Specification tests for partially identified models defined by moment inequalities

Federico A. Bugni
Ivan A. Canay
Xiaoxia Shi

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP01/13

Specification Tests for Partially Identified Models defined by Moment Inequalities*

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

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

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

January 17, 2013

Abstract

This paper studies the problem of specification testing in partially identified models defined by a finite number of moment equalities and inequalities (i.e., (in)equalities). Under the null hypothesis, there is at least one parameter value that simultaneously satisfies all of the moment (in)equalities whereas under the alternative hypothesis there is no such parameter value. While this problem has not been directly addressed in the literature (except in particular cases), several papers have suggested implementing this inferential problem by checking whether confidence intervals for the parameters of interest are empty or not.

We propose two hypothesis tests that use the infimum of the sample criterion function over the parameter space as the test statistic together with two different critical values. We obtain two main results. First, we show that the two tests we propose are asymptotically size correct in a uniform sense. Second, we show our tests are more powerful than the test that checks whether the confidence set for the parameters of interest is empty or not.

Keywords: Partial Identification, Moment Inequalities, Specification Tests, Hypothesis Testing.

JEL Classification: C01, C12, C15.

*We 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, and seminar participants at IUPUI, UC Berkeley, UC Davis, Yale, Columbia, Rochester, and Ohio State University for helpful comments. Bugni and Canay thank the National Science Foundation for research support via grants SES-1123771 and SES-1123586, respectively.

1 Introduction

This paper studies the problem of specification testing in partially identified models defined by a finite number of moment equalities and inequalities (i.e. (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

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

where $\{W_i\}_{i=1}^n$ is an i.i.d. sequence of random variables with distribution F and $m : \mathbb{R}^d \times \Theta \rightarrow \mathbb{R}^k$ is a known measurable function. This model is said to be *partially identified* because the sampling process and the maintained assumptions (i.e. 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 the entire space Θ but potentially larger than a single point. In this paper we propose two hypothesis tests for the null hypothesis of correct specification of the model in Eq. (1.1) that are based on the same test statistic but employ different critical values. We show that our two tests are asymptotically size correct in a uniform sense. Furthermore, we compare the power properties of our tests and the existent test proposed in the literature, which consists in testing whether confidence sets for θ are empty or not, and obtain two main findings. In finite samples, our tests have more or equal power than the existing test (for any alternative hypothesis). In addition, we show that there exist sequences of local alternative hypotheses for which our tests have strictly higher asymptotic power than the existing test.

A model is said to be correctly specified (or, statistically adequate) when the underlying assumptions are supported by the observed data.[1] The motivation behind the 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.[2] 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, Canay, and Guggenberger, 2012).

Specification tests for partially identified models have been studied in Guggenberger, Hahn, and Kim (2008), Romano and Shaikh (2008), Andrews and Guggenberger (2009), Andrews and Soares (2010), and 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 non-parametric instrumental variable model and, thus, his results are not directly applicable to the model in Eq. (1.1). To the best of our knowledge, there is only one specification test for the model in Eq. (1.1) that has been described in the

[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).

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

literature and has been shown to have correct (uniform) asymptotic size. This test arises as a by-product of confidence sets for partially identified parameters and has been proposed by Romano and Shaikh (2008, Remark 3.7), Andrews and Guggenberger (2009, Section 7), and Andrews and Soares (2010, Section 5). More specifically, the test rejects the null hypothesis of correct specification whenever the confidence set for θ is empty. We take this test as the main competitor of the new tests proposed in this paper and so we describe it formally in the next section. In what follows we will call this test "Test BP", to emphasize that it comes as a *by-product* of confidence sets for θ.

This paper makes two main contributions. Our first contribution is to introduce two new specification tests and show that both of them have correct asymptotic size and can be significantly less conservative than Test BP. Our second contribution is to show that our new tests have asymptotic power that is at worst equal to that of Test BP against all sequences of local alternatives, and that can be strictly higher than the asymptotic power of Test BP for certain sequences of local alternatives. We therefore conclude that the hypothesis tests proposed in this paper have better statistical properties that Test BP.

The central reason for this size and power advantages of our specification tests is that they were designed with the sole purpose of testing the specification of the model, in contrast to Test BP which was conceived while pursuing a different goal. In particular, our tests involve a different test statistic (which is the infimum over Θ of the usual test statistic used to construct confidence sets for θ) in the same spirit of the popular J-test in point-identified moment equality models. The main technical challenge in constructing tests based on this test statistic lies in the construction of valid and computationally feasible critical values. We propose two methods with relative merits depending on the context that result in tests that have correct asymptotic size and are less conservative than Test BP.

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 therefore 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 remainder 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. Section 3 introduces the test statistic we use in the construction of the two new specification tests. The description of the tests is then completed by combining this test statistic with appropriate critical values that are introduced 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 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 mild assumptions that have been assumed elsewhere already) and regular assumptions indexed by the letter A (to denote the assumptions that are fundamental for our results).

2 Framework

The objective of our inferential procedure is to test whether the moment conditions in Eq. (1.1) are valid or not, 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, when paired with some $\theta \in \Theta$, the following conditions hold:

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

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

(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, 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, when paired with some $\theta \in \Theta$, the following conditions hold:

- 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 the parameter vector (θ, F) satisfies all conditions 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,

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

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

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

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

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

For any $\alpha \in (0,1)$, the test is said to be asymptotically level α if $AsySz \leq \alpha$ and it is said to be asymptotically size correct if $AsySz = \alpha$. The importance of the distinction between tests that satisfy $AsySz \leq \alpha$ rather than pointwise requirement

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

has been emphasized in much of the recent literature on inference in partially identified models. See, e.g., Imbens and Manski (2004), Romano and Shaikh (2008), Andrews and Soares (2010), and Mikusheva (2010).

2.1 The existent specification test

This section formally introduces the 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, Hong, and Tamer (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

$$Q_F(\theta) = 0 \text{ if and only if } \theta \in \Theta_I(F) \ . \tag{2.4}$$

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.[3] 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

[3]See Assumptions M.4-M.8 in the Appendix for these regularity conditions.

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

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

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

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

where $c_n(\theta,1-\alpha)$ is such that $CS_n(1-\alpha)$ has the correct asymptotic coverage, i.e.,

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

The condition in Eq. (2.9) guarantees uniform coverage over the space $\mathcal{F}_0$, which denotes all 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

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

Given Eq. (2.9), it follows that Test BP has an asymptotic size bounded above by α (see Theorem C.3 in the Appendix). However, as pointed out in Andrews and Guggenberger (2009) and Andrews and Soares (2010), this test is admittedly conservative. 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 the construction of $CS_n(1-\alpha)$. 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) 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, the 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 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 to produce Figure 1, as it captures the types of situations where there are power gains of implementing the specification tests proposed in this paper instead of Test BP (see Figure 1).

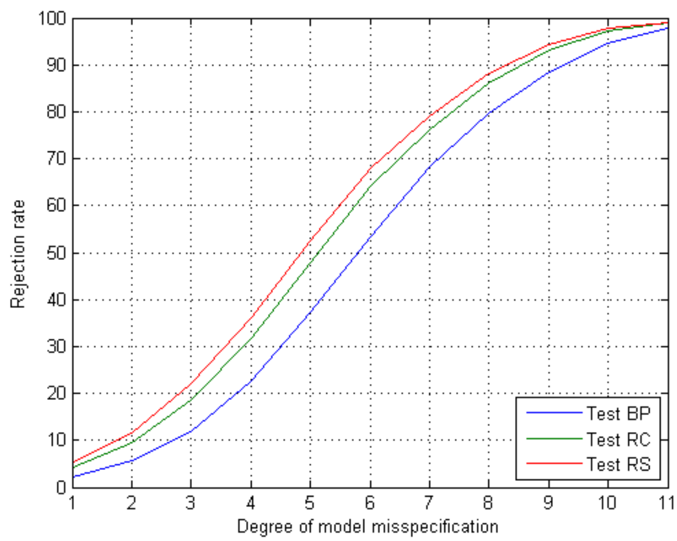

Figure 1: Rejection rates in the Example 2.1. Each line corresponds to the rejection rates in Table 2 in the Monte Carlo section.

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

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

where G is a known parametric 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 $l = 1, \ldots, d_x$:

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

We now choose a simple parametrization that we use in our Monte Carlo simulations. Suppose that $S_X = \{x_1 = (1,0,0), x_2 = (-1,0,1), x_3 = (0,1,0)\}$, that Y represents a non-negative outcome variable without an upper bound, i.e., $Y_L(x) = 0$ and $Y_H(x) = \infty$, that G is the linear model $G(x,\theta) = x'\theta$, $\theta = (\theta_1, \theta_2, 1)$, 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

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

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

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

It follows that this model is strictly partially identified (i.e., if a solution exists, it is always multiple) and it is correctly specified 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 share a common test statistic, which is related to the test statistic $Q_n(\theta)$ defined in Eq. (2.6). The justification for the new test statistic we propose follows immediately under the following two mild assumptions, that 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 hypothesis test in Eq. (2.1) can be re-written as

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

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., $\inf_{\theta\in\Theta} Q_n(\theta)$. 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

$$\phi_n = 1\left\{ \inf_{\theta\in\Theta} Q_n(\theta) > \hat{c}_n(1-\alpha) \right\} \ , \tag{3.3}$$

where $\hat{c}_n(1-\alpha)$ is an approximation to the $(1-\alpha)$ quantile of the asymptotic distribution of $\inf_{\theta\in\Theta} Q_n(\theta)$.

In order to make the test in Definition 3.3 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. (3.3) that results in a test that: (a) controls asymptotic size, (b) has desirable power properties, and (c) is amenable to computation. Towards this end, we propose two critical values that result in two hypothesis tests that satisfy these requirements. The first critical value is based on a bootstrap approximation of the distribution of the test statistic under the null hypothesis. The second critical value is based on a simple bounding argument based on "recycling" GMS critical values. We describe each of these critical values in the next two sections.

Before introducing the critical values, it is convenient to describe the $(1-\alpha)$ quantile we wish to approximate. In order to do this, we need to describe the limit distribution of the test statistic $\inf_{\theta\in\Theta} Q_n(\theta)$ along sequences of distributions $\{F_n\}_{n\geq 1}$ with $F_n \in \mathcal{P}$ for all $n \geq 1$. We do this by imposing the following four assumptions (Appendix A contains further details on the notation used below).

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 $v_n(\cdot)$ with j-component

$$v_{n,j}(\theta) = \sqrt{n}\sigma_{F_n,j}^{-1}(\theta)\sum_{i=1}^{n}(\bar{m}_j(W_i,\theta) - \bar{m}_n(\theta)), \quad j = 1, \ldots, k \ , \tag{3.4}$$

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 $j = 1, \ldots, k$,

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

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

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

The matrix Ω_F satisfies

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

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, for all $j = 1, \ldots, k$, the class of functions $\{\sigma_{F,j}^{-1}(\theta) m_j(\cdot, \theta) : \mathcal{W} \to \mathbb{R}\}$ is Donsker and pre-Gaussian uniformly in $F \in \mathcal{P}$ (see Lemma D.2). 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 $\overset{u}{\to}$ and $\overset{H}{\to}$ to denote uniform convergence and convergence in Hausdorff distance (see Appendix A). Finally, we let $D_F(\theta) \equiv Diag(\Sigma_F(\theta))$, and define for any $n \in \mathbb{N}$ and $F \in \mathcal{P}$ the following subset of $\Theta \times \mathbb{R}^k$,

$$\Lambda_{n,F} \equiv \left\{ (\theta, \ell) \in \Theta \times \mathbb{R}^k : \ell = \sqrt{n} D_F^{-1/2}(\theta) E_F[m(W, \theta)] \right\} . \tag{3.5}$$

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) $\Omega_{F_n}(\theta, \theta') \overset{u}{\to} \Omega(\theta, \theta')$, and (iii) $\Lambda_{n,F_n} \overset{H}{\to} \Lambda$. Then, along the (sub)sequence $\{F_n\}_{n\geq 1}$*

$$T_n \equiv \inf_{\theta \in \Theta} Q_n(\theta) \overset{d}{\to} J(\Lambda, \Omega) \equiv \inf_{(\theta,\ell) \in \Lambda} S(v_\Omega(\theta) + \ell, \Omega(\theta, \theta)), \text{ as } n \to \infty , \tag{3.6}$$

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 (for additional details, see Appendix C).

The limit distribution $J(\Lambda, \Omega)$ in Theorem 3.1 depends on the set Λ and the function Ω, and so does its $1 - \alpha$ quantile, which we denote by $c_\Lambda(\Omega, 1 - \alpha)$. Our goal is to construct feasible critical values that approximate $c_\Lambda(\Omega, 1 - \alpha)$ asymptotically. Unfortunately, this is a difficult task as Λ cannot be consistently estimated in a uniform way. In the next two sections, we propose two approaches to circumvent this problem.

4 Test RS: Re-Sampling critical value

The first approach is based on the use of resampling methods to approximate $J(\Lambda, \Omega)$, which require 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 two main difficulties. The first one can be related to the difficulty described in Andrews and Soares (2010) of estimating the slackness parameter for the moment inequalities. The second one is related to the fact that the set Λ includes limit points of random sequences that contain tuples (θ_n, F_n) such that $\theta_n \notin \Theta_I(F_n)$ for some or all $n \in \mathbb{N}$. This second difficulty is novel to this paper.

To be more precise, Andrews and Soares (2010) denote the slackness parameter that indicates whether a moment inequality is binding, close to be binding, or not binding, by

$$h_{1,j} = \lim_{n\to\infty} \sqrt{n}\sigma^{-1}_{F_n,j}(\theta_n)\, E_{F_n}[m_j(W,\theta_n)],\ \text{ for } j = 1,\ldots,p\ . \tag{4.1}$$

This slackness parameter is only defined for (non-random) sequences of parameters $\{(\theta_n, F_n)\}_{n\geq 1}$ with $\theta_n \in \Theta_I(F_n)$ for all $n \in \mathbb{N}$ (implying $h_{1,j} \in [0,\infty]$) and it cannot be consistently estimated. The GMS approach proposed by Andrews and Soares (2010) takes advantage of the weak monotonicity of the test function and then replaces $h_{1,j}$ with a function $\varphi_j(\cdot)$ of the following sample measure of slackness,

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

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

Our problem is similar in the sense that Λ contains the cluster points of the sequence

$$\{(\theta_n, \sqrt{n}D^{-1/2}_{F_n}(\theta_n)E_{F_n}[m(W,\theta_n)])\}_{n\geq 1}\ . \tag{4.3}$$

However, the presence of an infimum over θ in our problem implies that the sequence θ_n is now random. This is an important technical difficulty, as most of the results in Andrews and Soares (2010) cannot be extended to random sequences. More importantly, in our setup all we know under the null hypothesis is that $F_n \in \mathcal{P}_0$ for all $n \in \mathbb{N}$. By definition of $\mathcal{P}_0$ (see Definition 2.2), this means that there exists $\theta^\star_n$ such that $\theta^\star_n \in \Theta_I(F_n)$ for all $n \in \mathbb{N}$. There is, however, no guarantee that the random sequence θ_n in Eq. (4.3) satisfies $\theta_n \in \Theta_I(F_n)$. In fact, in most standard models one can construct cases in which $\theta_n \notin \Theta_I(F_n)$ for all $n \in \mathbb{N}$. This is problematic because it implies that the set Λ contains tuples (θ,ℓ) such that $\ell_j < 0$ for $j = 1,\ldots,p$, or $\ell_j \neq 0$ for $j = p+1,\ldots,k$ and so, if an infimum is attained, it could be attained at a value of θ that is not associated with $\ell_j \geq 0$ for $j = 1,\ldots,p$ and $\ell_j = 0$ for $j + p + 1,\ldots,k$.

Despite the aforementioned difficulties, we can approximate the quantiles of $J(\Omega,\Lambda)$ much in the spirit of Andrews and Soares (2010), with the addition of an approximation of the identified set $\Theta_I(F)$ to guarantee that the limit points of the elements (θ,ℓ) in the resampling approximation of the set Λ are such that $\ell_j \geq 0$ for $j = 1,\ldots,p$ and $\ell_j = 0$ for $j = p+1,\ldots,k$. The exact approximation of $\Theta_I(F)$ we use is described in the next definition.

Definition 4.1. Let $\{\tau_n\}_{n\geq 1}$ be a non-stochastic sequence defined as $\tau_n = \kappa_n^r$ for some $r \in (0,1)$ where $\{\kappa_n\}_{n\geq 1}$ is as in Assumption M.1, and let $q^\star_n \equiv \inf_{\theta\in\Theta} S(\tau_n^{-1}\sqrt{n}\bar{m}_n(\theta), \hat{\Sigma}_n(\theta))$. The approximation to the identified set is defined as

$$\hat{\Theta}_I \equiv \{\theta \in \Theta : S(\tau_n^{-1}\sqrt{n}\bar{m}_n(\theta), \hat{\Sigma}_n(\theta)) - q^\star_n \leq 1\}\ . \tag{4.4}$$

The set $\hat{\Theta}_I$ is the standard estimator of the identified set, based on a normalized sample criterion function centered at the infimum. This normalization is useful in cases where the infimum of the sample criterion function is not zero in finite samples (see Chernozhukov et al., 2007, p. 1247) and avoids having an estimator of the identified set that can be empty with positive probability. The estimator $\hat{\Theta}_I$ is non-empty by definition.

Now we can define the resampling test statistic that we use to construct an approximation to $c_\Lambda(\Omega, 1-\alpha)$. 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. Consider the following test statistic

$$T^*_n = \inf_{\theta\in\hat{\Theta}_I} S(\hat{v}^*_n(\theta) + \varphi(\xi_n(\theta), \hat{\Omega}_n(\theta)), \hat{\Omega}_n(\theta))\ ,$$

where for $\hat{D}_n(\theta) \equiv Diag(\hat{\Sigma}_n(\theta))$, and $\hat{\Sigma}_n(\theta)$ the estimator in Eq. (2.7),

$$\hat{\Omega}_n(\theta) \equiv \hat{D}_n^{-1/2}(\theta)\hat{\Sigma}_n(\theta)\hat{D}_n^{-1/2}(\theta) \ ,$$

$\xi_n(\theta) = \{\xi_{n,j}(\theta)\}_{j=1}^p$ with $\xi_{n,j}(\theta)$ is as in Eq. (4.2), and $\varphi = (\varphi_1, \ldots, \varphi_p, \mathbf{0}_{k-p})' \in \mathbb{R}^k_{[+\infty]}$ is the function in Assumption M.1 that is assumed to satisfy the assumptions in Andrews and Soares (2010). Examples of φ include $\varphi_j(\xi, \Omega) = \infty I(\xi_j > 1)$ (where we use the convention $\infty 0 = 0$), $\varphi_j(\xi, \Omega) = \max\{\xi_j, 0\}$, and $\varphi_j(\xi, \Omega) = \xi_j$ for $j = 1, \ldots, p$ (see Andrews and Soares, 2010, for additional examples).

Conditional on the sample, the distribution of T_n^* is known and its quantiles can be approximated by Monte Carlo simulation. For example, the stochastic process $\hat{v}_n^*$ can be simulated via a bootstrap approximation, in which case

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

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

$$\hat{v}_n^*(\theta) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \hat{D}_n^{-1/2}(\theta)(m(W_i, \theta) - \bar{m}_n(\theta))\zeta_i \ , \tag{4.6}$$

where $\{\zeta_i\}_{i=1}^n$ is an i.i.d. sample satisfying $\zeta_i \sim N(0, 1)$. We can now define 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

$$\phi_n^{RS} = 1\left\{\inf_{\theta \in \Theta} Q_n(\theta) > \hat{c}_n^{RS}(1 - \alpha)\right\} \ , \tag{4.7}$$

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

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

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

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

In order to understand the result in 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. (3.5). This can be done by noting that

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

where

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

Therefore, the resampling approach behind test RS in Definition 4.2 consists in replacing the set Λ with the approximation $\hat{\Lambda}_n^*$, which is (generally) not consistent for Λ. The important distinction here, relative to a standard resampling method, is that the set $\hat{\Lambda}_n^*$ restricts $\theta \in \hat{\Theta}_I$ as opposed to $\theta \in \Theta$ in $\Lambda_{n,F}$. In fact, it is possible to show that if we were to define the set $\hat{\Lambda}_n^*$ with Θ replacing $\hat{\Theta}_I$, this would result in an approach that would not control asymptotic size as in Theorem 4.1 for functions φ satisfying Assumption A.5. In other words, under the assumptions of the theorem, using a similar statistic to T_n^* but with an infimum over Θ (as it is the case for the original test statistic) does not provide a valid asymptotic approximation.[4]

Once we re-write the test statistic as above, the result in Theorem 4.1 follows from three intermediate steps that use the following expansion of $\Theta_I(F)$.

Definition 4.3. Let $\eta_n \equiv \tau_n \log \kappa_n$ and $\Theta_I^{\eta_n}(F)$ be defined as

$$\Theta_I^{\eta_n}(F) \equiv \{\theta \in \Theta : S(\eta_n^{-1}\sqrt{n}E_F[m(W,\theta)], \Sigma_F(\theta)) \leq 1\} \ .$$

Remark 4.1. $\Theta_I^{\eta_n}(F)$ is a blow-up of $\Theta_I(F)$, with an amount of expansion proportional to $\eta_n/\sqrt{n} = o(1)$. Moreover, Lemma D.13 in Appendix C shows that

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

The first step is immediate and consists in noticing that $T_n \leq \tilde{T}_n$ for all $n \in \mathbb{N}$, where $\tilde{T}_n$ is defined similarly to T_n but with an infimum over $\theta \in \Theta_I^{\eta_n}(F)$. If we let $J(\Lambda', \Omega)$ denote the limit distribution of $\tilde{T}_n$, it follows from $\Theta_I^{\eta_n}(F) \subseteq \Theta$ that

$$J(\Lambda, \Omega) \leq J(\Lambda', \Omega) \ . \tag{4.10}$$

The advantage of working with this (infeasible) test statistic, is that now the limit set Λ' does not contain limit points of sequences $\{(\theta_n, F_n)\}_{n\geq 1}$ with θ_n far from $\Theta_I(F_n)$.

The second step is to note that Remark 4.1 implies that, asymptotically, $\tilde{T}_n^* \leq T_n^*$, where $\tilde{T}_n^*$ is defined as T_n^* with $\hat{\Lambda}_n^*$ replaced by

$$\{(\theta, \ell) : \theta \in \Theta_I^{\eta_n}(F_n),\ \ell = \varphi^*(\xi_n(\theta))\} \ . \tag{4.11}$$

This representation is useful for two reasons: the set $\Theta_I^{\eta_n}(F_n)$ is non-random and the function φ^* is continuous (which is not necessarily the case for the original function φ) and does not depend on $\hat{\Omega}_n(\theta)$ (see Theorem C.2 for details). If we denote by $J(\Lambda^\dagger, \Omega)$ the limit distribution of the test statistic $\tilde{T}_n^*$, we then show that

$$J(\Lambda', \Omega) \leq J(\Lambda^\dagger, \Omega) \ . \tag{4.12}$$

The combination of Eqs. (4.10) and (4.12) ensure that the critical value $\hat{c}_n^{RS}(1-\alpha)$ is uniformly valid.

Remark 4.2. Note that the sequence η_n and the set $\Theta_I^{\eta_n}(F_n)$ are used in the proof at intermediate steps but are not needed for the implementation of the test ϕ_n^{RS}.

Remark 4.3. The definition of the set in Eq. (4.11) assumes the existence of a function $\varphi^*(\cdot)$ that is continuous and satisfies $\varphi \leq \varphi^*$. Such assumption is not restrictive. For example, it is satisfied for the functions $\varphi^{(1)} - \varphi^{(4)}$ described in Andrews and Soares (2010) and Andrews and Barwick (2012) (see Remark B.1 in Appendix B).

Remark 4.4. Test RS, although feasible, might be computationally demanding in some applications. The main complexity is that it requires to simulate a stochastic process in θ with a covariance kernel that converges to $\Omega(\theta, \theta')$ asymptotically. In practice, this is achieved by simulating $\hat{v}_n^*(\theta)$ for each $\theta \in \hat{\Theta}_I$ as in

[4]It is worth pointing out that a special choice of the function φ can be shown to circumvent the problem and result in a test that controls asymptotic size. 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.

Equation (4.6). Letting B denote the number of simulation draws, this means that for each $\theta \in \hat{\Theta}_I$ Test RS requires $n \times B$ simulations from a standard normal random variable, which increases proportionally to the sample size.

5 Test RC: Re-Cycling existent critical values

This section proposes a second method to compute critical values that can be computationally attractive in many models and generates power advantages vis-à-vis Test RS. The critical value for the Test RS is based on approximating the quantile of the infimum of the criterion function $Q_n(\theta)$ over all $\theta \in \Theta$. In contrast, the main idea behind the critical value for the Test RC is to bound the quantile of the infimum with the infimum of the approximated quantiles, which are already available in the literature. Formally, the Test RC is defined as follows.

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

$$\phi_n^{RC} = 1\left\{\inf_{\theta \in \Theta} Q_n(\theta) > \hat{c}_n^{RC}(1-\alpha)\right\} , \tag{5.1}$$

where $\hat{c}_n^{RC}(1-\alpha)$ is given by

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

and $\hat{c}_n(\theta, 1-\alpha)$ is the GMS critical value defined in Assumption M.1.

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

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

$$\limsup_{n\to\infty} \sup_{F \in \mathcal{P}_0} E_F[\phi_n^{RC}] \leq \alpha . \tag{5.3}$$

Remark 5.1. The computational cost of Test RC is only marginally higher than that of the Test BP, as Test BP already requires $\hat{c}_n(\theta, 1-\alpha)$ for all $\theta \in \Theta$ and Test RC just takes an infimum over $\hat{\Theta}_I$ of these critical values. In practice the infimum in Eq. (5.2) is just a minimum over a finite grid, as $\hat{\Theta}_I$ is typically a finite collection of points. The additional cost is therefore limited to the computation of $\hat{\Theta}_I$.

Remark 5.2. The computational advantage of Test RC over Test RS comes from the fact that Test RC requires to simulate the Gaussian process $\hat{v}_n^*(\theta)$ at each particular θ, which is a normal random variable with zero mean and a variance $\Omega(\theta)$. In other words, the correlation of $\hat{v}_n^*(\theta)$ and $\hat{v}_n^*(\theta')$ for $\theta \neq \theta'$ is not needed to compute Test RC. In practice, this means that for each $\theta \in \hat{\Theta}_I$, Test RC requires $k \times B$ simulations from a standard normal random variable, which does not increase with the sample size, c.f. Remark 4.4.

Remark 5.3. Theorem 5.1 follows immediately from Theorem 4.1 once we notice that

$$\hat{c}_n^{RS}(1-\alpha) \leq \inf_{\theta \in \hat{\Theta}_I} \hat{c}_n(\theta, 1-\alpha) = \hat{c}_n^{RC}(1-\alpha) , \tag{5.4}$$

as the quantile of an infimum is weakly smaller than the infimum of the quantiles. See the proof of Theorem 6.1 in the next section for details.

6 Power analysis

It follows from previous results that the three specification tests we study in this paper are asymptotically size correct. In this section we show that the two tests we propose, Test RS and test RC, weakly dominate Test BP in terms of finite sample power for all $n \in \mathbb{N}$, and can even strictly dominate Test BP in terms of asymptotic power for certain types of local alternatives. These results are summarized in the next two theorems.

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

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

Corollary 6.1. *For any sequence of local alternatives* $\{F_n \in \mathcal{P}/\mathcal{P}_0\}_{n \geq 1}$*,*

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

The proof of Theorem 6.1 is in Appendix C, while 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.

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 and 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. Let $\hat{\Theta}_I$ be the set defined in Definition 4.1. The sequence $\{F_n \in \mathcal{P}\}_{n \geq 1}$ satisfies the following properties: (i) there is a sequence $\{\theta_n^* \in \Theta\}_{n \geq 1}$ s.t. $Q_n(\theta_n^*) = T_n$ and $\hat{c}_n(\theta_n^*, 1-\alpha) \xrightarrow{p} c_2$, (ii) there is a sequence $\{\theta_n \in \hat{\Theta}_I\}_{n \geq 1}$ s.t. $\hat{c}_n(\theta_n, 1-\alpha) \xrightarrow{p} c_1$, and (iii) $T_n \xrightarrow{d} J$ and $P(J \in (c_1, c_2)) > 0$.

The intuition behind Assumption A.9 is illustrated in Figure 2, which is based on Example 2.1. In words, Assumption A.9 requires that the standard estimator of the identified set includes at least two points asymptotically along the sequence of alternatives, and that the quantiles of the limit distribution of $Q_n(\theta)$ are different at each of these two points. For example, in Figure 2 the set of minimizers of $Q_n(\theta)$ is a line and therefore $\hat{\Theta}_I$ is a rectangle. In particular, the critical value at the point labeled "1, 2, 3 active" is one for which there are three moment inequalities that are active (and it converges to c_2 in the notation of the assumption), while the critical value at the point labeled "1, 2 active" is one for which there are only two moment inequalities that are active (and so it converges to c_1 in the notation of the assumption). The value of $Q_n(\theta)$ is the same at these two points. It is therefore clear that whenever the set of minimizers of $Q_n(\theta)$ is not a singleton and the distribution of $Q_n(\theta)$ is not the same along the minimizers, Assumption A.9 will be satisfied.

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

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

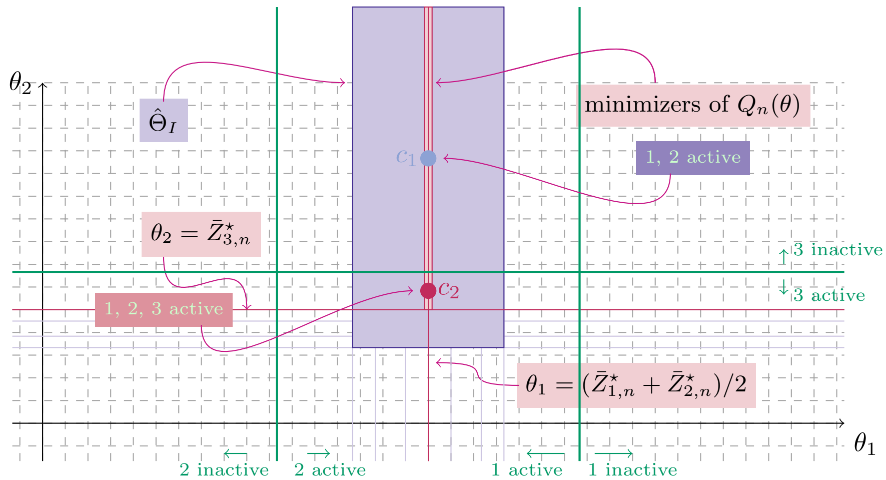

Figure 2: Illustration of Assumption A.9 . The identified set is defined in Eq. (7.2) and $Z^\star_j \equiv \{Y_i Z_i | X = x_j\}$ for $j = 1, 3$ and $Z^\star_2 \equiv 1 - \{Y_i Z_i | X = x_2\}$, where $\{Y_i Z_i | X = x_j\}$ are defined as in Eq. (7.1).

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 theorem with Theorem 6.1, it follows that Test RS is also strictly more powerful than Test BP asymptotically.

Remark 6.1. If one considers sequences of alternatives under which the inequality in Eq. (5.4) becomes strict, it is then possible that Test RS becomes strictly more powerful than Test RC for such alternatives.

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 $W_i \equiv (X_i, Y_i Z_i)$ is distributed as follows

$$\begin{aligned} &P(X_i = x_1) = P(X_i = x_2) = P(X_i = x_3) = 1/3 \ , \\ &\{Y_i Z_i | X = x_1\} \sim N(0,1), \quad \{Y_i Z_i | X = x_2\} \sim N(1+\eta, 1), \quad \{Y_i Z_i | X = x_3\} \sim N(0,1) \ , \end{aligned} \tag{7.1}$$

for different values of $\eta \in \mathbb{R}$. By plugging in this information into Eq. (2.14), we can derive the following closed form solution for the identified set

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

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 the model is misspecified, i.e., the identified set is empty.

The results from the simulations are collected in Tables 1 and 2. The parameters we use to produce both tables are as follows: $\alpha = 10\%$, $\kappa_n = \sqrt{\log n}$, $n \in \{100, 500\}$, and $\tau_n \in \{\sqrt{\log \log n}, \kappa_n/3, 2\kappa_n/3\}$. The number of replications is set to $MCsize = 2,000$. Our simulation results are consistent with our theoretical findings.

Under the null hypothesis (i.e. $\eta \leq 0$) all tests considered in this paper provide size control (i.e. rejection rate does not exceed α). In fact, these tests appear to be slightly conservative in the current setup (i.e.

Test	τ_n	η -0.1	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
BP	N/A	0.75	2.20	5.65	11.90	22.55	37.15	53.40	68.30	79.75	88.55	94.50	97.85
RC	$\sqrt{\log\log n}$	1.65	4.05	9.55	18.55	31.85	47.95	64.25	76.15	86.05	93.00	97.35	99.00
RC	$(1/3)\kappa_n$	1.65	4.05	9.55	18.55	31.85	47.95	64.25	76.15	86.05	93.00	97.35	99.00
RC	$(2/3)\kappa_n$	1.65	4.05	9.55	18.55	31.85	47.95	64.25	76.15	86.05	93.00	97.35	99.00
RS	$\sqrt{\log\log n}$	1.95	5.45	11.50	22.00	36.00	52.55	68.00	79.15	88.10	94.25	97.85	99.05
RS	$(1/3)\kappa_n$	1.80	4.75	10.65	20.45	34.10	50.15	66.30	77.70	87.00	93.65	97.55	99.05
RS	$(2/3)\kappa_n$	2.00	5.65	12.05	22.65	36.90	53.30	68.80	79.95	88.60	94.50	97.90	99.05

Table 1: Rejection rate (in %) of Test BP, Test RC, and Test RS for the model in Eq. (7.2). Parameter values are $n = 100$, $\alpha = 10\%$, and $\kappa_n = \sqrt{\log n}$. Results based on 2,000 Monte Carlo replications.

Test	τ_n	η -0.1	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
BP	N/A	0.15	2.50	14.80	44.75	78.75	95.45	99.55	100	100	100	100	100
RC	$\sqrt{\log\log n}$	0.40	4.15	20.75	53.90	85.05	97.20	99.70	100	100	100	100	100
RC	$(1/3)\kappa_n$	0.40	4.15	20.75	53.90	85.05	97.20	99.70	100	100	100	100	100
RC	$(2/3)\kappa_n$	0.40	4.15	20.75	53.90	85.05	97.20	99.70	100	100	100	100	100
RS	$\sqrt{\log\log n}$	0.50	4.45	22.40	55.80	86.50	97.70	99.70	100	100	100	100	100
RS	$(1/3)\kappa_n$	0.40	4.35	21.90	55.20	86.05	97.60	99.65	100	100	100	100	100
RS	$(2/3)\kappa_n$	0.50	4.55	23.20	56.40	86.65	97.80	99.70	100	100	100	100	100

Table 2: Rejection rate (in %) of Test BP, Test RC, and Test RS for the model in Eq. (7.2). Parameter values are $n = 500$, $\alpha = 10\%$, and $\kappa_n = \sqrt{\log n}$. Results based on 2,000 Monte Carlo replications.

rejection rate less than α). As expected, under the alternative hypothesis (i.e. $\eta > 0$) the rejection rates increase, with an amount of rejection increasing with the amount of misspecification, measured by η.

In accordance to Theorem 6.1, the tests proposed in this paper should reject more or equal than Test BP. As a corollary, Tests RS and RC are (a) less or equally conservative and (b) more or equally powerful than Test BP. Our Tables reveal that both of these findings occur strictly in finite samples. For concreteness, consider the results for a small sample size $n = 100$ in Table 1 and concentrate on implementing Tests RC and RS using the recommended value of $\tau_n = \sqrt{\log\log n}$. For a null hypothesis with $\eta = 0$, Test BP is significantly more conservative than Tests RC and RS (rejection rates of 2.2% versus 4.05% and 5.45%, respectively). For an alternative hypothesis with $\eta = 0.3$, Test BP is significantly less powerful than Tests RC and RS (rejection rates of 22.55% versus 31.85% and 36%, respectively). The results for a moderate sample size $n = 500$ in Table 2 are qualitatively similar. Our results indicate that the rejection rates of Tests RC and RS are relatively insensitive to reasonable choices of τ_n (especially for $n = 500$).

Section 6 shows that a simplified version of our model used in these Monte Carlo simulations satisfies the assumptions of Theorem 6.2. By using similar arguments, it is possible to show that these findings extends to the current setup. As a consequence, the strict power advantages that of the tests proposed in this paper vis-à-vis Test BP can be understood as the finite sample consequence of our asymptotic results.

One interesting feature of our simulation results is that the rejection rates for Test RC and RS are very similar. This is not a coincidence. In the context of Example 2.1, it is possible to show thats Test RC and RS have the same asymptotic power for all (local) alternative hypotheses considered in our simulations. Using a different example, we show that the Test RS can have a higher asymptotic power than Test RC. In Monte Carlo simulations based on this alternative example, we show that Test RS can have significantly more power in finite samples than test RC.[5] This example illustrates a situation in which the additional computational cost of Test RS relative to Test RC is compensated by better power properties.

8 Conclusions

This paper studies the problem of specification testing in partially identified models with special focus on models defined by a finite number of moment equalities and inequalities (i.e. (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 values. While this problem has not been directly addressed in the literature (except in particular cases), several papers in the literature have suggested implementing this inferential problem by checking whether confidence sets for the parameters of interest are empty or not.

We propose two hypothesis tests that use the infimum of the sample criterion function over the parameter space as test statistic. The difference between these tests lies in their critical values. We show that both of these hypothesis tests are: (a) asymptotically size correct in a uniform sense and (b) more powerful than tests that check whether the confidence intervals for the parameters of interest are empty or not. Our numerical results show that the gains in power can be substantial even for small sample sizes.

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 $x_1, x_2 \in \mathbb{R}^u_{[\pm\infty]}$,

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

where $G : \mathbb{R}_{[\pm\infty]} \to [0,1]$ is such that $G(-\infty) = 0$, $G(\infty) = 1$, and $G(y) = \Phi(y)$ for $y \in \mathbb{R}$, where Φ is the standard normal CDF. 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. (4.5) and (4.6) 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 $A, B \in \Theta \times \mathbb{R}^k_{[\pm\infty]}$,

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

We use "$\overset{H}{\to}$" to denote convergence in the Hausdorff metric, i.e., $A_n \overset{H}{\to} B \iff d_H(A_n, B) \to 0$. Finally, for non-stochastic functions of $\theta \in \Theta$, we use "$\overset{u}{\to}$" to denote uniform in θ convergence, e.g., $\Omega_{F_n} \overset{u}{\to} \Omega \iff \sup_{\theta,\theta' \in \Theta} d(\Omega_{F_n}(\theta,\theta'), \Omega(\theta,\theta')) \to 0$. Also, we use $\Omega(\theta)$ and $\Omega(\theta,\theta)$ equivalently.

[5] These Monte Carlo simulations are available upon request.

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

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

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 sets

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

$$\Lambda^*_{n,F_n} \equiv \left\{ (\theta,\ell) \in \Theta_I^{\eta_n}(F_n) \times \mathbb{R}^k : \ell = \kappa_n^{-1}\sqrt{n} D_{F_n}^{-1/2}(\theta) E_{F_n}[m(W,\theta)] \right\} . \tag{A-4}$$

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 $\{F_n \in \mathcal{P}_0\}_{n\geq 1}$, let Λ^* and Λ' be such that $\Lambda^*_{n,F_n} \overset{H}{\to} \Lambda^*$ and $\Lambda'_{n,F_n} \overset{H}{\to} \Lambda'$, where $\Lambda'_{n,F}$ and Λ^*_{n,F_n} are defined in Eqs. (A-3) and (A-4), 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 $\{F_n \in \mathcal{P}_0\}_{n\geq 1}$, let (Ω,Λ') be such that $\Omega_{F_n} \overset{u}{\to} \Omega$ and $\Lambda'_{n,F_n} \overset{H}{\to} \Lambda'$ 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. (A-3). Let $q_{(\Omega,\Lambda')}(1-\alpha)$ be the $(1-\alpha)$-quantile of $J(\Omega,\Lambda') \equiv \inf_{(\theta,\ell)\in\Lambda'} S(v_\Omega(\theta) + \ell, \Omega(\theta))$. Then,

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

(b) If $q_{(\Omega,\Lambda')}(1-\alpha) = 0$, $\liminf_{n\to\infty} P_{F_n}(\tilde{T}_n = 0) \geq (1-\alpha)$, where $\tilde{T}_n \equiv \inf_{\theta\in\Theta_I^{\eta_n}(F_n)} Q_n(\theta)$.

Assumption A.8. The following conditions hold.

(a) For all $(\theta,F) \in \Theta \times \mathcal{P}_0$, $Q_F(\theta) \geq c\min\{\delta, d_H(\{\theta\},\Theta_I(F))^{\chi}\}$ for constants $c,\delta > 0$.

(b) Θ is convex.

(c) For any $F \in \mathcal{P}_0$, let $G_F(\theta)$ denote the matrix conformed by collecting the gradient of each coordinate of $D_F^{-1/2}(\theta)E_F[m(W,\theta)] : \Theta \to \mathbb{R}^k$. There exists a constant $K > 0$ s.t. G_F satisfies the following Lipschitz type condition

$$\sup_{F\in\mathcal{P}_0} ||G_F(\theta') - G_F(\theta)|| \leq K||\theta' - \theta|| .$$

(d) $\tau_n^2(\log \kappa_n)^2/\sqrt{n} \to 0$, where τ_n and κ_n are as in Definition 4.1 and Assumption M.1, respectively.

Remark B.1. Assumption A.5 is satisfied if the function φ is any of the the functions $\varphi^{(1)}-\varphi^{(4)}$ described in Andrews and Soares (2010) or Andrews and Barwick (2012). This follows from Lemma D.8, as the functions $\varphi^{(1)}-\varphi^{(4)}$ satisfy the conditions of the lemma.

Remark B.2. Assumption A.6 is a mild assumption if we assume that the (studentized) moment conditions are smooth. In particular, Lemma D.9 shows that Assumption A.8 implies Assumption A.6.

Remark B.3. 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.

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 $m \in \mathbb{R}^p_{[+\infty]} \times \mathbb{R}^{k-p}$ 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 $m_1 < m_1^*$.

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. The proof is similar to that of Theorem C.1 in this section and therefore omitted. □

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

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

We divide the argument into steps. Steps 1-3 hold for $\eta \geq 0$, step 4 uses that $\eta > 0$, and step 5 proves the result for $\eta = 0$ under Assumption A.7.

Step 1. Recall that from Definition 4.3 and Assumption A.7, $\eta_n \equiv \tau_n \log \kappa_n$ and $\tilde{T}_n \equiv \inf_{\theta\in\Theta_I^{\eta_n}(F_n)} Q_n(\theta)$. Let $\tilde{c}_n^{RS}(1-\alpha)$ denote the conditional $(1-\alpha)$-quantile of

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

where φ^* is the function defined in Assumption A.5. For any $F \in \mathcal{P}_0$ consider the following derivation

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

where the first inequality follows from $F \in \mathcal{P}_0$ which implies that $\Theta_I^{\eta_n}(F) \supseteq \Theta_I(F) \neq \emptyset$ and so $T_n \leq \tilde{T}_n$, the second inequality is elementary, and the third 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

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

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

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

By Lemma D.6, there is a further sequence $\{b_n\}_{n\geq1}$ of $\{a_n\}_{n\geq1}$ 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. (A-3) and (A-4), 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 C.1 implies that $\tilde{T}_{b_n} \xrightarrow{d} J(\Lambda',\Omega) \equiv \inf_{(\theta,\ell)\in\Lambda'} S(v_\Omega(\theta)+\ell, \Omega(\theta))$.

Step 3. We now show that for $q_{(\Lambda',\Omega)}(1-\alpha)$ being the $(1-\alpha)$-quantile of $J(\Lambda',\Omega)$ and any $\varepsilon > 0$,

$$\lim P_{F_{b_n}}(\tilde{c}_{b_n}^{RS}(1-\alpha) \leq q_{(\Lambda',\Omega)}(1-\alpha)-\varepsilon) = 0 \ . \tag{C-2}$$

We begin by showing that $J^*(\Lambda^*,\Omega) \geq J(\Lambda',\Omega)$, where $J^*(\Lambda^*,\Omega)$ is defined in the statement of Theorem C.2. Suppose not, that is, suppose that $J^*(\Lambda^*,\Omega) < J(\Lambda',\Omega)$. It follows that $\exists(\theta,\ell) \in \Lambda^*$ 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 $j \leq p$. It then follows that

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

which is a contradiction to $(\theta,\ell') \in \Lambda'$. Now let $\varepsilon > 0$ be chosen so that $q_{(\Lambda',\Omega)}(1-\alpha)-\varepsilon$ is a continuity point of the CDF of $J^*(\Lambda^*,\Omega)$. By Theorem C.2, $\{\tilde{T}^*_{b_n}|\{W_i\}_{i=1}^{b_n}\} \xrightarrow{d} J^*(\Lambda^*,\Omega)$ for almost all sample sequences. As a consequence, for almost all sample sequences,

$$\begin{aligned}\lim P_{F_{b_n}}\left(\tilde{T}^*_{b_n} \leq q_{(\Lambda',\Omega)}(1-\alpha)-\varepsilon \,\middle|\, \{W_i\}_{i=1}^{b_n}\right) &= P(J^*(\Lambda^*,\Omega) \leq q_{(\Lambda',\Omega)}(1-\alpha)-\varepsilon) \\ &\leq P(J(\Lambda',\Omega) \leq q_{(\Lambda',\Omega)}(1-\alpha)-\varepsilon) \\ &< 1-\alpha \ ,\end{aligned}$$

where the first equality holds because $\{\tilde{T}^*_{b_n}|\{W_i\}_{i=1}^{b_n}\} \xrightarrow{d} J^*(\Lambda^*,\Omega)$ (for almost all sample sequences) and that $q_{(\Omega,\Lambda)}(1-\alpha)-\varepsilon$ is a continuity point of the CDF of $J^*(\Lambda^*,\Omega)$, the second weak inequality is a consequence of $J^*(\Lambda^*,\Omega) \geq J(\Lambda',\Omega)$, and the final strict inequality follows from the fact that $q_{(\Omega,\Lambda)}(1-\alpha)$ is the $(1-\alpha)$ quantile of $J(\Lambda',\Omega)$. Next, notice that

$$\left\{\lim P_{F_{b_n}}\left(\tilde{T}^*_{b_n} \leq q_{(\Lambda',\Omega)}(1-\alpha)-\varepsilon \,\middle|\, \{W_i\}_{i=1}^{b_n}\right) < 1-\alpha\right\} \subseteq \left\{\liminf\{\tilde{c}^{RS}_{b_n}(1-\alpha) > q_{(\Lambda',\Omega)}(1-\alpha)-\varepsilon\}\right\} \ .$$

Since the RHS occurs for almost all sample sequences, then the LHS must also occur for almost all sample sequences. Then, Eq. (C-2) is a consequence of this and Fatou's Lemma.

Step 4. For $\eta > 0$, we can define $\varepsilon > 0$ in step 3 so that $\eta - \varepsilon > 0$ and $q_{(\Lambda',\Omega)}(1-\alpha)+\eta-\varepsilon$ is a continuity point of the CDF of $J(\Lambda',\Omega)$. It then follows that

$$\begin{aligned}P_{F_{b_n}}\left(\tilde{T}_{b_n} > \tilde{c}^{RS}_{b_n}(1-\alpha)+\eta\right) &= P_{F_{b_n}}\left(\left\{\tilde{T}_{b_n} > \tilde{c}^{RS}_{b_n}(1-\alpha)+\eta\right\} \cap \left\{\tilde{c}^{RS}_{b_n}(1-\alpha) \leq q_{(\Lambda',\Omega)}(1-\alpha)-\varepsilon\right\}\right) \\ &\quad + P_{F_{b_n}}\left(\left\{\tilde{T}_{b_n} > \tilde{c}^{RS}_{b_n}(1-\alpha)+\eta\right\} \cap \left\{\tilde{c}^{RS}_{b_n}(1-\alpha) > q_{(\Lambda',\Omega)}(1-\alpha)-\varepsilon\right\}\right) \\ &\leq P_{F_{b_n}}\left(\tilde{c}^{RS}_{b_n}(1-\alpha) \leq q_{(\Lambda',\Omega)}(1-\alpha)-\varepsilon\right) + 1 - P_{F_{b_n}}\left(\tilde{T}_{b_n} \leq q_{(\Lambda',\Omega)}(1-\alpha)+\eta-\varepsilon\right) \ .\end{aligned}$$

Taking limit supremum on both sides, using steps 2 and 3, and that $\eta - \varepsilon > 0$,

$$\limsup_{n\to\infty} P_{F_{b_n}}\left(\tilde{T}_{b_n} > \tilde{c}^{RS}_{b_n}(1-\alpha)+\eta\right) \leq 1 - P\left(J(\Lambda',\Omega) \leq q_{(\Lambda',\Omega)}(1-\alpha)+\eta-\varepsilon\right) \leq \alpha \ .$$

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

Step 5. For $\eta = 0$, there are two cases to consider. First, $q_{(\Lambda',\Omega)}(1-\alpha) = 0$. In this case, note that

$$\begin{aligned}P_{F_{b_n}}(\tilde{T}_{b_n} > \tilde{c}^{RS}_{b_n}(1-\alpha)) &= P_{F_{b_n}}(\tilde{T}_{b_n} > \tilde{c}^{RS}_{b_n}(1-\alpha) \cap \tilde{T}_{b_n} = 0) + P_{F_{b_n}}(\tilde{T}_{b_n} > \tilde{c}^{RS}_{b_n}(1-\alpha) \cap \tilde{T}_{b_n} \neq 0) \\ &\leq P_{F_{b_n}}(\tilde{T}_{b_n} \neq 0) \ .\end{aligned}$$

By computing limit supremum on both sides and Assumption A.7, we deduce that

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

The proof is completed by combining the previous equation with steps 1 and 2. Second, $q_{(\Omega,\Lambda)}(1-\alpha) > 0$. Consider a sequence $\{\varepsilon_m\}_{m\geq 1}$ s.t. $\varepsilon_m \downarrow 0$ and $q_{(\Omega,\Lambda)}(1-\alpha)-\varepsilon_m$ is a continuity point of the CDF of $J(\Lambda',\Omega)$ for all $m \in \mathbb{N}$. For any $m \in \mathbb{N}$, consider the following argument

$$\begin{aligned}P_{F_{b_n}}(\tilde{T}_{b_n} > \tilde{c}^{RS}_{b_n}(1-\alpha)) &= P_{F_{b_n}}(\{\tilde{T}_{b_n} > \tilde{c}^{RS}_{b_n}(1-\alpha)\} \cap \{\tilde{c}^{RS}_{b_n}(1-\alpha) \leq q_{(\Lambda',\Omega)}(1-\alpha)-\varepsilon_m\}) \\ &\quad + P_{F_{b_n}}(\{\tilde{T}_{b_n} > \tilde{c}^{RS}_{b_n}(1-\alpha)\} \cap \{\tilde{c}^{RS}_{b_n}(1-\alpha) > q_{(\Lambda',\Omega)}(1-\alpha)-\varepsilon_m\}) \\ &\leq P_{F_{b_n}}(\tilde{c}^{RS}_{b_n}(1-\alpha) \leq q_{(\Lambda',\Omega)}(1-\alpha)-\varepsilon_m) + 1 - P_{F_{b_n}}(\tilde{T}_{b_n} \leq q_{(\Lambda',\Omega)}(1-\alpha)-\varepsilon_m) \ .\end{aligned}$$

Taking limit supremum on both sides, using steps 3 and 4,

$$\limsup_{n\to\infty} P_{F_{b_n}}(\tilde{T}_{b_n} > \tilde{c}^{RS}_{b_n}(1-\alpha)) \leq 1 - P(J(\Lambda',\Omega) \leq q_{(\Lambda',\Omega)}(1-\alpha)-\varepsilon_m) \ .$$

Now take $\varepsilon_m \downarrow 0$ and use continuity to deduce that the RHS is equal to α. The proof is completed by combining the previous equation with steps 1 and 2. □

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

$$S(\hat{v}_n^*(\theta) + \varphi(\kappa_n^{-1}\sqrt{n}\hat{D}_n^{1/2}(\theta)\bar{m}_n(\theta), \hat{\Omega}_n(\theta)), \hat{\Omega}_n(\theta)) \ . \tag{C-3}$$

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-3), 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

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

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

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

where the first equality holds by $\hat{\Theta}_I \neq \emptyset$ and $\inf_{\theta\in\Theta} Q_n(\theta) = \inf_{\theta\in\hat{\Theta}_I} Q_n(\theta)$ (see Lemma D.11) and the definition of $\hat{c}_n^{RC}(1-\alpha)$, 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 θ_n^* and θ_n denote the sequences in Assumption A.9. Assume n is large enough so that $\hat{c}_n(\theta_n^*, 1-\alpha) > \hat{c}_n(\theta_n, 1-\alpha)$, which holds by the same assumption, and recall that $T_n = Q_n(\theta_n^*)$. Consider the following derivation:

$$\begin{aligned} E_{F_n}[\phi_n^{BP}] &= P_{F_n}(Q_n(\theta) > \hat{c}_n(\theta, 1-\alpha),\ \forall \theta \in \Theta) \\ &\leq P_{F_n}\left(Q_n(\theta_n^*) > \hat{c}_n(\theta_n^*, 1-\alpha)\right) \\ &= P_{F_n}\left(Q_n(\theta_n^*) > \hat{c}_n(\theta_n, 1-\alpha)\right) - P_{F_n}(\hat{c}_n(\theta_n, 1-\alpha) < Q_n(\theta_n^*) \leq \hat{c}_n(\theta_n^*, 1-\alpha)) \\ &\leq P_{F_n}(T_n > \inf\nolimits_{\theta\in\hat{\Theta}_I} \hat{c}_n(\theta, 1-\alpha)) - P_{F_n}(\hat{c}_n(\theta_n, 1-\alpha) < T_n \leq \hat{c}_n(\theta_n^*, 1-\alpha)) \ , \end{aligned}$$

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

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

It suffices to show that the expression on the RHS is positive. To do this, fix $\varepsilon \in (0, (c_2 - c_1)/3)$ and consider the following argument

$$\begin{aligned} & P_{F_n}(\hat{c}_n(\theta_n, 1-\alpha) < T_n \leq \hat{c}_n(\theta_n^*, 1-\alpha)) \\ & \geq P_{F_n}(\hat{c}_n(\theta_n, 1-\alpha) < c_1 + \varepsilon < T_n < c_2 - \varepsilon \leq \hat{c}_n(\theta_n^*, 1-\alpha)) \\ & \geq P_{F_n}(c_1 + \varepsilon < T_n < c_2 - \varepsilon) + P_{F_n}(\hat{c}_n(\theta_n, 1-\alpha) < c_1 + \varepsilon) + P_{F_n}(c_2 - \varepsilon \leq \hat{c}_n(\theta_n^*, 1-\alpha)) - 2 \ , \end{aligned}$$

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

$$\liminf_{n\to\infty} \; P_{F_n}(\hat{c}_n(\theta_n, 1-\alpha) < T_n \le \hat{c}_n(\theta_n^*, 1-\alpha)) \ge \lim_{\varepsilon\downarrow 0} P(J \in (c_1+\varepsilon, c_2-\varepsilon)) = P(J \in (c_1, c_2)) > 0 \ ,$$

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

C.1 Auxiliary Theorems

Theorem C.1. *Assume Assumptions A.1-A.4. Let $\{F_n \in \mathcal{P}_0\}_{n\ge1}$ 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) $\Omega_{F_n} \overset{u}{\to} \Omega$ and (ii) $\Lambda'_{n,F_n} \overset{H}{\to} \Lambda'$, where Λ'_{n,F_n} is as in Eq. (A-3), for $\{\kappa_n\}_{n\ge1}$ and $\{\tau_n\}_{n\ge1}$ as in Assumption M.1 and Definition 4.1, respectively. Then, along the sequence $\{F_n\}_{n\ge1}$,*

$$\tilde{T}_n \equiv \inf_{\theta\in\Theta_I^{\eta_n}(F_n)} Q_n(\theta) \overset{d}{\to} J\left(\Lambda', \Omega\right) \equiv \inf_{(\theta,\ell)\in\Lambda'} S(v_\Omega(\theta) + \ell, \Omega(\theta,\theta)) \ ,$$

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

Proof. 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

$$\begin{aligned}
\tilde{T}_n &= \inf_{\theta\in\Theta_I^{\eta_n}(F_n)} S(\sqrt{n}\bar{m}_n(\theta), \Sigma_n(\theta)) \\
&= \inf_{\theta\in\Theta_I^{\eta_n}(F_n)} S(\sqrt{n}D_{F_n}^{-1/2}(\theta)\bar{m}_n(\theta), \tilde{\Omega}_n(\theta)) \\
&= \inf_{\theta\in\Theta_I^{\eta_n}(F_n)} S(v_n(\theta) + \sqrt{n}D_{F_n}^{-1/2}(\theta)E_{F_n}[m(W,\theta)], \tilde{\Omega}_n(\theta)) \\
&= \inf_{(\theta,\ell)\in\Lambda'_{n,F_n}} S(v_n(\theta) + \ell, \tilde{\Omega}_n(\theta)) \ .
\end{aligned}$$

Step 2. Let $\mathcal{D}$ denote the 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 $\{g_n\}_{n\ge1}$ with $g_n : \mathcal{D} \to \mathbb{R}$ be defined by

$$g_n(v(\cdot), \Omega(\cdot)) \equiv \inf_{(\theta,l)\in\Lambda'_{n,F_n}} S(v(\theta) + \ell, \Omega(\theta)) \ . \tag{C-4}$$

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 $\{(v_n(\cdot), \Omega_n(\cdot))\}_{n\ge1}$ with $(v_n(\cdot), \Omega_n(\cdot)) \in \mathcal{D}$ for all $n \in \mathbb{N}$ satisfies

$$\lim_{n\to\infty} \sup_{\theta\in\Theta} ||(v_n(\theta), \Omega_n(\theta)) - (v(\theta), \Omega(\theta))|| = 0 \ , \tag{C-5}$$

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 $\liminf g_n(v_n(\cdot), \Omega_n(\cdot)) \ge g(v(\cdot), \Omega(\cdot))$. The argument to show that $\limsup g_n(v_n(\cdot), \Omega_n(\cdot)) \le g(v(\cdot), \Omega(\cdot))$ is very similar and is therefore omitted. Suppose not, i.e., suppose that $\exists\delta > 0$ and a subsequence $\{a_n\}_{n\ge1}$ of $\{n\}_{n\ge1}$ s.t. $\forall n \in \mathbb{N}$,

$$g_{a_n}(v_{a_n}(\cdot), \Omega_{a_n}(\cdot)) < g(v(\cdot), \Omega(\cdot)) - \delta \ . \tag{C-6}$$

By definition, there exists a sequence $\{(\theta_{a_n}, \ell_{a_n})\}_{n\ge1}$ that approximately achieves the infimum in Eq. (C-4), i.e., $\forall n \in \mathbb{N}$, $(\theta_{a_n}, \ell_{a_n}) \in \Lambda_{a_n,F_{a_n}}$ and

$$||g_{a_n}(v_{a_n}(\cdot), \Omega_{a_n}(\cdot)) - S(v_{a_n}(\theta_{a_n}) + \ell_{a_n}, \Omega_{a_n}(\theta_{a_n}))|| \le \delta/2 \ . \tag{C-7}$$

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 $\{b_n\}_{n\geq 1}$ of $\{a_n\}_{n\geq 1}$ 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

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

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 $\Lambda'_{b_n,F_{b_n}} \overset{H}{\to} \Lambda'$, 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, $\lim \Xi(\ell_{b_n}) = \Xi(\ell^*) = \infty$.

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,

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

where the last convergence holds by Eq. (C-5), $\theta_{b_n} \to \theta^*$, 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 $\lim \Xi(\ell_{b_n}) = \infty$ and $v(\theta^*) \in \mathbb{R}^k$, it then follows that $\lim \Xi(\ell_{b_n})^{-1}||v_{b_n}(\theta_{b_n})|| = 0$. By construction, $\{\Xi(\ell_{b_n})^{-1}\ell_{b_n}\}_{n\geq 1}$ is s.t. $\lim \Xi(\ell_{b_n})^{-1}\, [\ell_{b_n,j}]_- = 1$ for some $j \leq p$ or $\lim \Xi(\ell_{b_n})^{-1}\, |\ell_{b_n,j}| = 1$ for some $j > p$. We conclude that $\lim \Xi(\ell_{b_n})^{-1}[v_{b_n,j}(\theta_{b_n}) + \ell_{b_n,j}]_- = 1$ for some $j \leq p$ or $\lim \Xi(h_{b_n})^{-1}\, |v_{b_n,j}(\theta_{b_n}) + \ell_{b_n,j}| = 1$ 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-4), it then follows that

$$g_n(v_n(\cdot), \Omega_n(\cdot)) \to \infty \ . \tag{C-8}$$

We now show that Eq. (C-8) 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.

$$\begin{aligned} \liminf_{n\to\infty} \sqrt{n}\sigma^{-1}_{F_n,j}(\theta_n) E_{F_n}[m_j(W,\theta_n)] &\equiv \ell^*_j \geq 0, \text{ for } j \leq p \\ \lim_{n\to\infty} \sqrt{n}\sigma^{-1}_{F_n,j}(\theta_n) |E_{F_n}[m_j(W,\theta_n)]| &\equiv \ell^*_j = 0, \text{ for } j > p \ . \end{aligned}$$

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 $\lim(v_{k_n}(\theta_{k_n}), \Omega_{k_n}(\tilde{\theta}_{b_n})) = (v(\tilde{\theta}^*), \Omega(\tilde{\theta}^*)) \in \mathbb{R}^k \times \Psi$, which implies that

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

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-8).

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

$$||S(v_{b_n}(\theta_{b_n}) + \ell_{b_n}, \Omega_{b_n}(\theta_{b_n})) - S(v(\theta^*) + \ell^*, \Omega(\theta^*))|| \leq \delta/2 \ . \tag{C-9}$$

By combining Eqs. (C-7), (C-9), 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-6).

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 $(v_n(\cdot), \tilde{\Omega}(\cdot)) \overset{d}{\to} (v_\Omega(\cdot), \Omega(\cdot))$ and $(v_\Omega(\cdot), \Omega(\cdot)) \in \mathcal{D}_0$ a.s. □

Theorem C.2. *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) $\Omega_{F_n} \overset{u}{\to} \Omega$ and (ii) $\Lambda^*_{n,F_n} \overset{H}{\to} \Lambda^*$, where Λ^*_{n,F_n} is as in Eq.* (A-4)*, for $\{\kappa_n\}_{n\geq1}$ and $\{\tau_n\}_{n\geq1}$ as in Assumption M.1 and Definition 4.1, respectively. Then, there is a subsequence $\{a_n\}_{n\geq1}$ of $\{n\}_{n\geq1}$ s.t., along the sequence $\{F_{a_n}\}_{n\geq1}$,*

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

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:

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

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

Step 2. We now show that there is a subsequence $\{a_n\}_{n\geq1}$ of $\{n\}_{n\geq1}$ s.t. $\{\{(\hat{v}^*_{a_n}, \mu_{a_n}, \hat{\Omega}_{a_n})|\{W_i\}_{i=1}^{a_n}\} \overset{d}{\to} (v_\Omega, (\mathbf{0}_k, \mathbf{1}_k), \Omega)$ in $l^\infty(\theta)$, for almost all sample sequences $\{W_i\}_{i=1}^\infty$. By part 8 in Lemma D.2, $\{\hat{v}^*_n|\{W_i\}_{i=1}^n\} \overset{d}{\to} v_\Omega$ in $l^\infty(\theta)$. Then the result would follow from finding a subsequence $\{a_n\}_{n\geq1}$ of $\{n\}_{n\geq1}$ s.t. $\{\{(\mu_{a_n}, \hat{\Omega}_{a_n})|\{W_i\}_{i=1}^{a_n}\} \to ((\mathbf{0}_k, \mathbf{1}_k), \Omega)$ in $l^\infty(\theta)$, for almost all sample sequences $\{W_i\}_{i=1}^\infty$. Since $(\mu_n, \hat{\Omega}_n)$ is conditionally non-random, this is equivalent to finding a subsequence $\{a_n\}_{n\geq1}$ of $\{n\}_{n\geq1}$ s.t. $(\mu_{a_n}, \hat{\Omega}_{a_n}) \overset{a.s.}{\to} ((\mathbf{0}_k, \mathbf{1}_k), \Omega)$ in $l^\infty(\theta)$. In turn, this follows from step 1, part 5 of Lemma D.2, and Lemma D.7.

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

$$g_n(v(\cdot), \mu(\cdot), \Omega(\cdot)) \equiv \inf_{(\theta,\ell)\in\Lambda^*_{n,F_n}} S(v(\theta) + \varphi^*(\mu_1(\theta) + \mu_2(\theta)'\ell), \Omega(\theta)) \ . \tag{C-10}$$

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

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

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

$$\lim_{n\to\infty} \sup_{\theta\in\Theta} ||(v_n(\theta), \mu_n(\theta), \Omega_n(\theta)) - (v(\theta), (\mathbf{0}_k, \mathbf{1}_k), \Omega(\theta))|| = 0 \ , \tag{C-11}$$

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

$$g_{a_n}(v_{a_n}(\cdot), \mu_{a_n}(\cdot), \Omega_{a_n}(\cdot)) < g(v(\cdot), (\mathbf{0}_k, \mathbf{1}_k), \Omega(\cdot)) - \delta \ . \tag{C-12}$$

By definition, there exists a sequence $\{(\theta_{a_n}, \ell_{a_n})\}_{n\geq 1}$ that approximately achieves the infimum in Eq. (C-10), i.e., $\forall n \in \mathbb{N}$, $(\theta_{a_n}, \ell_{a_n}) \in \Lambda_{a_n,F_{a_n}}$ and

$$|g_{a_n}(v_{a_n}(\cdot), \mu_{a_n}(\cdot), \Omega_{a_n}(\cdot)) - S(v_{a_n}(\theta_{a_n}) + \varphi^*(\mu_1(\theta_{a_n}) + \mu_2(\theta_{a_n})'\ell_{a_n}), \Omega_{a_n}(\theta_{a_n}))| \leq \delta/2 \ . \tag{C-13}$$

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 $\{b_n\}_{n\geq 1}$ of $\{a_n\}_{n\geq 1}$ 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

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

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 $\Lambda^*_{b_n,F_{b_n}} \overset{H}{\to} \Lambda^*$, 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:

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

where the last convergence holds by Eq. (C-11), $\theta_{b_n} \to \theta^*$, and $(v(\cdot), \Omega(\cdot)) \in \mathcal{D}_0$.

By continuity of φ^* and Eq. (C-11), 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 $j \leq p$, 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-11) 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, $\lim(v_{b_n}(\theta_{b_n})+\varphi^*(\mu_1(\theta_{b_n})+\mu_2(\theta_{b_n})\ell_{b_n}),\Omega_{b_n}(\theta_{b_n}))=(v(\theta^*)+\varphi^*(\ell^*),\Omega(\theta^*))\in(\mathbb{R}^k_{[\pm\infty]}\times\Psi)$ and so $\lim S(v_{b_n}(\theta_{b_n})+\varphi^*(\mu_1(\theta_{b_n})+\mu_2(\theta_{b_n})\ell_{b_n}),\Omega_{b_n}(\theta_{b_n}))=S(v(\theta^*)+\varphi^*(\ell^*),\Omega(\theta^*))$, i.e, $\exists N\in\mathbb{N}$ s.t. $\forall n\geq N$,

$$||S(v_{b_n}(\theta_{b_n})+\varphi^*(\mu_1(\theta_{b_n})+\mu_2(\theta_{b_n})\ell_{b_n}),\Omega_{b_n}(\theta_{b_n}))-S(v(\theta^*)+\varphi^*(\ell^*),\Omega(\theta^*))||\leq\delta/2\ . \tag{C-14}$$

By combining Eqs. (C-13), (C-14), and the fact that $(\theta^*,\ell^*)\in\Lambda^*$, it follows that $\exists N\in\mathbb{N}$ s.t. $\forall n\geq N$,

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

which is a contradiction to Eq. (C-12).

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.3. *Let ϕ_n^{BP} be the test defined in Definition 2.4. Then,* $\limsup_{n\to\infty}\sup_{F\in\mathcal{P}_0}E_F[\phi_n^{BP}]\leq\alpha$.

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 $\theta\in\Theta$. This implies that

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

This and Eq. (2.9) imply that

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

and the result follows. □

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:

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

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. $\cup_{s=1}^S\{\theta\in\Theta : ||\theta_s-\theta||\leq\delta'\}=\Theta$. 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 $\rho_F(\theta_s,\theta)\leq\delta$ for some $s=1,\ldots,S$. By the previous construction, $\exists s\in\{1,\ldots,S\}$ s.t. $\{||\theta_s-\theta||\leq\delta'\}\subseteq\{\rho_F(\theta_s,\theta)\leq\delta\}$. 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.

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

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

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

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. $\{||\theta-\theta'||\leq\delta'\}\subseteq\{\rho_F(\theta_s,\theta)\leq\delta\}$. From this, it follows that,

$$P_F\left(\sup_{\theta,\theta'\in\Theta:||\theta-\theta'||\leq\delta'}||v_n(\theta)-v_n(\theta')||>\varepsilon\right)\leq P_F\left(\sup_{\theta,\theta'\in\Theta:\rho(\theta,\theta')\leq\delta}||v_n(\theta)-v_n(\theta')||>\varepsilon\right)\ .$$

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

Lemma D.2. *Assume Assumptions A.1-A.4. Let $\{F_n\in\mathcal{P}\}_{n\geq1}$ be a (sub)sequence of distributions s.t. $\Omega_{F_n}\overset{u}{\to}\Omega$ for some $\Omega\in\mathcal{C}(\Theta^2)$. Then, the following results hold:*

1. *$v_n\overset{d}{\to}v_\Omega$ in $l^\infty(\Theta)$, where $v_\Omega:\Theta\to\mathbb{R}^k$ is a tight zero-mean Gaussian process with covariance (correlation) kernel Ω. In addition, v_Ω is a uniformly continuous function, a.s.*
2. *$\tilde{\Omega}_n\overset{p}{\to}\Omega$ in $l^\infty(\Theta)$.*
3. *$D_{F_n}^{-1/2}(\cdot)\hat{D}_n^{1/2}(\cdot)-I_k\overset{p}{\to}\boldsymbol{0}_k$ in $l^\infty(\Theta)$.*
4. *$\hat{D}_n^{-1/2}(\cdot)D_{F_n}^{1/2}(\cdot)-I_k\overset{p}{\to}\boldsymbol{0}_k$ in $l^\infty(\Theta)$.*
5. *$\hat{\Omega}_n\overset{p}{\to}\Omega$ in $l^\infty(\Theta)$.*
6. *For any arbitrary sequence $\{\lambda_n\in\mathbb{R}_{++}\}_{n\geq1}$ s.t. $\lambda_n\to\infty$, $\lambda_n^{-1}v_n\overset{p}{\to}\boldsymbol{0}_k$ in $l^\infty(\Theta)$.*
7. *For any arbitrary sequence $\{\lambda_n\in\mathbb{R}_{++}\}_{n\geq1}$ s.t. $\lambda_n\to\infty$, $\lambda_n^{-1}\tilde{v}_n\overset{p}{\to}\boldsymbol{0}_k$ in $l^\infty(\Theta)$.*
8. *$\{v_n^*|\{W_i\}_{i=1}^n\}\overset{d}{\to}v_\Omega$ in $l^\infty(\Theta)$, for almost all sample sequences $\{W_i\}_{i=1}^\infty$, 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 $\{F_n\in\mathcal{P}_0\}_{n\geq1}$, (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 $\Omega_{F_n}\overset{u}{\to}\Omega$, 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 $\Omega_{F_n} \overset{u}{\to} \Omega$ 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 Gliveko-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, $\forall \theta, \theta_s \in \Theta$,

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

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

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 al $j = 1,\ldots,k$. 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) $\tilde{D}_n^{-1/2}(\theta)D_{F_n}^{1/2}(\theta) - I_k \overset{p}{\to} \mathbf{0}_k$ in $l^\infty(\Theta)$. Under the previous two results, we conclude that $\hat{D}_n^{-1/2}(\theta)D_{F_n}^{1/2}(\theta) - I_k \overset{p}{\to} \mathbf{0}_k$ 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

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

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 $\tilde{D}_n^{-1/2}(\theta)D_{F_n}^{1/2}(\theta) - I_k \overset{p}{\to} \mathbf{0}_k$ 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

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

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 $\hat{\Omega}_n(\theta) - \tilde{\Omega}_n(\theta) \overset{p}{\to} \mathbf{0}_{k\times k}$ in $l^\infty(\Theta)$. To show this, consider the following argument:

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

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

$$||\hat{\Omega}_n(\theta) - \tilde{\Omega}_n(\theta)|| \leq 2||D_{F_n}^{1/2}(\theta)\hat{D}_n^{-1/2}(\theta) - I_k|| + ||D_{F_n}^{1/2}(\theta)\hat{D}_n^{-1/2}(\theta) - I_k||^2 .$$

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

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

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. $\cup_{s=1}^S\{\theta \in \Theta : ||\theta_s - \theta|| \leq \delta\} = \Theta$. Based on this, consider the following derivation:

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

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

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

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

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

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 $\Omega_{F_n} \stackrel{u}{\to} \Omega$ implies $\Omega_{F_n}(\theta, \theta) \to \Omega(\theta, \theta)$ which, in turn, implies that $v_n(\theta) \stackrel{d}{\to} N(\mathbf{0}_k, \Omega(\theta, \theta))$. 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:

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

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 $\{F_n \in \mathcal{P}\}_{n\geq1}$. The original result proves that three statements are equal: (i), (ii), and (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 $F \in \mathcal{P}$. $\square$

D.2 Auxiliary results on S

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

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

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 $j \leq p$ 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, $\exists\{(x_n,\Omega_n) \in A \times \Psi\}_{n\geq 1}$ (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 $\{a_n\}_{n\geq 1}$ of $\{n\}_{n\geq 1}$, $\{(x_{a_n},\Omega_{a_n})\}_{n\geq 1}$ 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 exist a constant $\varpi > 0$ such that $S(x,\Omega) \leq 1$ for any $\Omega \in \Psi$ implies $x_j \geq -\varpi$ for all $j \leq p$ and $|x_s| \leq \varpi$ for all $s > p$.*

Proof. Let $(x,\Omega) \in \mathbb{R}^{p}_{[+\infty]} \times \mathbb{R}^{k-p} \times \Psi$ be arbitrary s.t. $S(x,\Omega) \leq 1$. Set $\tilde{x} \equiv (\{[x_j]_-\}_{j=1}^{p}, \{x_s\}_{s=p+1}^{k})$ and note that $x_j \geq -\varpi$ for all $j \leq p$ and $|x_s| \leq \varpi$ for all $s > p$ is equivalent to $\max_{j=1,\ldots,k} |\tilde{x}_j| \leq \varpi$. Since $S((x_1,x_2),\Sigma)$ is non-increasing in $x_1 \in \mathbb{R}^{p}_{[+\infty]}$ and $\{x_j\}_{j=1}^{p} \geq \{[x_j]_-\}_{j=1}^{p}$, it follows that $S(x,\Omega) \leq S(\tilde{x},\Omega)$. Thus, it suffices to find $\varpi > 0$ s.t. $S(\tilde{x},\Omega) \leq 1$ implies that $\max_{j=1,\ldots,k} |\tilde{x}_j| \leq \varpi$. 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 the following,

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

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

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

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

Lemma D.5. *Let $\{(x_n,\Omega_n) \in \mathbb{R}^{k}_{[\pm\infty]} \times \Psi\}_{n\geq 1}$ be a sequence s.t.* $\liminf_{n\to\infty} x_{n,j} \geq 0$ *for $j \leq p$ and* $\lim_{n\to\infty} x_{n,j} = 0$ *for $j > p$. Then,* $\lim_{n\to\infty} S(x_n,\Omega_n) = 0$.

Proof. Suppose not, i.e., suppose that $\liminf_n |S(x_n,\Omega_n)| > 0$. Since $(\mathbb{R}^{k}_{[\pm\infty]} \times \Psi, d)$ is compact, there is a subsequence $\{a_n\}_{n\geq 1}$ of $\{n\}_{n\geq 1}$ s.t. $d((x_{a_n},\Omega_{a_n}),(x,\Omega)) \to 0$ 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, $\lim_n |S(x_{a_n},\Omega_{a_n})| = |S(x,\Omega)| = 0$, which is a contradiction. □

D.3 Auxiliary results on subsequences

Lemma D.6. *Let Assumption A.4 hold. Let $\{F_n \in \mathcal{P}\}_{n\geq 1}$ be an arbitrary sequence of distributions, and let $\{\eta_n \in \mathbb{R}_{++}\}_{n\geq 1}$ and $\{\kappa_n \in \mathbb{R}_{++}\}_{n\geq 1}$ be non-stochastic sequences. Then, there exists a subsequence $\{u_n\}_{n\geq 1}$ of $\{n\}_{n\geq 1}$ s.t. $\Lambda'_{u_n,F_{u_n}} \overset{H}{\to} \Lambda'$, $\Lambda^*_{u_n,F_{u_n}} \overset{H}{\to} \Lambda^*$, and $\Omega_{F_{u_n}} \overset{u}{\to} \Omega$ 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 Eqs.* (A-3) *and* (A-4)*, 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 $\{\Omega_{F_n}(\theta,\theta')\}_{n\geq 1}$ 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 $\{a_n\}_{n\geq 1}$ of $\{n\}_{n\geq 1}$ and $\Omega \in \mathcal{C}(\Theta^2)$ s.t. $\Omega_{F_{a_n}} \overset{u}{\to} \Omega$.

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 $\{b_n\}_{n\geq 1}$ of $\{a_n\}_{n\geq 1}$ 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, one would define a subsequence $\{c_n\}_{n\geq 1}$ of $\{b_n\}_{n\geq 1}$ s.t. $\Lambda^*_{c_n,F_{c_n}} \overset{H}{\to} \Lambda^*$ using an identical argument to the one used before. The proof is then concluded by setting $\{u_n\}_{n\geq 1} \equiv \{c_n\}_{n\geq 1}$. □

Lemma D.7. *Let $\{F_n \in \mathcal{P}\}_{n\geq 1}$ be an arbitrary (sub)sequence of distributions and let $X_n(\theta) : \Omega \to l^\infty(\Theta)$ be any stochastic process s.t. $X_n \overset{p}{\to} 0$ in $l^\infty(\Theta)$. Then, there exists a subsequence $\{u_n\}_{n\geq 1}$ of $\{n\}_{n\geq 1}$ s.t. $X_{u_n} \overset{a.s.}{\to} 0$ in $l^\infty(\Theta)$.*

Proof. Throughout this proof, we consider an arbitrary sequence $\{\varepsilon_n \in \mathbb{R}_{++}\}_{n\geq 1}$ with $\varepsilon_n \downarrow 0$. Then, for arbitrary $\delta > 0$ and arbitrary subsequence $\{u_n\}_{n\geq 1}$ of $\{n\}_{n\geq 1}$, it follows that:

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

Then, in order to complete the proof, it suffices to construct a subsequence $\{u_n\}_{n\geq 1}$ of $\{n\}_{n\geq 1}$ (solely dependent on $\{\varepsilon_n \in \mathbb{R}_{++}\}_{n\geq 1}$) 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:

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

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 $X_n \overset{p}{\to} 0$ 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 right hand size 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 section B to hold.

Lemma D.8. *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 such that, 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, this function 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 $\xi_j^{\star\star}$. It follows that $\varphi_j(\xi_j)$ is a continuous function for all $\xi_j \in [0, \xi_j^{\star\star})$. By continuity at zero, $\exists \xi_j^\star \in (0, \xi_j^{\star\star})$ s.t. for some real number $\delta > 0$, $|\varphi_j(\xi_j)| \leq \delta$ for all $\xi_j \in [0, \xi_j^\star]$. We divide the rest of the proof into two cases.

Case 1. $\exists \delta \in (0,1)$ s.t. $|\varphi_j(\delta\xi_j^\star)| > 0$. In this case, define the following constants: $A \equiv (G(\delta) - \delta)/(1-\delta)$ and $B \equiv \delta/|\varphi_j(\delta\xi_j^\star)|$, 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

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

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

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

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

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

Lemma D.9. *Let Assumption A.8 hold and let* $\eta_n = \tau_n \log \kappa_n$, *where* $\{\kappa_n\}_{n\geq1}$ *and* $\{\tau_n\}_{n\geq1}$ *are as in Assumption M.1 and Definition 4.1, respectively. Then, for any* $\{\theta_n \in \Theta_I^{\eta_n}(F_n)\}_{n\geq1}$ *and* $\gamma \in (0,1)$, *there is a subsequence* $\{u_n\}_{n\geq1}$ *of* $\{n\}_{n\geq1}$, *and a sequence* $\{\hat\theta_{u_n} \in \Theta_I^{\eta_{u_n}}(F_{u_n})\}_{n\geq1}$ *such that*

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

Proof. By definition, $\{\theta_n \in \Theta_I^{\eta_n}(F_n)\}_{n\geq1}$ implies that $S(\eta_n^{-1}\sqrt{n}D_{F_n}^{-1/2}(\theta_n)E_{F_n}[m(W,\theta_n)], \Omega_{F_n}(\theta_n)) \leq 1$ or, equivalently, $Q_{F_n}(\theta_n) \leq (\eta_n/\sqrt{n})^\chi$ and, thus, $\lim_{n\to\infty} Q_{F_n}(\theta_n) = 0$. We now claim that $\eta_n^{-1}\sqrt{n}d_H(\{\theta_n\}, \Theta_I(F_n)) = O(1)$. To see why, consider the following derivation:

$$\min\{\delta, d_H(\theta_n, \Theta_I(F_n))^\chi\} < c^{-1}Q_{F_n}(\theta_n) \leq c^{-1}(\eta_n/\sqrt{n})^\chi \ ,$$

where the first inequality holds by Assumption A.8(a). Because the right hand size converges to zero as $n \to \infty$, it is smaller than δ for sufficiently large n. Thus, the above display implies, for sufficiently large n,

$$d_H(\theta_n, \Theta_I(F_n))^{\chi} \leq c^{-1}(\eta_n/\sqrt{n})^{\chi} \ , \tag{D-6}$$

which implies $\eta_n^{-1}\sqrt{n}d_H(\theta_n, \Theta_I(F_n)) \leq c^{-1/\chi} = O(1)$. Since $\eta_n^{-1}\sqrt{n}d_H(\{\theta_n\}, \Theta_I(F_n)) = O(1)$, there is a sequence $\{\tilde{\theta}_n \in \Theta_I(F_n)\}_{n\geq 1}$ with $(\eta_n^{-1}\sqrt{n})||\tilde{\theta}_n - \theta_n|| = O(1)$.

By the convexity of Θ and the continuous differentiability of $D_{F_n}^{-1/2}(\cdot)E_{F_n}[m(W,\cdot)]$, the intermediate value theorem implies that there is a sequence $\{\theta_n^* \in \Theta\}_{n\geq 1}$ with θ_n^* in the linear combination between θ_n and $\tilde{\theta}_n$ such that

$$\kappa_n^{-\gamma}\sqrt{n}D_{F_n}^{-1/2}(\theta_n)E_{F_n}[m(W,\theta_n)] = \sqrt{n}G_{F_n}(\theta_n^*)\kappa_n^{-\gamma}(\theta_n - \tilde{\theta}_n) + \kappa_n^{-\gamma}\sqrt{n}D_{F_n}^{-1/2}(\tilde{\theta}_n)E_{F_n}[m(W,\tilde{\theta}_n)] \ . \tag{D-7}$$

Define $\hat{\theta}_n \equiv \kappa_n^{-\gamma}\theta_n + (1-\kappa_n^{-\gamma})\tilde{\theta}_n$ or, equivalently,

$$\hat{\theta}_n - \tilde{\theta}_n = \kappa_n^{-\gamma}(\theta_n - \tilde{\theta}_n) \ . \tag{D-8}$$

By convexity of Θ, $\{\hat{\theta}_n \in \Theta\}_{n\geq 1}$. By Eq. (D-8), $\limsup_{n\to\infty}(\eta_n^{-1}\sqrt{n})||\tilde{\theta}_n - \theta_n|| < \infty$, and $\kappa_n^{-\gamma} \to 0$, we conclude that $(\eta_n^{-1}\sqrt{n})||\hat{\theta}_n - \tilde{\theta}_n|| \to 0$. Furthermore, Eqs. (D-7) and (D-8) imply

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

By the convexity of Θ and by the continuous differentiability of $D_{F_n}^{-1/2}(\cdot)E_{F_n}[m(W,\cdot)]$, the intermediate value theorem implies that there is a sequence $\{\theta_n^{**} \in \Theta\}_{n\geq 1}$ with θ_n^{**} in the linear combination between $\hat{\theta}_n$ and $\tilde{\theta}_n$ such that

$$\sqrt{n}D_{F_n}^{-1/2}(\hat{\theta}_n)E_{F_n}[m(W,\hat{\theta}_n)] = \sqrt{n}G_{F_n}(\theta_n^{**})(\hat{\theta}_n - \tilde{\theta}_n) + \sqrt{n}D_{F_n}^{-1/2}(\tilde{\theta}_n)E_{F_n}[m(W,\tilde{\theta}_n)] \ . \tag{D-10}$$

By definition of $\{\theta_n^* \in \Theta\}_{n\geq 1}$ and $(\eta_n^{-1}\sqrt{n})||\tilde{\theta}_n - \theta_n|| = O(1)$, $(\eta_n^{-1}\sqrt{n})||\tilde{\theta}_n - \theta_n^*|| = O(1)$. Similarly, by definition of $\{\theta_n^{**} \in \Theta\}_{n\geq 1}$ and $(\eta_n^{-1}\sqrt{n})||\hat{\theta}_n - \tilde{\theta}_n|| \to 0$, $(\eta_n^{-1}\sqrt{n})||\tilde{\theta}_n - \theta_n^{**}|| \to 0$. Then, triangle inequality implies that $(\eta_n^{-1}\sqrt{n})||\theta_n^* - \theta_n^{**}|| = O(1)$.

Now consider the following derivation:

$$\begin{aligned}\sqrt{n}D_{F_n}^{-1/2}(\hat{\theta}_n)E_{F_n}[m(W,\hat{\theta}_n)] &= \sqrt{n}G_{F_n}(\theta_n^*)(\hat{\theta}_n - \tilde{\theta}_n) + \sqrt{n}D_{F_n}^{-1/2}(\tilde{\theta}_n)E_{F_n}[m(W,\tilde{\theta}_n)] + \varepsilon_{n,1} \\ &= \kappa_n^{-\gamma}\sqrt{n}D_{F_n}^{-1/2}(\theta_n)E_{F_n}[m(W,\theta_n)] + \varepsilon_{n,1} + \varepsilon_{n,2} \ ,\end{aligned} \tag{D-11}$$

where the first equality follows from Eq. (D-10) and defining $\varepsilon_{n,1} \equiv \sqrt{n}(G_{F_n}(\theta_n^{**}) - G_{F_n}(\theta_n^*))(\hat{\theta}_n - \tilde{\theta}_n)$, and the second equality follows from Eq. (D-9) and by defining $\varepsilon_{n,2} \equiv (1-\kappa_n^{-\gamma})\sqrt{n}D_{F_n}^{-1/2}(\tilde{\theta}_n)E_{F_n}[m(W,\tilde{\theta}_n)]$.

On the one hand, $\tilde{\theta}_n \in \Theta_I(F_n)$ and $\kappa_n^{-\gamma} \to 0$ implies that $\varepsilon_{n,2}$ is s.t. $\varepsilon_{n,2,j} \geq 0$ for $j \leq p$ and $\varepsilon_{n,2,j} = 0$ for $j > p$. On the other hand, $||\varepsilon_{n,1}|| \to 0$. To see this, notice that

$$||\varepsilon_{n,1}|| \leq \sqrt{n}\sup_{F\in\mathcal{P}_0}||G_F(\theta_n^*) - G_F(\theta_n^{**})|| \times ||\hat{\theta}_n - \tilde{\theta}_n|| = O(\eta_n^2/\sqrt{n}) \to 0 \ ,$$

where the result combines our assumptions with $(\eta_n^{-1}\sqrt{n})||\theta_n^* - \theta_n^{**}|| = O(1)$ and $(\eta_n^{-1}\sqrt{n})||\hat{\theta}_n - \tilde{\theta}_n|| = O(1)$.

We now verify that $\hat{\theta}_n \in \Theta_I^{\eta_n}(F_n)$ for all sufficiently large n. By Eq. (D-11), $\{\theta_n \in \Theta_I^{\eta_n}(F_n)\}_{n\geq 1}$, Lemma D.4, $\kappa_n^{-\gamma} \to 0$, and the properties shown for $\varepsilon_{n,1}$ and $\varepsilon_{n,2}$, we have that

$$\begin{aligned}\eta_n^{-1}\sqrt{n}\sigma_{F_n,j}^{-1}(\hat{\theta}_n)E_{F_n}[m_j(W,\hat{\theta}_n)] &\geq \eta_n^{-1}\kappa_n^{-\gamma}\sqrt{n}\sigma_{F_n,j}^{-1}(\theta_n)E_{F_n}[m_j(W,\theta_n)] + \varepsilon_{n,1,j} \geq -\kappa_n^{-\gamma}\varpi \to 0, \ j \leq p \ , \\ \eta_n^{-1}\sqrt{n}\sigma_{F_n,j}^{-1}(\hat{\theta}_n)|E_{F_n}[m_j(W,\hat{\theta}_n)]| &\leq \eta_n^{-1}\kappa_n^{-\gamma}\sqrt{n}\sigma_{F_n,j}^{-1}(\theta_n)|E_{F_n}[m_j(W,\theta_n)]| + |\varepsilon_{n,1}| \leq \kappa_n^{-\gamma}\varpi \to 0, \ j > p \ .\end{aligned}$$

By these findings and Lemma D.5, we conclude that $S(\eta_n^{-1}\sqrt{n}D_{F_n}^{-1/2}(\hat{\theta}_n)E_{F_n}[m(W,\hat{\theta}_n)], \Omega_{F_n}(\hat{\theta}_n)) \to 0$ and, thus, $\theta_n \in \Theta_I^{\eta_n}(F_n)$ for all sufficiently large n. As a consequence, there is a subsequence $\{a_n\}_{n\geq 1}$ of $\{n\}_{n\geq 1}$ s.t. $\{\hat{\theta}_{a_n} \in$

$\Theta_I^{\eta_{a_n}}(F_{a_n})\}_{n\geq 1}$.

To conclude the proof, consider the following argument. Since $(\mathbb{R}^k_{[\pm\infty]}, d)$ is compact, there is a further subsequence $\{b_n\}_{n\geq 1}$ of $\{a_n\}_{n\geq 1}$ s.t. $\sqrt{b_n}D_{F_{b_n}}^{-1/2}(\hat{\theta}_{b_n})E_{F_{b_n}}[m(W,\hat{\theta}_{b_n})]$ and $\kappa_{b_n}^{-\gamma}\sqrt{b_n}D_{F_{b_n}}^{-1/2}(\theta_{b_n})E_{F_{b_n}}[m(W,\theta_{b_n})]$ converge. By Eq. (D-11), combined with the properties of $\varepsilon_{n,1}$ and $\varepsilon_{n,2}$, we conclude that

$$\begin{aligned}\lim_{n\to\infty}\eta_{b_n}^{-1}\sqrt{b_n}\sigma_{F_{b_n},j}^{-1}(\hat{\theta}_{b_n})E_{F_{b_n}}[m_j(W,\hat{\theta}_{b_n})] &\geq \lim_{n\to\infty}\eta_{b_n}^{-1}\kappa_{b_n}^{-\gamma}\sqrt{b_n}\sigma_{F_{b_n},j}^{-1}(\theta_{b_n})E_{F_{b_n}}[m_j(W,\theta_{b_n})],\ \text{for } j\leq p\ ,\\ \lim_{n\to\infty}\eta_{b_n}^{-1}\sqrt{b_n}\sigma_{F_{b_n},j}^{-1}(\hat{\theta}_{b_n})E_{F_{b_n}}[m_j(W,\hat{\theta}_{b_n})] &= \lim_{n\to\infty}\eta_{b_n}^{-1}\kappa_{b_n}^{-\gamma}\sqrt{b_n}\sigma_{F_{b_n},j}^{-1}(\theta_{b_n})E_{F_{b_n}}[m_j(W,\theta_{b_n})],\ \text{for } j> p\ .\end{aligned}$$

The proof is completed by considering the subsequence $\{u_n\}_{n\geq 1}$ with $u_n\equiv b_n$. □

Lemma D.10. *Let $\{F_n\in\mathcal{P}_0\}_{n\geq 1}$ be s.t. $\Lambda'_{n,F_n}\overset{u}{\to}\Lambda'\in\mathcal{S}(\Theta\times\mathbb{R}^k_{[\pm\infty]})$. Then, Assumptions A.5 and A.8 imply Assumption A.6.*

Proof. First note that by Lemma D.9, Assumption A.8 implies that Eq. (D-5) holds. 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. $\lim_{n\to\infty}d((\theta_n,\ell_n),(\theta^*,\ell^*))=0$. Note that $\ell_n\equiv\kappa_n^{-1}\sqrt{n}D_{F_n}^{-1/2}(\theta_n)E_{F_n}[m(W,\theta_n)]$. Since $(\theta_n,\ell_n)\in\Lambda^*_{n,F_n}$ then $\theta_n\in\Theta_I^{\eta_n}(F_n)$, i.e.,

$$S(\eta_n^{-1}\sqrt{n}D_{F_n}^{-1/2}(\theta_n)E_{F_n}[m(W,\theta_n)],\Omega_n(\theta_n))\leq 1\ .$$

By Lemma D.4, $\exists\varpi>0$ such that

$$\begin{aligned}\kappa_n\eta_n^{-1}\ell_{n,j} &= \eta_n^{-1}\sqrt{n}\sigma_{F_n,j}^{-1}(\theta_n)E_{F_n}[m_j(W,\theta_n)]\geq-\varpi,\ \text{for } j\leq p\ ,\\ \kappa_n\eta_n^{-1}|\ell_{n,j}| &= \eta_n^{-1}\sqrt{n}\sigma_{F_n,j}^{-1}(\theta_n)|E_{F_n}[m_j(W,\theta_n)]|\leq\varpi,\ \text{for } j>p\ .\end{aligned}\tag{D-12}$$

By Definition 4.1, $\tau_n=\kappa_n^r$ for $r\in(0,1)$ and by Lemma D.9 we can choose $\gamma\in(r,1)\subset(0,1)$. By Eq. (D-12) and $\eta_n\equiv\tau_n\log\kappa_n$ it then follows that

$$\begin{aligned}\kappa_n^{1-\gamma}\ell_{n,j} &= \kappa_n^{-\gamma}\sqrt{n}\sigma_{F_n,j}^{-1}(\theta_n)E_{F_n}[m_j(W,\theta_n)]\geq-\kappa_n^{-\gamma}\eta_n\varpi\to 0,\ \text{for } j\leq p\ ,\\ \kappa_n^{1-\gamma}|\ell_{n,j}| &= \kappa_n^{-\gamma}\sqrt{n}\sigma_{F_n,j}^{-1}(\theta_n)|E_{F_n}[m_j(W,\theta_n)]|\leq\kappa_n^{-\gamma}\eta_n\varpi\to 0,\ \text{for } j>p\ .\end{aligned}$$

By the previous equations, $\kappa_n^{1-\gamma}\to\infty$, and $d(\ell_n,\ell^*)\to 0$, we conclude that $\ell^*\in\mathbb{R}^p_{[+\infty]}\times\{\mathbf{0}_{k-p}\}$.

Also, Eq. (D-12) and Lemma D.9 imply that there is a subsequence $\{a_n\}_{n\geq 1}$ of $\{n\}_{n\geq 1}$ and a sequence $\{\hat{\theta}_n\in\Theta_I^{\eta_n}(F_n)\}_{n\geq 1}$ that satisfies:

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

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, $(\hat{\theta}_{a_n},\hat{\ell}_{a_n})\in\Lambda'_{a_n,F_{a_n}}$. Since $(\Theta\times\mathbb{R}^k_{[\pm\infty]},d)$ is compact, there is a subsequence $\{b_n\}_{n\geq 1}$ of $\{a_n\}_{n\geq 1}$ s.t. $d((\hat{\theta}_{b_n},\hat{\ell}_{b_n}),(\theta,\ell'))\to 0$. Finally, since $\Lambda'_{n,F_n}\to\Lambda'\in\mathcal{S}(\Theta\times\mathbb{R}^k_{[\pm\infty]})$, we conclude that $(\theta,\ell')\in\Lambda'$. We can summarize the previous construction as follows:

$$\begin{aligned}\ell'_j=\lim_{n\to\infty}\hat{\ell}_{b_n,j} &\geq \lim_{n\to\infty}\kappa_{b_n}^{1-\gamma}\ell_{b_n,j}\geq 0,\ \text{for } j\leq p\ ,\\ \ell'_j=\lim_{n\to\infty}\hat{\ell}_{b_n,j} &= \lim_{n\to\infty}\kappa_{b_n}^{1-\gamma}\ell_{b_n,j}=0,\ \text{for } j>p\ .\end{aligned}\tag{D-13}$$

To conclude the proof, we show that (θ,ℓ') satisfies the requirements in Assumption A.6. First, for $j>p$, Eq. (D-13) implies that $\ell'_j=\lim_n\hat{\ell}_{b_n,j}=0$. Next, consider $j\leq p$. If $\ell^*_j=0$, then $\varphi^*_j(\ell^*_j)=0$ by Assumption A.5. Eq. (D-13) then implies $\ell'_j\geq 0=\varphi^*_j(\ell^*_j)$. If $\ell^*_j>0$, then $\kappa_{b_n}^{1-\gamma}\ell_{b_n,j}\to\infty$ and so Eq. (D-13) 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.11. *The set $\hat{\Theta}_I$ in Definition 4.1 satisfies the following properties: (i) $\inf_{\theta\in\Theta} Q_n(\theta) = \inf_{\theta\in\hat{\Theta}_I} Q_n(\theta)$, (ii) $\hat{\Theta}_I \neq \emptyset$.*

Proof. For the first property note that $\hat{\Theta}_I \subseteq \Theta$ implies that $\inf_{\theta\in\Theta} Q_n(\theta) \leq \inf_{\theta\in\hat{\Theta}_I} Q_n(\theta)$. Now suppose that the inequality is strict, i.e., $\inf_{\theta\in\Theta} Q_n(\theta) < \inf_{\theta\in\hat{\Theta}_I} Q_n(\theta)$. Then, there is $\tilde{\theta} \in \Theta\backslash\hat{\Theta}_I$ such that $Q_n(\tilde{\theta}) < \inf_{\theta\in\hat{\Theta}_I} Q_n(\theta)$. Since $\tilde{\theta} \notin \hat{\Theta}_I$, we conclude that $Q_n(\tilde{\theta}) > \inf_{\theta\in\Theta} Q_n(\theta) + \tau_n^\chi \geq \inf_{\theta\in\hat{\Theta}_I} Q_n(\theta)$, which together with $\tau_n \geq 0$ implies a contradiction.

For the second property, fix $n \in \mathbb{N}$ arbitrarily. Notice that $\tau_n > 0$ implies that $\exists\tilde{\theta} \in \Theta$ such that $Q_n(\tilde{\theta}) < \inf_{\theta\in\Theta} Q_n(\theta) + \tau_n^\chi$ which, in turn, implies that $\tilde{\theta} \in \hat{\Theta}_I$. □

Lemma D.12. *Let $\{F_n \in \mathcal{P}_0\}_{n\geq1}$ 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 $\{\lambda_n \in \mathbb{R}_{++}\}_{n\geq1}$ 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

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

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., $\exists\bar{\varepsilon} > 0$ such that

$$\limsup_{n\to\infty} P_{F_n}\left(\left|S(\lambda_n^{-1/\chi}v_n(\theta_n), \tilde{\Omega}_n(\theta_n))\right| > \bar{\varepsilon}\right) > 0 \ . \tag{D-14}$$

Based on this, notice that

$$\begin{aligned}\limsup_{n\to\infty} P_{F_n}\left(\left|S(\lambda_n^{-1/\chi}v_n(\theta_n), \tilde{\Omega}_n(\theta_n))\right| > \bar{\varepsilon}\right) &= \lim_{n\to\infty} P_{F_{a_n}}\left(\left|S(\lambda_{a_n}^{-1/\chi}v_{a_n}(\theta_{a_n}), \tilde{\Omega}_{a_n}(\theta_{a_n}))\right| > \bar{\varepsilon}\right)\\ &= \lim_{n\to\infty} P_{F_{b_n}}\left(\left|S(\lambda_{b_n}^{-1/\chi}v_{b_n}(\theta_{b_n}), \tilde{\Omega}_{b_n}(\theta_{b_n}))\right| > \bar{\varepsilon}\right) \ ,\end{aligned} \tag{D-15}$$

where the first equality holds for a subsequence $\{a_n\}_{n\geq1}$ of $\{n\}_{n\geq1}$ that achieves the limit supremum, the second equality holds for a subsequence $\{b_n\}_{n\geq1}$ of $\{a_n\}_{n\geq1}$ s.t. $\Omega(\theta_{b_n}) \to \Omega^*$. By Lemma D.2 (parts 5 and 6) and $\{\lambda_n^{1/\chi} \in \mathbb{R}_{++}\}_{n\geq1}$ s.t. $\lambda_n^{1/\chi} \to \infty$, 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-15) is zero, contradicting Eq. (D-14). □

Lemma D.13. *Assume Assumptions A.1-A.4. Then,*

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

where $\eta_n \equiv \tau_n \log\kappa_n$, for $\{\kappa_n\}_{n\geq1}$ and $\{\tau_n\}_{n\geq1}$ as in Assumption M.1 and Definition 4.1, respectively.

Proof. Throughout this proof, let

$$\begin{aligned}\hat{\Theta}_I^{LB} &\equiv \{\theta \in \Theta : Q_n(\theta) \leq \tau_n^\chi\} = \{\theta \in \Theta : S(\tau_n^{-1}\sqrt{n}\bar{m}_n(\theta), \hat{\Sigma}_n(\theta)) \leq 1\}\\ \hat{\Theta}_I^{UB} &\equiv \{\theta \in \Theta : Q_n(\theta) \leq \tau_n^\chi(1 + (\log\kappa_n)^{\chi/2})\} = \{\theta \in \Theta : S((\tau_n(1 + (\log\kappa_n)^{\chi/2})^{1/\chi})^{-1}\sqrt{n}\bar{m}_n(\theta), \hat{\Sigma}_n(\theta)) \leq 1\}\end{aligned}$$

where we have used the definition of Q_n and that S is homogeneous of degree χ.

Step 1. Show that $\inf_{F\in\mathcal{P}_0} P_F(\Theta_I(F)\subseteq\hat{\Theta}_I^{LB})\to 1$. Fix $(n,F)\in\mathbb{N}\times\mathcal{P}_0$ arbitrarily. By Lemma D.4 there exists $\varpi>0$ such that

$$\begin{aligned}
\hat{\Theta}_I^{LB} &\equiv \{\theta\in\Theta : S(\tau_n^{-1}\sqrt{n}\bar{m}(\theta),\hat{\Sigma}_n(\theta))\le 1\}\\
&= \{\theta\in\Theta : S(\tau_n^{-1}\sqrt{n}\hat{D}_n^{-1/2}(\theta)\bar{m}(\theta),\hat{\Omega}_n(\theta))\le 1\}\\
&\subseteq \left\{\theta\in\Theta : \left\{\begin{array}{c}\{\sqrt{n}\hat{\sigma}_{n,j}^{-1}(\theta)\bar{m}_j(\theta)\ge -\varpi\tau_n\}_{j=1}^{p}\cap\\ \{\sqrt{n}\hat{\sigma}_{n,j}^{-1}(\theta)|\bar{m}_j(\theta)|\le \varpi\tau_n\}_{j=p+1}^{k}\end{array}\right\}\right\} .
\end{aligned}$$

Based on this, consider the following derivation:

$$\begin{aligned}
\left\{\Theta_I(F)\subseteq\hat{\Theta}_I^{LB}\right\} &\supseteq \left\{\bigcap_{\theta\in\Theta_I(F)}\left\{\begin{array}{c}\{\sqrt{n}\hat{\sigma}_{n,j}^{-1}(\theta)\bar{m}_j(\theta)\ge -\varpi\tau_n\}_{j=1}^{p}\cap\\ \{\sqrt{n}\hat{\sigma}_{n,j}^{-1}(\theta)|\bar{m}_j(\theta)|\le \varpi\tau_n\}_{j=p+1}^{k}\end{array}\right\}\right\}\\
&= \left\{\begin{array}{c}\{\inf_{\theta\in\Theta_I(F)}(\sqrt{n}\hat{\sigma}_{n,j}^{-1}(\theta)\bar{m}_j(\theta))\ge -\varpi\tau_n\}_{j=1}^{p}\cap\\ \{\sup_{\theta\in\Theta_I(F)}(\sqrt{n}\hat{\sigma}_{n,j}^{-1}(\theta)|\bar{m}_j(\theta)|)\le \varpi\tau_n\}_{j=p+1}^{k}\end{array}\right\}\\
&\supseteq \left\{\left\{\inf_{\theta\in\Theta_I(F)}\tilde{v}_{n,j}(\theta)\ge -\varpi\tau_n\right\}_{j=1}^{p}\cap\left\{\sup_{\theta\in\Theta_I(F)}|\tilde{v}_{n,j}(\theta)|\le\varpi\tau_n\right\}_{j=p+1}^{k}\right\}\\
&\supseteq \left\{\sup_{\theta\in\Theta}||\tilde{v}_n(\theta)||\le\varpi\tau_n\right\} ,
\end{aligned}$$

where the second last inclusion follows from the fact that $\theta\in\Theta_I(F)$ implies $E_F[m_j(W,\theta)]\ge 0$ for $j\le p$ and $E_F[m_j(W,\theta)]=0$ for $j>p$, and the last inclusion follows from $\Theta_I(F)\subseteq\Theta$. From this, it follows that

$$\begin{aligned}
\liminf_{n\to\infty}\inf_{F\in\mathcal{P}_0} P_F\left(\Theta_I(F)\subseteq\hat{\Theta}_I^{LB}\right) &\ge \liminf_{n\to\infty}\inf_{F\in\mathcal{P}_0} P_F\left(\sup_{\theta\in\Theta}||\tilde{v}_n(\theta)||\le\varpi\tau_n\right)\\
&= \lim_{n\to\infty} P_{F_{a_n}}\left(\sup_{\theta\in\Theta}||\tilde{v}_{a_n}(\theta)||\le\varpi\tau_{a_n}\right)\\
&= \lim_{n\to\infty} P_{F_{b_n}}\left(\sup_{\theta\in\Theta}||\tilde{v}_{b_n}(\theta)||\le\varpi\tau_{b_n}\right) = 1 ,
\end{aligned}$$

where first equality holds for a subsequence $\{a_n\}_{n\ge1}$ of $\{n\}_{n\ge1}$ that achieves the limit and the infimum, the second equality holds for a subsequence $\{b_n\}_{n\ge1}$ of $\{a_n\}_{n\ge1}$ such that $\Omega_{F_{b_n}}\overset{u}{\to}\Omega$ for some $\Omega\in\mathcal{C}(\Theta^2)$ (which can be found by Lemma D.6), and the final equality follows from part 7 of Lemma D.2.

Step 2. Show that $\inf_{F\in\mathcal{P}_0} P_F(\hat{\Theta}_I^{LB}\subseteq\hat{\Theta}_I)\to 1$. Fix $(n,F)\in\mathbb{N}\times P_0$ arbitrarily. By $q_n\equiv\inf_{\theta\in\Theta}Q_n(\theta)\ge 0$, we conclude that

$$\hat{\Theta}_I^{LB} = \{\theta\in\Theta : Q_n(\theta)\le\tau_n^{\chi}\}\subseteq\{\theta\in\Theta : Q_n(\theta)-q_n\le\tau_n^{\chi}\}=\hat{\Theta}_I ,$$

which implies that $P_F(\hat{\Theta}_I^{LB}\subseteq\hat{\Theta}_I)=1$.

Step 3. Show that $\inf_{F\in\mathcal{P}_0} P_F(\hat{\Theta}_I\subseteq\hat{\Theta}_I^{UB})\to 1$. Fix $(n,F)\in\mathbb{N}\times P_0$ arbitrarily. Notice that:

$$\{q_n\le\tau_n^{\chi}\}\subseteq\left\{\{\theta\in\Theta : Q_n(\theta)\le q_n+\tau_n^{\chi}\}\subseteq\left\{\theta\in\Theta : Q_n(\theta)\le\tau_n^{\chi}(1+(\log\kappa_n)^{\chi/2})\right\}\right\}=\hat{\Theta}_I\subseteq\hat{\Theta}_I^{UB} .$$

Based on this, it suffices to show that $\inf_{F\in\mathcal{P}_0} P_F(q_n\le\tau_n^{\chi})\to 1$. To show this, notice that:

$$\liminf_{n\to\infty}\inf_{F\in\mathcal{P}_0} P_F\left(q_n\le\tau_n^{\chi}(\log\kappa_n)^{\chi/2}\right) = \lim_{n\to\infty} P_{F_{a_n}}(q_{a_n}\le\tau_{a_n}^{\chi}(\log\kappa_{a_n})^{\chi/2}) = \lim_{n\to\infty} P_{F_{b_n}}(q_{b_n}\le\tau_{b_n}^{\chi}(\log\kappa_{b_n})^{\chi/2}) = 1 ,$$

where the first equality holds for a subsequence $\{a_n\}_{n\ge1}$ of $\{n\}_{n\ge1}$ that achieves the limit and the infimum, the second equality holds for a subsequence $\{b_n\}_{n\ge1}$ of $\{a_n\}_{n\ge1}$ s.t. $\Omega_{F_{a_n}}\overset{u}{\to}\Omega$ for some $\Omega\in\mathcal{C}(\Theta^2)$, and the third equality holds by Lemma D.12.

Step 4. Show that $\inf_{F\in\mathcal{P}_0} P_F(\hat{\Theta}_I^{UB}\subseteq\Theta_I^{\eta_n}(F))\to 1$. Fix $(n,F)\in\mathbb{N}\times\mathcal{P}_0$ arbitrarily. By Lemma D.4 there exists

$\varpi > 0$ such that

$$\begin{aligned}\Theta_I^{\eta_n}(F) &\equiv \{\theta \in \Theta : S(\eta_n^{-1}\sqrt{n}E_F[m(W,\theta)], \Sigma_F(\theta)) \leq 1\} \\ &= \{\theta \in \Theta : S(\eta_n^{-1}\sqrt{n}D_F^{-1}(\theta)E_F[m(W,\theta)], \Omega_F(\theta)) \leq 1\} \\ &\subseteq \left\{\theta \in \Theta : \left\{ \begin{array}{c} \{\sqrt{n}\sigma_{F,j}^{-1}(\theta)E_F[m_j(W,\theta)] \geq -\varpi\eta_n\}_{j=1}^{p} \cap \\ \{\sqrt{n}\sigma_{F,j}^{-1}(\theta)|E_F[m_j(W,\theta)]| \leq \varpi\eta_n\}_{j=p+1}^{k} \end{array} \right\}\right\} .\end{aligned}$$

Based on this, consider the following derivation:

$$\begin{aligned}\left\{\hat{\Theta}_I \subseteq \Theta_I^{\eta_n}(F)\right\} &\supseteq \left\{ \bigcap_{\theta \in \hat{\Theta}_I^{UB}} \left\{ \begin{array}{c} \{\sqrt{n}\sigma_{F,j}^{-1}(\theta)E_F[m_j(W,\theta)] \geq -\varpi\eta_n\}_{j=1}^{p} \cap \\ \{\sqrt{n}\sigma_{F,j}^{-1}(\theta)|E_F[m_j(W,\theta)]| \leq \varpi\eta_n\}_{j=p+1}^{k} \end{array} \right\}\right\} \\ &= \left\{ \begin{array}{c} \{\inf_{\theta \in \hat{\Theta}_I^{UB}} \sqrt{n}\sigma_{F,j}^{-1}(\theta)E_F[m_j(W,\theta)] \geq -\varpi\eta_n\}_{j=1}^{p} \cap \\ \{\sup_{\theta \in \hat{\Theta}_I^{UB}} \sqrt{n}\sigma_{F,j}^{-1}(\theta)|E_F[m_j(W,\theta)]| \leq \varpi\eta_n\}_{j=p+1}^{k} \end{array} \right\} \\ &\supseteq \left\{ \max_{j=1,\dots,k} \sup_{\theta \in \hat{\Theta}_I^{UB}} |\sqrt{n}\sigma_{F,j}^{-1}(\theta)E_F[m_j(W,\theta)]| \leq \varpi\eta_n \right\} .\end{aligned}$$

From this, it follows that

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

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

Fix $(n, F, \theta, j) \in \mathbb{N} \times \mathcal{P}_0 \times \hat{\Theta}_I^{UB} \times \{1, \dots, k\}$ 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 Lemma D.4 then implies that

$$\inf_{\theta \in \hat{\Theta}_I^{UB}} \sqrt{n}\sigma_{F,j}^{-1}(\theta)E_F[m_j(W,\theta)] \geq -\sup_{\tilde{\theta}\in\Theta} |v_{n,j}(\tilde{\theta})| - \varpi\tau_n \sup_{\breve{\theta}\in\Theta} |\sigma_{F,j}^{-1}(\breve{\theta})\hat{\sigma}_{n,j}(\breve{\theta})| .$$

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

$$\sup_{\theta \in \hat{\Theta}_I^{UB}} \sqrt{n}\sigma_{F,j}^{-1}(\theta)E_F[m_j(W,\theta)] \leq \sup_{\tilde{\theta}\in\Theta} |v_{n,j}(\tilde{\theta})| + \varpi\tau_n \sup_{\breve{\theta}\in\Theta} |\sigma_{F,j}^{-1}(\breve{\theta})\hat{\sigma}_{n,j}(\breve{\theta})| .$$

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

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

From this, it follows that

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

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

References

Andrews, D. W. K. and P. Barwick (2012): "Inference for Parameters Defined by Moment Inequalities: A Recommended Moment Selection Procedure," Forthcoming: Econometrica.

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

Andrews, D. W. K. and G. Soares (2010): "Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection," *Econometrica*, 78, 119–157.

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

Bugni, F., I. Canay, and P. Guggenberger (2012): "Distortions of Asymptotic Confidence Size in Locally Misspecified Moment Inequality Models," *Econometrica*, 80, 1741–1768.

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

Chernozhukov, V., H. Hong, and E. Tamer (2007): "Parameter Set Inference in a Class of Econometric Models," *Econometrica*, 75, 1243–1284.

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

Guggenberger, P., J. Hahn, and K. Kim (2008): "Specification Testing Under Moment Inequalities," *Economics Letters*, 2, 375–378.

Haavelmo, T. (1944): "The probability approach in econometrics," *Econometrica*, 12, iii–vi+1–115.

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

Kitamura, Y. and J. Stoye (2012): "Nonparametric Analysis of Random Utility Models: Testing," Working Paper, Cornell University.

Koopmans, T. (1937): *Linear Regression Analysis of Economic Time Series*, Netherlands Economic Institute, Haarlem.

Lehman, E. and J. Romano (2005): *Testing Statistical Hypothesis: Third edition*, Sringer.

Manski, C. F. (1989): "Anatomy of the Selection Problem," *The Journal of Human Resources*, 24, 343–360.

——— (2003): *Partial Identification of Probability Distributions*, Springer-Verlag.

Mikusheva, A. (2010): "Robust Confidence Sets in the Presence of Weak Instruments," *Journal of Econometrics*, 157, 234–247.

Ponomareva, M. and E. Tamer (2011): "Misspecification in moment inequality models: back to moment equalities?" *Econometrics Journal*, 14, 186–203.

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

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," *Annual Review of Economics*, 2, 167–195.

van der Vaart, A. and J. Wellner (1996): *Weak Convergence and Empirical Processes*, Springer.