$$\Omega_F(\theta, \theta')_{[j_1,j_2]} \equiv E_F \left[\left(\tfrac{m_{j_1}(W,\theta) - E_F[m_{j_1}(W,\theta)]}{\sigma_{F,j_1}(\theta)} \right) \left(\tfrac{m_{j_2}(W,\theta') - E_F[m_{j_2}(W,\theta')]}{\sigma_{F,j_2}(\theta')} \right) \right] \ .$$

The matrix Ω_F satisfies

$$\lim_{\delta \downarrow 0} \sup_{\left\|(\theta_1,\theta_1') - (\theta_2,\theta_2')\right\| < \delta} \sup_{F \in \mathcal{P}} \left\| \Omega_F(\theta_1, \theta_1') - \Omega_F(\theta_2, \theta_2') \right\| = 0 \ .$$

Remark B.1. Assumption A.1 is a mild measurability condition. In fact, the kind of uniform laws large numbers we need for our analysis would not hold without this basic requirement (see van der Vaart and Wellner, 1996, page 110). Assumption A.2 is a uniform stochastic equicontinuity assumption which, in combination with the other three assumptions, is used to show that, for all $j = 1, \ldots, k$, the class of functions $\{\sigma_{F,j}^{-1}(\theta) m_j(\cdot, \theta) : \mathcal{W} \to \mathbb{R}\}$ is Donsker and pre-Gaussian uniformly in $F \in \mathcal{P}$ (see Lemma C.1). Assumption A.3 provides a uniform (in F and θ) envelope function that satisfies a uniform integrability condition. This is essential to obtain uniform versions of the laws of large numbers and central limit theorems. Finally, Assumption A.4 requires the correlation matrices to be uniformly equicontinuous, which is used to show pre-Gaussianity.

Assumption A.5. Given the function $\varphi(\cdot)$ in (2.6), there is a function $\varphi^* : \mathbb{R}^k_{[\pm\infty]} \to \mathbb{R}^k_{[+\infty]}$ that takes the form $\varphi^*(\xi) = (\varphi_1^*(\xi_1), \ldots, \varphi_p^*(\xi_p), \mathbf{0}_{k-p})$ and, for all $j = 1, \ldots, p$,

(a) $\varphi_j^*(\xi_j) \geq \varphi_j(\xi_j)$ for all $\xi_j \in \mathbb{R}_{[+\infty]}$.

(b) $\varphi_j^*(\cdot)$ is continuous.

(c) $\varphi_j^*(\xi_j) = 0$ for all $\xi_j \leq 0$ and $\varphi_j^*(\infty) = \infty$.

Remark B.2. Assumption A.5 is satisfied when φ is any of the the functions $\varphi^{(1)} - \varphi^{(4)}$ described in Andrews and Soares (2010) or Andrews and Barwick (2012). This follows from Bugni et al. (2013, Lemma D.8).

Assumption A.6. For any $\{(\gamma_n, F_n) \in \mathcal{L}_0\}_{n \geq 1}$, let (Λ, Ω) be such that $\Omega_{F_n} \overset{u}{\to} \Omega$ and $\Lambda_{n,F_n}(\gamma_n) \overset{H}{\to} \Lambda$ with $(\Omega, \Lambda) \in \mathcal{C}(\theta) \times \mathcal{S}(\Theta \times \mathbb{R}^k_{[\pm\infty]})$ and $\Lambda_{n,F_n}(\gamma_n)$ as in Table 2. Let $c_{(1-\alpha)}(\Lambda, \Omega)$ be the $(1-\alpha)$-quantile of $J(\Lambda, \Omega) \equiv \inf_{(\theta,\ell) \in \Lambda} S(v_\Omega(\theta) + \ell, \Omega(\theta))$. Then,

(a) If $c_{(1-\alpha)}(\Lambda, \Omega) > 0$, the distribution of $J(\Lambda, \Omega)$ is continuous at $c_{(1-\alpha)}(\Lambda, \Omega)$.

(b) If $c_{(1-\alpha)}(\Lambda, \Omega) = 0$, $\liminf_{n \to \infty} P_{F_n}(T_n(\gamma_n) = 0) \geq 1 - \alpha$, where $T_n(\gamma_n)$ is as in (2.3).

Remark B.3. Without Assumption A.6 the asymptotic distribution of the test statistic could be discontinuous at the probability limit of the critical value, resulting in asymptotic over-rejection under the null hypothesis. One could add an infinitesimal constant to the critical value and avoid introducing such assumption, but this introduces an additional tuning parameter that needs to be chosen by the researcher. Note that this assumption holds in Examples 3.1 and 3.2 where $J(\cdot)$ is continuous at $x \in \mathbb{R}$.

Assumption A.7. The following conditions hold.

(a) For all $(\gamma, F) \in \mathcal{L}_0$ and $\theta \in \Theta(\gamma)$, $Q_F(\theta) \geq c \min\{\delta, \inf_{\tilde{\theta} \in \Theta_I(F,\gamma)} ||\theta - \tilde{\theta}||\}^{\chi}$ for constants $c, \delta > 0$ and for χ as in Assumption M.1.

(b) $\Theta(\gamma)$ is convex.

(c) The function $g_F(\theta) \equiv D_F^{-1/2}(\theta) E_F[m(W, \theta)]$ is differentiable in θ for any $F \in \mathcal{P}_0$, and the class of functions $\{G_F(\theta) \equiv \partial g_F(\theta)/\partial \theta' : F \in \mathcal{P}_0\}$ is equicontinuous, i.e.,

$$\lim_{\delta \to 0} \sup_{F \in \mathcal{P}_0, (\theta, \theta') : ||\theta - \theta'|| \leq \delta} ||G_F(\theta) - G_F(\theta')|| = 0 .$$

Remark B.4. Assumption A.7(a) states that $Q_F(\theta)$ can be bounded below in a neighborhood of the null identified set $\Theta_I(F, \gamma)$ and so it is analogous to the polynomial minorant condition in (Chernozhukov et al., 2007, Eqs. (4.1) and (4.5)). The convexity in Assumption A.7(b) would be implied by the parameter space Θ being convex and the function $f(\cdot)$ being linear. Finally, A.7(c) is a smoothness condition that would be implied by the class of functions $\{G_F(\theta) \equiv \partial g_F(\theta)/\partial \theta' : F \in \mathcal{P}_0\}$ being Lipschitz. These three parts are a sufficient conditions for our test to be asymptotically valid (see Lemmas C.7 and C.8). One could create examples in which Assumption A.7 is violated and our test still controls asymptotic size. We however present the results this way as the sufficient conditions in Assumption A.7 are easier to interpret than the conditions that appear in the conclusions of Lemmas C.7 and C.8.

Assumption A.8. The sequences $\{\kappa_n\}_{n \geq 1}$ and $\{b_n\}_{n \geq 1}$ in Assumption M.2 satisfy $\limsup_{n \to \infty} \kappa_n \sqrt{b_n/n} \leq 1$.

Remark B.5. Assumption A.8 is a weaker version of Andrews and Soares (2010, Assumption GMS5) and it holds for all typical choices of κ_n and b_n. For example, it holds if we use the recommended choice of $\kappa_n = \sqrt{\ln n}$ in Andrews and Soares (2010, Page 131) and $b_n = n^c$ for any $c \in (0, 1)$. Note that the latter includes as a special case $b_n = n^{2/3}$, which has been shown to be optimal according to the rate of convergence of the error in the coverage probability (see Politis and Romano, 1994; Bugni, 2010, 2014).

Assumption A.9. For $\gamma_0 \in \Gamma$, there is $\{\gamma_n \in \Gamma\}_{n \geq 1}$ such that $\{(\gamma_n, F_n) \in \mathcal{L}_0\}_{n \geq 1}$ satisfies

(a) For all $n \in \mathbb{N}$, $\Theta_I(F_n) \cap \Theta(\gamma_0) = \emptyset$ (i.e. $(\gamma_0, F_n) \notin \mathcal{L}_0$),

(b) $d_H(\Theta(\gamma_n), \Theta(\gamma_0)) = O(n^{-1/2})$,

(c) For any $\theta \in \Theta$, $\kappa_n^{-1} G_{F_n}(\theta) = o(1)$.

Assumption A.10. For $\gamma_0 \in \Gamma$ and $\{\gamma_n \in \Gamma\}_{n \geq 1}$ as in Assumption A.9, let $(\Omega, \Lambda, \Lambda^{SS}, \Lambda^{R2}) \in \mathcal{C}(\Theta^2) \times \mathcal{S}(\Theta \times \mathbb{R}^k_{[\pm\infty]})^3$ be such that $\Omega_{F_n} \xrightarrow{u} \Omega$, $\Lambda_{n,F_n}(\gamma_0) \xrightarrow{H} \Lambda$, $\Lambda^{R2}_{n,F_n}(\gamma_0) \xrightarrow{H} \Lambda^{R2}$, $\Lambda^{SS}_{b_n,F_n}(\gamma_0) \xrightarrow{H} \Lambda^{SS}$ for $\Lambda_{n,F_n}(\gamma_0)$, $\Lambda^{R2}_{n,F_n}(\gamma_0)$, and $\Lambda^{SS}_{b_n,F_n}(\gamma_0)$ as in Table 2. Then,

(a) The distribution of $J(\Lambda, \Omega)$ is continuous at $c_{1-\alpha}(\Lambda^{SS}, \Omega)$.

(b) The distributions of $J(\Lambda, \Omega)$, $J(\Lambda^{SS}, \Omega)$, and $J(\Lambda^{R2}, \Omega)$ are strictly increasing at $x > 0$.

Assumption A.11. For $\gamma_0 \in \Gamma$, there is $\{\gamma_n \in \Gamma\}_{n \geq 1}$ such that $\{(\gamma_n, F_n) \in \mathcal{L}_0\}_{n \geq 1}$ satisfies

(a) The conditions in Assumption A.9.

(b) There are (possibly random) sequences $\{\tilde{\theta}_n \in \Theta_I(F_n)\}_{n \geq 1}$ and $\{\hat{\theta}_n \in \Theta(\gamma_0)\}_{n \geq 1}$ such that,

  i. $S(\sqrt{n}\bar{m}_n(\hat{\theta}_n), \hat{\Sigma}_n(\hat{\theta}_n)) - T_n(\gamma_0) = o_p(1)$.

  ii. $\lambda_n \equiv \sqrt{n}\left(D_{F_n}^{-1/2}(\hat{\theta}_n) E_{F_n}[m(W, \hat{\theta}_n)] - D_{F_n}^{-1/2}(\tilde{\theta}_n) E_{F_n}[m(W, \tilde{\theta}_n)]\right) \to \lambda \in \mathbb{R}^k$.

iii. $\sqrt{n}D_{F_n}^{-1/2}(\tilde{\theta}_n)E_{F_n}[m(W,\tilde{\theta}_n)] \xrightarrow{p} (h,\mathbf{0}_{k-p})$ with $h\in\mathbb{R}^p_{[+\infty]}$, and $\kappa_n^{-1}\sqrt{n}D_{F_n}^{-1/2}(\tilde{\theta}_n)E_{F_n}[m(W,\tilde{\theta}_n)] \xrightarrow{p} (\pi,\mathbf{0}_{k-p})$ with $\pi\in\mathbb{R}^p_{[+\infty]}$. Also, $\tilde{\theta}_n-\theta^*=o_p(1)$ for some $\theta^*\in\Theta$.

(c) There are (possibly random) sequences $\{\tilde{\theta}_n^{SS}\in\Theta_I(F_n)\}_{n\geq1}$ and $\{\hat{\theta}_n^{SS}\in\Theta(\gamma_0)\}_{n\geq1}$ such that,

i. Conditionally on $\{W_i\}_{i=1}^n$, $S(\sqrt{b_n}\bar{m}_{b_n}^{SS}(\hat{\theta}_n^{SS}),\hat{\Sigma}_{b_n}^{SS}(\hat{\theta}_n^{SS}))-T_n^{SS}(\gamma_0)=o_p(1)$ a.s.

ii. $\lambda_n^{SS}\equiv\sqrt{b_n}\left(D_{F_n}^{-1/2}(\hat{\theta}_n^{SS})E_{F_n}[m(W,\hat{\theta}_n^{SS})]-D_{F_n}^{-1/2}(\tilde{\theta}_n^{SS})E_{F_n}[m(W,\tilde{\theta}_n^{SS})]\right)\to\mathbf{0}_k$.

iii. $\sqrt{b_n}D_{F_n}^{-1/2}(\tilde{\theta}_n^{SS})E_{F_n}[m(W,\tilde{\theta}_n^{SS})] \xrightarrow{p} (g,\mathbf{0}_{k-p})$ with $g\in\mathbb{R}^p_{[+\infty]}$. Also, conditionally on $\{W_i\}_{i=1}^n$, $\hat{\theta}_n^{SS}-\theta^*=o_p(1)$ a.s., where θ^* is as in part (i).

(d) $g_j<\pi_j$ for some $j=1,\ldots,p$.

(e) $\lambda_j<-h_j$ for some $j\leq p$ or $|\lambda_j|\neq0$ for some $j>p$.

The literature routinely assumes that the function $S(\cdot)$ in (2.1) satisfies the following assumptions (see, e.g., Andrews and Soares (2010), Andrews and Guggenberger (2009), and Bugni et al. (2012)). We therefore treat the assumptions below as maintained. We note in particular that the constant χ in Assumption M.1 equals 2 when the function $S(\cdot)$ is either the modified methods of moments in (3.14) or the quasi-likelihood ratio.

Assumption M.1. For some $\chi>0$, $S(am,\Omega)=a^{\chi}S(m,\Omega)$ for all scalars $a>0$, $m\in\mathbb{R}^k$, and $\Omega\in\Psi$.

Assumption M.2. The sequence $\{\kappa_n\}_{n\geq1}$ satisfies $\kappa_n\to\infty$ and $\kappa_n/\sqrt{n}\to0$. The sequence $\{b_n\}_{n\geq1}$ satisfies $b_n\to\infty$ and $b_n/n\to0$.

Assumption M.3. For each $\gamma\in\Gamma$, $\Theta(\gamma)$ is a nonempty and compact subset of $\mathbb{R}^{d_\theta}$ ($d_\theta<\infty$).

Assumption M.4. Test BP is computed using the GMS approach in Andrews and Soares (2010). This is, $\phi_n^{BP}(\cdot)$ in (3.5) is based on $CS_n(1-\alpha)=\{\theta\in\Theta: Q_n(\theta)\leq\hat{c}_n(\theta,1-\alpha)\}$ where $\hat{c}_n(\theta,1-\alpha)$ is the GMS critical value constructed using the GMS function $\varphi(\cdot)$ in (2.6) and thresholding sequence $\{\kappa_n\}_{n\geq1}$ satisfying Assumption M.2.

Assumption M.5. The function $S(\cdot)$ satisfies the following conditions.

(a) $S((m_1,m_2),\Sigma)$ is non-increasing in m_1, for all $(m_1,m_2)\in\mathbb{R}^p_{[+\infty]}\times\mathbb{R}^{k-p}$ and all variance matrices $\Sigma\in\mathbb{R}^{k\times k}$.

(b) $S(m,\Sigma)=S(\Delta m,\Delta\Sigma\Delta)$ for all $m\in\mathbb{R}^k$, $\Sigma\in\mathbb{R}^{k\times k}$, and positive definite diagonal $\Delta\in\mathbb{R}^{k\times k}$.

(c) $S(m,\Omega)\geq0$ for all $m\in\mathbb{R}^k$ and $\Omega\in\Psi$,

(d) $S(m,\Omega)$ is continuous at all $m\in\mathbb{R}^k_{[\pm\infty]}$ and $\Omega\in\Psi$.

Assumption M.6. For all $h_1\in\mathbb{R}^p_{[+\infty]}\times\mathbb{R}^{k-p}$, all $\Omega\in\Psi$, and $Z\sim N(\mathbf{0}_k,\Omega)$, the distribution function of $S(Z+h_1,\Omega)$ at $x\in\mathbb{R}$

(a) is continuous for $x>0$,

(b) is strictly increasing for $x>0$ unless $p=k$ and $h_1=\infty^p$,

(c) is less than or equal to 1/2 at $x=0$ when $k>p$ or when $k=p$ and $h_{1,j}=0$ for some $j=1,\ldots,p$.

(d) is degenerate at $x=0$ when $p=k$ and $h_1=\infty^p$.

(e) satisfies $P(S(Z+(m_1,\mathbf{0}_{k-p}),\Omega)\leq x)<P(S(Z+(m_1^*,\mathbf{0}_{k-p}),\Omega)\leq x)$ for all $x>0$ and all $m_1,m_1^*\in\mathbb{R}^p_{[+\infty]}$ with $m_{1,j}\leq m_{1,j}^*$ for all $j=1,\ldots,p$ and $m_{1,j}<m_{1,j}^*$ for some $j=1,\ldots,p$.

Assumption M.7. The function $S(\cdot)$ satisfies the following conditions.

(a) The distribution function of $S(Z,\Omega)$ is continuous at its $(1-\alpha)$-quantile, denoted $c_{(1-\alpha)}(\Omega)$, for all $\Omega \in \Psi$, where $Z \sim N(\mathbf{0}_k, \Omega)$ and $\alpha \in (0, 0.5)$,

(b) $c_{(1-\alpha)}(\Omega)$ is continuous in Ω uniformly for $\Omega \in \Psi$.

Assumption M.8. $S(m,\Omega) > 0$ if and only if $m_j < 0$ for some $j = 1,\ldots,p$ or $m_j \neq 0$ for some $j = p+1,\ldots,k$, where $m = (m_1,\ldots,m_k)'$ and $\Omega \in \Psi$. Equivalently, $S(m,\Omega) = 0$ if and only if $m_j \geq 0$ for all $j = 1,\ldots,p$ and $m_j = 0$ for all $j = p+1,\ldots,k$, where $m = (m_1,\ldots,m_k)'$ and $\Omega \in \Psi$.

Assumption M.9. For all $n \geq 1$, $S(\sqrt{n}\bar{m}_n(\theta), \hat{\Sigma}(\theta))$ is a lower semi-continuous function of $\theta \in \Theta$.

Appendix C Auxiliary results

C.1 Auxiliary theorems

Theorem C.1. *Suppose Assumptions A.1-A.4 hold. Let $\Lambda^{R2}_{n,F}(\gamma)$ be as in Table 2 and $T^{R2}_n(\gamma)$ be as in (2.12). Let $\{(\gamma_n, F_n) \in \mathcal{L}_0\}_{n\geq 1}$ be a (sub)sequence of parameters such that for some $(\Omega, \Lambda^{R2}) \in \mathcal{C}(\Theta^2) \times \mathcal{S}(\Theta \times \mathbb{R}^k_{[\pm\infty]})$: (i) $\Omega_{F_n} \overset{u}{\to} \Omega$ and (ii) $\Lambda^{R2}_{n,F_n}(\gamma_n) \overset{H}{\to} \Lambda^{R2}$. Then, there exists a further subsequence $\{u_n\}_{n\geq 1}$ of $\{n\}_{n\geq 1}$ such that, along $\{F_{u_n}\}_{n\geq 1}$,*

$$\{T^{R2}_{u_n}(\gamma_{u_n}) | \{W_i\}^n_{i=1}\} \overset{d}{\to} J(\Lambda^{R2}, \Omega) \equiv \inf_{(\theta,\ell)\in\Lambda^{R2}} S(v_\Omega(\theta) + \ell, \Omega(\theta)), \text{ a.s. },$$

where $v_\Omega : \Theta \to \mathbb{R}^k$ is a tight Gaussian process with covariance (correlation) kernel Ω.

Theorem C.2. *Suppose Assumptions A.1-A.5 hold. Let $\Lambda^{R2}_{n,F}(\gamma)$ and $\Lambda^{R1}_{n,F}(\gamma)$ be as in Table 2, and let $T^{R2}_n(\gamma)$ and $\tilde{T}^{R1}_n(\gamma)$ be as in (2.12) and*

$$\tilde{T}^{R1}_n(\gamma) \equiv \inf_{\theta\in\Theta^{\delta_n}_I(\gamma)} S(v^*_n(\theta) + \varphi^*(\kappa_n^{-1}\sqrt{n}\hat{D}_n^{-1/2}(\theta)\bar{m}_n(\theta)), \hat{\Omega}_n(\theta)) , \tag{C-1}$$

where $v^*_n(\theta)$ is as in (2.8), $\varphi^*(\cdot)$ is as in Assumption A.5, and $\Theta^{\delta_n}_I(\gamma)$ is as in Table 2. Let $\{(\gamma_n, F_n) \in \mathcal{L}_0\}_{n\geq 1}$ be a (sub)sequence of parameters such that for some $(\Omega, \Lambda^{R1}, \Lambda^{R2}) \in \mathcal{C}(\Theta^2) \times \mathcal{S}(\Theta \times \mathbb{R}^k_{[\pm\infty]})^2$: (i) $\Omega_{F_n} \overset{u}{\to} \Omega$, (ii) $\Lambda^{R1}_{n,F_n}(\gamma_n) \overset{H}{\to} \Lambda^{R1}$, and (iii) $\Lambda^{R2}_{n,F_n}(\gamma_n) \overset{H}{\to} \Lambda^{R2}$. Then, there exists a further subsequence $\{u_n\}_{n\geq 1}$ of $\{n\}_{n\geq 1}$ such that, along $\{F_{u_n}\}_{n\geq 1}$,

$$\{\min\{\tilde{T}^{R1}_{u_n}(\gamma_{u_n}), T^{R2}_{u_n}(\gamma_{u_n})\} | \{W_i\}^n_{i=1}\} \overset{d}{\to} J(\Lambda^{MR}, \Omega) \equiv \inf_{(\theta,\ell)\in\Lambda^{MR}} S(v_\Omega(\theta) + \ell, \Omega(\theta)), \text{ a.s. },$$

where $v_\Omega : \Theta \to \mathbb{R}^k$ is a tight Gaussian process with covariance (correlation) kernel Ω,

$$\Lambda^{MR} \equiv \Lambda^{R1}_* \cup \Lambda^{R2} \quad \text{and} \quad \Lambda^{R1}_* \equiv \{(\theta,\ell) \in \Theta \times \mathbb{R}^k_{[\pm\infty]} : \ell = \varphi^*(\ell') \text{ for some } (\theta,\ell') \in \Lambda^{R1}\} . \tag{C-2}$$

Theorem C.3. *Suppose Assumptions A.1-A.4 hold. Let $\Lambda^{SS}_{b_n,F}(\gamma)$ be as in Table 2 and $T^{SS}_{b_n}(\gamma)$ be as in (3.8). Let $\{(\gamma_n, F_n) \in \mathcal{L}_0\}_{n\geq 1}$ be a (sub)sequence of parameters such that for some $(\Omega, \Lambda^{SS}) \in \mathcal{C}(\Theta^2) \times \mathcal{S}(\Theta \times \mathbb{R}^k_{[\pm\infty]})$: (i) $\Omega_{F_n} \overset{u}{\to} \Omega$ and (ii) $\Lambda^{SS}_{b_n,F_n}(\gamma_n) \overset{H}{\to} \Lambda^{SS}$. Then, there exists a further subsequence $\{u_n\}_{n\geq 1}$ of $\{n\}_{n\geq 1}$ such that, along $\{F_{u_n}\}_{n\geq 1}$,*

$$\{T^{SS}_{u_n}(\gamma_{u_n}) | \{W_i\}^n_{i=1}\} \overset{d}{\to} J(\Lambda^{SS}, \Omega) \equiv \inf_{(\theta,\ell)\in\Lambda^{SS}} S(v_\Omega(\theta) + \ell, \Omega(\theta,\theta)), \text{ a.s. },$$

where $v_\Omega : \Theta \to \mathbb{R}^k$ is a tight Gaussian process with covariance (correlation) kernel Ω.

Theorem C.4. *Suppose Assumptions A.1-A.4 hold. Let $\Lambda_{n,F}(\gamma)$ be as in Table 2 and $T_n(\gamma)$ be as in (2.3). Let $\{(\gamma_n, F_n) \in \mathcal{L}_0\}_{n\geq 1}$ be a (sub)sequence of parameters such that for some $(\Omega, \Lambda) \in \mathcal{C}(\Theta^2) \times \mathcal{S}(\Theta \times \mathbb{R}^k_{[\pm\infty]})$: (i) $\Omega_{F_n} \overset{u}{\to} \Omega$*

and (ii) $\Lambda_{n,F_n}(\gamma_n) \overset{H}{\to} \Lambda$. Then, there exists a further subsequence $\{u_n\}_{n\geq 1}$ of $\{n\}_{n\geq 1}$ such that, along $\{F_{u_n}\}_{n\geq 1}$,

$$T_{u_n}(\gamma_{u_n}) \overset{d}{\to} J(\Lambda, \Omega) \equiv \inf_{(\theta,\ell)\in\Lambda} S(v_\Omega(\theta) + \ell, \Omega(\theta)), \text{ as } n \to \infty ,$$

where $v_\Omega : \Theta \to \mathbb{R}^k$ is a tight Gaussian process with zero-mean and covariance (correlation) kernel Ω.

C.2 Auxiliary lemmas

Lemma C.1. *Suppose Assumptions A.1-A.4 hold. Let $\{F_n \in \mathcal{P}\}_{n\geq 1}$ be a (sub)sequence of distributions s.t. $\Omega_{F_n} \overset{u}{\to} \Omega$ for some $\Omega \in \mathcal{C}(\Theta^2)$. Then, the following results hold:*

1. $v_n \overset{d}{\to} v_\Omega$ in $l^\infty(\Theta)$, where $v_\Omega : \Theta \to \mathbb{R}^k$ is a tight zero-mean Gaussian process with covariance (correlation) kernel Ω. In addition, v_Ω is a uniformly continuous function, a.s.
2. $\tilde{\Omega}_n \overset{p}{\to} \Omega$ in $l^\infty(\Theta)$.
3. $D_{F_n}^{-1/2}(\cdot)\hat{D}_n^{1/2}(\cdot) - I_k \overset{p}{\to} \boldsymbol{0}_k$ in $l^\infty(\Theta)$.
4. $\hat{D}_n^{-1/2}(\cdot)D_{F_n}^{1/2}(\cdot) - I_k \overset{p}{\to} \boldsymbol{0}_k$ in $l^\infty(\Theta)$.
5. $\hat{\Omega}_n \overset{p}{\to} \Omega$ in $l^\infty(\Theta)$.
6. For any arbitrary sequence $\{\lambda_n \in \mathbb{R}_{++}\}_{n\geq 1}$ s.t. $\lambda_n \to \infty$, $\lambda_n^{-1} v_n \overset{p}{\to} \boldsymbol{0}_k$ in $l^\infty(\Theta)$.
7. For any arbitrary sequence $\{\lambda_n \in \mathbb{R}_{++}\}_{n\geq 1}$ s.t. $\lambda_n \to \infty$, $\lambda_n^{-1} \tilde{v}_n \overset{p}{\to} \boldsymbol{0}_k$ in $l^\infty(\Theta)$.
8. $\{v_n^* | \{W_i\}_{i=1}^n\} \overset{d}{\to} v_\Omega$ in $l^\infty(\Theta)$ a.s., where v_Ω is the tight Gaussian process described in part 1.
9. $\{\tilde{v}_{b_n}^{SS} | \{W_i\}_{i=1}^n\} \overset{d}{\to} v_\Omega$ in $l^\infty(\Theta)$ a.s., where

$$\tilde{v}_{b_n}^{SS}(\theta) \equiv \frac{1}{\sqrt{b_n}} \sum_{i=1}^{b_n} D_{F_n}^{-1/2}(\theta)(m(W_i^{SS}, \theta) - \bar{m}_n(\theta)) , \tag{C-3}$$

   $\{W_i^{SS}\}_{i=1}^{b_n}$ is a subsample of size b_n from $\{W_i\}_{i=1}^n$, and v_Ω is the tight Gaussian process described in part 1.
10. For $\tilde{\Omega}_{b_n}^{SS}(\theta) \equiv D_{F_n}^{-1/2}(\theta)\hat{\Sigma}_{b_n}^{SS}(\theta)D_{F_n}^{-1/2}(\theta)$, $\{\tilde{\Omega}_{b_n}^{SS} | \{W_i\}_{i=1}^n\} \overset{p}{\to} \Omega$ in $l^\infty(\Theta)$ a.s.

Lemma C.2. *Let Assumptions A.1-A.4 hold. Then, for any sequence $\{(\gamma_n, F_n) \in \mathcal{L}\}_{n\geq 1}$ there exists a subsequence $\{u_n\}_{n\geq 1}$ of $\{n\}_{n\geq 1}$ s.t. $\Omega_{F_{u_n}} \overset{u}{\to} \Omega$, $\Lambda_{u_n,F_{u_n}}(\gamma_{u_n}) \overset{H}{\to} \Lambda$, $\Lambda_{u_n,F_{u_n}}^{R2}(\gamma_{u_n}) \overset{H}{\to} \Lambda^{R2}$, and $\Lambda_{u_n,F_{u_n}}^{R1}(\gamma_{u_n}) \overset{H}{\to} \Lambda^{R1}$, for some $(\Omega, \Lambda, \Lambda^{R1}, \Lambda^{R2}) \in \mathcal{C}(\theta) \times \mathcal{S}(\Theta \times \mathbb{R}_{[\pm\infty]}^k)^3$, where $\Lambda_{n,F_n}(\gamma)$, $\Lambda_{n,F_n}^{R1}(\gamma)$, and $\Lambda_{n,F_n}^{R2}(\gamma)$ are defined in Table 2.*

Lemma C.3. *Let $\{F_n \in \mathcal{P}\}_{n\geq 1}$ be an arbitrary (sub)sequence of distributions and let $X_n(\theta) : \Omega \to l^\infty(\Theta)$ be any stochastic process such that $X_n \overset{p}{\to} 0$ in $l^\infty(\Theta)$. Then, there exists a subsequence $\{u_n\}_{n\geq 1}$ of $\{n\}_{n\geq 1}$ such that $X_{u_n} \overset{a.s.}{\to} 0$ in $l^\infty(\Theta)$.*

Lemma C.4. *Let the set A be defined as follows:*

$$A \equiv \left\{ x \in \mathbb{R}_{[+\infty]}^p \times \mathbb{R}^{k-p} : \max\left\{ \max_{j=1,\dots,p} \{[x_j]_-\}, \max_{s=p+1,\dots,k} \{|x_s|\} \right\} = 1 \right\} . \tag{C-4}$$

Then, $\inf_{(x,\Omega)\in A\times\Psi} S(x, \Omega) > 0$.

Lemma C.5. *If $S(x, \Omega) \leq 1$ then there exist a constant $\varpi > 0$ such that $x_j \geq -\varpi$ for all $j \leq p$ and $|x_j| \leq \varpi$ for all $j > p$.*

Lemma C.6. *The function S satisfies the following properties: (i) $x \in (-\infty, \infty]^p \times \mathbb{R}^{k-p}$ implies $\sup_{\Omega\in\Psi} S(x, \Omega) < \infty$, (ii) $x \notin (-\infty, \infty]^p \times \mathbb{R}^{k-p}$ implies $\inf_{\Omega\in\Psi} S(x, \Omega) = \infty$.*

Lemma C.7. *Let $(\Omega, \Lambda, \Lambda^{R1}) \in \mathcal{C}(\Theta^2) \times \mathcal{S}(\Theta \times \mathbb{R}_{[\pm\infty]}^k)^2$ be such that $\Omega_{F_n} \overset{u}{\to} \Omega$, $\Lambda_{n,F_n}(\gamma_n) \overset{H}{\to} \Lambda$, and $\Lambda_{n,F_n}^{R1}(\gamma_n) \overset{H}{\to} \Lambda^{R1}$, for some $\{(\gamma_n, F_n) \in \mathcal{L}_0\}_{n\geq 1}$. Then, Assumptions A.5 and A.7 imply that for all $(\theta, \ell) \in \Lambda^{R1}$ there exists $(\theta, \tilde{\ell}) \in \Lambda$ with $\tilde{\ell}_j \geq \varphi_j^*(\ell_j)$ for $j \leq p$ and $\tilde{\ell}_j = \ell_j \equiv 0$ for $j > p$, where $\varphi^*(\cdot)$ is defined in Assumption A.5.*

Lemma C.8. *Let $(\Omega,\Lambda,\Lambda^{R2})\in\mathcal{C}(\Theta^2)\times\mathcal{S}(\Theta\times\mathbb{R}^k_{[\pm\infty]})^2$ be such that $\Omega_{F_n}\overset{u}{\to}\Omega$, $\Lambda_{n,F_n}(\gamma_n)\overset{H}{\to}\Lambda$, and $\Lambda^{R2}_{n,F_n}(\gamma_n)\overset{H}{\to}\Lambda^{R2}$, for some $\{(\gamma_n,F_n)\in\mathcal{L}_0\}_{n\geq1}$. Then, Assumption A.7 implies that for all $(\theta,\ell)\in\Lambda^{R2}$ with $\ell\in\mathbb{R}^p_{[+\infty]}\times\mathbb{R}^{k-p}$, there exists $(\theta,\tilde{\ell})\in\Lambda$ with $\tilde{\ell}_j\geq\ell_j$ for $j\leq p$ and $\tilde{\ell}_j=\ell_j$ for $j>p$.*

Lemma C.9. *Let Assumptions A.1-A.4 and A.7-A.9 hold. For $\gamma_0\in\Gamma$ and $\{\gamma_n\in\Gamma\}_{n\geq1}$ as in Assumption A.9, assume that $\Omega_{F_n}\overset{u}{\to}\Omega$, $\Lambda_{n,F_n}(\gamma_0)\overset{H}{\to}\Lambda$, $\Lambda^{R2}_{n,F_n}(\gamma_0)\overset{H}{\to}\Lambda^{R2}$, $\Lambda^{SS}_{b_n,F_n}(\gamma_0)\overset{H}{\to}\Lambda^{SS}$, $\Lambda^{R2}_{n,F_n}(\gamma_n)\overset{H}{\to}\Lambda^{R2}_A$, and $\Lambda^{SS}_{b_n,F_n}(\gamma_n)\overset{H}{\to}\Lambda^{SS}_A$ for some $(\Omega,\Lambda,\Lambda^{SS},\Lambda^{R2},\Lambda^{SS}_A,\Lambda^{R2}_A)\in\mathcal{C}(\Theta^2)\times\mathcal{S}(\Theta\times\mathbb{R}^k_{[\pm\infty]})^5$. Then,*

$$c_{(1-\alpha)}(\Lambda^{R2},\Omega)\leq c_{(1-\alpha)}(\Lambda^{SS},\Omega)\ .$$

Lemma C.10. *Let Assumptions A.1-A.4 and A.7-A.11 hold. Then,*

$$\liminf_{n\to\infty}\ (E_{F_n}[\phi^{R2}_n(\gamma_0)]-E_{F_n}[\phi^{SS}_n(\gamma_0)])\ >\ 0\ .$$

Appendix D Proofs

D.1 Proofs of the main theorems

Proof of Theorem 2.1. We divide the proof in six steps and show that for $\eta\geq0$,

$$\limsup_{n\to\infty}\sup_{(\gamma,F)\in\mathcal{L}_0}P_F(T_n(\gamma)>\hat{c}^{MR}_n(\gamma,1-\alpha)+\eta)\leq\alpha\ .$$

Steps 1-4 hold for $\eta\geq0$, step 5 needs $\eta>0$, and step 6 holds for $\eta=0$ under Assumption A.6.

Step 1. For any $(\gamma,F)\in\mathcal{L}_0$, let $\tilde{T}^{R1}_n(\gamma)$ be as in (C-1) and $\tilde{c}^{MR}_n(\gamma,1-\alpha)$ be the conditional $(1-\alpha)$-quantile of $\min\{\tilde{T}^{R1}_n(\gamma),T^{R2}_n(\gamma)\}$. Consider the following derivation

$$\begin{aligned}P_F(T_n(\gamma)>\hat{c}^{MR}_n(\gamma,1-\alpha)+\eta)&\leq P_F(T_n(\gamma)>\tilde{c}^{MR}_n(\gamma,1-\alpha)+\eta)+P_F(\hat{c}^{MR}_n(\gamma,1-\alpha)<\tilde{c}^{MR}_n(\gamma,1-\alpha))\\&\leq P_F(T_n(\gamma)>\tilde{c}^{MR}_n(\gamma,1-\alpha)+\eta)+P_F(\hat{\Theta}_I(\gamma)\not\subseteq\Theta_I^{\ln\kappa_n}(\gamma))\ ,\end{aligned}$$

where the second inequality follows from the fact that Assumption A.5 and $\hat{c}^{MR}_n(\gamma,1-\alpha)<\tilde{c}^{MR}_n(\gamma,1-\alpha)$ imply that $\hat{\Theta}_I(\gamma)\not\subseteq\Theta_I^{\ln\kappa_n}(\gamma)$. By this and Lemma D.13 in Bugni et al. (2013) (with a redefined parameter space equal to $\Theta(\gamma)$), it follows that

$$\limsup_{n\to\infty}\sup_{(\gamma,F)\in\mathcal{L}_0}P_F(T_n(\gamma)>\hat{c}^{MR}_n(\gamma,1-\alpha)+\eta)\leq\limsup_{n\to\infty}\sup_{(\gamma,F)\in\mathcal{L}_0}P_F(T_n(\gamma)>\tilde{c}^{MR}_n(\gamma,1-\alpha)+\eta)\ .$$

Step 2. By definition, there exists a subsequence $\{a_n\}_{n\geq1}$ of $\{n\}_{n\geq1}$ and a subsequence $\{(\gamma_{a_n},F_{a_n})\}_{n\geq1}$ s.t.

$$\limsup_{n\to\infty}\sup_{(\gamma,F)\in\mathcal{L}_0}P_F(T_n(\gamma)>\hat{c}^{MR}_n(\gamma,1-\alpha)+\eta)=\lim_{n\to\infty}P_{F_{a_n}}(T_{a_n}(\gamma_{a_n})>\tilde{c}^{MR}_{a_n}(\gamma_{a_n},1-\alpha)+\eta)\ .\qquad\text{(D-1)}$$

By Lemma C.2, there is a further sequence $\{u_n\}_{n\geq1}$ of $\{a_n\}_{n\geq1}$ s.t. $\Omega_{F_{u_n}}\overset{u}{\to}\Omega$, $\Lambda_{u_n,F_{u_n}}(\gamma_{u_n})\overset{H}{\to}\Lambda$, $\Lambda^{R1}_{u_n,F_{u_n}}(\gamma_{u_n})\overset{H}{\to}\Lambda^{R1}$, and $\Lambda^{R2}_{u_n,F_{u_n}}(\gamma_{u_n})\overset{H}{\to}\Lambda^{R2}$, for some $(\Omega,\Lambda,\Lambda^{R1},\Lambda^{R2})\in\mathcal{C}(\theta)\times\mathcal{S}(\Theta\times\mathbb{R}^k_{[\pm\infty]})^3$. Since $\Omega_{F_{u_n}}\overset{u}{\to}\Omega$ and $\Lambda_{u_n,F_{u_n}}(\gamma_{u_n})\overset{H}{\to}\Lambda$, Theorem C.4 implies that $T_{u_n}(\gamma_{u_n})\overset{d}{\to}J(\Lambda,\Omega)\equiv\inf_{(\theta,\ell)\in\Lambda}S(v_\Omega(\theta)+\ell,\Omega(\theta))$. Similarly, Theorem C.2 implies that $\{\min\{\tilde{T}^{R1}_{u_n}(\gamma_{u_n}),T^{R2}_{u_n}(\gamma_{u_n})\}|\{W_i\}_{i=1}^{u_n}\}\overset{d}{\to}J(\Lambda^{MR},\Omega)$ a.s.

Step 3. We show that $J(\Lambda^{MR},\Omega)\geq J(\Lambda,\Omega)$. Suppose not, i.e., $\exists(\theta,\ell)\in\Lambda^{R1}_*\cup\Lambda^{R2}$ s.t. $S(v_\Omega(\theta)+\ell,\Omega(\theta))<J(\Lambda,\Omega)$. If $(\theta,\ell)\in\Lambda^{R1}_*$ then by definition $\exists(\theta,\ell')\in\Lambda^{R1}$ s.t. $\varphi^*(\ell')=\ell$ and $S(v_\Omega(\theta)+\varphi^*(\ell'),\Omega(\theta))<J(\Lambda,\Omega)$. By Lemma C.7, $\exists(\theta,\tilde{\ell})\in\Lambda$ where $\tilde{\ell}_j\geq\varphi^*_j(\ell'_j)$ for $j\leq p$ and $\tilde{\ell}_j=0$ for $j>p$. Thus

$$S(v_\Omega(\theta)+\tilde{\ell},\Omega(\theta))\leq S(v_\Omega(\theta)+\varphi^*(\ell'),\Omega(\theta))<J(\Lambda,\Omega)\equiv\inf_{(\theta,\ell)\in\Lambda}S(v_\Omega(\theta)+\ell,\Omega(\theta))\ ,$$

which is a contradiction to $(\theta, \tilde{\ell}) \in \Lambda$. If $(\theta, \ell) \in \Lambda^{R2}$, we first need to show that $\ell \in \mathbb{R}^p_{[+\infty]} \times \mathbb{R}^{k-p}$. Suppose not, i.e., suppose that $\ell_j = -\infty$ for some $j \leq p$ or $|\ell_j| = \infty$ for some $j > p$. Since $v_\Omega : \Theta \to \mathbb{R}^k$ is a tight Gaussian process, it follows that $v_{\Omega,j}(\theta) + \ell_j = -\infty$ for some $j \leq p$ or $|v_{\Omega,j}(\theta) + \ell_j| = \infty$ for some $j > p$. By Lemma C.6, we have $S(v_\Omega(\theta) + \ell, \Omega(\theta)) = \infty$ which contradicts $S(v_\Omega(\theta) + \ell, \Omega(\theta)) < J(\Lambda, \Omega)$. Since $\ell \in \mathbb{R}^p_{[+\infty]} \times \mathbb{R}^{k-p}$, Lemma C.8 implies that $\exists(\theta, \tilde{\ell}) \in \Lambda$ where $\tilde{\ell}_j \geq \ell_j$ for $j \leq p$ and $\tilde{\ell}_j = \ell_j$ for $j > p$. We conclude that

$$S(v_\Omega(\theta) + \tilde{\ell}, \Omega(\theta)) \leq S(v_\Omega(\theta) + \ell, \Omega(\theta)) < J(\Lambda, \Omega) \equiv \inf_{(\theta,\ell)\in\Lambda} S(v_\Omega(\theta) + \ell, \Omega(\theta)) \ ,$$

which is a contradiction to $(\theta, \tilde{\ell}) \in \Lambda$.

Step 4. We now show that for $c_{(1-\alpha)}(\Lambda, \Omega)$ being the $(1-\alpha)$-quantile of $J(\Lambda, \Omega)$ and any $\varepsilon > 0$,

$$\lim P_{F_{u_n}}(\tilde{c}^{MR}_{u_n}(\gamma_{u_n}, 1-\alpha) \leq c_{(1-\alpha)}(\Lambda, \Omega) - \varepsilon) = 0 \ . \tag{D-2}$$

Let $\varepsilon > 0$ be s.t. $c_{(1-\alpha)}(\Lambda, \Omega) - \varepsilon$ is a continuity point of the CDF of $J(\Lambda, \Omega)$. Then,

$$\begin{aligned} \lim P_{F_{u_n}} \left(\min\{\tilde{T}^{R1}_{u_n}(\gamma_{u_n}), T^{R2}_{u_n}(\gamma_{u_n})\} \leq c_{(1-\alpha)}(\Lambda, \Omega) - \varepsilon \middle| \{W_i\}_{i=1}^{u_n} \right) &= P\left(J(\Lambda^{MR}, \Omega) \leq c_{(1-\alpha)}(\Lambda, \Omega) - \varepsilon\right) \\ &\leq P\left(J(\Lambda, \Omega) \leq c_{(1-\alpha)}(\Lambda, \Omega) - \varepsilon\right) < 1 - \alpha \ , \end{aligned}$$

where the first equality holds because $\{\min\{\tilde{T}^{R1}_{u_n}(\gamma_{u_n}), T^{R2}_{u_n}(\gamma_{u_n})\} | \{W_i\}_{i=1}^{u_n}\} \stackrel{d}{\to} J(\Lambda^{MR}, \Omega)$ a.s., the second weak inequality is a consequence of $J(\Lambda^{MR}, \Omega) \geq J(\Lambda, \Omega)$, and the final strict inequality follows from $c_{(1-\alpha)}(\Lambda, \Omega)$ being the $(1-\alpha)$-quantile of $J(\Lambda, \Omega)$. Next, notice that

$$\left\{ \lim P_{F_{u_n}} \left(\min\{\tilde{T}^{R1}_{u_n}(\gamma_{u_n}), T^{R2}_{u_n}(\gamma_{u_n})\} \leq c_{(1-\alpha)}(\Lambda, \Omega) - \varepsilon \middle| \{W_i\}_{i=1}^{u_n} \right) < 1 - \alpha \right\} \subseteq \left\{ \liminf \{\tilde{c}^{MR}_{u_n}(1-\alpha) > c_{(1-\alpha)}(\Lambda, \Omega) - \varepsilon\} \right\} \ .$$

Since the RHS occurs a.s., then the LHS must also occur a.s. Then, (D-2) is a consequence of this and Fatou's Lemma.

Step 5. For $\eta > 0$, we can define $\varepsilon > 0$ in step 4 so that $\eta - \varepsilon > 0$ and $c_{(1-\alpha)}(\Lambda, \Omega) + \eta - \varepsilon$ is a continuity point of the CDF of $J(\Lambda, \Omega)$. It then follows that

$$\begin{aligned} P_{F_{u_n}} \left(T_{u_n}(\gamma_{u_n}) > \tilde{c}^{MR}_{u_n}(\gamma_{u_n}, 1-\alpha) + \eta \right) \leq {} & P_{F_{u_n}} \left(\tilde{c}^{MR}_{u_n}(\gamma_{u_n}, 1-\alpha) \leq c_{(1-\alpha)}(\Lambda, \Omega) - \varepsilon \right) \\ & + 1 - P_{F_{u_n}} \left(T_{u_n}(\gamma_{u_n}) \leq c_{(1-\alpha)}(\Lambda, \Omega) + \eta - \varepsilon \right) \ . \end{aligned} \tag{D-3}$$

Taking limit supremum on both sides, using steps 2 and 4, and that $\eta - \varepsilon > 0$,

$$\limsup_{n\to\infty} P_{F_{u_n}} \left(T_{u_n}(\gamma_{u_n}) > \hat{c}^{MR}_{u_n}(\gamma_{u_n}, 1-\alpha) + \eta \right) \leq 1 - P\left(J(\Lambda, \Omega) \leq c_{(1-\alpha)}(\Lambda, \Omega) + \eta - \varepsilon\right) \leq \alpha \ .$$

This combined with steps 1 and 2 completes the proof under $\eta > 0$.

Step 6. For $\eta = 0$, there are two cases to consider. First, suppose $c_{(1-\alpha)}(\Lambda, \Omega) = 0$. Then, by Assumption A.6,

$$\limsup_{n\to\infty} P_{F_{u_n}}(T_{u_n}(\gamma_{u_n}) > \tilde{c}^{MR}_{u_n}(\gamma_{u_n}, 1-\alpha)) \leq \limsup_{n\to\infty} P_{F_{u_n}}(T_{u_n}(\gamma_{u_n}) \neq 0) \leq \alpha \ .$$

The proof is completed by combining the previous equation with steps 1 and 2. Second, suppose $c_{(1-\alpha)}(\Lambda, \Omega) > 0$. Consider a sequence $\{\varepsilon_m\}_{m\geq 1}$ s.t. $\varepsilon_m \downarrow 0$ and $c_{(1-\alpha)}(\Lambda, \Omega) - \varepsilon_m$ is a continuity point of the CDF of $J(\Lambda, \Omega)$ for all $m \in \mathbb{N}$. For any $m \in \mathbb{N}$, it follows from (D-3) and steps 2 and 3 that

$$\limsup_{n\to\infty} P_{F_{u_n}}(T_{u_n}(\gamma_{u_n}) > \tilde{c}^{MR}_{u_n}(\gamma_{u_n}, 1-\alpha)) \leq 1 - P(J(\Lambda, \Omega) \leq c_{(1-\alpha)}(\Lambda, \Omega) - \varepsilon_m) \ .$$

Taking $\varepsilon_m \downarrow 0$ and using continuity gives the RHS equal to α. Combining the previous equation with steps 1 and 2 completes the proof. □

Proof of Theorem 3.1. This proof follows identical steps to those in the proof of Bugni et al. (2013, Theorem 6.1) and is therefore omitted. ☐

Proof of Theorem 3.2. Suppose not, i.e., suppose that $\liminf(E_{F_n}[\phi_n^{R2}(\gamma_0)] - E_{F_n}[\phi_n^{SS}(\gamma_0)]) \equiv -\delta < 0$. Consider a subsequence $\{k_n\}_{n\geq 1}$ of $\{n\}_{n\geq 1}$ such that,

$$P_{F_{k_n}}(T_{k_n}(\gamma_0) > c_{k_n}^{R2}(\gamma_0, 1-\alpha)) = E_{F_{k_n}}[\phi_{k_n}^{R2}(\gamma_0)] < E_{F_{k_n}}[\phi_{k_n}^{SS}(\gamma_0)] - \delta/2 = P_{F_{k_n}}(T_{k_n}(\gamma_0) > c_{k_n}^{SS}(\gamma_0, 1-\alpha)) - \delta/2 \ ,$$

or, equivalently,

$$P_{F_{k_n}}(T_{k_n}(\gamma_0) \leq c_{k_n}^{SS}(\gamma_0, 1-\alpha)) + \delta/2 < P_{F_{k_n}}(T_{k_n}(\gamma_0) \leq c_{k_n}^{R2}(\gamma_0, 1-\alpha)) \ . \tag{D-4}$$

Lemma C.2 implies that for some $(\Omega, \Lambda, \Lambda^{R2}, \Lambda^{SS}, \Lambda_A^{R2}, \Lambda_A^{SS}) \in \mathcal{C}(\Theta^2) \times \mathcal{S}(\Theta \times \mathbb{R}_{[\pm\infty]}^k)^5$, $\Omega_{F_{k_n}} \stackrel{u}{\to} \Omega$, $\Lambda_{k_n,F_{k_n}}^{R2}(\gamma_0) \stackrel{H}{\to} \Lambda^{R2}$, $\Lambda_{b_{k_n},F_{k_n}}^{SS}(\gamma_0) \stackrel{H}{\to} \Lambda^{SS}$, $\Lambda_{k_n,F_{k_n}}^{R2}(\gamma_{k_n}) \stackrel{H}{\to} \Lambda_A^{R2}$, and $\Lambda_{b_{k_n},F_{k_n}}^{SS}(\gamma_{k_n}) \stackrel{H}{\to} \Lambda_A^{SS}$. Then, Theorems C.4, C.1, and C.3 imply that $T_{k_n}(\gamma_0) \stackrel{d}{\to} J(\Lambda, \Omega)$, $\{T_{k_n}^{R2}(\gamma_0)|\{W_i\}_{i=1}^{k_n}\} \stackrel{d}{\to} J(\Lambda^{R2}, \Omega)$ a.s., and $\{T_{k_n}^{SS}(\gamma_0)|\{W_i\}_{i=1}^{k_n}\} \stackrel{d}{\to} J(\Lambda^{SS}, \Omega)$ a.s.

We next show that $c_{k_n}^{R2}(\gamma_0, 1-\alpha) \stackrel{a.s.}{\to} c_{1-\alpha}(\Lambda^{R2}, \Omega)$. Let $\varepsilon > 0$ be arbitrary and pick $\tilde{\varepsilon} \in (0, \varepsilon)$ s.t. $c_{(1-\alpha)}(\Lambda^{R2}, \Omega) + \tilde{\varepsilon}$ and $c_{(1-\alpha)}(\Lambda^{R2}, \Omega) - \tilde{\varepsilon}$ are both a continuity points of the CDF of $J(\Lambda^{R2}, \Omega)$. Then,

$$\lim_{n\to\infty} P_{F_{k_n}}(T_{k_n}^{R2}(\gamma_0) \leq c_{(1-\alpha)}(\Lambda^{R2}, \Omega) + \tilde{\varepsilon}|\{W_i\}_{i=1}^n) = P(J(\Lambda^{R2}, \Omega) \leq c_{(1-\alpha)}(\Lambda^{R2}, \Omega) + \tilde{\varepsilon}) > 1-\alpha \quad a.s. \ , \tag{D-5}$$

where the first equality holds because of $\{T_{k_n}^{R2}(\gamma_0)|\{W_i\}_{i=1}^{k_n}\} \stackrel{d}{\to} J(\Lambda^{R2}, \Omega)$ a.s., and the strict inequality is due to $\tilde{\varepsilon} > 0$ and $c_{(1-\alpha)}(\Lambda^{R2}, \Omega) + \tilde{\varepsilon}$ being a continuity point of the CDF of $J(\Lambda^{R2}, \Omega)$. Similarly,

$$\lim_{n\to\infty} P_{F_{k_n}}(T_{k_n}^{R2}(\gamma_0) \leq c_{(1-\alpha)}(\Lambda^{R2}, \Omega) - \tilde{\varepsilon}|\{W_i\}_{i=1}^n) = P(J(\Lambda^{R2}, \Omega) \leq c_{(1-\alpha)}(\Lambda^{R2}, \Omega) - \tilde{\varepsilon}) < 1-\alpha \ . \tag{D-6}$$

Next, notice that,

$$\{\lim_{n\to\infty} P_{F_{k_n}}(T_{k_n}^{R2}(\gamma_0) \leq c_{(1-\alpha)}(\Lambda^{R2}, \Omega)+\tilde{\varepsilon}|\{W_i\}_{i=1}^n) > 1-\alpha\} \subseteq \{\liminf_{n\to\infty}\{c_{k_n}^{R2}(\gamma_0, 1-\alpha) < c_{(1-\alpha)}(\Lambda^{R2}, \Omega)+\tilde{\varepsilon}\}\} \ , \tag{D-7}$$

with the same result holding with $-\tilde{\varepsilon}$ replacing $\tilde{\varepsilon}$. From (D-5), (D-6), (D-7), we conclude that

$$P_{Fn}(\liminf_{n\to\infty}\{|c_{k_n}^{R2}(\gamma_0, 1-\alpha) - c_{(1-\alpha)}(\Lambda^{R2}, \Omega)| \leq \varepsilon\}) = 1 \ ,$$

which is equivalent to $c_{k_n}^{R2}(\gamma_0, 1-\alpha) \stackrel{a.s.}{\to} c_{(1-\alpha)}(\Lambda^{R2}, \Omega)$. By similar arguments, $c_{k_n}^{SS}(\gamma_0, 1-\alpha) \stackrel{a.s.}{\to} c_{(1-\alpha)}(\Lambda^{SS}, \Omega)$.

Let $\varepsilon > 0$ be s.t. $c_{(1-\alpha)}(\Lambda^{SS}, \Omega) - \varepsilon$ is a continuity point of the CDF of $J(\Lambda, \Omega)$ and note that

$$\begin{aligned} P_{F_{k_n}}(T_{k_n}(\gamma_0) \leq c_{k_n}^{SS}(\gamma_0, 1-\alpha)) \geq\ & P_{F_{k_n}}(\{T_{k_n}(\gamma_0) \leq c_{(1-\alpha)}(\Lambda^{SS}, \Omega) - \varepsilon\} \cap \{c_{k_n}^{SS}(\gamma_0, 1-\alpha) \geq c_{(1-\alpha)}(\Lambda^{SS}, \Omega) - \varepsilon\}) \\ &+ P_{F_{k_n}}(\{T_{k_n}(\gamma_0) \leq c_{k_n}^{SS}(\gamma_0, 1-\alpha)\} \cap \{c_{k_n}^{SS}(\gamma_0, 1-\alpha) < c_{(1-\alpha)}(\Lambda^{SS}, \Omega) - \varepsilon\}) \ . \end{aligned}$$

Taking lim inf and using that $T_{k_n}(\gamma_0) \stackrel{d}{\to} J(\Lambda, \Omega)$ and $c_{k_n}^{SS}(\gamma_0, 1-\alpha) \stackrel{a.s.}{\to} c_{(1-\alpha)}(\Lambda^{SS}, \Omega)$, we deduce that

$$\liminf_{n\to\infty} P_{F_{k_n}}(T_{k_n}(\gamma_0) \leq c_{k_n}^{SS}(\gamma_0, 1-\alpha)) \geq P(J(\Lambda, \Omega) \leq c_{(1-\alpha)}(\Lambda^{SS}, \Omega) - \varepsilon) \ . \tag{D-8}$$

Fix $\varepsilon > 0$ arbitrarily and pick $\tilde{\varepsilon} \in (0, \varepsilon)$ s.t. $c_{(1-\alpha)}(\Lambda^{R2}, \Omega) + \tilde{\varepsilon}$ is a continuity point of the CDF of $J(\Lambda, \Omega)$. Then,

$$P_{F_{k_n}}(T_{k_n}(\gamma_0) \leq c_{k_n}^{R2}(\gamma_0, 1-\alpha)) \leq P_{F_{k_n}}(T_{k_n}(\gamma_0) \leq c_{(1-\alpha)}(\Lambda^{R2}, \Omega) + \tilde{\varepsilon}) + P_{F_{k_n}}(c_{k_n}^{R2}(\gamma_0, 1-\alpha) > c_{(1-\alpha)}(\Lambda^{R2}, \Omega) + \tilde{\varepsilon}) \ .$$

Taking lim sup on both sides, and using that $T_{k_n}(\gamma_0) \stackrel{d}{\to} J(\Lambda, \Omega)$, $c_{k_n}^{R2}(\gamma_0, 1-\alpha) \stackrel{a.s.}{\to} c_{(1-\alpha)}(\Lambda^{R2}, \Omega)$, and $\tilde{\varepsilon} \in (0, \varepsilon)$,

$$\limsup_{n\to\infty} P_{F_{k_n}}(T_{k_n}(\gamma_0) \leq c_{k_n}^{R2}(\gamma_0, 1-\alpha)) \leq P(J(\Lambda, \Omega) \leq c_{(1-\alpha)}(\Lambda^{R2}, \Omega) + \tilde{\varepsilon}) \ . \tag{D-9}$$

Next consider the following derivation

$$\begin{aligned}P(J(\Lambda,\Omega)\leq c_{(1-\alpha)}(\Lambda^{SS},\Omega)-\varepsilon)+\delta/2 &\leq \liminf P_{F_{k_n}}(T_{k_n}(\gamma_0)\leq c^{SS}_{k_n}(\gamma_0,1-\alpha))+\delta/2\\ &\leq \limsup P_{F_{k_n}}(T_{k_n}(\gamma_0)\leq c^{R2}_{k_n}(\gamma_0,1-\alpha))\\ &\leq P(J(\Lambda,\Omega)\leq c_{(1-\alpha)}(\Lambda^{R2},\Omega)+\varepsilon)\\ &\leq P(J(\Lambda,\Omega)\leq c_{(1-\alpha)}(\Lambda^{SS},\Omega)+\varepsilon)\ ,\end{aligned}$$

where the first inequality follows from (D-8), the second inequality follows from (D-4), the third inequality follows from (D-9), and the fourth inequality follows from $c_{(1-\alpha)}(\Lambda^{R2},\Omega)\leq c_{(1-\alpha)}(\Lambda^{SS},\Omega)$ by Lemma C.9. We conclude that

$$P(J(\Lambda,\Omega)\leq c_{(1-\alpha)}(\Lambda^{SS},\Omega)+\varepsilon)-P(J(\Lambda,\Omega)\leq c_{(1-\alpha)}(\Lambda^{SS},\Omega)-\varepsilon)\ \geq\ \delta/2\ >\ 0\ .$$

Taking $\varepsilon\downarrow 0$ and using Assumption A.10, the LHS converges to zero, which is a contradiction. □

D.2 Proofs of theorems in Appendix C

Proof of Theorem C.1. Step 1. To simplify expressions, let $\Lambda^{R2}_n\equiv\Lambda^{R2}_{n,F_n}(\gamma_n)$. Consider the following derivation,

$$\begin{aligned}T^{R2}_n(\gamma_n)&=\inf_{\theta\in\Theta(\gamma_n)}S\left(v^*_n(\theta)+\mu_{n,1}(\theta)+\mu_{n,2}(\theta)'\kappa_n^{-1}\sqrt{n}D^{-1/2}_{F_n}(\theta)E_{F_n}[m(W,\theta)],\hat{\Omega}_n(\theta)\right)\\ &=\inf_{(\theta,\ell)\in\Lambda^{R2}_n}S\left(v^*_n(\theta)+\mu_{n,1}(\theta)+\mu_{n,2}(\theta)'\ell,\hat{\Omega}_n(\theta)\right)\ ,\end{aligned}$$

where $\mu_n(\theta)=(\mu_{n,1}(\theta),\mu_{n,2}(\theta))$, $\mu_{n,1}(\theta)\equiv\kappa_n^{-1}\tilde{v}_n(\theta)$, $\mu_{n,2}(\theta)\equiv\{\hat{\sigma}^{-1}_{n,j}(\theta)\sigma_{F_n,j}(\theta)\}^k_{j=1}$, and $\tilde{v}_n(\theta)\equiv\sqrt{n}\hat{D}^{-1}_n(\theta)(\bar{m}_n(\theta)-E_F[m(W,\theta)])$. Note that $\hat{D}^{-1/2}_n(\theta)$ and $D^{1/2}_{F_n}(\theta)$ are both diagonal matrices.

Step 2. We now show that there is a subsequence $\{a_n\}_{n\geq1}$ of $\{n\}_{n\geq1}$ s.t. $\{(v^*_{a_n},\mu_{a_n},\hat{\Omega}_{a_n})|\{W_i\}^{a_n}_{i=1}\}\overset{d}{\to}(v_\Omega,(\mathbf{0}_k,\mathbf{1}_k),\Omega)$ in $l^\infty(\theta)$ a.s. By part 8 in Lemma C.1, $\{v^*_n|\{W_i\}^n_{i=1}\}\overset{d}{\to}v_\Omega$ in $l^\infty(\theta)$. Then the result would follow from finding a subsequence $\{a_n\}_{n\geq1}$ of $\{n\}_{n\geq1}$ s.t. $\{(\mu_{a_n},\hat{\Omega}_{a_n})|\{W_i\}^{a_n}_{i=1}\}\to((\mathbf{0}_k,\mathbf{1}_k),\Omega)$ in $l^\infty(\theta)$ a.s. Since $(\mu_n,\hat{\Omega}_n)$ is conditionally non-random, this is equivalent to finding a subsequence $\{a_n\}_{n\geq1}$ of $\{n\}_{n\geq1}$ s.t. $(\mu_{a_n},\hat{\Omega}_{a_n})\overset{a.s.}{\to}((\mathbf{0}_k,\mathbf{1}_k),\Omega)$ in $l^\infty(\theta)$. In turn, this follows from step 1, part 5 of Lemma C.1, and Lemma C.3.

Step 3. Since $\Theta_I(F_n,\gamma_n)\neq\emptyset$, there is a sequence $\{\theta_n\in\Theta(\gamma_n)\}_{n\geq1}$ s.t. for $\ell_{n,j}\equiv\kappa_n^{-1}\sqrt{n}\sigma^{-1}_{F_n,j}(\theta_n)E_{F_n}[m_j(W,\theta_n)]$,

$$\limsup_{n\to\infty}\ell_{n,j}\equiv\bar{\ell}_j\geq0,\quad\text{for }j\leq p,\qquad\text{and}\quad\lim_{n\to\infty}|\ell_{n,j}|\equiv\bar{\ell}_j=0,\quad\text{for }j>p\ .\tag{D-10}$$

By compactness of $(\Theta\times\mathbb{R}^k_{[\pm\infty]},d)$, there is a subsequence $\{k_n\}_{n\geq1}$ of $\{a_n\}_{n\geq1}$ s.t. $d((\theta_{k_n},\ell_{k_n}),(\bar{\theta},\bar{\ell}))\to0$ for some $(\bar{\theta},\bar{\ell})\in\Theta\times\mathbb{R}^p_{+,\infty}\times\mathbf{0}_{k-p}$. By step 2, $\lim(v_{k_n}(\theta_{k_n}),\mu_{k_n}(\theta_{k_n}),\Omega_{k_n}(\theta_{k_n}))=(v_\Omega(\bar{\theta}),(\mathbf{0}_k,\mathbf{1}_k),\Omega(\bar{\theta}))$, and so

$$T^{R2}_{k_n}(\gamma_{k_n})\leq S(v_{k_n}(\theta_{k_n})+\mu_{k_n,1}(\theta_{k_n})+\mu_{k_n,2}(\theta_{k_n})'\ell_{k_n},\Omega_{k_n}(\theta_{k_n}))\to S(v_\Omega(\bar{\theta})+\bar{\ell},\Omega(\bar{\theta}))\ ,\tag{D-11}$$

where the convergence occurs because by the continuity of $S(\cdot)$ and the convergence of its argument. Since $(v_\Omega(\bar{\theta})+\bar{\ell},\Omega(\bar{\theta}))\in\mathbb{R}^p_{[+\infty]}\times\mathbb{R}^{k-p}\times\Psi$, we conclude that $S(v_\Omega(\bar{\theta})+\bar{\ell},\Omega(\bar{\theta}))$ is bounded.

Step 4. Let $\mathcal{D}$ denote the space of functions that map Θ onto $\mathbb{R}^k\times\Psi$ and let $\mathcal{D}_0$ be the space of uniformly continuous functions that map Θ onto $\mathbb{R}^k\times\Psi$. Let the sequence of functionals $\{g_n\}_{n\geq1}$ with $g_n:\mathcal{D}\to\mathbb{R}$ given by

$$g_n(v(\cdot),\mu(\cdot),\Omega(\cdot))\equiv\inf_{(\theta,\ell)\in\Lambda^{R2}_n}S(v(\theta)+\mu_1(\theta)+\mu_2(\theta)'\ell,\Omega(\theta))\ .\tag{D-12}$$

Let the functional $g:\mathcal{D}_0\to\mathbb{R}$ be defined by

$$g(v(\cdot),\mu(\cdot),\Omega(\cdot))\equiv\inf_{(\theta,\ell)\in\Lambda^{R2}}S(v(\theta)+\mu_1(\theta)+\mu_2(\theta)'\ell,\Omega(\theta))\ .$$

We now show that if the sequence of (deterministic) functions $\{(v_n(\cdot), \mu_n(\cdot), \Omega_n(\cdot)) \in \mathcal{D}\}_{n\geq 1}$ satisfies

$$\lim_{n\to\infty} \sup_{\theta\in\Theta} ||(v_n(\theta), \mu_n(\theta), \Omega_n(\theta)) - (v(\theta), (\mathbf{0}_k, \mathbf{1}_k), \Omega(\theta))|| = 0 \ , \tag{D-13}$$

for some $(v(\cdot), \Omega(\cdot)) \in \mathcal{D}_0$, then $\lim_{n\to\infty} g_n(v_n(\cdot), \mu_n(\cdot), \Omega_n(\cdot)) = g(v(\cdot), (\mathbf{0}_k, \mathbf{1}_k), \Omega(\cdot))$. To prove this we show that $\liminf_{n\to\infty} g_n(v_n(\cdot), \mu_n(\cdot), \Omega_n(\cdot)) \geq g(v(\cdot), (\mathbf{0}_k, \mathbf{1}_k), \Omega(\cdot))$. Showing the reverse inequality for the lim sup is similar and therefore omitted. Suppose not, i.e., suppose that $\exists \delta > 0$ and a subsequence $\{a_n\}_{n\geq 1}$ of $\{n\}_{n\geq 1}$ s.t. $\forall n \in \mathbb{N}$,

$$g_{a_n}(v_{a_n}(\cdot), \mu_{a_n}(\cdot), \Omega_{a_n}(\cdot)) < g(v(\cdot), (\mathbf{0}_k, \mathbf{1}_k), \Omega(\cdot)) - \delta \ . \tag{D-14}$$

By definition, $\exists \ \{(\theta_{a_n}, \ell_{a_n})\}_{n\geq 1}$ that approximates the infimum in (D-12), i.e., $\forall n \in \mathbb{N}$, $(\theta_{a_n}, \ell_{a_n}) \in \Lambda^{R2}_{a_n}$ and

$$|g_{a_n}(v_{a_n}(\cdot), \mu_{a_n}(\cdot), \Omega_{a_n}(\cdot)) - S(v_{a_n}(\theta_{a_n}) + \mu_1(\theta_{a_n}) + \mu_2(\theta_{a_n})'\ell_{a_n}, \Omega_{a_n}(\theta_{a_n}))| \leq \delta/2 \ . \tag{D-15}$$

Since $\Lambda^{R2}_{a_n} \subseteq \Theta \times \mathbb{R}^k_{[\pm\infty]}$ and $(\Theta \times \mathbb{R}^k_{[\pm\infty]}, d)$ is a compact metric space, there exists a subsequence $\{u_n\}_{n\geq 1}$ of $\{a_n\}_{n\geq 1}$ and $(\theta^*, \ell^*) \in \Theta \times \mathbb{R}^k_{[\pm\infty]}$ s.t. $d((\theta_{u_n}, \ell_{u_n}), (\theta^*, \ell^*)) \to 0$. We first show that $(\theta^*, \ell^*) \in \Lambda^{R2}$. Suppose not, i.e. $(\theta^*, \ell^*) \notin \Lambda^{R2}$, and consider the following argument

$$\begin{aligned} d((\theta_{u_n}, \ell_{u_n}), (\theta^*, \ell^*)) + d_H(\Lambda^{R2}_{u_n}, \Lambda^{R2}) \quad &\geq \quad d((\theta_{u_n}, \ell_{u_n}), (\theta^*, \ell^*)) + \inf_{(\theta,\ell)\in\Lambda^{R2}} d((\theta, \ell), (\theta_{u_n}, \ell_{u_n})) \\ &\geq \quad \inf_{(\theta,\ell)\in\Lambda^{R2}} d((\theta, \ell), (\theta^*, \ell^*)) > 0 \ , \end{aligned}$$

where the first inequality follows from the definition of Hausdorff distance and the fact that $(\theta_{u_n}, \ell_{u_n}) \in \Lambda^{R2}_{u_n}$, and the second inequality follows by the triangular inequality. The final strict inequality follows from the fact that $\Lambda^{R2} \in \mathcal{S}(\Theta \times \mathbb{R}_{[\pm\infty]})$, i.e., it is a compact subset of $(\Theta \times \mathbb{R}^k_{[\pm\infty]}, d)$, $d((\theta, \ell), (\theta^*, \ell^*))$ is a continuous real-valued function, and Royden (1988, Theorem 7.18). Taking limits as $n \to \infty$ and using that $d((\theta_{u_n}, \ell_{u_n}), (\theta^*, \ell^*)) \to 0$ and $\Lambda^{R2}_{u_n} \overset{H}{\to} \Lambda^{R2}$, we reach a contradiction.

We now show that $\ell^* \in \mathbb{R}^p_{[+\infty]} \times \mathbb{R}^{k-p}$. Suppose not, i.e., suppose that $\exists j = 1, \ldots, k$ s.t. $l^*_j = -\infty$ or $\exists j > p$ s.t. $\ell^*_j = \infty$. Let J denote the set of indices $j = 1, \ldots, k$ s.t. this occurs. For any $\ell \in \mathbb{R}^k_{[\pm\infty]}$ define $\Xi(\ell) \equiv \max_{j\in J} ||\ell_j||$. By definition of $\Lambda^{R2}_{u_n,F_{u_n}}$, $\ell_{u_n} \in \mathbb{R}^k$ and thus, $\Xi(\ell_{u_n}) < \infty$. By the case under consideration, $\lim \Xi(\ell_{u_n}) = \Xi(\ell^*) = \infty$. Since $(\Theta, ||\cdot||)$ is a compact metric space, $d((\theta_{u_n}, \ell_{u_n}), (\theta^*, \ell^*)) \to 0$ implies that $\theta_{u_n} \to \theta^*$. Then,

$$\begin{aligned} &||(v_{u_n}(\theta_{u_n}), \mu_{u_n}(\theta_{u_n}), \Omega_{u_n}(\theta_{u_n})) - (v(\theta^*), (\mathbf{0}_k, \mathbf{1}_k), \Omega(\theta^*))|| \\ &\leq ||(v_{u_n}(\theta_{u_n}), \mu_{u_n}(\theta_{u_n}), \Omega_{u_n}(\theta_{u_n})) - (v(\theta_{u_n}), (\mathbf{0}_k, \mathbf{1}_k), \Omega(\theta_{u_n}))|| + ||(v(\theta_{u_n}), \Omega(\theta_{u_n})) - (v(\theta^*), \Omega(\theta^*))|| \\ &\leq \sup_{\theta\in\Theta} ||(v_{u_n}(\theta), \mu_{u_n}(\theta), \Omega_{u_n}(\theta)) - (v(\theta), (\mathbf{0}_k, \mathbf{1}_k), \Omega(\theta))|| + ||(v(\theta_{u_n}), \Omega(\theta_{u_n})) - (v(\theta^*), \Omega(\theta^*))|| \to 0 \ , \end{aligned}$$

where the last convergence holds by (D-13), $\theta_{u_n} \to \theta^*$, and $(v(\cdot), \Omega(\cdot)) \in \mathcal{D}_0$.

Since $(v(\cdot), \Omega(\cdot)) \in \mathcal{D}_0$, the compactness of Θ implies that $(v(\theta^*), \Omega(\theta^*)$ is bounded. Since $\lim \Xi(\ell_{u_n}) = \Xi(\ell^*) = \infty$ and $\lim v_{u_n}(\theta_{u_n}) = v(\theta^*) \in \mathbb{R}^k$, it then follows that $\lim \Xi(\ell_{u_n})^{-1}||v_{u_n}(\theta_{u_n})|| = 0$. By construction, $\{\Xi(\ell_{u_n})^{-1}\ell_{u_n}\}_{n\geq 1}$ is s.t. $\lim \Xi(\ell_{u_n})^{-1} [\ell_{u_n,j}]_- = 1$ for some $j \leq p$ or $\lim \Xi(\ell_{u_n})^{-1} |\ell_{u_n,j}| = 1$ for some $j > p$. By this, it follows that $\{\Xi(\ell_{u_n})^{-1}(v_{u_n}(\theta_{u_n}) + \ell_{u_n}), \Omega_{u_n}(\theta_{u_n})\}_{n\geq 1}$ with $\lim \Omega_{u_n}(\theta_{u_n}) = \Omega(\theta^*) \in \Psi$ and $\lim \Xi(\ell_{u_n})^{-1}[v_{u_n,j}(\theta_{u_n}) + \ell_{u_n,j}]_- = 1$ for some $j \leq p$ or $\lim \Xi(h_{u_n})^{-1} |v_{u_n,j}(\theta_{u_n}) + \ell_{u_n,j}| = 1$ for some $j > p$. This implies that,

$$S(v_{u_n}(\theta_{u_n}) + \ell_{u_n}, \Omega_{u_n}(\theta_{u_n})) = \Xi(\ell_{u_n})^\chi S(\Xi(\ell_{u_n})^{-1}(v_{u_n}(\theta_{u_n}) + \ell_{u_n}), \Omega_{u_n}(\theta_{u_n})) \to \infty \ .$$

Since $\{(\theta_{u_n}, \ell_{u_n})\}_{n\geq 1}$ is a subsequence of $\{(\theta_{a_n}, \ell_{a_n})\}_{n\geq 1}$ that approximately achieves the infimum in (D-12),

$$g_n(v_n(\cdot), \mu_n(\cdot), \Sigma_n(\cdot)) \to \infty \ . \tag{D-16}$$

However, (D-16) violates step 3 and is therefore a contradiction.

We then know that $d((\theta_{a_n}, \ell_{a_n}), (\theta^*, \ell^*)) \to 0$ with $\ell^* \in \mathbb{R}^p_{[+\infty]} \times \mathbb{R}^{k-p}$. By repeating previous arguments, we conclude that $\lim(v_{u_n}(\theta_{u_n}), \mu_{u_n}(\theta_{u_n}), \Omega_{u_n}(\theta_{u_n})) = (v(\theta^*), (\mathbf{0}_k, \mathbf{1}_k), \Omega(\theta^*)) \in \mathbb{R}^k \times \Psi$. This implies that $\lim(v_{u_n}(\theta_{u_n}) + \mu_{u_n,1}(\theta_{u_n}) + \mu_{u_n,2}(\theta_{u_n})'\ell_{u_n}, \Omega_{u_n}(\theta_{u_n})) = (v(\theta^*) + \ell^*, \Omega(\theta^*)) \in (\mathbb{R}^k_{[\pm\infty]} \times \Psi)$, i.e., $\exists N \in \mathbb{N}$ s.t. $\forall n \geq N$,

$$||S(v_{u_n}(\theta_{u_n}) + \mu_{u_n,1}(\theta_{u_n}) + \mu_{u_n,2}(\theta_{u_n})'\ell_{u_n}, \Omega_{u_n}(\theta_{u_n})) - S(v(\theta^*) + \ell^*, \Omega(\theta^*))|| \leq \delta/2 \; . \tag{D-17}$$

By combining (D-15), (D-17), and the fact that $(\theta^*, \ell^*) \in \Lambda^{R2}$, it follows that $\exists N \in \mathbb{N}$ s.t. $\forall n \geq N$,

$$g_{u_n}(v_{u_n}(\cdot), \mu_{u_n}(\cdot), \Omega_{u_n}(\cdot)) \geq S(v_\Omega(\theta^*) + \ell^*, \Omega(\theta^*)) - \delta \geq g(v(\cdot), (\mathbf{0}_k, \mathbf{1}_k), \Omega(\cdot)) - \delta \; ,$$

which is a contradiction to (D-14).

Step 5. The proof is completed by combining the representation in step 1, the convergence result in step 2, the continuity result in step 4, and the extended continuous mapping theorem (see, e.g., van der Vaart and Wellner, 1996, Theorem 1.11.1). In order to apply this result, it is important to notice that parts 1 and 5 in Lemma C.1 and standard convergence results imply that $(v(\cdot), \Omega(\cdot)) \in \mathcal{D}_0$ a.s. □

Proof of Theorem C.2. Step 1. To simplify expressions let $\Lambda_n^{R2} \equiv \Lambda_{n,F_n}^{R2}(\gamma_n)$, $\Lambda_n^{R1} \equiv \Lambda_{n,F_n}^{R1}(\gamma_n)$, and consider the following derivation,

$$\begin{aligned}
&\min\{\tilde{T}_n^{R1}(\gamma_n), T_n^{R2}(\gamma_n)\} \\
&= \min\left\{ \inf_{\theta \in \Theta_I^{\ln \kappa_n}(\gamma_n)} S(v_n^*(\theta) + \varphi^*(\kappa_n^{-1}\sqrt{n}\hat{D}_n^{-1/2}(\theta)\bar{m}_n(\theta)), \hat{\Omega}_n(\theta)), \inf_{\theta \in \Theta(\gamma_n)} S(v_n^*(\theta) + \kappa_n^{-1}\sqrt{n}\hat{D}_n^{-1/2}(\theta)\bar{m}_n(\theta), \hat{\Omega}_n(\theta)) \right\} \\
&= \min\left\{ \begin{array}{c} \inf_{\theta \in \Theta_I^{\ln \kappa_n}(\gamma_n)} S(v_n^*(\theta) + \varphi^*(\mu_{n,1}(\theta) + \mu_{n,2}(\theta)'\kappa_n^{-1}D_{F_n}^{-1/2}(\theta)\sqrt{n}(E_{F_n}m(W,\theta))), \hat{\Omega}_n(\theta)), \\ \inf_{\theta \in \Theta(\gamma_n)} S(v_n^*(\theta) + \mu_{n,1}(\theta) + \mu_{n,2}(\theta)'\kappa_n^{-1}D_{F_n}^{-1/2}(\theta)\sqrt{n}(E_{F_n}m(W,\theta)), \hat{\Omega}_n(\theta)) \end{array} \right\} \\
&= \min\left\{ \inf_{(\theta,\ell) \in \Lambda_n^{R1}} S(v_n^*(\theta) + \varphi^*(\mu_{n,1}(\theta) + \mu_{n,2}(\theta)'\ell), \hat{\Omega}_n(\theta)), \inf_{(\theta,\ell) \in \Lambda_n^{R2}} S(v_n^*(\theta) + \mu_{n,1}(\theta) + \mu_{n,2}(\theta)'\ell, \hat{\Omega}_n(\theta)) \right\}
\end{aligned}$$

where $\mu_n(\theta) \equiv (\mu_{n,1}(\theta), \mu_{n,2}(\theta))$, $\mu_{n,1}(\theta) \equiv \kappa_n^{-1}\hat{D}_n^{-1/2}(\theta)\sqrt{n}(\bar{m}_n(\theta) - E_{F_n}m(W,\theta)) \equiv \kappa_n^{-1}\tilde{v}_n(\theta)$, and $\mu_{n,2}(\theta) \equiv \{\sigma_{n,j}^{-1}(\theta)\sigma_{F_n,j}(\theta)\}_{j=1}^k$. Note that we used that $D_{F_n}^{-1/2}(\theta)$ and $\hat{D}_n^{-1/2}(\theta)$ are both diagonal matrices.

Step 2. There is a subsequence $\{a_n\}_{n\geq 1}$ of $\{n\}_{n\geq 1}$ s.t. $\{(\hat{v}_{a_n}^*, \mu_{a_n}, \hat{\Omega}_{a_n})|\{W_i\}_{i=1}^{a_n}\} \to^d (v_\Omega, (\mathbf{0}_k, \mathbf{1}_k), \Omega)$ in $l^\infty(\Theta)$ a.s. This step is identical to Step 2 in the proof of Theorem C.1.

Step 3. Let $\mathcal{D}$ denote the space of bounded functions that map Θ onto $\mathbb{R}^{2k} \times \Psi$ and let $\mathcal{D}_0$ be the space of bounded uniformly continuous functions that map Θ onto $\mathbb{R}^{2k} \times \Psi$. Let the sequence of functionals $\{g_n\}_{n\geq 1}$, $\{g_n^1\}_{n\geq 1}$, $\{g_n^2\}_{n\geq 1}$ with $g_n : \mathcal{D} \to \mathbb{R}$, $g_n^1 : \mathcal{D} \to \mathbb{R}$, and $g_n^2 : \mathcal{D} \to \mathbb{R}$ be defined by

$$\begin{aligned}
g_n(v(\cdot), \mu(\cdot), \Omega(\cdot)) &\equiv \min\left\{g_n^1(v(\cdot), \mu(\cdot), \Omega(\cdot)), g_n^2(v(\cdot), \mu(\cdot), \Omega(\cdot))\right\} \\
g_n^1(v(\cdot), \mu(\cdot), \Omega(\cdot)) &\equiv \inf_{(\theta,\ell) \in \Lambda_n^{R1}} S(v_n^*(\theta) + \varphi^*(\mu_{n,1}(\theta) + \mu_{n,2}(\theta)'\ell), \Omega(\theta)) \\
g_n^2(v(\cdot), \mu(\cdot), \Omega(\cdot)) &\equiv \inf_{(\theta,\ell) \in \Lambda_n^{R2}} S(v_n^*(\theta) + \mu_{n,1}(\theta) + \mu_{n,2}(\theta)'\ell, \Omega(\theta)) \; .
\end{aligned}$$

Let the functional $g : \mathcal{D}_0 \to \mathbb{R}$, $g^1 : \mathcal{D}_0 \to \mathbb{R}$, and $g^2 : \mathcal{D}_0 \to \mathbb{R}$ be defined by:

$$\begin{aligned}
g(v(\cdot), \mu(\cdot), \Omega(\cdot)) &\equiv \min\left\{g^1(v(\cdot), \mu(\cdot), \Omega(\cdot)), g^2(v(\cdot), \mu(\cdot), \Omega(\cdot))\right\} \\
g^1(v(\cdot), \mu(\cdot), \Omega(\cdot)) &\equiv \inf_{(\theta,\ell) \in \Lambda^{R1}} S(v_\Omega(\theta) + \varphi^*(\mu_1(\theta) + \mu_2(\theta)'\ell), \Omega(\theta)) \\
g^2(v(\cdot), \mu(\cdot), \Omega(\cdot)) &\equiv \inf_{(\theta,l) \in \Lambda^{R2}} S(v_\Omega(\theta) + \mu_1(\theta) + \mu_2(\theta)'\ell, \Omega(\theta)) \; .
\end{aligned}$$

If the sequence of deterministic functions $\{(v_n(\cdot), \mu_n(\cdot), \Omega_n(\cdot))\}_{n\geq 1}$ with $(v_n(\cdot), \mu_n(\cdot), \Omega_n(\cdot)) \in \mathcal{D}$ for all $n \in \mathbb{N}$ satisfies

$$\lim_{n\to\infty} \sup_{\theta \in \Theta} ||(v_n(\theta), \mu_n(\theta), \Omega_n(\theta)) - (v_\Omega(\theta), (\mathbf{0}_k, \mathbf{1}_k), \Omega(\theta))|| = 0 \; ,$$

for some $(v(\cdot),(\mathbf{0}_k,\mathbf{1}_k),\Omega(\cdot))\in\mathcal{D}_0$ then $\lim_{n\to\infty}||g_n^s(v_n(\cdot),\mu_n(\cdot),\Omega_n(\cdot))-g^s(v(\cdot),(\mathbf{0}_k,\mathbf{1}_k),\Omega(\cdot))||=0$ for $s=1,2$, respectively. This follows from similar steps to those in the proof of Theorem C.1, step 4. By continuity of the minimum function,

$$\lim_{n\to\infty}||g_n(v_n(\cdot),\mu_n(\cdot),\Omega_n(\cdot))-g(v(\cdot),(\mathbf{0}_k,\mathbf{1}_k),\Omega(\cdot))||=0\ .$$

Step 4. By combining the representation of $\min\{\tilde{T}_n^{R1}(\gamma_n),T_n^{R2}(\gamma_n)\}$ in step 1, the convergence results in steps 2 and 3, Theorem C.1, and the extended continuous mapping theorem (see, e.g., Theorem 1.11.1 of van der Vaart and Wellner (1996)) we conclude that

$$\{\min\{\tilde{T}_n^{R1}(\gamma_n),T_n^{R2}(\gamma_n)\}|\{W_i\}_{i=1}^n\}\overset{d}{\to}\min\left\{J(\Lambda_*^{R1},\Omega),J(\Lambda^{R2},\Omega)\right\}\ a.s.,$$

where

$$J(\Lambda_*^{R1},\Omega)\equiv\inf_{(\theta,\ell)\in\Lambda_*^{R1}}S(v_\Omega(\theta)+\ell,\Omega(\theta))=\inf_{(\theta,\ell')\in\Lambda^{R1}}S(v_\Omega(\theta)+\varphi^*(\ell'),\Omega(\theta))\ . \tag{D-18}$$

The result then follows by noticing that,

$$\begin{aligned}\min\left\{J(\Lambda_*^{R1},\Omega),J(\Lambda^{R2},\Omega)\right\}&=\min\left\{\inf_{(\theta,\ell)\in\Lambda_*^{R1}}S(v_\Omega(\theta)+\ell,\Omega(\theta)),\inf_{(\theta,\ell)\in\Lambda^{R2}}S(v_\Omega(\theta)+\ell,\Omega(\theta))\right\}\\&=\inf_{(\theta,\ell)\in\Lambda_*^{R1}\cup\Lambda^{R2}}S(v_\Omega(\theta)+\ell,\Omega(\theta))=J(\Lambda^{MR},\Omega)\ .\end{aligned}$$

This completes the proof. □

Proof of Theorem C.3. This proof is similar to that of Theorem C.1. For the sake of brevity, we only provide a sketch that focuses on the main differences. From the definition of $T_{b_n}^{SS}(\gamma_n)$, we can consider the following derivation,

$$\begin{aligned}T_{b_n}^{SS}(\gamma_n)&\equiv\inf_{\theta\in\Theta(\gamma_n)}Q_{b_n}^{SS}(\theta)=\inf_{\theta\in\Theta(\gamma_n)}S(\sqrt{b_n}\bar{m}_{b_n}^{SS}(\theta),\hat{\Sigma}_{b_n}^{SS}(\theta))\\&=\inf_{\theta\in\Theta(\gamma_n)}S(\tilde{v}_{b_n}^{SS}(\theta)+\sqrt{b_n}D_{F_n}^{-1/2}(\theta)(\bar{m}_n(\theta)-E_{F_n}[m(W,\theta)])+\sqrt{b_n}D_{F_n}^{-1/2}(\theta)E_{F_n}[m(W,\theta)],\tilde{\Omega}_{b_n}^{SS}(\theta))\\&=\inf_{(\theta,\ell)\in\Lambda_{b_n}^{SS}}S(\tilde{v}_{b_n}^{SS}(\theta)+\mu_n(\theta)+\ell,\tilde{\Omega}_{b_n}^{SS}(\theta))\ ,\end{aligned}$$

where $\mu_n(\theta)\equiv\sqrt{b_n}D_{F_n}^{-1/2}(\theta)(\bar{m}_n(\theta)-E_{F_n}[m(W,\theta)])$, $\tilde{v}_{b_n}^{SS}(\theta)$ is as in (C-3), and $\tilde{\Omega}_n^{SS}(\theta)\equiv D_{F_n}^{-1/2}(\theta)\hat{\Sigma}_{b_n}^{SS}(\theta)D_{F_n}^{-1/2}(\theta)$. From here, we can repeat the arguments used in the proof of Theorem C.1. The main difference in the argument is that the reference to parts 2 and 8 in Lemma C.1 need to be replaced by parts 10 and 9, respectively. □

Proof of Theorem C.4. The proof of this theorem follows by combining arguments from the proof of Theorem C.1 with those from Bugni et al. (2013, Theorem 3.1). It is therefore omitted. □

D.3 Proofs of lemmas in Appendix C

We note that Lemmas C.2-C.5 correspond to Lemmas D3-D7 in Bugni et al. (2013) and so we do not include the proofs of those lemmas in this paper.

Proof of Lemma C.1. The proof of parts 1-8 follow from similar arguments to those used in the proof of Bugni et al. (2013, Theorem D.2). Therefore, we now focus on the proof of parts 9-10.

Part 9. By the argument used to prove Bugni et al. (2013, Theorem D.2 (part 1)), $\mathcal{M}(F)\equiv\{D_F^{-1/2}(\theta)m(\cdot,\theta):\mathcal{W}\to\mathbb{R}^k\}$ is Donsker and pre-Gaussian, both uniformly in $F\in\mathcal{P}$. Thus, we can extend the arguments in the proof of van der Vaart and Wellner (1996, Theorem 3.6.13 and Example 3.6.14) to hold under a drifting sequence of distributions $\{F_n\}_{n\geq1}$ along the lines of van der Vaart and Wellner (1996, Section 2.8.3). From this, it follows that:

$$\left\{\sqrt{\frac{n}{(n-b_n)}}\tilde{v}_{b_n}^{SS}(\theta)\middle|\{W_i\}_{i=1}^n\right\}\ \overset{d}{\to}\ v_\Omega(\theta)\quad\text{in } l^\infty(\Theta)\quad a.s. \tag{D-19}$$

To conclude the proof, note that,

$$\sup_{\theta\in\Theta}\left\|\sqrt{\frac{n}{(n-b_n)}}\tilde{v}^{SS}_{b_n}(\theta)-\tilde{v}^{SS}_{b_n}(\theta)\right\|=\sup_{\theta\in\Theta}\|\tilde{v}^{SS}_{b_n}(\theta)\|\sqrt{\frac{b_n/n}{(1-b_n/n)}}~.$$

In order to complete the proof, it suffices to show that the RHS of the previous equation is $o_p(1)$ a.s. In turn, this follows from $b_n/n=o(1)$ and (D-19) as they imply that $\{\sup_{\theta\in\Theta}||\tilde{v}^{SS}_{b_n}(\theta)|||\{W_i\}_{i=1}^n\}=O_p(1)$ a.s.

Part 10. This result follows from considering the subsampling analogue of the arguments used to prove Bugni et al. (2013, Theorem D.2 (part 2)). □

Proof of Lemma C.6. Part 1. Suppose not, that is, suppose that $\sup_{\Omega\in\Psi}S(x,\Omega)=\infty$ for $x\in(-\infty,\infty]^p\times\mathbb{R}^{k-p}$. By definition, there exists a sequence $\{\Omega_n\in\Psi\}_{n\geq1}$ s.t. $S(x,\Omega_n)\to\infty$. By the compactness of Ψ, there exists a subsequence $\{k_n\}_{n\geq1}$ of $\{n\}_{n\geq1}$ s.t. $\Omega_{k_n}\to\Omega^*\in\Psi$. By continuity of S on $(-\infty,\infty]^p\times\mathbb{R}^{k-p}\times\Psi$ it then follows that $\lim S(x,\Omega_{k_n})=S(x,\Omega^*)=\infty$ for $(x,\Omega^*)\in(-\infty,\infty]^p\times\mathbb{R}^{k-p}\times\Psi$, which is a contradiction to $S:(-\infty,\infty]^p\times\mathbb{R}^{k-p}\to\mathbb{R}_+$.

Part 2. Suppose not, that is, suppose that $\sup_{\Omega\in\Psi}S(x,\Omega)=B<\infty$ for $x\notin(-\infty,\infty]^p\times\mathbb{R}^{k-p}$. By definition, there exists a sequence $\{\Omega_n\in\Psi\}_{n\geq1}$ s.t. $S(x,\Omega_n)\to\infty$. By the compactness of Ψ, there exists a subsequence $\{k_n\}_{n\geq1}$ of $\{n\}_{n\geq1}$ s.t. $\Omega_{k_n}\to\Omega^*\in\Psi$. By continuity of S on $\mathbb{R}^k_{[\pm\infty]}\times\Psi$ it then follows that $\lim S(x,\Omega_{k_n})=S(x,\Omega^*)=B<\infty$ for $(x,\Omega^*)\in\mathbb{R}^k_{[\pm\infty]}\times\Psi$. Let $J\in\{1,\dots k\}$ be set of coordinates s.t. $x_j=-\infty$ for $j\leq p$ or $|x_j|=\infty$ for $j>p$. By the case under consideration, there is at least one such coordinate. Define $M\equiv\max\{\max_{j\notin J,j\leq p}[x_j]_-,\max_{j\notin J,j>p}|x_j|\}<\infty$. For any $C>M$, let $x'(C)$ be defined as follows. For $j\notin J$, set $x'_j(C)=x_j$ and for $j\in J$, set $x'_j(C)$ as follows $x'_j(C)=-C$ for $j\leq p$ and $|x'_j(C)|=C$ for $j>p$. By definition, $\lim_{C\to\infty}x'(C)=x$ and by continuity properties of the function S, $\lim_{C\to\infty}S(x'(C),\Omega^*)=S(x,\Omega^*)=B<\infty$. By homogeneity properties of the function S and by Lemma C.4, we have that

$$S(x'(C),\Omega^*)=C^{\chi}S(C^{-1}x'(C),\Omega^*)\geq C^{\chi}\inf_{(x,\Omega)\in A\times\Psi}S(x,\Omega)>0,$$

where A is the set in Lemma C.4. Taking $C\to\infty$ the RHS diverges to infinity, producing a contradiction. □

Proof of Lemma C.7. The result follows from similar steps to those in Bugni et al. (2013, Lemma D.10) and is therefore omitted. □

Proof of Lemma C.8. Let $(\theta,\ell)\in\Lambda^{R2}$ with $\ell\in\mathbb{R}^p_{[+\infty]}\times\mathbb{R}^{k-p}$. Then, there is a subsequence $\{a_n\}_{n\geq1}$ of $\{n\}_{n\geq1}$ and a sequence $\{(\theta_n,\ell_n)\}_{n\geq1}$ such that $\theta_n\in\Theta(\gamma_n)$, $\ell_n\equiv\kappa_n^{-1}\sqrt{n}D^{-1/2}_{F_n}(\theta_n)E_{F_n}[m(W,\theta_n)]$, $\lim_{n\to\infty}\ell_{a_n}=\ell$, and $\lim_{n\to\infty}\theta_{a_n}=\theta$. Also, by $\Omega_{F_n}\overset{u}{\to}\Omega$ we get $\Omega_{F_n}(\theta_n)\to\Omega(\theta)$. By continuity of $S(\cdot)$ at $(\ell,\Omega(\theta))$ with $\ell\in\mathbb{R}^p_{[+\infty]}\times\mathbb{R}^{k-p}$,

$$\kappa_{a_n}^{-\chi}{a_n}^{\chi/2}Q_{F_{a_n}}(\theta_{a_n})=S(\kappa_{a_n}^{-1}\sqrt{a_n}\sigma^{-1}_{F_{a_n},j}(\theta_{a_n})E_{F_{a_n}}[m_j(W,\theta_{a_n})],\Omega_{F_{a_n}}(\theta_{a_n}))\to S(\ell,\Omega(\theta))<\infty~. \tag{D-20}$$

Hence $Q_{F_{a_n}}(\theta_{a_n})=O(\kappa^{\chi}_{a_n}{a_n}^{-\chi/2})$. By this and Assumption A.7(a), it follows that

$$O(\kappa^{\chi}_{a_n}{a_n}^{-\chi/2})=c^{-1}Q_{F_{a_n}}(\theta_{a_n})\geq\min\{\delta,\inf_{\tilde{\theta}\in\Theta_I(F_{a_n},\gamma_{a_n})}||\theta_{a_n}-\tilde{\theta}||\}^{\chi}\quad\Rightarrow\quad||\theta_{a_n}-\tilde{\theta}_{a_n}||\leq O(\kappa_{a_n}/\sqrt{a_n})~, \tag{D-21}$$

for some sequence $\{\tilde{\theta}_{a_n}\in\Theta_I(F_{a_n},\gamma_{a_n})\}_{n\geq1}$. By the convexity of $\Theta(\gamma_n)$ and Assumption A.7(c), the intermediate value theorem implies that there is a sequence $\{\theta^*_n\in\Theta(\gamma_n)\}_{n\geq1}$ with θ^*_n in the line between θ_n and $\tilde{\theta}_n$ such that

$$\kappa_n^{-1}\sqrt{n}D^{-1/2}_{F_n}(\theta_n)E_{F_n}[m(W,\theta_n)]=G_{F_n}(\theta^*_n)\kappa_n^{-1}\sqrt{n}(\theta_n-\tilde{\theta}_n)+\kappa_n^{-1}\sqrt{n}D^{-1/2}_{F_n}(\tilde{\theta}_n)E_{F_n}[m(W,\tilde{\theta}_n)]~.$$

Define $\hat{\theta}_n\equiv(1-\kappa_n^{-1})\tilde{\theta}_n+\kappa_n^{-1}\theta_n$ or, equivalently, $\hat{\theta}_n-\tilde{\theta}_n\equiv\kappa_n^{-1}(\theta_n-\tilde{\theta}_n)$. We can write the above equation as

$$G_{F_n}(\theta^*_n)\sqrt{n}(\hat{\theta}_n-\tilde{\theta}_n)=\kappa_n^{-1}\sqrt{n}D^{-1/2}_{F_n}(\theta_n)E_{F_n}[m(W,\theta_n)]-\kappa_n^{-1}\sqrt{n}D^{-1/2}_{F_n}(\tilde{\theta}_n)E_{F_n}[m(W,\tilde{\theta}_n)]~. \tag{D-22}$$

By convexity of $\Theta(\gamma_n)$ and $\kappa_n^{-1} \to 0$, $\{\hat{\theta}_n \in \Theta(\gamma_n)\}_{n\geq 1}$ and by (D-21), $\sqrt{a_n}||\hat{\theta}_{a_n} - \tilde{\theta}_{a_n}|| = O(1)$. By the intermediate value theorem again, there is a sequence $\{\theta_n^{**} \in \Theta(\gamma_n)\}_{n\geq 1}$ with θ_n^{**} in the line between $\hat{\theta}_n$ and $\tilde{\theta}_n$ such that

$$\begin{aligned}\sqrt{n}D_{F_n}^{-1/2}(\hat{\theta}_n)E_{F_n}[m(W,\hat{\theta}_n)] &= G_{F_n}(\theta_n^{**})\sqrt{n}(\hat{\theta}_n - \tilde{\theta}_n) + \sqrt{n}D_{F_n}^{-1/2}(\tilde{\theta}_n)E_{F_n}[m(W,\tilde{\theta}_n)] \\ &= G_{F_n}(\theta_n^{*})\sqrt{n}(\hat{\theta}_n - \tilde{\theta}_n) + \sqrt{n}D_{F_n}^{-1/2}(\tilde{\theta}_n)E_{F_n}[m(W,\tilde{\theta}_n)] + \epsilon_{1,n} \ , \end{aligned} \quad \text{(D-23)}$$

where the second equality holds by $\epsilon_{1,n} \equiv (G_{F_n}(\theta_n^{**}) - G_{F_n}(\theta_n^{*}))\sqrt{n}(\hat{\theta}_n - \tilde{\theta}_n)$. Combining (D-22) with (D-23) we get

$$\sqrt{n}D_{F_n}^{-1/2}(\hat{\theta}_n)E_{F_n}[m(W,\hat{\theta}_n)] = \kappa_n^{-1}\sqrt{n}D_{F_n}^{-1/2}(\theta_n)E_{F_n}[m(W,\theta_n)] + \epsilon_{1,n} + \epsilon_{2,n} \ , \quad \text{(D-24)}$$

where $\epsilon_{2,n} \equiv (1-\kappa_n^{-1})\sqrt{n}D_{F_n}^{-1/2}(\tilde{\theta}_n)E_{F_n}[m(W,\tilde{\theta}_n)]$. From $\{\tilde{\theta}_{a_n} \in \Theta_I(F_{a_n}, \gamma_{a_n})\}_{n\geq 1}$ and $\kappa_n^{-1} \to 0$, it follows that $\epsilon_{2,a_n,j} \geq 0$ for $j \leq p$ and $\epsilon_{2,a_n,j} = 0$ for $j > p$. Moreover, Assumption A.7(c) implies that $||G_{F_{a_n}}(\theta_{a_n}^{**}) - G_{F_{a_n}}(\theta_{a_n}^{*})|| = o(1)$ for any sequence $\{F_{a_n} \in \mathcal{P}_0\}_{n\geq 1}$ whenever $||\theta_{a_n}^{*} - \theta_{a_n}^{**}|| = o(1)$. Using $\sqrt{a_n}||\hat{\theta}_{a_n} - \tilde{\theta}_{a_n}|| = O(1)$, we have

$$||\epsilon_{1,a_n}|| \leq ||G_{F_{a_n}}(\theta_{a_n}^{**}) - G_{F_{a_n}}(\theta_{a_n}^{*})||\sqrt{a_n}||\hat{\theta}_{a_n} - \tilde{\theta}_{a_n}|| = o(1) \ . \quad \text{(D-25)}$$

Finally, since $(\mathbb{R}^k_{[\pm\infty]}, d)$ is compact, there is a further subsequence $\{u_n\}_{n\geq 1}$ of $\{a_n\}_{n\geq 1}$ s.t. $\sqrt{u_n}D_{F_{u_n}}^{-1/2}(\hat{\theta}_{u_n})E_{F_{u_n}}[m(W,\hat{\theta}_{u_n})]$ and $\kappa_{u_n}^{-1}\sqrt{u_n}D_{F_{u_n}}^{-1/2}(\theta_{u_n})E_{F_{u_n}}[m(W,\theta_{u_n})]$ converge. Then, from (D-24), (D-25), and the properties of ϵ_{2,a_n} we conclude that

$$\begin{aligned}\lim_{n\to\infty}\tilde{\ell}_{u_n,j} &\equiv \lim_{n\to\infty}\sqrt{u_n}\sigma^{-1}_{F_{u_n},j}(\hat{\theta}_{u_n})E_{F_{u_n}}[m_j(W,\hat{\theta}_{u_n})] \geq \lim_{n\to\infty}\kappa_{u_n}^{-1}\sqrt{u_n}\sigma^{-1}_{F_{u_n},j}(\theta_{u_n})E_{F_{u_n}}[m_j(W,\theta_{u_n})], \quad \text{for } j \leq p \ , \\ \lim_{n\to\infty}\tilde{\ell}_{u_n,j} &\equiv \lim_{n\to\infty}\sqrt{u_n}\sigma^{-1}_{F_{u_n},j}(\hat{\theta}_{u_n})E_{F_{u_n}}[m_j(W,\hat{\theta}_{u_n})] = \lim_{n\to\infty}\kappa_{u_n}^{-1}\sqrt{u_n}\sigma^{-1}_{F_{u_n},j}(\theta_{u_n})E_{F_{u_n}}[m_j(W,\theta_{u_n})], \quad \text{for } j > p \ ,\end{aligned}$$

which completes the proof, as $\{(\hat{\theta}_{u_n}, \tilde{\ell}_{u_n}) \in \Lambda_{u_n,F_{u_n}}(\gamma_{u_n})\}_{n\geq 1}$ and $\hat{\theta}_{u_n} \to \theta$. □

Proof of Lemma C.9. We divide the proof into four steps.

Step 1. We show that $\inf_{(\theta,\ell)\in\Lambda^{SS}} S(v_\Omega(\theta) + \ell, \Omega(\theta)) < \infty$ a.s. By Assumption A.9, there exists a sequence $\{\tilde{\theta}_n \in \Theta_I(F_n,\gamma_n)\}_{n\geq 1}$, where $d_H(\Theta(\gamma_n), \Theta(\gamma_0)) = O(n^{-1/2})$. Then, there exists another sequence $\{\theta_n \in \Theta(\gamma_0)\}_{n\geq 1}$ s.t. $\sqrt{n}||\theta_n - \tilde{\theta}_n|| = O(1)$ for all $n \in \mathbb{N}$. Since Θ is compact, there is a subsequence $\{k_n\}_{n\geq 1}$ s.t. $\sqrt{k_n}(\theta_{k_n} - \tilde{\theta}_{k_n}) \to \lambda \in \mathbb{R}^{d_\theta}$, and $\theta_{k_n} \to \theta^*$ and $\tilde{\theta}_{k_n} \to \theta^*$ for some $\theta^* \in \Theta$. For any $n \in \mathbb{N}$, let $\ell_{k_n,j} \equiv \sqrt{b_{k_n}}\sigma^{-1}_{F_{k_n},j}(\theta_{k_n})E_{F_{k_n}}[m_j(W,\theta_{k_n})]$ for $j = 1,\ldots,k$, and note that

$$\ell_{k_n,j} = \sqrt{b_{k_n}}\sigma^{-1}_{F_{k_n},j}(\tilde{\theta}_{k_n})E_{F_{k_n}}[m_j(W,\tilde{\theta}_{k_n})] + \Delta_{k_n,j} \quad \text{(D-26)}$$

by the intermediate value theorem, where $\hat{\theta}_{k_n}$ lies between θ_{k_n} and $\tilde{\theta}_{k_n}$ for all $n \in \mathbb{N}$, and

$$\Delta_{k_n,j} \equiv \frac{\sqrt{b_{k_n}}}{\sqrt{k_n}}(G_{F_{k_n},j}(\hat{\theta}_{k_n}) - G_{F_{k_n},j}(\theta^*))\sqrt{k_n}(\theta_{k_n} - \tilde{\theta}_{k_n}) + \frac{\sqrt{b_{k_n}}}{\sqrt{k_n}}G_{F_{k_n},j}(\theta^*)\sqrt{k_n}(\theta_{k_n} - \tilde{\theta}_{k_n}) \ .$$

Letting $\Delta_{k_n} = \{\Delta_{k_n,j}\}_{j=1}^k$, it follows that

$$||\Delta_{k_n}|| \leq \frac{\sqrt{b_{k_n}}}{\sqrt{k_n}}||G_{F_{k_n}}(\hat{\theta}_{k_n}) - G_{F_{k_n}}(\theta^*)|| \times ||\sqrt{k_n}(\theta_{k_n} - \tilde{\theta}_{k_n})|| + ||\frac{\sqrt{b_{k_n}}}{\sqrt{k_n}}G_{F_{k_n}}(\theta^*)|| \times ||\sqrt{k_n}(\theta_{k_n} - \tilde{\theta}_{k_n})|| = o(1) \ , \quad \text{(D-27)}$$

where $b_n/n \to 0$, $\sqrt{k_n}(\theta_{k_n} - \tilde{\theta}_{k_n}) \to \lambda$, $\sqrt{b_{k_n}}G_{F_{k_n}}(\theta^*)/\sqrt{k_n} = o(1)$, $\hat{\theta}_{k_n} \to \theta^*$, and $||G_{F_{k_n}}(\hat{\theta}_{k_n}) - G_{F_{k_n}}(\theta^*)|| = o(1)$ for any sequence $\{F_{k_n} \in \mathcal{P}_0\}_{n\geq 1}$ by Assumption A.7(c). Thus, for all $j \leq k$,

$$\lim_{n\to\infty}\ell_{k_n,j} \equiv \lim_{n\to\infty}\sqrt{b_{k_n}}\sigma^{-1}_{F_{k_n},j}(\theta_{k_n})E_{F_{k_n}}[m_j(W,\theta_{k_n})] = \ell_j^* \equiv \lim_{n\to\infty}\sqrt{b_{k_n}}\sigma^{-1}_{F_{k_n},j}(\tilde{\theta}_{k_n})E_{F_{k_n}}[m_j(W,\tilde{\theta}_{k_n})] \ .$$

Since $\{\tilde{\theta}_n \in \Theta_I(F_n,\gamma_n)\}_{n\geq 1}$, $\ell_j^* \geq 0$ for $j \leq p$ and $\ell_j^* = 0$ for $j > p$. Let $\ell^* \equiv \{\ell_j^*\}_{j=1}^k$. By definition, $\{(\theta_{k_n}, \ell_{k_n}) \in$

$\Lambda^{SS}_{b_{k_n},F_{k_n}}(\gamma_0)\}_{n\geq 1}$ and $d((\theta_{k_n},\ell_{k_n}),(\theta^*,\ell^*))\to 0$, which implies that $(\theta^*,\ell^*)\in\Lambda^{SS}$. From here, we conclude that

$$\inf_{(\theta,\ell)\in\Lambda^{SS}} S(v_\Omega(\theta)+\ell,\Omega(\theta))\leq S(v_\Omega(\theta^*)+\ell^*,\Omega(\theta^*))\leq S(v_\Omega(\theta^*),\Omega(\theta^*))\ ,$$

where the first inequality follows from $(\theta^*,\ell^*)\in\Lambda^{SS}$, the second inequality follows from the fact that $\ell^*_j\geq 0$ for $j\leq p$ and $\ell^*_j=0$ for $j>p$ and the properties of $S(\cdot)$. Finally, the RHS is bounded as $v_\Omega(\theta^*)$ is bounded a.s.

Step 2. We show that if $(\bar\theta,\bar\ell)\in\Lambda^{SS}$ with $\bar\ell\in\mathbb{R}^p_{[+\infty]}\times\mathbb{R}^{k-p}$, $\exists(\bar\theta,\ell^*)\in\Lambda^{R2}$ where $\ell^*_j\geq\bar\ell_j$ for $j\leq p$ and $\ell^*_j=\bar\ell_j$ for $j>p$. As an intermediate step, we use the limit sets under the sequence $\{(\gamma_n,F_n)\}_{n\geq 1}$, denoted by Λ^{SS}_A and Λ^{R2}_A in the statement of the lemma.

We first show that $(\bar\theta,\bar\ell)\in\Lambda^{SS}_A$. Since $\Lambda^{SS}_{b_n,F_n}(\gamma_0)\overset{H}{\to}\Lambda^{SS}$, there exist a subsequence $\{(\theta_{k_n},\ell_{k_n})\in\Lambda^{SS}_{b_{k_n},F_{k_n}}(\gamma_0)\}_{n\geq 1}$, $\theta_{k_n}\to\bar\theta$, and $\ell_{k_n}\equiv\sqrt{b_{k_n}}D^{-1/2}_{F_{k_n}}(\theta_{k_n})E_{F_{k_n}}[m(W,\theta_{k_n})]\to\bar\ell$. To show that $(\bar\theta,\bar\ell)\in\Lambda^{SS}_A$, we now find a subsequence $\{(\theta'_{k_n},\ell'_{k_n})\in\Lambda^{SS}_{b_{k_n},F_{k_n}}(\gamma_n)\}_{n\geq 1}$, $\theta'_{k_n}\to\bar\theta$, and $\ell'_{k_n}\equiv\sqrt{b_{k_n}}D^{-1/2}_{F_{k_n}}(\theta'_{k_n})E_{F_{k_n}}[m(W,\theta'_{k_n})]\to\bar\ell$. Notice that $\{(\theta_{k_n},\ell_{k_n})\in\Lambda^{SS}_{b_{k_n},F_{k_n}}(\gamma_0)\}_{n\geq 1}$ implies that $\{\theta_{k_n}\in\Theta(\gamma_0)\}_{n\geq 1}$. This and $d_H(\Theta(\gamma_n),\Theta(\gamma_0))=O(n^{-1/2})$ implies that there is $\{\theta'_{k_n}\in\Theta(\gamma_{k_n})\}_{n\geq 1}$ s.t. $\sqrt{k_n}||\theta'_{k_n}-\theta_{k_n}||=O(1)$ which implies that $\theta'_{k_n}\to\bar\theta$. By the intermediate value theorem there exists a sequence $\{\theta^*_n\in\Theta\}_{n\geq 1}$ with θ^*_n in the line between θ_n and θ'_n such that

$$\begin{aligned}\ell'_{k_n}&\equiv\sqrt{b_{k_n}}D^{-1/2}_{F_{k_n}}(\theta'_{k_n})E_{F_{k_n}}[m(W,\theta'_{k_n})]=\sqrt{b_{k_n}}D^{-1/2}_{F_{k_n}}(\theta_{k_n})E_{F_{k_n}}[m(W,\theta_{k_n})]+\sqrt{b_{k_n}}G_{F_{k_n}}(\theta^*_{k_n})(\theta'_{k_n}-\theta_{k_n})\\&=\ell_{k_n}+\Delta_{k_n}\to\bar\ell\ ,\end{aligned}$$

where we have defined $\Delta_{k_n}\equiv\sqrt{b_{k_n}}G_{F_{k_n}}(\theta^*_{k_n})(\theta'_{k_n}-\theta_{k_n})$ and $\Delta_{k_n}=o(1)$ holds by similar arguments to those in (D-27). This proves $(\bar\theta,\bar\ell)\in\Lambda^{SS}_A$.

We now show that $\exists(\bar\theta,\ell^*)\in\Lambda^{R2}_A$ where $\ell^*_j\geq\bar\ell_j$ for $j\leq p$ and $\ell^*_j=\bar\ell_j$ for $j>p$. Using similar arguments to those in (D-20) and (D-21) in the proof of Lemma C.8, we have that $Q_{F_{k_n}}(\theta'_{k_n})=O(b^{-\chi/2}_{k_n})$ and that there is a sequence $\{\tilde\theta_n\in\Theta_I(F_n,\gamma_n)\}_{n\geq 1}$ s.t. $\sqrt{b_{k_n}}||\theta'_{k_n}-\tilde\theta_{k_n}||=O(1)$.

Following similar steps to those leading to (D-22) in the proof of Lemma C.8, it follows that

$$\kappa^{-1}_n\sqrt{n}G_{F_n}(\theta^*_n)(\hat\theta_n-\tilde\theta_n)=\sqrt{b_n}D^{-1/2}_{F_n}(\theta'_n)E_{F_n}[m(W,\theta'_n)]-\sqrt{b_n}D^{-1/2}_{F_n}(\tilde\theta_n)E_{F_n}[m(W,\tilde\theta_n)]\ ,\tag{D-28}$$

where $\{\theta^*_n\in\Theta(\gamma_n)\}_{n\geq 1}$ lies in the line between θ'_n and $\tilde\theta_n$, and $\hat\theta_n\equiv(1-\kappa_n\sqrt{b_n/n})\tilde\theta_n+\kappa_n\sqrt{b_n/n}\theta'_n$. By Assumption A.8, $\hat\theta_n$ is a convex combination of $\tilde\theta_n$ and θ'_n for n sufficiently large. Note also that $\sqrt{b_{k_n}}||\hat\theta_{k_n}-\tilde\theta_{k_n}||=o(1)$. By doing yet another intermediate value theorem expansion, there is a sequence $\{\theta^{**}_n\in\Theta(\gamma_n)\}_{n\geq 1}$ with θ^{**}_n in the line between $\hat\theta_n$ and $\tilde\theta_n$ such that

$$\kappa^{-1}_n\sqrt{n}D^{-1/2}_{F_n}(\hat\theta_n)E_{F_n}[m(W,\hat\theta_n)]=\kappa^{-1}_n\sqrt{n}G_{F_n}(\theta^{**}_n)(\hat\theta_n-\tilde\theta_n)+\kappa^{-1}_n\sqrt{n}D^{-1/2}_{F_n}(\tilde\theta_n)E_{F_n}[m(W,\tilde\theta_n)]\ .\tag{D-29}$$

Since $\sqrt{b_{k_n}}||\theta^*_{k_n}-\tilde\theta_{k_n}||=O(1)$ and $\sqrt{b_{k_n}}||\tilde\theta_{k_n}-\theta^{**}_{k_n}||=o(1)$, it follows that $\sqrt{b_{k_n}}||\theta^*_{k_n}-\theta^{**}_{k_n}||=O(1)$. Next,

$$\begin{aligned}\kappa^{-1}_n\sqrt{n}D^{-1/2}_{F_n}(\hat\theta_n)E_{F_n}[m(W,\hat\theta_n)]&=\kappa^{-1}_n\sqrt{n}G_{F_n}(\theta^*_n)(\hat\theta_n-\tilde\theta_n)+\kappa^{-1}_n\sqrt{n}D^{-1/2}_{F_n}(\tilde\theta_n)E_{F_n}[m(W,\tilde\theta_n)]+\Delta_{n,1}\\&=\sqrt{b_n}D^{-1/2}_{F_n}(\theta'_n)E_{F_n}[m(W,\theta'_n)]+\Delta_{n,1}+\Delta_{n,2}\ ,\end{aligned}\tag{D-30}$$

where the first equality follows from (D-29) and $\Delta_{n,1}\equiv\kappa^{-1}_n\sqrt{n}(G_{F_n}(\theta^{**}_n)-G_{F_n}(\theta^*_n))(\hat\theta_n-\tilde\theta_n)$, and the second holds by (D-28) and $\Delta_{n,2}\equiv\kappa^{-1}_n\sqrt{n}(1-\kappa_n\sqrt{b_n/n})D^{-1/2}_{F_n}(\tilde\theta_n)E_{F_n}[m(W,\tilde\theta_n)]$. By similar arguments to those in the proof of Lemma C.8, $||\Delta_{k_n,1}||=o(1)$. In addition, Assumption A.8 and $\{\tilde\theta_n\in\Theta_I(F_n,\gamma_n)\}_{n\geq 1}$ imply that $\Delta_{n,2,j}\geq 0$ for $j\leq p$ and n sufficiently large, and that $\Delta_{n,2,j}=0$ for $j>p$ and all $n\geq 1$.

Now define $\ell''_{k_n}\equiv\kappa^{-1}_{k_n}\sqrt{k_n}D^{-1/2}_{F_{k_n}}(\hat\theta_{k_n})E_{F_n}[m(W,\hat\theta_{k_n})]$ so that by compactness of $(\mathbb{R}^k_{[\pm\infty]},d)$ there is a further subsequence $\{u_n\}_{n\geq 1}$ of $\{k_n\}_{n\geq 1}$ s.t. $\ell''_{u_n}=\kappa^{-1}_{u_n}\sqrt{u_n}D^{-1/2}_{F_{u_n}}(\hat\theta_{u_n})E_{F_{u_n}}[m(W,\hat\theta_{u_n})]$ and $\Delta_{u_n,1}$ converges. We define $\ell^*\equiv\lim_{n\to\infty}\ell''_{u_n}$. By (D-30) and properties of $\Delta_{n,1}$ and $\Delta_{n,2}$, we conclude that

$$\lim_{n\to\infty}\ell''_{u_n,j}=\lim_{n\to\infty}\kappa^{-1}_{u_n}\sqrt{u_n}\sigma^{-1}_{F_{u_n},j}(\hat\theta_{u_n})E_{F_{u_n}}[m_j(W,\hat\theta_{u_n})]\geq\lim_{n\to\infty}\sqrt{b_{u_n}}\sigma^{-1}_{F_{u_n},j}(\theta'_{u_n})E_{F_{u_n}}[m_j(W,\theta'_{u_n})]=\bar\ell_j,\ \text{for } j\leq p\ ,$$

$$\lim_{n\to\infty} \ell''_{u_n,j} = \lim_{n\to\infty} \kappa_{u_n}^{-1}\sqrt{u_n}\sigma^{-1}_{F_{u_n},j}(\hat\theta_{u_n})E_{F_{u_n}}[m_j(W,\hat\theta_{u_n})] = \lim_{n\to\infty} \sqrt{b_{u_n}}\sigma^{-1}_{F_{u_n},j}(\theta'_{u_n})E_{F_{u_n}}[m_j(W,\theta'_{u_n})] = \bar\ell_j, \text{ for } j>p ,$$

Thus, $\{(\hat\theta_{u_n},\ell''_{u_n}) \in \Lambda^{R2}_{u_n,F_{u_n}}(\gamma_n)\}_{n\geq 1}$, $\hat\theta_{u_n} \to \bar\theta$, and $\ell''_{u_n} \to \ell^*$ where $\ell^*_j \geq \bar\ell_j$ for $j\leq p$ and $\ell^*_j = \bar\ell_j$ for $j>p$, and $(\bar\theta,\ell^*) \in \Lambda^{R2}_A$.

We conclude the step by showing that $(\bar\theta,\ell^*) \in \Lambda^{R2}$. To this end, find a subsequence $\{(\theta^\dagger_{u_n},\ell^\dagger_{u_n}) \in \Lambda^{R2}_{b_{u_n},F_{u_n}}(\gamma_0)\}_{n\geq 1}$, $\theta^\dagger_{u_n} \to \bar\theta$, and $\ell^\dagger_{u_n} \equiv \kappa^{-1}_{u_n}\sqrt{u_n}D^{-1/2}_{F_{u_n}}(\theta^\dagger_{u_n})E_{F_{u_n}}[m(W,\theta^\dagger_{u_n})] \to \ell^*$. Notice that $\{(\hat\theta_{u_n},\ell''_{u_n}) \in \Lambda^{R2}_{u_n,F_{u_n}}(\gamma_n)\}_{n\geq1}$ implies that $\{\hat\theta_{u_n} \in \Theta(\gamma_{u_n})\}_{n\geq 1}$. This and $d_H(\Theta(\gamma_n),\Theta(\gamma_0)) = O(n^{-1/2})$ implies that there is $\{\theta^\dagger_{u_n} \in \Theta(\gamma_0)\}_{n\geq1}$ s.t. $\sqrt{u_n}||\hat\theta_{u_n}-\theta^\dagger_{u_n}|| = O(1)$ which implies that $\theta^\dagger_{u_n} \to \bar\theta$. By the intermediate value theorem there exists a sequence $\{\theta^{***}_n \in \Theta\}_{n\geq1}$ with θ^{***}_n in the line between $\hat\theta_n$ and $\theta^\dagger_n$ such that

$$\begin{aligned}\ell^\dagger_{u_n} &\equiv \kappa^{-1}_{u_n}\sqrt{u_n}D^{-1/2}_{F_{u_n}}(\theta^\dagger_{u_n})E_{F_{u_n}}[m(W,\theta^\dagger_{u_n})] = \kappa^{-1}_{u_n}\sqrt{u_n}D^{-1/2}_{F_{u_n}}(\theta^\dagger_{u_n})E_{F_{u_n}}[m(W,\theta^\dagger_{u_n})] + \kappa^{-1}_{u_n}\sqrt{u_n}G_{F_{u_n}}(\theta^{***}_{u_n})(\theta^\dagger_{u_n}-\hat\theta_{u_n})\\ &= \ell''_{u_n} + \Delta_{u_n} \to \ell^* ,\end{aligned}$$

where we have define $\Delta_{u_n} \equiv \kappa^{-1}_{u_n}\sqrt{u_n}G_{F_{u_n}}(\theta^{***}_{u_n})(\theta^\dagger_{u_n}-\hat\theta_{u_n})$ and $\Delta_{u_n} = o(1)$ holds by similar arguments to those used before. By definition, this proves that $(\bar\theta,\ell^*) \in \Lambda^{R2}$.

Step 3. We show that $\inf_{(\theta,\ell)\in\Lambda^{SS}} S(v_\Omega(\theta)+\ell,\Omega(\theta)) \geq \inf_{(\theta,\ell)\in\Lambda^{R2}} S(v_\Omega(\theta)+\ell,\Omega(\theta))$ a.s. Since v_Ω is a tight stochastic process, there is a subset of the sample space $\mathcal{W}$, denoted $\mathcal{A}_1$, s.t. $P(\mathcal{A}_1)=1$ and $\forall\omega\in\mathcal{A}_1$, $\sup_{\theta\in\Theta}||v_\Omega(\omega,\theta)||<\infty$. By step 1, there is a subset of $\mathcal{W}$, denoted $\mathcal{A}_2$, s.t. $P(\mathcal{A}_2)=1$ and $\forall\omega\in\mathcal{A}_2$,

$$\inf_{(\theta,\ell)\in\Lambda^{SS}} S(v_\Omega(\omega,\theta)+\ell,\Omega(\theta)) < \infty .$$

Define $\mathcal{A}\equiv\mathcal{A}_1\cap\mathcal{A}_2$ and note that $P(\mathcal{A})=1$. In order to complete the proof, it then suffices to show that $\forall\omega\in\mathcal{A}$,

$$\inf_{(\theta,\ell)\in\Lambda^{SS}} S(v_\Omega(\omega,\theta)+\ell,\Omega(\theta)) \geq \inf_{(\theta,\ell)\in\Lambda^{R2}} S(v_\Omega(\omega,\theta)+\ell,\Omega(\theta)) . \quad \text{(D-31)}$$

Fix $\omega\in\mathcal{A}$ arbitrarily and suppose that (D-31) does not occur, i.e.,

$$\Delta \equiv \inf_{(\theta,\ell)\in\Lambda^{R2}} S(v_\Omega(\omega,\theta)+\ell,\Omega(\theta)) - \inf_{(\theta,\ell)\in\Lambda^{SS}} S(v_\Omega(\omega,\theta)+\ell,\Omega(\theta)) > 0 . \quad \text{(D-32)}$$

By definition of infimum, $\exists(\bar\theta,\bar\ell)\in\Lambda^{SS}$ s.t. $\inf_{(\theta,\ell)\in\Lambda^{SS}} S(v_\Omega(\omega,\theta)+\ell,\Omega(\theta)) + \Delta/2 \geq S(v_\Omega(\omega,\bar\theta)+\bar\ell,\Omega(\bar\theta))$, and so, from this and (D-32) it follows that

$$S(v_\Omega(\omega,\bar\theta)+\bar\ell,\Omega(\bar\theta)) \leq \inf_{(\theta,\ell)\in\Lambda^{R2}} S(v_\Omega(\omega,\theta)+\ell,\Omega(\theta)) - \Delta/2 . \quad \text{(D-33)}$$

We now show that $\bar\ell\in\mathbb{R}^p_{[+\infty]}\times\mathbb{R}^{k-p}$. Suppose not, i.e., suppose that $\bar\ell_j=-\infty$ for some $j<p$ and $|\bar\ell_j|=\infty$ for some $j>p$. Since $\omega\in\mathcal{A}\subseteq\mathcal{A}_1$, $||v_\Omega(\omega,\bar\theta)||<\infty$. By part 2 of Lemma C.6 it then follows that $S(v_\Omega(\omega,\bar\theta)+\bar\ell,\Omega(\bar\theta))=\infty$. By (D-33), $\inf_{(\theta,\ell)\in\Lambda^{SS}} S(v_\Omega(\omega,\theta)+\ell,\Omega(\theta))=\infty$, which is a contradiction to $\omega\in\mathcal{A}_2$.

Since $\bar\ell\in\mathbb{R}^p_{[+\infty]}\times\mathbb{R}^{k-p}$, step 2 implies that $\exists(\bar\theta,\ell^*)\in\Lambda^{R2}$ where $\ell^*_j\geq\bar\ell_j$ for $j\leq p$ and $\ell^*_j=\bar\ell_j$ for $j>p$. By properties of $S(\cdot)$,

$$S(v_\Omega(\omega,\bar\theta)+\ell^*,\Omega(\bar\theta)) \leq S(v_\Omega(\omega,\bar\theta)+\bar\ell,\Omega(\bar\theta)) . \quad \text{(D-34)}$$

Combining (D-32), (D-33), (D-34), and $(\bar\theta,\ell^*)\in\Lambda^{R2}$, we reach the following contradiction,

$$\begin{aligned}0<\Delta/2 &\leq \inf_{(\theta,\ell)\in\Lambda^{R2}} S(v_\Omega(\omega,\theta)+\ell,\Omega(\theta)) - S(v_\Omega(\omega,\bar\theta)+\bar\ell,\Omega(\bar\theta))\\ &\leq \inf_{(\theta,\ell)\in\Lambda^{R2}} S(v_\Omega(\omega,\theta)+\ell,\Omega(\theta)) - S(v_\Omega(\omega,\bar\theta)+\ell^*,\Omega(\bar\theta)) \leq 0 .\end{aligned}$$

Step 4. Suppose the conclusion of the lemma is not true. This is, suppose that $c_{(1-\alpha)}(\Lambda^{R2},\Omega) > c_{(1-\alpha)}(\Lambda^{SS},\Omega)$.

Consider the following derivation

$$\begin{aligned}&\alpha < P(J(\Lambda^{R2},\Omega) > c_{(1-\alpha)}(\Lambda^{SS},\Omega))\\&\quad\leq P(J(\Lambda^{SS},\Omega) > c_{(1-\alpha)}(\Lambda^{SS},\Omega)) + P(J(\Lambda^{R2},\Omega) > J(\Lambda^{SS},\Omega)) = 1 - P(J(\Lambda^{SS},\Omega) \leq c_{(1-\alpha)}(\Lambda^{SS},\Omega)) \leq \alpha\ ,\end{aligned}$$

where the first strict inequality holds by definition of quantile and $c_{(1-\alpha)}(\Lambda^{R2},\Omega) > c_{(1-\alpha)}(\Lambda^{SS},\Omega)$, the last equality holds by step 3, and all other relationships are elementary. Since the result is contradictory, the proof is complete. □

Proof of Lemma C.10. By Theorem 3.2, $\liminf(E_{F_n}[\phi_n^{R2}(\gamma_0)] - E_{F_n}[\phi_n^{SS}(\gamma_0)]) \geq 0$. Suppose that the desired result is not true. Then, there is a further subsequence $\{u_n\}_{n\geq 1}$ of $\{n\}_{n\geq 1}$ s.t.

$$\lim E_{F_{u_n}}[\phi_{u_n}^{R2}(\gamma_0)] = \lim E_{F_{u_n}}[\phi_{u_n}^{SS}(\gamma_0)]\ . \tag{D-35}$$

This sequence $\{u_n\}_{n\geq 1}$ will be referenced from here on. We divide the remainder of the proof into steps.

Step 1. Asymptotic distribution of $T_n^{SS}(\gamma_0)$. We show that there is subsequence $\{a_n\}_{n\geq 1}$ of $\{u_n\}_{n\geq 1}$ s.t.

$$\{T_{a_n}^{SS}(\gamma_0)|\{W_i\}_{i=1}^{a_n}\} \xrightarrow{d} S(v_\Omega(\theta^*) + (g,\mathbf{0}_{k-p}),\Omega(\theta^*)),\ \text{a.s.} \tag{D-36}$$

Conditionally on $\{W_i\}_{i=1}^n$, Assumption A.11(c) implies that

$$T_n^{SS}(\gamma_0) = S(\sqrt{b_n}D_{F_n}^{-1}(\hat{\theta}_n^{SS})\bar{m}_{b_n}^{SS}(\hat{\theta}_n^{SS}),\tilde{\Omega}_{b_n}^{SS}(\hat{\theta}_n^{SS})) + o_p(1),\ \text{a.s.} \tag{D-37}$$

Then, (D-36) would follow from (D-37) provided that there is a subsequence $\{a_n\}_{n\geq 1}$ of $\{u_n\}_{n\geq 1}$ s.t.

$$\{S(\sqrt{b_{a_n}}D_{F_{a_n}}^{-1/2}(\hat{\theta}_{a_n}^{SS})\bar{m}_{b_{a_n}}^{SS}(\hat{\theta}_{a_n}^{SS}),\tilde{\Omega}_{b_{a_n}}^{SS}(\hat{\theta}_{a_n}^{SS}))|\{W_i\}_{i=1}^{a_n}\} \xrightarrow{d} S(v_\Omega(\theta^*) + (g,\mathbf{0}_{k-p}),\Omega(\theta^*)),\ \text{a.s.}$$

This follows, by the maintained assumptions, from finding a subsequence $\{a_n\}_{n\geq 1}$ of $\{u_n\}_{n\geq 1}$ s.t.

$$\{\tilde{\Omega}_{a_n}^{SS}(\hat{\theta}_{a_n}^{SS})|\{W_i\}_{i=1}^{a_n}\} \xrightarrow{p} \Omega(\theta^*),\ \text{a.s.} \tag{D-38}$$

$$\{\sqrt{b_{a_n}}D_{F_{a_n}}^{-1/2}(\hat{\theta}_{a_n}^{SS})\bar{m}_{a_n}^{SS}(\hat{\theta}_{a_n}^{SS})|\{W_i\}_{i=1}^{a_n}\} \xrightarrow{d} v_\Omega(\theta^*) + (g,\mathbf{0}_{k-p}),\ \text{a.s.} \tag{D-39}$$

To show (D-38), note that

$$||\tilde{\Omega}_n^{SS}(\hat{\theta}_n^{SS}) - \Omega(\theta^*)|| \leq \sup_{\theta\in\Theta}||\tilde{\Omega}_n^{SS}(\theta,\theta) - \Omega(\theta,\theta)|| + ||\Omega(\hat{\theta}_n^{SS}) - \Omega(\theta^*)||\ .$$

The RHS is a sum of two terms. Lemma C.1 (part 5) implies that the first term is conditionally $o_p(1)$ a.s. By $\Omega \in \mathcal{C}(\Theta^2)$, and that, conditionally, $\hat{\theta}_n^{SS} \xrightarrow{p} \theta^*$ a.s., the second term is conditionally $o_p(1)$ a.s. This implies that (D-38) holds for the original sequence $\{n\}_{n\geq 1}$ and thus it also holds for its subsequence $\{a_n\}_{n\geq 1}$.

To show (D-39), note that

$$\sqrt{b_n}D_{F_n}^{-1/2}(\hat{\theta}_n^{SS})\bar{m}_n^{SS}(\hat{\theta}_n^{SS}) = \tilde{v}_n^{SS}(\theta^*) + (g,\mathbf{0}_{k-p}) + \mu_{n,1} + \mu_{n,2}\ ,$$

where

$$\begin{aligned}\mu_{n,1} &\equiv (\tilde{v}_{b_n}^{SS}(\hat{\theta}_n^{SS}) - \tilde{v}_{b_n}^{SS}(\theta^*)) + (\sqrt{b_n}D_{F_n}^{-1/2}(\hat{\theta}_n^{SS})E_{F_n}[m(W,\hat{\theta}_n^{SS})] - \sqrt{b_n}D_{F_n}^{-1/2}(\tilde{\theta}_n^{SS})E_{F_n}[m(W,\tilde{\theta}_n^{SS})])\\&\quad+(\sqrt{b_n}D_{F_n}^{-1/2}(\tilde{\theta}_n^{SS})E_{F_n}[m(W,\tilde{\theta}_n^{SS})] - (g,\mathbf{0}_{k-p}))\\\mu_{n,2} &\equiv \tilde{v}_n(\hat{\theta}_n^{SS})\sqrt{b_n/n}\ .\end{aligned}$$

Lemma C.1 (part 9) implies that $\{\tilde{v}_{b_n}^{SS}(\theta^*)|\{W_i\}_{i=1}^n\} \xrightarrow{d} v_\Omega(\theta^*)$ a.s. The proof is then completed by showing that

$$\{\mu_{a_n,1}|\{W_i\}_{i=1}^{a_n}\} = o_p(1),\ \text{a.s.} \tag{D-40}$$

$$\{\mu_{a_n,2}|\{W_i\}_{i=1}^{a_n}\} \quad = \quad o_p(1), \text{ a.s.} \tag{D-41}$$

By Assumption A.11(c), (D-40) follows from showing that $\{\tilde{v}_n^{SS}(\theta^*) - \tilde{v}_n^{SS}(\hat{\theta}_n^{SS})|\{W_i\}_{i=1}^n\} = o_p(1)$ a.s., which we now show. Fix $\mu > 0$ arbitrarily, we need to show that

$$\limsup P_{F_n}(||\tilde{v}_n^{SS}(\theta^*) - \tilde{v}_n^{SS}(\hat{\theta}_n^{SS})|| > \varepsilon|\{W_i\}_{i=1}^n) < \mu \text{ a.s.} \tag{D-42}$$

Fix $\delta > 0$ arbitrarily. As a preliminary step, we first show that

$$\lim P_{F_n}(\rho_{F_n}(\theta^*, \hat{\theta}_n^{SS}) \geq \delta|\{W_i\}_{i=1}^n) = 0 \;\; \text{a.s.} \;, \tag{D-43}$$

where ρ_{F_n} is the intrinsic variance semimetric in (A-1). Then, for any $j = 1, \ldots, k$,

$$V_{F_n}(\sigma_{F_n,j}^{-1}(\hat{\theta}_n^{SS})m_j(W, \hat{\theta}_n^{SS}) - \sigma_{F_n,j}^{-1}(\theta^*)m_j(W, \theta^*)) = 2(1 - \Omega_{F_n}(\theta^*, \hat{\theta}_n^{SS})_{[j,j]}) \;.$$

By (A-1), this implies that

$$P_{F_n}(\rho_{F_n}(\theta^*, \hat{\theta}_n^{SS}) \geq \delta|\{W_i\}_{i=1}^n) \leq \sum\nolimits_{j=1}^k P_{F_n}(1 - \Omega_{F_n}(\theta^*, \hat{\theta}_n^{SS})_{[j,j]} \geq \delta^2 2^{-1}k^{-1}|\{W_i\}_{i=1}^n) \;. \tag{D-44}$$

Fix $j = 1, \ldots, k$ arbitrarily and note that

$$\begin{aligned} P_{F_n}(1 - \Omega_{F_n}(\theta^*, \hat{\theta}_n^{SS})_{[j,j]} \geq \delta^2 2^{-1}k^{-1}|\{W_i\}_{i=1}^n) \quad &\leq \quad P_{F_n}(1 - \Omega(\theta^*, \hat{\theta}_n^{SS})_{[j,j]} \geq \delta^2 2^{-2}k^{-1}|\{W_i\}_{i=1}^n) + o(1) \\ &\leq \quad P_{F_n}(||\theta^* - \hat{\theta}_n^{SS}|| > \tilde{\delta}|\{W_i\}_{i=1}^n) + o(1) = o_{a.s.}(1) \;, \end{aligned}$$

where we have used that $\Omega_{F_n} \stackrel{u}{\to} \Omega$ and so $\sup_{\theta,\theta' \in \Theta} ||\Omega(\theta, \theta')_{[j,j]} - \Omega_{F_n}(\theta, \theta')_{[j,j]}|| < \delta^2 2^{-2}k^{-1}$ for all sufficiently large n, that $\Omega \in \mathcal{C}(\Theta^2)$ and so $\exists\tilde{\delta} > 0$ s.t. $||\theta^* - \hat{\theta}_n^{SS}|| \leq \tilde{\delta}$ implies that $1 - \Omega(\theta^*, \hat{\theta}_n^{SS})_{[j,j]} \leq \delta^2 2^{-2}k^{-1}$, and that $\{\hat{\theta}_n^{SS}|\{W_i\}_{i=1}^n\} \stackrel{p}{\to} \theta^*$ a.s. Combining this with (D-44), (D-43) follows.

Lemma C.1 (part 1) implies that $\{\tilde{v}_n^{SS}(\cdot)|\{W_i\}_{i=1}^n\}$ is asymptotically ρ_F-equicontinuous uniformly in $F \in \mathcal{P}$ (a.s.) in the sense of van der Vaart and Wellner (1996, page 169). Then, $\exists\delta > 0$ s.t.

$$\limsup_{n\to\infty} P_{F_n}^*(\sup_{\rho_{F_n}(\theta,\theta')<\delta} ||\tilde{v}_n^{SS}(\theta) - \tilde{v}_n^{SS}(\theta')|| > \varepsilon|\{W_i\}_{i=1}^n) < \mu \;\; \text{a.s.} \tag{D-45}$$

Based on this choice, consider the following argument:

$$\begin{aligned} P_{F_n}^*(||\tilde{v}_n^{SS}(\theta^*) - \tilde{v}_n^{SS}(\hat{\theta}_n^{SS})|| > \varepsilon|\{W_i\}_{i=1}^n) \leq P_{F_n}^*(&\sup_{\rho_{F_n}(\theta,\theta')<\delta} ||\tilde{v}_n^{SS}(\theta^*) - \tilde{v}_n^{SS}(\hat{\theta}_n)|| > \varepsilon|\{W_i\}_{i=1}^n) \\ &+ P_{F_n}^*(\rho_{F_n}(\theta^*, \hat{\theta}_n) \geq \delta|\{W_i\}_{i=1}^n) \;. \end{aligned}$$

From this, (D-43), and (D-45), (D-42) follows. To conclude this step, it suffices to show (D-41). By Lemma C.1 (part 7), $\sup_{\theta\in\Theta} ||\tilde{v}_n(\theta)||\sqrt{b_n/n} \stackrel{p}{\to} 0$, and by taking a further subsequence $\{a_n\}_{n\geq1}$ of $\{n\}_{n\geq1}$, $\sup_{\theta\in\Theta} ||\tilde{v}_{a_n}(\theta)||\sqrt{b_{a_n}/a_n} \stackrel{a.s.}{\to} 0$. Since $\tilde{v}_n(\cdot)$ is conditionally non-stochastic, $\{\sup_{\theta\in\Theta} ||\tilde{v}_{a_n}(\theta)||\sqrt{b_{a_n}/a_n}|\{W_i\}_{i=1}^{a_n}\} \stackrel{p}{\to} 0$ a.s. From this, (D-41) follows.

Step 2. Analyze the asymptotic distribution of $c_n^{SS}(\gamma_0, 1-\alpha)$. For arbitrary $\varepsilon > 0$ and for the subsequence $\{a_n\}_{n\geq1}$ of $\{u_n\}_{n\geq1}$ in step 1 we want show that

$$\lim P_{F_{a_n}}(|c_{a_n}^{SS}(\gamma_0, 1-\alpha) - c_{(1-\alpha)}(g, \Omega(\theta^*))| \leq \varepsilon) = 1 \;, \tag{D-46}$$

where $c_{(1-\alpha)}(g, \Omega(\theta^*))$ denotes the $(1-\alpha)$-quantile of $S(v_\Omega(\theta^*) + (g, \mathbf{0}_{k-p}), \Omega(\theta^*))$. We now show that $c_{(1-\alpha)}(g, \Omega(\theta^*)) > 0$. If $k > p$, it follows from our maintained assumptions. If $k = p$, Assumption A.11(b.ii,e) implies $\infty > -\lambda_j > h_j$ for some $j \leq p$, which implies $g_j \leq 0$. By our maintained assumptions, the result then follows.

Fix $\bar{\varepsilon} \in (0, \min\{\varepsilon, c_{(1-\alpha)}(g, \Omega(\theta^*))\})$. By our maintained assumptions, $c_{(1-\alpha)}(g, \Omega(\theta^*)) - \bar{\varepsilon}$ and $c_{(1-\alpha)}(g, \Omega(\theta^*)) + \bar{\varepsilon}$

are continuity points of the CDF of $S(v_\Omega(\theta^*) + (g, \mathbf{0}_{k-p}), \Omega(\theta^*))$. Then,

$$\lim P_{F_{a_n}}(T^{SS}_{a_n}(\gamma_0) \leq c_{(1-\alpha)}(g, \Omega(\theta^*)) + \bar{\varepsilon} | \{W_i\}_{i=1}^{a_n}) = P(S(v_\Omega(\theta^*) + (g, \mathbf{0}_{k-p}), \Omega(\theta^*)) \leq c_{(1-\alpha)}(g, \Omega(\theta^*)) + \bar{\varepsilon}) > 1-\alpha \ , \quad \text{(D-47)}$$

where the equality holds a.s. by part 1, and the strict inequality holds by $\bar{\varepsilon} > 0$. By a similar argument,

$$\lim P_{F_{a_n}}(T^{SS}_{a_n}(\gamma_0) \leq c_{(1-\alpha)}(g, \Omega(\theta^*)) - \bar{\varepsilon} | \{W_i\}_{i=1}^{n}) = P(S(v_\Omega(\theta^*) + (g, \mathbf{0}_{k-p}), \Omega(\theta^*)) \leq c_{(1-\alpha)}(g, \Omega(\theta^*)) - \bar{\varepsilon}) < 1-\alpha \ a.s. \quad \text{(D-48)}$$

Next, notice that

$$\{\lim P_{F_{a_n}}(T^{SS}_{a_n}(\gamma_0) \leq c_{(1-\alpha)}(g, \Omega(\theta^*)), \Omega(\theta^*)) + \bar{\varepsilon} | \{W_i\}_{i=1}^{a_n}) > 1-\alpha\} \subseteq \{\liminf\{c^{SS}_{a_n}(\gamma_0, 1-\alpha) < c_{(1-\alpha)}(g, \Omega(\theta^*)) + \bar{\varepsilon}\}\} \ ,$$

with the same result holding with $-\bar{\varepsilon}$ replacing $+\bar{\varepsilon}$. By combining this result with (D-47) and (D-48), we get

$$\{\liminf\{|c^{SS}_{a_n}(\gamma_0, 1-\alpha) - c_{(1-\alpha)}(g, \Omega(\theta^*))| \leq \bar{\varepsilon}\}\} \quad \text{a.s.}$$

From this result, $\bar{\varepsilon} < \varepsilon$, and Fatou's Lemma, (D-46) follows.

Step 3. Analyze the asymptotic distribution of $c^{R2}_n(\gamma_0, 1-\alpha)$. For any $\theta \in \Theta(\gamma_0)$, define $\tilde{T}^{R2}_n(\theta) \equiv S(v^*_n(\theta) + \kappa_n^{-1}\sqrt{n}\hat{D}_n^{-1}(\theta)\bar{m}_n(\theta), \hat{\Omega}_n(\theta))$ and let $\tilde{c}^{R2}_n(\theta, 1-\alpha)$ denote the conditional $(1-\alpha)$-quantile of $\tilde{T}^{R2}_n(\theta)$. By definition, $T^{R2}_n(\gamma_0) \equiv \inf_{\theta \in \Theta(\gamma_0)} \tilde{T}^{R2}_n(\theta)$ and so $c^{R2}_n(\gamma_0, 1-\alpha) \leq \tilde{c}^{R2}_n(\hat{\theta}^{SS}_n, 1-\alpha)$.

Fix $\varepsilon > 0$ arbitrarily. By arguments similar to steps 1 and 2, we deduce that there is subsequence $\{k_n\}_{n\geq 1}$ of $\{a_n\}_{n\geq 1}$ s.t. $\lim P_{F_{k_n}}(|\tilde{c}^{R2}_{k_n}(\hat{\theta}^{SS}_{k_n}, 1-\alpha) - c_{(1-\alpha)}(\pi, \Omega(\theta^*))| \leq \varepsilon) = 1$. This implies that

$$\lim P_{F_{k_n}}(c_{(1-\alpha)}(\pi, \Omega(\theta^*)) + \varepsilon \geq c^{R2}_{k_n}(\gamma_0, 1-\alpha)) = 1 \ . \quad \text{(D-49)}$$

Furthermore, by Assumption A.11(d) and the first part of step 2, we can conclude that

$$0 < c_{(1-\alpha)}(\pi, \Omega(\theta^*)) < c_{(1-\alpha)}(g, \Omega(\theta^*)) \ . \quad \text{(D-50)}$$

Step 4. By using an argument analogous to that used in step 1 we deduce that

$$T_{k_n}(\gamma_0) \xrightarrow{d} S(v_\Omega(\theta^*) + (h, \mathbf{0}_{k-p}), \Omega(\theta^*)) \ . \quad \text{(D-51)}$$

Fix $\varepsilon \in (0, min\{c_{(1-\alpha)}(g, \Omega(\theta^*)), (c_{(1-\alpha)}(g, \Omega(\theta^*)) - c_{(1-\alpha)}(\pi, \Omega(\theta^*)))/2\}$ (possible by (D-50)). By using elementary arguments, we conclude that

$$P_{F_{k_n}}(T_{k_n}(\gamma_0) \leq c^{SS}_{k_n}(\gamma_0, 1-\alpha)) \leq P_{F_{k_n}}(T_{k_n}(\gamma_0) \leq c_{(1-\alpha)}(g, \Omega(\theta^*)) + \varepsilon) + P_{F_{k_n}}(|c^{SS}_{k_n}(\gamma_0, 1-\alpha) - c_{(1-\alpha)}(g, \Omega(\theta^*))| > \varepsilon) \ ,$$

Taking limits and using (D-46), (D-51), and that the CDF of $S(v_\Omega(\theta^*) + (h, \mathbf{0}_{k-p}), \Omega(\theta^*))$ is continuous on positive values, it follows that

$$\limsup P_{F_{k_n}}(T_{k_n}(\gamma_0) \leq c^{SS}_{k_n}(\gamma_0, 1-\alpha)) \leq P(S(v_\Omega(\theta^*) + (h, \mathbf{0}_{k-p}), \Omega(\theta^*)) \leq c_{(1-\alpha)}(g, \Omega(\theta^*)) + \varepsilon) \ . \quad \text{(D-52)}$$

By a completely analogous argument, we conclude that

$$\liminf P_{F_{k_n}}(T_{k_n}(\gamma_0) \leq c^{SS}_{k_n}(\gamma_0, 1-\alpha)) \geq P(S(v_\Omega(\theta^*) + (h, \mathbf{0}_{k-p}), \Omega(\theta^*)) \leq c_{(1-\alpha)}(g, \Omega(\theta^*)) - \varepsilon) \ . \quad \text{(D-53)}$$

Since (D-52) and (D-53) are valid for all sufficiently small $\varepsilon > 0$ and the CDF of $S(v_\Omega(\theta^*) + (h, \mathbf{0}_{k-p}), \Omega(\theta^*))$ is continuous on positive values,

$$\lim E_{F_{k_n}}[\phi^{SS}_{k_n}(\gamma_0)] = P(S(v_\Omega(\theta^*) + (h, \mathbf{0}_{k-p}), \Omega(\theta^*)) > c_{(1-\alpha)}(g, \Omega(\theta^*))) \ . \quad \text{(D-54)}$$

We can now repeat the same arguments used to deduce (D-54) for Test SS in order to deduce an analogous result for Test R2. The main difference is that for Test R2 we do not have a characterization of the minimizer, which is not problematic as we can simply bound the asymptotic rejection rate. This is,

$$\lim E_{F_{k_n}}[\phi^{R2}_{k_n}(\gamma_0)] \geq P(S(v_\Omega(\theta^*) + (h, \mathbf{0}_{k-p}), \Omega(\theta^*)) > c_{(1-\alpha)}(\pi, \Omega(\theta^*)))\ . \tag{D-55}$$

By our maintained assumptions and (D-50), (D-54), and (D-55), we conclude that

$$\begin{aligned} \lim E_{F_{k_n}}[\phi^{R2}_{k_n}(\gamma_0)] &\geq P(S(v_\Omega(\theta^*) + (h, \mathbf{0}_{k-p}), \Omega(\theta^*)) > c_{(1-\alpha)}(\pi, \Omega(\theta^*))) \\ &> P(S(v_\Omega(\theta^*) + (h, \mathbf{0}_{k-p}), \Omega(\theta^*)) > c_{(1-\alpha)}(g, \Omega(\theta^*))) = \lim E_{F_{k_n}}[\phi^{SS}_{k_n}(\gamma_0)]\ . \end{aligned}$$

Since $\{k_n\}_{n\geq 1}$ is a subsequence of $\{u_n\}_{n\geq 1}$, this is a contradiction to (D-35) and concludes the proof. □

References

Andrews, D. W. K. and P. J. Barwick (2012): "Inference for Parameters Defined by Moment Inequalities: A Recommended Moment Selection Procedure," *Econometrica*, 80, 2805–2826.

Andrews, D. W. K. and P. Guggenberger (2009): "Validity of Subsampling and "Plug-in Asymptotic" Inference for Parameters Defined by Moment Inequalities," *Econometric Theory*, 25, 669–709.

Andrews, D. W. K. and S. Han (2009): "Invalidity of the bootstrap and the m out of n bootstrap for confidence interval endpoints defined by moment inequalities," *The Econometrics Journal*, 12, S172–S199.

Andrews, D. W. K. and X. Shi (2013): "Inference based on conditional moment inequalities," *Econometrica*, 81, 609–666.

Andrews, D. W. K. and G. Soares (2010): "Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection," *Econometrica*, 78, 119–158.

Armstrong, T. B. (2011): "Weighted KS statistics for inference on conditional moment inequalities," *arXiv:1112.1023*.

Beresteanu, A. and F. Molinari (2008): "Asymptotic Properties for a Class of Partially Identified Models," *Econometrica*, 76, 763–814.

Bontemps, C., T. Magnac, and E. Maurin (2012): "Set Identified Linear Models," *Econometrica*, 80, 1129–1155.

Bugni, F. A. (2009): "Bootstrap Inference in Partially Identified Models Defined by Moment Inequalities: Coverage of the Elements of the Identified Set," Manuscript, Duke University.

——— (2010): "Bootstrap Inference in Partially Identified Models Defined by Moment Inequalities: Coverage of the Identified Set," *Econometrica*, 78, 735–753.

——— (2014): "A Comparison of Inferential Methods in Partially Identified Models in Terms of Error in Coverage Probability," Manuscript.

Bugni, F. A., I. A. Canay, and P. Guggenberger (2012): "Distortions of asymptotic confidence size in locally misspecified moment inequality models," *Econometrica*, 80, 1741–1768.

Bugni, F. A., I. A. Canay, and X. Shi (2013): "Specification Tests for Partially Identified Models Defined by Moment Inequalities," CeMMAP working paper CWP01/13.

Canay, I. A. (2010): "EL Inference for Partially Identified Models: Large Deviations Optimality and Bootstrap Validity," *Journal of Econometrics*, 156, 408–425.

Chen, X., E. Tamer, and A. Torgovitsky (2011): "Sensitivity Analysis in Partially Identified Semiparametric Models," Manuscript, Northwestern University.

Chernozhukov, V., H. Hong, and E. Tamer (2007): "Estimation and Confidence Regions for Parameter Sets in Econometric Models," *Econometrica*, 75, 1243–1284.

Chernozhukov, V., S. Lee, and A. M. Rosen (2013): "Intersection bounds: estimation and inference," *Econometrica*, 81, 667–737.

Chetverikov, D. (2013): "Adaptive test of conditional moment inequalities," *arXiv:1201.0167*.

Ciliberto, F. and E. Tamer (2010): "Market Structure and Multiple Equilibria in Airline Industry," *Econometrica*, 77, 1791–1828.

Galichon, A. and M. Henry (2011): "Set identification in models with multiple equilibria," *The Review of Economic Studies*, 78, 1264–1298.

Gandhi, A., Z. Lu, and X. Shi (2013): "Estimating Demand for Differentiated Products with Error in Market Shares: A Moment Inequalities Approach," Manuscript, University of Wisconsin - Madison.

Grieco, P. (2013): "Discrete Games with Flexible Information Structures: An Application to Local Grocery Markets," Manuscript, Penn State University.

Imbens, G. and C. F. Manski (2004): "Confidence Intervals for Partially Identified Parameters," *Econometrica*, 72, 1845–1857.

Kline, B. and E. Tamer (2013): "Default Bayesian Inference in a Class of Partially Identified Models," *manuscript, Northwestern University*.

Pakes, A., J. Porter, K. Ho, and J. Ishii (2011): "Moment Inequalities and Their Application," Manuscript, Harvard University.

Politis, D. N. and J. P. Romano (1994): "Large Sample Confidence Regions Based on Subsamples Under Minimal Assumptions," *Annals of Statistics*, 22, 2031–2050.

Politis, D. N., J. P. Romano, and M. Wolf (1999): *Subsampling*, Springer, New York.

Romano, J. P. and A. M. Shaikh (2008): "Inference for Identifiable Parameters in Partially Identified Econometric Models," *Journal of Statistical Planning and Inference*, 138, 2786–2807.

——— (2010): "Inference for the Identified Set in Partially Identified Econometric Models," *Econometrica*, 78, 169–212.

Romano, J. P., A. M. Shaikh, and M. Wolf (2013): "A Practical Two-Step Method for Testing Moment Inequalities," *Available at SSRN 2137550*.

Rosen, A. M. (2008): "Confidence Sets for Partially Identified Parameters that Satisfy a Finite Number of Moment Inequalities," *Journal of Econometrics*, 146, 107–117.

Royden, H. L. (1988): *Real Analysis*, Prentice-Hall.

Stoye, J. (2009): "More on Confidence Intervals for Partially Identified Parameters," *Econometrica*, 77, 1299–1315.

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

Wan, Y. Y. (2013): "An Integration-based Approach to Moment Inequality Models," *Manuscript. University of Toronto*.