Journal of Econometrics 185 (2015) 259–282

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Specification tests for partially identified models defined by moment inequalities[☆]

Federico A. Bugni[a,*], Ivan A. Canay[b], Xiaoxia Shi[c]

[a] Department of Economics, Duke University, 213 Social Sciences Road, Box 90097, Durham, NC, 27708, USA

[b] Department of Economics, Northwestern University, 2001 Sheridan Road, Evanston, IL, 60208, USA

[c] Department of Economics, University of Wisconsin, Madison, 1180 Observatory Drive, Madison, WI, 53706, USA

ARTICLE INFO

Article history:
Received 25 July 2013
Received in revised form
4 September 2014
Accepted 6 October 2014
Available online 18 November 2014

JEL classification:
C01
C12
C15

Keywords:
Partial identification
Moment inequalities
Specification tests
Hypothesis testing

ABSTRACT

This paper studies the problem of specification testing in partially identified models defined by moment (in)equalities. This problem has not been directly addressed in the literature, although several papers have suggested a test based on checking whether confidence sets for the parameters of interest are empty or not, referred to as Test BP. We propose two new specification tests, denoted Test RS and Test RC, that achieve uniform asymptotic size control and dominate Test BP in terms of power in any finite sample and in the asymptotic limit.

1. Introduction

This paper studies the problem of specification testing in partially identified models defined by a finite number of moment equalities and inequalities (henceforth, referred to as (in)equalities). The model can be written as follows. For a parameter vector (θ, F), where $\theta \in \Theta$ is a finite dimensional parameter of interest and F denotes the distribution of the observed data, the model states that

$$E_F[m_j(W_i, \theta)] \geq 0 \quad \text{for } j = 1, \ldots, p,$$
$$E_F[m_j(W_i, \theta)] = 0 \quad \text{for } j = p+1, \ldots, k, \tag{1.1}$$

where $\{W_i\}_{i=1}^n$ is an i.i.d. sequence of random variables with distribution F and $m : \mathbb{R}^d \times \Theta \to \mathbb{R}^k$ is a known measurable function. This model is *partially identified* because the sampling process and the maintained assumptions (that is, Eq. (1.1) together with regularity conditions) restrict the value of the parameter of interest θ to a set, called the *identified set*, which is smaller than Θ but potentially larger than a single point.

The model is said to be *correctly specified* (or statistically adequate) when the moment (in)equalities hold for at least one parameter value, i.e., when the identified set is non-empty.[1] A specification test takes correct specification of the model as the null hypothesis and rejects if the data seem to be inconsistent

☆ We thank Han Hong and two anonymous referees whose valuable suggestions helped greatly improve the paper. We also thank Adam Rosen, the participants at the 2011 Triangle Econometrics Conference, the 2012 Junior Festival on New Developments in Microeconometrics at Northwestern University, the 2012 Summer Meetings of the Econometrics Society, the Workshop on Inference in Nonstandard Problems at Princeton University, the Workshop on Mathematical Statistics of Partially Identified at Oberwolfach, and seminar participants at IUPUI, UC Berkeley, UC Davis, Yale, Columbia, Rochester, Ohio State, UT Austin, Wisconsin, UC San Diego, and Stanford 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.

* Corresponding author.
E-mail addresses: federico.bugni@duke.edu (F.A. Bugni), iacanay@northwestern.edu (I.A. Canay), xshi@ssc.wisc.edu (X. Shi).

[1] The concept of statistical adequacy was introduced by Koopmans (1937) and referred to as the Fisher's axiom of correct specification. The discussion of the importance of a correct specification for inference purposes dates back to Haavelmo (1944).

http://dx.doi.org/10.1016/j.jeconom.2014.10.013
0304-4076/

with it. Specification tests for partially identified models have been studied by a small number of authors (reviewed below), but the only existent test applicable to the general specification of Eq. (1.1) is the one based on checking whether a confidence set for θ is empty or not. We refer to this procedure as "Test BP", to emphasize that it is a *by-product* of confidence sets for θ, and describe it formally in the next section.

In this paper, we propose two new specification tests for the model above and show that they have the following properties. First, our tests achieve uniform size control, just like Test BP. Second, our tests dominate Test BP in terms of power in any finite sample and in the asymptotic limit. Specifically, our tests have more or equal power than Test BP in all finite samples, and there are sequences of local alternative hypotheses for which our tests have strictly higher asymptotic power.

Both of our tests use the same "infimum" test statistic $\inf_{\theta\in\Theta} Q_n(\theta)$, where $Q_n(\theta)$ is the criterion function typically used to construct confidence sets for θ, much in the spirit of the popular J-test in (point-identified) GMM models (see Remark 4.1). The difference between them lies in the critical value used to implement the test. Computing one of these critical values requires little additional work beyond the computation involved in the confidence set construction, just like in Test BP. We therefore always recommend the use of this test, as it attains better power at almost no additional cost. On the other hand, our second test has even better power, but it requires a separate resampling procedure to implement. For this reason, we recommend its use when one has serious interest in the statistical adequacy of the model.[2]

From a methodological point of view, there are two aspects of our paper worth highlighting. First, we derive the limiting distribution of the "infimum" test statistic under drifting sequences of data distributions and provide two methods to approximate its quantiles. To the best of our knowledge, we are the first ones to obtain these kinds of results in partially identified moment (in)equality models. These methodological contributions are relevant in problems that go well beyond specification testing. For example, Bugni et al. (2014) show that hypothesis tests based on the "infimum" test statistic can be adapted to address a large class of interesting new problems, which includes inference on a particular coordinate of a multivariate parameter θ. Second, the asymptotic framework we use is one where the tuning parameter κ_n that determines if a moment inequality is binding, diverges to infinity at an appropriate rate, c.f. Andrews and Soares (2010). In this framework, the arguably best possible implementation of Test BP is the one we use, see Definition 2.4. Recent contributions to the literature have used an alternative asymptotic framework where this tuning parameter κ_n converges to a constant $\kappa < \infty$ that affects the limiting distribution, see Andrews and Barwick (2012), Romano et al. (2014), McCloskey (2014). One could potentially use these methods to define another version of Test BP, and then study the behavior of our tests using fixed-κ asymptotics. We do not pursue this strategy as it involves technical tools that are well beyond those developed here.[3]

The motivation behind our interest in misspecified models stems from the view that most econometric models are only approximations to the underlying phenomenon of interest. This is also the case for partially identified models, where strong and usually unrealistic assumptions are replaced by weaker and more credible ones (see, e.g., Manski, 1989, 2003). In other words, the partial identification approach to inference allows the researcher to conduct inference on the parameter of interest without imposing assumptions on certain fundamental aspects of the model, typically related to the behavior of economic agents. Still, for computational or analytical convenience, the researcher has to impose certain other assumptions, that are typically related to functional forms or distributional assumptions.[4] If these assumptions are not supported by the data, and so the model is misspecified, the resulting statistical inferences are usually invalid (see, e.g., Ponomareva and Tamer, 2011; Bugni et al., 2012).

Specification tests for partially identified models have been studied in Guggenberger et al. (2008), Romano and Shaikh (2008), Andrews and Guggenberger (2009), Andrews and Soares (2010), Santos (2012). Guggenberger et al. (2008) propose to transform a linear moment (in)equality model into a dual form that does not involve parameters and, in this way, eliminate the partial-identification problem. Innovative as it is, their approach only applies to linear models and is not practical when the dimension of the parameter is large because the dimension of the dual form grows exponentially with the dimension of the parameters. Santos (2012) defines specification tests in a partially identified nonparametric instrumental variable model and, thus, his results are not directly applicable to the model in Eq. (1.1). To the best of our knowledge, the only valid specification test for the model in Eq. (1.1) that has been described in the literature is Test BP. This specification test has been proposed by Romano and Shaikh (2008, Remark 3.7), Andrews and Guggenberger (2009, Section 7), and Andrews and Soares (2010, Section 5).[5]

It is worth mentioning that the specification tests we propose in this paper are a type of omnibus tests, in the sense that the specific structure of certain nonparametric alternatives is unknown. However, a partially identified model is typically the result of removing undesirable restrictions in a certain point identified model. As a consequence, refuting the partially identified model leaves the researcher with a reduced set of assumptions that could potentially be wrong. In addition, in some cases testing the specification of a partially identified model can be analogous to directly testing an interesting economic behavior. For example, Kitamura and Stoye (2012) recently proposed a specification test for the Axiom of Revealed Stochastic Preference that shares similarities to our specification tests. In their case, rejecting the specification of the model through their non-parametric test directly means rejection of the Axiom of Revealed Stochastic Preferences. We note, however, that there are substantial differences between our approach and that in Kitamura and Stoye (2012) in terms of the nature of the model, the construction of the test statistic, and the range of applications in which each of these tests can be applied.

The rest of the paper is organized as follows. Section 2 introduces the basic notation we use in our formal analysis and describes the aforementioned Test BP. The tests proposed in this paper compare a test statistic with a critical value. Section 3 introduces our test statistic. The description of our tests is then completed by introducing appropriate critical values that are presented in the succeeding sections. Section 4 describes a critical value based on the asymptotic approximation or bootstrap approximation of the limiting distribution of the test statistic. We call this test the *re-sampling* test or "Test RS". Section 5 describes a critical value that is based on recycling critical values that have already been

2 It is worth pointing out that a version of our second test has been used in Gandhi et al. (2013), with $p = 401$ and a parameter θ with more than 20 coordinates, which illustrates the feasibility of this test in real scale applications.

3 For example, all tests would suffer from asymptotic size distortion and size correction would be needed.

4 See Manski (2003) and Tamer (2003) for a discussion on the role of different assumptions and partial identification.

5 It is important to clarify that Test BP was conceived by papers whose main objective was the construction of confidence sets and not the design of a specification test. In addition, Test BP has some robustness properties, see Remark 6.7.

considered in the literature. We call this test the *re-cycling* test or "Test RC". Section 6 compares the asymptotic size and power of the new tests we propose and the existing test, Test BP. Finally, Section 7 presents evidence from Monte Carlo simulations and Section 8 concludes. The Appendix includes all of the proofs of the paper and several intermediate results. Finally, throughout the paper we divide the assumptions in two groups: maintained assumptions indexed by the letter M (to denote the assumptions that have been already assumed by the literature) and regular assumptions indexed by the letter A (to denote the assumptions that introduced by this paper).

2. Framework

The objective of our inferential procedure is to test whether the moment conditions in Eq. (1.1) are valid or not for at least one parameter value, while maintaining a set of regularity conditions that we use to derive uniform asymptotic statements. We assume throughout the paper that F, the distribution of the observed data, belongs to a *baseline probability space* that we define below. Given this baseline space, we define an appropriate subset where the null hypothesis holds, denoted *null probability space*. These two spaces are the main pieces in the description of our testing problem. We then introduce more technical assumptions in Section 3 before presenting the main results. The next three definitions provide the basic framework of our problem.

Definition 2.1 (*Baseline Probability 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 for some $\theta \in \Theta$, (θ, F) satisfies:

(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 $j = 1, \ldots, k$,
(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[6], and M and a are fixed positive constants.

Definition 2.2 (*Null Probability Space*)**.** The null space of probability measures, denoted by $\mathcal{P}_0 \equiv \mathcal{P}_0(a, M, \Psi)$, is the set of distributions F such that for some $\theta \in \Theta$, (θ, F) satisfies:

- Conditions (i)–(iv) in Definition 2.1,

(v) $E_F[m_j(W_i, \theta)] \geq 0$ for $j = 1, \ldots, p$,
(vi) $E_F[m_j(W_i, \theta)] = 0$ for $j = p + 1, \ldots, k$,

where Ψ, M, and a are as in Definition 2.1.

Definition 2.3 (*Identified Set*)**.** For any distribution $F \in \mathcal{P}$, the corresponding identified set $\Theta_I(F)$ is the set of parameters $\theta \in \Theta$ such that (θ, F) satisfies the moment (in)equalities in Eq. (1.1) or, equivalently, conditions (v)–(vi) in Definition 2.2.

We can now use these definitions to describe the null and alternative hypothesis of our test in a concise way. Under the maintained hypothesis that $F \in \mathcal{P}$, our objective is to conduct the following hypothesis test,

$$H_0 : F \in \mathcal{P}_0 \quad \text{vs.} \quad H_1 : F \notin \mathcal{P}_0. \tag{2.1}$$

By Definitions 2.2 and 2.3, it follows that $F \in \mathcal{P}_0$ if and only if $\Theta_I(F) \neq \emptyset$, and thus the hypotheses in Eq. (2.1) can be alternatively expressed as

$$H_0 : \Theta_I(F) \neq \emptyset \quad \text{vs.} \quad H_1 : \Theta_I(F) = \emptyset, \tag{2.2}$$

which is a convenient representation to characterize the existing test, Test BP, in the next subsection.

To test the hypothesis in Eq. (2.1), we use ϕ_n to denote a non-randomized test that maps data into a binary decision, where $\phi_n = 1$ ($\phi_n = 0$) denotes rejection (non-rejection) of the null hypothesis. The exact size of the test ϕ_n is given by $\sup_{F \in \mathcal{P}_0} E_F[\phi_n]$, while the asymptotic size is

$$AsySz \equiv \limsup_{n \to \infty} \sup_{F \in \mathcal{P}_0} E_F[\phi_n]. \tag{2.3}$$

Given a significance level $\alpha \in (0, 1)$, the test is said to be asymptotically level α if $AsySz \leq \alpha$ and is said to be asymptotically size α or asymptotically size correct if $AsySz = \alpha$. In order to adequately capture the finite sample behavior, the recent literature on inference in partially identified models has emphasized the importance that hypothesis tests satisfy $AsySz \leq \alpha$ rather than pointwise requirement

$$\limsup_{n \to \infty} E_F[\phi_n] \leq \alpha, \quad \forall F \in \mathcal{P}_0.$$

See, e.g., Imbens and Manski (2004); Romano and Shaikh (2008); Andrews and Guggenberger (2009); Andrews and Soares (2010), and Mikusheva (2010).

2.1. The existent specification test

This section formally introduces Test BP, which is currently used by the literature as the specification test in partially identified models. As we have already explained, this test arises as a by-product of confidence sets for partially identified parameters and has been described in Romano and Shaikh (2008, Remark 3.7), Andrews and Guggenberger (2009, Section 7), and Andrews and Soares (2010, Section 5). Before describing this test, we need additional notation.

All the specification tests that this paper considers build upon the *criterion function approach* developed by Chernozhukov et al. (2007). In this approach, we define a non-negative function of the parameter space, $Q_F : \Theta \to \mathbb{R}_+$, referred to as *population criterion function*, with the property that

$$Q_F(\theta) = 0 \iff \theta \in \Theta_I(F). \tag{2.4}$$

As the notation suggests, $Q_F(\theta)$ depends on the unknown probability distribution $F \in \mathcal{P}$ and, thus, it is unknown. We therefore use a sample criterion function, denoted by Q_n, that approximates the population criterion function and can be used for inference. In the context of the moment (in)equality model in Eq. (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),

$$Q_F(\theta) = S(E_F[m(W, \theta)], \Sigma_F(\theta)), \tag{2.5}$$

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.[7] The (properly scaled) sample analogue criterion function is given by

$$Q_n(\theta) = S(\sqrt{n}\bar{m}_n(\theta), \hat{\Sigma}_n(\theta)), \tag{2.6}$$

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 $\Sigma_F(\theta)$. A natural choice for this estimator is

$$\hat{\Sigma}_n(\theta) = n^{-1}\sum_{i=1}^n (m(W_i, \theta) - \bar{m}_n(\theta))(m(W_i, \theta) - \bar{m}_n(\theta))'. \tag{2.7}$$

---

6 See Andrews and Soares (2010) or Bugni et al. (2012) for a description of the parameter space Ψ.

7 See Assumptions M.4–M.8 in the Appendix for these regularity conditions. Two popular functions that satisfy these conditions are the Modified Method of Moments (MMM) and the Quasi-Likelihood ratio (QLR), see Andrews and Guggenberger, 2009.

Using this notation, we can now define a generic $1-\alpha$ confidence set for θ as

$$CS_n(1-\alpha) = \{\theta \in \Theta : Q_n(\theta) \leq \hat{c}_n(\theta, 1-\alpha)\}, \tag{2.8}$$

where $\hat{c}_n(\theta, 1-\alpha)$ is such that $CS_n(1-\alpha)$ has the correct asymptotic coverage, i.e.,

$$\liminf_{n\to\infty} \inf_{(\theta,F)\in\mathcal{F}_0} P_F(\theta \in CS_n(1-\alpha)) \geq 1-\alpha, \tag{2.9}$$

where $\mathcal{F}_0$ denotes the set of parameters (θ, F) that satisfy the conditions in Definition 2.2.

Confidence sets that have the structure in Eq. (2.8) and satisfy Eq. (2.9) have been proposed by Romano and Shaikh (2008); Andrews and Guggenberger (2009); Andrews and Soares (2010); Canay (2010); and Bugni (2010), among others. In particular, Andrews and Soares (2010) consider confidence sets using plug-in asymptotics, subsampling, or generalized moment selection (GMS), and show that all of these methods satisfy Eq. (2.9). We are now ready to define Test BP.

Definition 2.4 (*Test BP*)**.** Let $CS_n(1-\alpha)$ be a confidence set for θ that satisfies Eq. (2.9). The specification Test BP rejects the null hypothesis in Eq. (2.1) according to the following rejection rule

$$\phi_n^{BP} \equiv 1\{CS_n(1-\alpha) = \emptyset\}. \tag{2.10}$$

Given Eq. (2.9), it follows that Test BP is asymptotically level α (see Theorem C.2 in the Appendix). However, as pointed out in Andrews and Guggenberger (2009) and Andrews and Soares (2010), this test is admittedly conservative, i.e., its asymptotic size may be strictly smaller than α. Although it has not been formally established in the literature, one might also suspect that this test suffers from low (asymptotic) power. Our formal analysis shows that Test BP can have strictly less power than the new specification tests developed in this paper.

Definition 2.4 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 its construction. All the tests we consider in this paper are functions of the sample criterion function defined in Eq. (2.6) 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 the two tests we propose in this paper does depend on the choice of critical value used in the construction of $CS_n(1-\alpha)$. Bugni (2010, 2014) 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 confidence set will be less conservative and more powerful than the analogous test implemented with Plug-in asymptotics or subsampling. Since our objective is to propose new specification tests on the grounds of better asymptotic size control and asymptotic power improvements, we adopt the GMS version of the specification test in Definition 2.4 as the "benchmark version" of Test BP. This is summarized in the following assumption, maintained throughout the paper.

Assumption M.1. Test BP is computed using the GMS approach in Andrews and Soares (2010). In other words, ϕ_n^{BP} in Eq. (2.10) is based on

$$CS_n(1-\alpha) = \{\theta \in \Theta : Q_n(\theta) \leq \hat{c}_n(\theta, 1-\alpha)\}, \tag{2.11}$$

where $\hat{c}_n(\theta, 1-\alpha)$ is the GMS critical value constructed using a function φ and a positive thresholding sequence $\{\kappa_n\}_{n\geq 1}$ satisfying $\kappa_n \to \infty$ and $\kappa_n/\sqrt{n} \to 0$.

We conclude this section by presenting a simple example that illustrates how the identified set can be empty under misspecification. The example is also used in Section 7, as it captures the types of situations where there are power gains of implementing the specification tests we propose.

Example 2.1 (*Missing Data*)**.** The economic model states that the true parameters (θ, F) satisfy

$$E_F[Y|X=x] = \mathcal{H}(x,\theta) \quad \forall x \in S_X, \tag{2.12}$$

where $\mathcal{H}$ is a known continuous function specified by the researcher and $S_X = \{x_l\}_{l=1}^{d_x}$ is the (finite) support of X. As there is missing data on Y, we let Z denote the binary variable that takes value of one if Y is observed and zero if Y is missing. Conditional on $X=x$, Y has logical lower and upper bounds given by $Y_L(x)$ and $Y_H(x)$, respectively. The observed data are $\{W_i\}_{i=1}^n$, where $\forall i = 1, \ldots, n$, $W_i = (Y_iZ_i, Z_i, X_i)$. The model in Eq. (2.12) therefore results in the following moment inequalities for $l = 1, \ldots, d_x$:

$$\begin{aligned} &E_F[m_{l,L}(W,\theta)] \\ &\quad \equiv E_F[(\mathcal{H}(x_l,\theta) - YZ - Y_L(x_l)(1-Z))1\{X=x_l\}] \geq 0, \\ &E_F[m_{l,H}(W,\theta)] \\ &\quad \equiv E_F[(YZ + Y_H(x_l)(1-Z) - \mathcal{H}(x_l,\theta))1\{X=x_l\}] \geq 0. \end{aligned} \tag{2.13}$$

We now choose a simple parametrization that we can use in our Monte Carlo simulations. Suppose that $S_X = \{(1,0,0), (-1,0,1), (0,1,0)\}$, that Y represents a non-negative outcome variable without an upper bound, i.e., $Y_L(x) = 0$ and $Y_H(x) = \infty$, that $\mathcal{H}$ is the linear model $\mathcal{H}(x,\theta) = x'\theta$, $\theta = (\theta_1, \theta_2, 1)$, and that there are missing data for all covariate values, i.e., $P(Z=1|X=x_l) < 1\ \forall l = 1, 2, 3$. In this context, Eq. (2.13) is equivalent to

$$\begin{aligned} &E_F[m_{1,L}(W,\theta)] \equiv E_F[(\theta_1 - YZ)1\{X=x_1\}] \geq 0, \\ &E_F[m_{2,L}(W,\theta)] \equiv E_F[(1-\theta_1 - YZ)1\{X=x_2\}] \geq 0, \\ &E_F[m_{3,L}(W,\theta)] \equiv E_F[(\theta_2 - YZ)1\{X=x_3\}] \geq 0. \end{aligned} \tag{2.14}$$

It is straightforward to show that for any distribution $F \in \mathcal{P}$, the identified set $\Theta_I(F)$ is given by

$$\Theta_I(F) = \left\{(\theta_1,\theta_2) \in \Theta : \begin{cases} \theta_1 \in [E_F[YZ|X=x_1], E_F[1-YZ|X=x_2]], \\ \theta_2 \geq E_F[YZ|X=x_3] \end{cases}\right\}. \tag{2.15}$$

It follows that this model is strictly partially identified (i.e. if $\Theta_I(F)$ is non-empty, it is not a singleton) and it is correctly specified (i.e. $\Theta_I(F)$ is non-empty) if and only if $E_F[YZ|X = x_1] \leq E_F[1-YZ|X=x_2]$. □

3. The new test statistic

The specification tests we present in this paper use the natural test statistic for specification testing, namely, the infimum of the sample criterion function $Q_n(\theta)$ defined in Eq. (2.6). The justification for this test statistic follows immediately from the following two mild assumptions which we maintain throughout the paper.

Assumption M.2. Θ is a nonempty and compact subset of $\mathbb{R}^{d_\theta}$ ($d_\theta < \infty$).

Assumption M.3. For any $F \in \mathcal{P}$, Q_F is a lower semi-continuous function.

Under Assumptions M.2 and M.3, the population criterion function achieves a minimum value in Θ. This minimum value is zero when the identified set is non-empty. More precisely, $\inf_{\theta\in\Theta} Q_F(\theta) \geq 0$ and

$$\inf_{\theta\in\Theta} Q_F(\theta) = 0 \iff \Theta_I(F) \neq \emptyset. \tag{3.1}$$

It then follows that the hypotheses in Eq. (2.1) can be re-written as

$$H_0 : \inf_{\theta\in\Theta} Q_F(\theta) = 0 \quad \text{vs.} \quad H_1 : \inf_{\theta\in\Theta} Q_F(\theta) > 0. \tag{3.2}$$

Based on this formulation of the problem, it is natural to suggest implementing the test using the infimum of the sample analogue criterion function as a test statistic, i.e.,

$$T_n \equiv \inf_{\theta\in\Theta} Q_n(\theta). \tag{3.3}$$

In particular, the specification of the model should be rejected whenever the test statistic exceeds a certain critical value. This leads to the following hypothesis testing procedure.

Definition 3.1 (*New Specification Test*)**.** The new specification test rejects the null hypothesis in Eq. (2.1) according to the following rejection rule

$$\phi_n = 1\left\{T_n > \hat{c}_n(1-\alpha)\right\}, \tag{3.4}$$

where T_n is as in Eq. (3.3) and $\hat{c}_n(1-\alpha)$ is an approximation to the $(1-\alpha)$-quantile of the asymptotic distribution of T_n.

In order to make the test in Definition 3.1 feasible, we need to specify the critical value $\hat{c}_n(1-\alpha)$. The challenging part of our analysis is to propose a critical value in Eq. (3.4) that results in a test that: (a) controls asymptotic size, (b) has superior power properties, and (c) is amenable to computation. We propose two critical values that result in two hypothesis tests that satisfy these requirements. The first critical value is based on an approximation of the distribution of the test statistic under the null hypothesis using resampling methods. This critical value gives rise to "Test RS". The second critical value is based on "recycling" GMS critical values described in the previous section. This critical value gives rise to "Test RC". We describe each of these tests in the next two sections.

Before introducing the new tests, it is convenient to first derive the asymptotic distribution of $\inf_{\theta\in\Theta} Q_n(\theta)$ along (relevant) sequences of data generating processes $\{F_n\}_{n\geq 1}$.

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

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

$$v_{n,j}(\theta) = n^{-1/2}\sigma_{F,j}^{-1}(\theta)\sum_{i=1}^{n}(m_j(W_i,\theta) - E_F[m_j(W_i,\theta)]),$$
$$\text{for } j = 1, \ldots, k, \tag{3.5}$$

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 denotes the coordinate-wise version of the intrinsic variance semimetric (see Eq. (A.2) in Appendix A for details).

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 $\Omega_F\left(\theta,\theta'\right)$ be a $k \times k$ correlation matrix with typical $[j_1, j_2]$-component

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

The matrix Ω_F satisfies

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

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, the class of functions $\{\sigma_{F,j}^{-1}(\theta)m_j(\cdot,\theta) : \mathcal{W} \to \mathbb{R}\}_{j\leq k}$ is Donsker and pre-Gaussian uniformly in $F \in \mathcal{P}$ (see Lemma D.2 and van der Vaart and Wellner (1996, Theorem 2.8.2)). For interpretable sufficient conditions for uniform stochastic equicontinuity, consider the uniform version of Examples 19.6–19.11 in van der Vaart (1998). 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. This condition implies that the Euclidean metric for θ is uniformly stronger than the variance semimetric (see van der Vaart and Wellner, 1996 problem 3, page 93).

The next theorem derives the limit distribution of our test statistic under the above assumptions. In the theorem, we 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 $(\Theta \times \mathbb{R}^k_{[\pm\infty]}, d(\cdot))$, where $d(\cdot)$ is the metric defined in Appendix A, Eq. (A.1). We use the symbols $\xrightarrow{u}$ and $\xrightarrow{H}$ to denote uniform convergence and convergence in Hausdorff distance (see Appendix A). Finally, we let $D_F(\theta) \equiv Diag(\Sigma_F(\theta))$ and

$$\Lambda_{n,F} \equiv \left\{(\theta,\ell) \in \Theta \times \mathbb{R}^k : \ell = \sqrt{n}D_F^{-1/2}(\theta)E_F[m(W,\theta)]\right\}. \tag{3.6}$$

Theorem 3.1. *Let Assumptions A.1–A.4 hold. Let $\{F_n\}_{n\geq 1}$ be a (sub)sequence of distributions such that for some $(\Omega, \Lambda) \in \mathcal{C}(\Theta^2) \times \mathcal{S}(\Theta \times \mathbb{R}^k_{[\pm\infty]})$, (i) $F_n \in \mathcal{P}_0$ for all $n \in \mathbb{N}$, (ii) $\Omega_{F_n}(\theta,\theta') \xrightarrow{u} \Omega(\theta,\theta')$, and (iii) $\Lambda_{n,F_n} \xrightarrow{H} \Lambda$. Then, along the (sub)sequence $\{F_n\}_{n\geq 1}$,*

$$T_n \xrightarrow{d} J(\Lambda,\Omega) \equiv \inf_{(\theta,\ell)\in\Lambda} S(v_\Omega(\theta) + \ell, \Omega(\theta,\theta)), \tag{3.7}$$

where $v_\Omega : \Theta \to \mathbb{R}^k$ is a $\mathbb{R}^k$-valued tight Gaussian process with covariance (correlation) kernel $\Omega \in \mathcal{C}(\Theta^2)$.

Theorem 3.1 gives the asymptotic distribution of our test statistic under a (sub)sequence of distributions that satisfies certain properties. It turns out that these types of (sub)sequences are the relevant ones to determine the asymptotic size of our tests (see Appendix C for additional details).

Having an expression for $J(\Lambda,\Omega)$, our goal is to construct feasible critical values that asymptotically approximate the $1-\alpha$ quantile of this distribution, denoted by $c_{(1-\alpha)}(\Lambda,\Omega)$. This requires approximating the limiting set Λ and the limiting correlation 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 novel difficulties. In the next sections we propose two approaches to circumvent these challenges.

4. Test RS: Re-Sampling

The critical value of Test RS is based on directly approximating the quantiles of $J(\Lambda,\Omega)$. The main challenge in approximating these quantiles lies in the approximation of the set Λ which, by definition, is composed of the cluster points of the sequences of the form

$$\{(\theta_n, \sqrt{n}D_{F_n}^{-1/2}(\theta_n)E_{F_n}[m(W,\theta_n)])\}_{n\geq 1}. \tag{4.1}$$

Notice that the second component in Eq. (4.1) represents the slackness parameter for the moment (in)equalities.

Approximating the limiting behavior of these sequences presents two main difficulties. The first one is a typical problem in this literature: approximating the slackness parameter for the moment inequalities. This problem has been described by Andrews and Soares (2010), which argues that the limit of the slackness parameter cannot be uniformly consistently estimated at a suitable rate of convergence. Their paper overcomes this problem by proposing the GMS method. The idea of this method is to take advantage of the monotonicity of the test function and replacing the slackness parameter with a function of the following sample measure of slackness

$$\xi_{n,j}(\theta_n) = \kappa_n^{-1}\sqrt{n}\hat{\sigma}_{n,j}^{-1}(\theta_n)\bar{m}_{n,j}(\theta_n), \quad \text{for } j=1,\ldots,p, \tag{4.2}$$

where $\{\kappa_n\}_{n\geq 1}$ is a thresholding sequence that satisfies $\kappa_n \to \infty$ and $\kappa_n/\sqrt{n} \to 0$.

In their GMS approximation, Andrews and Soares (2010) consider sequences of $\{(\theta_n, F_n)\}_{n\geq 1}$ that are (a) deterministic and (b) such that $\theta_n \in \Theta_I(F_n)$ for all $n \in \mathbb{N}$. While these are sequences are suitable for their GMS approximation, they prove to be insufficient for our analysis. To be more precise, the second difficulty in approximating the limit of the sequences in Eq. (4.1) is that, by the nature of our test statistic, we are specifically interested in sequences for which θ_n is the infimum of $Q_n(\theta)$ over Θ. This requires us to consider sequences of $\{(\theta_n, F_n)\}_{n\geq 1}$ that are (a) random and (b) such that $\theta_n \in \Theta/\Theta_I(F_n)$ for some $n \in \mathbb{N}$. This second difficulty is completely novel to this paper and cannot be addressed by the approximation methods currently available in the literature.

Despite the aforementioned difficulties, we show that one can approximate the quantiles of $J(\Omega,\Lambda)$ much in the spirit of the resampling GMS procedure in Andrews and Soares (2010), provided that the relevant values of θ are restricted to the argmin set of $Q_n(\theta)$. The role of the restriction is to guarantee that the resampling procedure does not consider excessive violations of the sample moment (in)equalities.

Definition 4.1. For T_n as in Eq. (3.3), the approximation to the identified set is given by

$$\hat{\Theta}_I \equiv \{\theta \in \Theta : Q_n(\theta) \leq T_n\}. \tag{4.3}$$

By definition, $\hat{\Theta}_I$ is the argmin set of $Q_n(\theta)$, which is non-empty. As in other M-estimation problems, it is not necessary to impose that $\hat{\Theta}_I$ is the set of exact minimizers of $Q_n(\theta)$. This set could be replaced with an "approximate" set of minimizers, i.e., $\hat{\Theta}_I \equiv \{\theta \in \Theta : Q_n(\theta) \leq T_n + o_p(1)\}$, without affecting our results. In addition, it is important to note that $\hat{\Theta}_I$ does not coincide with the consistent estimator of $\Theta_I(F)$ proposed in Chernozhukov et al. (2007, see p. 1247 and Theorem 3.1). In fact, $\hat{\Theta}_I$ is not generally consistent for $\Theta_I(F)$ in the Hausdorff distance, which is not a problem in our setting. All we need is for $\hat{\Theta}_I$ to lie in the expansion of $\Theta_I(F)$ specified in Definition 4.3 below.

Now we can define the resampling test statistic that we use to construct an approximation to $c_{(1-\alpha)}(\Lambda,\Omega)$. In order to do this, let $\hat{\Omega}_n(\theta) \equiv \hat{D}_n^{-1/2}(\theta)\hat{\Sigma}_n(\theta)\hat{D}_n^{-1/2}(\theta)$, where $\hat{D}_n(\theta) \equiv Diag(\hat{\Sigma}_n(\theta))$ and $\hat{\Sigma}_n(\theta)$ is as in Eq. (2.7). In addition, let $\{\hat{v}_n^*(\theta) : \theta \in \Theta\}$ be a stochastic process indexed by θ, whose conditional distribution given the original sample is known and can be simulated. For example, this can be done via a bootstrap approximation, in which case

$$\hat{v}_n^*(\theta) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\hat{D}_n^{-1/2}(\theta)(m(W_i^*,\theta) - \bar{m}_n(\theta)), \tag{4.4}$$

where $\{W_i^*\}_{i=1}^n$ is an i.i.d. sample drawn with replacement from original sample $\{W_i\}_{i=1}^n$, or via an asymptotic approximation, in which case

$$\hat{v}_n^*(\theta) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\hat{D}_n^{-1/2}(\theta)(m(W_i,\theta) - \bar{m}_n(\theta))\zeta_i, \tag{4.5}$$

and $\{\zeta_i\}_{i=1}^n$ is an i.i.d. sample satisfying $\zeta_i \sim N(0,1)$. Now consider the following test statistic

$$T_n^* \equiv \inf_{\theta\in\hat{\Theta}_I} S(\hat{v}_n^*(\theta) + \varphi(\xi_n(\theta), \hat{\Omega}_n(\theta)), \hat{\Omega}_n(\theta)), \tag{4.6}$$

and $\xi_n(\theta) = \{\xi_{n,j}(\theta)\}_{j=1}^p$ with $\xi_{n,j}(\theta)$ is as in Eq. (4.2), and $\varphi = (\varphi_1,\ldots,\varphi_p,\mathbf{0}_{k-p})' \in \mathbb{R}_{[+\infty]}^k$ is the function in Assumption M.1 that is assumed to satisfy the assumptions in Andrews and Soares (2010). Examples of φ include $\varphi_j(\xi,\Omega) = \infty 1\{\xi_j > 1\}$ (with the convention that $\infty 0 = 0$), $\varphi_j(\xi,\Omega) = \max\{\xi_j, 0\}$, and $\varphi_j(\xi,\Omega) = \xi_j$ for $j=1,\ldots,p$ (see Andrews and Soares, 2010 for other examples). Conditional on the sample, the distribution of T_n^* is known and its quantiles can be approximated by Monte Carlo simulation. This leads us to Test RS.

Definition 4.2 (*Test RS*)**.** The specification Test RS rejects the null hypothesis in Eq. (2.1) according to the following rejection rule

$$\phi_n^{RS} \equiv 1\left\{T_n > \hat{c}_n^{RS}(1-\alpha)\right\}, \tag{4.7}$$

where T_n is as in Eq. (3.3) and $\hat{c}_n^{RS}(1-\alpha)$ is a resampling approximation to the $(1-\alpha)$-quantile of T_n^*.

Remark 4.1. In the special case of point identified moment equality models, Test RS reduces to a standard J-test. In particular, if $S(\cdot)$ is the QLR test statistic it follows that $T_n = \inf_{\theta\in\Theta} n\bar{m}_n(\theta)'\hat{\Sigma}_n^{-1}(\theta)\bar{m}_n(\theta)$, so that Test RS is a J-test implemented with Continuously Updating GMM and a bootstrapped critical value.

The following result shows that the test proposed in Definition 4.2 is asymptotically level correct.

Theorem 4.1. *Let Assumptions A.1–A.7 hold. Then, for any $\alpha \in (0,1)$,*

$$\limsup_{n\to\infty}\sup_{F\in\mathcal{P}_0} E_F[\phi_n^{RS}] \leq \alpha. \tag{4.8}$$

Remark 4.2. Theorem 4.1 requires Assumption A.6 which is a high level assumption. In Lemma D.10 we show that Assumption A.8 is sufficient for Assumption A.6. Assumption A.8(a) states that $Q_F(\theta)$ can be bounded below in a neighborhood of the identified set $\Theta_I(F)$

and so it corresponds to the polynomial minorant condition in Chernozhukov et al. (2007, Eqs. (4.1) and (4.5)). The convexity in Assumption A.8(b) and the equicontinuity in Assumption A.8(c) are both used exclusively when applying the intermediate value theorem in the proof of Lemma D.10. Assumption A.8 is certainly easier to interpret than Assumption A.6, but it is mildly stronger.

In order to provide intuition for Theorem 4.1, it is convenient to re-write the test statistic T_n^* in a way that facilitates comparisons with the set $\Lambda_{n,F}$ defined in Eq. (3.6). This can be done by noting that

$$T_n^* = \inf_{(\theta,\ell)\in\hat{\Lambda}_n^*} S(\hat{v}_n^*(\theta)+\ell, \hat{\Omega}_n(\theta))$$

$$\text{where } \hat{\Lambda}_n^* = \left\{(\theta,\ell) : \theta \in \hat{\Theta}_I,\ \ell = \varphi(\xi_n(\theta), \hat{\Omega}_n(\theta))\right\}. \tag{4.9}$$

Test RS therefore consists in replacing the set Λ with the approximation $\hat{\Lambda}_n^*$, which is (generally) not consistent for Λ. The important aspect here is that $\hat{\Lambda}_n^*$ restricts $\theta \in \hat{\Theta}_I$ as opposed to $\theta \in \Theta$ in $\Lambda_{n,F}$. Since $\varphi_j(\cdot) \geq 0$ for $j = 1, \ldots, p$ and $\varphi_j(\cdot) = 0$ for $j = p+1, \ldots, k$, using such random set in the definition of T_n^* guarantees that the (in)equality restrictions are not violated by much when evaluated at the θ that approximates the infimum in Eq. (4.6). This makes the function $\varphi(\cdot)$ a valid replacement for ℓ and plays an important role in establishing the consistency in level of our test. In fact, if we were to define the set $\hat{\Lambda}_n^*$ with Θ instead of $\hat{\Theta}_I$, we would not obtain a test that controls asymptotic size as in Theorem 4.1 for functions φ satisfying Assumption A.5. In other words, using a similar statistic to T_n^* but with an infimum over Θ (as it is the case for the original test statistic) would not result in a valid asymptotic approximation.[8]

The result in Theorem 4.1 follows from arguments that use the following expansion of $\Theta_I(F)$.

Definition 4.3. Let $\Theta_I^{\ln\kappa_n}(F)$ be defined as

$$\Theta_I^{\ln\kappa_n}(F) \equiv \{\theta \in \Theta : S(\sqrt{n}E_F[m(W,\theta)], \Sigma_F(\theta)) \leq \ln\kappa_n\}.$$

Note that $\Theta_I^{\ln\kappa_n}(F)$ is a non-random expansion of $\Theta_I(F)$. Lemma D.13 in the Appendix shows that, asymptotically, our approximation of the identified set is included in this expansion uniformly over $\mathcal{P}_0$, i.e.,

$$\lim_{n\to\infty} \inf_{F\in\mathcal{P}_0} P_F(\hat{\Theta}_I \subseteq \Theta_I^{\ln\kappa_n}(F)) = 1. \tag{4.10}$$

Now consider an auxiliary random variable $\tilde{T}_n^*$, which is defined as T_n^* but with $\hat{\Lambda}_n^*$ replaced by

$$\left\{(\theta,\ell) : \theta \in \Theta_I^{\ln\kappa_n}(F_n), \ell = \varphi^*(\xi_n(\theta))\right\}, \tag{4.11}$$

where $\varphi^*(\cdot)$ is a continuous function that satisfies $\varphi(\cdot) < \varphi^*(\cdot,\Omega)$ for all Ω (see Assumption A.5). Notice that $\tilde{T}_n^*$ is a "hybrid" object in the sense that it depends both on the data sample and on unknown population parameters. It is more convenience to work $\tilde{T}_n^*$ rather than T_n^* for the following reasons: (a) $\tilde{T}_n^*$ uses a non-random set $\Theta_I^{\ln\kappa_n}(F_n)$ instead of the random set $\hat{\Theta}_I$ and (b) the function φ^* is continuous and does not depend on $\hat{\Omega}_n(\theta)$, which is not necessarily the case for the original GMS function φ. We denote by $J^*(\Lambda^*, \Omega)$ the conditional limiting distribution of $\tilde{T}_n^*$, which is characterized in Theorem C.1 in the Appendix.

Having defined these objects, Theorem 4.1 is the result of the following argument. We show first that

$$J(\Lambda,\Omega) \leq J^*(\Lambda^*,\Omega), \tag{4.12}$$

meaning that, asymptotically, the quantiles of T_n could be approximated by the infeasible conditional quantiles of $\tilde{T}_n^*$. Second, we note that Eq. (4.10) implies that, asymptotically,

$$\tilde{T}_n^* \leq T_n^*. \tag{4.13}$$

This allows us to replace the infeasible conditional quantiles of $\tilde{T}_n^*$ with the feasible conditional quantiles of T_n^*. In summary, Eqs. (4.12) and (4.13) ensure that, asymptotically, the resampling approximation of the $(1-\alpha)$-quantile of T_n^*, $\hat{c}_n^{RS}(1-\alpha)$, is a uniformly valid approximation to the $(1-\alpha)$-quantile of T_n. From this, the uniform asymptotic validity of Test RS follows.

Remark 4.3. The non-random expansion $\Theta_I^{\ln\kappa_n}(F_n)$ is used in intermediate steps of the proof of Theorem 4.1 but is *not needed* to implement Test RS.

Remark 4.4. The set in Eq. (4.11) assumes the existence of the function $\varphi^*(\cdot)$. This assumption is not restrictive as it is satisfied for the functions $\varphi^{(1)}(\cdot) - \varphi^{(4)}(\cdot)$ described in Andrews and Soares (2010) and Andrews and Barwick (2012) (see Remark B.1 in Appendix B).

Remark 4.5. To put the computational feasibility of Test RS into perspective, we compare it to the computation of confidence sets for θ, a problem that the literature has become familiar with. Typically, Test RS is easier than confidence set construction. To construct the confidence set as described in Section 2, one needs to compute $\hat{c}_n(\theta, 1-\alpha)$ for "enough" number of grid points on Θ. This is generally considered very difficult to do accurately unless Θ is low dimensional (3 dimensions or less). On the other hand, to implement Test RS, the challenging part is to compute T_n^* in Eq. (4.6) a large number of times (say, 1000 times), each time for a different simulation draw of $\hat{v}_n^*(\cdot)$. Although this amounts to solving a minimization problem accurately a large number of times, the task often is quite feasible because the objective functions to be minimized often are well-behaved, especially for smooth versions of $\varphi(\cdot)$.

5. Test RC: Re-cycling existent critical values

In practice, the researcher often needs to compute the confidence set $CS_n(1-\alpha)$ for reasons other than specification testing. In that case, it is reasonable to take the computation of the confidence set as given when implementing a model specification test. From this perspective, Test BP becomes more attractive than Test RS computation-wise because it is an immediate by-product of the confidence set construction. In this section, we propose a new specification test that involves a simple transformation of exactly the same critical values used for Test BP, therefore marginally increasing the computational effort. We call it the re-cycling test or Test RC precisely for the reason that it recycles existing critical values. Even with such a simple modification, Test RC presents power advantages over Test BP that we formalize in Section 6.

Definition 5.1 (*Test RC*)**.** The specification Test RC rejects the null hypothesis in Eq. (2.1) according to the following rejection rule

$$\phi_n^{RC} \equiv 1\left\{T_n > \hat{c}_n^{RC}(1-\alpha)\right\}, \tag{5.1}$$

[8] We note that a special choice of the function φ can be shown to circumvent the problem and result in a test that controls asymptotic size (see Bugni et al., 2014 Remark 2.2 for details). However, such function does not belong to the class of functions considered in Andrews and Soares (2010) and thus not suitable for the type of power comparisons we study in this paper.

where T_n is as in Eq. (3.3), $\hat{c}_n^{RC}(1-\alpha)$ is given by

$$\hat{c}_n^{RC}(1-\alpha) = \inf_{\theta \in \hat{\Theta}_I} \hat{c}_n(\theta, 1-\alpha), \tag{5.2}$$

where $\hat{\Theta}_I$ is as in Eq. (4.3) and $\hat{c}_n(\theta, 1-\alpha)$ is the GMS critical value used by Test BP, see Assumption M.1.

Remark 5.1. Test BP requires computation of the sample criterion function $Q_n(\theta)$ and the GMS quantile $\hat{c}_n(\theta, 1-\alpha)$ for every $\theta \in \Theta$. With this information in hand, it is relatively easy to compute the approximation to the identified set $\hat{\Theta}_I$. Thus, relative to Test BP, implementing Test RC requires little additional work.

Remark 5.2. Test RC is defined as a test whose critical value is the minimum of the critical values used by Test BP (c.f. Eq. (5.2)). This implies that Test RC and Test BP are implemented with the *same choice* of GMS function $\varphi(\cdot)$ and tuning parameter κ_n, as Test RC inherits this choice from Test BP. We use this fact in the power comparisons of Section 6.

Remark 5.3. For a given $\theta \in \Theta$, the GMS quantile $\hat{c}_n(\theta, 1-\alpha)$ coincides with the $(1-\alpha)$-quantile of the random variable $J_n(\theta) \equiv S(\hat{v}_n^*(\theta) + \varphi(\xi_n(\theta), \hat{\Omega}_n(\theta)), \hat{\Omega}_n(\theta))$, used in to implement Test RS.[9] From this observation, it follows that the critical value of Test RS is the $(1-\alpha)$-quantile of the infimum of $J_n(\theta)$ over $\hat{\Theta}_I$, while the critical value of Test RC is the infimum of the $(1-\alpha)$-quantiles of $J_n(\theta)$ over $\hat{\Theta}_I$. Since the quantile of an infimum is weakly smaller than the infimum of the quantiles, we deduce that

$$\hat{c}_n^{RS}(1-\alpha) \leq \inf_{\theta \in \hat{\Theta}_I} \hat{c}_n(\theta, 1-\alpha) = \hat{c}_n^{RC}(1-\alpha). \tag{5.3}$$

The following result shows that the test proposed in Definition 4.2 is asymptotically level correct and it is an immediate consequence of Theorem 4.1 and Eq. (5.3).

Theorem 5.1. *Let Assumptions A.1–A.7 hold. Then, for any $\alpha \in (0,1)$,*

$$\limsup_{n\to\infty} \sup_{F\in\mathcal{P}_0} E_F[\phi_n^{RC}] \leq \alpha. \tag{5.4}$$

Remark 5.4. Theorems 4.1 and 5.1 show that Test RS and Test RC are asymptotically level correct but are silent about the type of conditions that could make these test asymptotically non-conservative. Unfortunately, we could not find such conditions with sufficient level of generality.

6. Power analysis

Previous results reveal that the existing test, Test BP, and the ones proposed in this paper, Test RS and Test RC, are all asymptotically level correct. The goal of this section is to compare these procedures in terms of power. We show that the tests proposed in this paper have weakly more power than Test BP in all finite samples, and there are sequences of local alternative hypotheses for which they have strictly higher asymptotic power. We open the section with the finite sample findings.

Theorem 6.1. *For any $(n, F) \in \mathbb{N} \times \mathcal{P}$,*

$$\phi_n^{RS} \geq \phi_n^{RC} \geq \phi_n^{BP}.$$

[9] See the proof of Theorem 6.1 for the details.

Corollary 6.1. *For any sequence of local alternatives $\{F_n \in \mathcal{P}/\mathcal{P}_0\}_{n\geq1}$,*

$$\liminf_{n\to\infty}(E_{F_n}[\phi_n^{RS}] - E_{F_n}[\phi_n^{RC}]) \geq 0, \quad \textit{and}$$

$$\liminf_{n\to\infty}(E_{F_n}[\phi_n^{RC}] - E_{F_n}[\phi_n^{BP}]) \geq 0.$$

The proof of Theorem 6.1 is in Appendix C and Corollary 6.1 follows directly from Theorem 6.1. Note that Theorem 6.1 is a statement that holds for all $n \in \mathbb{N}$ and $F \in \mathcal{P}$. This is not only a finite sample power result, but it is also a relationship that holds for distributions $F \in \mathcal{P}_0$. It follows that the two tests we propose cannot be more conservative than the existing Test BP.

Remark 6.1. The tests considered in this paper are only shown to control size *asymptotically*. Thus, for any distribution $F \in \mathcal{P}_0$ and any sample size n, it is certainly possible that all of these tests over-reject the null hypothesis, i.e., $E_F[\phi_n^s] > \alpha$ for $s = \{BP, RC, RS\}$. In any case, the fact that these tests provide asymptotic uniform size control means that, for any $\epsilon > 0$, there exists a sample size $N(\epsilon)$ (not dependent on F) such that for all $n \geq N(\epsilon)$,

$$E_F[\phi_n^s] \leq \alpha + \epsilon \quad \text{for } s = \{BP, RC, RS\}. \tag{6.1}$$

In other words, to the extent that the sample size is reasonably large, the amount of over-rejection of all these test is *uniformly* bounded.

Remark 6.2. It would be ideal to have finite sample results for *both* size and power. Unfortunately, constructing test with finite sample size control in this type of problems is extremely hard, which explains why the literature resorts exclusively to asymptotic approximations. In light of this, we view the result in Theorem 6.1 as particularly important and novel, especially if we take into account that it requires no assumptions beyond the maintained ones.

Remark 6.3. The first inequality in Theorem 6.1 uses that Test RS and Test RC are implemented with the same choice of GMS function $\varphi(\cdot)$ and the same tuning parameter κ_n. We recommend this practice as these objects play exactly the same role in all of these tests. The second inequality follows by definition, as Test RC and Test BP share the same $\varphi(\cdot)$ and κ_n by construction (see Remark 5.2).[10]

Theorem 6.1 and Corollary 6.1 show that Test BP will never do better (in terms of power or asymptotic conservativeness) than Test RS or Test RC. However, there is nothing that prevents a situation in which all these tests provide exactly the same power. The last result in this section therefore provides a type of local alternatives for which both of our tests have strictly higher asymptotic power than Test BP. The result relies on the following condition.

Assumption A.9. For T_n as in Eq. (3.3) and $\hat{\Theta}_I$ as in Eq. (4.3), $\{F_n \in \mathcal{P}\}_{n\geq1}$ satisfies the following:

(i) There is a (possibly random) sequence $\{\theta_n^* \in \hat{\Theta}_I\}_{n\geq1}$ such that $\hat{c}_n(\theta_n^*, 1-\alpha) \xrightarrow{p} c_H$,
(ii) There is a (possibly random) sequence $\{\theta_n \in \hat{\Theta}_I\}_{n\geq1}$ such that $\hat{c}_n(\theta_n, 1-\alpha) \xrightarrow{p} c_L$,
(iii) $T_n \xrightarrow{d} J$ and $P(J \in (c_L, c_H)) > 0$.

[10] As a referee pointed out, if Test BP were to be implemented with a κ_n smaller than the one used by Test RC, it is possible that Test BP delivers higher finite sample power than Test RC. This does not contradict our results, as they hold for the same choice of κ_n. Importantly, our results do not rely on the fact that the power of these tests is decreasing in κ_n.

We illustrate how Assumption A.9 holds in the context of Example 2.1 below, where we have simplified the example slightly to keep the derivations as short as possible.[11]

Example 6.1. Let $W = (W_1, W_2, W_3) \in \mathbb{R}^3$ be a random vector with distribution F_n, $V_{F_n}[W] = I_3$, $E_{F_n}[W_1] = 0$, $E_{F_n}[W_2] = -\mu/\sqrt{n}$, and $E_{F_n}[W_3] = 0$ for some $\mu \in \mathbb{R}$. Consider the following model with $\Theta = [-B, B]^2$ for some $B > 0$,

$$\begin{aligned} E_{F_n}[m_1(W_i,\theta)] &= E_{F_n}[\theta_1 - W_{i,1}] \geq 0, \\ E_{F_n}[m_2(W_i,\theta)] &= E_{F_n}[W_{i,2} - \theta_1] \geq 0, \\ E_{F_n}[m_3(W_i,\theta)] &= E_{F_n}[\theta_2 - W_{i,3}] \geq 0. \end{aligned} \tag{6.2}$$

The identified set is $\Theta_I(F_n) = \{\theta \in \Theta : \theta_1 \in [0, -\mu/\sqrt{n}], \theta_2 \geq 0\}$, which is non-empty if and only if $\mu \leq 0$. This identified set has the same structure as in Example 2.1 with $E_{F_n}[YZ|X = x_1] = 0$, $E_{F_n}[YZ|X = x_2] = 1 + \mu/\sqrt{n}$, and $E_{F_n}[YZ|X = x_3] = 0$.

The model in Eq. (6.2) is linear in θ, and hence many relevant parameters and estimators do not depend on θ. These include $\hat{\sigma}_j(\theta) = \hat{\sigma}_j$ for $j = 1, 2, 3$, $\hat{D}_n^{-1/2}(\theta) = \hat{D}_n^{-1/2}$, $\tilde{v}_{n,j}(\theta) = \tilde{v}_{n,j} = \sqrt{n}\hat{\sigma}_j^{-1}(E_{F_n}[W_j] - \bar{W}_{n,j})$ for $j = 1, 3$, $\tilde{v}_{n,2} = \sqrt{n}\hat{\sigma}_2^{-1}(\bar{W}_{n,2} - E_{F_n}[W_2])$, and

$$v_n^*(\theta) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \hat{D}_n^{-1/2}(\theta)(m(W_i,\theta) - \bar{m}_n(\theta))\zeta_i = v_n^*, \tag{6.3}$$

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 (MMM) criterion function given by

$$S(m, \Sigma) = \sum_{j=1}^{p}[m_j/\sigma_j]_-^2 + \sum_{j=p+1}^{k}(m_j/\sigma_j)^2, \tag{6.4}$$

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

$$\begin{aligned} &\varphi_j(x) = \infty 1\{x > 1\} \quad \text{for } j = 1, \ldots, p \quad \text{and} \\ &\varphi_j(x) = 0 \quad \text{for } j = p+1, \ldots, k. \end{aligned} \tag{6.5}$$

The sample criterion function is given by

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

It is easy to verify Assumptions A.1–A.7 in this context. We now explicitly verify Assumption A.9. To do this, we 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 the notation $Z \sim N(\mathbf{0}_4, I_4)$.

Assumption A.9 (i): The set of minimizers of $Q_n(\theta)$ over Θ is $\hat{\Theta}_I \supseteq \{\theta_n^* : \theta_{n,1}^* = (\hat{\sigma}_1^{-1}\bar{W}_1 + \hat{\sigma}_2^{-1}\bar{W}_2)/(\hat{\sigma}_1^{-1} + \hat{\sigma}_2^{-1}), \theta_{n,2}^* \in [\bar{W}_3, B]\}$. Let us take $\theta_{n,2}^* = \bar{W}_3$ for concreteness. Notice that the sequence $\{\theta_n^* \in \hat{\Theta}_I\}_{n\geq 1}$ is random, which is allowed by condition (i). Simple algebra shows that

$$\begin{aligned} \sqrt{n}\theta_{n,1}^* &= \frac{\tilde{v}_{n,2} - \tilde{v}_{n,1}}{2} + \frac{\sqrt{n}E_{F_n}[W_2]}{2} + o_p(1) \\ &= \frac{\tilde{v}_{n,2} - \tilde{v}_{n,1}}{2} - \frac{\mu}{2} + o_p(1). \end{aligned} \tag{6.6}$$

The test statistic T_n in Eq. (3.3) therefore satisfies

$$\begin{aligned} T_n = Q_n(\theta_n^*) &= \left[\tilde{v}_{n,1} + \sqrt{n}\hat{\sigma}_1^{-1}\theta_{n,1}^*\right]_-^2 \\ &\quad + \left[\tilde{v}_{n,2} + \sqrt{n}\hat{\sigma}_2^{-1}(E_{F_n}[W_2] - \theta_{n,1}^*)\right]_-^2 , \\ &= \left[\tilde{v}_{n,1} + \frac{\tilde{v}_{n,2} - \tilde{v}_{n,1}}{2} - \frac{\mu}{2} + o_p(1)\right]_-^2 \\ &\quad + \left[\tilde{v}_{n,2} - \frac{\tilde{v}_{n,2} - \tilde{v}_{n,1}}{2} + \frac{\mu}{2} - \mu + o_p(1)\right]_-^2 , \\ &= 2\left[\frac{\tilde{v}_{n,1} + \tilde{v}_{n,2}}{2} - \frac{\mu}{2} + o_p(1)\right]_-^2 \xrightarrow{d} J \\ &\equiv \left[Z_1 - \frac{\mu}{\sqrt{2}}\right]_-^2 . \end{aligned} \tag{6.7}$$

Next note that the GMS critical value along θ_n^* is the (conditional) $(1 - \alpha)$-quantile of

$$\begin{aligned} Q_n^*(\theta_n^*) = &\left[v_{n,1}^* + \infty 1\{\kappa_n^{-1}\tilde{v}_{n,1} + \kappa_n^{-1}\sqrt{n}\hat{\sigma}_1^{-1}\theta_{n,1}^* > 1\}\right]_-^2 \\ &+ \left[v_{n,2}^* + \infty 1\{\kappa_n^{-1}\tilde{v}_{n,2} - \kappa_n^{-1}\sqrt{n}\hat{\sigma}_2^{-1}\theta_{n,1}^*\right. \\ &\left. - \kappa_n^{-1}\hat{\sigma}_2^{-1}\mu > 1\}\right]_-^2 + \left[v_{n,3}^*\right]_-^2 , \\ &\xrightarrow{d} [Z_2]_-^2 + [Z_3]_-^2 + [Z_4]_-^2 \quad \text{w.p.a.1}, \end{aligned}$$

since $\kappa_n^{-1}\sqrt{n}\theta_{n,1}^* \xrightarrow{p} 0$ by Eq. (6.6). If we let c_H denote the $(1 - \alpha)$-quantile of RHS of the previous display, it follows that $\hat{c}_n(\theta_n^*, 1 - \alpha) \xrightarrow{p} c_H$ and condition (i) holds.

Assumption A.9 (ii): Let $\theta_n = (\theta_{n,1}^*, \theta_{n,2})$ where $\theta_{n,2} = \bar{W}_3 + C\kappa_n/\sqrt{n}$ for $C > 1$. As before, the sequence $\{\theta_n \in \hat{\Theta}_I\}_{n\geq 1}$ is random, which is allowed by condition (ii). The GMS critical value evaluated at θ_n is the (conditional) $(1 - \alpha)$-quantile of

$$\begin{aligned} Q_n^*(\theta_n) = &\left[v_{n,1}^* + \infty 1\{\kappa_n^{-1}\tilde{v}_{n,1} + \kappa_n^{-1}\sqrt{n}\hat{\sigma}_1^{-1}\theta_{n,1}^* > 1\}\right]_-^2 \\ &+ \left[v_{n,2}^* + \infty 1\{\kappa_n^{-1}\tilde{v}_{n,2} - \kappa_n^{-1}\sqrt{n}\hat{\sigma}_2^{-1}\theta_{n,1}^*\right. \\ &\left. - \kappa_n^{-1}\hat{\sigma}_2^{-1}\mu > 1\}\right]_-^2 + \left[v_{n,3}^* + \infty 1\{\hat{\sigma}_3^{-1}C > 1\}\right]_-^2 , \\ &\xrightarrow{d} [Z_2]_-^2 + [Z_3]_-^2 \quad \text{w.p.a.1}, \end{aligned}$$

since $\kappa_n^{-1}\sqrt{n}\theta_{n,1}^* \xrightarrow{p} 0$ by Eq. (6.6) and $C > 1$. If we let c_L denote the $(1 - \alpha)$-quantile of RHS of the previous display, it follows that $\hat{c}_n(\theta_n, 1 - \alpha) \xrightarrow{p} c_L$ and condition (ii) holds.

Assumption A.9 (iii): $T_n \xrightarrow{d} J$ follows from Eq. (6.7). In addition, $c_L < c_H$ is immediate from the previous derivations whenever $\alpha < 50\%$. For example, when $\alpha = 10\%$, we have $c_L = 2.95$ and $c_H = 4.01$. Finally, $P(J \in (c_L, c_H)) > 0$ holds as the distribution of J is continuous at $x > 0$. For example, when $\alpha = 10\%$ and $\mu = 2$, $P(J \in (c_L, c_H)) = 11\%$. We conclude that condition (iii) holds. □

Loosely speaking, Assumption A.9 considers a sequence of local alternatives where the set of minimizers $\hat{\Theta}_I$ includes at least two points for which the quantiles of the limit distribution of $Q_n^*(\theta)$ are different. In Example 6.1, the critical value along the sequence θ_n^* has three moments binding, while the critical value under the sequence θ_n has two moments binding. It follows that the GMS critical values satisfy $\hat{c}_n(\theta_n^*, 1 - \alpha) > \hat{c}_n(\theta_n, 1 - \alpha)$ with high probability as n gets large. At the same time, these sequences are such that $T_n = \inf_{\theta\in\Theta} Q_n(\theta) = Q_n(\theta_n^*) = Q_n(\theta_n)$. Putting all this together, we can informally anticipate the result in Theorem 6.2 as

[11] It is worth emphasizing that Example 6.1 is the simplest example we can construct to generate strict power differences between the specification tests. One could use more sophisticated and realistic examples (e.g. including non-linear moment conditions or moment equalities) to generate similar results but this additional complexity would severely complicate the derivations.

follows,

$$\begin{aligned}\phi_n^{BP} &= 1\{\forall\theta \in \Theta : Q_n(\theta) > \hat{c}_n(\theta, 1-\alpha)\}\\ &\leq 1\{Q_n(\theta_n^*) > \hat{c}_n(\theta_n^*, 1-\alpha)\}\langle 1\{Q_n(\theta_n)\rangle \hat{c}_n(\theta_n, 1-\alpha)\}\\ &\leq 1\{T_n > \inf_{\theta\in\hat{\Theta}_I} \hat{c}_n(\theta, 1-\alpha)\} = \phi_n^{RC},\end{aligned}$$

where the strict inequality holds with positive probability. Assumption A.9 is satisfied whenever the set of minimizers of $Q_n(\theta)$ is not a singleton and the limiting distribution of $Q_n^*(\theta)$ is not the same along the different sequences of minimizers. On the other hand, the assumption does not hold if the local alternatives are such that the argmin set of $Q_n(\cdot)$ converges to a singleton (e.g. point identification).

Theorem 6.2. *For any sequence of local alternatives* $\{F_n \in \mathcal{P}/\mathcal{P}_0\}_{n\geq 1}$ *that satisfies Assumption A.9,*

$$\liminf_{n\to\infty}(E_{F_n}[\phi_n^{RC}] - E_{F_n}[\phi_n^{BP}]) > 0.$$

Theorem 6.2 shows that Test RC is asymptotically strictly more powerful than Test BP for sequence of alternatives satisfying Assumption A.9. Combining this result with Theorem 6.1, it follows that Test RS is also strictly more powerful than Test BP asymptotically.

Remark 6.4. We can use Example 6.1 to illustrate the asymptotic power gains. For $\mu = 2$ and $\alpha = 10\%$, the asymptotic local power of Test RC and Test BP are 38.1% and 27.8%, respectively. Clearly, the power differences could be significant. In addition, the same example illustrates how all these tests could be asymptotically conservative. For $\mu = 0$ and $\alpha = 10\%$, the asymptotic size of Test RC and Test BP are 4.4% and 2.2%, respectively, which are consistent with the simulation results in Section 7.

Remark 6.5. We can also use Example 6.1 to show that a modification of Test RS that replaces $\hat{\Theta}_I$ by Θ in Eq. (4.6) would not control asymptotic size. In this case, simple algebra shows that, conditionally

$$\begin{aligned}\inf_{\theta\in\Theta} Q_n^*(\theta) &\leq \min\{Q_n^*(\theta_n^L), Q_n^*(\theta_n^H)\} = \min\{[v_{n,1}^*]_-^2, [v_{n,2}^*]_-^2\}\\ &\xrightarrow{d} \min\{[Z_2]_-^2, [Z_3]_-^2\} \quad \text{w.p.a.1},\end{aligned} \tag{6.8}$$

where $\theta_n^L = (\theta_{n,1}^L, \theta_{n,2})$ and $\theta_n^H = (\theta_{n,1}^H, \theta_{n,2})$ are two sequences in Θ such that

$$\theta_{n,1}^L = \bar{W}_2 - \frac{2\kappa_n}{\sqrt{n}}, \qquad \theta_{n,1}^H = \bar{W}_1 + \frac{2\kappa_n}{\sqrt{n}},$$
$$\text{and} \quad \theta_{n,2} = \bar{W}_3 + \frac{2\kappa_n}{\sqrt{n}}. \tag{6.9}$$

Note that the $(1-\alpha)$-quantile of $\min\{[Z_2]_-^2, [Z_3]_-^2\}$ is smaller than the $(1-\alpha)$-quantile of J in Eq. (6.7) and, thus, this *modified* Test RS suffers from over-rejection. For $\mu = 0$ and $\alpha = 10\%$, the asymptotic size of is 31.70%. The Test RS from Definition 4.2 avoids this problem by restricting θ to $\hat{\Theta}_I$ and thus guaranteeing that sequences with values of θ_1 that are "too" big or "too" small (like those in Eq. (6.9)) are not feasible.

Remark 6.6. If one considers sequences of alternatives under which the inequality in Eq. (5.3) becomes strict (asymptotically), it is then possible that Test RS becomes strictly more powerful than Test RC.

Remark 6.7. Test BP requires fewer assumptions to obtain asymptotic size control than the tests we propose here. It is fair to say then that Test BP is more "robust" than Test RC and Test RS, in the sense that if some of the Assumptions A.1–A.7 fail, Test BP would still control asymptotic size.

7. Monte Carlo simulations

We now present Monte Carlo simulations that illustrate the finite sample properties of the specification tests considered in this paper. We simulate data according to the simple parametrization presented in Example 2.1, i.e., Eq. (2.14). The data $\{W_i\}_{i=1}^n$ are i.i.d., where $W_i \equiv (Y_iZ_i, Z_i, X_i)$ is distributed such that

$$\begin{aligned}&\{Y_iZ_i|X = x_1\} \sim N(0, 1), \qquad \{Y_iZ_i|X = x_2\} \sim N(1+\eta_n, 1),\\ &\{Y_iZ_i|X = x_3\} \sim N(0, 1),\end{aligned} \tag{7.1}$$

for $\eta_n = \eta/n^{1/2} \in \mathbb{R}$, and $P(X_i = x_s) = 1/3$ for $s \in \{1, 2, 3\}$. By plugging in this information into Eq. (2.14), we get

$$\Theta_I(F) = \{(\theta_1, \theta_2) \in \Theta : \theta_1 \in [0, -\eta_n],\ \theta_2 \geq 0\}. \tag{7.2}$$

The parameter $\eta \in \mathbb{R}$ measures the amount of model misspecification. On the one hand, $\eta \leq 0$ implies that the model is correctly specified and strictly partially identified, i.e., the identified set includes multiple values. On the other hand, $\eta > 0$ implies that the model is misspecified, i.e., the identified set is empty.

The simulation results are collected in Tables 1 and 2. The parameters we use to produce both tables are as follows: $\alpha = 10\%$, $n \in \{100, 1000\}$, $\kappa_n = C\sqrt{\ln n}$ for $C \in \{0.01, 0.1, 0.8, 0.9, 1, 10\}$, $\varphi(\cdot)$ as in Eq. (6.5), and $S(\cdot)$ as in Eq. (6.4).[12] The number of replications is set to 5000.

The simulation results are consistent with the theoretical findings. Under the null hypothesis (i.e. $\eta = 0$) all tests are asymptotically level correct (i.e. the asymptotic rejection rate does not exceed α). In fact, Remark 6.4 shows that these tests are asymptotically conservative in this example, which is consistent with the results in Tables 1 and 2. Under the alternative hypothesis (i.e. $\eta > 0$) the rejection rates increase monotonically with the amount of misspecification, measured by η. Comparing the rejection rates across methods, we see that Test RS shows better power than Test RC, and that Test RC has better power than Test BP. The differences can be substantial, with the power of Test RS being almost twice the power of Test BP for some alternatives (e.g. $\eta = 0.3$ and $n = 100$). These results are consistent with Theorem 6.1 and the analytical derivations in Example 6.1. However, as a referee pointed out, if Test BP were to be implemented with a smaller κ_n than the one used for our tests, it could deliver similar finite sample power.[13] Table 2 illustrates this possibility by comparing Test BP with $C = 0.1$ and Test RS with $C = 10$. In this case, both tests have the same rejection probability for $\eta = 0$ (i.e. 3.52) and similar power for all values of $\eta > 0$. This does not contradict our results, as they are established for the same choice of κ_n, see Remarks 5.2 and 6.3.[14]

8. Conclusions

This paper studies the problem of specification testing in partially identified models defined by a finite number of moment (in)equalities. Under the null hypothesis of the test, there is at least one parameter value that simultaneously satisfies all of the moment (in)equalities whereas under the alternative hypothesis of the test there is no such parameter value. While this problem has not been directly addressed in the literature (except in particular

[12] Additional simulations for $C \in \{0.05, 0.5\}$ show similar results and are therefore omitted. Note that for $n = 100$, the parameter κ_n ranges from 0.021 (for $C = 0.01$) to 21.4 (for $C = 10$).

[13] This is why we consider such a wide range for the parameter κ_n in our simulations.

[14] In Example 2.1 one can show that the power gains of Test RS over Test RC vanish asymptotically. However, there are models in which Test RS has strictly higher asymptotic power than Test RC. We have constructed one such example and the results of the simulations are available upon request.

Table 1
Rejection rate (in %) of Test BP, Test RC, and Test RS for the model in Eq. (7.2). Parameter values are $n = 100$, $\alpha = 10\%$, $\kappa_n = C\sqrt{\ln n}$. Results based on 5000 Monte Carlo replications.

C	Method	η										
		0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
	BP	1.88	2.18	3.24	5.32	9.24	16.52	28.88	47.68	69.92	88.78	97.50
0.01	RC	3.60	4.06	5.68	8.74	13.80	23.08	38.78	58.06	78.36	92.42	98.48
	RS	3.62	4.22	5.62	8.64	13.80	22.84	38.58	58.14	78.28	92.56	98.46
	BP	1.82	2.08	3.20	5.18	9.04	16.26	28.42	47.02	69.26	88.38	97.40
0.1	RC	3.58	4.00	5.64	8.66	13.76	22.98	38.64	57.96	78.26	92.38	98.48
	RS	3.60	4.18	5.58	8.56	13.78	22.72	38.42	58.02	78.20	92.52	98.46
	BP	1.76	2.02	3.08	4.90	8.86	15.72	27.94	46.44	68.66	88.14	97.22
0.8	RC	3.32	3.70	5.32	8.42	13.24	22.46	37.88	56.60	77.52	92.02	98.42
	RS	3.36	3.84	5.24	8.36	13.30	22.16	37.62	56.68	77.52	92.16	98.40
	BP	1.76	2.02	3.06	4.88	8.84	15.72	27.94	46.40	68.66	88.12	97.18
0.9	RC	3.28	3.64	5.32	8.40	13.18	22.30	37.72	56.44	77.44	91.98	98.40
	RS	3.30	3.80	5.24	8.32	13.24	22.04	37.42	56.54	77.44	92.12	98.40
	BP	1.76	2.02	3.06	4.86	8.80	15.72	27.94	46.38	68.66	88.10	97.18
1	RC	3.24	3.60	5.32	8.32	13.08	22.24	37.66	56.30	77.34	91.94	98.38
	RS	3.26	3.76	5.22	8.24	13.20	22.00	37.36	56.40	77.32	92.08	98.38
	BP	1.76	1.98	3.02	4.84	8.78	15.66	27.80	46.38	68.58	88.02	97.16
10	RC	1.82	2.02	3.08	5.02	8.86	15.86	28.26	46.90	69.06	88.32	97.30
	RS	1.76	2.06	3.14	5.04	8.70	15.82	28.30	47.10	69.08	88.26	97.34

Table 2
Rejection rate (in %) of Test BP, Test RC, and Test RS for the model in Eq. (7.2). Parameter values are $n = 1000$, $\alpha = 10\%$, $\kappa_n = C\sqrt{\ln n}$. Results based on 5000 Monte Carlo replications.

C	Method	η										
		0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
	BP	3.62	4.18	5.48	8.24	13.08	22.02	35.82	56.24	75.36	90.82	97.92
0.01	RC	4.24	4.96	6.34	9.20	14.32	24.08	38.66	58.74	77.72	91.82	98.28
	RS	4.30	4.92	6.32	9.16	14.30	24.20	38.82	58.82	77.84	91.86	98.24
	BP	3.52	4.06	5.34	7.80	12.80	21.44	35.26	55.62	74.80	90.44	97.78
0.1	RC	4.24	4.96	6.34	9.20	14.32	24.08	38.62	58.70	77.70	91.80	98.26
	RS	4.30	4.92	6.32	9.16	14.30	24.20	38.78	58.78	77.82	91.86	98.24
	BP	2.62	3.04	4.52	6.34	10.34	17.88	30.80	50.36	71.12	88.42	97.16
0.8	RC	4.20	4.94	6.32	9.14	14.26	24.00	38.58	58.64	77.66	91.70	98.22
	RS	4.28	4.90	6.30	9.10	14.24	24.14	38.70	58.72	77.74	91.78	98.22
	BP	2.60	3.00	4.52	6.26	10.32	17.76	30.68	50.22	71.10	88.36	97.16
0.9	RC	4.20	4.94	6.32	9.14	14.26	24.00	38.58	58.62	77.66	91.70	98.22
	RS	4.28	4.90	6.30	9.10	14.24	24.14	38.70	58.70	77.74	91.78	98.22
	BP	2.58	2.98	4.44	6.20	10.10	17.72	30.58	50.16	70.96	88.34	97.12
1	RC	4.20	4.94	6.32	9.10	14.22	23.96	38.58	58.56	77.62	91.70	98.22
	RS	4.28	4.90	6.30	9.06	14.20	24.10	38.68	58.62	77.70	91.78	98.22
	BP	2.10	2.50	3.66	5.58	9.14	16.40	28.62	47.26	68.82	86.98	96.80
10	RC	3.44	4.00	5.34	8.02	12.18	21.66	35.16	54.48	74.22	90.12	97.64
	RS	3.52	4.00	5.38	7.98	12.44	21.64	35.28	54.72	74.26	90.08	97.64

cases), several papers in the literature have suggested addressing it by checking whether confidence sets for the parameters of interest are empty or not. We refer to this procedure as Test BP.

We propose two new specification tests that achieve uniform asymptotic size control, which we refer to as Test RS and Test RC. Both of these tests dominate Test BP in terms of power. In particular, we show that Test RS and Test RC have more or equal power than Test BP in all finite samples, and we characterize sequences of local alternative hypotheses for which they have strictly higher asymptotic power. Our numerical results reveal that these power differences can be substantial, even in small sample sizes.

This paper also compares these specification tests in terms of their computational costs. By definition, Test BP requires the computation of a confidence set for the parameter of interest. If one is willing to compute this confidence set, then Test RC is particularly convenient: it requires almost no additional work and can potentially lead to significant power gains vis-à-vis Test BP. On the other hand, implementing Test RS requires a separate resampling procedure that is typically easier than the computation of the confidence sets (especially in high dimensional problems encountered in practice). In reward for this extra computation, Test RS can lead to power gains relative to the other two procedures.

We point out that the methodological contributions in this paper can be used to address a wide range of inferential problems that are different from specification testing. In particular, Bugni et al. (2014) use inferential procedures along the lines of Test RS to conduct inference on functions of partially identified parameters in a moment (in)equality model, i.e.,

$$H_0 : f(\theta_0) = \gamma_0 \quad \text{vs.} \quad H_1 : f(\theta_0) \neq \gamma_0, \tag{8.1}$$

where θ_0 now denotes the true parameter value in the moment inequality model, f is a known function, and γ_0 is an arbitrary number. We also point out that there are other interesting extensions that we did not pursue. First, our paper does not consider conditional moment restrictions, c.f. Andrews and Shi (2013); Chernozhukov et al. (2013); Armstrong (2014), 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); Romano et al. (2014), and McCloskey (2014). These two extensions are well beyond the scope of this paper and are left for future research.

Appendix A. Notation

Throughout the Appendix we use the following 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 $\mathbb{R}_{[\pm\infty]} = \mathbb{R} \cup \{\pm\infty\}$. For any $u \in \mathbb{N}$, we equip $\mathbb{R}^u_{[\pm\infty]}$ with the following metric d. For any $x_1, x_2 \in \mathbb{R}^u_{[\pm\infty]}$,

$$d(x_1, x_2) = \left(\sum_{i=1}^{u} (\vartheta(x_{1,i}) - \vartheta(x_{2,i}))^2 \right)^{1/2}, \tag{A.1}$$

where $\vartheta : \mathbb{R}_{[\pm\infty]} \to [0, 1]$ is such that $\vartheta(-\infty) = 0$, $\vartheta(\infty) = 1$, and $\vartheta(y) = \Phi(y)$ for $y \in \mathbb{R}$, where Φ is the standard normal CDF. Finally, $\hat{D}_n(\theta) \equiv Diag(\hat{\Sigma}_n(\theta))$, $\hat{\Omega}_n(\theta) \equiv \hat{D}_n^{-1/2}(\theta)\hat{\Sigma}_n(\theta)\hat{D}_n^{-1/2}(\theta)$, $v_n(\theta) \equiv \sqrt{n} D_F^{-1/2}(\theta)(\bar{m}_n(\theta) - E_F[m(W, \theta)])$, $\tilde{v}_n(\theta) \equiv \sqrt{n} \hat{D}_n^{-1/2}(\theta)(\bar{m}_n(\theta) - E_F[m(W, \theta)])$, and $v_n^*(\theta)$ is defined as $\hat{v}_n^*(\theta)$ in Eqs. (4.4) and (4.5) with $D_F^{-1/2}(\theta)$ replacing $\hat{D}_n^{-1/2}(\theta)$.

Remark A.1. The space $(\mathbb{R}^u_{[\pm\infty]}, d)$ constitutes a compact metric space. Also, if a sequence in $(\mathbb{R}^u_{[\pm\infty]}, d)$ converges to an element in $\mathbb{R}^u$, such a sequence will also converge in $(\mathbb{R}^u, \|\cdot\|)$, where $\|\cdot\|$ denotes the Euclidean norm.

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 to d, i.e., for any sets $A, B \in \Theta \times \mathbb{R}^k_{[\pm\infty]}$,

$$d_H(A, B) \equiv \max\left\{ \sup_{(\theta_1,h_1)\in A} \inf_{(\theta_2,h_2)\in B} d((\theta_1, h_1), (\theta_2, h_2)), \right.$$
$$\left. \sup_{(\theta_2,h_2)\in B} \inf_{(\theta_1,h_1)\in A} d((\theta_1, h_1), (\theta_2, h_2)) \right\}.$$

We use "$\xrightarrow{H}$" to denote convergence in the Hausdorff metric, i.e., $A_n \xrightarrow{H} B \iff d_H(A_n, B) \to 0$. Finally, for non-stochastic functions of $\theta \in \Theta$, we use "$\xrightarrow{u}$" to denote uniform in θ convergence, e.g., $\Omega_{F_n} \xrightarrow{u} \Omega \iff \sup_{\theta,\theta' \in \Theta} d(\Omega_{F_n}(\theta, \theta'), \Omega(\theta, \theta')) \to 0$. 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 $\{F_n \in \mathcal{P}\}_{n\geq1}$ determine a sequence of probability spaces $\{(\mathcal{W}, \mathcal{A}, F_n)\}_{n\geq1}$. Stochastic processes are then random maps $X : \mathcal{W} \to l^\infty(\Theta)$. In this context, we use "$\xrightarrow{d}$", "$\xrightarrow{p}$", and "$\xrightarrow{a.s.}$" 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.,

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

It is easy to show that $\rho_F(\theta, \theta') = \sqrt{2}\|[I_k - Diag(\Omega_F(\theta, \theta'))]^{1/2}\|$.

Finally, the assumptions in the next section and some of the auxiliary results make use of the set

$$\Lambda^*_{n,F_n} \equiv \left\{ (\theta, \ell) \in \Theta_I^{\eta_n}(F_n) \times \mathbb{R}^k : \right.$$
$$\left. \ell = \kappa_n^{-1}\sqrt{n} D_{F_n}^{-1/2}(\theta) E_{F_n}[m(W, \theta)] \right\}, \tag{A.3}$$

where $\Theta_I^{\eta_n}(F_n)$ and $\{\eta_n\}_{n\geq1}$ are as in Definition 4.3 and $\{\kappa_n\}_{n\geq1}$ is as in Assumption M.1.

Appendix B. Additional assumptions

This section collects several assumptions that are routinely assumed in the literature of partially identified models defined by moment (in)equalities, and some additional ones required by this paper.

Assumption A.5. Given the function $\varphi : \mathbb{R}^p_{[+\infty]} \times \mathbb{R}^{k-p}_{[\pm\infty]} \times \Psi \to \mathbb{R}^k_{[+\infty]}$ in Assumption M.1, 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) $\varphi_j^*(\xi_j) \geq \varphi_j(\xi, \Omega)$ for all $(\xi, \Omega) \in \mathbb{R}^p_{[+\infty]} \times \mathbb{R}^{k-p}_{[\pm\infty]} \times \Psi$.
(b) φ_j^* is continuous.
(c) $\varphi_j^*(\xi_j) = 0$ for all $\xi_j \leq 0$ and $\varphi_j^*(\infty) = \infty$.

Assumption A.6. For any $\{F_n \in \mathcal{P}_0\}_{n\geq1}$, let Λ and Λ^* be such that $\Lambda_{n,F_n} \xrightarrow{H} \Lambda$ and $\Lambda^*_{n,F_n} \xrightarrow{H} \Lambda^*$, where $\Lambda_{n,F}$ and Λ^*_{n,F_n} are defined in Eqs. (3.6) and (A.3), respectively. Then, for all $(\theta, \ell^*) \in \Lambda^*$ there exists $(\theta, \ell) \in \Lambda$ where $\ell_j = 0$ for all $j > p$, $\ell_j \geq \varphi_j^*(\ell_j^*)$ for all $j \leq p$, and φ^* is defined as in Assumption A.5.

Assumption A.7. For any $\{F_n \in \mathcal{P}_0\}_{n\geq1}$, let (Ω, Λ) be such that $\Omega_{F_n} \xrightarrow{u} \Omega$ and $\Lambda_{n,F_n} \xrightarrow{H} \Lambda$ with $(\Omega, \Lambda) \in \mathcal{C}(\theta) \times \mathcal{S}(\Theta \times \mathbb{R}^k_{[\pm\infty]})$ and Λ_{n,F_n} as in Eq. (3.6). 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 $c_{(1-\alpha)}(\Lambda, \Omega) = 0$, $\liminf_{n\to\infty} P_{F_n}(T_n = 0) \geq (1-\alpha)$, where T_n is as in Eq. (3.3).

Assumption A.8. The following conditions hold.

(a) For all $(\theta, F) \in \Theta \times \mathcal{P}_0$, $Q_F(\theta) \geq c \min\{\delta, \inf_{\tilde{\theta}\in\Theta_I(F)} \|\theta - \tilde{\theta}\|\}^\chi$ for constants $c, \delta > 0$ and χ as in Assumption M.8.
(b) Θ 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.,

$$\lim_{\delta\to0} \sup_{F\in\mathcal{P}_0,(\theta,\theta'):\|\theta-\theta'\|\leq\delta} \|G_F(\theta) - G_F(\theta')\| = 0.$$

Remark B.1. Assumption A.5 is satisfied if the function φ is any of the functions $\varphi^{(1)} - \varphi^{(4)}$ described in Andrews and Soares (2010) or Andrews and Barwick (2012). This follows from Lemma D.9, as the functions $\varphi^{(1)} - \varphi^{(4)}$ satisfy the conditions of this result.

Remark B.2. Without Assumption A.7 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 way to address this problem is by adding an infinitesimal constant to the critical value, which introduces an additional tuning parameter that needs to be chosen by the researcher. Another way is to impose Assumption A.7, so that the limiting distribution is either continuous or has a discontinuity that does not cause asymptotic over-rejection. Note that this assumption holds in Example 6.1, where J in Eq. (6.7) is continuous at $x > 0$.

The literature routinely assumes that the test function S in Eq. (2.5) 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.

Assumption M.4. The function S 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) $S(m, \Omega) = S(\Delta m, \Delta\Sigma\Delta)$ 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) $S(m, \Omega) \geq 0$ 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.5. 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 $x \in \mathbb{R}$ is:

(a) continuous for $x > 0$,
(b) strictly increasing for $x > 0$ unless $p = k$ and $h_1 = \infty^p$,
(c) 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) $P(S(Z + (m_1, \mathbf{0}_{k-p}), \Omega) \leq x) < P(S(Z + (m_1^*, \mathbf{0}_{k-p}), \Omega) \leq x)$ for all $x > 0$ and all $m_1, m_1^* \in \mathbb{R}^p_{[+\infty]}$ with $m_1 \leq m_1^*$ and $m_1 \neq m_1^*$.

Assumption M.6. The function S satisfies the following conditions.

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

Assumption M.7. $S(m, \Omega) > 0$ 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.8. 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.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. Proofs of the main theorems

Proof of Theorem 3.1. *Step 1.* Let $\tilde{\Omega}_n(\theta) \equiv D_{F_n}^{-1/2}(\theta)\hat{\Sigma}_n(\theta)D_{F_n}^{-1/2}(\theta)$ and consider the following derivation

$$\begin{aligned}
T_n &\equiv \inf_{\theta\in\Theta} S(\sqrt{n}\bar{m}_n(\theta), \hat{\Sigma}_n(\theta)) \\
&= \inf_{\theta\in\Theta} S(\sqrt{n}D_{F_n}^{-1/2}(\theta)\bar{m}_n(\theta), \tilde{\Omega}_n(\theta)) \\
&= \inf_{\theta\in\Theta} S(v_n(\theta) + \sqrt{n}D_{F_n}^{-1/2}(\theta)E_{F_n}[m(W, \theta)], \tilde{\Omega}_n(\theta)) \\
&= \inf_{(\theta,\ell)\in\Lambda_{n,F_n}} S(v_n(\theta) + \ell, \tilde{\Omega}_n(\theta)).
\end{aligned}$$

Step 2. Let $\mathcal{D}$ be 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 $g_n : \mathcal{D} \to \mathbb{R}$ be defined by

$$g_n(v(\cdot), \Omega(\cdot)) \equiv \inf_{(\theta,l)\in\Lambda_{n,F_n}} S(v(\theta) + \ell, \Omega(\theta)). \tag{C.1}$$

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

$$g(v(\cdot), \Omega(\cdot)) \equiv \inf_{(\theta,l)\in\Lambda} S(v(\theta) + \ell, \Omega(\theta)).$$

We now show that if the sequence of (deterministic) functions $\{(v_n(\cdot), \Omega_n(\cdot))\}_{n\geq1}$ with $(v_n(\cdot), \Omega_n(\cdot)) \in \mathcal{D}$ for all $n \in \mathbb{N}$ satisfies

$$\lim_{n\to\infty} \sup_{\theta\in\Theta} \|(v_n(\theta), \Omega_n(\theta)) - (v(\theta), \Omega(\theta))\| = 0, \tag{C.2}$$

for some $(v(\cdot), \Omega(\cdot)) \in \mathcal{D}_0$, then

$$\lim_{n\to\infty} g_n(v_n(\cdot), \Omega_n(\cdot)) = g(v(\cdot), \Omega(\cdot)).$$

We need to show that $\liminf_{n\to\infty} g_n(v_n(\cdot), \Omega_n(\cdot)) \geq g(v(\cdot), \Omega(\cdot))$. The argument to show that $\limsup_{n\to\infty} g_n(v_n(\cdot), \Omega_n(\cdot)) \leq g(v(\cdot), \Omega(\cdot))$ is similar and therefore omitted. Suppose not, i.e., suppose that $\exists\delta > 0$ and a subsequence $\{a_n\}_{n\geq1}$ of $\{n\}_{n\geq1}$ s.t. $\forall n \in \mathbb{N}$,

$$g_{a_n}(v_{a_n}(\cdot), \Omega_{a_n}(\cdot)) < g(v(\cdot), \Omega(\cdot)) - \delta. \tag{C.3}$$

By definition, there exists a sequence $\{(\theta_{a_n}, \ell_{a_n})\}_{n\geq1}$ that approximately achieves the infimum in Eq. (C.1), i.e., $\forall n \in \mathbb{N}$, $(\theta_{a_n}, \ell_{a_n}) \in \Lambda_{a_n,F_{a_n}}$ and

$$\|g_{a_n}(v_{a_n}(\cdot), \Omega_{a_n}(\cdot)) - S(v_{a_n}(\theta_{a_n}) + \ell_{a_n}, \Omega_{a_n}(\theta_{a_n}))\| \leq \delta/2. \tag{C.4}$$

Since $\Lambda_{a_n,F_{a_n}} \subseteq \Theta \times \mathbb{R}^k_{[\pm\infty]}$ and since $(\Theta \times \mathbb{R}^k_{[\pm\infty]}, d)$ is a compact metric space, there exists a subsequence $\{b_n\}_{n\geq1}$ of $\{a_n\}_{n\geq1}$ and $(\theta^*, \ell^*) \in \Theta \times \mathbb{R}^k_{[\pm\infty]}$ s.t. $d((\theta_{b_n}, \ell_{b_n}), (\theta^*, \ell^*)) \to 0$.

We first show that $(\theta^*, \ell^*) \in \Lambda$. Suppose not, i.e., $(\theta^*, \ell^*) \notin \Lambda$, and consider the following argument

$$\begin{aligned}
&d((\theta_{b_n}, \ell_{b_n}), (\theta^*, \ell^*)) + d_H(\Lambda_{b_n,F_{b_n}}, \Lambda) \\
&\quad\geq d((\theta_{b_n}, \ell_{b_n}), (\theta^*, \ell^*)) + \inf_{(\theta,\ell)\in\Lambda} d((\theta, \ell), (\theta_{b_n}, \ell_{b_n})) \\
&\quad\geq \inf_{(\theta,\ell)\in\Lambda'} d((\theta, \ell), (\theta^*, \ell^*)) > 0,
\end{aligned}$$

where the first inequality follows from the definition of Hausdorff distance and the fact that $(\theta_{b_n}, \ell_{b_n}) \in \Lambda_{b_n,F_{b_n}}$, and the second inequality follows by the triangular inequality. The final strict inequality follows from the fact that $\Lambda \in \mathcal{S}(\Theta \times \mathbb{R}^k_{[\pm\infty]})$, i.e., it is a compact subset of $(\Theta \times \mathbb{R}^k_{[\pm\infty]}, d)$, $f(\theta, \ell) = d((\theta, \ell), (\theta^*, \ell^*))$ is a continuous real-valued function, and Royden (1988, Theorem 7.18). Taking limits as $n \to \infty$ and using that $d((\theta_{b_n}, \ell_{b_n}), (\theta^*, \ell^*)) \to 0$ and $\Lambda_{b_n,F_{b_n}} \xrightarrow{H} \Lambda$, 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. $\ell_j^* = -\infty$ 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 $\Lambda_{b_n,F_{b_n}}$, $\ell_{b_n} \in \mathbb{R}^k$ and thus, $\Xi(\ell_{b_n}) < \infty$. By the case under consideration, $\lim_{n\to\infty} \Xi(\ell_{b_n}) = \Xi(\ell^*) = \infty$.

Since $(\Theta, \|\cdot\|)$ is a compact metric space, $d((\theta_{b_n}, \ell_{b_n}), (\theta^*, \ell^*)) \to 0$ implies that $\theta_{b_n} \to \theta^*$. Then, consider the following derivation,

$$\begin{aligned}
&\|(v_{b_n}(\theta_{b_n}), \Omega_{b_n}(\theta_{b_n})) - (v(\theta^*), \Omega(\theta^*))\| \\
&\quad\leq \|(v_{b_n}(\theta_{b_n}), \Omega_{b_n}(\theta_{b_n})) - (v(\theta_{b_n}), \Omega(\theta_{b_n}))\| \\
&\qquad+ \|(v(\theta_{b_n}), \Omega(\theta_{b_n})) - (v(\theta^*), \Omega(\theta^*))\| \\
&\quad\leq \sup_{\theta\in\Theta} \|(v_{b_n}(\theta), \Omega_{b_n}(\theta)) - (v(\theta), \Omega(\theta))\| \\
&\qquad+ \|(v(\theta_{b_n}), \Omega(\theta_{b_n})) - (v(\theta^*), \Omega(\theta^*))\| \to 0,
\end{aligned}$$

where the last convergence holds by Eq. (C.2), $\theta_{b_n} \to \theta^*$, and $(v(\cdot), \Omega(\cdot)) \in \mathcal{D}_0$.

Notice that $(v(\cdot), \Omega(\cdot)) \in \mathcal{D}_0$ and the compactness of Θ imply that $(v(\theta^*), \Omega(\theta^*))$ is bounded. Since $\lim_{n\to\infty} \Xi(\ell_{b_n}) = \infty$ and $v(\theta^*) \in \mathbb{R}^k$, it then follows that $\lim_{n\to\infty} \Xi(\ell_{b_n})^{-1}\|v_{b_n}(\theta_{b_n})\| = 0$. By construction, $\{\Xi(\ell_{b_n})^{-1}\ell_{b_n}\}_{n\geq 1}$ is s.t. $\lim_{n\to\infty} \Xi(\ell_{b_n})^{-1}\left[\ell_{b_n,j}\right]_- = 1$ for some $j \leq p$ or $\lim_{n\to\infty} \Xi(\ell_{b_n})^{-1}\left|\ell_{b_n,j}\right| = 1$ for some $j > p$. We then conclude that $\lim_{n\to\infty} \Xi(\ell_{b_n})^{-1}[v_{b_n,j}(\theta_{b_n}) + \ell_{b_n,j}]_- = 1$ for some $j \leq p$ or $\lim_{n\to\infty} \Xi(h_{b_n})^{-1}\left|v_{b_n,j}(\theta_{b_n}) + \ell_{b_n,j}\right| = 1$ for some $j > p$. This implies that

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

Since $\{(\theta_{b_n}, \ell_{b_n})\}_{n\geq 1}$ is a subsequence of $\{(\theta_{a_n}, \ell_{a_n})\}_{n\geq 1}$ which approximately achieves the infimum in Eq. (C.1), it then follows that

$$g_n(v_n(\cdot), \Omega_n(\cdot)) \to \infty. \tag{C.5}$$

We now show that Eq. (C.5) is a contradiction. Since $\{F_n \in \mathcal{P}_0\}_{n\geq 1}$ then there is a sequence $\{\theta_n\}_{n\geq 1}$ s.t.

$$\liminf_{n\to\infty} \sqrt{n}\sigma_{F_n,j}^{-1}(\theta_n)E_{F_n}[m_j(W, \theta_n)] \equiv \ell_j^* \geq 0, \quad \text{for } j \leq p$$
$$\lim_{n\to\infty} \sqrt{n}\sigma_{F_n,j}^{-1}(\theta_n)|E_{F_n}[m_j(W, \theta_n)]| \equiv \ell_j^* = 0, \quad \text{for } j > p.$$

By compactness of $(\Theta \times \mathbb{R}^k_{[\pm\infty]}, d)$, we can find a subsequence $\{k_n\}_{n\geq 1}$ of $\{n\}_{n\geq 1}$ s.t. $d((\tilde{\theta}_{k_n}, \tilde{\ell}_{k_n}), (\tilde{\theta}^*, \tilde{\ell}^*)) \to 0$ with $(\tilde{\theta}^*, \tilde{\ell}^*) \in \Theta \times \mathbb{R}^p_{[+\infty]} \times \mathbb{R}^{k-p}$. By repeating the previous arguments, we can show that $\lim_{n\to\infty}(v_{k_n}(\tilde{\theta}_{k_n}), \Omega_{k_n}(\tilde{\theta}_{k_n})) = (v(\tilde{\theta}^*), \Omega(\tilde{\theta}^*)) \in \mathbb{R}^k \times \Psi$, which implies that

$$\inf_{(\theta,l)\in\Lambda'_{k_n,F_{k_n}}} S(v_{k_n}(\theta) + \ell, \Omega_{k_n}(\theta))$$
$$\leq S(v_{k_n}(\tilde{\theta}_{k_n}) + \tilde{\ell}_{k_n}, \Omega_{k_n}(\tilde{\theta}_{k_n})) \to S(v(\tilde{\theta}^*) + \tilde{\ell}^*, \Omega(\tilde{\theta}^*)).$$

Since $(v(\tilde{\theta}^*) + \tilde{\ell}^*, \Omega(\tilde{\theta}^*)) \in \mathbb{R}^p_{[+\infty]} \times \mathbb{R}^{k-p} \times \Psi$, we conclude that $S(v(\tilde{\theta}^*) + \tilde{\ell}^*, \Omega(\tilde{\theta}^*))$ is bounded. Since $\{k_n\}_{n\geq 1}$ is a subsequence of $\{n\}_{n\geq 1}$, this is a contradiction to Eq. (C.5).

Since $d((\theta_{b_n}, \ell_{b_n}), (\theta^*, \ell^*)) \to 0$, we can conclude that $\lim_{n\to\infty}(v_{b_n}(\theta_{b_n}), \Omega_{b_n}(\theta_{b_n})) = (v(\theta^*), \Omega(\theta^*)) \in \mathbb{R}^k \times \Psi$ repeating previous arguments. This implies that $\lim_{n\to\infty}(v_{b_n}(\theta_{b_n}) + \ell_{b_n}, \Omega_{b_n}(\theta_{b_n})) = (v(\theta^*) + \ell^*, \Omega(\theta^*)) \in (\mathbb{R}^k_{[\pm\infty]} \times \Psi)$ and, so, gives us that $\lim_{n\to\infty} S(v_{b_n}(\theta_{b_n}) + \ell_{b_n}, \Omega_{b_n}(\theta_{b_n})) = S(v(\theta^*) + \ell^*, \Omega(\theta^*))$, i.e, $\exists N \in \mathbb{N}$ s.t. $\forall n \geq N$,

$$\|S(v_{b_n}(\theta_{b_n}) + \ell_{b_n}, \Omega_{b_n}(\theta_{b_n})) - S(v(\theta^*) + \ell^*, \Omega(\theta^*))\| \leq \delta/2. \tag{C.6}$$

By combining Eqs. (C.4) and (C.6), and the fact that $(\theta^*, \ell^*) \in \Lambda$, it follows that $\exists N \in \mathbb{N}$ s.t. $\forall n \geq N$,

$$g_{b_n}(v_{b_n}(\cdot), \Omega_{b_n}(\cdot)) \geq S(v(\theta^*) + \ell^*, \Omega(\theta^*)) - \delta$$
$$\geq g(v(\cdot), \Omega(\cdot)) - \delta,$$

which is a contradiction to Eq. (C.3).

Step 3. The proof is completed by combining the representation in step 1, the convergence result in step 2, Lemma D.2, 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 D.2 and standard convergence results imply that $(v_n(\cdot), \tilde{\Omega}(\cdot)) \xrightarrow{d} (v_\Omega(\cdot), \Omega(\cdot))$ and $(v_\Omega(\cdot), \Omega(\cdot)) \in \mathcal{D}_0$ a.s. □

Proof of Theorem 4.1. We start by proving that for $\eta \geq 0$,

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

Steps 1–4 hold for $\eta \geq 0$, step 5 needs that $\eta > 0$, and step 6 holds for $\eta = 0$ under Assumption A.7.

Step 1. For any $F \in \mathcal{P}_0$, let $\tilde{T}_n^*$ be defined by as follows

$$\tilde{T}_n^* \equiv \inf_{\theta\in\Theta_I^{\eta_n}(F)} S(\hat{v}_n^*(\theta) + \varphi^*(\kappa_n^{-1}\sqrt{n}\hat{D}_n^{1/2}(\theta)\bar{m}_n(\theta)), \hat{\Omega}_n(\theta)),$$

and let $\tilde{c}_n^{RS}(1-\alpha)$ denote its conditional $(1-\alpha)$-quantile. Consider the following derivation

$$P_F(T_n > \hat{c}_n^{RS}(1-\alpha) + \eta)$$
$$\leq P_F(T_n > \tilde{c}_n^{RS}(1-\alpha) + \eta) + P_F(\hat{c}_n^{RS}(1-\alpha) < \tilde{c}_n^{RS}(1-\alpha))$$
$$\leq P_F(T_n > \tilde{c}_n^{RS}(1-\alpha) + \eta) + P_F(\hat{\Theta}_I \not\subseteq \Theta_I^{\eta_n}(F)),$$

where the first inequality is elementary and the second inequality follows from the fact that Assumption A.5 and $\hat{c}_n^{RS}(1-\alpha) < \tilde{c}_n^{RS}(1-\alpha)$ implies $\hat{\Theta}_I \not\subseteq \Theta_I^{\eta_n}(F)$. By this and Lemma D.13, it follows that

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

Step 2. By definition, there exists a subsequence $\{a_n\}_{n\geq 1}$ of $\{n\}_{n\geq 1}$ and a subsequence $\{F_{a_n}\}_{n\geq 1}$ s.t.

$$\limsup_{n\to\infty} \sup_{F\in\mathcal{P}_0} P_F(T_n > \tilde{c}_n^{RS}(1-\alpha) + \eta)$$
$$= \lim_{n\to\infty} P_{F_{a_n}}(T_{a_n} > \tilde{c}_{a_n}^{RS}(1-\alpha) + \eta). \tag{C.7}$$

By Lemma D.7, there is a further sequence $\{b_n\}_{n\geq 1}$ of $\{a_n\}_{n\geq 1}$ s.t. $\Omega_{F_{b_n}} \xrightarrow{u} \Omega$, $\Lambda_{b_n,F_{b_n}} \xrightarrow{H} \Lambda$, and $\Lambda^*_{b_n,F_{b_n}} \xrightarrow{H} \Lambda^*$, where $\Lambda_{b_n,F_{b_n}}$ and $\Lambda^*_{b_n,F_{b_n}}$ are as in Eqs. (3.6) and (A.3), respectively, for some $(\Omega, \Lambda, \Lambda^*) \in \mathcal{C}(\theta) \times \mathcal{S}(\Theta \times \mathbb{R}^k_{[\pm\infty]})^2$. Since $\Omega_{F_{b_n}} \xrightarrow{u} \Omega$ and $\Lambda_{b_n,F_{b_n}} \xrightarrow{H} \Lambda$, Theorem 3.1 implies that $T_{b_n} \xrightarrow{d} J(\Lambda, \Omega) \equiv \inf_{(\theta,\ell)\in\Lambda} S(v_\Omega(\theta) + \ell, \Omega(\theta))$. Similarly, Theorem C.1 implies that $\{\tilde{T}^*_{a_n}|\{W_i\}_{i=1}^{a_n}\} \xrightarrow{d} J^*(\Lambda^*, \Omega) \equiv \inf_{(\theta,\ell)\in\Lambda^*} S(v_\Omega(\theta) + \varphi^*(\ell), \Omega(\theta))$ a.s.

Step 3. We now show that $J^*(\Lambda^*, \Omega) \geq J(\Lambda, \Omega)$. Suppose not, i.e., $\exists(\theta, \ell^*) \in \Lambda^*$ s.t. $S(v_\Omega(\theta) + \varphi^*(\ell^*), \Omega(\theta)) < J(\Lambda, \Omega)$. By Assumption A.6, $\exists(\theta, \ell) \in \Lambda$ where $\ell_j = 0$ for all $j > p$ and $\ell_j \geq \varphi_j^*(\ell_j^*)$ for all $j \leq p$. Thus

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

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

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

$$\lim P_{F_{b_n}}(\tilde{c}_{b_n}^{RS}(1-\alpha) \leq c_{(1-\alpha)}(\Lambda, \Omega) - \varepsilon) = 0. \tag{C.8}$$

Let $\tilde{\varepsilon} \in (0, \varepsilon)$ be chosen s.t. $c_{(1-\alpha)}(\Lambda, \Omega) - \tilde{\varepsilon}$ is a continuity point of the CDF of $J^*(\Lambda^*, \Omega)$. Then, for almost all sample sequences,

$$\lim P_{F_{b_n}}(\tilde{T}^*_{b_n} \leq c_{(1-\alpha)}(\Lambda, \Omega) - \tilde{\varepsilon}|\{W_i\}_{i=1}^{b_n})$$
$$= P(J^*(\Lambda^*, \Omega) \leq c_{(1-\alpha)}(\Lambda, \Omega) - \tilde{\varepsilon})$$
$$\leq P(J(\Lambda, \Omega) \leq c_{(1-\alpha)}(\Lambda, \Omega) - \tilde{\varepsilon}) < 1 - \alpha,$$

where the equality holds a.s. by step 2 and that $c_{(1-\alpha)}(\Lambda, \Omega) - \tilde{\varepsilon}$ is a continuity point of the CDF of $J^*(\Lambda^*, \Omega)$, the weak inequality is a consequence of $J^*(\Lambda^*, \Omega) \geq J(\Lambda, \Omega)$, and the strict inequality follows from the fact that $c_{(1-\alpha)}(\Lambda, \Omega)$ is the $(1-\alpha)$-quantile of $J(\Lambda, \Omega)$. From here, the definition of quantile and $\tilde{\varepsilon} < \varepsilon$ imply that

$$\{\lim P_{F_{b_n}}(\tilde{T}^*_{b_n} \leq c_{(1-\alpha)}(\Lambda, \Omega) - \tilde{\varepsilon} | \{W_i\}_{i=1}^{b_n}) < 1-\alpha\}$$
$$\subseteq \{\lim \inf\{\tilde{c}^{RS}_{b_n}(1-\alpha) > c_{(1-\alpha)}(\Lambda, \Omega) - \varepsilon\}\}.$$

Since the RHS occurs for almost all sample sequences, then the LHS must also occur for almost all sample sequences. Then, Eq. (C.8) 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

$$\begin{aligned} &P_{F_{b_n}}(T_{b_n} > \tilde{c}^{RS}_{b_n}(1-\alpha) + \eta) \\ &\leq P_{F_{b_n}}(\tilde{c}^{RS}_{b_n}(1-\alpha) \leq c_{(1-\alpha)}(\Lambda, \Omega) - \varepsilon) \\ &+ 1 - P_{F_{b_n}}(T_{b_n} \leq c_{(1-\alpha)}(\Lambda, \Omega) + \eta - \varepsilon). \end{aligned} \tag{C.9}$$

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

$$\begin{aligned} &\limsup_{n\to\infty} P_{F_{b_n}}(T_{b_n} > \tilde{c}^{RS}_{b_n}(1-\alpha) + \eta) \\ &\leq 1 - P(J(\Lambda, \Omega) \leq c_{(1-\alpha)}(\Lambda, \Omega) + \eta - \varepsilon) \leq \alpha. \end{aligned}$$

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.7,

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

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 Eq. (C.9) and steps 2 and 4 that

$$\begin{aligned} &\limsup_{n\to\infty} P_{F_{b_n}}(T_{b_n} > \tilde{c}^{RS}_{b_n}(1-\alpha)) \leq 1 - P(J(\Lambda, \Omega) \\ &\leq c_{(1-\alpha)}(\Lambda, \Omega) - \varepsilon_m). \end{aligned}$$

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

Proof of Theorem 5.1. The proof follows directly from Theorem 6.1. □

Proof of Theorem 6.1. This is a non-stochastic result that holds for every sample $\{W_i\}_{i=1}^n$.

Part 1. Show that $\phi^{RS}_n \geq \phi^{RC}_n$. This result follows immediately from $\hat{c}^{RS}_n(1-\alpha) \leq \hat{c}^{RC}_n(1-\alpha)$. To show this, note that by definition $\hat{c}^{RS}_n(1-\alpha) \leq \tilde{c}_n(\theta, 1-\alpha)$ $\forall \theta \in \hat{\Theta}_I$, where $\tilde{c}_n(\theta, 1-\alpha)$ is the conditional $(1-\alpha)$-quantile of

$$S(\hat{v}^*_n(\theta) + \varphi(\kappa_n^{-1}\sqrt{n}\hat{D}_n^{1/2}(\theta)\bar{m}_n(\theta), \hat{\Omega}_n(\theta)), \hat{\Omega}_n(\theta)). \tag{C.10}$$

By definition, $\hat{c}_n(\theta, 1-\alpha)$ denotes the GMS critical value, which is defined as the conditional $(1-\alpha)$-quantile of Eq. (C.10), except that $\hat{v}^*_n(\theta)$ is replaced by $\hat{\Omega}_n^{1/2}(\theta)Z^*$, with $Z^* \sim N(\mathbf{0}_k, I_k)$ and Z^* independent of $\{W_i\}_{i=1}^n$. Since $\hat{v}^*_n(\theta)$ and $\hat{\Omega}_n^{1/2}(\theta)Z^*$ have the same conditional distribution, it follows that $\tilde{c}_n(\theta, 1-\alpha) = \hat{c}_n(\theta, 1-\alpha)$ $\forall \theta \in \hat{\Theta}_I$. We conclude that

$$\hat{c}^{RS}_n(1-\alpha) \leq \inf_{\theta\in\hat{\Theta}_I} \tilde{c}_n(\theta, 1-\alpha) = \inf_{\theta\in\hat{\Theta}_I} \hat{c}_n(\theta, 1-\alpha) = \hat{c}^{RC}_n(1-\alpha).$$

Part 2. Show that $\phi^{RC}_n \geq \phi^{BP}_n$. This result is a consequence of the following argument

$$\begin{aligned} \left\{\inf_{\theta\in\Theta} Q_n(\theta) \leq \hat{c}^{RC}_n(1-\alpha)\right\} &= \left\{\inf_{\theta\in\hat{\Theta}_I} Q_n(\theta) \leq \inf_{\theta\in\hat{\Theta}_I} \hat{c}_n(\theta', 1-\alpha)\right\} \\ &\subseteq \left\{\inf_{\theta\in\hat{\Theta}_I} Q_n(\theta) \leq \hat{c}_n(\theta', 1-\alpha), \forall \theta' \in \hat{\Theta}_I\right\} \\ &\subseteq \left\{\exists \theta \in \hat{\Theta}_I : Q_n(\theta) \leq \hat{c}_n(\theta, 1-\alpha)\right\} \\ &\subseteq \left\{\exists \theta \in \Theta : Q_n(\theta) \leq \hat{c}_n(\theta, 1-\alpha)\right\}, \end{aligned}$$

where the first equality holds by $\inf_{\theta\in\Theta} Q_n(\theta) = \inf_{\theta\in\hat{\Theta}_I} Q_n(\theta)$ and the definition of $\hat{c}^{RC}_n(1-\alpha)$, the first inclusion is elementary, the second inclusion holds by the lower semi-continuity of Q_n (implies that Q_n achieves a minimum in Θ and, hence, a minimum in $\hat{\Theta}_I$), and the final inclusion holds by $\hat{\Theta}_I \subseteq \Theta$. □

Proof of Theorem 6.2. Let $\{F_n \in \mathcal{P}\}_{n\geq 1}$, $\{\theta^*_n \in \Theta\}_{n\geq 1}$, and $\{\theta_n \in \Theta\}_{n\geq 1}$ be as in Assumption A.9. Then,

$$\begin{aligned} &E_{F_n}[\phi^{BP}_n] = P_{F_n}(Q_n(\theta) > \hat{c}_n(\theta, 1-\alpha), \ \forall \theta \in \Theta) \\ &\leq P_{F_n}\left(Q_n(\theta^*_n) > \hat{c}_n(\theta^*_n, 1-\alpha)\right) \\ &= \left\{\begin{array}{l} P_{F_n}(Q_n(\theta^*_n) > \hat{c}_n(\theta_n, 1-\alpha) \cap \hat{c}_n(\theta^*_n, 1-\alpha) \\ \quad \geq \hat{c}_n(\theta_n, 1-\alpha)) - P_{F_n}(\hat{c}_n(\theta^*_n, 1-\alpha) \\ \quad \geq Q_n(\theta^*_n) > \hat{c}_n(\theta_n, 1-\alpha) \cap \hat{c}_n(\theta^*_n, 1-\alpha) \\ \quad \geq \hat{c}_n(\theta_n, 1-\alpha)) + P_{F_n}(Q_n(\theta^*_n) > \hat{c}_n(\theta^*_n, 1-\alpha) \\ \quad \cap \hat{c}_n(\theta^*_n, 1-\alpha) < \hat{c}_n(\theta_n, 1-\alpha)) \end{array}\right\} \\ &\leq P_{F_n}(T_n > \inf_{\theta\in\hat{\Theta}_I} \hat{c}_n(\theta, 1-\alpha)) - P_{F_n}(\hat{c}_n(\theta^*_n, 1-\alpha) \\ &\geq T_n > \hat{c}_n(\theta_n, 1-\alpha)) + P_{F_n}\left(\hat{c}_n(\theta^*_n, 1-\alpha) < \hat{c}_n(\theta_n, 1-\alpha)\right), \end{aligned}$$

where the first equality holds by definition, the first inequality and second equality are elementary, and the final inequality follows from $\hat{c}_n(\theta_n, 1-\alpha) \geq \inf_{\theta\in\hat{\Theta}_I} \hat{c}_n(\theta, 1-\alpha)$ and $Q_n(\theta^*_n) = T_n$. Note that $P_{F_n}(T_n > \inf_{\theta\in\hat{\Theta}_I} \hat{c}_n(\theta, 1-\alpha)) = E_{F_n}[\phi^{RC}_n]$, and so

$$\begin{aligned} &\limsup_{n\to\infty} (E_{F_n}[\phi^{RC}_n] - E_{F_n}[\phi^{BP}_n]) \\ &\geq \limsup_{n\to\infty} P_{F_n}\left(\hat{c}_n(\theta_n, 1-\alpha) < T_n \leq \hat{c}_n(\theta^*_n, 1-\alpha)\right) \\ &\quad - \limsup_{n\to\infty} P_{F_n}\left(\hat{c}_n(\theta^*_n, 1-\alpha) < \hat{c}_n(\theta_n, 1-\alpha)\right). \end{aligned}$$

It then suffices to show that the first expression on the RHS is positive and the second expression on the RHS is zero.

We begin with the first expression. To do this, fix $\varepsilon \in (0, (c_H - c_L)/3)$ and consider the following argument

$$\begin{aligned} &P_{F_n}(\hat{c}_n(\theta_n, 1-\alpha) < T_n \leq \hat{c}_n(\theta^*_n, 1-\alpha)) \\ &\geq P_{F_n}(\hat{c}_n(\theta_n, 1-\alpha) < c_L + \varepsilon < T_n < c_H - \varepsilon \leq \hat{c}_n(\theta^*_n, 1-\alpha)) \\ &\geq P_{F_n}(c_L + \varepsilon < T_n < c_H - \varepsilon) + P_{F_n}(\hat{c}_n(\theta_n, 1-\alpha) < c_L + \varepsilon) \\ &\quad + P_{F_n}(c_H - \varepsilon \leq \hat{c}_n(\theta^*_n, 1-\alpha)) - 2, \end{aligned}$$

where all the inequalities are elementary. Using Assumption A.9 and taking sequential limits lim inf as $n \to \infty$ and $\varepsilon \downarrow 0$ we conclude that

$$\begin{aligned} &\liminf_{n\to\infty} P_{F_n}(\hat{c}_n(\theta_n, 1-\alpha) < T_n \leq \hat{c}_n(\theta^*_n, 1-\alpha)) \\ &\geq \lim_{\varepsilon\downarrow 0} P(J \in (c_L + \varepsilon, c_H - \varepsilon)) = P(J \in (c_L, c_H)) > 0, \end{aligned}$$

where the equality follows from Fatou's Lemma and the strict inequality is due to Assumption A.9.

Now consider the second expression. To do this, fix $\varepsilon \in (0, (c_H - c_L)/3)$ and consider the following argument

$$\begin{aligned}&P_{F_n}\left(\hat{c}_n(\theta_n^*, 1-\alpha) \geq \hat{c}_n(\theta_n, 1-\alpha)\right)\\&\quad\geq P_{F_n}\left(\hat{c}_n(\theta_n^*, 1-\alpha) \geq c_H - \varepsilon > c_L + \varepsilon \geq \hat{c}_n(\theta_n, 1-\alpha)\right)\\&\quad\geq P_{F_n}\left(\hat{c}_n(\theta_n^*, 1-\alpha) \geq c_H - \varepsilon\right)\\&\qquad+P_{F_n}\left(c_L+\varepsilon \geq \hat{c}_n(\theta_n, 1-\alpha)\right) - 1,\end{aligned}$$

where all the inequalities are elementary. Using Assumption A.9 and taking sequential limits lim inf as $n \to \infty$ and $\varepsilon \downarrow 0$ we conclude that the two expression on the RHS converge to one, which leads to the desired result. □

C.1. Auxiliary theorems

Theorem C.1. *Assume Assumptions A.1–A.5. Let $\{F_n \in \mathcal{P}_0\}_{n\geq1}$ be a (sub)sequence of distributions s.t. for some $(\Omega, \Lambda^*) \in \mathcal{C}(\Theta^2) \times \mathcal{S}(\Theta \times \mathbb{R}^k_{[\pm\infty]})$, (i) $\Omega_{F_n} \xrightarrow{u} \Omega$, and (ii) $\Lambda^*_{n,F_n} \xrightarrow{H} \Lambda^*$, where Λ^*_{n,F_n} is as in Eq. (A.3) (implies that $\Theta_I^{\eta_n}(F_n)$ is as in Definition 4.3). Then, there is a subsequence $\{a_n\}_{n\geq1}$ of $\{n\}_{n\geq1}$ s.t., along the sequence $\{F_{a_n}\}_{n\geq1}$,*

$$\begin{aligned}&\left\{\inf_{\theta\in\Theta_I^{\eta_{a_n}}(F_{a_n})} S\Bigg(\hat{v}^*_{a_n}(\theta) + \varphi^*(\kappa_{a_n}^{-1}\sqrt{a_n}\hat{D}_{a_n}^{1/2}(\theta)\bar{m}_{a_n}(\theta)),\right.\\&\qquad\left.\hat{\Omega}_{a_n}(\theta)\Bigg)\middle|\{W_i\}_{i=1}^{a_n}\right\} \xrightarrow{d} J^*(\Lambda^*, \Omega)\\&\quad\equiv \inf_{(\theta,\ell)\in\Lambda^*} S(v_\Omega(\theta) + \varphi^*(\ell), \Omega(\theta)),\end{aligned}$$

for almost all sample sequences $\{W_i\}_{i\geq1}$, where $v_\Omega : \Theta \to \mathbb{R}^k$ is a tight zero-mean Gaussian process with covariance (correlation) kernel $\Omega \in \mathcal{C}(\Theta^2)$.

Proof. *Step 1.* Consider the following derivation:

$$\begin{aligned}&\inf_{\theta\in\Theta_I^{\eta_n}(F_n)} S\left(\hat{v}^*_n(\theta) + \varphi^*(\kappa_n^{-1}\sqrt{n}\hat{D}_n^{1/2}(\theta)\bar{m}_n(\theta)), \hat{\Omega}_n(\theta)\right)\\&\quad= \inf_{\theta\in\Theta_I^{\eta_n}(F_n)} S\Bigg(\hat{v}^*_n(\theta)\\&\qquad+\varphi^*(\mu_{n,1}(\theta)+\mu_{n,2}(\theta)'\kappa_n^{-1}\sqrt{n}D_{F_n}^{-1/2}(\theta)E_{F_n}[m(W,\theta)]), \hat{\Omega}_n(\theta)\Bigg)\\&\quad= \inf_{(\theta,\ell)\in\Lambda^*_{n,F_n}} S\left(\hat{v}^*_n(\theta) + \varphi^*(\mu_{n,1}(\theta) + \mu_{n,2}(\theta)'\ell), \hat{\Omega}_n(\theta)\right),\end{aligned}$$

where $\mu_n(\theta) = (\mu_{n,1}(\theta), \mu_{n,2}(\theta))$, $\mu_{n,1}(\theta) \equiv \kappa_n^{-1}\tilde{v}_n(\theta)$ and $\mu_{n,2}(\theta) \equiv \{\hat{\sigma}_{n,j}^{-1}(\theta)\sigma_{F_n,j}(\theta)\}_{j=1}^k$. In order to obtain this expression, we have used that $\hat{D}_n^{-1/2}(\theta)$ and $D_{F_n}^{1/2}(\theta)$ are both diagonal matrices.

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

Step 3. 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 $g_n : \mathcal{D} \to \mathbb{R}$ be defined by

$$\begin{aligned}&g_n(v(\cdot), \mu(\cdot), \Omega(\cdot))\\&\quad\equiv \inf_{(\theta,\ell)\in\Lambda^*_{n,F_n}} S(v(\theta) + \varphi^*(\mu_1(\theta) + \mu_2(\theta)'\ell), \Omega(\theta)).\end{aligned}\tag{C.11}$$

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

$$\begin{aligned}&g(v(\cdot), \mu(\cdot), \Omega(\cdot))\\&\quad\equiv \inf_{(\theta,\ell)\in\Lambda^*} S(v(\theta) + \varphi^*(\mu_1(\theta) + \mu_2(\theta)'\ell), \Omega(\theta)).\end{aligned}$$

We now show that if the sequence of (deterministic) functions $\{(v_n(\cdot), \mu_n(\cdot), \Omega_n(\cdot))\}_{n\geq1}$ with $(v_n(\cdot), \mu_n(\cdot), \Omega_n(\cdot)) \in \mathcal{D}$ for all $n \in \mathbb{N}$ satisfies

$$\begin{aligned}&\lim_{n\to\infty}\sup_{\theta\in\Theta}\|(v_n(\theta), \mu_n(\theta), \Omega_n(\theta))\\&\quad-(v(\theta), (\mathbf{0}_k, \mathbf{1}_k), \Omega(\theta))\| = 0,\end{aligned}\tag{C.12}$$

for some $(v(\cdot), \Omega(\cdot)) \in \mathcal{D}_0$, then

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

We now show $\liminf_{n\to\infty} g_n(v_n(\cdot), \mu_n(\cdot), \Omega_n(\cdot)) \geq g(v(\cdot), (\mathbf{0}_k, \mathbf{1}_k), \Omega(\cdot))$. Showing $\limsup_{n\to\infty} g_n(v_n(\cdot), \mu_n(\cdot), \Omega_n(\cdot)) \leq g(v(\cdot), (\mathbf{0}_k, \mathbf{1}_k), \Omega(\cdot))$ is very similar and therefore omitted. Suppose not, i.e., suppose that $\exists\delta > 0$ and a subsequence $\{a_n\}_{n\geq1}$ of $\{n\}_{n\geq1}$ s.t. $\forall n \in \mathbb{N}$,

$$g_{a_n}(v_{a_n}(\cdot), \mu_{a_n}(\cdot), \Omega_{a_n}(\cdot)) < g(v(\cdot), (\mathbf{0}_k, \mathbf{1}_k), \Omega(\cdot)) - \delta.\tag{C.13}$$

By definition, there exists a sequence $\{(\theta_{a_n}, \ell_{a_n})\}_{n\geq1}$ that approximately achieves the infimum in Eq. (C.11), i.e., $\forall n \in \mathbb{N}$, $(\theta_{a_n}, \ell_{a_n}) \in \Lambda^*_{a_n,F_{a_n}}$ and

$$\begin{aligned}&|g_{a_n}(v_{a_n}(\cdot), \mu_{a_n}(\cdot), \Omega_{a_n}(\cdot)) - S(v_{a_n}(\theta_{a_n}) + \varphi^*(\mu_1(\theta_{a_n})\\&\quad+\mu_2(\theta_{a_n})'\ell_{a_n}), \Omega_{a_n}(\theta_{a_n}))| \leq \delta/2.\end{aligned}\tag{C.14}$$

Since $\Lambda^*_{a_n,F_{a_n}} \subseteq \Theta \times \mathbb{R}^k_{[\pm\infty]}$ and since $(\Theta \times \mathbb{R}^k_{[\pm\infty]}, d)$ is a compact metric space, there exists a subsequence $\{b_n\}_{n\geq1}$ of $\{a_n\}_{n\geq1}$ and $(\theta^*, \ell^*) \in \Theta \times \mathbb{R}^k_{[\pm\infty]}$ s.t. $d((\theta_{b_n}, \ell_{b_n}), (\theta^*, \ell^*)) \to 0$.

We first show that $(\theta^*, \ell^*) \in \Lambda^*$. Suppose not, i.e. $(\theta^*, \ell^*) \notin \Lambda^*$, and consider the following argument

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

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

Since $(\Theta, \|\cdot\|)$ is a compact metric space, $d((\theta_{b_n}, \ell_{b_n}), (\theta^*, \ell^*)) \to 0$ implies that $\theta_{b_n} \to \theta^*$. Then, consider the following derivation:

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

where the last convergence holds by Eq. (C.12), $\theta_{b_n} \to \theta^*$, and $(v(\cdot), \Omega(\cdot)) \in \mathcal{D}_0$.

By continuity of φ^* and Eq. (C.12), it follows that $\varphi^*(\mu_{b_n,1}(\theta_{b_n}) + \mu_{b_n,2}(\theta_{b_n})'\ell_{b_n}) \to \varphi^*(\ell^*)$. To see why, it suffices to show that $\varphi_j^*(\mu_{b_n,1,j}(\theta_{b_n}) + \mu_{b_n,2,j}(\theta_{b_n})'\ell_{b_n,j}) \to \varphi_j^*(\ell_j^*)$ for any $j = 1,\ldots,k$. For $j > p$, the result holds because $\varphi_j^* = 0$. For $j \le p$, we consider the following argument. On the one hand, $d((\theta_{b_n},\ell_{b_n}),(\theta^*,\ell^*)) \to 0$ implies $\ell_{b_n,j} \to \ell_j^* \in \mathbb{R}_{[\pm\infty]}$ and on the other hand, Eq. (C.12) implies $(\mu_{b_n,1,j}(\theta_{b_n}),\mu_{b_n,2,j}(\theta_{b_n})) \to (0,1)$. Combining this, we conclude that $\mu_{b_n,1,j}(\theta_{b_n}) + \mu_{b_n,2,j}(\theta_{b_n})\ell_{b_n,j} \to \ell_j^*$, where $\ell_j^* \in \mathbb{R}_{[\pm\infty]}$. Assumption A.5 then implies that $\varphi_j^*(\mu_{b_n,1,j}(\theta_{b_n}) + \mu_{b_n,2,j}(\theta_{b_n})\ell_{b_n,j}) \to \varphi_j^*(\ell_j^*)$.

Notice that $(v(\cdot),\Omega(\cdot)) \in \mathcal{D}_0$ and the compactness of Θ imply that $(v(\theta^*),\Omega(\theta^*))$ is bounded. Then, regardless of whether $\varphi^*(\ell^*)$ is bounded or not, $\lim_{n\to\infty}(v_{b_n}(\theta_{b_n}) + \varphi^*(\mu_1(\theta_{b_n}) + \mu_2(\theta_{b_n})\ell_{b_n}),\Omega_{b_n}(\theta_{b_n})) = (v(\theta^*)+\varphi^*(\ell^*),\Omega(\theta^*)) \in (\mathbb{R}^k_{[\pm\infty]}\times\Psi)$ and so $\lim_{n\to\infty} S(v_{b_n}(\theta_{b_n})+\varphi^*(\mu_1(\theta_{b_n})+\mu_2(\theta_{b_n})\ell_{b_n}),\Omega_{b_n}(\theta_{b_n})) = S(v(\theta^*)+\varphi^*(\ell^*),\Omega(\theta^*))$, i.e, $\exists N\in\mathbb{N}$ s.t. $\forall n\ge N$,

$$\begin{aligned}&\|S(v_{b_n}(\theta_{b_n}) + \varphi^*(\mu_1(\theta_{b_n}) + \mu_2(\theta_{b_n})\ell_{b_n}),\Omega_{b_n}(\theta_{b_n}))\\ &\quad - S(v(\theta^*)+\varphi^*(\ell^*),\Omega(\theta^*))\| \le \delta/2.\end{aligned} \tag{C.15}$$

By combining Eqs. (C.14) and (C.15), and the fact that $(\theta^*,\ell^*)\in\Lambda^*$, it follows that $\exists N\in\mathbb{N}$ s.t. $\forall n \ge N$,

$$\begin{aligned}&g_{b_n}(v_{b_n}(\cdot),\mu_{b_n}(\cdot),\Omega_{b_n}(\cdot)) \ge S(v(\theta^*)+\varphi^*(\ell^*),\Omega(\theta^*)) - \delta\\ &\quad \ge g(v(\cdot),(\mathbf{0}_k,\mathbf{1}_k),\Omega(\cdot)) - \delta,\end{aligned}$$

which is a contradiction to Eq. (C.13).

Step 4. The proof is completed by combining the representation in step 1, the convergence result in step 2, the continuity result in step 3, 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 D.2 and standard convergence results imply that $(v_\Omega(\cdot),\Omega(\cdot))\in\mathcal{D}_0$ a.s. □

Theorem C.2. $\limsup_{n\to\infty}\sup_{F\in\mathcal{P}_0} E_F[\phi_n^{BP}] \le \alpha$.

Proof. Fix $(n,F)\in\mathbb{N}\times\mathcal{P}_0$ arbitrarily. By definition, $F\in\mathcal{P}_0$ if and only if $(\theta,F)\in\mathcal{F}_0$ for some $\theta\in\Theta$. Then,

$$E_F[1-\phi_n^{BP}] = P_F(CS_n(1-\alpha)\ne\emptyset) \ge P_F(\theta\in CS_n(1-\alpha)).$$

The result follows by noting that this and Eq. (2.9) imply that

$$\begin{aligned}\liminf_{n\to\infty}\inf_{F\in\mathcal{P}_0} E_F[1-\phi_n^{BP}] &\ge \liminf_{n\to\infty}\inf_{F\in\mathcal{P}_0}\inf_{\theta\in\Theta_I(F)} P_F(\theta\in CS_n(1-\alpha))\\ &\ge 1-\alpha. \quad \square\end{aligned}$$

Appendix D. Auxiliary lemmas

D.1. Auxiliary convergence results

Lemma D.1. *Assumptions A.1–A.4 imply that:*

1. *$(\mathcal{M}(F),\rho_F)$ being totally bounded uniformly in $F\in\mathcal{P}$.*
2. *$\mathcal{M}(F)$ is Donsker and pre-Gaussian, both uniformly in $F\in\mathcal{P}$.*
3. *$(\Theta,\|\cdot\|)$ is a totally bounded metric space.*
4. *$\forall\varepsilon>0$, $\lim_{\delta\downarrow 0}\limsup_{n\to\infty}\sup_{F\in\mathcal{P}} P_F(\sup_{\|\theta-\theta'\|<\delta}\|v_n(\theta)-v_n(\theta')\|>\varepsilon)=0$.*

Proof. *Part 1.* Fix $\delta>0$ arbitrarily and consider the following derivation:

$$\begin{aligned}&\{\rho_F(\theta,\theta')\le\delta\}\\ &\equiv\left\{\left\|\left\{V_F[\sigma_{F,j}^{-1}(\theta)m_j(W,\theta)-\sigma_{F,j}^{-1}(\theta')m_j(W,\theta')]^{1/2}\right\}_{j=1}^k\right\|\le\delta\right\}\\ &=\left\{\|[I_k-Diag(\Omega_F(\theta,\theta'))]^{1/2}\|\le\delta/\sqrt{2}\right\}\\ &\supseteq\left\{\|\theta-\theta'\|\le\delta'\right\},\end{aligned}$$

where the identity follows from the definition of the "intrinsic" variance semimetric, the second equality is elementary, and the inclusion holds for some $\delta'>0$ independent of F due to Assumption A.4.

By compactness of $(\Theta,\|\cdot\|)$, $\exists\{\theta_s\}_{s=1}^S$ s.t. $\cup_{s=1}^S\{\theta\in\Theta:\|\theta_s-\theta\|\le\delta'\}=\Theta$. Based on this, we can define $\{f_s\in\mathcal{M}(F)\}_{s=1}^S$ s.t. $f_s\equiv D_F^{-1/2}(\theta_s)m(\cdot,\theta_s)$ for all $s=1,\ldots,S$. Let $D_F^{-1/2}(\theta)m(\cdot,\theta)\in\mathcal{M}(F)$ be arbitrarily chosen.

We now claim that $\rho_F(\theta_s,\theta)\le\delta$ for some $s=1,\ldots,S$. By the previous construction, $\exists s\in\{1,\ldots,S\}$ s.t. $\{\|\theta_s-\theta\|\le\delta'\}\subseteq\{\rho_F(\theta_s,\theta)\le\delta\}$. Since the choice of $\delta>0$ was arbitrary and independent of F, the result holds.

Part 2. This follows from van der Vaart and Wellner (1996, Theorem 2.8.2). Assumption A.1 implies that $\mathcal{M}(F)$ is a measurable class. We take the envelope function to be $\{\sup_{\theta\in\Theta}|\sigma_{F,j}^{-1}(\theta)m_j(\cdot,\theta)|^2\}_{j=1}^k$, which is square integrable uniformly in $F\in\mathcal{P}$ due to Assumption A.3.

Under these conditions, the desired result is equivalent to the following: (i) v_n being asymptotically ρ_F-equicontinuous uniformly in $F\in\mathcal{P}$ and (ii) $(\mathcal{M}(F),\rho_F)$ being totally bounded uniformly in $F\in\mathcal{P}$. The first condition is *exactly* assumed by Assumption A.2 and the second condition follows from part 1.

Part 3. This result follows trivially from the fact that $(\Theta,\|\cdot\|)$ is a compact metric space. See, e.g., Royden (1988, pages 154–155).

Part 4. Fix $\varepsilon>0$ arbitrarily. By elementary arguments, it suffices to show $\exists\delta'>0$ s.t.

$$\limsup_{n\to\infty}\sup_{F\in\mathcal{P}} P_F\left(\sup_{\theta,\theta'\in\Theta:\|\theta-\theta'\|\le\delta'}\|v_n(\theta)-v_n(\theta')\|>\varepsilon\right)\le\varepsilon. \tag{D.1}$$

By Assumption A.2, $\exists\delta>0$ s.t.

$$\limsup_{n\to\infty}\sup_{F\in\mathcal{P}} P_F\left(\sup_{\theta,\theta'\in\Theta:\rho(\theta,\theta')\le\delta}\|v_n(\theta)-v_n(\theta')\|>\varepsilon\right)\le\varepsilon. \tag{D.2}$$

In turn, for this choice of δ, we can use the argument in Part 1 to prove $\exists\delta'>0$ (independent of F) s.t. $\{\|\theta-\theta'\|\le\delta'\}\subseteq\{\rho_F(\theta_s,\theta)\le\delta\}$. From this, it follows that,

$$\begin{aligned}&P_F\left(\sup_{\theta,\theta'\in\Theta:\|\theta-\theta'\|\le\delta'}\|v_n(\theta)-v_n(\theta')\|>\varepsilon\right)\\ &\quad\le P_F\left(\sup_{\theta,\theta'\in\Theta:\rho(\theta,\theta')\le\delta}\|v_n(\theta)-v_n(\theta')\|>\varepsilon\right).\end{aligned}$$

By combining the previous equation with Eq. (D.2), Eq. (D.1) follows. □

Lemma D.2. *Assume Assumptions A.1–A.4. Let $\{F_n\in\mathcal{P}\}_{n\ge1}$ be a (sub)sequence of distributions s.t. $\Omega_{F_n}\xrightarrow{u}\Omega$ for some $\Omega\in\mathcal{C}(\Theta^2)$. Then, the following results hold:*

1. *$v_n\xrightarrow{d}v_\Omega$ 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. *$\tilde\Omega_n\xrightarrow{p}\Omega$ in $l^\infty(\Theta)$.*
3. *$D_{F_n}^{-1/2}(\cdot)\hat D_n^{1/2}(\cdot)-I_k\xrightarrow{p}\mathbf{0}_k$ in $l^\infty(\Theta)$.*
4. *$\hat D_n^{-1/2}(\cdot)D_{F_n}^{1/2}(\cdot)-I_k\xrightarrow{p}\mathbf{0}_k$ in $l^\infty(\Theta)$.*
5. *$\hat\Omega_n\xrightarrow{p}\Omega$ in $l^\infty(\Theta)$.*
6. *For any arbitrary sequence $\{\lambda_n\in\mathbb{R}_{++}\}_{n\ge1}$ s.t. $\lambda_n\to\infty$, $\lambda_n^{-1}v_n\xrightarrow{p}\mathbf{0}_k$ in $l^\infty(\Theta)$.*
7. *For any arbitrary sequence $\{\lambda_n\in\mathbb{R}_{++}\}_{n\ge1}$ s.t. $\lambda_n\to\infty$, $\lambda_n^{-1}\tilde v_n\xrightarrow{p}\mathbf{0}_k$ in $l^\infty(\Theta)$.*

8. $\{v_n^*|\{W_i\}_{i=1}^n\} \xrightarrow{d} v_\Omega$ *in* $l^\infty(\Theta)$ *for almost all sample sequences* $\{W_i\}_{i\geq 1}$, *where* v_Ω *is the tight Gaussian process described in part 1.*

Proof. *Part 1.* The first part of the result follows from van der Vaart and Wellner (1996, Lemma 2.8.7), which requires three conditions: (i) $\mathcal{M}(F)$ is Donsker and pre-Gaussian, both uniformly in $\{F_n \in \mathcal{P}_0\}_{n\geq 1}$, (ii) van der Vaart and Wellner (1996, Eq. (2.8.5)), and (iii) van der Vaart and Wellner (1996, Eq. (2.8.6)). Condition (i) follows from part 1 in Lemma D.1, condition (ii) follows from $\Omega_{F_n} \xrightarrow{u} \Omega$, and condition (iii) follows from Assumption A.3.

To show the second part, consider the following arguments. On the one hand, Assumption A.4 and $\Omega_{F_n} \xrightarrow{u} \Omega$ imply that $\forall \varepsilon_1 > 0$, $\exists \delta_1 > 0$ (independent of $\theta, \theta' \in \Theta$) s.t. $\|\theta - \theta'\| \leq \delta_1$ implies that $\|Diag(\Omega(\theta,\theta')) - I_k\| \leq \varepsilon_1$ and this, in turn, implies that: $\rho_\Omega(\theta,\theta') = \sqrt{2}\|[Diag(\Omega(\theta,\theta')) - I_k]^{1/2}\| \leq \sqrt{2\varepsilon_1}$ where ρ_Ω is the "intrinsic" variance semimetric when the variance–covariance function is Ω. On the other hand, the fact that v_Ω is a tight Gaussian process and the argument in van der Vaart and Wellner (1996, page 41) implies that $\forall \varepsilon_2 > 0$, $\exists \delta_2 > 0$ (independent of $\theta, \theta' \in \Theta$) s.t. $\rho_\Omega(\theta,\theta') \leq \delta_2$ implies that $P(\|v_\Omega(\theta) - v_\Omega(\theta')\| \leq \varepsilon_2) = 1$. Fix $\varepsilon > 0$ arbitrarily. By setting $\varepsilon = \varepsilon_2$, $\varepsilon_1 = \delta_2$, and $\delta = \delta_1$, we conclude from both of these arguments that $\forall \varepsilon > 0$, $\exists \delta > 0$ (independent of $\theta, \theta' \in \Theta$) s.t. $\|\theta - \theta'\| \leq \delta$ implies that $P(\|v_\Omega(\theta) - v_\Omega(\theta')\| \leq \varepsilon) = 1$, as required.

Part 2. For any $j_1, j_2 = \{1,\ldots,k\}$, define the classes of functions $\mathcal{M}_{j_1,j_2}(F) \equiv \{\sigma_{F,j_1}^{-1}(\theta)m_{j_1}(\cdot,\theta)\sigma_{F,j_2}^{-1}(\theta)m_{j_2}(\cdot,\theta) : \mathcal{W} \to \mathbb{R}^k\}$ and $\mathcal{M}_{j_1}(F) \equiv \{\sigma_{F,j_1}^{-1}(\theta)m_{j_1}(\cdot,\theta) : \mathcal{W} \to \mathbb{R}^k\}$. The desired result can be shown by verifying that, $\forall j_1, j_2 = \{1,\ldots,k\}$, $\mathcal{M}_{j_1,j_2}(F)$ and $\mathcal{M}_{j_1}(F)$ are both Glivenko–Cantelli uniformly in $F \in \mathcal{P}$. In order to show such a result, we apply van der Vaart and Wellner (1996, Theorem 2.8.1) to each of these classes. We only verify the conditions of the theorem for $\mathcal{M}_{j_1,j_2}(F)$ (the result for $\mathcal{M}_{j_1}(F)$ follows from using very similar arguments).

Consider $\mathcal{M}_{j_1,j_2}(F)$ for any $j_1, j_2 = \{1,\ldots,k\}$. Assumption A.1 implies that $\mathcal{M}_{j_1,j_2}(F)$ is a measurable class for all $F \in \mathcal{P}$. For this class, the function $\max_{j\leq k}\sup_{\theta\in\Theta}(\sigma_{F,j}^{-1}(\theta)m_j(W,\theta))^2$ is an envelope function.

We now argue the envelope satisfies the first condition of the theorem. Under Assumption A.3, we follow the argument in Lehman and Romano (2005, page 463) to deduce that,

$$\lim_{\lambda\to\infty}\sup_{F\in\mathcal{P}} E_F\left[\left(\sup_{\theta\in\Theta}\left|\frac{m_j(W,\theta)}{\sigma_{F,j}(\theta)}\right|^2\right)1\left[\sup_{\theta\in\Theta}\left|\frac{m_j(W,\theta)}{\sigma_{F,j}(\theta)}\right| > \lambda\right]\right] < \infty,$$
$$\text{for } j = 1,\ldots,k,$$

which implies that the envelope function satisfies the first condition of the theorem.

We now verify the second condition for $\mathcal{M}_{j_1,j_2}(F)$. By Assumption A.3, the envelope is bounded in the $L_1(F)$-norm, uniformly in $F \in \mathcal{P}$. Consequently, a sufficient requirement to verify the second condition is that $(\mathcal{M}_{j_1,j_2}(F), L_1(F))$ is totally bounded uniformly in $F \in \mathcal{P}$, i.e., for all $\delta > 0$ there is a set $\{\theta_s \in \Theta\}_{s=1}^S$ s.t. for all $\theta \in \Theta$, $\exists s \leq S$ s.t.

$$E_F\left[\left|\frac{m_{j_1}(W,\theta)}{\sigma_{F,j_1}^{-1}(\theta)}\frac{m_{j_2}(W,\theta)}{\sigma_{F,j_2}^{-1}(\theta)} - \frac{m_{j_1}(W,\theta_s)}{\sigma_{F,j_1}^{-1}(\theta_s)}\frac{m_{j_2}(W,\theta_s)}{\sigma_{F,j_2}^{-1}(\theta_s)}\right|\right] < \delta.$$

Now notice that, $\forall \theta, \theta_s \in \Theta$,

$$\begin{aligned}
&E_F\left[\left|\frac{m_{j_1}(W,\theta)}{\sigma_{F,j_1}(\theta)}\frac{m_{j_2}(W,\theta)}{\sigma_{F,j_2}(\theta)} - \frac{m_{j_1}(W,\theta_s)}{\sigma_{F,j_1}(\theta_s)}\frac{m_{j_2}(W,\theta_s)}{\sigma_{F,j_2}(\theta_s)}\right|\right]\\
&\leq E_F\left[\left|\frac{m_{j_1}(W,\theta)}{\sigma_{F,j_1}(\theta)} - \frac{m_{j_1}(W,\theta_s)}{\sigma_{F,j_1}(\theta_s)}\right|\left|\frac{m_{j_2}(W,\theta)}{\sigma_{F,j_2}(\theta)}\right|\right]\\
&\quad + E_F\left[\left|\frac{m_{j_2}(W,\theta)}{\sigma_{F,j_2}(\theta)} - \frac{m_{j_2}(W,\theta_s)}{\sigma_{F,j_2}(\theta_s)}\right|\left|\frac{m_{j_1}(W,\theta_s)}{\sigma_{F,j_1}(\theta_s)}\right|\right]\\
&\leq \left\{\max_{j\in\{j_1,j_2\}}\left(E_F\left[\left|\frac{m_j(W,\theta)}{\sigma_{F,j}(\theta)} - \frac{m_j(W,\theta_s)}{\sigma_{F,j}(\theta_s)}\right|^2\right]\right)^{1/2}\right\}\\
&\quad \times \left\{2\max_{j'\in\{j_1,j_2\}}\left(E_F\left[\left|\frac{m_{j'}(W,\theta)}{\sigma_{F,j'}(\theta)}\right|^2\right]\right)^{1/2}\right\},
\end{aligned}$$

where the first inequality is elementary and the second inequality follows Hölder's inequality. The RHS is a product of two terms. By Assumption A.3, the second term is finite. Hence, the LHS can be arbitrarily small by choosing the first term of the RHS small enough. As a consequence, $(\mathcal{M}_{j_1,j_2}(F), L_1(F))$ is totally bounded uniformly in $F \in \mathcal{P}$ follows from $(\mathcal{M}_{j_1}(F), L_2(F))$ and $(\mathcal{M}_{j_2}(F), L_2(F))$ being totally bounded uniformly in $F \in \mathcal{P}$. By using the argument in van der Vaart and Wellner (1996, Exercise 1, Page 93), we can show this follows from $(\mathcal{M}(F), \rho_F)$ being totally bounded uniformly in $F \in \mathcal{P}$, which has already been shown in part 1 of Lemma D.1.

Part 3. By part 2 and that $Diag(\tilde{\Omega}_n(\theta)) = D_{F_n}^{-1}(\theta)\hat{D}_n(\theta)$ and $Diag(\Omega(\theta)) = I_k$, it follows that $D_{F_n}^{-1}(\theta)\hat{D}_n(\theta) - I_k \xrightarrow{p} \mathbf{0}_k$ in $l^\infty(\Theta)$, i.e., $\sup_{\theta\in\Theta}|\sigma_{F_n,j}^{-2}(\theta)\hat{\sigma}_{n,j}^2(\theta) - 1| \xrightarrow{p} 0\ \forall j = 1,\ldots,k$.

For any $(a,\tilde{\varepsilon}) \in \mathbb{R} \times (0,1)$, $|a^2 - 1| \leq \tilde{\varepsilon}$ implies $||a| - 1| \leq \max\{\sqrt{1+\tilde{\varepsilon}}-1, 1-\sqrt{1-\tilde{\varepsilon}}\} = 1-\sqrt{1-\tilde{\varepsilon}}$. Based on this, choose $\varepsilon \in (0, \min\{1, 2/k\})$ arbitrarily, set $\tilde{\varepsilon} = 1 - (1-k\varepsilon)^2 > 0$, and consider the following argument,

$$\begin{aligned}
&\left\{\max_{\theta\in\Theta}\|D_{F_n}^{-1}(\theta)\hat{D}_n(\theta) - I_k\| \leq \tilde{\varepsilon}\right\}\\
&\subseteq \bigcap_{j=1,\ldots,k}\left\{\max_{\theta\in\Theta}|\sigma_{F_n,j}^{-2}(\theta)\hat{\sigma}_{n,j}^2(\theta) - 1| \leq \tilde{\varepsilon}\right\}\\
&\subseteq \bigcap_{j=1,\ldots,k}\left\{\max_{\theta\in\Theta}|\sigma_{F_n,j}^{-1}(\theta)\hat{\sigma}_{n,j}(\theta) - 1| \leq \varepsilon/k\right\}\\
&\subseteq \left\{\max_{\theta\in\Theta}\|D_{F_n}^{-1/2}(\theta)\hat{D}_n^{1/2}(\theta) - I_k\| \leq \varepsilon\right\}.
\end{aligned}$$

The result then follows from part 2 and ε being arbitrarily chosen.

Part 4. For a finite sample size, it is possible that $\hat{\sigma}_{n,j}(\theta) = 0$ for some $(\theta, j) \in \Theta \times \{1,\ldots,k\}$, in which case $\hat{D}_n^{1/2}(\theta)$ would not be invertible. Let $A_n = \{\hat{D}_n^{1/2}(\theta)$ is invertible $\forall\theta \in \Theta\}$ and define $\tilde{D}_n^{1/2}(\theta) \equiv \hat{D}_n^{1/2}(\theta)$ if A_n occurs and $\tilde{D}_n^{1/2}(\theta) \equiv I_k$ otherwise. Note that $\tilde{D}_n^{1/2}(\theta)$ and $\tilde{D}_n^{-1/2}(\theta)$ are both diagonal matrices, and denote $\tilde{\sigma}_n(\theta) \equiv \tilde{D}_n^{1/2}(\theta)_{[j,j]}$ and $\tilde{\sigma}_n^{-1}(\theta) \equiv \tilde{D}_n^{-1/2}(\theta)_{[j,j]}$ for all $j = 1,\ldots,k$. Since $\hat{D}_n^{1/2}(\theta)$ may not always be invertible, we prove instead that: (i) $\inf_{F\in\mathcal{P}} P_F(\{\tilde{D}_n^{-1/2}(\theta) = \hat{D}_n^{-1/2}(\theta)\ \forall\theta \in \Theta\}) \to 1$ and (ii) $\tilde{D}_n^{-1/2}(\theta)D_{F_n}^{1/2}(\theta) - I_k \xrightarrow{p} \mathbf{0}_k$ in $l^\infty(\Theta)$. Under the previous two results, we conclude that $\hat{D}_n^{-1/2}(\theta)D_{F_n}^{1/2}(\theta) - I_k \xrightarrow{p} \mathbf{0}_k$ in $l^\infty(\Theta)$ by a slight abuse of notation.

We first show that $\inf_{F\in\mathcal{P}} P_F(\{\tilde{D}_n^{-1/2}(\theta) = \hat{D}_n^{-1/2}(\theta)\forall\theta \in \Theta\}) \to 1$. Fix $(n,\varepsilon) \in \mathbb{N} \times (0,1)$ arbitrarily. Notice that $\sup_{\theta\in\Theta}\|D_{F_n}^{-1/2}(\theta)\hat{D}_n^{1/2}(\theta) - I_k\| \leq \varepsilon$ implies that $\hat{\sigma}_{n,j}(\theta) > 0$ for all $(\theta, j) \in \Theta \times \{1,\ldots,k\}$ which is equivalent to $\hat{D}_n^{1/2}(\theta)$ being invertible $\forall\theta \in \Theta$, i.e., A_n. From this, we conclude that

$$\begin{aligned}
&\left\{\sup_{\theta\in\Theta}\|D_{F_n}^{-1/2}(\theta)\hat{D}_n^{1/2}(\theta) - I_k\| \leq \varepsilon\right\}\\
&\subseteq \{\tilde{D}_n^{-1/2}(\theta) = \hat{D}_n^{-1/2}(\theta)\ \forall\theta \in \Theta\}.
\end{aligned}$$

The result then follows from part 3. The result reveals that the matrix $\hat{D}_n^{1/2}(\theta)$ is invertible $\forall\theta \in \Theta$, uniformly in $F \in \mathcal{P}$, for n large enough.

We now show $\tilde{D}_n^{-1/2}(\theta)D_{F_n}^{1/2}(\theta) - I_k \xrightarrow{p} \mathbf{0}_k$ in $l^\infty(\Theta)$. For any arbitrarily chosen $\varepsilon \in (0,1)$ we set $\varepsilon' \equiv k\varepsilon/(1-\varepsilon) > 0$ s.t. $\varepsilon = \varepsilon'/(k+\varepsilon') > 0$. In this case, elementary arguments imply that

$$\begin{aligned}
&\left\{\sup_{\theta\in\Theta}\|D_{F_n}^{-1/2}(\theta)\hat{D}_n^{1/2}(\theta) - I_k\| \le \varepsilon\right\}\\
&\subseteq \bigcap_{j=1,\ldots,k}\left\{\sup_{\theta\in\Theta}\left|\tilde{\sigma}_{n,j}(\theta)\sigma_{F_n,j}^{-1}(\theta) - 1\right| \le \varepsilon\right\}\\
&= \bigcap_{j=1,\ldots,k}\left\{\sup_{\theta\in\Theta}\left|\tilde{\sigma}_{n,j}^{-1}(\theta)\sigma_{F_n,j}(\theta) - 1\right| \le \frac{\varepsilon'}{k}\right\}\\
&\subseteq \left\{\sup_{\theta\in\Theta}\|\tilde{D}_n^{-1/2}(\theta)D_{F_n}^{1/2}(\theta) - I_k\| \le \varepsilon'\right\}.
\end{aligned}$$

Since the arbitrary choice of $\varepsilon \in (0,1)$ induced a constant $\varepsilon' > 0$, the result then follows from part 3.

Part 5. By the triangular inequality and part 2, it suffices to show that $\hat{\Omega}_n(\theta) - \tilde{\Omega}_n(\theta) \xrightarrow{p} \mathbf{0}_{k\times k}$ in $l^\infty(\Theta)$. To show this, consider the following argument:

$$\begin{aligned}
\hat{\Omega}_n(\theta) - \tilde{\Omega}_n(\theta) &\equiv \hat{D}_n^{-1/2}(\theta)\hat{\Sigma}_n(\theta)\hat{D}_n^{-1/2}(\theta) - \tilde{\Omega}_n(\theta)\\
&= \hat{D}_n^{-1/2}(\theta)D_{F_n}^{1/2}(\theta)\tilde{\Omega}_n(\theta)D_{F_n}^{1/2}(\theta)\hat{D}_n^{-1/2}(\theta) - \tilde{\Omega}_n(\theta)\\
&= ((D_{F_n}^{1/2}(\theta)\hat{D}_n^{-1/2}(\theta) - I_k) + I_k)\tilde{\Omega}_n(\theta)\\
&\quad\times ((D_{F_n}^{1/2}(\theta)\hat{D}_n^{-1/2}(\theta) - I_k) + I_k) - \tilde{\Omega}_n(\theta)\\
&= 2(D_{F_n}^{1/2}(\theta)\hat{D}_n^{-1/2}(\theta) - I_k)\tilde{\Omega}_n(\theta)\\
&\quad + (D_{F_n}^{1/2}(\theta)\hat{D}_n^{-1/2}(\theta) - I_k)\tilde{\Omega}_n(\theta)(D_{F_n}^{1/2}(\theta)\hat{D}_n^{-1/2}(\theta) - I_k).
\end{aligned}$$

By the previous equation, the submultiplicative property of the matrix norm and the fact that $\tilde{\Omega}_n(\theta)$ is a correlation matrix, it follows that

$$\begin{aligned}
\|\hat{\Omega}_n(\theta) - \tilde{\Omega}_n(\theta)\| \le\ & 2\|D_{F_n}^{1/2}(\theta)\hat{D}_n^{-1/2}(\theta) - I_k\|\\
& + \|D_{F_n}^{1/2}(\theta)\hat{D}_n^{-1/2}(\theta) - I_k\|^2.
\end{aligned}$$

Fix $\varepsilon > 0$ arbitrarily and set $\varepsilon' > 0$ s.t. $2\varepsilon' + (\varepsilon')^2 \le \varepsilon$. Then, the previous equation implies that

$$\begin{aligned}
&\left\{\sup_{\theta\in\Theta}\|D_{F_n}^{-1/2}(\theta)\hat{D}_n^{1/2}(\theta) - I_k\| \le \varepsilon'\right\}\\
&\subseteq \left\{\sup_{\theta\in\Theta}\|\hat{\Omega}_n(\theta) - \tilde{\Omega}_n(\theta)\| \le \varepsilon\right\}.
\end{aligned}$$

The result then follows from part 3 and ε being arbitrarily chosen.

Part 6. Fix $\varepsilon, \delta > 0$ arbitrarily. By part 3 in Lemma D.1, $\exists\{\theta_s\}_{s=1}^S$ s.t. $\cup_{s=1}^S\{\theta \in \Theta : \|\theta_s - \theta\| \le \delta\} = \Theta$. Based on this, consider the following derivation:

$$\begin{aligned}
&P_{F_n}\left(\sup_{\theta\in\Theta}\|v_n(\theta)\| > \lambda_n\varepsilon\right)\\
&= P_{F_n}\left(\max_{s\le S}\sup_{\{\theta\in\Theta:\|\theta_s-\theta\|\le\delta\}}\|(v_n(\theta) - v_n(\theta_s)) + v_n(\theta_s)\| > \lambda_n\varepsilon\right)\\
&\le P_{F_n}\left(\max_{s\le S}\sup_{\{\theta\in\Theta:\|\theta_s-\theta\|\le\delta\}}\|v_n(\theta) - v_n(\theta_s)\| > \lambda_n\varepsilon/2\right)\\
&\quad + P_{F_n}\left(\max_{s\le S}\|v_n(\theta_s)\| > \lambda_n\varepsilon/2\right)\\
&\le P_{F_n}\left(\sup_{\{\theta,\theta'\in\Theta:\|\theta'-\theta\|\le\delta\}}\|v_n(\theta) - v_n(\theta')\| > \lambda_n\varepsilon/2\right)\\
&\quad + \sum_{s=1}^S P_{F_n}(\|v_n(\theta_s)\| > \lambda_n\varepsilon/2).
\end{aligned}$$

Since $\lambda_n \to \infty$, $\lambda_n\varepsilon/2 > \varepsilon$ for all $n \in \mathbb{N}$ and, so

$$\begin{aligned}
&\limsup_{n\to\infty} P_{F_n}\left(\sup_{\theta\in\Theta}\|v_n(\theta)\| > \lambda_n\varepsilon\right)\\
&\le \limsup_{n\to\infty} P_{F_n}\left(\sup_{\{\theta,\theta'\in\Theta:\|\theta'-\theta\|\le\delta\}}\|v_n(\theta) - v_n(\theta')\| > \varepsilon\right)\\
&\quad + \sum_{s=1}^S \limsup_{n\to\infty} P_{F_n}(\|v_n(\theta_s)\| > \lambda_n\varepsilon/2).
\end{aligned}$$

By taking limits as $\delta \downarrow 0$ and part 4 in Lemma D.1, we conclude that

$$\begin{aligned}
&\limsup_{n\to\infty} P_{F_n}\left(\sup_{\theta\in\Theta}\|v_n(\theta)\| > \lambda_n\varepsilon\right)\\
&\le \sum_{s=1}^S \limsup_{n\to\infty} P_{F_n}(\|v_n(\theta_s)\| > \lambda_n\varepsilon/2),
\end{aligned}$$

and it then suffices to show $P_{F_n}(\|v_n(\theta)\| > \lambda_n\varepsilon/2) \to 0\ \forall\theta \in \Theta$. To show this, notice that $\Omega_{F_n} \xrightarrow{u} \Omega$ implies $\Omega_{F_n}(\theta,\theta) \to \Omega(\theta,\theta)$ which, in turn, implies that $v_n(\theta) \xrightarrow{d} N(\mathbf{0}_k, \Omega(\theta,\theta))$. Since $\lambda_n \to \infty$, the result follows.

Part 7. Fix $\varepsilon > 0$ arbitrarily. By definition, $\tilde{v}_n(\theta) \equiv \hat{D}_n^{-1/2}(\theta)D_{F_n}^{1/2}(\theta)v_n(\theta)\ \forall\theta \in \Theta$ and, so the next derivation follows:

$$\begin{aligned}
&P_{F_n}\left(\sup_{\theta\in\Theta}\|\tilde{v}_n(\theta)\| > \lambda_n\varepsilon\right)\\
&= P_{F_n}\left(\sup_{\theta\in\Theta}\|((\hat{D}_n^{-1/2}(\theta)D_{F_n}^{1/2}(\theta) - I_k) + I_k)v_n(\theta)\| > \lambda_n\varepsilon\right)\\
&\le P_{F_n}\left(\sup_{\theta\in\Theta}\|\hat{D}_n^{-1/2}(\theta)D_{F_n}^{1/2}(\theta) - I_k\|\sup_{\tilde{\theta}\in\Theta}\|v_n(\tilde{\theta})\| > \lambda_n\varepsilon\right)\\
&\quad + P_{F_n}\left(\sup_{\theta\in\Theta}\|v_n(\theta)\| > \lambda_n\varepsilon\right)\\
&\le P_{F_n}\left(\sup_{\theta\in\Theta}\|\hat{D}_n^{-1/2}(\theta)D_{F_n}^{1/2}(\theta) - I_k\| > \sqrt{\lambda_n\varepsilon}\right)\\
&\quad + P_{F_n}\left(\sup_{\theta\in\Theta}\|v_n(\theta)\| > \sqrt{\lambda_n\varepsilon}\right)\\
&\quad + P_{F_n}\left(\sup_{\theta\in\Theta}\|v_n(\theta)\| > \lambda_n\varepsilon\right).
\end{aligned}$$

By parts 4 and 6, the three terms on the RHS converge to zero, concluding the proof.

Part 8. This result follows from a modification of van der Vaart and Wellner (1996, Theorem 3.6.2) to allow for drifting sequences of probability measures $\{F_n \in \mathcal{P}\}_{n\ge1}$. The original result proves that three statements are equal: (i)–(iii). For the purpose of this part, it suffices to prove that (i) still implies (iii) in the case of drifting sequences of probability measures. In order to complete the proof, one could follow the steps of the original proof: (i) implies (ii), and (i) plus (ii) imply (iii).

Provided that the assumptions of the original theorem are valid uniformly in $F \in \mathcal{P}$, then it is natural that the conclusions of such theorem are also hold uniformly. Based on this argument, we limit ourselves to show that condition (i) is uniformly valid. First, part 2 of Lemma D.1 indicates that $\mathcal{M}(F)$ is Donsker and pre-Gaussian, both uniformly in $F \in \mathcal{P}$. Second, Assumption A.3 is a finite $(2+a)$-moment condition uniformly in $F \in \mathcal{P}$. □

D.2. Auxiliary results on S

Lemma D.3. *Let the set A be defined as follows:*

$$A \equiv \left\{x \in \mathbb{R}^p_{[+\infty]} \times \mathbb{R}^{k-p} : \max\left\{\max_{j=1,\ldots,p}\{[x_j]_-\}, \max_{s=p+1,\ldots,k}\{|x_s|\}\right\} = 1\right\}. \quad \text{(D.3)}$$

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

Proof. First, notice that $(x, \Omega) \in A \times \Psi$ implies that either $x_j < 0$ for $j \le p$ or $x_s \neq 0$ for $s > p$, and so $S(x, \Omega) > 0$. So suppose not, i.e., suppose that $\inf_{(x,\Omega)\in A\times\Psi} S(x, \Omega) = 0$. Then, $\exists\{(x_n, \Omega_n) \in A \times \Psi\}_{n\ge1}$ (and so, $S(x_n, \Omega_n) > 0$) s.t. $\lim_{n\to\infty} S(x_n, \Omega_n) = 0$. By taking a further subsequence $\{a_n\}_{n\ge1}$ of $\{n\}_{n\ge1}$, $\{(x_{a_n}, \Omega_{a_n})\}_{n\ge1}$ converges to $(\bar{x}, \bar{\Omega}) \in cl(A \times \Psi) = A \times \Psi$ and so $S(\bar{x}, \bar{\Omega}) > 0$. This implies that $(x_{a_n}, \Omega_{a_n}) \to (\bar{x}, \bar{\Omega})$ and $\lim_{n\to\infty} S(x_{a_n}, \Omega_{a_n}) = 0 < S(\bar{x}, \bar{\Omega})$, which is a contradiction to the continuity of S on $\mathbb{R}^p_{[+\infty]} \times \mathbb{R}^{k-p} \times \Psi$. □

Lemma D.4. *There exists a constant* $\varpi_1 > 0$ *s.t.* $S(x, \Omega) \le 1$ *and* $\Omega \in \Psi$ *implies* $x_j \ge -\varpi_1$ *for all* $j \le p$ *and* $|x_s| \le \varpi_1$ *for all* $s > p$.

Proof. Let $(x, \Omega) \in \mathbb{R}^p_{[+\infty]} \times \mathbb{R}^{k-p} \times \Psi$ be arbitrary s.t. $S(x, \Omega) \le 1$. Set $\tilde{x} \equiv (\{[x_j]_-\}_{j=1}^p, \{x_s\}_{s=p+1}^k)$ and note that $x_j \ge -\varepsilon$ for all $j \le p$ and $|x_s| \le \varepsilon$ for all $s > p$ is equivalent to $\max_{j=1,\ldots,k} |\tilde{x}_j| \le \varepsilon$. Since $S((x_1, x_2), \Sigma)$ is non-increasing in $x_1 \in \mathbb{R}^p_{[+\infty]}$ and $\{x_j\}_{j=1}^p \ge \{[x_j]_-\}_{j=1}^p$, it follows that $S(x, \Omega) \le S(\tilde{x}, \Omega)$. Thus, it suffices to find $\varpi_1 > 0$ s.t. $S(\tilde{x}, \Omega) \le 1$ implies that $\max_{j=1,\ldots,k} |\tilde{x}_j| \le \varpi_1$. If $\max_{j=1,\ldots,k} |\tilde{x}_j| = 0$, the result trivially follows so consider the case where $\max_{j=1,\ldots,k} |\tilde{x}_j| > 0$. In this case, the maintained assumptions on S imply that

$$1 \ge S(\tilde{x}, \Omega) = S\left(\frac{\tilde{x}}{\max_{j=1,\ldots,k} |\tilde{x}_j|}, \Omega\right)\left(\max_{j=1,\ldots,k} |\tilde{x}_j|\right)^{\chi} \ge \inf_{(x,\Omega)\in A\times\Psi} S(x, \Omega)\left(\max_{j=1,\ldots,k} |\tilde{x}_j|\right)^{\chi},$$

where the set A is as in Eq. (D.3). Lemma D.3 then implies that

$$\max_{j=1,\ldots,k} |\tilde{x}_j| \le \left(\inf_{(x,\Omega)\in A\times\Psi} S(x, \Omega)\right)^{-1/\chi},$$

and the result then holds for $\varpi_1 \equiv (\inf_{(x,\Omega)\in A\times\Psi} S(x, \Omega))^{-1/\chi} > 0$. □

Lemma D.5. *There exists a constant* $\varpi_2 > 0$ *s.t.* $(x, \Omega) \in \mathbb{R}^p_{[+\infty]} \times \mathbb{R}^{k-p} \times \Psi$ *with* $x_j \ge -\varpi_2$ *for all* $j \le p$ *and* $|x_s| \le \varpi_2$ *for all* $s > p$ *implies* $S(x, \Omega) \le 1$.

Proof. Suppose not. If so, for any sequence $\{\varepsilon_m\}_{m\ge1}$ with $\varepsilon_m \downarrow 0$, we can find a sequence $\{(x_m, \Omega_m) \in \mathbb{R}^p_{[+\infty]} \times \mathbb{R}^{k-p} \times \Psi\}_{m\ge1}$ with $x_{m,j} \ge -\varepsilon_m$ for all $j \le p$, $|x_{m,s}| \le \varepsilon_m$ for all $s > p$, and $S(x_m, \Omega_m) > 1$. By definition, then, $\liminf_{m\to\infty} S(x_m, \Omega_m) > 1$. Since $(\mathbb{R}^p_{[+\infty]} \times \mathbb{R}^{k-p} \times \Psi, d)$ is compact, we can always consider a subsequence $\{a_m\}_{m\ge1}$ of $\{m\}_{m\ge1}$ s.t. $\lim_{m\to\infty} d((x_{a_m}, \Omega_{a_m}), (x, \Omega)) = 0$ for some $(x, \Omega) \in \mathbb{R}^p_{[+\infty]} \times \mathbb{R}^{k-p} \times \Psi$. By the behavior of the limits, $(x, \Omega) \in \mathbb{R}^p_{+,\infty} \times \{\mathbf{0}_{k-p}\} \times \Psi$. By continuity of S, $\lim_{m\to\infty} S(x_{a_m}, \Omega_{a_m}) = S(x, \Omega) = 0$, which is a contradiction. □

Lemma D.6. *Let* $\{(x_m, \Omega_m) \in \mathbb{R}^k_{[\pm\infty]} \times \Psi\}_{m\ge1}$ *be a sequence s.t.* $\liminf_{m\to\infty} x_{m,j} \ge 0$ *for* $j \le p$ *and* $\lim_{m\to\infty} x_{m,j} = 0$ *for* $j > p$. *Then,* $\lim_{m\to\infty} S(x_m, \Omega_n) = 0$.

Proof. Suppose not, i.e., suppose that $\liminf_{m\to\infty} S(x_m, \Omega_m) > 0$. Since $(\mathbb{R}^k_{[\pm\infty]} \times \Psi, d)$ is compact, there is a subsequence $\{a_m\}_{m\ge1}$ of $\{m\}_{m\ge1}$ s.t. $\lim_{m\to\infty} d((x_{a_m}, \Omega_{a_m}), (x, \Omega)) = 0$ for some $(x, \Omega) \in \mathbb{R}^k_{[\pm\infty]} \times \Psi$. By the behavior of the limits, $x \in \mathbb{R}^p_{[+\infty]} \times \{\mathbf{0}_{k-p}\} \subseteq \mathbb{R}^p_{[+\infty]} \times \mathbb{R}^{k-p}$. By continuity of S, $\lim_{m\to\infty} S(x_{a_m}, \Omega_{a_m}) = S(x, \Omega) = 0$, which is a contradiction. □

D.3. Auxiliary results on subsequences

Lemma D.7. *Let Assumption A.4 hold. For any arbitrary* $\{F_n \in \mathcal{P}\}_{n\ge1}$, *there exists a subsequence* $\{u_n\}_{n\ge1}$ *of* $\{n\}_{n\ge1}$ *s.t.* $\Lambda_{u_n,F_{u_n}} \xrightarrow{H} \Lambda$, $\Lambda^*_{u_n,F_{u_n}} \xrightarrow{H} \Lambda^*$, *and* $\Omega_{F_{u_n}} \xrightarrow{u} \Omega$ *for some* $(\Omega, \Lambda, \Lambda^*) \in \mathcal{C}(\Theta^2) \times \mathcal{S}(\Theta \times \mathbb{R}^k_{[\pm\infty]})^2$, *where* $\Lambda_{n,F}$ *and* Λ^*_{n,F_n} *are defined in (3.6) and (A.3), respectively.*

Proof. By Assumption A.4, $\{\Omega_F(\theta, \theta') \in \mathcal{C}(\Theta^2)\}_{F\in\mathcal{P}}$ is an equicontinuous family of functions. Since $\{\Omega_{F_n}(\theta, \theta')\}_{n\ge1}$ is a bounded sequence in $\mathbb{R}^{k\times k}$, and its closure is compact. Then, by the Arzelà-Ascoli theorem (see, e.g., Royden (1988, page 169)), there is a subsequence $\{a_n\}_{n\ge1}$ of $\{n\}_{n\ge1}$ and $\Omega \in \mathcal{C}(\Theta^2)$ s.t. $\Omega_{F_{a_n}} \xrightarrow{u} \Omega$.

Since $(\Theta \times \mathbb{R}^k_{[\pm\infty]}, d)$ is a compact metric space, $\Lambda_{a_n,F_{a_n}} \in \Theta \times \mathbb{R}^k_{[\pm\infty]}$, and the fact that any closed subset of a compact space is compact (see, e.g., Royden (1988, page 156)), $cl(\Lambda_{a_n,F_{a_n}})$ is a compact subset of $\Theta \times \mathbb{R}^k_{[\pm\infty]}$, i.e., $cl(\Lambda_{a_n,F_{a_n}}) \in \mathcal{S}(\Theta \times \mathbb{R}^k_{[\pm\infty]})$. By Corbae et al. (2009, Theorem 6.1.16), $\mathcal{S}(\Theta \times \mathbb{R}^k_{[\pm\infty]})$ is compact under the Hausdorff metric. As a consequence, there is a subsequence $\{b_n\}_{n\ge1}$ of $\{a_n\}_{n\ge1}$ and $\Lambda \in \mathcal{S}(\Theta \times \mathbb{R}^k_{[\pm\infty]})$ s.t. $d_H(cl(\Lambda_{b_n,F_{b_n}}), \Lambda) \to 0$. To conclude, it suffices to show that $d_H(\Lambda_{b_n,F_{b_n}}, \Lambda) \to 0$, which follows from $d_H(\Lambda_{b_n,F_{b_n}}, cl(\Lambda_{b_n,F_{b_n}})) = 0$ and the triangular inequality.

As a next step, we define a subsequence $\{c_n\}_{n\ge1}$ of $\{b_n\}_{n\ge1}$ s.t. $\Lambda^*_{c_n,F_{c_n}} \xrightarrow{H} \Lambda^*$ using an identical argument to the one used before. The proof is then concluded by setting $\{u_n\}_{n\ge1} \equiv \{c_n\}_{n\ge1}$. □

Lemma D.8. *For any arbitrary* $\{F_n \in \mathcal{P}\}_{n\ge1}$, *let* $X_n(\theta) : \Omega \to l^\infty(\Theta)$ *be any stochastic process s.t.* $X_n \xrightarrow{p} 0$ *in* $l^\infty(\Theta)$. *Then, there exists a subsequence* $\{u_n\}_{n\ge1}$ *of* $\{n\}_{n\ge1}$ *s.t.* $X_{u_n} \xrightarrow{a.s.} 0$ *in* $l^\infty(\Theta)$.

Proof. Throughout this proof, we consider an arbitrary sequence $\{\varepsilon_n \in \mathbb{R}_{++}\}_{n\ge1}$ with $\varepsilon_n \downarrow 0$. Then, for arbitrary $\delta > 0$ and arbitrary subsequence $\{u_n\}_{n\ge1}$ of $\{n\}_{n\ge1}$, it follows that:

$$\left\{\limsup_{n\to\infty}\left\{\sup_{\theta\in\Theta}\|X_{u_n}(\theta)\| > \varepsilon_{u_n}\right\}\right\}^c = \left\{\liminf_{n\to\infty}\left\{\sup_{\theta\in\Theta}\|X_{u_n}(\theta)\| \le \varepsilon_{u_n}\right\}\right\} \subseteq \left\{\liminf_{n\to\infty}\left\{\sup_{\theta\in\Theta}\|X_{u_n}(\theta)\| \le \delta\right\}\right\}.$$

Then, in order to complete the proof, it suffices to construct a subsequence $\{u_n\}_{n\ge1}$ of $\{n\}_{n\ge1}$ (solely dependent on $\{\varepsilon_n \in \mathbb{R}_{++}\}_{n\ge1}$) s.t.

$$P\left(\limsup_{n\to\infty}\left\{\sup_{\theta\in\Theta}\|X_{u_n}(\theta)\| > \varepsilon_{u_n}\right\}\right) = 0.$$

Consider the following elementary argument:

$$\begin{aligned}
&P\left(\limsup_{n\to\infty}\left\{\sup_{\theta\in\Theta}\|X_{u_n}(\theta)\|>\varepsilon_{u_n}\right\}\right)\\
&\quad\equiv P\left(\cap_{n\geq 1}\left\{\cup_{m\geq n}\left\{\sup_{\theta\in\Theta}\|X_{k_m}(\theta)\|>\varepsilon_{k_m}\right\}\right\}\right)\\
&\quad\leq \limsup_{n\to\infty} P\left(\left\{\cup_{m\geq n}\left\{\sup_{\theta\in\Theta}\|X_{k_m}(\theta)\|>\varepsilon_{k_m}\right\}\right\}\right)\\
&\quad\leq \limsup_{n\to\infty}\sum_{m\geq n}P_{F_{k_m}}\left(\left\{\sup_{\theta\in\Theta}\|X_{k_m}(\theta)\|>\varepsilon_{k_m}\right\}\right).
\end{aligned}\tag{D.4}$$

It suffices to show that we can construct a subsequence $\{u_n\}_{n\geq 1}$ of $\{n\}_{n\geq 1}$ (solely dependent on $\{\varepsilon_n\}_{n\geq 1}$) s.t. the limit supremum on the RHS of Eq. (D.4) is zero.

Set $u_0 = 1$. By the fact that $X_n \xrightarrow{p} 0$ in $l^\infty(\Theta)$ and for each $n \in \mathbb{N}$, we can find $u_n \geq max\{n, u_{n-1}\}$ s.t.

$$P_{F_{u_n}}\left(\sup_{\theta\in\Theta}\|X_{u_n}(\theta)\|>\varepsilon_n\right)\leq\frac{1}{2^n}.$$

As a corollary of this, we would have constructed a subsequence $\{u_n\}_{n\geq 1}$ of $\{n\}_{n\geq 1}$ s.t.

$$\sum_{m\geq 1}P_{F_{u_m}}\left(\sup_{\theta\in\Theta}\|X_{u_m}(\theta)\|>\varepsilon_m\right)<\infty.$$

It follows that the RHS of Eq. (D.4) is zero, completing the proof. □

D.4. Auxiliary results on sufficient conditions for our assumptions

In this section we present some sufficient conditions for the assumptions in Appendix B to hold.

Lemma D.9. *Let $\varphi : \mathbb{R}^p_{[+\infty]} \times \mathbb{R}^{k-p}_{[\pm\infty]} \times \Psi \to \mathbb{R}^k_{[+\infty]}$ take the form $\varphi(\xi) = (\varphi_1(\xi_1), \ldots, \varphi_p(\xi_p), 0_{k-p})$ and be s.t., for all $j = 1, \ldots, p$,*

a. *$\varphi_j(\xi_j) \leq 0$ for all $\xi_j < 0$.*
b. *$\varphi_j(\xi_j) = 0$ at $\xi_j = 0$.*
c. *$\varphi_j(\xi_j) \to \infty$ as $\xi_j \to \infty$.*
d. *$\varphi_j(\xi_j)$ has finitely many discontinuity points and $\xi_j = 0$ is not one of them.*

Then, φ satisfies Assumption A.5.

Proof. Consider the following argument $\forall j = 1, \ldots, p$. If φ_j is continuous, then set $\varphi^*_j(\xi_j) = \max\{\varphi_j(\xi_j), 0\}$ for all $\xi_j \in \mathbb{R}_{[\pm\infty]}$. Otherwise, we split the constructive argument into the following cases.

First, suppose that all its points of discontinuity are negative. In this case, define $\varphi^*_j(\xi_j) = 0$ for all $\xi_j < 0$ and $\varphi^*_j(\xi_j) = \varphi^*_j(\xi_j)$ for all $\xi_j \geq 0$. It is now easy to verify that this function satisfies all the desired properties.

Second, suppose not all points of discontinuity are negative. By condition (d), zero is not a discontinuity point and we can find the minimum discontinuity point, which we denote by $\xi^{\star\star}_j$. It follows that $\varphi_j(\xi_j)$ is a continuous function for all $\xi_j \in [0, \xi^{\star\star}_j)$. By continuity at zero, $\exists\xi^\star_j \in (0, \xi^{\star\star}_j)$ s.t. for some real number $\delta > 0$, $|\varphi_j(\xi_j)| \leq \delta$ for all $\xi_j \in [0, \xi^\star_j]$. We divide the rest of the proof into two cases.

Case 1. $\exists\delta \in (0, 1)$ s.t. $|\varphi_j(\delta\xi^\star_j)| > 0$. In this case, define the following constants: $A \equiv (G(\delta) - \delta)/(1 - \delta)$ and $B \equiv \delta/|\varphi_j(\delta\xi^\star_j)|$, where $G : \mathbb{R}_{[\pm\infty]} \to [0, 1]$ is the function defined in Eq. (A.1). Since $\delta \in (0, 1)$, it follows that $A \in (0, 1)$ and $B \geq 1$. In this case, define

$$\varphi^*_j(\xi_j)=\begin{cases}0 & \text{if } \xi_j\in[-\infty,0)\\ B|\varphi_j(\xi_j)| & \text{if } \xi_j\in[0,\delta\xi^\star_j)\\ G^{-1}(A\xi_j/\xi^\star_j+(1-A)) & \text{if } \xi_j\in[\delta\xi^\star_j,\xi^\star_j)\\ \infty & \text{if } \xi_j\in[\xi^\star_j,\infty]\end{cases}.$$

It is now easy to verify that this function satisfies all the desired properties.

Case 2. $\nexists\delta \in (0, 1)$ s.t. $|\varphi_j(\delta\xi^\star_j)| > 0$, i.e., $\varphi_j(\xi_j) = 0\ \forall\xi_j \in [0, \xi^\star_j)$. In this case, define:

$$\varphi^*_j(\xi_j)=\begin{cases}0 & \text{if } \xi_j\in[-\infty,0)\\ G^{-1}(\xi_j/(2\xi^\star_j)+1/2) & \text{if } \xi_j\in[0,\xi^\star_j)\\ \infty & \text{if } \xi_j\in[\xi^\star_j,\infty]\end{cases}.$$

It is now easy to verify that this function satisfies all the desired properties. □

Lemma D.10. *Let Assumption A.8 hold. Then, for any $\{F_n \in \mathcal{P}_0\}_{n\geq 1}$, $\gamma \in (0, 1)$, and $\{\theta_n \in \Theta^{\eta_n}_I(F_n)\}_{n\geq 1}$ for $\eta_n = \ln\kappa_n$, there is a subsequence $\{u_n\}_{n\geq 1}$ of $\{n\}_{n\geq 1}$ and a sequence $\{\hat\theta_{u_n} \in \Theta\}_{n\geq 1}$ s.t. $\|\hat\theta_{u_n} - \theta_{u_n}\| \to 0$ and*

$$\begin{aligned}
&\lim_{n\to\infty}\sqrt{u_n}\sigma^{-1}_{F_{u_n},j}(\hat\theta_{u_n})E_{F_{u_n}}[m_j(W,\hat\theta_{u_n})]\\
&\quad\geq\lim_{n\to\infty}\kappa^{-\gamma}_{u_n}\sqrt{u_n}\sigma^{-1}_{F_{u_n},j}(\theta_{u_n})E_{F_{u_n}}[m_j(W,\theta_{u_n})]\quad \textit{for } j\leq p,\\
&\lim_{n\to\infty}\sqrt{u_n}\sigma^{-1}_{F_{u_n},j}(\hat\theta_{u_n})E_{F_{u_n}}[m_j(W,\tilde\theta_{u_n})]\\
&\quad=\lim_{n\to\infty}\kappa^{-\gamma}_{u_n}\sqrt{u_n}\sigma^{-1}_{F_{u_n},j}(\theta_{u_n})E_{F_{u_n}}[m_j(W,\theta_{u_n})]\quad \textit{for } j> p.
\end{aligned}\tag{D.5}$$

Proof. By definition, $\{\theta_n \in \Theta^{\eta_n}_I(F_n)\}_{n\geq 1}$ implies that $S(\sqrt{n}D^{-1/2}_{F_n}(\theta_n)E_{F_n}[m(W, \theta_n)], \Omega_{F_n}(\theta_n)) \leq \eta_n$ and, therefore, $Q_{F_n}(\theta_n) = S(D^{-1/2}_{F_n}(\theta_n)E_{F_n}[m(W, \theta_n)], \Omega_{F_n}(\theta_n)) \leq (\eta_n^{1/\chi}/\sqrt{n})^\chi \to 0$, where the convergence occurs by $\eta_n = \ln\kappa_n$. By this and Assumption A.8(a), it follows that

$$\begin{aligned}
&O((\eta_n^{1/\chi}/\sqrt{n})^\chi)=c^{-1}Q_{F_n}(\theta_n)\geq\min\{\delta,\inf_{\tilde\theta\in\Theta_I(F_n)}\|\theta_n-\tilde\theta\|\}^\chi\\
&\quad\Rightarrow\|\theta_n-\tilde\theta_n\|\leq O(\eta_n^{1/\chi}/\sqrt{n}),
\end{aligned}\tag{D.6}$$

for some sequence $\{\tilde\theta_n \in \Theta_I(F_n)\}_{n\geq 1}$. By the convexity of Θ and Assumption A.8(c), the intermediate value theorem implies that there is a sequence $\{\theta^*_n \in \Theta\}_{n\geq 1}$ with θ^*_n in the line between θ_n and $\tilde\theta_n$ s.t. for all $\gamma > 0$,

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

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

$$\begin{aligned}
&G_{F_n}(\theta^*_n)\sqrt{n}(\hat\theta_n-\tilde\theta_n)=\kappa_n^{-\gamma}\sqrt{n}D^{-1/2}_{F_n}(\theta_n)E_{F_n}[m(W,\theta_n)]\\
&\quad-\kappa_n^{-\gamma}\sqrt{n}D^{-1/2}_{F_n}(\tilde\theta_n)E_{F_n}[m(W,\tilde\theta_n)].
\end{aligned}\tag{D.7}$$

By convexity of Θ and $\kappa_n^{-\gamma} \to 0$, $\{\hat\theta_n \in \Theta\}_{n\geq 1}$. By Eq. (D.6), $\sqrt{n}(\hat\theta_n - \tilde\theta_n) = \sqrt{n}\kappa_n^{-\gamma}(\theta_n - \tilde\theta_n) \leq O(\kappa_n^{-\gamma}\eta_n^{1/\chi}) = O(1)$. Notice that this implies that $\|\hat\theta_n - \tilde\theta_n\| \to 0$. Also, notice that this also implies that $\|\theta_n - \tilde\theta_n\| \to 0$ as $\sqrt{n}\kappa_n^{-\gamma} \to \infty$ because of $\gamma \leq 1$. Since θ^*_n is in the line between θ_n and $\tilde\theta_n$, this also means that $\|\theta^*_n - \tilde\theta_n\| \to 0$.

By using the intermediate value theorem once again, there is a sequence $\{\theta^{**}_n \in \Theta\}_{n\geq 1}$ with θ^{**}_n in the line between $\hat\theta_n$ and $\tilde\theta_n$ s.t.

$$\begin{aligned}
&\sqrt{n}D^{-1/2}_{F_n}(\hat\theta_n)E_{F_n}[m(W,\hat\theta_n)]\\
&\quad=G_{F_n}(\theta^{**}_n)\sqrt{n}(\hat\theta_n-\tilde\theta_n)+\sqrt{n}D^{-1/2}_{F_n}(\tilde\theta_n)E_{F_n}[m(W,\tilde\theta_n)]\\
&\quad=G_{F_n}(\theta^*_n)\sqrt{n}(\hat\theta_n-\tilde\theta_n)+\sqrt{n}D^{-1/2}_{F_n}(\tilde\theta_n)E_{F_n}[m(W,\tilde\theta_n)]\\
&\qquad+\epsilon_{1,n},
\end{aligned}\tag{D.8}$$

where the second equality holds by $\epsilon_{1,n} \equiv (G_{F_n}(\theta_n^{**}) - G_{F_n}(\theta_n^*))\sqrt{n}(\hat{\theta}_n - \tilde{\theta}_n)$. Since θ_n^{**} is in the line between $\hat{\theta}_n$ and $\tilde{\theta}_n$ and $\|\hat{\theta}_n - \tilde{\theta}_n\| \to 0$, this means that $\|\theta_n^{**} - \tilde{\theta}_n\| \to 0$. Then, we conclude that $\|\theta_n^{**} - \theta_n^*\| \leq \|\theta_n^{**} - \tilde{\theta}_n\| + \|\theta_n^* - \tilde{\theta}_n\| \to 0$.

Combining Eqs. (D.7) and (D.8) we get

$$\sqrt{n}D_{F_n}^{-1/2}(\hat{\theta}_n)E_{F_n}[m(W,\hat{\theta}_n)] = \kappa_n^{-\gamma}\sqrt{n}D_{F_n}^{-1/2}(\theta_n)E_{F_n}[m(W,\theta_n)] + \epsilon_{1,n} + \epsilon_{2,n}, \quad (D.9)$$

where $\epsilon_{2,n} \equiv (1 - \kappa_n^{-\gamma})\sqrt{n}D_{F_n}^{-1/2}(\tilde{\theta}_n)E_{F_n}[m(W,\tilde{\theta}_n)]$. From $\{\tilde{\theta}_n \in \Theta_I(F_n)\}_{n\geq 1}$ and $\kappa_n^{-\gamma} \to 0$, it follows that $\epsilon_{2,n,j} \geq 0$ for $j \leq p$ and $\epsilon_{2,n,j} = 0$ for $j > p$. Moreover, Assumption A.8(c) implies that $\|G_{F_n}(\theta_n^{**}) - G_{F_n}(\theta_n^*)\| = o(1)$ for any sequence $\{F_n \in \mathcal{P}_0\}_{n\geq 1}$ whenever $\|\theta_n^{**} - \theta_n^*\| \to 0$. Using $\sqrt{n}(\hat{\theta}_n - \tilde{\theta}_n) = O(1)$, we have

$$\|\epsilon_{1,n}\| \equiv \|G_{F_n}(\theta_n^{**}) - G_{F_n}(\theta_n^*)\| \times \sqrt{n}\|\hat{\theta}_n - \tilde{\theta}_n\| = o(1). \quad (D.10)$$

Finally, since $(\mathbb{R}_{[\pm\infty]}^k, d)$ is compact, there is a subsequence $\{u_n\}_{n\geq 1}$ of $\{n\}_{n\geq 1}$ 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 $\kappa_{u_n}^{-\gamma}\sqrt{u_n}D_{F_{u_n}}^{-1/2}(\theta_{u_n})E_{F_{u_n}}[m(W,\theta_{u_n})]$ converge. Then, from Eqs. (D.9), (D.10), and the properties of $\epsilon_{2,n}$, we conclude that

$$\lim_{n\to\infty} \sqrt{u_n}\sigma_{F_{u_n},j}^{-1}(\hat{\theta}_{u_n})E_{F_{u_n}}[m_j(W,\hat{\theta}_{u_n})] \geq \lim_{n\to\infty} \kappa_{u_n}^{-\gamma}\sqrt{u_n}\sigma_{F_{u_n},j}^{-1}(\theta_{u_n})E_{F_{u_n}}[m_j(W,\theta_{u_n})] \quad \text{for } j \leq p,$$

$$\lim_{n\to\infty} \sqrt{u_n}\sigma_{F_{u_n},j}^{-1}(\hat{\theta}_{u_n})E_{F_{u_n}}[m_j(W,\hat{\theta}_{u_n})] = \lim_{n\to\infty} \kappa_{u_n}^{-\gamma}\sqrt{u_n}\sigma_{F_{u_n},j}^{-1}(\theta_{u_n})E_{F_{u_n}}[m_j(W,\theta_{u_n})] \quad \text{for } j > p.$$

To conclude the proof, notice that $\|\hat{\theta}_n - \theta_n\| \leq \|\hat{\theta}_n - \tilde{\theta}_n\| + \|\tilde{\theta}_n - \theta_n\| \to 0$. □

Lemma D.11. *Let $\{F_n \in \mathcal{P}_0\}_{n\geq 1}$ be s.t. $\Lambda_{n,F_n} \xrightarrow{u} \Lambda$ and $\Lambda_{n,F_n}^* \xrightarrow{u} \Lambda^*$ for some $\Lambda, \Lambda^* \in \mathcal{S}(\Theta \times \mathbb{R}_{[\pm\infty]}^k)$. Then, Assumptions A.5 and A.8 imply Assumption A.6.*

Proof. Let $\eta_n = \ln \kappa_n$. By definition, $(\theta^*, \ell^*) \in \Lambda^*$ implies that there is a (sub)sequence $\{(\theta_n, \ell_n) \in \Lambda_{n,F_n}^*\}_{n\geq 1}$ s.t. $d((\theta_n, \ell_n), (\theta^*, \ell^*)) \to 0$ with $\ell_n \equiv \kappa_n^{-1}\sqrt{n}D_{F_n}^{-1/2}(\theta_n)E_{F_n}[m(W,\theta_n)]$. The fact that $(\theta_n, \ell_n) \in \Lambda_{n,F_n}^*$, implies that $\theta_n \in \Theta_I^{\eta_n}(F_n)$ or, equivalently, $S(\sqrt{n}D_{F_n}^{-1/2}(\theta_n)E_{F_n}[m(W,\theta_n)], \Omega_n(\theta_n)) \leq \eta_n$. By Lemma D.4, $\exists \varpi_1 > 0$ s.t.

$$\kappa_n^{1-\gamma}\ell_{n,j} = \kappa_n^{-\gamma}\sqrt{n}\sigma_{F_n,j}^{-1}(\theta_n)E_{F_n}[m_j(W,\theta_n)] \geq -\kappa_n^{-\gamma}\eta_n^{1/\chi}\varpi_1 \to 0, \quad \text{for } j \leq p,$$

$$\kappa_n^{1-\gamma}|\ell_{n,j}| = \kappa_n^{-\gamma}\sqrt{n}\sigma_{F_n,j}^{-1}(\theta_n)|E_{F_n}[m_j(W,\theta_n)]| \leq \kappa_n^{-\gamma}\eta_n^{1/\chi}\varpi_1 \to 0, \quad \text{for } j > p, \quad (D.11)$$

where $\gamma \in (0,1)$ is as in Lemma D.10 and the convergence occurs by $\eta_n = \ln \kappa_n$. By the previous equations, $\gamma < 1$, and the fact that $d(\ell_n, \ell^*) \to 0$, we conclude that $\ell^* \in \mathbb{R}_{[+\infty]}^p \times \{\mathbf{0}_{k-p}\}$.

Lemma D.10 implies that Eq. (D.5) holds. By combining this with Eq. (D.11), we conclude that there is a subsequence $\{u_n\}_{n\geq 1}$ of $\{n\}_{n\geq 1}$ and a sequence $\{\hat{\theta}_{u_n} \in \Theta\}_{n\geq 1}$ s.t. $\|\hat{\theta}_{u_n} - \theta_{u_n}\| \to 0$ and

$$\lim_{n\to\infty} \sqrt{u_n}\sigma_{F_{u_n},j}^{-1}(\hat{\theta}_{u_n})E_{F_{u_n}}[m_j(W,\hat{\theta}_{u_n})] \geq \lim_{n\to\infty} \kappa_{u_n}^{1-\gamma}\ell_{u_n,j} \geq 0, \quad \text{for } j \leq p,$$

$$\lim_{n\to\infty} \sqrt{u_n}\sigma_{F_{u_n},j}^{-1}(\hat{\theta}_{u_n})E_{F_{u_n}}[m_j(W,\hat{\theta}_{u_n})] = \lim_{n\to\infty} \kappa_{u_n}^{1-\gamma}\ell_{u_n,j} = 0, \quad \text{for } j > p.$$

We define $\hat{\ell}_n \equiv \sqrt{n}D_{F_n}^{-1/2}(\hat{\theta}_n)E_{F_n}[m(W,\hat{\theta}_n)]$ and notice that, by definition, $(\hat{\theta}_n, \hat{\ell}_n) \in \Lambda_{n,F_n}$. By the compactness of $(\Theta \times \mathbb{R}_{[\pm\infty]}^k, d)$, there is a subsequence $\{k_n\}_{n\geq 1}$ of $\{u_n\}_{n\geq 1}$ s.t. $d((\hat{\theta}_{k_n}, \hat{\ell}_{k_n}), (\theta, \ell)) \to 0$. Finally, since $\Lambda_{n,F_n} \to \Lambda \in \mathcal{S}(\Theta \times \mathbb{R}_{[\pm\infty]}^k)$, we conclude that $(\theta, \ell) \in \Lambda$. We can summarize the previous construction as follows:

$$\ell_j = \lim_{n\to\infty} \hat{\ell}_{k_n,j} \geq \lim_{n\to\infty} \kappa_{k_n}^{1-\gamma}\ell_{k_n,j} \geq 0, \quad \text{for } j \leq p,$$

$$\ell_j = \lim_{n\to\infty} \hat{\ell}_{k_n,j} = \lim_{n\to\infty} \kappa_{k_n}^{1-\gamma}\ell_{k_n,j} = 0, \quad \text{for } j > p. \quad (D.12)$$

To conclude, we show that (θ, ℓ) satisfies the requirements in Assumption A.6. First, $d((\hat{\theta}_{k_n}, \hat{\ell}_{k_n}), (\theta, \ell)) \to 0$ and $\|\hat{\theta}_{k_n} - \theta_{k_n}\| \to 0$ and imply that $\lim_{n\to\infty}\hat{\theta}_{k_n} = \lim_{n\to\infty}\theta_{k_n} = \theta$. Second, for $j > p$, Eq. (D.12) implies that $\ell_j = \lim_{n\to\infty}\hat{\ell}_{k_n,j} = 0$. Next, consider $j \leq p$ for which we know that $\ell_j^* \in \mathbb{R}_{[+\infty]}$. If $\ell_j^* = 0$, then $\varphi_j^*(\ell_j^*) = 0$ by Assumption A.5. Eq. (D.12) then implies $\ell_j \geq 0 = \varphi_j^*(\ell_j^*)$. If $\ell_j^* > 0$, then $\kappa_{k_n}^{1-\gamma}\ell_{k_n,j} \to \infty$ and so Eq. (D.12) implies $\ell_j = \infty$. It follows that $\ell_j \geq \varphi^*(\ell_j^*)$ in this case as well. □

D.5. Auxiliary results on $\hat{\Theta}_I$

Lemma D.12. *Let $\{F_n \in \mathcal{P}_0\}_{n\geq 1}$ be a (sub)sequence of distributions s.t. $\Omega_{F_n} \xrightarrow{u} \Omega$ for some $\Omega \in \mathcal{C}(\Theta^2)$. For any arbitrary sequence $\{\lambda_n \in \mathbb{R}_{++}\}_{n\geq 1}$ s.t. $\lambda_n \to \infty$, $\lambda_n^{-1}\inf_{\theta\in\Theta} Q_n(\theta) \xrightarrow{p} 0$.*

Proof. Fix $n \in \mathbb{N}$ arbitrarily. By definition, $F_n \in \mathcal{P}_0$ implies that $\theta_n \in \Theta_I(F_n)$, which implies that $E_{F_n}[m_j(W,\theta_n)] \geq 0$ for $j \leq p$ and $E_{F_n}[m_j(W,\theta_n)] = 0$ for $j > p$. Therefore

$$0 \leq \lambda_n^{-1}\inf_{\theta\in\Theta} Q_n(\theta) \leq \lambda_n^{-1}Q_n(\theta_n) = S(\lambda_n^{-1/\chi}\sqrt{n}m_n(\theta_n), \hat{\Sigma}_n(\theta_n)) \leq S(\lambda_n^{-1/\chi}v_n(\theta_n), \tilde{\Omega}_n(\theta_n)),$$

where the first two inequalities are elementary, the first equality is by definition of Q_n and by the fact that S is homogeneous of degree χ, and the second equality follows from monotonicity properties of S and $\theta_n \in \Theta_I(F_n)$, which implies that $E_{F_n}[m_j(W,\theta_n)] \geq 0$ for $j \leq p$ and $E_{F_n}[m_j(W,\theta_n)] = 0$ for $j > p$.

The proof is completed by showing that $S(\lambda_n^{-1/\chi}v_n(\theta_n), \tilde{\Omega}_n(\theta_n)) \xrightarrow{p} 0$. Suppose not, i.e., $\exists\bar{\varepsilon} > 0$ s.t.

$$\limsup_{n\to\infty} P_{F_n}\left(\left|S(\lambda_n^{-1/\chi}v_n(\theta_n), \tilde{\Omega}_n(\theta_n))\right| > \bar{\varepsilon}\right) > 0. \quad (D.13)$$

Based on this, notice that

$$\limsup_{n\to\infty} P_{F_n}\left(\left|S(\lambda_n^{-1/\chi}v_n(\theta_n), \tilde{\Omega}_n(\theta_n))\right| > \bar{\varepsilon}\right) = \lim_{n\to\infty} P_{F_{a_n}}\left(\left|S(\lambda_{a_n}^{-1/\chi}v_{a_n}(\theta_{a_n}), \tilde{\Omega}_{a_n}(\theta_{a_n}))\right| > \bar{\varepsilon}\right) = \lim_{n\to\infty} P_{F_{b_n}}\left(\left|S(\lambda_{b_n}^{-1/\chi}v_{b_n}(\theta_{b_n}), \tilde{\Omega}_{b_n}(\theta_{b_n}))\right| > \bar{\varepsilon}\right), \quad (D.14)$$

where the first equality holds for a subsequence $\{a_n\}_{n\geq 1}$ of $\{n\}_{n\geq 1}$ that achieves the limit supremum, the second equality holds for a subsequence $\{b_n\}_{n\geq 1}$ of $\{a_n\}_{n\geq 1}$ s.t. $\Omega(\theta_{b_n}) \to \Omega^*$. By Lemma D.2 (parts 5 and 6) and $\lambda_n^{1/\chi} \to \infty$, we conclude that $\lambda_{b_n}^{-1/\chi}v_{b_n}(\theta_{b_n}) \xrightarrow{p} \mathbf{0}_k$ and $\tilde{\Omega}_{b_n}(\theta_{b_n}) - \Omega_{b_n}(\theta_{b_n}) \xrightarrow{p} \mathbf{0}_k$. This, combined with $\Omega(\theta_{b_n}) - \Omega^* \to \mathbf{0}_k$ and assumed properties of S, implies that $S(\lambda_{b_n}^{-1/\chi}v_{b_n}(\theta_{b_n}), \tilde{\Omega}_{b_n}(\theta_{b_n})) \xrightarrow{p} S(\mathbf{0}_k, \Omega^*) = 0$. As a result, the RHS of Eq. (D.14) is zero, contradicting Eq. (D.13). □

Lemma D.13. *Assume Assumptions A.1–A.4 and let $\Theta_I^{\ln\kappa_n}(F_n)$ be as in Definition 4.3. Then,*

$$\lim_{n\to\infty} \inf_{F\in\mathcal{P}_0} P_F\left(\hat{\Theta}_n \subseteq \Theta_I^{\ln\kappa_n}(F)\right) = 1.$$

Proof. Throughout this proof, let

$$\hat{\Theta}_I^{UB} \equiv \{\theta \in \Theta : Q_n(\theta) \leq \sqrt{\ln \kappa_n}\}$$
$$= \{\theta \in \Theta : S((\ln \kappa_n)^{-1/(2\chi)}\sqrt{n}\hat{D}_n^{-1/2}(\theta)\bar{m}(\theta), \hat{\Omega}_n(\theta)) \leq 1\},$$

where the equality relies on the definition of $Q_n(\theta)$ and the maintained properties of S.

Step 1. Show that $\inf_{F\in\mathcal{P}_0} P_F(\hat{\Theta}_I \subseteq \hat{\Theta}_I^{UB}) \to 1$. Fix $(n, F) \in \mathbb{N} \times \mathcal{P}_0$ arbitrarily. By definition, $T_n \leq \sqrt{\ln \kappa_n}$ implies $\hat{\Theta}_I \subseteq \hat{\Theta}_I^{UB}$ and, thus, it suffices to show that $\inf_{F\in\mathcal{P}_0} P_F(T_n \leq \sqrt{\ln \kappa_n}) \to 1$. To show this, notice that:

$$\liminf_{n\to\infty} \inf_{F\in\mathcal{P}_0} P_F(T_n \leq \sqrt{\ln \kappa_n}) = \lim_{n\to\infty} P_{F_{a_n}}(T_{a_n} \leq \sqrt{\ln \kappa_{a_n}}) = 1,$$

where the first equality holds for a subsequence $\{a_n\}_{n\geq1}$ of $\{n\}_{n\geq1}$ that achieves the limit, the infimum, and is s.t. $\Omega_{F_{a_n}} \xrightarrow{u} \Omega$ for some $\Omega \in \mathcal{C}(\Theta^2)$ (which can be found by Lemma D.7), and the second equality holds by Lemma D.12.

Step 2. Show that $\inf_{F\in\mathcal{P}_0} P_F(\hat{\Theta}_I^{UB} \subseteq \Theta_I^{\ln \kappa_n}(F)) \to 1$. Fix $(n, F) \in \mathbb{N} \times \mathcal{P}_0$ arbitrarily. By Lemma D.5, there exists $\varpi_2 > 0$ s.t.

$$\Theta_I^{\ln \kappa_n}(F) \equiv \{\theta \in \Theta : S((\ln \kappa_n)^{-1/\chi}\sqrt{n}E_F[m(W,\theta)], \Sigma_F(\theta)) \leq 1\}$$
$$\supseteq \left\{\theta \in \Theta : \begin{Bmatrix} \{\sqrt{n}\sigma_{F,j}^{-1}(\theta)E_F[m_j(W,\theta)] \geq -\varpi_2(\ln \kappa_n)^{1/\chi}\}_{j=1}^{p} \cap \\ \{\sqrt{n}\sigma_{F,j}^{-1}(\theta)|E_F[m_j(W,\theta)]| \leq \varpi_2(\ln \kappa_n)^{1/\chi}\}_{j=p+1}^{k} \end{Bmatrix}\right\}.$$

It then follows that

$$\left\{\hat{\Theta}_I^{UB} \subseteq \Theta_I^{\ln \kappa_n}(F)\right\}$$
$$\supseteq \left\{\bigcap_{\theta\in\hat{\Theta}_I^{UB}} \begin{Bmatrix} \{\sqrt{n}\sigma_{F,j}^{-1}(\theta)E_F[m_j(W,\theta)] \geq -\varpi_2(\ln \kappa_n)^{-1/\chi}\}_{j=1}^{p} \cap \\ \{\sqrt{n}\sigma_{F,j}^{-1}(\theta)|E_F[m_j(W,\theta)]| \leq \varpi_2(\ln \kappa_n)^{-1/\chi}\}_{j=p+1}^{k} \end{Bmatrix}\right\}$$
$$= \begin{Bmatrix} \{\inf_{\theta\in\hat{\Theta}_I^{UB}} \sqrt{n}\sigma_{F,j}^{-1}(\theta)E_F[m_j(W,\theta)] \geq -\varpi_2(\ln \kappa_n)^{-1/\chi}\}_{j=1}^{p} \cap \\ \{\sup_{\theta\in\hat{\Theta}_I^{UB}} \sqrt{n}\sigma_{F,j}^{-1}(\theta)|E_F[m_j(W,\theta)]| \leq \varpi_2(\ln \kappa_n)^{-1/\chi}\}_{j=p+1}^{k} \end{Bmatrix}$$
$$\supseteq \left\{\max_{j=1,\ldots,k} \sup_{\theta\in\hat{\Theta}_I^{UB}} |\sqrt{n}\sigma_{F,j}^{-1}(\theta)E_F[m_j(W,\theta)]| \leq \varpi_2(\ln \kappa_n)^{-1/\chi}\right\}.$$

In turn, this implies that

$$\liminf_{n\to\infty} \inf_{F\in\mathcal{P}_0} P_F(\hat{\Theta}_I^{UB} \subseteq \Theta_I^{\ln \kappa_n}(F)) \geq \liminf_{n\to\infty} \inf_{F\in\mathcal{P}_0} P_F$$
$$\times \left(\max_{j=1,\ldots,k} \sup_{\theta\in\hat{\Theta}_I^{UB}} |\sqrt{n}\sigma_{F,j}^{-1}(\theta)E_F[m_j(W,\theta)]| \leq \varpi_2(\ln \kappa_n)^{-1/\chi}\right).$$

The proof is completed by showing that the RHS is equal to one.

By Lemma D.4, there exists $\varpi_1 > 0$ s.t.

$$\hat{\Theta}_I^{UB} = \{\theta \in \Theta : S((\sqrt{\ln \kappa_n})^{-1/\chi}\sqrt{n}\hat{D}_n^{-1/2}(\theta)\bar{m}(\theta), \hat{\Omega}_n(\theta)) \leq 1\}$$
$$\subseteq \left\{\theta \in \Theta : \begin{Bmatrix} \{\sqrt{n}\hat{\sigma}_{n,j}^{-1}(\theta)\bar{m}_j(\theta) \geq -\varpi_1(\ln \kappa_n)^{1/(2\chi)}\}_{j=1}^{p} \cap \\ \{\sqrt{n}\hat{\sigma}_{n,j}^{-1}(\theta)|\bar{m}_j(\theta)| \leq \varpi_1(\ln \kappa_n)^{1/(2\chi)}\}_{j=p+1}^{k} \end{Bmatrix}\right\}. \quad \text{(D.15)}$$

Now, fix $(n, F, \theta, j) \in \mathbb{N} \times \mathcal{P}_0 \times \hat{\Theta}_I^{UB} \times \{1, \ldots, k\}$ arbitrarily. By definition,

$$\sqrt{n}\sigma_{F,j}^{-1}(\theta)E_F[m_j(W,\theta)]$$
$$= -v_{n,j}(\theta) + \sqrt{n}\hat{\sigma}_{n,j}^{-1}(\theta)\bar{m}_j(\theta)\sigma_{F,j}^{-1}(\theta)\hat{\sigma}_{n,j}(\theta).$$

In the case of $j \leq p$, $\theta \in \hat{\Theta}_I^{UB} \subseteq \Theta$, and Eq. (D.15) then implies that

$$\inf_{\theta\in\hat{\Theta}_I^{UB}} \sqrt{n}\sigma_{F,j}^{-1}(\theta)E_F[m_j(W,\theta)]$$
$$\geq -\sup_{\tilde{\theta}\in\Theta} |v_{n,j}(\tilde{\theta})| - \varpi_1(\ln \kappa_n)^{1/(2\chi)} \sup_{\check{\theta}\in\Theta} |\sigma_{F,j}^{-1}(\check{\theta})\hat{\sigma}_{n,j}(\check{\theta})|.$$

In the case of $j > p$, the same argument implies that

$$\sup_{\theta\in\hat{\Theta}_I^{UB}} \sqrt{n}\sigma_{F,j}^{-1}(\theta)E_F[m_j(W,\theta)]$$
$$\leq \sup_{\tilde{\theta}\in\Theta} |v_{n,j}(\tilde{\theta})| + \varpi_1(\ln \kappa_n)^{1/(2\chi)} \sup_{\check{\theta}\in\Theta} |\sigma_{F,j}^{-1}(\check{\theta})\hat{\sigma}_{n,j}(\check{\theta})|.$$

One can combine the information $\forall j \in \{1, \ldots, k\}$ to deduce that

$$\max_{j=1,\ldots,k} \sup_{\theta\in\hat{\Theta}_I^{UB}} |\sqrt{n}\sigma_{F,j}^{-1}(\theta)E_F[m_j(W,\theta)]|$$
$$\leq \sup_{\tilde{\theta}\in\Theta} \|v_n(\tilde{\theta})\| + \varpi_1(\ln \kappa_n)^{1/(2\chi)} \sup_{\check{\theta}\in\Theta} \|D_F^{-1/2}(\check{\theta})\hat{D}_n^{1/2}(\check{\theta})\|.$$

From this, it follows that

$$\liminf_{n\to\infty} \inf_{F\in\mathcal{P}_0} P_F\left(\max_{j=1,\ldots,k} \sup_{\theta\in\hat{\Theta}_I^{UB}} |\sqrt{n}\sigma_{F,j}^{-1}(\theta)E_F[m_j(W,\theta)]|\right.$$
$$\left.\leq \varpi_2(\ln \kappa_n)^{-1/\chi}\right)$$
$$\geq \liminf_{n\to\infty} \inf_{F\in\mathcal{P}_0} P_F\left(\sup_{\tilde{\theta}\in\Theta} \|v_n(\tilde{\theta})\|\right.$$
$$\left.+ \varpi_1\eta_n^{1/(2\chi)} \sup_{\theta\in\Theta} \|D_F^{-1/2}(\theta)\hat{D}_n^{1/2}(\theta)\| \leq \varpi_2(\ln \kappa_n)^{-1/\chi}\right)$$
$$\geq \liminf_{n\to\infty} \inf_{F\in\mathcal{P}_0} P_F\left(\sup_{\theta\in\Theta} \|v_n(\theta)\| \leq \varpi_2(\ln \kappa_n)^{-1/\chi}/2\right)$$
$$+ \liminf_{n\to\infty} \inf_{F\in\mathcal{P}_0} P_F\left(\sup_{\theta\in\Theta} \|D_F^{-1/2}(\theta)\hat{D}_n^{1/2}(\theta)\|\right.$$
$$\left.\leq \varpi_2(\ln \kappa_n)^{1/(2\chi)}/(2\varpi_1)\right) - 1$$
$$= \lim_{n\to\infty} P_{F_{a_n}}\left(\sup_{\theta\in\Theta} \|v_{a_n}(\theta)\| \leq \varpi_2(\ln \kappa_{a_n})^{1/\chi}/2\right)$$
$$+ \lim_{n\to\infty} P_{F_{a_n}}\left(\sup_{\theta\in\Theta} \|D_{F_{a_n}}^{-1/2}(\theta)\hat{D}_{a_n}^{1/2}(\theta)\|\right.$$
$$\left.\leq \varpi_2(\ln \kappa_{a_n})^{1/(2\chi)}/(2\varpi_1)\right) - 1 = 1,$$

where the first two inequalities are elementary, the first equality holds for a subsequence $\{a_n\}_{n\geq1}$ of $\{n\}_{n\geq1}$ that achieves the limit, the infimum, and is s.t. $\Omega_{F_{a_n}} \xrightarrow{u} \Omega$ for some $\Omega \in \mathcal{C}(\Theta^2)$ (which can be found by Lemma D.7), and the final equality follows from $\ln \kappa_n$, and parts 3 and 6 of Lemma D.2. □

D.6. Auxiliary results on consistency

Theorem D.1. *Assume Assumptions A.1–A.4 and let $F \in \mathcal{P}/\mathcal{P}_0$ be s.t. $h_F(\theta) \equiv D_F^{-1/2}(\theta)E_F[m(W,\theta)] : \Theta \to \mathbb{R}^k$ is continuous. Then,*

$$\lim E_F[\phi_n^{BP}] = \lim E_F[\phi_n^{RC}] = \lim E_F[\phi_n^{RS}] = 1.$$

Proof. By Theorem 6.1, it suffices to show $\lim E_F[\phi_n^{BP}] = 1$. For any $C > 0$, consider the following derivation:

$$\phi_n^{BP} = 1[\forall\theta \in \Theta : Q_n(\theta) > \hat{c}_n(\theta, 1-\alpha)]$$
$$\geq 1[\sup_{\theta\in\Theta} \hat{c}_n(\theta, 1-\alpha) \leq C \cap C < T_n],$$

where $T_n \equiv \inf_{\theta\in\Theta} Q_n(\theta)$ and $\hat{c}_n(\theta, 1-\alpha)$ is the (conditional) $(1-\alpha)$-quantile of $S(\hat{\Omega}_n^{1/2}(\theta)Z^* + \varphi(\xi_n(\theta), \hat{\Omega}_n(\theta)), \hat{\Omega}_n(\theta))$ with $Z^* \sim N(\mathbf{0}_k, I_k)$. From here it follows that

$$E_F[\phi_n^{BP}] \geq P_F(\sup_{\theta\in\Theta} \hat{c}_n(\theta, 1-\alpha) \leq C) + P_F(T_n > C) - 1.$$

To complete the proof, it suffices to find $\bar{C} > 0$ s.t. both expressions on the RHS are equal to one.

First, we show that $P_F(\sup_{\theta\in\Theta}\hat{c}_n(\theta, 1-\alpha) \leq \bar{C}) = 1$ for some constant $\bar{C}$. By monotonicity of $S(\cdot)$, $\hat{c}_n(\theta, 1-\alpha) \leq \hat{c}_n^{PA}(\theta, 1-\alpha)$, where $\hat{c}_n^{PA}(\theta, 1-\alpha)$ is the (conditional) $(1-\alpha)$-quantile of $S(\hat{\Omega}_n^{1/2}(\theta)Z^*, \hat{\Omega}_n(\theta))$ with $Z^* \sim N(\mathbf{0}_k, I_k)$. Since $\hat{\Omega}_n(\theta) \in \Psi$, $\hat{c}_n^{PA}(\theta, 1-\alpha) \leq \bar{C}$ where $\bar{C}$ is the (conditional) $(1-\alpha)$-quantile of $\sup_{\Omega\in\Psi} S(\Omega^{1/2}Z^*, \Omega)$ with $Z^* \sim N(\mathbf{0}_k, I_k)$. By definition, $\sup_{\theta\in\Theta}\hat{c}_n(\theta, 1-\alpha) \leq \bar{C} \in (0,\infty)$, which implies the desired result.

Second, we show that $P_F(T_n > \bar{C}) \to 1$. To show this, consider the following derivation:

$$\begin{aligned}
\inf_{\theta\in\Theta} S(\bar{m}_n(\theta), \hat{\Sigma}_n(\theta)) &= \inf_{\theta\in\Theta} S(D_F^{-1/2}(\theta)\bar{m}_n(\theta), \tilde{\Omega}_n(\theta)) \\
&\xrightarrow{p} \inf_{\theta\in\Theta} S(D_F^{-1/2}(\theta)E_F[m(W,\theta)], \Omega_F(\theta)) \\
&= \inf_{\theta\in\Theta} Q_F(\theta) > 0, \qquad \text{(D.16)}
\end{aligned}$$

where we use $\tilde{\Omega}_n(\theta) \equiv D_F^{-1/2}(\theta)\hat{\Sigma}_n(\theta)D_F^{-1/2}(\theta)$ and $\Omega_F(\theta) \equiv D_F^{-1/2}(\theta)\Sigma_F(\theta)D_F^{-1/2}(\theta)$. The two equalities are elementary, the strict inequality follows from $F \in \mathcal{P}/\mathcal{P}_0$, and the convergence in probability is shown in the next paragraph. From Eq. (D.16), it follows that $T_n = n^{\chi/2}\inf_{\theta\in\Theta} S(\bar{m}_n(\theta), \hat{\Sigma}_n(\theta)) \xrightarrow{p} \infty$, which implies the desired result.

To complete the proof, it suffices to show the convergence in probability in Eq. (D.16). The main steps of this argument are similar to the ones used to prove Theorem 3.1. We now provide a basic sketch of the main ideas. By part 6 in Lemma D.2 (with $\lambda_n = \sqrt{n}$) we conclude that $D_F^{-1/2}(\cdot)\bar{m}_n(\cdot) \xrightarrow{p} D_F^{-1/2}(\cdot)E_F[m(W,\cdot)]$ in $l^\infty(\Theta)$. By part 2 in Lemma D.2 we conclude that $\tilde{\Omega}_n \xrightarrow{p} \Omega_F$ in $l^\infty(\Theta)$. Elementary arguments in convergence in probability then imply that $(D_F^{-1/2}(\cdot)\bar{m}_n(\cdot), \tilde{\Omega}_n) \xrightarrow{p} (D_F^{-1/2}(\cdot)E_F[m(W,\cdot)], \Omega_F)$ in $l^\infty(\Theta)$. Furthermore, under the assumptions of the result, $(D_F^{-1/2}(\cdot)E_F[m(W,\cdot)], \Omega_F) : \Theta \to \mathbb{R}^k_{[\pm\infty]} \times \Psi$ is continuous function. By these findings and the extended continuous mapping theorem (e.g. van der Vaart and Wellner (1996, Theorem 1.11.1)), the result follows. □

References

Andrews, D.W.K., Barwick, P.J., 2012. Inference for parameters defined by moment inequalities: A recommended moment selection procedure. Econometrica 80, 2805–2826.

Andrews, D.W.K., Guggenberger, P., 2009. Validity of subsampling and plug-in asymptotic inference for parameters defined by moment inequalities. Econometric Theory 25, 669–709.

Andrews, D.W.K., Shi, X., 2013. Inference based on conditional moment inequalities. Econometrica 81, 609–666.

Andrews, D.W.K., Soares, G., 2010. Inference for parameters defined by moment inequalities using generalized moment selection. Econometrica 78, 119–157.

Armstrong, T.B., 2014. Weighted KS statistics for inference on conditional moment inequalities. J. Econometrics 181, 92–116.

Bugni, F.A., 2010. Bootstrap inference in partially identified models defined by moment inequalities: Coverage of the identified set. Econometrica 78, 735–753.

Bugni, F.A., 2014. Comparison of inferential methods in partially identified models in terms of error in coverage probability. Econometric Theory First View, 1–56. http://dx.doi.org/10.1017/S0266466614000826.

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

Bugni, F.A., Canay, I.A., Shi, X., 2014. Inference for Functions of Partially Identified Parameters in Moment Inequality Models, CeMMAP Working Paper CWP05/14.

Canay, I.A., 2010. E.L. inference for partially identified models: Large deviations optimality and bootstrap validity. J. Econometrics 156, 408–425.

Chernozhukov, V., Hong, H., Tamer, E., 2007. Estimation and confidence regions for parameter sets in econometric models. Econometrica 75, 1243–1284.

Chernozhukov, V., Lee, S., Rosen, A.M., 2013. Intersection bounds: Estimation and inference. Econometrica 81, 667–737.

Chetverikov, D., 2013. Adaptive test of conditional moment inequalities, arXiv:1201.0167.

Corbae, D., Stinchcombe, M.B., Zeman, J., 2009. An Introduction to Mathematical Analysis for Economic Theory and Econometrics. Princeton University Press.

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

Guggenberger, P., Hahn, J., Kim, K., 2008. Specification testing under moment inequalities. Econom. Lett. 2, 375–378.

Haavelmo, T., 1944. The probability approach in econometrics. Econometrica 12, iii–vi+1–115.

Imbens, G., Manski, C.F., 2004. Confidence intervals for partially identified parameters. Econometrica 72, 1845–1857.

Kitamura, Y., Stoye, Y., 2012. Nonparametric Analysis of Random Utility Models, Working Paper, Cornell University.

Koopmans, T., 1937. Linear Regression Analysis of Economic Time Series. Netherlands Economic Institute, Haarlem.

Lehman, E., Romano, J.P., 2005. Testing Statistical Hypothesis, 3rd ed.. Springer.

Manski, C.F., 1989. Anatomy of the Selection Problem. J. Hum. Resour. 24, 343–360.

Manski, C.F., 2003. Partial Identification of Probability Distributions. Springer-Verlag.

McCloskey, A., 2014. On the Computation of Size-Correct Power-Directed Tests with Null Hypotheses Characterized by Inequalities (Preliminary Draft, Comments Welcome), Manuscript, Brown University.

Mikusheva, A., 2010. Robust Confidence Sets in the Presence of Weak Instruments. J. Econometrics 157, 234–247.

Ponomareva, M., Tamer, E., 2011. Misspecification in moment inequality models: Back to moment equalities?. Econom. J. 14, 186–203.

Romano, J.P., Shaikh, A.M., 2008. Inference for identifiable parameters in partially identified econometric models. J. Statist. Plann. Inference 138, 2786–2807.

Romano, J.P., Shaikh, A.M., Wolf, M., 2014. A practical two-step method for testing moment inequalities. Econometrica 82, 1979–2002.

Royden, H.L., 1988. Real Analysis. Prentice-Hall.

Santos, A., 2012. Inference in nonparametric instrumental variables with partial identification. Econometrica 80, 213–275.

Tamer, E., 2003. Partial identification in econometrics. Annu. Rev. Econ. 2, 167–195.

van der Vaart, A., 1998. Asymptotic Statistics. Cambridge University Press.

van der Vaart, A., Wellner, J., 1996. Weak Convergence and Empirical Processes. Springer.