Supplemental Material to
Misspecified Moment Inequality Models: Inference and Diagnostics

By

Donald W.K. Andrews, and Soonwon Kwon

March 2017

Revised July 2022

COWLES FOUNDATION DISCUSSION PAPER NO. 2184R2

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

http://cowles.yale.edu/

Online Appendix A

to

Misspecified Moment Inequality Models: Inference and Diagnostics

Contents: Simulation Results and Related Results

Donald W. K. Andrews
Cowles Foundation for Research in Economics
Yale University

Soonwoo Kwon
Department of Economics
Brown University

First Version: March 12, 2017
Revised: July 29, 2022

7 Outline of Online Appendix A

References to sections with section numbers 6 or less refer to sections of the main paper. Similarly, all equations, theorems, and lemmas with section numbers 6 or less refer to results in the main paper.

Section 8 provides simulation results for the size and power of the misspecification index (MI), SPUR1, and SPUR2 tests in misspecified and correctly-specified versions of two models. (The MI test is the two-sided test obtained by inverting the MI CI $CI_{n,\Delta}(\alpha)$ defined in (5.3).) The first model is a lower/upper bound model. The second model is a missing data model. Section 8 also assesses the sensitivity to the tuning parameters of the rejection probabilities of the two-sided MI and SPUR2 tests under the null and alternative hypotheses in the lower/upper bound model.

Section 9 provides derivations for two formulae for the missing data model that are employed in Section 8.

Section 10 concerns the empirical illustration in the main paper. It assesses the sensitivity of the MI and SPUR2 CI's to changes in the tuning parameters, provides simulated power results for a simplified version of the model, shows how the moment inequalities in (6.2) are obtained, and describes the initial values used in the optimization problems that deliver GMS and SPUR projection CI's.

Section 11 establishes the uniform consistency, under correct model specification and misspecification, of an estimator of the MR-identified set. Rate of convergence results for this set estimator are also given using arguments similar to those in Chernozhukov, Hong, and Tamer (2007).

Section 12 shows, using the simple lower/upper model, that subsampling a SPUR test statistic does not necessarily deliver correct asymptotic size under model misspecification.

We note that Appendix D of Andrews and Kwon (2022) provides some additional results: (i) an alternative interpretation of the identified set $\Theta_I^{MR}(F)$, (ii) the equivalence of the SPUR test statistic to a recentered test statistic, as has been considered in Chernozhukov, Hong, and Tamer (2007), when the "max" S function is employed, and (iii) extensions of the results of the paper to non-i.i.d. observations, to tests with weighted moment inequalities, and to tests without the standard-deviation normalization.

Let $[x]_- := \max\{-x, 0\}$ (≥ 0) for $x \in R$.

8 Simulation Results

In this section, we provide Monte Carlo simulation results that illustrate the performance of the misspecification index tests, SPUR1 tests, and SPUR2 tests. When the model under consideration is correctly specified, we compare SPUR1 and SPUR2 tests to the standard GMS test. We consider two simple models under various levels of misspecification (i.e., different values of $r_{F_n}^{\inf}$). All simulation results are based on 1,000 simulation repetitions, 500 bootstrap replications, a sample size of $n = 250$, $\kappa_n = \tau_n = (\ln n)^{1/2}$, and $S(\cdot) = S_1(\cdot)$. The GMS function $\varphi(\cdot)$ employed is $\varphi(\xi) = \infty 1(\xi_j > 1)$ for $j \leq k$. The significance level is fixed at $\alpha = .05$ with $\alpha_1 = .005$ and $\alpha_2 = .045$ for the SPUR2 test.

8.1 Lower/Upper Bound Model

8.1.1 The Model

We consider a simple model where the means of the observations impose lower and upper bounds on a scalar parameter. The data $\{W_i\}_{i \leq n}$ are i.i.d. with $W_i = (W_{i1}, \ldots, W_{ik})' \sim N(\mu, I_k)$, where $\mu = (\mu_1, \ldots, \mu_k)' \in R^k$ and I_k denotes the $k \times k$ identity matrix. We consider $k = 2$, 4, and 8. The parameter space Θ is taken to be $[-20, 20]$. We consider various configurations of μ. When $r_F^{\inf} > 0$, the MR-identified set is always a singleton in this model, but it may have different lengths when $r_F^{\inf} = 0$. Accordingly, when $r_F^{\inf} = 0$ we consider configurations that correspond to different lengths of the MR-identified set.

For $k = 2$, the population moment inequalities are

$$E_F W_{i1} \leq \theta \,\text{and}\, \theta \leq E_F W_{i2}. \tag{8.1}$$

The model is identifiably misspecified (i.e., $\Theta_I(F)$ is empty) if and only if $\mu_1 > \mu_2$. In this model, $\Delta_F^{\inf} = (\mu_1 - \mu_2)/2$ and $r_F^{\inf} = [\mu_1 - \mu_2]_+/2$, where $[x]_+ := \max\{x, 0\}$. We take $\mu = (r, -r)'$ for each $r \in \{.5, 1, 2, 5\}$ as the misspecified cases. We have $r_F^{\inf} = r$ and $\Theta_I^{MR}(F) = \{0\}$ in these cases. For the correctly-specified cases, we take $\mu = (-\ell, 0)'$ for each $\ell \in \{0, .5, 1, 2\}$. Here the MR-identified set is $\Theta_I^{MR}(F) = [-\ell, 0]$, which has length ℓ, and $\Delta_F^{\inf} = -\ell/2$.

For $k = 4$, the moment inequalities are

$$E_F W_{i1} \leq \theta,\; E_F W_{i2} \leq \theta,\; \theta \leq E_F W_{i3},\; \text{and}\, \theta \leq E_F W_{i4}. \tag{8.2}$$

Identifiable misspecification arises if and only if $\max\{\mu_1, \mu_2\} > \min\{\mu_3, \mu_4\}$. In this model, $\Delta_F^{\inf} = (\max\{\mu_1, \mu_2\} - \min\{\mu_3, \mu_4\})/2$ and $r_F^{\inf} = [\max\{\mu_1, \mu_2\} - \min\{\mu_3, \mu_4\}]_+/2$. For $k = 4$, many different configurations of μ are possible for a given value of $r_F^{\inf} > 0$ or a given length of the MR-identified set when $r_F^{\inf} = 0$. Accordingly, we consider several scenarios for $k = 4$. For the misspecified cases, we consider five different scenarios: "binding," "almost binding," "somewhat slack," "very slack," and "slack/almost binding."[28] In each scenario, we consider $r_F^{\inf} = .5$ and 1. Regardless of the scenario and the value of $r_F^{\inf}$, the MR-identified set is $\Theta_I^{MR}(F) = \{0\}$. For the correctly-specified cases and $k = 4$, we consider the same five scenarios as for the misspecified cases. However, the definitions are slightly different in the correctly-specified cases.[29] The MR-identified set takes the form $\Theta_I^{MR}(F) = [-\ell, 0]$ for each $\ell \in \{0, .5, 1\}$.

For $k = 8$, the moment inequalities are

$$\begin{aligned} &E_F W_{ij} \leq \theta \text{ for } 1 \leq j \leq 4 \text{ and} \\ &\theta \leq E_F W_{ij} \text{ for } 5 \leq j \leq 8. \end{aligned} \tag{8.3}$$

The definition of each scenario is analogous to the $k = 4$ cases, with each entry repeated twice. That is, if $\mu^4 = (\mu_1, \mu_2, \mu_3, \mu_4)' \in R^4$ is the mean vector used under some scenario for $k = 4$, then $\mu^8 = (\mu_1, \mu_1, \mu_2, \mu_2, \mu_3, \mu_3, \mu_4, \mu_4)' \in R^8$ is the mean vector used in the same scenario for $k = 8$.

8.1.2 Rejection Probabilities of the Misspecification Index Test and the SPUR1 and SPUR2 Tests

Figure 8.1 gives the simulated rejection probabilities of nominal .05 two-sided tests concerning the misspecification index $\Delta_F^{\inf}$ for $k = 2$. Each plot shows, for different values of $\Delta_F^{\inf} \in \{-5, -2, -1, -.5, , 0, .5, 1, 2, 5\}$, the rejection probabilities of the MI test of $H_0 : \Delta_F^{\inf} = \Delta_0$ versus $H_1 : \Delta_F^{\inf} \neq \Delta_0$ for a range of Δ_0 values and a fixed $\Delta_F^{\inf}$ value.[30] The two-sided MI test rejects H_0 if $\Delta_0 \notin CI_{n,\Delta}(.05)$, which is defined in (5.3). Figure 8.2 gives simulation results for the MI test for $k = 4$ under the "binding" and "very slack" scenarios, which are the two extreme scenarios. For brevity, the MI test results for the "almost binding," "somewhat slack," and "slack/almost binding" scenarios are not reported because they lie between the two extreme scenarios and the results for the latter two scenarios do not differ very much. Similarly, and for the same reasons, for $k = 8$, Figure 8.3 gives simulation results for the MI test for the "binding" and "very slack" scenarios only, which are the two extreme scenarios.

[28] For given $r > 0$, the mean vectors μ in the five misspecified scenarios are (i) "binding": $\mu = (r, r, -r, -r)'$, (ii) "almost binding": $\mu = (r, r - .1, -r + .1, -r)'$, (iii) "somewhat slack": $\mu = (r, r - .5, -r + .5, -r)'$, (iv) "very slack": $\mu = (r, r - 1, -r + 1, -r)'$, and (v) "slack/almost binding": $\mu = (r, r - .1, -r + 1, -r)'$. In each scenario, $r_F^{\inf} = r$ and the MR-identified set is $\Theta_I^{MR}(F) = \{0\}$.

[29] For given $\ell > 0$, the mean vectors μ in the five correctly-specified scenarios are (i) "binding": $\mu = (-\ell, -\ell, 0, 0)'$, (ii) "almost binding": $\mu = (-\ell - .1, -\ell, 0, .1)'$, (iii) "somewhat slack": $\mu = (-\ell - .5, -\ell, 0, .5)'$, (iv) "very slack": $\mu = (-\ell - 1, -\ell, 0, 1)'$, and (v) "slack/almost binding": $\mu = (-\ell - 1, -\ell, 0, .1)'$. In all scenarios, $\Theta_I^{MR}(F) = [-\ell, 0]$ and the MR-identified set has length ℓ.

[30] That is, Figure 8.1 reports power for a fixed true $\Delta_F^{\inf}$ value and the null value being Δ_0 for a range of Δ_0 values. This differs from, but is no less informative than, a conventional power function that considers a fixed null value and a range of true alternative values.

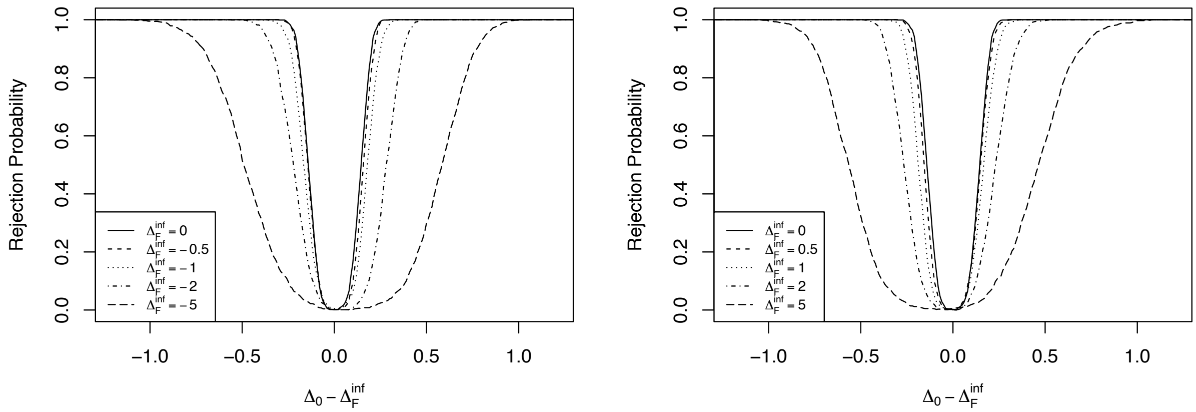

Figure 8.1: Rejection probabilities of tests concerning the misspecification index Δ_F^{inf} for $k = 2$. Each plot shows, for different values of Δ_F^{inf}, the rejection probabilities of the nominal .05 two-sided misspecification index test of the null hypothesis $H_0 : \Delta_F^{\text{inf}} = \Delta_0$ for a range of Δ_0 values and a fixed Δ_F^{inf} value. The two-sided misspecification index test rejects H_0 if $\Delta_0 \notin CI_{n,\Delta}(.05)$, which is defined in (4.3).

Figures 8.1–8.3 show that the rejection probabilities of the MI test are well behaved. They monotonically increase in $|\Delta_0 - \Delta_F^{\text{inf}}|$ to 1 for positive values of $\Delta_0 - \Delta_F^{\text{inf}}$ and negative values of $\Delta_0 - \Delta_F^{\text{inf}}$. In several scenarios, the plots are relatively flat to the immediate right of the point $\Delta_0 - \Delta_F^{\text{inf}} = 0$. This implies that it is difficult to reject the null hypothesis of a low level of misspecification when the truth is a correctly-specified singleton identified set. For the cases with $k = 4$ and 8, power is highest in the "binding" scenarios, but the differences across the different slackness scenarios are not very large. The MI test has noticeably higher power for Δ_F^{inf} close to zero, i.e., for $\Delta_F^{\text{inf}} \in \{-1, -.5, 0, .5, 1\}$, than for large $|\Delta_F^{\text{inf}}|$, i.e., for $\Delta_F^{\text{inf}} \in \{-5, -2, 2, 5\}$. This is advantageous when one is interested in determining whether Δ_F^{inf} is nonnegative versus positive. Looking at the rejection probabilities for $\Delta_0 - \Delta_F^{\text{inf}} = 0$, we see that the MI test has correct size, but under-rejects with the null rejection probabilities being close to 0 in the cases considered.

Figure 8.4 gives the simulated rejection probabilities, i.e., power, of the SPUR1 and SPUR2 tests for a range of null values $\theta_0 \geq 0$ for the misspecified cases for $k = 2$.[31] Figure 8.5 provides the simulated rejection probabilities of the SPUR1, SPUR2, and standard GMS tests in the correctly-specified models for $k = 2$ for fixed $\Theta_I^{MR}(F) = [-\ell, 0]$ for a range of null hypothesis values $\theta_0 \geq 0$ for $\ell \in \{0, .5, 1, 2\}$. Figure 8.6 gives the simulation results for the SPUR1 and SPUR2 tests for the

[31] That is, Figure 8.4 reports power for the true θ being 0, which is in $\Theta_I(F) = \{0\}$, and the null being $\theta_0 > 0$ for a range of θ_0 values. This differs from, but is no less informative than, a conventional power function that considers a fixed null value and a range of true alternative values.

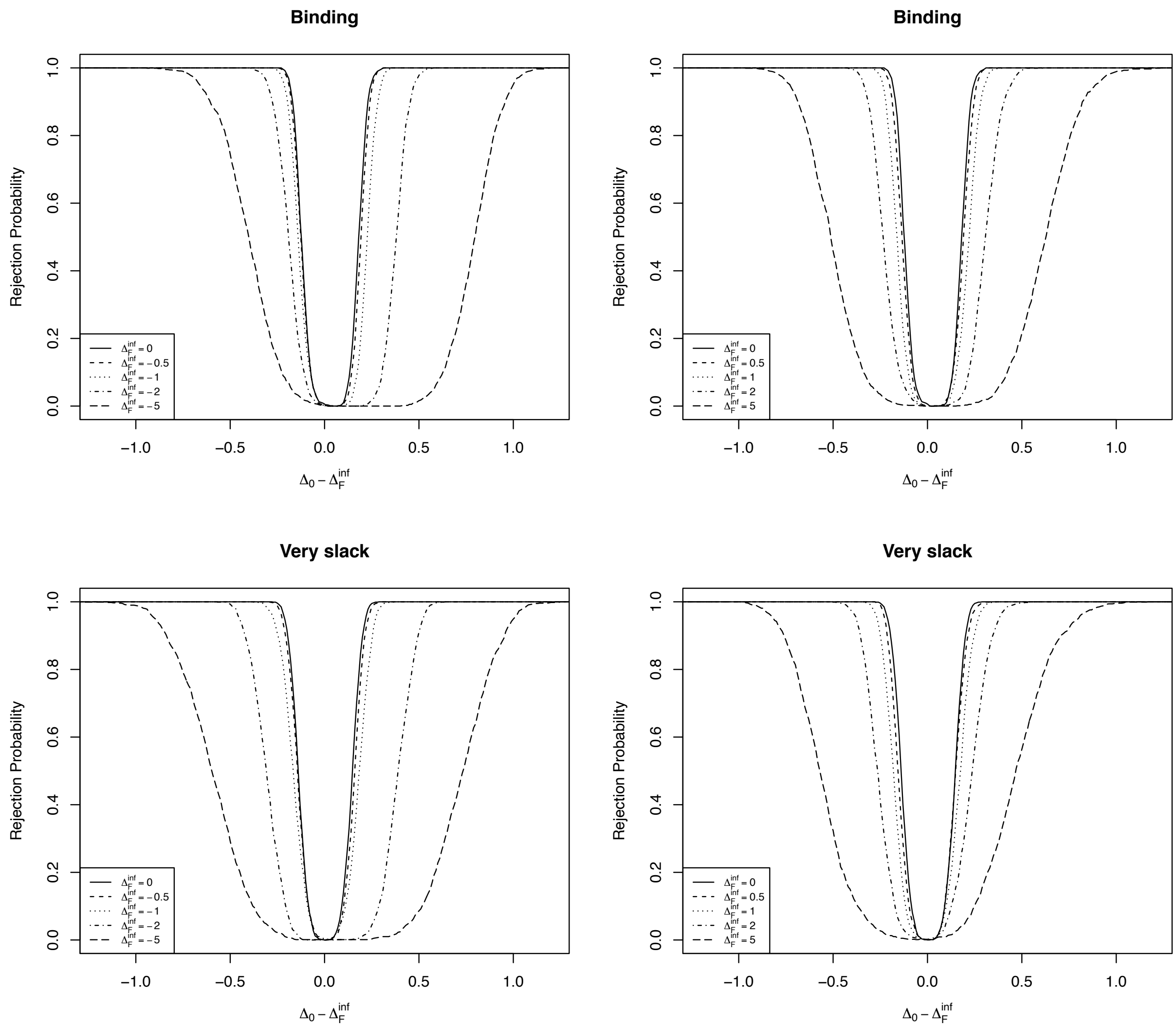

Figure 8.2: Rejection probabilities of tests concerning the misspecification index Δ_F^{inf} for $k = 4$. Each plot shows, for different values of Δ_F^{inf}, the rejection probabilities of the nominal .05 two-sided misspecification index test of the null hypothesis $H_0 : \Delta_F^{\text{inf}} = \Delta_0$ for a range of Δ_0 values and a fixed Δ_F^{inf} value. The two-sided misspecification index test rejects H_0 if $\Delta_0 \notin CI_{n,\Delta}(.05)$, which is defined in (4.3). Results are given for the binding and very slack scenarios.

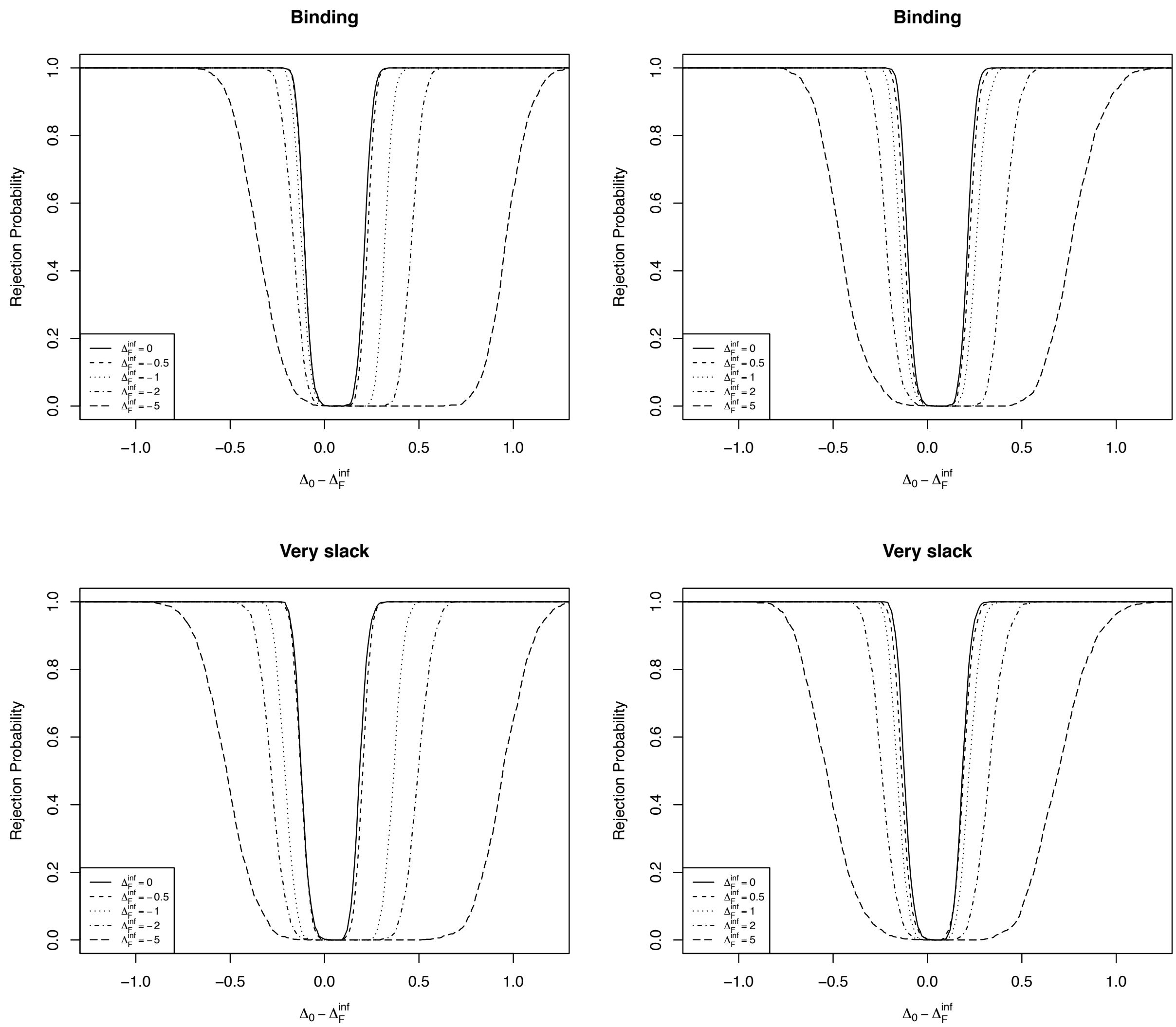

Figure 8.3: Rejection probabilities of tests concerning the misspecification index Δ_F^{inf} for $k = 8$. Each plot shows, for different values of Δ_F^{inf}, the rejection probabilities of the nominal .05 two-sided misspecification index test of the null hypothesis $H_0 : \Delta_F^{\text{inf}} = \Delta_0$ for a range of Δ_0 values and a fixed Δ_F^{inf} value. The two-sided misspecification index test rejects H_0 if $\Delta_0 \notin CI_{n,\Delta}(.05)$, which is defined in (4.3). Results are given for the binding and very slack scenarios.

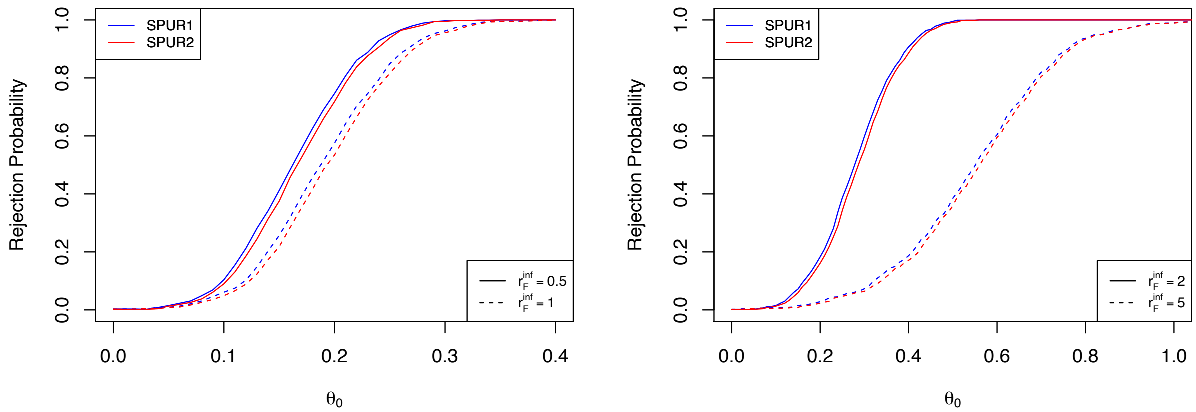

Figure 8.4: Rejection probabilities of tests concerning θ for misspecified cases for $k = 2$. Each plot shows, for different values of r_F^{inf}, the rejection probabilities of the nominal .05 SPUR1 and SPUR2 tests for the null hypothesis $H_0 : \theta = \theta_0$ for a range of θ_0 values and fixed identified set $\Theta_I(F) = \{0\}$.

misspecified cases for $k = 4$ in the "binding," "almost binding," "somewhat slack," "very slack," and "slack/almost binding" scenarios. Figure 8.7 gives the corresponding results for the correctly-specified models for $k = 4$. Figures 8.8 and 8.9 give the simulation results for the SPUR1 and SPUR2 tests for $k = 8$ for all five slackness scenarios in the misspecified and correctly specified scenarios, respectively.

Figures 8.4, 8.6, and 8.8 show that the performance of the two tests, SPUR1 and SPUR2, is quite similar under misspecification (i.e., $r_F^{\text{inf}} > 0$), which is what we expect given the discussion in Section 4.2. Looking at the rejection probability at $\theta_0 = 0$, we see that both tests have correct size, but under-reject with the null rejection probabilities being close to 0. The rejection probabilities increase to 1 fairly quickly as the distance between the null value and the MR-identified set increases. The tests perform better in terms of power when r_F^{inf} is smaller, but they perform reasonably well even when r_F^{inf} is as large as 5, which is five times the standard deviation of the moment functions. Additionally, for the cases with $k = 4$ and 8, we see that the performance of the tests does not differ much across the different scenarios.

For the correctly-specified cases, we focus on the comparison of the SPUR1 and SPUR2 tests with the standard GMS test, which is known to perform well in such cases. From the discussion in Section 4.2, we expect the SPUR2 and standard GMS tests to exhibit similar performance when the length of the identified set is large enough. Indeed, in Figure 8.5, we see that when the length

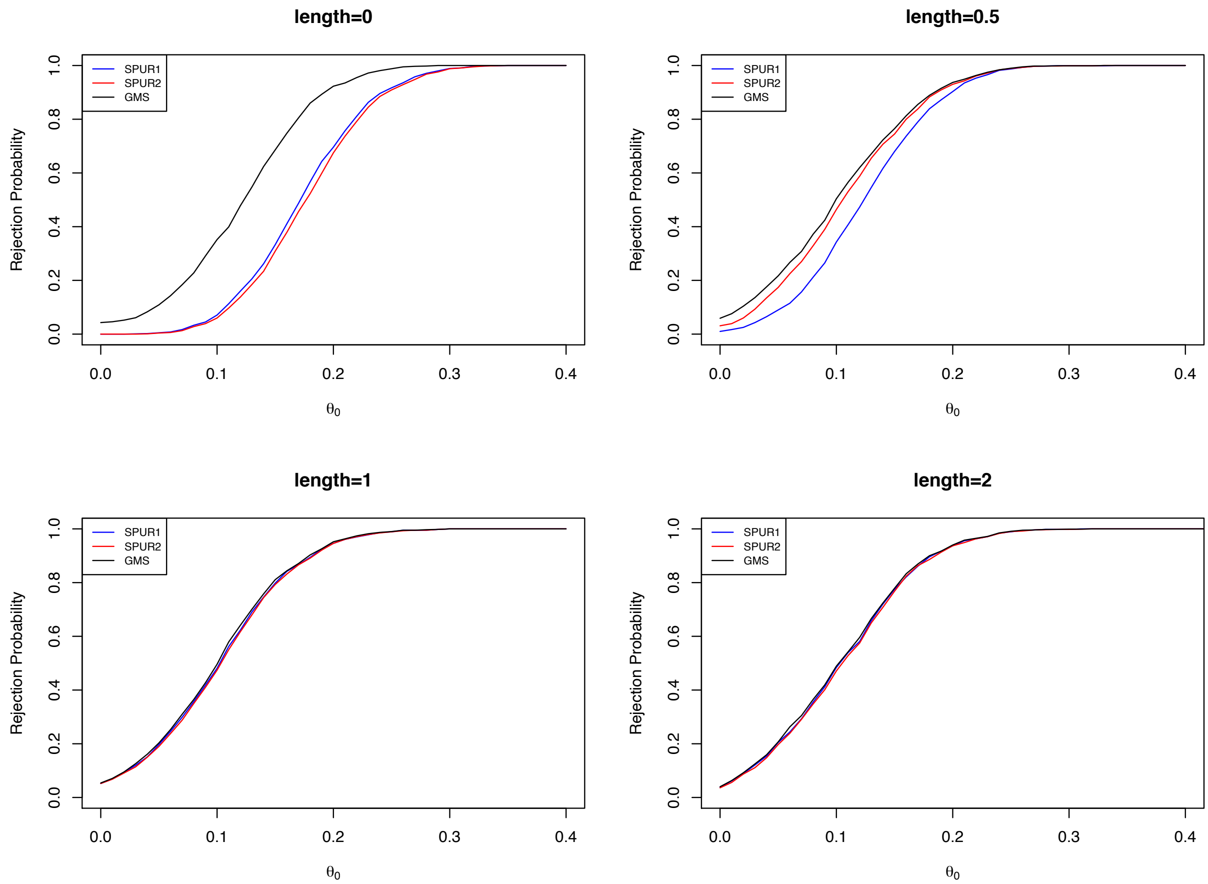

Figure 8.5: Rejection probabilities of tests concerning θ for correctly specified cases for $k = 2$. Each plot shows, for different lengths ℓ of the identified set, the rejection probabilities of the nominal .05 SPUR1, SPUR2, and standard GMS tests for the null hypothesis $H_0 : \theta = \theta_0$ for a range of θ_0 values and identified set $\Theta_I(F) = [-\ell, 0]$.

of the identified set is .5 the rejection probabilities of the two tests are very close to each other, and when the length is greater than .5 all three tests are essentially indistinguishable. We can also see that the SPUR2 test catches up to the standard GMS test under shorter identified sets than the SPUR1 test does, which shows its adaptive nature. However, when the identified set is a singleton, the SPUR1 and SPUR2 tests are more conservative than the standard GMS test under the null and have lower power over a wide range of positive θ_0 values. Essentially the same occurs when $k = 4$. That is, for each of the scenarios, the SPUR1 and SPUR2 tests are more conservative when the identified set has length 0, the SPUR2 test performs similarly to the standard GMS test when the length is .5, and all three tests are indistinguishable when the length is greater than .5. Again, this exhibits the adaptive nature of the SPUR2 test. The pattern for $k = 8$ is similar, although there is a larger gap between the power of the SPUR2 and GMS tests. When $k = 4$ and 8, the

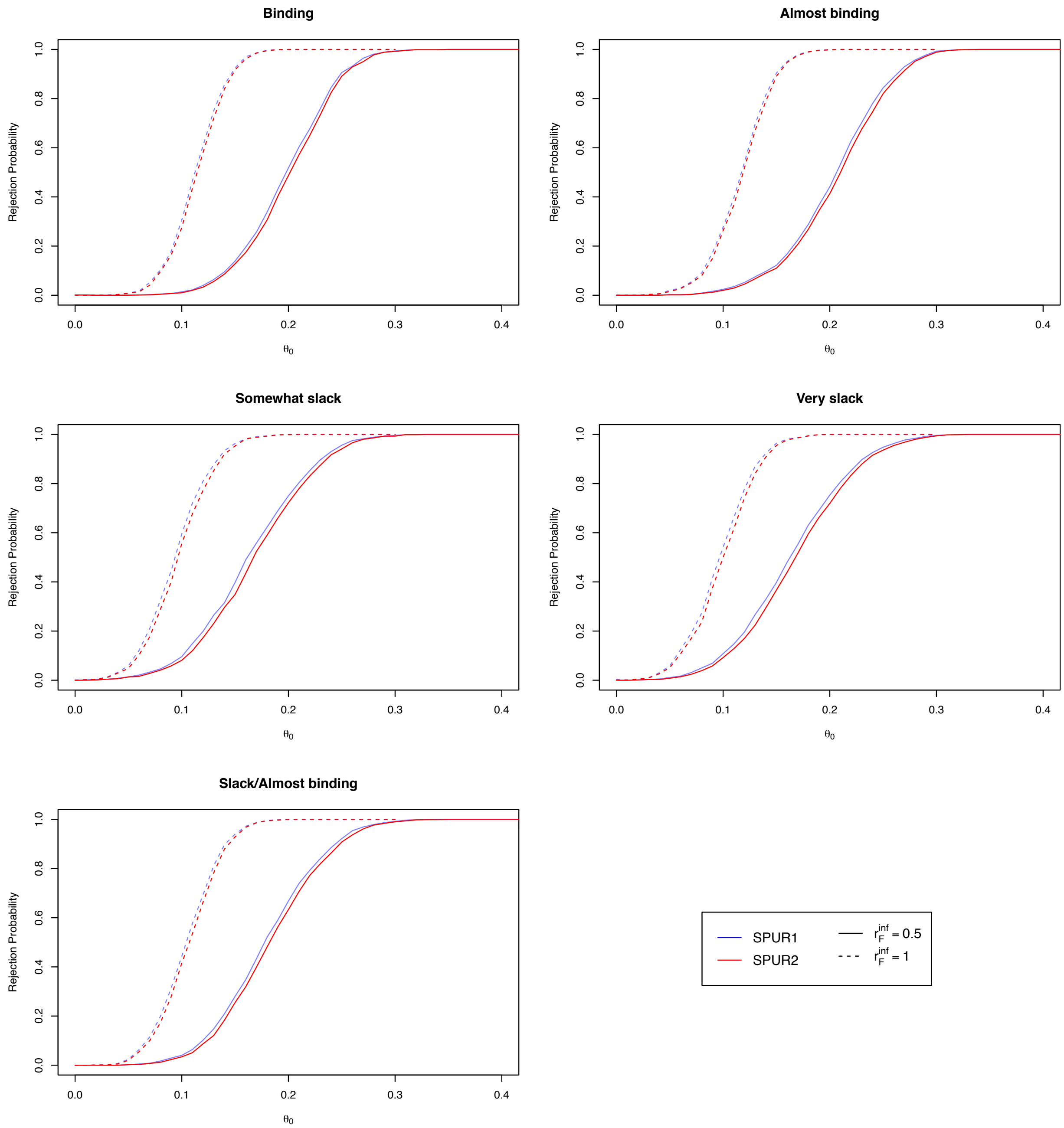

Figure 8.6: Rejection probabilities of tests concerning θ for misspecified cases for $k = 4$. Each plot shows, under different scenarios, the rejection probabilities of the nominal .05 SPUR1 and SPUR2 tests for the null hypothesis $H_0 : \theta = \theta_0$ for a range of θ_0 values, identified set $\Theta_I(F) = \{0\}$, and two different values of r_F^{inf}.

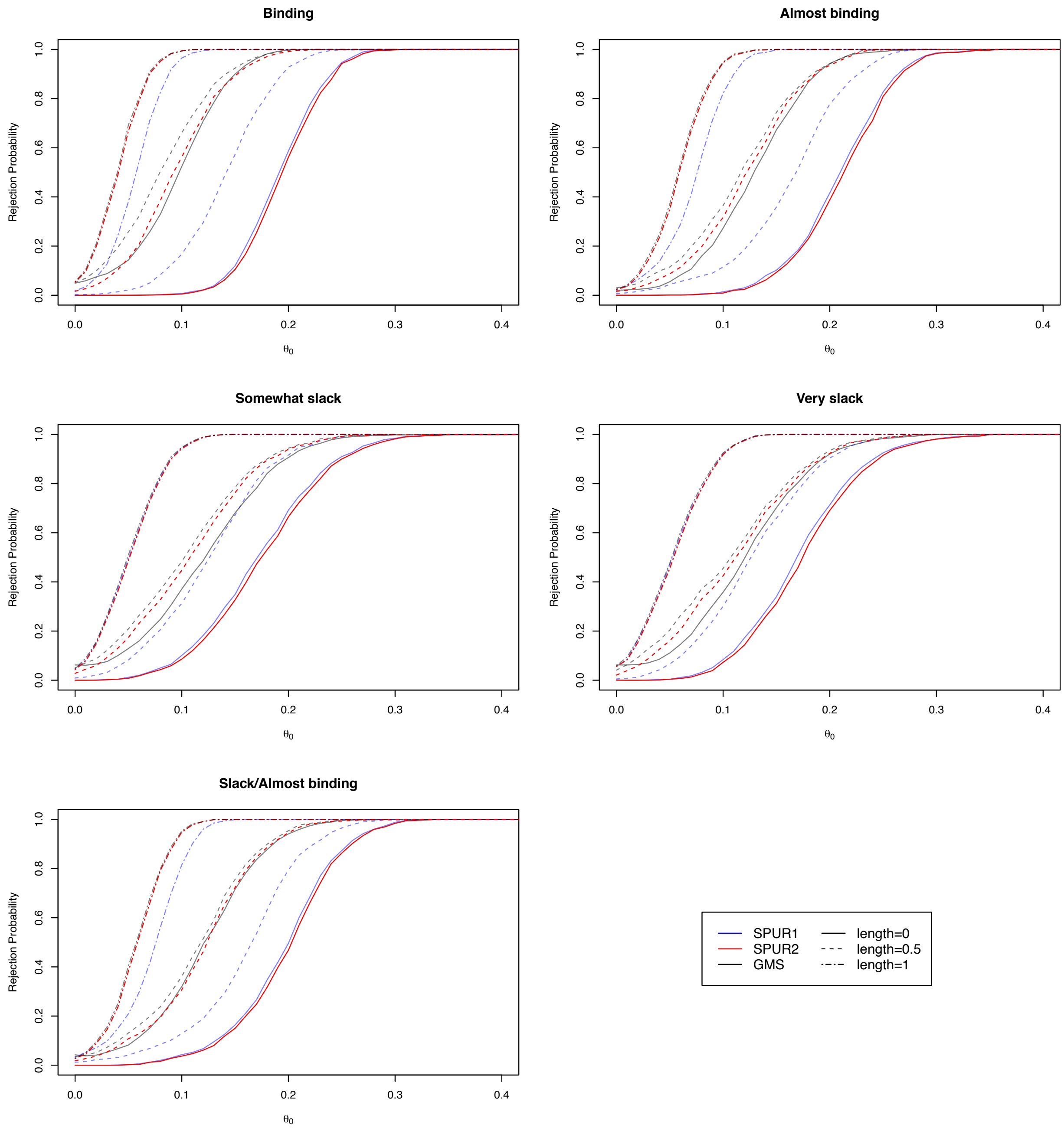

Figure 8.7: Rejection probabilities of tests concerning θ for correctly specified cases for $k = 4$. Each plot shows, under different scenarios, the rejection probabilities of the nominal .05 SPUR1, SPUR2, and standard GMS tests for the null hypothesis $H_0 : \theta = \theta_0$ for a range of θ_0 values and different lengths ℓ of the identified set $\Theta_I(F) = [-\ell, 0]$.

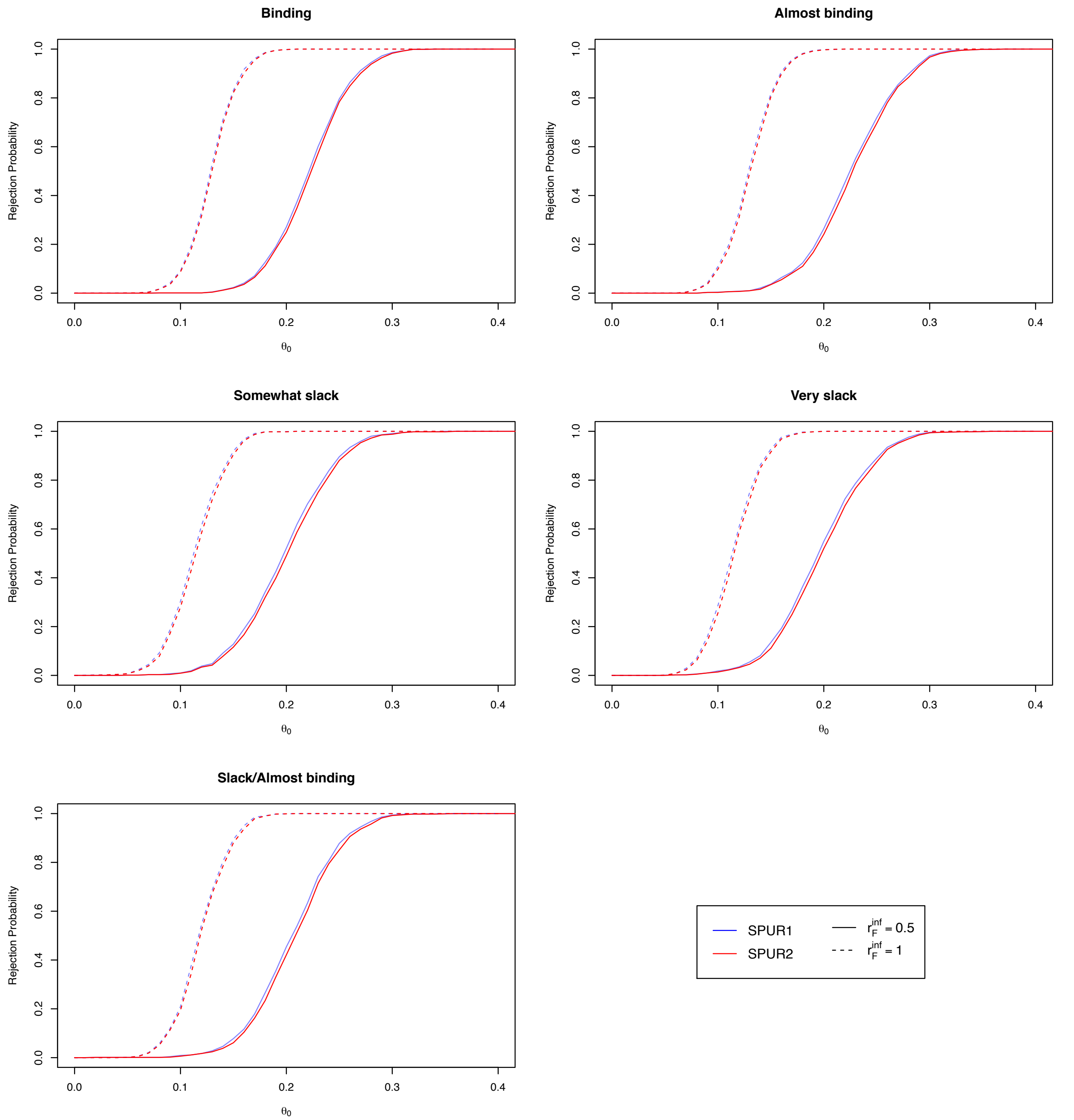

Figure 8.8: Rejection probabilities of tests concerning θ for misspecified cases for $k = 8$. Each plot shows, under different scenarios, the rejection probabilities of the nominal .05 SPUR1 and SPUR2 tests for the null hypothesis $H_0 : \theta = \theta_0$ for a range of θ_0 values, identified set $\Theta_I(F) = \{0\}$, and two different values of r_F^{inf}.

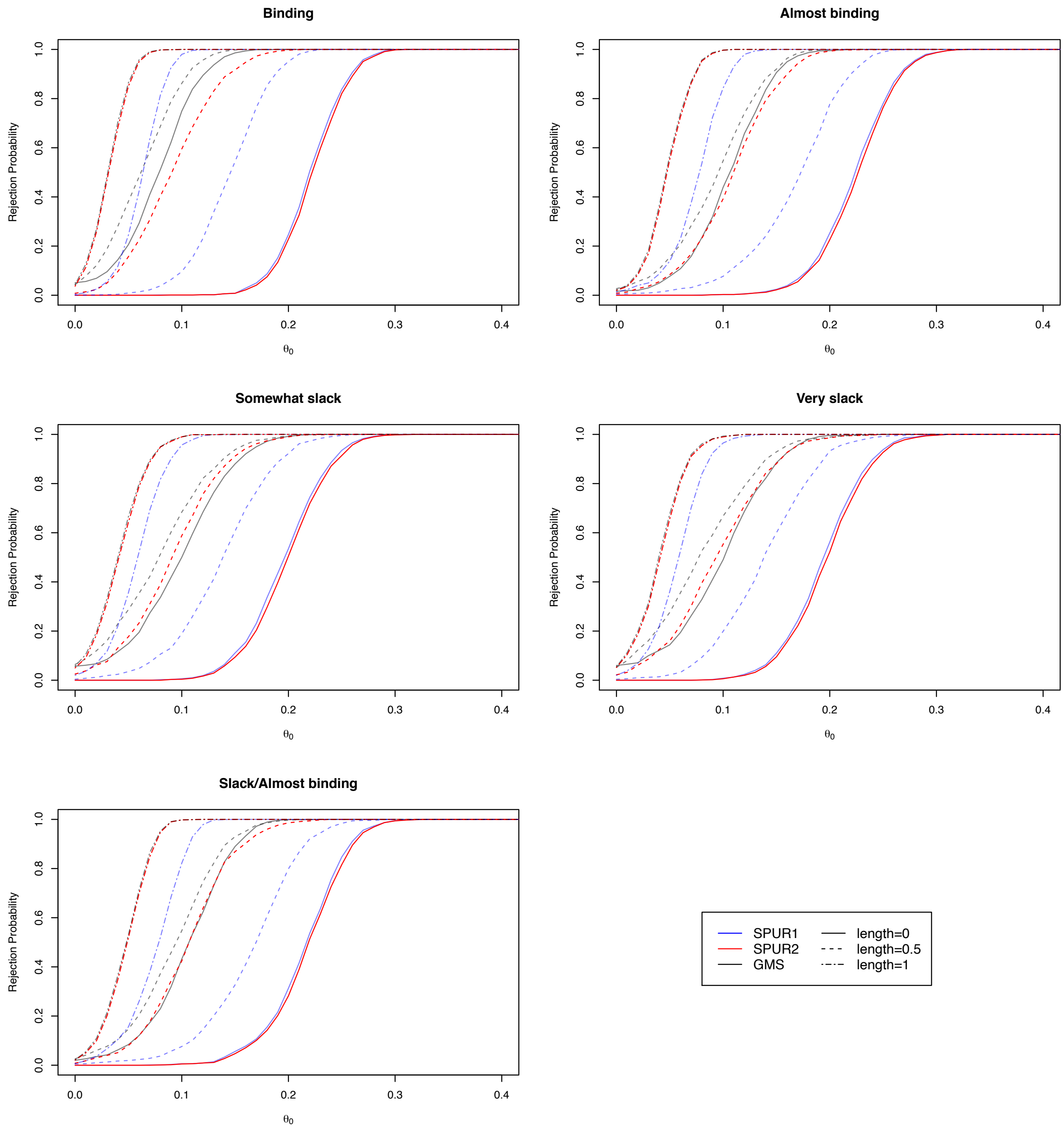

Figure 8.9: Rejection probabilities of tests concerning θ for correctly specified cases for $k = 8$. Each plot shows, under different scenarios, the rejection probabilities of the nominal .05 SPUR1, SPUR2 and standard GMS tests for the null hypothesis $H_0 : \theta = \theta_0$ for a range of θ_0 values and different lengths ℓ of the identified set $\Theta_I(F) = [-\ell, 0]$.

discrepancy between the standard GMS test and the SPUR1 and SPUR2 tests is largest in the "binding" scenario.

In sum, (i) when the model is misspecified, the SPUR1 and SPUR2 tests perform quite similarly, with their rejection probabilities reaching 1 fairly quickly as the distance between the null value and the MR-identified set increases, and (ii) when the model is correctly specified, the SPUR2 test performs similarly to the GMS test provided the identified set is not too short, and likewise for the SPUR1 test for somewhat longer identified sets. Thus, the SPUR2 test performs better than the SPUR1 test when the identified set is small, but not too small.

8.1.3 Sensitivity to Tuning Parameters

Now, we assess the sensitivity to the tuning parameters of the rejection probabilities of the two-sided MI and SPUR2 tests under the null and alternative hypotheses. The baseline values for the tuning parameters are: $\tau_{n,base} = \kappa_{n,base} = (\ln n)^{1/2}$, $\alpha_{1,base} = .005$ (when $\alpha = .05$), $\iota_{base} = 10^{-6}$, and $B_{base} = 1000$. We alter these tuning parameters one at a time. For τ_n, we consider $c_\tau \tau_{n,base}$ for $c_\tau = .5, 2$; for κ_n, we consider $c_\kappa \kappa_{n,base}$ for $c_\kappa = .5, 2$; for α_1, we consider $\alpha_1 = .0025, .01$; for ι, which affects both the standard deviation in (4.16) and the quantile, see the paragraph containing (4.15), we alter the value of ι separately using $\iota_{sd} = 0.5, 10^{-8}$ and $\iota_q = 0.0, 0.01$; and for B, we consider $B = 500, 2000$. The changes in the tuning parameters that we consider are relatively large. In most cases, they correspond to halving or doubling the value.

We consider 40 different data generating processes (DGP's) for the MI test results: 8 DGP's have $k = 2$ moment inequalities, 16 have $k = 4$, and 16 have $k = 8$; 20 DGP's are correctly specified (for which $\Delta^{\inf} \in \{-5, -1, -.5, 0\}$) and 20 DGP's are misspecified with $\Delta^{\inf} \in \{.5, 1, 2, 5\}$; and 25 DGP's have a singleton MR-identified set and 15 DGP's have an MR-identified set with positive length). For $k = 4, 8$, we consider two different scenarios defined above: b="binding" and vs="very slack."

Tables 8.1 and 8.2 provide the results for the MI test. The nominal .05 two-sided MI test rejects $H_0 : \Delta_F^{\inf} = \Delta_0$ if $\Delta_0 \notin CI_{n,\Delta}(.05)$. Each entry in these tables shows the difference in the average rejection probabilities between a given change in a tuning parameter and the baseline value. The average is taken over 1,000 (equally spaced) values of Δ_0 such that $\Delta_0 - \Delta_F^{\inf} \in [-1, 1]$. For the tuning parameters τ_n, ι_{sd}, ι_q, and B, Tables 8.1 and 8.2 show that there is very little sensitivity of the rejection probabilities. For these tuning parameters, the differences are .00 in 261 cases and .01 in absolute value in the remaining 59 cases. For the tuning parameter κ_n, there is some sensitivity, but it is relatively small in most cases. The differences are .03 or less in absolute value in 65 of 80 cases and .06 or less in absolute value in 73 of 80 cases. Differences in rejection probabilities that

Table 8.1: Tuning parameter sensitivity: Rejection probability differences for the misspecification index test based on the two-sided CI for the misspecification index.

	c_τ=0.5	c_τ=2	c_κ=0.5	c_κ=2	ι_{sd}=0.5	$\iota_{sd} = 10^{-8}$
k=2, $\Delta_F^{\inf}$=-5	0.01	−0.01	0.04	−0.03	0.00	0.00
k=2, $\Delta_F^{\inf}$=-1	0.00	0.00	0.01	−0.01	0.00	0.00
k=2, $\Delta_F^{\inf}$=-0.5	0.00	0.00	0.01	0.00	0.00	0.00
k=2, $\Delta_F^{\inf}$=0	0.00	0.00	0.01	0.00	0.00	0.00
k=2, $\Delta_F^{\inf}$=0.5	0.00	0.00	0.01	0.00	0.00	0.00
k=2, $\Delta_F^{\inf}$=1	0.00	0.00	0.01	0.00	0.00	0.00
k=2, $\Delta_F^{\inf}$=2	0.01	0.00	0.01	0.00	0.00	0.00
k=2, $\Delta_F^{\inf}$=5	0.01	−0.01	0.02	−0.01	0.00	0.00
k=4, $\Delta_F^{\inf}$=-5, b	0.01	−0.01	**0.08**	−0.06	0.00	0.00
k=4, $\Delta_F^{\inf}$=-5, vs	0.01	−0.01	0.05	**−0.10**	0.00	0.00
k=4, $\Delta_F^{\inf}$=-1, b	0.00	0.00	0.02	−0.02	0.00	0.00
k=4, $\Delta_F^{\inf}$=-1, vs	0.00	0.00	0.01	0.00	0.00	0.00
k=4, $\Delta_F^{\inf}$=-0.5, b	0.00	0.00	0.02	−0.01	0.00	0.00
k=4, $\Delta_F^{\inf}$=-0.5, vs	0.00	0.00	0.01	0.00	0.00	0.00
k=4, $\Delta_F^{\inf}$=0, b	0.00	0.00	0.02	−0.01	0.00	0.00
k=4, $\Delta_F^{\inf}$=0, vs	0.00	0.00	0.01	0.00	0.00	0.00
k=4, $\Delta_F^{\inf}$=0.5, b	0.00	0.00	0.02	−0.01	0.00	0.00
k=4, $\Delta_F^{\inf}$=0.5, vs	0.00	0.00	0.01	0.00	0.00	0.00
k=4, $\Delta_F^{\inf}$=1, b	0.00	0.00	0.02	−0.01	0.00	0.00
k=4, $\Delta_F^{\inf}$=1, vs	0.00	0.00	0.01	0.00	0.00	0.00
k=4, $\Delta_F^{\inf}$=2, b	0.00	0.00	0.03	−0.01	0.00	0.00
k=4, $\Delta_F^{\inf}$=2, vs	0.01	0.00	0.01	0.00	0.00	0.00
k=4, $\Delta_F^{\inf}$=5, b	0.01	−0.01	0.06	−0.03	0.00	0.00
k=4, $\Delta_F^{\inf}$=5, vs	0.01	−0.01	0.03	−0.02	0.00	0.00
k=8, $\Delta_F^{\inf}$=-5, b	0.01	0.00	**0.11**	**−0.08**	0.00	0.00
k=8, $\Delta_F^{\inf}$=-5, vs	0.01	0.00	**0.09**	**−0.13**	0.00	0.00
k=8, $\Delta_F^{\inf}$=-1, b	0.00	0.00	0.04	−0.03	0.00	0.00
k=8, $\Delta_F^{\inf}$=-1, vs	0.00	0.00	0.02	−0.01	0.00	0.00
k=8, $\Delta_F^{\inf}$=-0.5, b	0.00	0.00	0.03	−0.02	0.00	0.00
k=8, $\Delta_F^{\inf}$=-0.5, vs	0.00	0.00	0.02	−0.01	0.00	0.00
k=8, $\Delta_F^{\inf}$=0, b	0.00	0.00	0.03	−0.02	0.00	0.00
k=8, $\Delta_F^{\inf}$=0, vs	0.00	0.00	0.02	−0.01	0.00	0.00
k=8, $\Delta_F^{\inf}$=0.5, b	0.00	0.00	0.03	−0.01	0.00	0.00
k=8, $\Delta_F^{\inf}$=0.5, vs	0.00	0.00	0.02	−0.01	0.00	0.00
k=8, $\Delta_F^{\inf}$=1, b	0.00	0.00	0.03	−0.02	0.00	0.00
k=8, $\Delta_F^{\inf}$=1, vs	0.00	0.00	0.02	−0.01	0.00	0.00
k=8, $\Delta_F^{\inf}$=2, b	0.00	0.00	0.04	−0.02	0.00	0.00
k=8, $\Delta_F^{\inf}$=2, vs	0.00	0.00	0.03	−0.01	0.00	0.00
k=8, $\Delta_F^{\inf}$=5, b	0.01	0.00	**0.09**	−0.04	0.00	0.00
k=8, $\Delta_F^{\inf}$=5, vs	0.01	−0.01	0.06	−0.03	0.00	0.00

Table 8.2: Tuning parameter sensitivity: Rejection probability differences for the misspecification index test based on the two-sided CI for the misspecification index (continued).

	ι_q=0	ι_q=0.01	B=500	B=2,000
k=2, $\Delta_F^{\inf}$=-5	0.00	−0.01	0.00	0.00
k=2, $\Delta_F^{\inf}$=-1	0.00	−0.01	0.00	0.00
k=2, $\Delta_F^{\inf}$=-0.5	0.00	−0.01	0.00	0.00
k=2, $\Delta_F^{\inf}$=0	0.00	−0.01	0.00	0.00
k=2, $\Delta_F^{\inf}$=0.5	0.00	−0.01	0.00	0.00
k=2, $\Delta_F^{\inf}$=1	0.00	−0.01	0.00	0.00
k=2, $\Delta_F^{\inf}$=2	0.00	−0.01	0.00	0.00
k=2, $\Delta_F^{\inf}$=5	0.00	−0.01	0.00	0.00
k=4, $\Delta_F^{\inf}$=-5, b	0.00	−0.01	0.00	0.00
k=4, $\Delta_F^{\inf}$=-5, vs	0.00	−0.01	0.00	0.00
k=4, $\Delta_F^{\inf}$=-1, b	0.00	−0.01	0.00	0.00
k=4, $\Delta_F^{\inf}$=-1, vs	0.00	−0.01	0.00	0.00
k=4, $\Delta_F^{\inf}$=-0.5, b	0.00	−0.01	0.00	0.00
k=4, $\Delta_F^{\inf}$=-0.5, vs	0.00	−0.01	0.00	0.00
k=4, $\Delta_F^{\inf}$=0, b	0.00	−0.01	0.00	0.00
k=4, $\Delta_F^{\inf}$=0, vs	0.00	−0.01	0.00	0.00
k=4, $\Delta_F^{\inf}$=0.5, b	0.00	−0.01	0.00	0.00
k=4, $\Delta_F^{\inf}$=0.5, vs	0.00	−0.01	0.00	0.00
k=4, $\Delta_F^{\inf}$=1, b	0.00	−0.01	0.00	0.00
k=4, $\Delta_F^{\inf}$=1, vs	0.00	−0.01	0.00	0.00
k=4, $\Delta_F^{\inf}$=2, b	0.00	−0.01	0.00	0.00
k=4, $\Delta_F^{\inf}$=2, vs	0.00	−0.01	0.00	0.00
k=4, $\Delta_F^{\inf}$=5, b	0.00	−0.01	0.00	0.00
k=4, $\Delta_F^{\inf}$=5, vs	0.00	−0.01	0.00	0.00
k=8, $\Delta_F^{\inf}$=-5, b	0.00	−0.01	0.00	0.00
k=8, $\Delta_F^{\inf}$=-5, vs	0.00	−0.01	0.00	0.00
k=8, $\Delta_F^{\inf}$=-1, b	0.00	−0.01	0.00	0.00
k=8, $\Delta_F^{\inf}$=-1, vs	0.00	−0.01	0.00	0.00
k=8, $\Delta_F^{\inf}$=-0.5, b	0.00	−0.01	0.00	0.00
k=8, $\Delta_F^{\inf}$=-0.5, vs	0.00	−0.01	0.00	0.00
k=8, $\Delta_F^{\inf}$=0, b	0.00	−0.01	0.00	0.00
k=8, $\Delta_F^{\inf}$=0, vs	0.00	−0.01	0.00	0.00
k=8, $\Delta_F^{\inf}$=0.5, b	0.00	−0.01	0.00	0.00
k=8, $\Delta_F^{\inf}$=0.5, vs	0.00	−0.01	0.00	0.00
k=8, $\Delta_F^{\inf}$=1, b	0.00	−0.01	0.00	0.00
k=8, $\Delta_F^{\inf}$=1, vs	0.00	−0.01	0.00	0.00
k=8, $\Delta_F^{\inf}$=2, b	0.00	−0.01	0.00	0.00
k=8, $\Delta_F^{\inf}$=2, vs	0.00	−0.01	0.00	0.00
k=8, $\Delta_F^{\inf}$=5, b	0.00	−0.01	0.00	0.00
k=8, $\Delta_F^{\inf}$=5, vs	0.00	−0.01	0.00	0.00

are .007 or greater in absolute value are highlighted in boldface. There are 7 out of 80 cases for $c_\kappa = .5, 2$ in boldface.

Next, we assess the sensitivity to the tuning parameters of the rejection probabilities of the SPUR2 test under the null and alternative hypotheses. We consider 33 different DGP's for the SPUR2 test results: 8 DGP's have $k = 2$ moment inequalities and 25 have $k = 4$; 19 DGP's are correctly specified (for which $r := r^{\inf} = 0$) and 14 DGP's are misspecified with $r = r^{\inf} \in \{.5, 1, 2\}$; and 20 DGP's have a point-identified set (i.e., length=0) and 13 DGP's have a non-degenerate identified set (i.e., length $\in \{.5, 1, 2\}$). For $k = 4$, we consider the five different scenarios defined above: b="binding," ab="almost binding," ss="somewhat slack," vs="very slack," and sab="slack/almost binding.

In Tables 8.3-8.6, we report the differences between the rejection probability of the SPUR2 test based on the altered tuning parameter value and the rejection probability based on the baseline tuning parameter value for each DGP. The null rejection probabilities considered are those for the case where the null value $\theta_0 = 0$ is on the boundary of the MR-identified set. For rejection probabilities under the alternative hypothesis, we take the true value of θ to be 0 and report averages of the rejection probabilities over null θ_0 values in the interval $[0, .3]$, which corresponds to the relevant range of the rejection probabilities.

Tables 8.3 and 8.4 report null rejection probability differences. They show that there is very little sensitivity of the null rejection probabilities to the changes in τ_n, α_1, ι_{sd}, ι_q, and B. All values are .005 or less in absolute value, and all but 24 out of 330 are .002 or less in absolute value. There is more sensitivity in Tables 8.3 and 8.4 to κ_n than the other tuning parameters. There are 47 out of 66 cases that are .005 or less in absolute value. Differences in rejection probabilities that are .010 or greater in absolute value are highlighted in boldface. There are 10 out of 66 cases for $c_\kappa = .5, 2$ in boldface. Eight of these correspond to $c_\tau = .5$. In consequence, we recommend not using a κ_n value as small as $.5(\ln n)^{1/2}$. For $c_\kappa = 2$, only 5 out of 33 cases exceed .005 in absolute value, with a maximum of .015. So, the null rejection probabilities are not very sensitive to a doubling of the value of κ_n.

The results for the average rejection probabilities under the alternative hypothesis are reported in Tables 8.5 and 8.6. For the tuning parameters τ_n, α_1, ι_{sd}, ι_q, and B, the results are similar to those in Tables 8.3 and 8.4. That is, the average rejection probability differences are small. All are .02 or less in absolute value, and all but 10 out of 330 are .00 or .01 in absolute value. For κ_n, the average differences in some cases are substantially larger. Differences that exceed .05 are highlighted in boldface. There are 39 out of 66 values for $c_\kappa = .5, 2$ in boldface. The higher level of sensitivity to κ_n occurs in the DGP's that have point-identified sets. Decreasing κ_n increases

Table 8.3: Tuning parameter sensitivity: SPUR2 null rejection probability comparisons with baseline.

DGP	c_τ=0.5	c_τ=2	c_κ=0.5	c_κ=2	α_2=0.0475	α_2=0.04
k=2, r=0, length=0	0.000	0.000	0.001	-0.001	0.000	0.000
k=2, r=0, length=0.5	0.005	0.000	**0.011**	-0.002	-0.002	0.001
k=2, r=0, length=1	0.000	0.000	0.000	0.000	0.001	-0.004
k=2, r=0, length=2	0.000	0.000	0.000	0.000	0.001	-0.004
k=2, r=0.5, length=0	0.000	0.000	0.002	-0.001	0.000	0.000
k=2, r=1, length=0	0.000	0.000	0.007	0.000	0.000	0.000
k=2, r=2, length=0	0.000	0.000	0.005	0.000	0.000	0.000
k=2, r=5, length=0	0.000	0.000	0.006	-0.001	0.000	-0.001
k=4, r=0, length=0, b	0.000	0.000	0.000	0.000	0.000	0.000
k=4, r=0, length=0, ab	0.000	0.000	0.000	0.000	0.000	0.000
k=4, r=0, length=0, ss	0.000	0.000	0.001	-0.001	0.000	0.000
k=4, r=0, length=0, vs	0.000	0.000	0.001	-0.001	0.000	0.000
k=4, r=0, length=0, sab	0.000	0.000	0.000	-0.001	0.000	0.000
k=4, r=0, length=0.5, b	0.001	0.000	**0.023**	**-0.015**	-0.005	0.002
k=4, r=0, length=0.5, ab	0.000	0.000	**0.022**	**-0.011**	0.000	-0.002
k=4, r=0, length=0.5, ss	0.000	0.000	**0.014**	-0.005	-0.003	0.004
k=4, r=0, length=0.5, vs	0.000	0.000	**0.014**	-0.004	-0.003	0.004
k=4, r=0, length=0.5, sab	0.000	0.000	**0.021**	-0.009	0.002	-0.001
k=4, r=0, length=1, b	0.000	0.000	0.002	0.000	0.003	-0.002
k=4, r=0, length=1, ab	0.000	0.000	**0.015**	-0.007	0.003	-0.003
k=4, r=0, length=1, ss	0.000	0.000	0.000	0.000	0.001	-0.005
k=4, r=0, length=1, vs	0.000	0.000	0.000	0.000	0.001	-0.005
k=4, r=0, length=1, sab	0.000	0.000	**0.015**	-0.007	0.003	-0.003
k=4, r=0.5, length=0, b	0.000	0.000	0.000	0.000	0.000	0.000
k=4, r=0.5, length=0, ab	0.000	0.000	0.002	0.000	0.000	0.000
k=4, r=0.5, length=0, ss	0.000	0.000	0.008	0.000	0.000	0.000
k=4, r=0.5, length=0, vs	0.000	0.000	0.008	0.000	0.000	0.000
k=4, r=0.5, length=0, sab	0.000	0.000	0.005	0.000	0.000	0.000
k=4, r=1, length=0, b	0.000	0.000	0.000	0.000	0.000	0.000
k=4, r=1, length=0, ab	0.000	0.000	0.001	0.000	0.000	0.000
k=4, r=1, length=0, ss	0.000	0.000	0.004	-0.001	0.000	-0.001
k=4, r=1, length=0, vs	0.000	0.000	0.004	-0.001	0.000	-0.001
k=4, r=1, length=0, sab	0.000	0.000	0.002	-0.001	0.000	-0.001

power (and the null rejection probabilities), whereas increasing κ_n decreases power (and the null rejection probabilities). The recommended value of κ_n aims to achieve high power subject to the null rejection probability being less than or equal to α.

Table 8.4: Tuning parameter sensitivity: SPUR2 null rejection probability comparisons with baseline (continued).

DGP	ι_{sd}=0.5	ι_{sd}=10^{-8}	ι_q=0	ι_q=0.01	B=500	B=2,000
k=2, r=0, length=0	0.000	0.000	0.000	0.000	0.000	0.000
k=2, r=0, length=0.5	-0.001	0.000	0.000	-0.002	-0.001	-0.002
k=2, r=0, length=1	0.000	0.000	0.000	0.000	-0.002	-0.001
k=2, r=0, length=2	0.000	0.000	0.000	0.000	-0.002	-0.001
k=2, r=0.5, length=0	0.000	0.000	0.000	0.000	0.000	0.000
k=2, r=1, length=0	0.000	0.000	0.000	0.000	0.000	0.000
k=2, r=2, length=0	0.000	0.000	0.000	0.000	0.001	0.001
k=2, r=5, length=0	0.000	0.000	0.000	0.000	0.001	0.000
k=4, r=0, length=0, b	0.000	0.000	0.000	0.000	0.000	0.000
k=4, r=0, length=0, ab	0.000	0.000	0.000	0.000	0.000	0.000
k=4, r=0, length=0, ss	0.000	0.000	0.000	0.000	0.000	0.000
k=4, r=0, length=0, vs	0.000	0.000	0.000	0.000	0.000	0.000
k=4, r=0, length=0, sab	0.000	0.000	0.000	0.000	0.000	0.000
k=4, r=0, length=0.5, b	-0.001	0.000	0.000	-0.005	0.000	-0.003
k=4, r=0, length=0.5, ab	-0.001	0.000	0.000	-0.001	0.000	-0.001
k=4, r=0, length=0.5, ss	-0.001	0.000	0.000	-0.004	0.002	-0.003
k=4, r=0, length=0.5, vs	-0.001	0.000	0.000	-0.004	0.002	-0.003
k=4, r=0, length=0.5, sab	-0.001	0.000	0.000	-0.001	0.000	-0.001
k=4, r=0, length=1, b	0.000	0.000	0.000	0.000	-0.003	0.001
k=4, r=0, length=1, ab	0.000	0.000	0.000	0.000	0.001	0.000
k=4, r=0, length=1, ss	0.000	0.000	0.000	0.000	0.002	-0.003
k=4, r=0, length=1, vs	0.000	0.000	0.000	0.000	0.002	-0.003
k=4, r=0, length=1, sab	0.000	0.000	0.000	0.000	0.001	0.000
k=4, r=0.5, length=0, b	0.000	0.000	0.000	0.000	0.000	0.000
k=4, r=0.5, length=0, ab	0.000	0.000	0.000	0.000	0.000	0.000
k=4, r=0.5, length=0, ss	0.000	0.000	0.000	0.000	0.001	0.001
k=4, r=0.5, length=0, vs	0.000	0.000	0.000	0.000	0.001	0.001
k=4, r=0.5, length=0, sab	0.000	0.000	0.000	0.000	0.001	0.001
k=4, r=1, length=0, b	0.000	0.000	0.000	0.000	0.000	0.000
k=4, r=1, length=0, ab	0.000	0.000	0.000	0.000	0.000	0.000
k=4, r=1, length=0, ss	0.000	0.000	0.000	0.000	0.000	0.000
k=4, r=1, length=0, vs	0.000	0.000	0.000	0.000	0.000	0.000
k=4, r=1, length=0, sab	0.000	0.000	0.000	0.000	0.000	0.000

Table 8.5: Tuning parameter sensitivity: SPUR2 average power comparisons with baseline.

DGP	c_τ=0.5	c_τ=2	c_κ=0.5	c_κ=2	α_2=0.0475	α_2=0.04
k=2, r=0, length=0	0.00	0.01	**0.07**	**-0.09**	0.01	-0.02
k=2, r=0, length=0.5	0.00	0.00	0.01	-0.01	0.00	-0.01
k=2, r=0, length=1	0.00	0.00	0.00	0.00	0.01	-0.01
k=2, r=0, length=2	0.00	0.00	0.00	0.00	0.01	-0.01
k=2, r=0.5, length=0	0.00	0.00	**0.12**	**-0.09**	0.01	-0.02
k=2, r=1, length=0	0.00	0.00	**0.13**	**-0.09**	0.01	-0.02
k=2, r=2, length=0	0.00	0.00	**0.13**	**-0.07**	0.01	-0.02
k=2, r=5, length=0	0.00	0.00	0.03	-0.01	0.00	0.00
k=4, r=0, length=0, b	0.00	0.01	**0.15**	**-0.07**	0.00	-0.01
k=4, r=0, length=0, ab	0.00	0.01	**0.15**	**-0.09**	0.01	-0.01
k=4, r=0, length=0, ss	0.00	0.01	**0.07**	**-0.11**	0.01	-0.01
k=4, r=0, length=0, vs	0.00	0.01	**0.07**	**-0.09**	0.01	-0.01
k=4, r=0, length=0, sab	0.00	0.01	**0.14**	**-0.07**	0.01	-0.01
k=4, r=0, length=0.5, b	0.00	0.00	0.03	**-0.06**	-0.01	0.00
k=4, r=0, length=0.5, ab	0.00	0.00	0.02	-0.04	0.00	0.00
k=4, r=0, length=0.5, ss	0.00	0.00	0.01	-0.02	0.00	-0.01
k=4, r=0, length=0.5, vs	0.00	0.00	0.01	-0.01	0.00	-0.01
k=4, r=0, length=0.5, sab	0.00	0.00	0.02	-0.02	0.00	-0.01
k=4, r=0, length=1, b	0.00	0.00	0.00	0.00	0.00	-0.01
k=4, r=0, length=1, ab	0.00	0.00	0.01	0.00	0.00	-0.01
k=4, r=0, length=1, ss	0.00	0.00	0.00	-0.01	0.01	-0.01
k=4, r=0, length=1, vs	0.00	0.00	0.00	0.00	0.01	-0.01
k=4, r=0, length=1, sab	0.00	0.00	0.01	0.00	0.00	-0.01
k=4, r=0.5, length=0, b	0.00	0.00	**0.18**	**-0.07**	0.01	-0.01
k=4, r=0.5, length=0, ab	0.00	0.01	**0.20**	**-0.08**	0.01	-0.01
k=4, r=0.5, length=0, ss	0.00	0.00	**0.11**	**-0.14**	0.01	-0.02
k=4, r=0.5, length=0, vs	0.00	0.00	**0.11**	**-0.09**	0.01	-0.02
k=4, r=0.5, length=0, sab	0.00	0.02	**0.14**	**-0.14**	0.01	-0.01
k=4, r=1, length=0, b	0.00	0.00	**0.21**	**-0.07**	0.01	-0.01
k=4, r=1, length=0, ab	0.00	0.01	**0.22**	**-0.08**	0.01	-0.01
k=4, r=1, length=0, ss	0.00	0.00	**0.13**	**-0.18**	0.01	-0.02
k=4, r=1, length=0, vs	0.00	0.00	**0.13**	**-0.09**	0.01	-0.02
k=4, r=1, length=0, sab	0.00	0.01	**0.16**	**-0.14**	0.01	-0.02

Table 8.6: Tuning parameter sensitivity: SPUR2 average power comparisons with baseline (continued).

DGP	ι_{sd}=0.5	ι_{sd}=10^{-8}	ι_q=0	ι_q=0.01	B=500	B=2,000
k=2, r=0, length=0	-0.01	0.00	0.00	0.00	0.00	0.00
k=2, r=0, length=0.5	0.00	0.00	0.00	0.00	0.00	0.00
k=2, r=0, length=1	0.00	0.00	0.00	0.00	0.00	0.00
k=2, r=0, length=2	0.00	0.00	0.00	0.00	0.00	0.00
k=2, r=0.5, length=0	0.00	0.00	0.00	0.00	0.00	0.00
k=2, r=1, length=0	0.00	0.00	0.00	0.00	0.00	0.00
k=2, r=2, length=0	0.00	0.00	0.00	0.00	0.00	0.00
k=2, r=5, length=0	0.00	0.00	0.00	0.00	0.00	0.00
k=4, r=0, length=0, b	0.00	0.00	0.00	0.00	0.00	0.00
k=4, r=0, length=0, ab	-0.01	0.00	0.00	0.00	0.00	0.00
k=4, r=0, length=0, ss	-0.01	0.00	0.00	0.00	0.00	0.00
k=4, r=0, length=0, vs	-0.01	0.00	0.00	0.00	0.00	0.00
k=4, r=0, length=0, sab	-0.01	0.00	0.00	0.00	0.00	0.00
k=4, r=0, length=0.5, b	0.00	0.00	0.00	-0.01	0.00	0.00
k=4, r=0, length=0.5, ab	0.00	0.00	0.00	0.00	0.00	0.00
k=4, r=0, length=0.5, ss	0.00	0.00	0.00	0.00	0.00	0.00
k=4, r=0, length=0.5, vs	0.00	0.00	0.00	0.00	0.00	0.00
k=4, r=0, length=0.5, sab	0.00	0.00	0.00	-0.01	0.00	0.00
k=4, r=0, length=1, b	0.00	0.00	0.00	0.00	0.00	0.00
k=4, r=0, length=1, ab	0.00	0.00	0.00	0.00	0.00	0.00
k=4, r=0, length=1, ss	0.00	0.00	0.00	0.00	0.00	0.00
k=4, r=0, length=1, vs	0.00	0.00	0.00	0.00	0.00	0.00
k=4, r=0, length=1, sab	0.00	0.00	0.00	0.00	0.00	0.00
k=4, r=0.5, length=0, b	0.00	0.00	0.00	0.00	0.00	0.00
k=4, r=0.5, length=0, ab	0.00	0.00	0.00	0.00	0.00	0.00
k=4, r=0.5, length=0, ss	0.00	0.00	0.00	0.00	0.00	0.00
k=4, r=0.5, length=0, vs	0.00	0.00	0.00	0.00	0.00	0.00
k=4, r=0.5, length=0, sab	0.00	0.00	0.00	0.00	0.00	0.00
k=4, r=1, length=0, b	0.00	0.00	0.00	0.00	0.00	0.00
k=4, r=1, length=0, ab	0.00	0.00	0.00	0.00	0.00	0.00
k=4, r=1, length=0, ss	0.00	0.00	0.00	0.00	0.00	0.00
k=4, r=1, length=0, vs	0.00	0.00	0.00	0.00	0.00	0.00
k=4, r=1, length=0, sab	0.00	0.00	0.00	0.00	0.00	0.00

8.2 Missing Data Model

In this subsection, we revisit the missing data model that BCS use in their simulations. The specification of the model closely follows BCS, but we consider a somewhat different data generating process.[32] Example 2.1 of BCS provides motivation for the model. Let $\{W_i = (Y_iZ_i, Z_i, X_i)\}_{i\leq n}$ be the i.i.d data. Here, $Z_i \sim Bernoulli(p_z)$ is the indicator of whether the outcome variable Y_i is missing. It is independent of $(Y_i, X_i)'$. The conditional distribution of Y_i given X_i is

$$Y_i|X_i = x_1 \sim N(0,1),\ Y_i|X_i = x_2 \sim N((1+\widetilde{r})/p_z, 1), \text{ and } Y_i|X_i = x_3 \sim N(0,1), \tag{8.4}$$

with $P(X_i = x_1) = P(X_i = x_2) = P(X_i = x_3) = 1/3$. The parameter space is $\Theta = [-20, 20] \times [-20, 20]$. The moment functions are

$$\begin{aligned} m_1(W_i,\theta) &= (\theta_1 - Y_iZ_i)1\{X_i = x_1\}, \\ m_2(W_i,\theta) &= (1 - \theta_1 - Y_iZ_i)1\{X_i = x_2\}, \text{ and} \\ m_3(W_i,\theta) &= (\theta_2 - Y_iZ_i)1\{X_i = x_3\} \text{ for } \theta = (\theta_1,\theta_2)'. \end{aligned} \tag{8.5}$$

The value of $\widetilde{r}$ determines whether the model is misspecified. When $\widetilde{r} \leq 0$, the model is correctly specified, which implies that $r_F^{\inf} = 0$, and the MR-identified set is $\Theta_I^{MR}(F) = [0, -\widetilde{r}] \times [0, \infty)$. When $\widetilde{r} > 0$, the model is misspecified and some calculations show that

$$r_F^{\inf} = \left(\frac{\widetilde{r}^2/3}{\left(p_z^{1/2} + ((1+\widetilde{r})^2(1/p_z - 1) + p_z)^{1/2}\right)^2 + 2\widetilde{r}^2/3} \right)^{1/2}. \tag{8.6}$$

For $\widetilde{r} > 0$, it can be shown that the MR-identified set is $\Theta_I^{MR}(F) = \{\theta_1^I(\widetilde{r})\} \times [\theta_1^I(\widetilde{r}), \infty)$, where

$$\theta_1^I(\widetilde{r}) := -\frac{p_z^{1/2}\widetilde{r}}{p_z^{1/2} + ((1+\widetilde{r})^2(1/p_z - 1) + p_z)^{1/2}}. \tag{8.7}$$

See Section 9 below for the derivations of (8.6) and (8.7).

We take $p_z = .8$ throughout. We consider values of $\widetilde{r}$ that cover both misspecified and correctly-specified cases. As above, we simulate rejection probabilities for a fixed data generating process and a range of null hypothesis values $\theta_0 = (\theta_{01}, \theta_{02})'$, where $H_0 : \theta = \theta_0$. For the null values, we consider θ_{02} fixed at $\theta_1^I(\widetilde{r})$ when $\widetilde{r} > 0$ and at 0 when $\widetilde{r} \leq 0$, and we consider a range of θ_{01} values. Accordingly, the x–axes in Figures 8.10 and 8.11 correspond to the first element of the null vector.

[32] A different data generating process is employed to ensure that the random variable YZ is nonnegative, which is an implication of the structure of the missing data model.

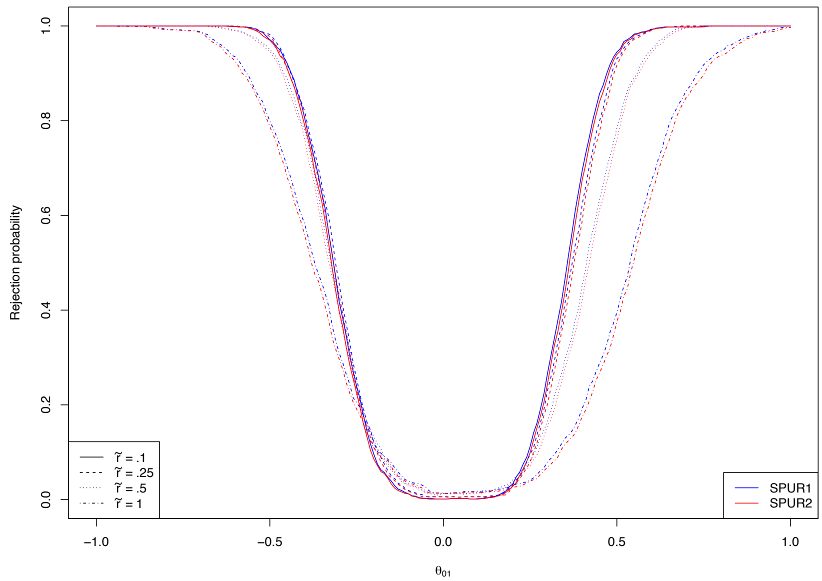

Figure 8.10: Rejection probabilities of tests concerning θ under misspecification for the missing data model. The figure shows the rejection probabilities the nominal .05 SPUR1 and SPUR2 tests for the null hypothesis $H_0 : \theta = \theta_0$ for a range of θ_{01} values and a fixed identified set, for four different $\widetilde{r}$ values.

Figure 8.10 reports the simulated rejection probabilities for the misspecified cases with $\widetilde{r} = .1$, .2, .5, and 1.[33] Here, the MR-identified set is $\{0\} \times [0, \infty)$. As in the lower/upper bound model, the SPUR1 and SPUR2 tests perform quite similarly, as expected. Also, the rejection probabilities increase to 1 fairly quickly as the distance between the null value and the MR-identified set increases, and the performance is better for smaller values of $\widetilde{r}$ (or, equivalently, smaller values of $r_F^{\inf}$).

Figure 8.11 provides the results under correct specification. Here, we see that when $\widetilde{r} = 0$, which implies that the identified set contains no slack points, the standard GMS test performs better than the SPUR1 and SPUR2 tests, which is expected. In this case, the SPUR1 and SPUR2 tests have almost identical rejection probabilities. Also, the difference between the standard GMS test and the SPUR2 test decreases quickly as the identified set gets larger (i.e., as $\widetilde{r}$ become more negative) and, hence, contains more slack points. The SPUR2 test is essentially on par with the standard GMS test when $\widetilde{r}$ is -1. The difference in power between the standard GMS test and the SPUR1 test also decreases to some extent as the identified set get larger. But, the SPUR1 test has lower power (similar to the $\widetilde{r} = -1$ case) even for $\widetilde{r}$ values in the range of $[-2, -5]$ (based on results not reported in Figure 8.11). Overall, the four plots show how the SPUR2 test adapts, and eventually behaves very much like the standard GMS test as the identified set gets larger.

[33] By (8.6), these $\widetilde{r}$ values correspond (approximately) to $r_F^{\inf} = .03$, .07, .14, and .24, respectively.

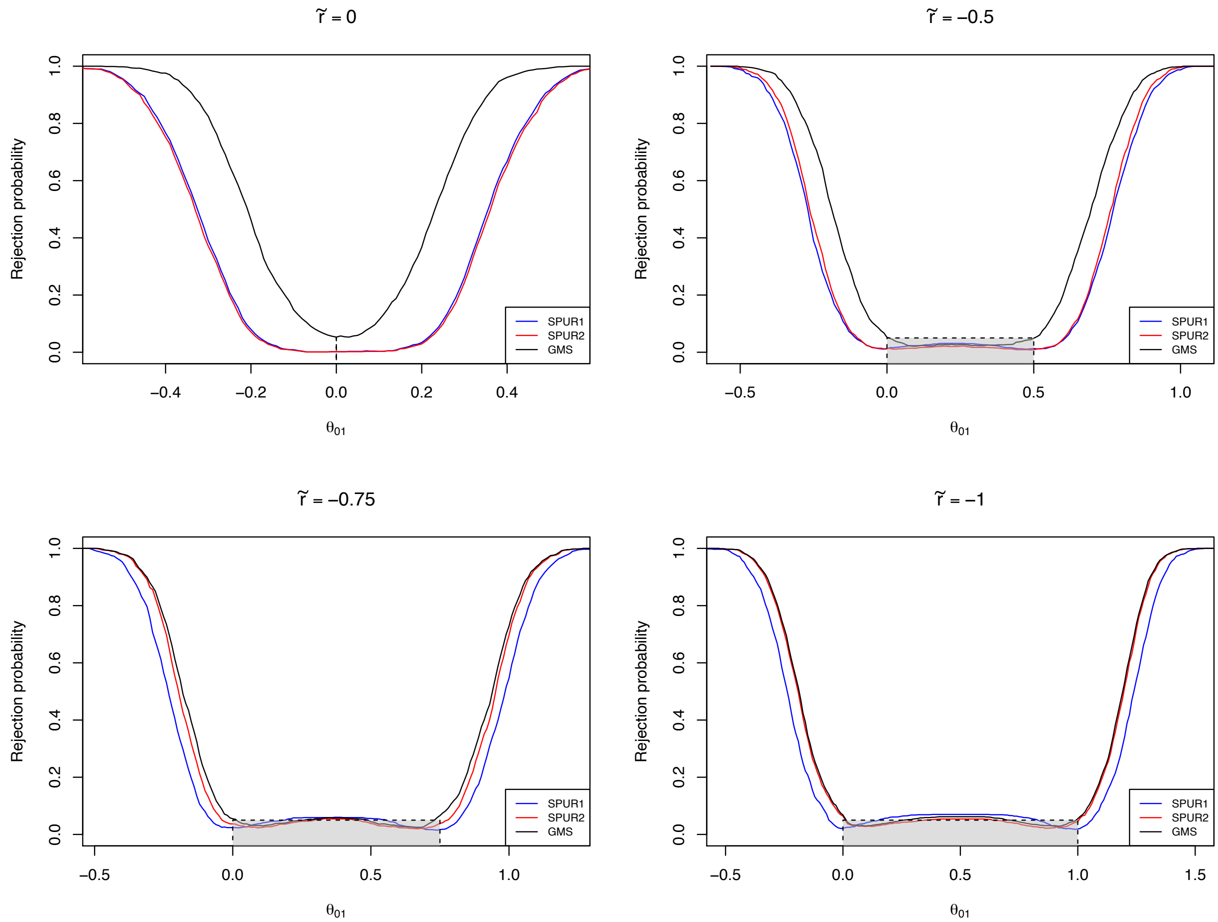

Figure 8.11: Rejection probabilities of tests concerning θ under correct specification for the missing data model. Each plot shows the rejection probabilities the nominal .05 SPUR1, SPUR2, and standard GMS tests for the null hypothesis $H_0 : \theta = \theta_0$ and a range of θ_{01} values, for one of the four $\widetilde{r}$ values considered. The shaded region in each plot delineates the identified set.

9 Details for the Missing Data Model

In this section, we provide additional details for the missing data model considered in Section 8.2. Specifically, we provide derivations for (8.6), (8.7), and the line following (8.7), which gives an expression for the MR-identified set.

Let $p_j := P(X_i = x_j) > 0$ for $j \leq 3$. In the simulations, we take $p_j = 1/3$ for $j \leq 3$. Some calculations give

$$\begin{aligned} E_F m_1(W, \theta) &= p_1 \theta_1, \\ E_F m_2(W, \theta) &= -p_2(\theta_1 + \widetilde{r}), \text{ and} \\ E_F m_3(W, \theta) &= p_3 \theta_2. \end{aligned} \tag{9.1}$$

In consequence, the model is misspecified if and only if $\widetilde{r} > 0$, as stated in Section 8. If $\widetilde{r} \leq 0$, $r_F^{\inf} = 0$.

Now, suppose $\widetilde{r} > 0$. Additional calculations give

$$\begin{aligned} Var_F\left(m_1(W,\theta)\right) &= (p_1 - p_1^2)\theta_1^2 + p_1 p_z, \\ Var_F\left(m_2(W,\theta)\right) &= (p_2 - p_2^2)(\theta_1 + \widetilde{r})^2 + p_2\left((1+\widetilde{r})^2(1/p_z - 1) + p_z\right), \text{ and} \\ Var_F\left(m_3(W,\theta)\right) &= (p_3 - p_3^2)\theta_2^2 + p_3 p_z. \end{aligned} \tag{9.2}$$

We relax the (standardized) inequalities by r. Then, by (9.1) and (9.2), the inequalities are

$$\begin{aligned} \frac{p_1\theta_1}{((p_1 - p_1^2)\theta_1^2 + p_1 p_z)^{1/2}} &\geq -r, \\ -\frac{p_2(\theta_1 + \widetilde{r})}{((p_2 - p_2^2)(\theta_1+\widetilde{r})^2 + p_2((1+\widetilde{r})^2(1/p_z - 1) + p_z))^{1/2}} &\geq -r, \text{ and} \\ \frac{p_3\theta_2}{((p_3 - p_3^2)\theta_2^2 + p_3 p_z)^{1/2}} &\geq -r. \end{aligned} \tag{9.3}$$

By definition, $r_F^{\inf}$ is the smallest $r > 0$ such that there exists some $\theta \in \Theta$ that satisfies (9.3). The third inequality does not play a role in determining $r_F^{\inf}$. Hence, we focus on finding the smallest $r > 0$ such that there exists some θ_1 that satisfies the first two inequalities.

For arbitrary numbers a, b, and c with $a > 0$ and $b > 0$, consider the function

$$h(\theta_1) = \frac{\theta_1 + c}{(a(\theta_1 + c)^2 + b)^{1/2}}. \tag{9.4}$$

Calculation of the first derivative of $h(\cdot)$ shows that $h(\cdot)$ is strictly increasing. This implies that the left-hand sides of the first and second inequalities in (9.3) are strictly increasing and strictly decreasing functions of θ_1, respectively. Hence, if we let $\underline{\theta}_1(r)$ and $\overline{\theta}_1(r)$ denote the θ_1 values that solve the first and second inequalities as equalities, respectively, then θ_1 satisfies the two inequalities if and only if θ_1 lies in $[\underline{\theta}_1(r), \overline{\theta}_1(r)]$, where this interval is defined to be empty if $\underline{\theta}_1(r) > \overline{\theta}_1(r)$.

Some algebra gives

$$\begin{aligned} \underline{\theta}_1(r) &= -\left(\frac{p_z}{(p_1/r^2 + p_1 - 1))}\right)^{1/2} \text{ and} \\ \overline{\theta}_1(r) &= \left(\frac{(1+\widetilde{r})^2(1/p_z - 1) + p_z}{p_2/r^2 + p_2 - 1)}\right)^{1/2} - \widetilde{r}. \end{aligned} \tag{9.5}$$

Hence, if r is such that

$$\widetilde{r} \le \left(\frac{(1+\widetilde{r})^2(1/p_z - 1) + p_z}{p_2/r^2 + p_2 - 1} \right)^{1/2} + \left(\frac{p_z}{p_1/r^2 + p_1 - 1)} \right)^{1/2}, \tag{9.6}$$

then the MR-identified set under the relaxation r is non-empty. Since the right-hand side is increasing in r, $r_F^{\inf}$ must solve (9.6) as an equality. That is, $r_F^{\inf}$ is the value of r that makes $\underline{\theta}_1(r) = \overline{\theta}_1(r)$. Assuming $p_1 = p_2$, this gives

$$r_F^{\inf} = \left(\frac{p_1 \widetilde{r}^2}{\left(p_z^{1/2} + ((1+\widetilde{r})^2(1/p_z - 1) + p_z)^{1/2}\right)^2 + (1 - p_1)\widetilde{r}^2} \right)^{1/2}. \tag{9.7}$$

Taking $p_1 = p_2 = 1/3$ gives (8.6).

Plugging the expression for $r_F^{\inf}$ in place of r in (9.5) gives

$$\underline{\theta}_1(r_F^{\inf}) = \overline{\theta}_1(r_F^{\inf}) = -\frac{p_z^{1/2}\widetilde{r}}{p_z^{1/2} + ((1+\widetilde{r})^2(1/p_z - 1) + p_z)^{1/2}} =: \theta_1^I(\widetilde{r}). \tag{9.8}$$

Thus, the only θ_1 value that satisfies (9.3) with $r = r_F^{\inf}$ is $\theta_1 = \theta_1^I(\widetilde{r})$. This gives (8.7).

Now, plugging in $r_F^{\inf}$ in place of r in the third inequality of (9.3) and taking $p_1 = p_2 = p_3 = 1/3$, one can see that any θ_2 such that $\theta_2 \ge \theta_1^I(\widetilde{r})$ satisfies (9.3) (with $r_F^{\inf}$ in place of r). This shows that $\Theta_I^{MR}(F) = \{\theta_1^I(\widetilde{r})\} \times [\theta_1^I(\widetilde{r}), \infty)$.

10 Empirical Illustration

10.1 Sensitivity to Tuning Parameters

The baseline and altered values of the tuning parameters are the same as in Section 8.1.3. Table 10.1 reports the differences between the MI CI lower and upper bounds when computed with altered values of the tuning parameters compared to the baseline tuning parameters, where the tuning parameters are altered one at a time. The results in Table 10.1 are for $\alpha = .025$. The corresponding results for $\alpha = .05$ are identical except that for $\Delta_{n,L}^{\inf}(.05)$ and $B = 2000$ the difference is .000, rather than $-.001$. Table 10.1 shows very little sensitivity to the tuning parameters τ_n, i_{sd}, ι_q, and B. There is very little sensitivity of the MI CI lower bound to κ_n. There is some sensitivity of the MI CI upper bound to κ_n. But the magnitudes are only $-.014$ and .013 for $c_\kappa = .5$ and 2, respectively, which is fairly small.

Table 10.2 reports the differences between the SPUR2 CI lower bounds when computed with altered values of the tuning parameters compared to the baseline tuning parameters, where the tuning parameters are altered one at a time. Table 10.3 provides analogous results for the SPUR2 CI upper bounds.

Tables 10.2 and 10.3 show relatively low sensitivity in general to the tuning parameters τ_n, α_1, ι_{sd}, ι_q, and B. There is more sensitivity to κ_n. Halving or doubling κ_n alters the SPUR2 CI lower and upper bounds by an amount typically in the range of 0.000 to 0.350 in absolute value.

Table 10.1: Tuning parameter sensitivity: Misspecification index CI lower and upper bound comparisons with baseline.

	c_τ=0.5	c_τ=2	c_κ=0.5	c_κ=2	$\iota_{sd}=10^{-8}$	ι_{sd}=0.5
$\widehat{\Delta}^{\inf}_{n,L}(.025)$	0.000	0.000	0.000	−0.001	0.000	0.000
$\widehat{\Delta}^{\inf}_{n,U}(.025)$	−0.001	0.001	−0.014	0.013	0.000	0.000

	ι_q=0	ι_q=0.01	B=500	B=2,000
$\widehat{\Delta}^{\inf}_{n,L}(.025)$	0.000	0.000	0.000	−0.001
$\widehat{\Delta}^{\inf}_{n,U}(.025)$	0.000	0.000	0.000	0.000

10.2 Power Results for a Simplified Entry Game Model

In this section, we report power results for MI and SPUR2 tests based on a simplified version of the entry game model considered in Section 6. The goal is to provide some numerical evidence that the SPUR2 test does not suffer from severe power issues, in a setting designed to mimic the empirical illustration. The model employed is the same as in Section 6, but it is simplified by assuming there is only one observed (market-level, binary) covariate $X_{i,t} = X_i^{\text{size}}$, which we refer to as “size,” and by assuming that the unobserved shocks are uncorrelated across types so that $\rho = 0$.

As in Section 6, the moment inequalities are

$$\begin{aligned}
&E[1(Y_i = (0,0)', X_i = x) - P_{00}(x,\theta)p_x] \geq 0,\\
&E[P_{00}(x,\theta)p_x - 1(Y_i = (0,0)', X_i = x)] \geq 0,\\
&E[1(Y_i = (0,1)', X_i = x) - \underline{P}_{01}(x,\theta)p_x] \geq 0,\\
&E[\overline{P}_{01}(x,\theta)p_x - 1(Y_i = (0,1)', X_i = x)] \geq 0,\\
&E[1(Y_i = (1,1)', X_i = x) - P_{11}(x,\theta)p_x] \geq 0,\\
&E[P_{11}(x,\theta)p_x - 1(Y_i = (1,1)', X_i = x)] \geq 0, \qquad (10.1)
\end{aligned}$$

but now with just $x \in \{0,1\}$, which results in 12 moment inequalities. Here, we define the quantities

Table 10.2: Tuning parameter sensitivity: SPUR2 CI lower bound comparisons with baseline.

	c_τ=0.5	c_τ=2	c_κ=0.5	c_κ=2	α_2=0.0475	α_2=0.04
$\beta_{\text{LCC}}^{\text{const}}$	0.000	0.000	0.000	0.000	0.000	0.000
$\beta_{\text{LCC}}^{\text{size}}$	−0.002	0.000	0.176	−0.216	0.027	−0.036
$\beta_{\text{LCC}}^{\text{pres}}$	0.070	0.000	0.327	−0.285	0.087	−0.019
γ_{LCC}	0.004	−0.020	0.207	−0.293	0.031	0.000
$\beta_{\text{OA}}^{\text{const}}$	0.018	0.000	0.088	−0.115	0.000	−0.024
$\beta_{\text{OA}}^{\text{size}}$	0.011	−0.009	0.280	−0.210	0.086	−0.008
$\beta_{\text{OA}}^{\text{pres}}$	0.034	0.000	0.150	−0.167	0.049	0.000
γ_{OA}	0.058	0.000	0.268	−0.262	0.116	−0.014
ρ	0.000	0.000	0.000	0.000	0.000	0.000

	ι_{sd}=10^{-8}	ι_{sd}=0.5	ι_q=0	ι_q=0.01	B=500	B=2,000
$\beta_{\text{LCC}}^{\text{const}}$	0.000	0.000	0.000	0.000	0.000	0.000
$\beta_{\text{LCC}}^{\text{size}}$	0.013	0.000	0.023	0.000	−0.055	0.010
$\beta_{\text{LCC}}^{\text{pres}}$	0.000	0.000	0.000	−0.005	−0.084	0.063
γ_{LCC}	0.009	−0.016	0.000	0.000	0.017	−0.031
$\beta_{\text{OA}}^{\text{const}}$	0.019	0.000	0.001	0.000	−0.029	0.024
$\beta_{\text{OA}}^{\text{size}}$	0.016	−0.003	0.066	−0.009	−0.006	0.029
$\beta_{\text{OA}}^{\text{pres}}$	0.000	−0.009	0.058	0.000	0.043	0.025
γ_{OA}	0.000	0.000	0.000	0.000	0.041	0.111
ρ	0.000	0.000	0.000	0.000	0.000	0.000

$P_{00}(x,\theta)$, $\underline{P}_{01}(x,\theta)$, etc. as in Section 6, but with the simplified version of x.

We consider the case where there is no intercept term in X_i, which yields $\beta_{LCC} = \beta_{LCC}^{size}$ and $\beta_{OA} = \beta_{OA}^{size}$, and we suppose that $\gamma_{LCC} = \gamma_{OA} = \gamma$. The parameter spaces for β_{LCC}, β_{OA}, and γ are $[-5,5]$, $[-5,5]$, and $[0,4]$, respectively, and thus $\Theta = [-5,5]\times[-5,5]\times[0,4]$.

For this model, we consider two different data generating processes (DGP's), i.e., two different joint distributions of Y_i and X_i. The observed covariates are drawn from a Bernoulli distribution with probability .6 (i.e., $p_1 = .6$) under both DGP's. The sample size is set to $n =$ 7,882, which is the same as the sample size in the empirical illustration. Data is simulated from a given DGP with 500 simulation repetitions. The MI tests/CI's and SPUR2 projection tests/CI's are constructed using the tuning parameters recommended in Sections 4.7.1 and 5.6.1.

Let $p_{y,x} := P(Y_i = y, X = x)$ denote the joint distribution of the outcome given the covariate. The first DGP is defined by $p_{(0,0)',0} = .1$, $p_{(0,0)',1} = .05$, $p_{(0,1)',0} = .2$, $p_{(0,1)',1} = .4$, $p_{(1,1)',0} = .06$, and $p_{(1,1)',1} = .1$. This DGP is chosen to match the empirical marginal distribution of Y_i in the empirical illustration. The marginal distribution of Y_i in the first DGP is $P(Y_i = (0,0)') = .15$, $P(Y_i = (0,0)') = .6$, and $P(Y_i = (1,1)') = .15$. The corresponding probabilities from the empirical

Table 10.3: Tuning parameter sensitivity: SPUR2 CI upper bound comparisons with baseline.

	c_τ=0.5	c_τ=2	c_κ=0.5	c_κ=2	α_2=0.0475	α_2=0.04
$\beta_{\text{LCC}}^{\text{const}}$	−0.015	0.001	−0.366	0.245	−0.019	0.000
$\beta_{\text{LCC}}^{\text{size}}$	−0.012	0.000	−0.213	0.056	0.000	0.007
$\beta_{\text{LCC}}^{\text{pres}}$	−0.001	0.000	−0.181	0.166	−0.072	0.068
γ_{LCC}	0.000	0.000	0.000	0.000	0.000	0.000
$\beta_{\text{OA}}^{\text{const}}$	−0.068	0.000	−0.150	0.060	−0.078	0.000
$\beta_{\text{OA}}^{\text{size}}$	0.000	0.019	−0.168	0.574	−0.003	0.088
$\beta_{\text{OA}}^{\text{pres}}$	0.000	0.017	−0.326	0.250	−0.060	0.000
γ_{OA}	0.000	0.000	0.000	0.000	0.000	0.000
ρ	0.000	0.000	0.000	0.000	0.000	0.000

	ι_{sd}=10^{-8}	ι_{sd}=0.5	ι_q=0	ι_q=0.01	B=500	B=2,000
$\beta_{\text{LCC}}^{\text{const}}$	−0.002	0.001	−0.024	0.001	0.027	−0.017
$\beta_{\text{LCC}}^{\text{size}}$	−0.007	0.000	−0.014	0.000	−0.049	−0.042
$\beta_{\text{LCC}}^{\text{pres}}$	−0.001	0.000	−0.042	0.000	−0.011	−0.041
γ_{LCC}	0.000	0.000	0.000	0.000	0.000	0.000
$\beta_{\text{OA}}^{\text{const}}$	0.000	0.000	−0.099	0.009	−0.049	−0.061
$\beta_{\text{OA}}^{\text{size}}$	−0.012	0.000	−0.012	0.000	0.173	0.117
$\beta_{\text{OA}}^{\text{pres}}$	−0.186	0.000	−0.174	0.028	0.089	−0.091
γ_{OA}	0.000	0.000	0.000	0.000	0.000	0.000
ρ	0.000	0.000	0.000	0.000	0.000	0.000

marginal distribution are .153, .614, and .160. Under the first DGP, the (population) MI is $\Delta_F^{\text{inf}} = 0.133$, approximately, and the model is misspecified. The smallest hyperrectangle that contains the true MR-identified set $\Theta_I^{MR}(F)$ is approximately $[-.473, .230] \times [.661, 1.648] \times [.624, .624]$. Note that, for example, $[-.473, .230]$ is simply the projection of $\Theta_I^{MR}(F)$ onto the first element of θ.

The MI in the first DGP is larger than the estimated MI value in the empirical illustration. In consequence, we consider a second DGP with a smaller value of Δ_F^{inf}. This second DGP is defined by $p_{(0,0)',0} = .1075$, $p_{(0,0)',1} = .0425$, $p_{(0,1)',0} = .1425$, $p_{(0,1)',1} = .4075$, $p_{(1,1)',0} = .0425$, and $p_{(1,1)',1} = .1075$. Here, too, the marginal distribution of Y_i is similar to what is observed in the empirical illustration. Under this DGP, the (population) MI is $\Delta_F^{\text{inf}} = .024$, approximately, which is similar to the estimate of .023 for the case of $\rho = 0$ in the empirical illustration. For the second DGP, the smallest hyperrectangle that contains the true MR-identified set $\Theta_I^{MR}(F)$ is approximately $[-.376, -.157] \times [1.101, 1.300] \times [.440, .505]$.

We simulate the power of the MI and SPUR2 tests under the two DGP's.

Figure 10.1 shows the simulated rejection probabilities of the nominal .05 two-sided MI test of the null hypothesis $H_0 : \Delta_F^{\text{inf}} = \Delta_0$ for varying values of Δ_0. As shown in the figure, under both

DGP's, the test (based on the two-sided MI CI) has correct size and its rejection probabilities approach 1 reasonably quickly as the difference between the null value Δ_0 and the true value Δ_F^{inf} becomes larger in absolute value.

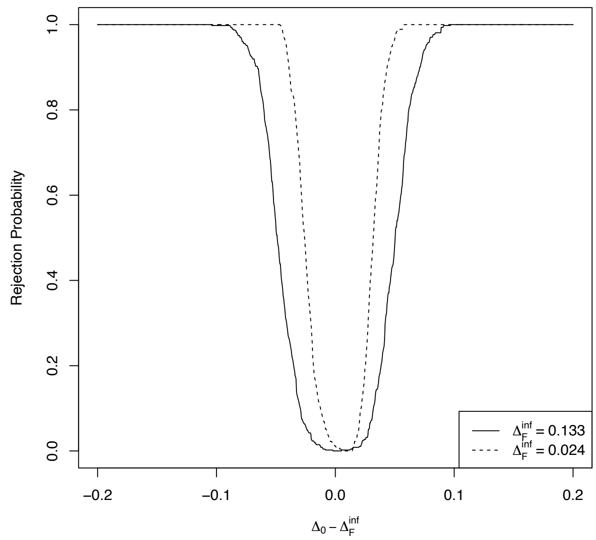

Figure 10.1: Rejection probabilities of tests concerning the misspecification index under the two DGP's. The two graphs show the rejection probabilities of the nominal .05 two-sided misspecification index test of the null hypothesis $H_0 : \Delta_F^{\text{inf}} = \Delta_0$ for 1,000 (equally spaced) values of Δ_0 such that $\Delta_0 - \Delta_F^{\text{inf}} \in [-.2, .2]$.

Next, we consider the SPUR2 tests. Under the first DGP with $\Delta_F^{\text{inf}} = .133$, the three graphs in Figure 10.2 show the rejection probabilities for the null hypotheses

$$\begin{aligned} H_{0,\beta_{LCC}} :& \quad (\beta_{LCC,0}, \beta_{OA}, \gamma)' \in \Theta_I^{MR}(F) \text{ for some } (\beta_{OA}, \gamma)', \\ H_{0,\beta_{OA}} :& \quad (\beta_{LCC}, \beta_{OA,0}, \gamma)' \in \Theta_I^{MR}(F) \text{ for some } (\beta_{LCC}, \gamma)' \text{ and} \\ H_{0,\gamma} :& \quad (\beta_{LCC}, \beta_{OA}, \gamma_0)' \in \Theta_I^{MR}(F) \text{ for some } (\beta_{LCC}, \beta_{OA})', \end{aligned} \tag{10.2}$$

respectively, for varying null values $\beta_{LCC,0}$, $\beta_{OA,0}$, and γ_0 and fixed true values. Note that inverting each of the tests gives the projection CI's for β_{LCC}, β_{OA}, and γ, respectively. The gray shaded regions show the (projected) MR-identified sets.[34] Figure 10.1 shows that the MI test has correct size and has nontrivial power against alternatives fairly close to the (projected) MR-identified set with the rejection probabilities approaching 1 reasonably quickly as the alternative becomes further away from the MR-identified set.

[34] When the (projected) MR-identified set is a singleton, the shaded area appears as a short vertical line below the power curve in the graph.

Under the second DGP with $\Delta_F^{\text{inf}} = .024$, the three graphs in Figure 10.3 show the rejection probabilities for the null hypotheses $H_{0,\beta_{LCC}}$, $H_{0,\beta_{OA}}$, and $H_{0,\gamma}$, respectively, for varying null values $\beta_{LCC,0}$, $\beta_{OA,0}$, and γ_0. Here too, the results demonstrate correct size and reasonable power properties of the test based on the SPUR2 projection CI's.

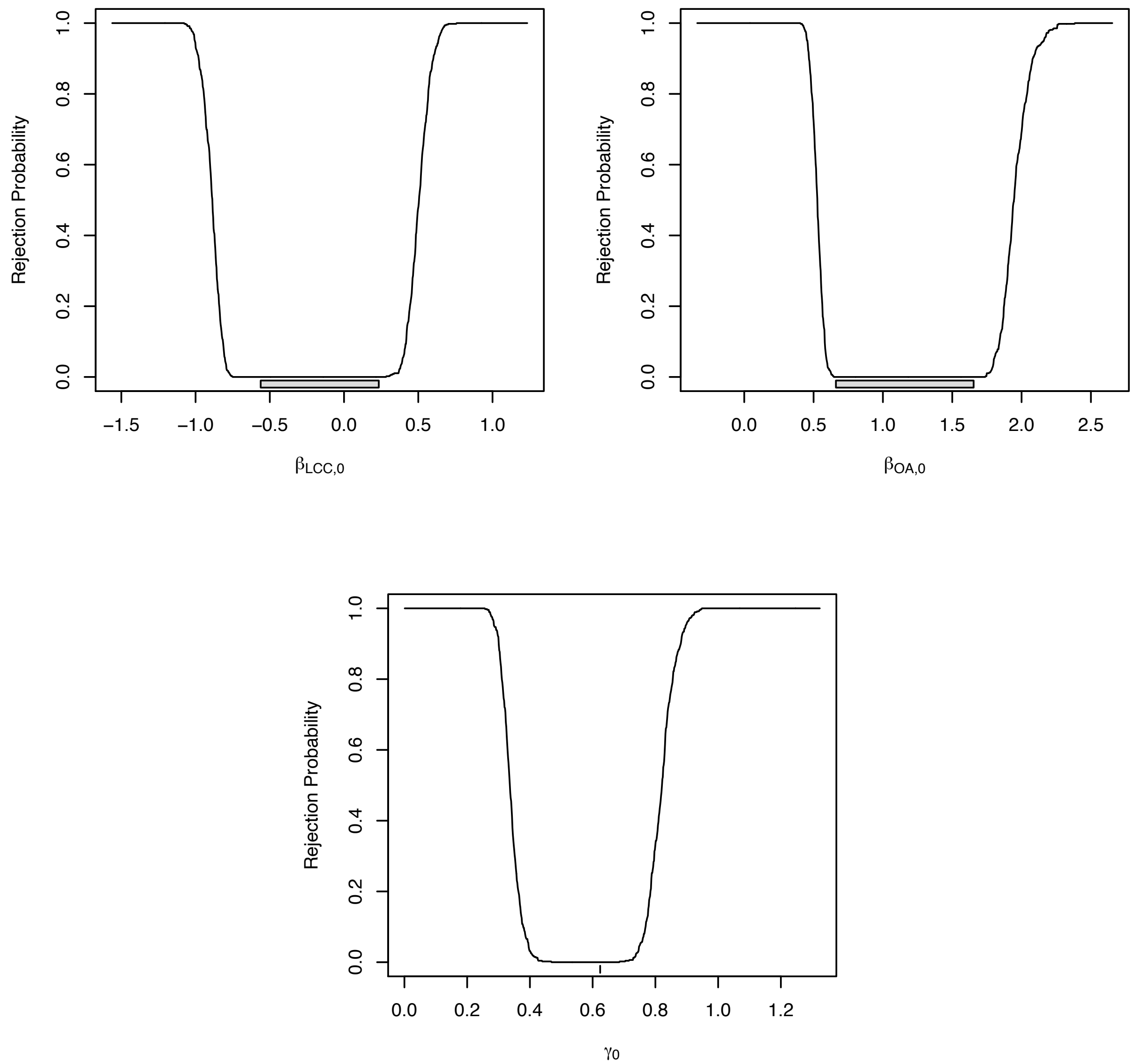

Figure 10.2: Rejection probabilities of the test based on the SPUR2 projection CI's for the null hypotheses $H_{0,\beta_{LCC}}$, $H_{0,\beta_{OA}}$ and $H_{0,\gamma}$ under the first DGP ($\Delta_F^{\text{inf}} = .133$).

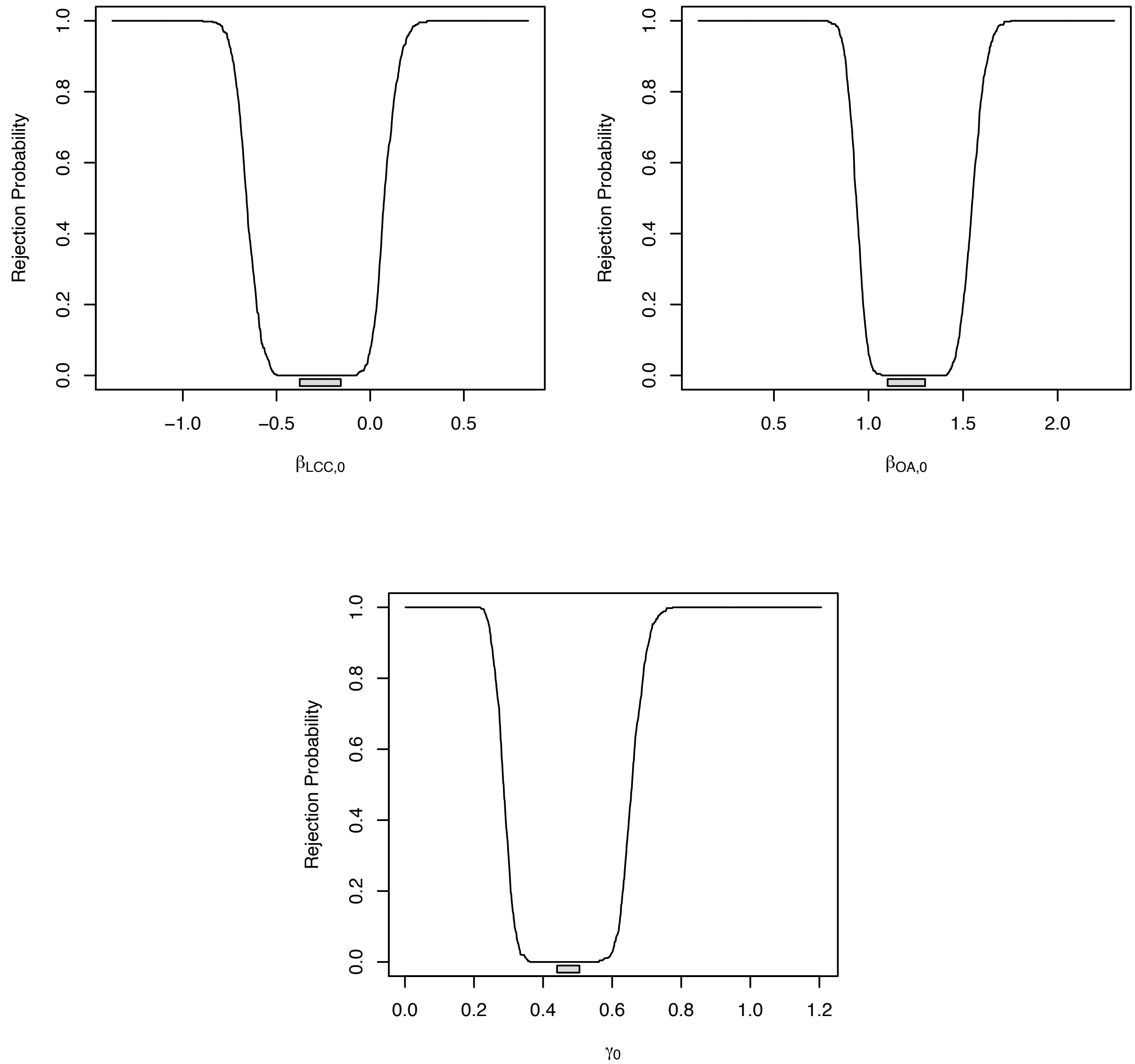

Figure 10.3: Rejection probabilities of the test based on the SPUR2 projection CI's for the null hypotheses $H_{0,\beta_{LCC}}$, $H_{0,\beta_{OA}}$ and $H_{0,\gamma}$ under the second DGP ($\Delta_F^{\text{inf}} = .024$).

10.3 Moment Inequalities in the Empirical Illustration

Here we show how the moment inequalities in (6.2) for the empirical illustration are obtained. As stated in the paper, we assume complete information so that the players observe ε_i in addition to everything the econometrician observes, and that the market outcome is determined by a pure strategy Nash equilibrium. Given these assumptions, the model implies the following (conditional) moment inequalities:

$$
\begin{aligned}
E[1(Y_i=(0,0)')|X_i] &= P(\varepsilon_{i,\mathrm{LCC}} \le -X_{i,\mathrm{LCC}}'\beta_{\mathrm{LCC}},\ \varepsilon_{i,\mathrm{OA}} \le -X_{i,\mathrm{OA}}'\beta_{\mathrm{OA}})\\
E[1(Y_i=(0,1)')|X_i] &\ge P(\varepsilon_{i,\mathrm{LCC}} \le -X_{i,\mathrm{LCC}}'\beta_{\mathrm{LCC}},\ \varepsilon_{i,\mathrm{OA}} \ge -X_{i,\mathrm{OA}}'\beta_{\mathrm{OA}})\\
&\quad +P(\varepsilon_{i,\mathrm{LCC}} \in [-X_{i,\mathrm{LCC}}'\beta_{\mathrm{LCC}}, -\gamma_{\mathrm{LCC}} - X_{i,\mathrm{LCC}}'\beta_{\mathrm{LCC}}],\ \varepsilon_{i,\mathrm{OA}} \ge -\gamma_{\mathrm{OA}} - X_{i,\mathrm{OA}}'\beta_{\mathrm{OA}})\\
E[1(Y_i=(0,1)')|X_i] &\le P(\varepsilon_{i,\mathrm{LCC}} \le -\gamma_{\mathrm{LCC}} - X_{i,\mathrm{LCC}}'\beta_{\mathrm{LCC}},\ \varepsilon_{i,\mathrm{OA}} \ge -X_{i,\mathrm{OA}}'\beta_{\mathrm{OA}})\\
E[1(Y_i=(1,0)')|X_i] &\ge P(\varepsilon_{i,\mathrm{LCC}} \ge -X_{i,\mathrm{LCC}}'\beta_{\mathrm{LCC}},\ \varepsilon_{i,\mathrm{OA}} \le -X_{i,\mathrm{OA}}'\beta_{\mathrm{OA}})\\
&\quad +P(\varepsilon_{i,\mathrm{LCC}} \ge -\gamma_{\mathrm{LCC}} - X_{i,\mathrm{LCC}}'\beta_{\mathrm{LCC}},\ \varepsilon_{i,\mathrm{OA}} \in [-X_{i,\mathrm{OA}}'\beta_{\mathrm{OA}}, -\gamma_{\mathrm{OA}} - X_{i,\mathrm{OA}}'\beta_{\mathrm{OA}}])\\
E[1(Y_i=(1,0)')|X_i] &\le P(\varepsilon_{i,\mathrm{LCC}} \ge -X_{i,\mathrm{LCC}}'\beta_{\mathrm{LCC}},\ \varepsilon_{i,\mathrm{OA}} \le -\gamma_{\mathrm{OA}} - X_{i,\mathrm{OA}}'\beta_{\mathrm{OA}})\\
E[1(Y_i=(1,1)')|X_i] &= P(\varepsilon_{i,\mathrm{LCC}} \ge -\gamma_{\mathrm{LCC}} - X_{i,\mathrm{LCC}}'\beta_{\mathrm{LCC}},\ \varepsilon_{i,\mathrm{OA}} \ge -\gamma_{\mathrm{OA}} - X_{i,\mathrm{OA}}'\beta_{\mathrm{OA}}). \qquad (10.3)
\end{aligned}
$$

Because

$$
\begin{aligned}
&E[1(Y_i=(1,0)')|X_i]\\
&= 1 - E[1(Y_i=(0,0)')|X_i] - E[1(Y_i=(0,1)')|X_i] - E[1(Y_i=(1,1)')|X_i], \qquad (10.4)
\end{aligned}
$$

we omit the moment inequalities corresponding to $E[1(Y_i=(1,0)')|X_i]$, which leaves us with two moment equalities and two moment inequalities. Writing the two moment equalities as four moment inequalities, the model can be written as six conditional moment inequalities.

Since X_i is discrete with its support $\mathcal{X}$ consisting of only $2^3 = 8$ different values, the six conditional moment inequalities can be transformed into $k = 48$ unconditional moment inequalities. For $x = (x_{\mathrm{LCC}}, x_{\mathrm{OA}})' \in \mathcal{X}$, define $p_x := P(X_i = x)$ and

$$
\begin{aligned}
P_{00}(x,\theta) &:= P(\varepsilon_{i,\mathrm{LCC}} \le -x_{\mathrm{LCC}}'\beta_{\mathrm{LCC}},\ \varepsilon_{i,\mathrm{OA}} \le -x_{\mathrm{OA}}'\beta_{\mathrm{OA}})\\
\underline{P}_{01}(x,\theta) &:= P(\varepsilon_{i,\mathrm{LCC}} \le -x_{\mathrm{LCC}}'\beta_{\mathrm{LCC}},\ \varepsilon_{i,\mathrm{OA}} \ge -x_{\mathrm{OA}}'\beta_{\mathrm{OA}})\\
&\quad + P(\varepsilon_{i,\mathrm{LCC}} \in [-x_{\mathrm{LCC}}'\beta_{\mathrm{LCC}}, -\gamma_{\mathrm{LCC}} - x_{\mathrm{LCC}}'\beta_{\mathrm{LCC}}],\ \varepsilon_{i,\mathrm{OA}} \ge -\gamma_{\mathrm{OA}} - x_{\mathrm{OA}}'\beta_{\mathrm{OA}})\\
\overline{P}_{01}(x,\theta) &:= P(\varepsilon_{i,\mathrm{LCC}} \le -\gamma_{\mathrm{LCC}} - x_{\mathrm{LCC}}'\beta_{\mathrm{LCC}},\ \varepsilon_{i,\mathrm{OA}} \ge -x_{\mathrm{OA}}'\beta_{\mathrm{OA}})\\
P_{11}(x,\theta) &:= P(\varepsilon_{i,\mathrm{LCC}} \ge -\gamma_{\mathrm{LCC}} - x_{\mathrm{LCC}}'\beta_{\mathrm{LCC}},\ \varepsilon_{i,\mathrm{OA}} \ge -\gamma_{\mathrm{OA}} - x_{\mathrm{OA}}'\beta_{\mathrm{OA}}). \qquad (10.5)
\end{aligned}
$$

Consider, for example, the conditional moment inequality

$$
E[1(Y_i=(0,0)') - P_{00}(X_i,\theta)|X_i] \ge 0,
$$

which corresponds to one of the moment inequalities resulting from rewriting the first line of (10.5) as two inequalities. This is equivalent to

$$\begin{aligned} & E[1(Y_i = (0,0)') - P_{00}(X_i,\theta)|X_i = x] \geq 0 \ \forall x \in \mathcal{X} \\ \Leftrightarrow \quad & E[(1(Y_i = (0,0)') - P_{00}(x,\theta))1(X_i = x)] \geq 0 \ \forall x \in \mathcal{X} \\ \Leftrightarrow \quad & E[1(Y_i = (0,0)', X_i = x) - P_{00}(x,\theta)p_x] \geq 0 \ \forall x \in \mathcal{X}, \end{aligned} \tag{10.6}$$

where the first equivalence holds because $P_{00}(X_i,\theta)1(X_i = x) = P_{00}(x,\theta)1(X_i = x)$ due to the independence between X_i and ε_i. Following Kaido, Molinari, and Stoye (2019), we take p_x to be known.[35] As is evident from the expression in the last line, an implication of this assumption is that the data and parameters become additively separable. Hence, (10.3) is equivalent to the following moment inequality model:

$$\begin{aligned} & E[1(Y_i = (0,0)', X_i = x) - P_{00}(x,\theta)p_x] \geq 0, \\ & E[P_{00}(x,\theta)p_x - 1(Y_i = (0,0)', X_i = x)] \geq 0, \\ & E[1(Y_i = (0,1)', X_i = x) - \underline{P}_{01}(x,\theta)p_x] \geq 0, \\ & E[\overline{P}_{01}(x,\theta)p_x - 1(Y_i = (0,1)', X_i = x)] \geq 0, \\ & E[1(Y_i = (1,1)', X_i = x) - P_{11}(x,\theta)p_x] \geq 0, \\ & E[P_{11}(x,\theta)p_x - 1(Y_i = (1,1)', X_i = x)] \geq 0, \end{aligned} \tag{10.7}$$

for all $x \in \mathcal{X}$, which are the moment inequalities given in (6.2). In practice, we take the empirical distribution of X_i to be the true distribution and plug it in place of p_x, as in Kaido, Molinari, and Stoye (2019).

10.4 Initial Values for Computation of the Projection CI's

This section describes the initial values that are used for computing projection CI's of the GMS, SPUR1, and SPUR2 types in the empirical illustration. These choices work well in the empirical illustration, however, in other models it's not clear how well they will work. Define the arg min

[35] If p_x is unknown, one can use the expression in the second line of (10.6) as the moment inequalities.

parameter values for the GMS and SPUR1 projection CI's:

$$\begin{aligned}
\theta_{1-\alpha}^{GMS,a,l} &:= \arg\min\nolimits_{\theta\in\Theta}\theta_a \text{ s.t. } S_{n,Std}(\theta) \leq \widehat{c}_{n,GMS}(\theta,1-\alpha), \\
\theta_{1-\alpha}^{GMS,a,u} &:= \arg\min\nolimits_{\theta\in\Theta} -\theta_a \text{ s.t. } S_{n,Std}(\theta) \leq \widehat{c}_{n,GMS}(\theta,1-\alpha), \\
\theta_{1-\alpha}^{SPUR1,a,l} &:= \arg\min\nolimits_{\theta\in\Theta}\theta_a \text{ s.t. } S_n(\theta) \leq \widehat{c}_n(\theta,1-\alpha), \text{ and} \\
\theta_{1-\alpha}^{SPUR1,a,u} &:= \arg\min\nolimits_{\theta\in\Theta} -\theta_a \text{ s.t. } S_n(\theta) \leq \widehat{c}_n(\theta,1-\alpha)
\end{aligned} \tag{10.8}$$

for $a = 1, ..., d_\theta$. We refer to the first two problems as the GMS projection problems and the latter two as the SPUR1 projection problems.

To calculate the projection CI's reported in Table 2 for the empirical illustration (and to calculate projection CI's in general, as described in Section 5.2), one must calculate $(\theta_{1-\alpha}^{GMS,a,l}, \theta_{1-\alpha}^{GMS,a,u})$ for $\alpha = .05$ (for the GMS projection CI) and $\alpha = .045$ (to construct the SPUR2 projection CI), and $(\theta_{1-\alpha}^{SPUR1,a,l}, \theta_{1-\alpha}^{SPUR1,a,u})$ for $\alpha = .045$. Calculating such quantities amounts to solving non-linear, non-convex constrained optimization problems. Hence, the choice of the initial values is relevant. An added difficulty is that finding points in the feasible set is not trivial in this context. Here, we introduce a systematic way to find feasible values, and make a recommendation on how to choose the initial values based on this method.

We make use of the following two quantities: $\theta_\Delta^{init} := \arg\min_{\theta\in\Theta}\max_{j\leq k}\widehat{\Delta}_{nj}(\theta)$ and $\theta_{S_1}^{init} := \arg\min_{\theta\in\Theta} S_{n,Std}(\theta)$ using the $S_1(\cdot)$ function.[36] For now, we presume that these quantities are well-defined in the sense that the arg min sets are singleton sets. Below, we discuss the choice of initial values when $\Theta_{S_1}^{init} := \arg\min_{\theta\in\Theta} S_{n,Std}(\theta)$ is found to contain multiple points, which typically holds when the model is correctly specified with a non-singleton identified set.[37] The quantities θ_Δ^{init} and $\theta_{S_1}^{init}$ are "likely" to lie in the feasible sets for the GMS and SPUR1 projection problems, respectively.

To calculate $\theta_{.95}^{GMS,a,l}$, we recommend using initial values θ_Δ^{init}, $\theta_{S_1}^{init}$, and $\theta_{1-\tau}^{GMS,a,l}$ for $\tau \in \mathcal{T}_{.95} = \{.1, .2, \ldots, .5\}$.[38] The last set of initial values always belong to the feasible set due to the smaller nominal coverage than the desired coverage of .95. We consider analogous initial values for the calculation of $\theta_{.95}^{GMS,a,u}$. To calculate $\theta_{.955}^{GMS,a,l}$, we recommend adding $\theta_{.95}^{GMS,a,l}$ to the set of initial values considered, and thus using θ_Δ^{init}, $\theta_{S_1}^{init}$, and $\theta_{1-\tau}^{GMS,a,l}$ for $\tau \in \mathcal{T}_{.955}$ as the initial values. To

[36] Calculation of $\theta_{S_1}^{init}$ can be done using standard software with initial values drawn from a Sobol sequence in Θ, as we do to calculate θ_Δ^{init}. For example, in the empirical illustration, we use a Sobol sequence with 100 points for both θ_Δ^{init} and $\theta_{S_1}^{init}$.

[37] We maintain the presumption that $\arg\min_{\theta\in\Theta}\max_{j\leq k}\widehat{\Delta}_{nj}(\theta)$ is a singleton set because this is typically the case. However, if this is not the case, one can use a similar method to the one described for $\arg\min_{\theta\in\Theta} S_{n,Std}(\theta)$.

[38] The points $\theta_{1-\tau}^{GMS,a,l}$ for $\tau \in \mathcal{T}_{.95} = \{.1, .2, ..., .5\}$ are computed by first calculating $\theta_{1-\tau}^{GMS,a,l}$ for $\tau = .5$ using the initial values θ_Δ^{init} and $\theta_{S_1}^{init}$. Then, $\theta_{1-\tau}^{GMS,a,l}$ for $\tau = .4$ is computed using the initial values θ_Δ^{init}, $\theta_{S_1}^{init}$, and $\theta_{.5}^{GMS,a,l}$. The remaining $\theta_{1-\tau}^{GMS,a,l}$ values are computed inductively, using θ_Δ^{init}, $\theta_{S_1}^{init}$, and $\{\theta_{1-\tau'}^{GMS,a,l} : \tau' > \tau, \tau' \in \mathcal{T}_{.95}\}$.

calculate $\theta^{SPUR1,a,l}_{.955}$, we consider the initial values used to calculate $\theta^{GMS,a,l}_{.955}$ and also $\theta^{SPUR1,a,l}_{1-\tau}$ for $\tau \in \mathcal{T}_{.955}$, where $\theta^{SPUR1,a,l}_{1-\tau}$ is defined analogously to $\theta^{GMS,a,l}_{1-\tau}$ using the SPUR test statistic in place of the GMS test statistic throughout.[39],[40]

Hence, to calculate $\theta^{GMS,a,l}_{.95}$, $\theta^{GMS,a,l}_{.955}$, and $\theta^{SPUR1,a,l}_{.955}$, one uses 7, 8, and 14 initial values, respectively. These initial values are also obtained by solving optimization problems using multiple initial values. Overall, one ends up running the optimizer 27 times to compute $\theta^{GMS,a,l}_{.95}$, an additional 8 times (35 total) for $\theta^{GMS,a,l}_{.955}$, and an additional 21 times (56 total) for $\theta^{SPUR1,a,l}_{.955}$.

When $\arg\min_{\theta\in\Theta} S_{n,Std}(\theta)$ is not a singleton, the procedure provided above is modified by choosing suitable points from $\arg\min_{\theta\in\Theta} S_{n,Std}(\theta)$.[41] Let $\widehat{\Theta}^{init}_{S_1}$ denote the set of parameter values that obtain the same minimum value based on some set of initial values and write $m^{init} := |\widehat{\Theta}^{init}_{S_1}|$. For example, in the adjusted empirical illustration, we use 100 initial values drawn according to a Sobol sequence in Θ to calculate $\min_{\theta\in\Theta} S_{n,Std}(\theta)$. Using these 100 initial values, we found $m^{init} = 17$ different optimal points that obtain the same minimum value of zero. Here, $\widehat{\Theta}^{init}_{S_1}$ is the set of these 17 points. Let $\theta^{init}_{S_1,a,m}$ denote the point in $\widehat{\Theta}^{init}_{S_1}$ that has the mth smallest ath component. For example, $\theta^{init}_{S_1,a,1}$ is the point in $\widehat{\Theta}^{init}_{S_1}$ with the smallest ath component.

When $\arg\min_{\theta\in\Theta} S_{n,Std}(\theta)$ is not a singleton, to calculate $\theta^{GMS,a,l}_{.95}$, we recommend using the following initial values: first 10 points from a Sobol sequence in Θ, θ^{init}_{Δ}, $\theta^{init}_{S_1,a,m}$ for $m = 1, \ldots, 5$, and $\theta^{GMS,a,l}_{1-\tau}$ for $\tau \in \mathcal{T}_{.95} = \{.1, .2, \ldots, .5\}$. The idea is to choose points from $\widehat{\Theta}^{init}_{S_1}$ that are likely to be closest to the optimum. Since the objective here is to minimize the ath component, we choose points that have small ath components. The quantity $\theta^{GMS,a,l}_{1-\tau}$ is calculated in a slightly different way than above, but we abuse notation and keep notation same as above.[42] We recommend analogous initial values for the calculation of $\theta^{GMS,a,u}_{.95}$: first 10 points from a Sobol sequence, θ^{init}_{Δ}, $\theta^{init}_{S_1,a,m}$ for $m = m^{init}, m^{init}-1, \ldots, m^{init}-4$, and $\theta^{GMS,a,u}_{1-\tau}$ for $\tau \in \mathcal{T}_{.95} = \{.1, .2, \ldots, .5\}$. To calculate $\theta^{GMS,a,l}_{.955}$, we recommend adding $\theta^{GMS,a,l}_{.95}$ to the set of initial values considered. To calculate $\theta^{SPUR1,a,l}_{.955}$, we recommend using the initial values used to calculate $\theta^{GMS,a,l}_{.955}$ and also $\theta^{SPUR1,a,l}_{1-\tau}$ for $\tau \in \mathcal{T}_{.955}$, where $\theta^{SPUR1,a,l}_{1-\tau}$ is defined analogously to $\theta^{GMS,a,l}_{1-\tau}$ using the SPUR test statistic in place of the

[39] Calculation of $\theta^{SPUR1,a,l}_{1-\tau}$ for $\tau \in \mathcal{T}_{.955}$ is done in an analogous "inductive" way to that of $\theta^{GMS,a,l}_{1-\tau}$ for $\tau \in \mathcal{T}_{.955}$, but here $\theta^{GMS,a,l}_{1-\tau'}$ with $\tau' > \tau$ are used as initial values as well. That is, $\theta^{SPUR1,a,l}_{1-\tau}$ is calculated using initial values θ^{init}_{Δ}, $\theta^{init}_{S_1}$, $\{\theta^{GMS,a,l}_{1-\tau'} : \tau' > \tau, \tau' \in \mathcal{T}_{.955}\}$, and $\{\theta^{SPUR1,a,l}_{1-\tau'} : \tau' > \tau, \tau' \in \mathcal{T}_{.955}\}$

[40] The GMS CS is typically contained in the SPUR1 CS, and thus the $\{\theta^{GMS,a,l}_{1-\tau}\}_{\tau\in\mathcal{T}_{.955}}$ initial values are typically feasible for the SPUR1 CS. The reverse is not true, which is why we do not recommend using, for example, $\theta^{SPUR1,a,l}_{.95}$ as an initial value when computing $\theta^{GMS,a,l}_{.95}$.

[41] Our recommendation is based on experimentation with an adjusted version of the empirical illustration. The adjustment is to add .05 to the (standardized) moments to force the model to be correctly specified with a moderately large identified set. Accordingly, $\arg\min_{\theta\in\Theta} S_{n,Std}(\theta)$ is not a singleton under this adjustment.

[42] The points $\theta^{GMS,a,l}_{1-\tau}$ for $\tau \in \mathcal{T}_{.95} = \{.1, .2, ..., .5\}$ are computed by first calculating $\theta^{GMS,a,l}_{1-\tau}$ for $\tau = .5$ using the initial values θ^{init}_{Δ} and $\theta^{init}_{S_1,a,1}$. Then, $\theta^{GMS,a,l}_{1-\tau}$ for $\tau = .4$ is computed using the initial values θ^{init}_{Δ}, $\theta^{init}_{S_1,a,1}$, and $\theta^{GMS,a,l}_{.5}$. The remaining $\theta^{GMS,a,l}_{1-\tau}$ values are computed inductively, using θ^{init}_{Δ}, $\theta^{init}_{S_1,a,1}$, and $\{\theta^{GMS,a,l}_{1-\tau'} : \tau' > \tau, \tau' \in \mathcal{T}_{.95}\}$.

GMS test statistic throughout.

11 Uniform Consistency and Rate of Convergence of $\widehat{\Theta}_n$

This section shows that the set estimator $\widehat{\Theta}_n$, defined in (4.20), is uniformly consistent for the MR-identified set $\Theta_I^{MR}(F)$). It also establishes the rate of convergence of $\widehat{\Theta}_n$ to $\Theta_I^{MR}(F_n)$ under the Hausdorff distance d_H.[43] These results are similar to results in Theorem 3.1 of Chernozhukov, Hong, and Tamer (2007).

All limits are as the sample size $n \to \infty$. Let $O_p^{\Theta}(1)$ denote random functions that are $O_p(1)$ uniformly over $\theta \in \Theta$.

11.1 Uniform Consistency of $\widehat{\Theta}_n$

The following result shows that the set estimator $\widehat{\Theta}_n$ is uniformly consistent for the MR-identified set $\Theta_I^{MR}(F)$ over $F \in \mathcal{P}$ with respect to the Hausdorff metric d_H. The result is similar to results in Theorem 3.1 of CHT except that it applies under both correct model specification and misspecification, and it establishes uniform over $F \in \mathcal{P}$ consistency, rather than pointwise in F consistency.

For $\theta \in \Theta$ and $A \subset \Theta$, define the distance between θ and A as $d(\theta, A) := \inf_{\theta' \in A} ||\theta - \theta'||$. For any $\varepsilon > 0$ and $F \in \mathcal{P}$, define

$$\Theta_{I,\varepsilon}^{MR}(F) := \{\theta \in \Theta : d(\theta, \Theta_I^{MR}(F)) \leq \varepsilon\}. \tag{11.1}$$

The set $\Theta_{I,\varepsilon}^{MR}(F)$ is an ε-expansion of the MR-identified set $\Theta_I^{MR}(F)$.

For any $F \in \mathcal{P}$, $\inf_{\theta \in \Theta \backslash \Theta_{I,\varepsilon}^{MR}(F)} \max_{j \leq k} [E_F \widetilde{m}_j(W, \theta)]_- - r_F^{\inf} > 0$ for all $\varepsilon > 0$ under Assumption A.0 by the definitions of $r_F^{\inf}$ and $\Theta_{I,\varepsilon}^{MR}(F)$. The following Assumption A.9 requires that this positive quantity is bounded away from zero over $F \in \mathcal{P}$.

Assumption A.9. For all $\varepsilon > 0$, $\inf_{F \in \mathcal{P}} \inf_{\theta \in \Theta \backslash \Theta_{I,\varepsilon}^{MR}(F)} \max_{j \leq k} [E_F \widetilde{m}_j(W, \theta)]_- - r_F^{\inf} > 0$.

We are not aware of any interesting models that fail Assumption A.9.

Uniform consistency of $\widehat{\Theta}_n$ for $\Theta_I^{MR}(F)$ is established in the following theorem.

Theorem 11.1 *Suppose Assumptions* A.0–A.5 *and* A.9 *hold and the positive constants* $\{\tau_n\}_{n \geq 1}$

[43] The Hausdorff distance between two non-empty sets Θ_1 and Θ_2 in Θ is $d_H(\Theta_1, \Theta_2) := \max\{\sup_{\theta_1 \in \Theta_1} \inf_{\theta_2 \in \Theta_2} ||\theta_1 - \theta_2||, \sup_{\theta_2 \in \Theta_2} \inf_{\theta_1 \in \Theta_1} ||\theta_1 - \theta_2||\}$.

that appear in (4.20) *satisfy* $\tau_n \to \infty$ *and* $\tau_n/n^{1/2} = o(1)$. *Then, for all* $\varepsilon > 0$,

$$\lim_{n\to\infty} \sup_{F\in\mathcal{P}} P_F(d_H(\widehat{\Theta}_n, \Theta_I^{MR}(F)) > \varepsilon) = 0.$$

Comments. (i). If Assumption A.9 fails to hold, the result of Theorem 11.1 holds with $\mathcal{P}_U$ in place of $\mathcal{P}$ for any $\mathcal{P}_U \subset \mathcal{P}$ for which Assumption A.9 holds with $\mathcal{P}_U$ in place of $\mathcal{P}$. In particular, for a fixed distribution $F \in \mathcal{P}$, the result of Theorem 11.1 holds with $\mathcal{P}_U = \{F\}$ in place of $\mathcal{P}$ because Assumption A.9 automatically holds in this case.

(ii). The proofs of Theorem 11.1 and Lemma 11.2 below are given in online Appendix B.

11.2 Consistency and Rate of Convergence of $\widehat{\Theta}_n$ under $\{F_n\}_{n\geq 1}$

Next, we establish consistency and rate of convergence results for $\widehat{\Theta}_n$ under a drifting sequence of distributions $\{F_n\}_{n\geq 1}$. These results are similar to results in Theorem 3.1 of Chernozhukov, Hong, and Tamer (2007), which apply to a fixed distribution F. The proofs also are similar.

The following assumption ensures that $\inf_{\theta\in\Theta\backslash\Theta_{I,\varepsilon}^{MR}(F_n)} \max_{j\leq k}\ [E_{F_n}\widetilde{m}_j(W,\theta)]_- - r_{F_n}^{\inf}$ is bounded away from zero under $\{F_n\}_{n\geq 1}$, where the set $\Theta_{I,\varepsilon}^{MR}(F_n)$ is defined in (11.1).

Assumption C.9. For all $\varepsilon > 0$,

$$\liminf_{n\to\infty} \left(\inf_{\theta\in\Theta\backslash\Theta_{I,\varepsilon}^{MR}(F_n)} \max_{j\leq k}\ [E_{F_n}\widetilde{m}_j(W,\theta)]_- - r_{F_n}^{\inf} \right) > 0.$$

The following minorant condition for the population moments is similar to (4.1) of Chernozhukov, Hong, and Tamer (2007). It is used to determine the rate of convergence of $d_H(\widehat{\Theta}_n,$ $\Theta_I^{MR}(F_n))$ to zero.

Assumption C.10. There exist positive constants C, ε, and γ such that for all $\theta \in \Theta$ and $n \geq 1$,

$$\max_{j\leq k} [E_{F_n}\widetilde{m}_j(W,\theta)]_- - r_{F_n}^{\inf} \geq C \cdot (\min\{d(\theta, \Theta_I^{MR}(F_n)), \varepsilon\})^{\gamma}.$$

Typically, Assumption C.10 holds with $\gamma = 1$.

Part (a) of the following lemma is used in the proof of Theorem 11.1 given below. Part (b) provides a rate of convergence result for $\widehat{\Theta}_n$.

Lemma 11.2 *Suppose Assumptions* A.0, C.4, C.5, C.7, *and* C.9 *hold under* $\{F_n\}_{n\geq 1}$. *Suppose the positive constants* $\{\tau_n\}_{n\geq 1}$ *that appear in* (4.20) *satisfy* $\tau_n \to \infty$ *and* $\tau_n/n^{1/2} = o(1)$. *Then,*

(a) $d_H(\widehat{\Theta}_n, \Theta_I^{MR}(F_n)) = o_p(1)$ *and*

(b) $d_H(\widehat{\Theta}_n, \Theta_I^{MR}(F_n)) = O_p((\tau_n/n^{1/2})^{1/\gamma})$ *provided Assumption* C.10 *also holds.*

Comment. When $F_n = F$ for all $n \geq 1$ for some $F \in \mathcal{P}$, Assumption C.9 holds by the definitions of $r_F^{\inf}$ and $\Theta_{I,\varepsilon}^{MR}(F)$ under Assumption A.0. In consequence, Lemma 11.2(a) establishes the result of Theorem 11.1 with $\sup_{F \in \mathcal{P}}$ deleted and without imposing Assumption A.9.

12 Problems with Subsampling SPUR and Recentered Test Statistics under Model Misspecification

Next, we show that subsampling a SPUR test statistic or a recentered test statistic does not necessarily deliver correct asymptotic size under identifiable model misspecification. We consider the simple lower/upper bound model on a scalar parameter θ discussed in Section 2. Thus, the observations $\{W_i\}_{i \leq n}$ are i.i.d. with $W_i = (W_{i1}, W_{i2})' \sim N(\mu, I_2)$, where $\mu = (\mu_1, \mu_2)' \in R^2$. The population moment inequalities are $E_F W_{i1} \leq \theta$ and $\theta \leq E_F W_{i2}$. In this model, $\Delta_F^{\inf} = (\mu_1 - \mu_2)/2$ and $r_F^{\inf} = [\mu_1 - \mu_2]_+/2$, where $[x]_+ := \max\{x, 0\}$. The model is misspecified when $\mu_1 > \mu_2$.

The null hypothesis of interest is $H_0 : \theta = 0$. We consider null distributions F for which the model is misspecified and $\mu = (c/n^{1/2}, -c/n^{1/2})'$ for some $c > 0$, which implies that $\Theta_I^{MR}(F) = \{0\}$ and $r_F^{\inf} = c/n^{1/2}$.

We consider a SPUR test statistic based on the "max" function S_4. As shown in Section 35, this is equivalent to a recentered "max" test statistic, i.e., $S_n(\theta) = S_{4n,Recen}(\theta)$, where the latter is defined in (35.1). In the present case, we have

$$\begin{aligned} S_n(\theta) &= n^{1/2} \max\{\overline{W}_{n1} - \theta, \theta - \overline{W}_{n2}, 0\} - n^{1/2} \inf_{\overline{\theta} \in \Theta} \max\{\overline{W}_{n1} - \overline{\theta}, \overline{\theta} - \overline{W}_{n2}, 0\} \\ &= n^{1/2} \max\{\overline{W}_{n1} - \theta, \theta - \overline{W}_{n1}, 0\} - n^{1/2} \max\left\{\frac{\overline{W}_{n1} - \overline{W}_{n2}}{2}, 0\right\}, \end{aligned} \tag{12.1}$$

where $\overline{W}_{nj} := n^{-1} \sum_{i=1}^{n} W_{ij}$ for $j = 1, 2$. Let $(Z_1, Z_2)' \sim N(0_2, I_2)$. Then, $n^{1/2}\overline{W}_{n1} \overset{d}{=} Z_1 + c$ and $n^{1/2}\overline{W}_{n2} \overset{d}{=} -Z_2 - c$, where "$\overset{d}{=}$" denotes equality in distribution, and the two variables are independent. The test statistic evaluated at the null value $\theta = 0$ satisfies

$$\begin{aligned} S_n(0) &\overset{d}{=} \max\{Z_1 + c, Z_2 + c, 0\} - \max\left\{\frac{Z_1 + Z_2}{2} + c, 0\right\} \\ &= \max\{Z_1, Z_2, -c\} - \max\left\{\frac{Z_1 + Z_2}{2}, -c\right\}. \end{aligned} \tag{12.2}$$

If c is very large, the two summands are essentially $\max\{Z_1, Z_2\}$ and $-\frac{Z_1+Z_2}{2}$, which simplifies the distribution. Hence, we consider the case where $c = c_n \to \infty$ as $n \to \infty$. In this case, under H_0,

$$S_n(0) \to_d \max\{Z_1, Z_2\} - \frac{Z_1 + Z_2}{2} =: S_\infty \text{ as } n \to \infty. \tag{12.3}$$

We consider subsampling with a subsampling size b_n that satisfies $b_n/n \to 0$. The subsample statistic $S_{b_n}(0)$ satisfies: under H_0,

$$\begin{aligned} S_{b_n}(0) &= b_n^{1/2} \max\{\overline{W}_{b_n 1}, -\overline{W}_{b_n 2}, 0\} - b_n^{1/2} \max\left\{\frac{\overline{W}_{b_n 1} - \overline{W}_{b_n 2}}{2}, 0\right\} \\ &\overset{d}{=} \max\{Z_1 + (b_n/n)^{1/2} c, Z_2 + (b_n/n)^{1/2} c, 0\} - \max\left\{\frac{Z_1 + Z_2}{2} + (b_n/n)^{1/2} c, 0\right\} \\ &= \max\{Z_1, Z_2, -(b_n/n)^{1/2} c\} - \max\left\{\frac{Z_1 + Z_2}{2}, -(b_n/n)^{1/2} c\right\} \\ &\to_d \max\{Z_1, Z_2, 0\} - \max\left\{\frac{Z_1 + Z_2}{2}, 0\right\} =: S_{Sub,\infty}, \end{aligned} \tag{12.4}$$

where the convergence in distribution holds for $c = c_n$ that satisfies $c_n \to \infty$ and $c_n = o((n/b_n)^{1/2})$. In consequence, the nominal level α subsampling critical value converges in probability to the $1-\alpha$ quantile of $S_{Sub,\infty}$, denoted by $cv_{Sub,\infty}(1-\alpha)$, see Andrews and Guggenberger (2010, Thm. 1(ii) and Lem. 5) for details.

Simulations of $NRP_{Sub,\infty}(\alpha) := P(S_\infty > cv_{Sub,\infty}(1-\alpha))$ yield $NRP_{Sub,\infty}(.10) = .152$, $NRP_{Sub,\infty}(.05) = .078$, and $NRP_{Sub,\infty}(.01) = .016$ (using 100 million simulation repetitions). These values give lower bounds on the asymptotic sizes of the subsampling test for $\alpha = .10$, $.05$, and $.01$. In each case, the lower bound is slightly larger than 150% of the nominal level of the test. Hence, the subsampling test does not have correct (uniform) asymptotic size in this model.

Because the subsampling test does not have correct asymptotic size under misspecification in one of the simplest moment inequality models in the literature, we conclude that subsampling a SPUR test statistic or recentered test statistic does not necessarily deliver correct asymptotic size under misspecification in moment inequality models.

Online Appendix A References

Allen R., and J. Rehbeck (2019), “Assessing Misspecification and Aggregation for Structured Preferences,” Department of Economics Research Report #2019-4, Western University, London, Ontario, Canada.

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 P. Guggenberger (2010), “Asymptotic Size and a Problem with Subsampling and with the m Out of n Bootstrap,” *Econometric Theory,* 26, 426–468.

Andrews, D. W. K., and S. Kwon (2022), “Misspecified Moment Inequality Models: Inference and Diagnostics” Cowles Foundation Discussion Paper Number 2184R2, Yale University.

Andrews, D. W. K., and G. Soares (2010), “Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection,” *Econometrica,* 78, 119–157.2.

Chernozhukov, V., H. Hong, and E. Tamer (2007), “Estimation and Confidence Regions for Parameter Sets in Econometric Models,” *Econometrica,* 75, 1243–284.

Online Appendix B

to

Misspecified Moment Inequality Models: Inference and Diagnostics

Contents: Proofs of the SPUR CS Results

Donald W. K. Andrews
Cowles Foundation for Research in Economics
Yale University

Soonwoo Kwon
Department of Economics
Brown University

First Version: March 12, 2017
Revised: July 29, 2022

13 Outline of Online Appendix B

Appendix B proves the results of the paper for SPUR1 and SPUR2 tests and confidence sets (CS's).

References to sections with section numbers 6 or less refer to sections of the main paper. Similarly, all equations, theorems, and lemmas with section numbers 6 or less refer to results in the main paper. BCS abbreviates Bugni, Canay, and Shi (2015).

Section 14 of Appendix B states some additional assumptions used in Sections 5.5 and 4.6. For ease of reference, all of the assumptions used in the paper and Appendix B are listed in the last section of Appendix B, Section 22.

Section 15 provides the asymptotic distribution of the SPUR test statistic under drifting sequences of distributions, using the approach in BCS.

Section 16 states Lemma 16.1, which gives sufficient conditions for Assumptions NLA and CA, and proves Lemmas 15.1, 15.2, and 16.1.

Section 17 proves Theorem 15.3, which gives the asymptotic distribution of the SPUR test statistic.

Sections 18–20 prove the main results of the paper for SPUR1 and SPUR2 tests and CS's.

Section 18 states Theorem 18.1, which is the key ingredient to the proof of Theorem 4.1 and the Comment to Theorem 4.1, which provide asymptotic level results for SPUR1 and SPUR2 tests and CS's. Theorem 18.1 provides asymptotic null rejection probability (NRP) results for the nominal level α SPUR1 test $\phi_{n,SPUR1}(\theta_n)$, defined in (4.7), under drifting subsequences of distributions and parameter values. Section 19 proves Lemmas 18.2–18.4, which are used in the proof of Theorem 18.1.

Section 20 proves Theorem 4.1, which shows that the SPUR2 tests and CS's have correct asymptotic level, using Theorem 18.1. Section 20 also proves analogous results for SPUR1 tests and CS's.

Section 21 proves Theorem 11.1 and Lemma 11.2 stated in online Appendix A, which give uniform consistency and rate of convergence results for the estimator $\widehat{\Theta}_n$ of the MR-identified set.

All limits are as the sample size $n \to \infty$. Let $R_{[\pm\infty]} := R \cup \{\pm\infty\}$ and $R_{[+\infty]} := R \cup \{+\infty\}$. Let $||\cdot||$ denote the Euclidean norm for vectors and the Frobenious norm for matrices. Let $[x]_- := \max\{-x, 0\}$ (≥ 0) and $[x]_+ := \max\{x, 0\}$ (≥ 0) for $x \in R$. Let $o_p^\Theta(1)$ and $O_p^\Theta(1)$ denote quantities that are $o_p(1)$ and $O_p(1)$, respectively, uniformly over $\theta \in \Theta$.

14 Additional Assumptions

Here, we state some additional assumptions used in Sections 5.5 and 4.6. Assumption S.1 is stated in a footnote in Section 4.1. The population-standard-deviation-normalized sample moments are

$$\widetilde{m}_{nj}(\theta) := n^{-1}\sum_{i=1}^{n}\widetilde{m}_j(W_i,\theta), \text{ where } \widetilde{m}_j(W,\theta) := \frac{m_j(W,\theta)}{\sigma_{Fj}(\theta)} \ \forall j \leq k, \tag{14.1}$$

and $\widetilde{m}_n(\theta) := (\widetilde{m}_{n1}(\theta), ..., \widetilde{m}_{nk}(\theta))'$. The corresponding population-normalized sample moment empirical process and sample second-central-moment empirical process are

$$\begin{aligned} \nu_n^m(\theta) &:= n^{1/2}(\widetilde{m}_n(\theta) - E_F\widetilde{m}_n(\theta)),\ \widehat{\sigma}^2_{Fnj}(\theta) := n^{-1}\sum_{i=1}^{n}(m_j(W_i,\theta) - E_F m_j(W,\theta))^2, \\ \nu_{nj}^{\sigma}(\theta) &:= n^{1/2}\left(\frac{\widehat{\sigma}^2_{Fnj}(\theta)}{\sigma^2_{Fj}(\theta)} - 1\right) = n^{-1/2}\sum_{i=1}^{n}[(\widetilde{m}_j(W_i,\theta) - E_F\widetilde{m}_j(W,\theta))^2 - 1] \ \forall j \leq k, \text{ and} \\ \nu_n(\theta) &:= \begin{pmatrix} \nu_n^m(\theta) \\ \nu_n^{\sigma}(\theta) \end{pmatrix}, \end{aligned} \tag{14.2}$$

where the superscripts m and σ denote mean and variance, respectively. Let $\nu_{nj}^m(\theta)$ and $\nu_{nj}^{\sigma}(\theta)$ denote the jth elements of $\nu_n^m(\theta)$ and $\nu_n^{\sigma}(\theta)$, respectively, for $j = 1, ..., k$. The variance matrix of $\nu_n(\theta)$ is $\Omega_{F+}(\theta)$, which is defined in (5.10).

The covariance kernel $\Omega_F(\theta,\theta')$ of $\nu_n(\theta)$ is defined as follows: for $\theta,\theta' \in \Theta$,

$$\Omega_F(\theta,\theta') := E_F\begin{pmatrix} \widetilde{m}(W,\theta) - E_F\widetilde{m}(W,\theta) \\ \widetilde{m}^{\sigma}(W,\theta) \end{pmatrix}\begin{pmatrix} \widetilde{m}(W,\theta') - E_F\widetilde{m}(W,\theta') \\ \widetilde{m}^{\sigma}(W,\theta') \end{pmatrix}' \in R^{2k\times 2k}, \tag{14.3}$$

where $\widetilde{m}(W,\theta)$ and $\widetilde{m}^{\sigma}(W,\theta)$ are defined in (5.10) and $E_F\widetilde{m}_j^{\sigma}(W,\theta) = 0$ for $j \leq k$, $\forall\theta \in \Theta$.

Assumption A.3. The empirical process $\nu_n(\cdot)$ is asymptotically ρ_F-equicontinuous on Θ uniformly in $F \in \mathcal{P}$.[44]

Assumption A.4. The covariance kernel $\Omega_F(\theta,\theta')$ satisfies: for all $F \in \mathcal{P}$,
$\lim_{\delta\to 0}\sup_{||(\theta_1,\theta_1')-(\theta_2,\theta_2')||<\delta}||\Omega_F(\theta_1,\theta_1') - \Omega_F(\theta_2,\theta_2')|| = 0.$

Assumption A.5. $E_F\widetilde{m}(W,\theta)$ is equicontinuous on Θ over $F \in \mathcal{P}$. That is, $\lim_{\delta\downarrow 0}\sup_{F\in\mathcal{P}}$ $\sup_{||\theta-\theta'||<\delta}||E_F\widetilde{m}(W,\theta) - E_F\widetilde{m}(W,\theta')|| = 0$.

In (4.19)–(4.24) and (5.7)–(5.12), the constants $\{\kappa_n\}_{n\geq 1}$ and $\{\tau_n\}_{n\geq 1}$ must satisfy:

[44] That is, $\lim_{\delta\to 0}\limsup_{n\to\infty}\sup_{F\in\mathcal{P}}P_F^*(\sup_{\rho_F(\theta,\theta')<\delta}||\nu_n(\theta) - \nu_n(\theta')||) = 0$, where P_F^* denotes outer probability and $\rho_F(\theta,\theta') := ||Var_F(\nu_n(\theta) - \nu_n(\theta'))||$.

Assumption A.6. (i) $\kappa_n \to \infty$ and (ii) $\tau_n \to \infty$.

For correct asymptotic level of $CI_{n,\Delta L}(\alpha)$, the constant κ_n in (5.15) must satisfy:

Assumption A.7. (i) $\kappa_n \to \infty$ and (ii) $\kappa_n/n^{1/2} \to 0$.

Let $\Delta_F(\theta) := \max_{j \leq k} \Delta_{Fj}(\theta)$. The set of minimizers of $\Delta_F(\theta)$ over Θ is $\Theta_{\min}(F) := \{\theta \in \Theta : \Delta_F(\theta) = \Delta_F^{\inf}\}$. For the lower-bound CI $CI_{n,\Delta L}(\alpha)$ only, we impose the following minorant condition on $\Theta_{\min}(F)$. It is analogous to the minorant conditions in CHT, BCS, and Bugni, Canay, and Shi (2017) for the identified set.

Assumption A.8. (i) For all $F \in \mathcal{P}$ and $\theta \in \Theta$, $\Delta_F(\theta) - \Delta_F^{\inf} \geq c \min\{\delta, \inf_{\overline{\theta} \in \Theta_{\min}(F)} ||\theta - \overline{\theta}||\}$ for constants $c, \delta > 0$, (ii) Θ is convex, and (iii) $E_F \widetilde{m}(W, \theta)$ is differentiable in θ for all $F \in \mathcal{P}$ and $\{\widetilde{M}_F(\theta) := (\partial/\partial\theta') E_F \widetilde{m}(W, \theta) : F \in \mathcal{P}\}$ is equicontinuous, i.e., $\lim_{\delta \to 0} \sup_{F \in \mathcal{P}} \sup_{(\theta, \overline{\theta}) : ||\theta - \overline{\theta}|| \leq \delta} ||\widetilde{M}_F(\theta) - \widetilde{M}_F(\overline{\theta})|| = 0$.

15 Asymptotic Distribution of the SPUR Test Statistic

The EGMS critical value for the SPUR1 test defined above is constructed based on the asymptotic distribution of $S_n(\theta_0)$ under drifting sequences of null distributions $\{F_n\}_{n \geq 1}$ for which $\theta_0 \in \Theta_I^{MR}(F_n)$ for $n \geq 1$. In this section, we establish this asymptotic distribution. For power properties, we also establish the asymptotic distribution under local and global alternatives as well.

One obtains a CS for $\theta \in \Theta_I^{MR}(F)$ by inverting tests based on $S_n(\theta_0)$ for $\theta_0 \in \Theta$. To obtain uniform asymptotic coverage probability results, we need the asymptotic distribution of $S_n(\theta_n)$ under drifting sequences of null values $\{\theta_n\}_{n \geq 1}$ and distributions $\{F_n\}_{n \geq 1}$. For this reason, in the results below, we consider the statistic

$$S_n := S_n(\theta_n) \text{ for testing } H_0 : \theta_n \in \Theta_I^{MR}(F_n). \tag{15.1}$$

The results cover models that may be correctly specified or misspecified. The form of the asymptotic null distribution is important in order to understand the definition of the EGMS critical value given in Section 4.4 above.

The proofs of the asymptotic level results for SPUR tests and CS's show that it suffices to determine the asymptotic null rejection probabilities of tests under sequences or subsequences of distributions F_n that satisfy certain conditions. These conditions are Assumptions C.1, C.3, C.4, C.7, and C.8 introduced below, which depend only on deterministic quantities and can be made to hold for certain subsequences using the fact that any sequence in a compact metric set has a convergent subsequence. For this reason, we do not provide sufficient conditions for these conditions and these conditions do not appear in the statements of the asymptotic level results in Theorem 4.1.

15.1 High-Level Convergence Assumptions

The components $T_n(\theta)$ and A_n of $S_n(\theta)$ in (4.12) are centered and scaled such that they have asymptotic distributions. We obtain the asymptotic distribution of A_n using a similar approach to that in BCS. The results are also closely related to the asymptotic distribution results for the supremum of a moment inequality objective function in CHT, Theorems 4.2 and 5.2. The results given below differ from these results in that they allow for model misspecification.

Let $R_{[\pm\infty]} := R \cup \{+\infty, -\infty\}$. As in BCS, for any $x_1, x_2 \in R^{a_*}_{[\pm\infty]}$ for some positive integer a_*, let $d(x_1, x_2) := (\sum_{j=1}^{a_*}(\Phi(x_{1,j}) - \Phi(x_{2,j}))^2)^{1/2}$, where $\Phi : R_{[\pm\infty]} \to [0,1]$, $\Phi(y)$ is the standard normal distribution function at y for $y \in R$, $\Phi(-\infty) := 0$, and $\Phi(\infty) := 1$. The space $(R^{a_*}_{[\pm\infty]}, d)$ is a compact metric space. Convergence in $(R^{a_*}_{[\pm\infty]}, d)$ to a point in R^{a_*} implies convergence under the Euclidean norm. Let $\mathcal{S}(\Theta \times R^{2k}_{[\pm\infty]})$ denote the space of non-empty compact subsets of the metric space $(\Theta \times R^{2k}_{[\pm\infty]}, d)$, where d is defined with $a_* = d_\theta + 2k$. Let $\Rightarrow$ denote weak convergence of a sequence of stochastic processes in the sense of van der Vaart and Wellner (1996). Let $\to_H$ denote convergence in Hausdorff distance (under d) for elements of $\mathcal{S}(\Theta \times R^{2k}_{[\pm\infty]})$. For any $b, \ell, m \in R^k$, including $b_n, b^*, \widetilde{b}, \ell_n$ which arise below, let b_j, ℓ_j, m_j denote the jth elements of b, ℓ, m, respectively.

To obtain the asymptotic distribution of A_n, we use the following sets:

$$\Lambda_{n,F} := \Big\{(\theta, b, \ell) \in \Theta \times R^{2k} : b_j = n^{1/2}([E_F \widetilde{m}_j(W,\theta)]_- - r_F^{\inf}), \ell_j = n^{1/2} E_F \widetilde{m}_j(W,\theta)\ \forall j \leq k.\Big\} \tag{15.2}$$

for $n \geq 1$. For $(\theta, b, \ell) \in \Lambda_{n,F}$, b_j is the difference between the magnitude of violation of the jth moment at θ, $[E_F \widetilde{m}_j(W,\theta)]_-$, and the minimal relaxation, $r_F^{\inf}$, scaled by $n^{1/2}$, and ℓ_j is the jth moment at θ scaled by $n^{1/2}$. The quantities b_j and ℓ_j can be positive, negative, or zero.

For $\eta > 0$, define

$$\Theta_I^\eta(F) := \{\theta \in \Theta : \max_{j \leq k}[E_F \widetilde{m}_j(W,\theta) + r_F^{\inf}]_- \leq \eta/n^{1/2}\}. \tag{15.3}$$

The set $\Theta_I^\eta(F)$ is an $\eta/n^{1/2}$-expansion of the MR-identified set $\Theta_I^{MR}(F)$. It depends on n, but this is suppressed. One can also write $\Theta_I^\eta(F)$ as $\{\theta \in \Theta : \max_{j \leq k}[E_F \widetilde{m}_j(W,\theta)]_- - r_F^{\inf} \leq \eta/n^{1/2}\}$.[45]

For $\eta > 0$, define Λ^η_{n,F_n} as in (15.2) with $\Theta_I^\eta(F_n)$ in place of Θ. By definition, $\Lambda^\eta_{n,F_n} \subset \Lambda_{n,F_n}$.

We employ the following "convergence" assumptions that apply to a drifting sequence of null

[45] This holds because for $b, c \geq 0$, $[a+b]_- \leq c$ if and only if $[a]_- - b \leq c$.

values $\{\theta_n\}_{n\geq 1}$, as in (15.1), and distributions $\{F_n\}_{n\geq 1}$.

Assumption C.1. $\theta_n \to \theta_\infty$ for some $\theta_\infty \in \Theta$.

Assumption C.2. $n^{1/2} E_{F_n} \widetilde{m}_j(W, \theta_n) \to \ell_{j\infty}$ for some $\ell_{j\infty} \in R_{[\pm\infty]}$ $\forall j \leq k$.

Assumption C.3. $n^{1/2}(E_{F_n} \widetilde{m}_j(W, \theta_n) + r^{\inf}_{F_n}) \to h_{j\infty}$ for some $h_{j\infty} \in R_{[\pm\infty]}$ $\forall j \leq k$.

Assumption C.4. $\sup_{\theta\in\Theta} ||E_{F_n} \widetilde{m}(W, \theta) - \widetilde{m}(\theta)|| \to 0$ for some nonrandom bounded continuous R^k-valued function $\widetilde{m}(\cdot)$ on Θ.

Assumption C.5. $\nu_n(\cdot) := (\nu_n^m(\cdot)', \nu_n^\sigma(\cdot)')' \Rightarrow G(\cdot) := (G^m(\cdot)', G^\sigma(\cdot)')'$ as $n \to \infty$, where $\{G(\theta) : \theta \in \Theta\}$ is a mean zero R^{2k}-valued Gaussian process with some covariance kernel $\Omega_\infty(\cdot, \cdot)$, bounded continuous sample paths a.s., and $G^m(\theta), G^\sigma(\theta) \in R^k$.

Assumption C.6. $\widehat{\Omega}_n(\theta_n) \to_p \Omega_\infty$ for some $\Omega_\infty \in \Psi$.

Assumption C.7. $\Lambda_{n,F_n} \to_H \Lambda$ for some non-empty set $\Lambda \in \mathcal{S}(\Theta \times R^{2k}_{[\pm\infty]})$.

Assumption C.8. $\Lambda^{\eta_n}_{n,F_n} \to_H \Lambda_I$ for some non-empty set $\Lambda_I \in \mathcal{S}(\Theta \times R^{2k}_{[\pm\infty]})$, where $\{\eta_n\}_{n\geq 1}$ is a sequence of positive constants for which $\eta_n \to \infty$.

All of the limit quantities above, i.e., θ_∞, $\{\ell_{j\infty}\}_{j\leq k}$, etc., depend on $\{\theta_n\}_{n\geq 1}$ and $\{F_n\}_{n\geq 1}$. Assumptions A.1–A.4, C.1, and uniform convergence of the covariance kernel $\Omega_{F_n}(\cdot, \cdot)$ to a continuous limit function $\Omega_\infty(\cdot, \cdot)$ are sufficient conditions for Assumptions C.5 and C.6, with Ω_∞ in Assumption C.6 equal to the upper left $k \times k$ submatrix of $\Omega_\infty(\theta_\infty, \theta_\infty)$, see Lemma 20.1 in online Appendix B. Assumption C.7 is a generalization of assumption (iii) in Theorem 3.1 of BCS to allow for model misspecification. Assumption C.8 is used to simplify the asymptotic distribution of S_n.

Let

$$\widetilde{m}_{j\infty} = \widetilde{m}_j(\theta_\infty) \text{ for } j \leq k \text{ and } \widetilde{m}(\theta) = (\widetilde{m}_1(\theta), ..., \widetilde{m}_k(\theta))'. \tag{15.4}$$

The values $\ell_{j\infty}$, $h_{j\infty}$, and $\widetilde{m}_{j\infty}$ in Assumptions C.2 and C.3 and (15.4) have the following properties.

Lemma 15.1 (a) *Under Assumption* C.3, *if* $\theta_n \in \Theta_I^{MR}(F_n)$ *for all* n *large, then* $h_{j\infty} \geq 0$ $\forall j \leq k$, (b) *under Assumptions* C.2 *and* C.3, $\ell_{j\infty} \leq h_{j\infty}$ $\forall j \leq k$, (c) *under Assumptions* C.1, C.2, *and* C.4, $|\widetilde{m}_{j\infty}| \leq |\ell_{j\infty}|$ *and if* $|\ell_{j\infty}| < \infty$, *then* $\widetilde{m}_{j\infty} = 0$ $\forall j \leq k$, *and* (d) *under Assumptions* C.1–C.4, *if* $\theta_n \in \Theta_I^{MR}(F_n)$ *for all* n *large and the model is correctly specified, then* $h_{j\infty} = \ell_{j\infty}$ *and* $h_{j\infty}, \ell_{j\infty}, \widetilde{m}_{j\infty} \geq 0$ $\forall j \leq k$.

Comment. By Lemma 15.1(a), under the null hypothesis H_0 in (4.1), $h_{j\infty} \geq 0$ $\forall j \leq k$.

The elements (θ, b, ℓ) of Λ in Assumption C.7 have the following properties.

Lemma 15.2 *Under* $\{F_n\}_{n\geq 1}$, (a) $\max_{j\leq k} b_{nj}(\theta) \geq 0 \ \forall \theta \in \Theta$, $\forall n \geq 1$, *where* $b_{nj}(\theta) := n^{1/2}([E_{F_n}\widetilde{m}_j(W,\theta)]_- - r_{F_n}^{\inf})$, (b) $\forall(\theta, b, \ell) \in \Lambda$, $\max_{j\leq k} b_j \geq 0$ *provided Assumption* C.7 *holds,* (c) $\exists \widetilde{\theta}_n \in \Theta$ *with* $\max_{j\leq k} b_{nj}(\widetilde{\theta}_n) = 0 \ \forall n \geq 1$ *provided Assumption* A.0 *holds,* (d) $\exists(\widetilde{\theta}, \widetilde{b}, \widetilde{\ell}) \in \Lambda$ *with* $\max_{j\leq k} \widetilde{b}_j = 0$ *provided Assumptions* A.0 *and* C.7 *hold, and* (e) $\forall(\theta, b, \ell) \in \Lambda$, $|\ell_j| < \infty$ *implies* $\widetilde{m}_j(\theta) = 0 \ \forall j \leq k$ *provided Assumptions* C.4 *and* C.7 *hold.*

Comment. Lemma 15.2(a)–(d) are used to show that the asymptotic distribution of A_n is in R a.s. Lemma 15.2(a) and (b) are key properties that are utilized when constructing a stochastic lower bound on the asymptotic distribution of A_n. Lemma 15.2(c) implies that the MR-identified set is non-empty under Assumption A.0 for all $n \geq 1$. Lemma 15.2(e) is used to show that the asymptotic distribution of A_n simplifies somewhat in some scenarios.

Next, we state assumptions that specify whether $\{\theta_n\}_{n\geq 1}$ is a sequence of parameter values (i) in the MR-identified set or $n^{-1/2}$-local to the MR-identified set, i.e., a null or $n^{-1/2}$-local alternative (NLA) sequence, or (ii) non-$n^{-1/2}$-local to the MR-identified set, which yields a consistent alternative (CA) sequence.

Assumption NLA. $\min_{j\leq k} h_{j\infty} > -\infty$.

Assumption CA. $\min_{j\leq k} h_{j\infty} = -\infty$.

Two alternative sufficient conditions for Assumption NLA are: Assumption N: $\theta_n \in \Theta_I^{MR}(F_n)$ $\forall n \geq 1$, and Assumption LA: The null values $\{\theta_n\}_{n\geq 1}$ and distributions $\{F_n\}_{n\geq 1}$ satisfy: (i) $||\theta_n - \theta_{In}|| = O(n^{-1/2})$ for some sequence $\{\theta_{In} \in \Theta_I^{MR}(F_n)\}_{n\geq 1}$, (ii) $n^{1/2}(E_{F_n}\widetilde{m}_j(W,\theta_{In}) + r_{F_n}^{\inf}) \to h_{Ij\infty}$ for some $h_{Ij\infty} \in R_{[\pm\infty]}$ $\forall j \leq k$, and (iii) $E_F\widetilde{m}(W,\theta)$ is Lipschitz on Θ uniformly over $\mathcal{P}$, i.e., there exists a constant $K < \infty$ such that $||E_F\widetilde{m}(W,\theta_1) - E_F\widetilde{m}(W,\theta_2)|| \leq K||\theta_1 - \theta_2|| \ \forall \theta_1, \theta_2 \in \Theta$, $\forall F \in \mathcal{P}$. Under Assumption N, $\min_{j\leq k} h_{j\infty} \geq 0$. A "fixed alternative" (FA) sufficient condition for Assumption CA is: Assumption FA: (i) $(\theta_n, F_n) = (\theta_*, F_*) \in \Theta \times \mathcal{P}$ does not depend on $n \geq 1$ and (ii) $E_{F_*}\widetilde{m}_j(W,\theta_*) + r_{F_*}^{\inf} < 0$ for some $j \leq k$.[46]

15.2 Asymptotic Distribution of S_n

For notational simplicity, we use the following conventions: for any scalars $\nu \in R$ and $c = \pm\infty$, where ν may be deterministic or random and c is deterministic, we let

$$\nu + c = c, \ [\nu + c]_- - [c]_- = 0 \text{ when } c = +\infty, \text{ and } [\nu + c]_- - [c]_- = -\nu \text{ when } c = -\infty.^{47} \quad (15.5)$$

[46] The sufficiency of these conditions is established in Section 16 in online Appendix B.

[47] This notation is motivated by the fact that for finite deterministic scalar constants ν and c, for ν fixed, $\lim_{c\to\pm\infty}(\nu + c) = \lim_{c\to\pm\infty} c$, $\lim_{c\to+\infty}([\nu + c]_- - [c]_-) = 0$, and $\lim_{c\to-\infty}([\nu + c]_- - [c]_-) = -\nu$, and analogous convergence in probability results hold when ν is random.

Let $G_j^m(\theta)$, $G_j^\sigma(\theta)$, $\nu_{nj}^m(\theta)$, and $\nu_{nj}^\sigma(\theta)$ denote the jth elements of $G^m(\theta)$, $G^\sigma(\theta)$, $\nu_n^m(\theta)$, and $\nu_n^\sigma(\theta)$, respectively. Let

$$\begin{aligned} G_{j\infty}^m &:= G_j^m(\theta_\infty),\ G_{j\infty}^\sigma := G_j^\sigma(\theta_\infty),\ G_{j\infty}^{m\sigma} := G_{j\infty}^m - \frac{1}{2}\widetilde{m}_{j\infty}G_{j\infty}^\sigma, \\ G_j^{m\sigma}(\theta) &:= G_j^m(\theta) - \frac{1}{2}\widetilde{m}_j(\theta)G_j^\sigma(\theta), \text{ and} \\ \nu_{nj}^{m\sigma}(\theta) &:= \nu_{nj}^m(\theta) - \frac{1}{2}\widetilde{m}_j(\theta)\nu_{nj}^\sigma(\theta) \end{aligned} \tag{15.6}$$

for $j \leq k$ and θ_∞ as in Assumption C.1. Define

$$T_{j\infty} := G_{j_\infty}^{m\sigma} + h_{j\infty} \text{ for } j \leq k \text{ and } T_\infty := (T_{1\infty}, ..., T_{k\infty})', \tag{15.7}$$

where we employ the notational convention in (15.5). Thus, we have: $T_{j\infty} = \infty$ if $\ell_{j\infty} = \infty$ (because $h_{j\infty} \geq \ell_{j\infty} = \infty$ by Lemma 15.1(c)), $T_{j\infty} = G_{j_\infty}^m + h_{j\infty}$ if $|\ell_{j\infty}| < \infty$ (because $|\ell_{j\infty}| < \infty$ implies that $\widetilde{m}_{j\infty} = 0$ by Lemma 15.1(c)), and $T_{j\infty}$ is finite and as in (15.7) with $\widetilde{m}_{j\infty} \neq 0$ if $\ell_{j\infty} = -\infty$ and $|h_{j\infty}| < \infty$. As noted above, under H_0, $h_{j\infty} \geq 0$ for $j \leq k$.

If the model is correctly specified and $\theta_n \in \Theta_I^{MR}(F_n)$ for n large, then T_{j_∞} simplifies to

$$T_{j_\infty} = G_{j_\infty}^m + \ell_{j\infty} \tag{15.8}$$

because, in this case, $h_{j\infty} = \ell_{j\infty}$ (by Lemma 15.1(d)), $\ell_{j\infty} \in [-\infty, 0)$ cannot occur (because $\ell_{j\infty} \geq 0$ by Lemma 15.1(d)), $|\ell_{j\infty}| < \infty$ implies that $\widetilde{m}_{j\infty} = 0$ (by Lemma 15.1(c)), and $\ell_{j\infty}$ $(= h_{j\infty}) = \infty$ implies $G_{j_\infty}^m - (\widetilde{m}_{j\infty}/2)G_{j_\infty}^\sigma + h_{j\infty} = \infty = G_{j_\infty}^m + \ell_{j\infty}$ (by the notational convention in (15.5)).

The following quantities arise with the asymptotic distribution of A_n:

$$\begin{aligned} A_n(\Lambda_{n,F_n}) &:= \inf_{(\theta,b,\ell)\in\Lambda_{n,F_n}} \max_{j\leq k} \left([\nu_{nj}^{m\sigma}(\theta) + \ell_j]_- - [\ell_j]_- + b_j\right), \\ A_\infty &:= A_\infty(\Lambda), \text{ and } A_{I\infty} := A_\infty(\Lambda_I), \text{ where} \\ A_\infty(\Lambda) &:= \inf_{(\theta,b,\ell)\in\Lambda} \max_{j\leq k} \left([G_j^{m\sigma}(\theta) + \ell_j]_- - [\ell_j]_- + b_j\right). \end{aligned} \tag{15.9}$$

We show that $A_n = A_n(\Lambda_{n,F_n}) + o_p(1) \to_d A_\infty$ as $n \to \infty$ in Lemma 17.1 in online Appendix B and Theorem 15.3 below. The term in parentheses in the definition of $A_\infty(\Lambda)$ equals b_j when $\ell_j = +\infty$ (because $[\nu + c]_- - [c]_- = 0$ for $\nu \in R$ and $c = +\infty$ by definition in (15.5)); equals $[G_j^m(\theta) + \ell_j]_- - [\ell_j]_- + b_j$ when $|\ell_j| < \infty$ (because $|\ell_j| < \infty$ implies $\widetilde{m}_j(\theta) = 0$ for $(\theta, b, \ell) \in \Lambda$ by Lemma 15.2(e)); and equals $-G_j^{m\sigma}(\theta) + b_j$ when $\ell_j = -\infty$ (because $[\nu + c]_- - [c]_- = -\nu$ for $\nu \in R$ and $c = -\infty$ by definition in (15.5)).

The asymptotic distribution of the SPUR statistic S_n under the null hypothesis and $n^{-1/2}$-local alternatives is the distribution of

$$S_\infty := S(T_\infty + A_\infty 1_k, \Omega_\infty), \text{ which is equal to } S_{I\infty} := S(T_\infty + A_{I\infty} 1_k, \Omega_\infty) \tag{15.10}$$

under Assumption C.8.

Theorem 15.3 (a) *Under* $\{F_n\}_{n\geq 1}$ *and Assumptions* C.1 *and* C.3–C.5, $T_n(\theta_n) \to_d T_\infty$,

(b) *under* $\{F_n\}_{n\geq 1}$ *and Assumptions* A.0, C.4, C.5, *and* C.7, $A_n \to_d A_\infty$,

(c) *under Assumptions* A.0 *and* C.7, $A_\infty \in R$ *a.s.,*

(d) *under Assumptions* C.1 *and* C.3–C.5 *and* NLA, $T_{j\infty} > -\infty$ *a.s.* $\forall j \leq k$,

(e) *under* $\{F_n\}_{n\geq 1}$ *and Assumptions* A.0, C.1 *and* C.3–C.7, NLA, *and* S.1(iii), $S_n \to_d S_\infty$,

(f) *under Assumptions* A.0, C.1, *and* C.3–C.8, $A_\infty = A_{I\infty}$ *a.s. and* $S_\infty = S_{I\infty}$ *a.s.,*

(g) *under Assumptions* C.1 *and* C.3–C.5, *and* CA, $T_{j\infty} = -\infty$ *a.s. for some* $j \leq k$,

(h) *under* $\{F_n\}_{n\geq 1}$ *and Assumptions* A.0, C.1 *and* C.3–C.7, CA, S.1(iii), S.2, *and* S.3, $S_n \to_p \infty$, *and*

(i) *the convergence results in parts* (a)–(e) *hold jointly.*

Comments. (i). Under correct model specification, $r_F^{\inf} = 0$, $A_n = n^{1/2}\widehat{r}_n^{\inf}$ (see (4.12)), $n^{1/2}\widehat{r}_n^{\inf}$ is the same as the model specification test statistic in BCS when their function $S(m, \Omega)$ equals $\max_{j\leq k}[m_j]_-$, and the asymptotic distribution of A_n given in Theorem 15.3(b) can be shown to reduce to the same distribution as the asymptotic null distribution of the specification test statistic given in Theorem 3.1 of BCS. In addition, in the correctly specified case, $A_n = n^{1/2}\widehat{r}_n^{\inf}$ equals CHT's statistic $\inf_{\theta\in\Theta} a_n Q_n(\theta)$ for moment inequality models when $Q_n(\theta)$ is the "max" sample objective function defined by $\max_{j\leq k}[\widehat{m}_{nj}(\theta)]_-$ (and $a_n = n^{1/2}$) and CHT provide the asymptotic distribution of $\inf_{\theta\in\Theta} a_n Q_n(\theta)$ under correct specification and for a fixed true distribution (rather than a drifting sequence of distributions as in Theorem 15.3(b)).[48] Theorem 15.3(b) extends these results to allow for model misspecification.

(ii). The asymptotic distributions in Theorem 15.3 depend on the localization parameters $h_{j\infty}$ and $\ell_{j\infty}$, which are not consistently estimable, and $\widetilde{m}_{j\infty}$, which is consistently estimable. Under

[48] The asymptotic distribution of Chernozhukov, Hong, and Tamer's (2007) statistic $\inf_{\theta\in\Theta} a_n Q_n(\theta)$ is given in their Theorems 4.2(2) and 5.2(2) by the difference between $\mathcal{C}$ in their (4.8) and (4.7) or the difference between $\mathcal{C}(\theta)$ in their (5.6) and (5.5). Their definition of the identified set on p. 1265 assumes correct model specification, as do their equation (4.5) and Assumption M.2. The function $\xi(\theta)$ in their Theorem 4.2 only takes values of $-\infty$ or 0 due to their asymptotics being for a fixed true distribution, as opposed to a drifting sequence of distributions. Because Chernozhukov, Hong, and Tamer (2007) consider "$\leq$" inequalities, whereas the present paper considers "$\geq$" inequalities, the sample moments enter the statistics with different signs in the two papers.

the null hypothesis H_0 in (4.1), $h_{j\infty} \geq 0$ for all $j \leq k$. The asymptotic distribution also depends on the (b_j, ℓ_j) values, which appear in the limit sets Λ and Λ_I, and are not consistently estimable. For the purposes of inference (i.e., obtaining a critical value), one needs a stochastic lower bound on the distribution of the vector sum $T_\infty + A_\infty 1_k$ for the case when $h_{j\infty} \geq 0$ for all $j \leq k$.

(iii). Theorem 15.3(c) is important because it implies that adding A_∞ to $T_{j\infty}$ cannot result in adding $+\infty$ to $-\infty$ or $-\infty$ to $+\infty$.

(iv). Theorem 15.3(f) is important because it implies that parameters $(\theta, b, \ell) \in \Lambda \backslash \Lambda_I$ do not contribute to the infimum in A_∞. This means that when constructing a critical value for a test based on S_n one only needs to find a lower bound on $A_{I\infty}$.

(v). The stochastic process $G_j^\sigma(\cdot)$ enters S_∞ (through $G_j^{m\sigma}(\cdot)$). Thus, the asymptotic distribution of S_n depends on the randomness due to the estimation of the standard deviation of the jth sample moment by $\widehat{\sigma}_{nj}(\theta)$ for $j \leq k$. Under correct model specification, this is not the case.

For any subsequence $\{q_n\}_{n\geq 1}$ of $\{n\}_{n\geq 1}$, Theorem 15.3 and its proof hold with q_n in place of n throughout, including the assumptions. To prove Theorem 15.3(b), we use a similar proof to the proof of Theorem 3.1 of BCS with $S(m, \Omega) = \max_{j\leq k}[m_j]_-$ in their proof. The statistic $A_n(\Lambda_{n,F_n})$ depends on $b_{nj}(\theta) := n^{1/2}[E_{F_n}\widetilde{m}_j(W, \theta)]_- - n^{1/2} r_{F_n}^{\inf}$, $\ell_{nj}(\theta) := n^{1/2} E_{F_n}\widetilde{m}_j(W, \theta)$, $\nu_{nj}^m(\theta)$, and $\nu_{nj}^\sigma(\theta)$, whereas the statistic in BCS only depends on $\ell_{nj}(\theta)$ and $\nu_{nj}^m(\theta)$.

The asymptotic distribution $S_{I\infty}$ of the SPUR statistic under H_0 explains the form of the critical value for the SPUR1 test in Section 4.4. The Gaussian quantities $G_{j\infty}^{m\sigma}$ and $G_j^{m\sigma}(\cdot)$ in (15.7) and (15.9) are approximated by the bootstrap quantities $\widehat{\nu}_{njb}^*(\theta)$ and $\widehat{\nu}_{njb}^*(\cdot)$ in (4.17). The constant $h_{j\infty}$ in (15.7) is nonnegative under H_0 and is lower bounded by the GMS quantity $\varphi(\xi_{nj}(\theta))$ in (4.19). The random quantities $[G_j^{m\sigma}(\theta) + \ell_j]_- - [\ell_j]_-$ for $(\theta, b, \ell) \in \Lambda_I$ in (15.9) are lower bounded by $\widehat{\chi}_{nj,b}^*(\theta)$ for $\theta \in \widehat{\Theta}_n$ defined in (4.21), which appears in $A_{n,b}^*$ in (4.25). The infimum is taken over $\widehat{\Theta}_n$ in $A_{n,b}^*$ because Λ_I depends on the limit of $\Theta_I^{\eta_n}(F_n)$ and it is shown that $\widehat{\Theta}_n \supset \Theta_I^{\eta_n}(F_n)$ wp$\to$1. The elements b_j in the vectors $b = (b_1, ..., b_k)'$ for $(\theta, b, \ell) \in \Lambda$ in (15.9) are lower bounded by $\widehat{b}_{nj}(\theta)$ in general and by the better GMS-type lower bound $\varphi(\xi_{nj}^A(\theta))$ when $j = j_1$ is such that $b_{j_1} \geq 0$ of which there is at least one by Lemma 15.2(b). We show that $j_1 \in \widehat{J}_{nB}(\theta)$ wp$\to$1, so $A_{n,b}^*$ is defined to allow j_1 to be any of the values in $\widehat{J}_{nB}(\theta)$ and a minimum over $j_1 \in \widehat{J}_{nB}(\theta)$ is taken to get a lower bound. Imposing the property in Lemma 15.2(b) is important because otherwise the EGMS critical value would slowly diverge in probability to ∞ as $n \to \infty$.

16 Lemma 16.1 and Proofs of Lemmas 15.1, 15.2, and 16.1

The following is a sufficient condition for Assumption NLA, which first appears in Section 15.1.

Assumption LA. The null values $\{\theta_n\}_{n\geq 1}$ and distributions $\{F_n\}_{n\geq 1}$ satisfy: (i) $||\theta_n - \theta_{In}|| = O(n^{-1/2})$ for some sequence $\{\theta_{In} \in \Theta_I^{MR}(F_n)\}_{n\geq 1}$, (ii) $n^{1/2}(E_{F_n}\widetilde{m}_j(W,\theta_{In})+r_{F_n}^{\inf}) \to h_{Ij\infty}$ for some $h_{Ij\infty} \in R_{[\pm\infty]}$ $\forall j \leq k$, and (iii) $E_F\widetilde{m}(W,\theta)$ is Lipschitz on Θ uniformly over $\mathcal{P}$, i.e., there exists a constant $K < \infty$ such that $||E_F\widetilde{m}(W,\theta_1) - E_F\widetilde{m}(W,\theta_2)|| \leq K||\theta_1 - \theta_2||$ $\forall \theta_1, \theta_2 \in \Theta$, $\forall F \in \mathcal{P}$.

Under Assumption LA, $\{\theta_n\}_{n\geq 1}$ is a sequence of $n^{-1/2}$-local alternatives to the null hypothesis $\forall n \geq 1$. Assumption LA(ii) is the same as Assumption C.3 with $\{\theta_n\}_{n\geq 1}$ replaced by some sequence $\{\theta_{In}\}_{n\geq 1}$ in the MR-identified set(s). Hence, by Lemma 15.1(a), $h_{Ij\infty} \geq 0$ $\forall j \leq k$.

A sufficient condition for Assumption CA is the following fixed alternative assumption.

Assumption FA. The null values $\{\theta_n\}_{n\geq 1}$ and distributions $\{F_n\}_{n\geq 1}$ satisfy: (i) The distributions $F_n = F_* \in \mathcal{P}$ and the null values $\theta_n = \theta_* \in \Theta$ do not depend on $n \geq 1$ and (ii) $E_{F_*}\widetilde{m}_j(W,\theta_*)+r_{F_*}^{\inf} < 0$ for some $j \leq k$.

Lemma 16.1 *Under Assumption* C.3, (a) *Assumption* N *implies Assumption* NLA, (b) *Assumption* LA *implies Assumption* NLA, *and* (c) *Assumption* FA *implies Assumption* CA.

Proof of Lemma 15.1. Part (a) holds because $r_{F_n}^{\inf} \geq 0$ by its definition in (3.5). The first result in part (b) holds because $n^{1/2} \geq 1$. The second result in part (b) holds because $|\ell_{j\infty}| < \infty$ implies $n^{1/2}E_{F_n}\widetilde{m}_j(W,\theta_n) = O(1)$, which implies that $\widetilde{m}_{j\infty} := \widetilde{m}_j(\theta_\infty) = \lim_{n\to\infty} E_{F_n}\widetilde{m}_j(W,\theta_n) = 0$, using Assumptions C.1, C.2, and C.4.

Now, we prove part (c). If $\theta \in \Theta_I^{MR}(F)$, then $r_F(\theta) = r_F^{\inf}$ (by the definition of $\Theta_I^{MR}(F)$ in (3.6)), $r_{Fj}(\theta) \leq r_F^{\inf}$ $\forall j \leq k$ (by the definition of $r_{Fj}(\theta)$ in (3.5)), and $r_{Fj}(\theta) = r_F^{\inf}$ for some $j \leq k$. In consequence,

$$\begin{aligned} 0 &= \max_{j\leq k}(r_{Fj}(\theta) - r_F^{\inf}) = \max_{j\leq k}(\max\{-E_F\widetilde{m}_j(W,\theta),0\} - r_F^{\inf}) \\ &\geq \max_{j\leq k}(-E_F\widetilde{m}_j(W,\theta) - r_F^{\inf}) = -\min_{j\leq k}(E_F\widetilde{m}_j(W,\theta) + r_F^{\inf}), \end{aligned} \tag{16.1}$$

where the second equality holds by the definition of $r_{Fj}(\theta)$ and the inequality is trivial.

Using (16.1), if $\theta_n \in \Theta_I^{MR}(F_n)$ for n large, then

$$\begin{aligned} 0 &\leq \liminf_{n\to\infty} \min_{j\leq k} n^{1/2}(E_{F_n}\widetilde{m}_j(W,\theta_n) + r_{F_n}^{\inf}) \\ &= \min_{j\leq k} \liminf_{n\to\infty} n^{1/2}(E_{F_n}\widetilde{m}_j(W,\theta_n) + r_{F_n}^{\inf}) = \min_{j\leq k} h_{j\infty}, \end{aligned} \tag{16.2}$$

where the first equality holds by a subsequence argument and the second equality uses Assumption C.3. This establishes part (c).

Lastly, we prove part (d). If $\theta \in \Theta_I^{MR}(F)$ and the model is correctly specified, then

$$r_F^{\inf} = \max_{j \leq k} r_{Fj}(\theta) = \max_{j \leq k} \max\{-E_F \widetilde{m}(W, \theta), 0\} = 0, \tag{16.3}$$

where the first two equalities hold by the definitions of $r_F^{\inf}$ and $r_{Fj}(\theta)$ in (3.5) and the last equality holds because $E_F \widetilde{m}(W, \theta) \geq 0_k \ \forall \theta \in \Theta_I^{MR}(F)$ by correct model specification, see (3.4).

Equation (16.3) implies that under correct model specification, if $\theta_n \in \Theta_I^{MR}(F_n)$ for all n large, then

$$h_{j\infty} = \lim n^{1/2}(E_{F_n} \widetilde{m}_j(W, \theta_n) + r_{F_n}^{\inf}) = \lim n^{1/2} E_{F_n} \widetilde{m}_j(W, \theta_n) = \ell_{j\infty} \ \forall j \leq k. \tag{16.4}$$

We have $h_{j\infty}, \ell_{j\infty}, \widetilde{m}_{j\infty} \geq 0$ under correct model specification when $\theta_n \in \Theta_I^{MR}(F_n)$ for all n large, because the moment inequalities all hold at $\theta_n \in \Theta_I^{MR}(F_n)$, i.e., $E_{F_n} \widetilde{m}_j(W, \theta_n) \geq 0$, under correct model specification. This completes the proof of part (d). □

Proof of Lemma 15.2. Because $r_F^{\inf} := \inf_{\theta \in \Theta} \max_{j \leq k} r_{Fj}(\theta)$ for all F and $\theta \in \Theta$, see (3.5), we have

$$\max_{j \leq k} (r_{Fj}(\theta) - r_F^{\inf}) \geq 0, \tag{16.5}$$

which establishes part (a).

Any $(\theta, b, \ell) \in \Lambda$ is the limit of some sequence $(\theta_n, b_n, \ell_n) \in \Lambda_{n,F_n}$ because $\Lambda_{n,F_n} \to_H \Lambda$ by Assumption C.7. That is, $b_n \to b$ and $\max_{j \leq k} b_{nj} \to \max_{j \leq k} b_j$. This and (16.5) applied with $(\theta, F) = (\theta_n, F_n)$ give

$$0 \leq \max_{j \leq k} n^{1/2}(r_{F_n j}(\theta_n) - r_{F_n}^{\inf}) = \max_{j \leq k} b_{nj} \to \max_{j \leq k} b_j, \tag{16.6}$$

which proves part (b) of the lemma.

Next, we prove part (c). The function $r_{F_n}(\theta) - r_{F_n}^{\inf}$ is lower semi-continuous on Θ (since $E_F \widetilde{m}_j(W, \theta)$ is upper semi-continuous on Θ by Assumption A.0(ii)) and $[x]_- := \max\{-x, 0\}$, Θ is compact by Assumption A.0(i), and a lower semi-continuous function on a compact set achieves its infimum. Hence, there exists $\widetilde{\theta}_n \in \Theta$ such that $r_F(\widetilde{\theta}_n) = r_F^{\inf} \ \forall n \geq 1$, which establishes part (c).

For part (d), let $(\widetilde{\theta}_n, \widetilde{b}_n, \widetilde{\ell}_n) \in \Lambda_{n,F_n}$ be such that $\widetilde{\theta}_n \in \Theta_I^{MR}(F_n) \ \forall n \geq 1$. Such $(\widetilde{\theta}_n, \widetilde{b}_n, \widetilde{\ell}_n)$ exist because $\Theta_I^{MR}(F_n)$ is non-empty $\forall n \geq 1$ by part (c). There exists a subsequence $\{q_n\}_{n \geq 1}$ of $\{n\}_{n \geq 1}$ and a $(\widetilde{\theta}, \widetilde{b}, \widetilde{\ell}) \in \Theta \times R_{[\pm\infty]}^{2k}$ such that $d((\widetilde{\theta}_{q_n}, \widetilde{b}_{q_n}, \widetilde{\ell}_{q_n}), (\widetilde{\theta}, \widetilde{b}, \widetilde{\ell})) \to 0$ because $(\Theta \times R_{[\pm\infty]}^{2k}, d)$ is a compact metric space under Assumption A.0(i). We have $(\widetilde{\theta}, \widetilde{b}, \widetilde{\ell}) \in \Lambda$ by the following argument:

$$0 \leq \inf_{(\theta,b,\ell)\in\Lambda} d((\theta,b,\ell),(\widetilde{\theta},\widetilde{b},\widetilde{\ell})) \leq \inf_{(\theta,b,\ell)\in\Lambda} d((\theta,b,\ell),(\widetilde{\theta}_{q_n},\widetilde{b}_{q_n},\widetilde{\ell}_{q_n})) + d((\widetilde{\theta}_{q_n},\widetilde{b}_{q_n},\widetilde{\ell}_{q_n}),(\widetilde{\theta},\widetilde{b},\widetilde{\ell})) \\ \to 0, \tag{16.7}$$

where the second inequality holds by the triangle inequality and the convergence holds using Assumption C.7 (i.e., $\Lambda_{n,F_n} \to_H \Lambda$). Thus, $\inf_{(\theta,b,\ell)\in\Lambda} d((\theta,b,\ell),(\widetilde{\theta},\widetilde{b},\widetilde{\ell})) = 0$. This implies that $(\widetilde{\theta},\widetilde{b},\widetilde{\ell}) \in \Lambda$, because Λ is a compact subset of $(\Theta \times R^{2k}_{[\pm\infty]}, d)$ by Assumption C.7, $d((\theta,b,\ell),(\widetilde{\theta},\widetilde{b},\widetilde{\ell}))$ is a continuous function of (θ,b,ℓ), and a continuous function on a compact set attains its infimum.

Since $\widetilde{\theta}_n \in \Theta_I^{MR}(F_n)$, $r_{F_n}(\widetilde{\theta}_n) = r_{F_n}^{\inf}$ $\forall n \geq 1$. Hence, for all $n \geq 1$,

$$\max_{j\leq k} \widetilde{b}_{nj} = \max_{j\leq k} n^{1/2}([E_{F_n}\widetilde{m}_j(W,\widetilde{\theta}_n)]_- - r_{F_n}^{\inf}) = n^{1/2}(r_{F_n}(\widetilde{\theta}_n) - r_{F_n}^{\inf}) = 0, \tag{16.8}$$

where the first equality holds by the definition of Λ_{n,F_n} in (15.2) and the second equality holds by the expression for $r_F(\theta)$ in (3.5). We obtain

$$\max_{j\leq k} \widetilde{b}_j = \lim_{n\to\infty} \max_{j\leq k} \widetilde{b}_{nj} = 0, \tag{16.9}$$

which proves part (d) of the lemma since $(\widetilde{\theta},\widetilde{b},\widetilde{\ell}) \in \Lambda$.

Given any $(\theta^*, b^*, \ell^*) \in \Lambda$, there exists a sequence $\{(\theta_n^*, b_n^*, \ell_n^*) \in \Lambda_{n,F_n}\}_{n\geq 1}$ such that $(\theta_n^*, b_n^*, \ell_n^*) \to (\theta^*, b^*, \ell^*)$ because $\Lambda_{n,F_n} \to_H \Lambda$ by Assumption C.7. Hence, if $|\ell_j^*| < \infty$, we have

$$|\widetilde{m}_j(\theta^*)| = \lim |E_{F_n}\widetilde{m}_j(W,\theta_n^*)| = \lim(n^{-1/2}(|\ell_j^*| + o(1))) = 0, \tag{16.10}$$

where the first equality uses Assumption C.4. This establishes part (e). $\square$

Proof of Lemma 16.1. Under Assumption N, Lemma 15.1(a) implies that $h_{j\infty} \geq 0$ $\forall j \leq k$, which establishes Assumption NLA and part (a).

Now, we establish part (b). Under Assumption LA, for all $j \leq k$, we have

$$n^{1/2}|E_{F_n}\widetilde{m}_j(W,\theta_n) - E_{F_n}\widetilde{m}_j(W,\theta_{In})| \leq Kn^{1/2}||\theta_n - \theta_{In}|| = O(1), \tag{16.11}$$

where the inequality holds by Assumption LA(iii) and the equality holds by Assumption LA(i). In consequence, for all $j \leq k$, we have

$$h_{j\infty} = \lim_{n\to\infty} n^{1/2}(E_{F_n}\widetilde{m}_j(W,\theta_n) + r^{\inf}_{F_n})$$
$$= \lim_{n\to\infty} n^{1/2}(E_{F_n}\widetilde{m}_j(W,\theta_{In}) + r^{\inf}_{F_n}) + O(1) = h_{Ij\infty} + O(1) \geq O(1), \tag{16.12}$$

where the first equality holds by Assumption C.3, the second equality holds by (16.11), the third equality holds by Assumption LA(ii), and the inequality holds by Lemma 15.1(a) with θ_{In} in place of θ_n using Assumption LA(ii) in place of Assumption C.3. This completes the proof of part (b).

Under Assumption FA, we have

$$\min_{j\leq k} h_{j\infty} = \min_{j\leq k} \lim n^{1/2}(E_{F_*}\widetilde{m}_j(W,\theta_*) + r^{\inf}_{F_*}) = -\infty, \tag{16.13}$$

where the second equality holds because $E_{F_*}\widetilde{m}_j(W,\theta_*) + r^{\inf}_{F_*} < 0$ for some $j \leq k$ by Assumption FA(ii). Thus, Assumption CA holds, which establishes part (c). □

17 Proof of Theorem 15.3

The proof of Theorem 15.3(b) uses the following lemma.

Lemma 17.1 *Suppose Assumptions* C.4 *and* C.5 *hold. Under* $\{F_n\}_{n\geq 1}$, *we have*

$$\text{(a) } \widehat{\nu}_{nj}(\theta) = \nu^{m\sigma}_{nj}(\theta) + o^{\Theta}_p(1)\ \forall j \leq k \text{ and (b) } A_n = A_n(\Lambda_{n,F_n}) + o_p(1).$$

Proof of Lemma 17.1. For a given distribution F, define

$$\nu^{\sigma\dagger}_n(\theta) := n^{1/2}\left(\left(\frac{\widehat{\sigma}^2_{n1}(\theta)}{\sigma^2_{F1}(\theta)} - 1\right), ..., \left(\frac{\widehat{\sigma}^2_{nk}(\theta)}{\sigma^2_{Fk}(\theta)} - 1\right)\right)'. \tag{17.1}$$

Note that $\nu^{\sigma\dagger}_n(\theta)$ differs from $\nu^{\sigma}_n(\theta)$ (defined in (14.2)) because the former depends on $\widehat{\sigma}^2_{nj}(\theta)$, which is centered at the sample quantity $\overline{m}_{nj}(\theta)$, see (4.2), whereas the latter depends on $\widehat{\sigma}^2_{Fnj}(\theta)$, which is centered at the population quantity $E_F m_j(W,\theta)$. The following calculations show that

$\nu_{nj}^{\sigma\dagger}(\theta) = \nu_{nj}^{\sigma}(\theta) - n^{-1/2}(\nu_{nj}^{m}(\theta))^2$:

$$\begin{aligned}
\nu_{nj}^{\sigma\dagger}(\theta) &:= n^{1/2}\left(\frac{\widehat{\sigma}_{nj}^2(\theta)}{\sigma_{F_nj}^2(\theta)} - 1\right) = n^{-1/2}\sum_{i=1}^{n}\left[(\widetilde{m}_j(W_i,\theta) - \widetilde{m}_{nj}(\theta))^2 - 1\right] \\
&= n^{-1/2}\sum_{i=1}^{n}\left[(\widetilde{m}_j(W_i,\theta) - E_{F_n}\widetilde{m}_j(W,\theta))^2 - 1\right] - n^{1/2}(\widetilde{m}_{nj}(\theta) - E_{F_n}\widetilde{m}_j(W,\theta))^2 \\
&= \nu_{nj}^{\sigma}(\theta) - n^{-1/2}(\nu_{nj}^{m}(\theta))^2, \text{ and} \\
\nu_{nj}^{\sigma\dagger}(\theta) &= \nu_{nj}^{\sigma}(\theta) + o_p^{\Theta}(1) \qquad (17.2)
\end{aligned}$$

$\forall j \leq k$, where the last equality holds by Assumption C.5.

By (17.2), Assumption C.5, and the continuous mapping theorem, for all $j \leq k$,

$$\begin{aligned}
&\sup_{\theta\in\Theta}\left|\frac{\widehat{\sigma}_{nj}^2(\theta)}{\sigma_{F_nj}^2(\theta)} - 1\right| =: \sup_{\theta\in\Theta} n^{-1/2}\left|\nu_{nj}^{\sigma\dagger}(\theta)\right| = \sup_{\theta\in\Theta} n^{-1/2}\left|\nu_{nj}^{\sigma}(\theta)\right| + o_p^{\Theta}(n^{-1/2}) \to_p 0, \text{ and so,} \\
&\sup_{\theta\in\Theta}\left|\frac{\sigma_{F_nj}(\theta)}{\widehat{\sigma}_{nj}(\theta)} - 1\right| \to_p 0. \qquad (17.3)
\end{aligned}$$

We have

$$\begin{aligned}
n^{1/2}\left(\frac{\widehat{\sigma}_{nj}(\theta)}{\sigma_{F_nj}(\theta)} - 1\right) &= n^{1/2}\left(\left(1 + \left(\frac{\widehat{\sigma}_{nj}^2(\theta)}{\sigma_{F_nj}^2(\theta)} - 1\right)\right)^{1/2} - 1\right) \\
&= \frac{1}{2}(1 + o_p^{\Theta}(1))^{-1/2} n^{1/2}\left(\frac{\widehat{\sigma}_{nj}^2(\theta)}{\sigma_{F_nj}^2(\theta)} - 1\right) \\
&= \frac{1}{2}\nu_{nj}^{\sigma}(\theta) + o_p^{\Theta}(1), \qquad (17.4)
\end{aligned}$$

where the second equality holds by the following mean-value expansion, $(1+x)^{1/2} = 1 + (1/2)(1+\widetilde{x})^{-1/2}x$, where $|\widetilde{x}| \leq |x|$, with $x := \widehat{\sigma}_{nj}^2(\theta)/\sigma_{F_nj}^2(\theta) - 1$ and $\sup_{\theta\in\Theta}|x| \leq \sup_{\theta\in\Theta}|\widehat{\sigma}_{nj}^2(\theta)/\sigma_{F_nj}^2(\theta) - 1| = o_p(1)$ by (17.3), and the last equality uses (17.2) and Assumption C.5.

For all $j \leq k$, we have

$$\begin{aligned}
\widehat{\nu}_{nj}(\theta) &:= n^{1/2}\left(\widehat{m}_{nj}(\theta) - E_{F_n}\widetilde{m}_j(W,\theta)\right) = \frac{\sigma_{F_nj}(\theta)}{\widehat{\sigma}_{nj}(\theta)}\left(\nu_{nj}^{m}(\theta) - E_{F_n}\widetilde{m}_j(W,\theta)n^{1/2}\left(\frac{\widehat{\sigma}_{nj}(\theta)}{\sigma_{F_nj}(\theta)} - 1\right)\right) \\
&= (1 + o_p^{\Theta}(1))\left(\nu_{nj}^{m}(\theta) - \frac{1}{2}E_{F_n}\widetilde{m}_j(W,\theta)\nu_{nj}^{\sigma}(\theta) + o_p^{\Theta}(1)\right) \\
&= \nu_{nj}^{m\sigma}(\theta) + o_p^{\Theta}(1), \qquad (17.5)
\end{aligned}$$

where $\nu_{nj}^{m}(\theta) := n^{1/2}(\widetilde{m}_{nj}(\theta) - E_{F_n}\widetilde{m}_j(W,\theta))$, $\widetilde{m}_{nj}(\theta) = (\widehat{\sigma}_{nj}(\theta)/\sigma_{F_nj}(\theta))\widehat{m}_{nj}(\theta)$ is defined in (14.1) in online Appendix A, the second equality holds by (17.4), and the third equality holds by the definition of $\nu_{nj}^{m\sigma}(\theta)$ in (15.6) and Assumptions C.4 and C.5. This proves part (a).

To prove part (b), we have

$$\sup_{\ell_j \in R} \left| \left[\nu_{nj}^{m\sigma}(\theta) + o_p^{\Theta}(1) + \ell_j\right]_- - \left[\nu_{nj}^{m\sigma}(\theta) + \ell_j\right]_- \right| = o_p^{\Theta}(1) \quad (17.6)$$

because the function $\chi(v, c) := [v + c]_- - [c]_-$ for $v, c \in R_{[\pm\infty]}$ satisfies

$$|\chi(v, c)| \leq |v|. \quad (17.7)$$

This holds because (i) if $c \leq 0$ and $\nu + c \leq 0$, then $\chi(\nu, c) = |\nu|$, (ii) if $c \leq 0$ and $\nu + c > 0$, then $\nu > -c$ and $\chi(\nu, c) = |c| \leq |\nu|$, and (iii) if $c > 0$, then $\chi(\nu, c) = [\nu + c]_- \leq [\nu]_- \leq |\nu|$.

We have

$$\begin{aligned}
& n^{1/2}\left(\widehat{r}_{nj}(\theta) - r_{F_n}^{\inf}\right) \\
:= {} & n^{1/2}\left([\widehat{m}_{nj}(\theta)]_- - r_{F_n}^{\inf}\right) \\
= {} & \left(\left[\nu_{nj}^{m\sigma}(\theta) + n^{1/2}E_{F_n}\widetilde{m}_j(W, \theta)\right]_- - \left[n^{1/2}E_{F_n}\widetilde{m}_j(W, \theta)\right]_- + s_{nj}(\theta, F_n)\right) + o_p^{\Theta}(1),
\end{aligned} \quad (17.8)$$

where $s_{nj}(\theta, F) := n^{1/2}([E_F\widetilde{m}_j(W, \theta)]_- - r_F^{\inf})$, using (17.5) and (17.6).

For given $(\theta, b, \ell) \in \Lambda_{n,F_n}$, where Λ_{n,F_n} is defined in (15.2), we have

$$n^{1/2}E_{F_n}\widetilde{m}(W, \theta) = \ell_j \text{ and } s_{nj}(\theta, F_n) = b_j. \quad (17.9)$$

Using (17.8) and (17.9), we obtain

$$\begin{aligned}
A_n := {} & \inf_{\theta\in\Theta} \max_{j \leq k} n^{1/2}\left(\widehat{r}_{nj}(\theta) - r_{F_n}^{\inf}\right) \\
= {} & \inf_{(\theta,b,\ell)\in\Lambda_{n,F_n}} \max_{j\leq k} \left([\nu_{nj}^{m\sigma}(\theta) + \ell_j]_- - [\ell_j]_- + b_j\right) + o_p(1) \\
=: {} & A_n(\Lambda_{n,F_n}) + o_p(1),
\end{aligned} \quad (17.10)$$

where the first equality holds by the definitions in (4.4) and (4.12) and the last equality holds by the definition in (15.9). This proves part (b). □

Proof of Theorem 15.3. First, we prove part (a). For $j \leq k$, we show that

$$n^{1/2}(\widehat{m}_{nj}(\theta_n) + r_{F_n}^{\inf}) \to_d T_{j\infty} \quad (17.11)$$

and the convergence holds jointly over $j \leq k$. Stacking these results for $j = 1, ..., k$ gives $T_n(\theta_n) \to_d$

T_∞ using the definitions of $T_n(\theta_n)$ and T_∞ in (4.12) and (15.7), respectively.

We have

$$\begin{aligned}
n^{1/2}\left(\widehat{m}_{nj}(\theta)+r_F^{\inf}\right) &= n^{1/2}\left(\frac{\overline{m}_{nj}(\theta)}{\widehat{\sigma}_{nj}(\theta)}+r_F^{\inf}\right)\\
&= \frac{\sigma_{Fj}(\theta)}{\widehat{\sigma}_{nj}(\theta)}\widehat{K}_{1nj}(\theta,F)+\frac{\sigma_{Fj}(\theta)}{\widehat{\sigma}_{nj}(\theta)}\widehat{K}_{2nj}(\theta,F)+K_{3nj}(\theta,F), \text{ where}\\
\widehat{K}_{1nj}(\theta,F) &:= n^{1/2}\left(\frac{\overline{m}_{nj}(\theta)}{\sigma_{Fj}(\theta)}-\frac{E_F m_j(W,\theta)}{\sigma_{Fj}(\theta)}\right),\\
\widehat{K}_{2nj}(\theta,F) &:= -n^{1/2}\left(\frac{\widehat{\sigma}_{nj}(\theta)}{\sigma_{Fj}(\theta)}-1\right)\frac{E_F m_j(W,\theta)}{\sigma_{Fj}(\theta)}, \text{ and}\\
K_{3nj}(\theta,F) &:= n^{1/2}\left(\frac{E_F m_j(W,\theta)}{\sigma_{Fj}(\theta)}+r_F^{\inf}\right). \qquad (17.12)
\end{aligned}$$

By Assumption C.3,

$$K_{3nj}(\theta_n,F_n)\to h_{j\infty}. \qquad (17.13)$$

By (17.4) and Assumption C.5,

$$\frac{\sigma_{F_nj}(\theta_n)}{\widehat{\sigma}_{nj}(\theta_n)}\to_p 1. \qquad (17.14)$$

Given (17.14), to prove part (a), it remains to determine the asymptotic distributions of $\widehat{K}_{1nj}(\theta_n,F_n)$ and $\widehat{K}_{2nj}(\theta_n,F_n)$.

We have

$$n^{1/2}\left(\frac{\widehat{\sigma}^2_{nj}(\theta_n)}{\sigma^2_{F_nj}(\theta_n)}-1\right)=:\nu^{\sigma\dagger}_{nj}(\theta_n)=\nu^{\sigma}_{nj}(\theta_n)+o_p^{\Theta}(1)\to_d G^{\sigma}_{j\infty}, \qquad (17.15)$$

where the two equalities hold by (17.2) and the convergence holds by Assumption C.5 (which implies stochastic equicontinuity of $\{\nu^\sigma_n(\cdot)\}_{n\geq 1}$) and Assumption C.1. Equation (17.15) and the δ-method applied with the function $g(x)=x^{1/2}$, for which $g'(x)|_{x=1}=1/2$, give

$$n^{1/2}\left(\frac{\widehat{\sigma}_{nj}(\theta_n)}{\sigma_{F_nj}(\theta_n)}-1\right)\to_d \frac{1}{2}G^{\sigma}_{j\infty}. \qquad (17.16)$$

By Assumptions C.1 and C.4, $E_{F_n}\widetilde{m}_j(W,\theta_n)=\widetilde{m}_j(\theta_n)+o(1)\to\widetilde{m}_j(\theta_\infty):=\widetilde{m}_{j\infty}$. This and (17.16) give

$$\widehat{K}_{2nj}(\theta_n,F_n)\to_d -\frac{\widetilde{m}_{j\infty}}{2}G^{\sigma}_{j\infty}. \qquad (17.17)$$

We have

$$\widehat{K}_{1nj}(\theta_n,F_n):=n^{1/2}\left(\widetilde{m}_{nj}(\theta_n)-E_{F_n}\widetilde{m}_{nj}(\theta_n)\right)=\nu^m_{nj}(\theta_n)\to_d G^m_{j\infty}, \qquad (17.18)$$

where $\nu^m_{nj}(\theta_n)$ denotes the jth element of $\nu^m_n(\theta_n)$ and the convergence holds by Assumption C.5.

Combining the results in (17.12)–(17.14), (17.17), (17.18) and, for the case where $h_{j\infty} = \pm\infty$, the fact that $G^m_{j\infty} - \widetilde{m}_{j\infty} G^\sigma_{j\infty}/2 = O_p(1)$ (by Assumptions C.4 and C.5), establishes (17.11). The results in (17.11) for $j \leq k$ hold jointly because they are all based on the convergence result in Assumption C.5. This completes the proof of part (a).

Next, we prove part (b). By Lemma 17.1(b), it suffices to show

$$A_n(\Lambda_{n,F_n}) \to_d A_\infty. \tag{17.19}$$

Let $\mathcal{D}$ be the space of functions from Θ to R^{2k}. Let $\mathcal{D}_0$ be the subset of uniformly continuous functions in $\mathcal{D}$. For a nonstochastic function $\nu(\cdot) \in \mathcal{D}$, let $\nu(\theta) = (\nu^m(\theta)', \nu^\sigma(\theta)')'$, and let $\nu^m_j(\theta)$ and $\nu^\sigma_j(\theta)$ denote the jth elements of $\nu^m(\theta)$ and $\nu^\sigma(\theta)$, respectively. Define

$$\begin{aligned} g_n(\nu(\cdot)) &:= \inf_{(\theta,b,\ell)\in\Lambda_{n,F_n}} \max_{j\leq k} \left[\tau_j(\nu(\cdot),\theta,\ell) + b_j\right], \\ g(\nu(\cdot)) &:= \inf_{(\theta,b,\ell)\in\Lambda} \max_{j\leq k} \left[\tau_j(\nu(\cdot),\theta,\ell) + b_j\right], \text{ where} \\ \tau_j(\nu(\cdot),\theta,\ell) &:= [\nu^{m\sigma}_j(\theta) + \ell_j]_- - [\ell_j]_- \text{ and} \\ \nu^{m\sigma}_j(\theta) &:= \nu^m_j(\theta) - \frac{1}{2}\widetilde{m}_j(\theta)\nu^\sigma_j(\theta). \end{aligned} \tag{17.20}$$

For the stochastic processes $\nu_n(\cdot)$ and $G(\cdot)$, we can write

$$A_n(\Lambda_{n,F_n}) = g_n(\nu_n(\cdot)) \text{ and } A_\infty = A_\infty(\Lambda) = g(G(\cdot)). \tag{17.21}$$

We want to show that $g_n(\nu_n(\cdot)) \to_d g(G(\cdot))$. By Assumption C.5, $\nu_n(\cdot) \Rightarrow G(\cdot)$ for $\nu_n(\cdot) \in \mathcal{D}$ a.s. and $G(\cdot) \in \mathcal{D}_0$ a.s. We use the extended CMT, see van der Vaart and Wellner (1996, Theorem 1.11.1), to establish the desired result, as in the proof of Theorem 3.1 in BCS. The extended CMT requires showing: for any deterministic sequence $\{\nu_n(\cdot) \in \mathcal{D}\}_{n\geq 1}$ and deterministic $\nu(\cdot) \in \mathcal{D}_0$ such that $\sup_{\theta\in\Theta} ||\nu_n(\theta) - \nu(\theta)|| \to 0$, we have $g_n(\nu_n(\cdot)) \to g(\nu(\cdot))$. (For notational simplicity, we abuse notation here and consider a deterministic $\nu_n(\cdot)$ that differs from the random $\nu_n(\cdot)$ in Assumption C.5.) Once we have shown this, the proof of part (b) is complete.

Let $\{\nu_n(\cdot) \in \mathcal{D}\}_{n\geq 1}$ and $\nu(\cdot) \in \mathcal{D}_0$ be deterministic and satisfy $\sup_{\theta\in\Theta} ||\nu_n(\theta) - \nu(\theta)|| \to 0$. We show

$$\text{(i) } \liminf_{n\to\infty} g_n(\nu_n(\cdot)) \geq g(\nu(\cdot)) \text{ and (ii) } \limsup_{n\to\infty} g_n(\nu_n(\cdot)) \leq g(\nu(\cdot)). \tag{17.22}$$

First, we establish (i) in (17.22). There exists a subsequence $\{a_n\}_{n\geq 1}$ of $\{n\}_{n\geq 1}$ and there exists a sequence $\{(\overline{\theta}_{a_n}, \overline{b}_{a_n}, \overline{\ell}_{a_n}) \in \Lambda_{a_n,F_{a_n}}\}_{n\geq 1}$ such that

$$\begin{aligned}\liminf_{n\to\infty} g_n(\nu_n(\cdot)) &= \lim_{n\to\infty} g_{a_n}(\nu_{a_n}(\cdot)) \text{ and} \\ \lim_{n\to\infty} g_{a_n}(\nu_{a_n}(\cdot)) &= \lim_{n\to\infty} \max_{j\leq k} \left[\tau_j(\nu_{a_n}(\cdot), \overline{\theta}_{a_n}, \overline{\ell}_{a_n}) + \overline{b}_{a_n j}\right], \end{aligned} \tag{17.23}$$

where $\overline{b}_{a_n j}$ denotes the jth element of $\overline{b}_{a_n}$. Also, there exists a subsequence $\{e_n\}_{n\geq 1}$ of $\{a_n\}_{n\geq 1}$ and $(\overline{\theta}, \overline{b}, \overline{\ell}) \in \Theta \times R^{2k}_{[\pm\infty]}$ such that

$$d\left((\overline{\theta}_{e_n}, \overline{b}_{e_n}, \overline{\ell}_{e_n}), (\overline{\theta}, \overline{b}, \overline{\ell})\right) \to 0, \tag{17.24}$$

where d is defined in Section 15.1, by compactness of the metric space $(\Theta \times R^{2k}_{[\pm\infty]}, d)$ under Assumption A.0(i). We have $(\overline{\theta}, \overline{b}, \overline{\ell}) \in \Lambda$ by the same argument as used to show $(\widetilde{\theta}, \widetilde{b}, \widetilde{\ell}) \in \Lambda$ in (16.7) (but without the requirement that $\overline{\theta}_{a_n} \in \Theta_I^{MR}(F_{a_n})$ $\forall n \geq 1$) using (17.24) and Assumption C.7.

For all $j \leq k$,

$$\begin{aligned}\lim_{n\to\infty} \tau_j(\nu_{e_n}(\cdot), \overline{\theta}_{e_n}, \overline{\ell}_{e_n}) &= \tau_{j\infty}(\nu(\cdot), \overline{\theta}, \overline{\ell}) \in R, \text{ where} \\ \tau_{j\infty}(\nu(\cdot), \overline{\theta}, \overline{\ell}) := &\begin{cases} [\nu_j^{m\sigma}(\overline{\theta}) + \overline{\ell}_j]_- - [\overline{\ell}_j]_- & \text{if } |\overline{\ell}_j| < \infty \\ -\nu_j^{m\sigma}(\overline{\theta}) & \text{if } \overline{\ell}_j = -\infty \\ 0 & \text{if } \overline{\ell}_j = +\infty \end{cases} \\ &= [\nu_j^{m\sigma}(\overline{\theta}) + \overline{\ell}_j]_- - [\overline{\ell}_j]_- \\ &:= \tau_j(\nu(\cdot), \overline{\theta}, \overline{\ell}), \end{aligned} \tag{17.25}$$

the equality on the first line holds by $\nu_{e_n}(\theta) \to \nu(\theta) = (\nu^m(\theta)', \nu^\sigma(\theta)')'$ uniformly over $\theta \in \Theta$ (by assumption), (17.24), $[\nu_n + c_n]_- - [c_n]_- \to -\nu$ as $(\nu_n, c_n) \to (\nu, -\infty)$ for $\nu \in R$, and $[\nu_n + c_n]_- - [c_n]_- \to 0$ as $(\nu_n, c_n) \to (\nu, +\infty)$ for $\nu \in R$, the equality on the third line holds using the notational convention in (15.5), the equality on the last line holds by the definition of $\tau_j(\nu(\cdot), \theta, \ell)$ in (17.20), and "$\in R$" in the first line holds using the right-hand side (rhs) expression on the second line because $\nu_j^{m\sigma}(\overline{\theta})$ is finite since $\nu(\cdot)$ is assumed to be in $\mathcal{D}$, $\chi(\nu, c) := [\nu + c]_- - [c]_-$ for $\nu, c \in R$ satisfies $|\chi(\nu, c)| \leq |\nu|$ as shown in (17.7), and $\widetilde{m}_j(\overline{\theta})$ is finite by Assumption C.4.

Now, we have

$$\begin{aligned}
\liminf_{n\to\infty} g_n(\nu_n(\cdot)) &= \lim_{n\to\infty} \max_{j\leq k} \left[\tau_j(\nu_{e_n}(\cdot), \overline{\theta}_{e_n}, \overline{\ell}_{e_n}) + \overline{b}_{e_n j}\right] \\
&= \max_{j\leq k} \left[\tau_j(\nu(\cdot), \overline{\theta}, \overline{\ell}) + \overline{b}_j\right] \\
&\geq \inf_{(\theta,b,\ell)\in\Lambda} \max_{j\leq k} \left[\tau_j(\nu(\cdot), \theta, \ell) + b_j\right] \\
&:= g(\nu(\cdot)),
\end{aligned} \tag{17.26}$$

where the first equality holds by (17.23) and the fact that $\{e_n\}_{n\geq 1}$ is a subsequence of $\{a_n\}_{n\geq 1}$, the second equality holds by (17.25) (using the notational convention in (15.5) if $\overline{b}_j = \pm\infty$ for any $j \leq k$), the inequality holds because $(\overline{\theta}, \overline{b}, \overline{\ell}) \in \Lambda$ by the paragraph containing (17.24), and the last equality holds by the definition of $g(\nu(\cdot))$ in (17.20). This establishes result (i) in (17.22).

Next, we establish result (ii) in (17.22). There exists $(\theta^\dagger, b^\dagger, \ell^\dagger) \in \Lambda$ such that

$$g(\nu(\cdot)) = \max_{j\leq k} \left[\tau_j(\nu(\cdot), \theta^\dagger, \ell^\dagger) + b_j^\dagger\right] \tag{17.27}$$

because Λ is compact under the metric d, defined in Section 15.1 (since it is assumed to be an element of $\mathcal{S}(\Theta \times R^{2k}_{[\pm\infty]})$) and $\tau_j(\nu(\cdot), \theta, \ell) + b_j$ is a continuous function of (θ, b, ℓ) under d that takes values in the extended real line. By Assumption C.7, $\Lambda_{n,F_n} \to_H \Lambda$. Hence, there is a sequence $\{(\theta_n^\dagger, b_n^\dagger, \ell_n^\dagger) \in \Lambda_{n,F_n}\}_{n\geq 1}$ such that $d((\theta_n^\dagger, b_n^\dagger, \ell_n^\dagger), (\theta^\dagger, b^\dagger, \ell^\dagger)) \to 0$. We obtain

$$\begin{aligned}
\limsup_{n\to\infty} g_n(\nu_n(\cdot)) &:= \limsup_{n\to\infty} \inf_{(\theta,b,\ell)\in\Lambda_{n,F_n}} \max_{j\leq k} \left[\tau_j(\nu_n(\cdot), \theta, \ell) + b_j\right] \\
&\leq \limsup_{n\to\infty} \max_{j\leq k} \left[\tau_j(\nu_n(\cdot), \theta_n^\dagger, \ell_n^\dagger) + b_{nj}^\dagger\right] \\
&= \max_{j\leq k} \left[\tau_j(\nu(\cdot), \theta^\dagger, \ell^\dagger) + b_j^\dagger\right] \\
&= g(\nu(\cdot)),
\end{aligned} \tag{17.28}$$

where the inequality holds because $(\theta_n^\dagger, b_n^\dagger, \ell_n^\dagger) \in \Lambda_{n,F_n}$ $\forall n \geq 1$, the second equality holds using $d((\theta_n^\dagger, b_n^\dagger, \ell_n^\dagger), (\theta^\dagger, b^\dagger, \ell^\dagger)) \to 0$ and (17.25) with $(\nu_n(\cdot), \theta_n^\dagger, \ell_n^\dagger)$ and $(\nu(\cdot), \theta^\dagger, \ell^\dagger)$ in place of $(\nu_{e_n}(\cdot), \overline{\theta}_{e_n}, \overline{\ell}_{e_n})$ and $(\nu(\cdot), \overline{\theta}, \overline{\ell})$, respectively, and the last equality holds by (17.27). This establishes result (ii) in (17.22) and completes the proof of part (b).

Now we prove part (c). We have

$$A_\infty := \inf_{(\theta,b,\ell)\in\Lambda} \max_{j\leq k} \left([G_j^{m\sigma}(\theta) + \ell_j]_- - [\ell_j]_- + b_j\right) > -\infty \text{ a.s.} \tag{17.29}$$

because (I) $\max_{j\leq k} b_j \geq 0$ $\forall(\theta, b, \ell) \in \Lambda$ by Lemma 15.2(b) and (II) $\sup_{(\theta,b,\ell)\in\Lambda} |[G_j^{m\sigma}(\theta) + \ell_j]_- -$

$[\ell_j]_-| \leq \sup_{\theta\in\Theta} |G_j^{m\sigma}(\theta)| < \infty$ a.s. (because $\chi(\nu, c) := [\nu + c]_- - [c]_-$ satisfies $|\chi(\nu, c)| \leq |\nu|$ as shown in (17.7), $|[\nu + c]_- - [c]_-| := 0$ if $\nu \in R$ and $c = +\infty$, $|[\nu + c]_- - [c]_-| := -\nu$ if $\nu \in R$ and $c = -\infty$ using (15.5), and $\sup_{\theta\in\Theta} |G_j^{m\sigma}(\theta)| < \infty$ a.s. since $G(\cdot)$ is bounded on Θ a.s. by Assumption C.5 and $\widetilde{m}_j(\cdot)$ is bounded on Θ by Assumption C.4).

To obtain the other half of part (c), i.e., $A_\infty < \infty$ a.s., we use Lemma 15.2(d). We have

$$\begin{aligned} A_\infty &:= \inf_{(\theta,b,\ell)\in\Lambda} \max_{j\leq k} \left([G_j^{m\sigma}(\theta) + \ell_j]_- - [\ell_j]_- + b_j\right) \\ &\leq \max_{j\leq k} \left([G_j^{m\sigma}(\widetilde{\theta}) + \widetilde{\ell}_j]_- - [\widetilde{\ell}_j]_- + \widetilde{b}_j\right) < \infty \text{ a.s.}, \end{aligned} \tag{17.30}$$

where $(\widetilde{\theta}, \widetilde{b}, \widetilde{\ell}) \in \Lambda$ is as in Lemma 15.2(d), the first equality holds by the definition of A_∞ in (15.9), the first inequality holds because $(\widetilde{\theta}, \widetilde{b}, \widetilde{\ell}) \in \Lambda$, and last inequality holds because (I) $\max_{j\leq k} \widetilde{b}_j = 0$ by Lemma 15.2(d) and (II) $\sup_{(\theta,b,\ell)\in\Lambda} |[G_j^{m\sigma}(\theta) + \ell_j]_- - [\ell_j]_-| < \infty$ a.s. by (II) following (17.29). This completes the proof of part (c).

Now we prove part (d). Under Assumption NLA, for all $j \leq k$, we have

$$T_{j\infty} := G_{j\infty}^{m\sigma} + h_{j\infty} > -\infty \text{ a.s.}, \tag{17.31}$$

where the first equality holds by (15.7) and the inequality holds because $|G_{j\infty}^{m\sigma}| < \infty$ a.s. by the definitions in (15.4) and (15.6) and Assumptions C.4 and C.5, and $h_{j\infty} > -\infty$ by Assumption NLA.

Part (e) follows from the convergence results for $T_n(\theta_n)$ and A_n in parts (a) and (b), the convergence result for $\widehat{\Omega}_n(\theta_n)$ in Assumption C.6, the definition of $S_n := S_n(\theta_n)$ in (4.5) and (4.12), the continuity of $S(m, \Omega)$ at all $m \in R^k_{[+\infty]}$ and $\Omega \in \Psi$ by Assumption S.1(iii), and the fact that $T_{j\infty} > -\infty$ $\forall j \leq k$ and $A_\infty \in R$ by parts (c) and (d).

Now, we establish part (f). If $\Lambda = \Lambda_I$, then part (f) holds immediately. So, we suppose that $\Lambda\backslash\Lambda_I$ is not empty. We show that for any $(\theta^*, b^*, \ell^*) \in \Lambda\backslash\Lambda_I$,

$$\max_{j\leq k} \left[\tau_j(G(\cdot), \theta^*, \ell^*) + b_j^*\right] = \infty \text{ a.s.}, \tag{17.32}$$

where $\tau_j(\nu(\cdot), \theta, \ell)$ is defined in (17.20). Since $A_\infty \in R$ a.s. by part (c), and $A_\infty := \inf_{(\theta,b,\ell)\in\Lambda}$ $\max_{j\leq k} [\tau_j(G(\cdot), \theta, \ell) + b_j]$ by (15.9), (17.32) implies that $A_\infty = A_{I\infty}$ a.s., which establishes the first result in part (f). The second result in part (f) follows from the first result provided the quantities θ_∞, T_∞, and Ω_∞ are well defined, which requires Assumptions C.1, C.3, and C.6.

For part (f), it remains to show (17.32). By Assumption C.8, Λ_I is compact. For any $(\theta^*, b^*, \ell^*) \in \Lambda\backslash\Lambda_I$, there is a neighborhood of (θ^*, b^*, ℓ^*) that lies in $\Lambda\backslash\Lambda_I$ and there exists a sequence $\{(\theta_n^*, b_n^*, \ell_n^*) \in \Lambda_{n,F_n}\}_{n\geq 1}$ such that $d((\theta_n^*, b_n^*, \ell_n^*), (\theta^*, b^*, \ell^*)) \to 0$ by Assumption C.7. In consequence, for n large, $(\theta_n^*, b_n^*, \ell_n^*) \notin \Lambda_{n,F_n}^{\eta_n}$. In turn, this implies that $\theta_n^* \notin \Theta_I^{\eta_n}(F_n)$ for n large using the definition of $\Lambda_{n,F_n}^{\eta_n}$ following (15.3).

Now, $\theta_n^* \notin \Theta_I^{\eta_n}(F_n)$ for all n large implies

$$\begin{aligned}
&\max_{j\leq k} n^{1/2}[E_{F_n}\widetilde{m}_j(W,\theta_n^*) + r_{F_n}^{\inf}]_- > \eta_n \text{ for all } n \text{ large,}\\
&\max_{j\leq k} n^{1/2}(-E_{F_n}\widetilde{m}_j(W,\theta_n^*) - r_{F_n}^{\inf}) \to \infty, \text{ and}\\
&\max_{j\leq k} b_j^* = \lim \max_{j\leq k} b_{n,j}^* := \lim \max_{j\leq k} n^{1/2}([E_{F_n}\widetilde{m}_j(W,\theta_n^*)]_- - r_{F_n}^{\inf}) = \infty, \qquad (17.33)
\end{aligned}$$

where the first line holds by the definition of $\Theta_I^{\eta}(F)$ in (15.3), the first line implies that $\min_{j\leq k} E_{F_n}\widetilde{m}_j(W,\theta_n^*) + r_{F_n}^{\inf} < 0$ for all n large, which is used to obtain the second line, the second line also uses $\eta_n \to \infty$ by Assumption C.8, the first equality in the third line holds by the convergence result for $\{(\theta_n^*, b_n^*, \ell_n^*)\}_{n\geq 1}$ in the previous paragraph, the second equality in the third line holds by $(\theta_n^*, b_n^*, \ell_n^*) \in \Lambda_{n,F_n}$ and the definition of $\Lambda_{n,F}$ in (15.2), and the third equality in the third line follows from the second line because $\min_{j\leq k} E_{F_n}\widetilde{m}_j(W,\theta_n^*) + r_{F_n}^{\inf} < 0$ for n large implies $\min_{j\leq k} E_{F_n}\widetilde{m}_j(W,\theta_n^*) < 0$ for n large, since $r_{F_n}^{\inf} \geq 0$ by (3.5).

The result $\max_{j\leq k} b_j^* = \infty$ in (17.33) implies that (17.32) holds because $|\tau_j(G(\cdot), \theta^*, \ell^*)| < \infty$ a.s. (using Assumptions C.4 and C.5, the definition of $\tau_j(\nu(\cdot), \theta, \ell)$ in (17.20), and explanation (II) following (17.29)). This completes the proof of part (f).

Part (g) holds because $T_{j\infty} := G_{j\infty}^{m\sigma} + h_{j\infty} = -\infty$ for some $j \leq k$ by (15.7), Assumption CA, and the notational convention in (15.5).

Next, we prove part (h). We have $T_{nj}(\theta_n) \to_p h_{j\infty} = -\infty$ for some $j \leq k$ by parts (a) and (g) and $A_n \to_d A_\infty \in R$ by parts (b) and (c). Thus,

$$\varsigma_n := \min_{j\leq k}(T_{nj}(\theta_n) + A_n) \to_p -\infty. \qquad (17.34)$$

Using this, we obtain

$$\begin{aligned}
S_n := S_n(\theta_n) &= S\left(T_n(\theta_n) + A_n 1_k, \widehat{\Omega}_n(\theta_n)\right) = |\varsigma_n|^\chi S\left([T_n(\theta_n) + A_n 1_k]/|\varsigma_n|, \widehat{\Omega}_n(\theta_n)\right)\\
&\geq |\varsigma_n|^\chi \min_{j\leq k} S\left(c_j, \widehat{\Omega}_n(\theta_n)\right) = |\varsigma_n|^\chi \left(\min_{j\leq k} S\left(c_j, \Omega_\infty\right) + o_p(1)\right) \to_p \infty, \qquad (17.35)
\end{aligned}$$

where c_j is a k-vector of ∞'s but with -1 as its jth element, the second equality holds by (4.12), the third equality holds with $\chi > 0$ by Assumption S.3, the inequality holds with probability that goes to one as $n \to \infty$ (wp$\to$1) because $(T_{nj}(\theta_n) + A_n)/|\varsigma_n| = -1$ for some $j \leq k$ wp$\to$1 by the definition of ς_n and $\varsigma_n \to_p -\infty$, $S(m, \Omega)$ is nonincreasing in m for all $\Omega \in \Psi$ by Assumption S.1(i), and $[T_n(\theta_n) + A_n 1_k]/|\varsigma_n| < \infty \; \forall j \leq k$, the last equality holds by Assumptions C.6 and S.1(iii), and the convergence holds because $\min_{j \leq k} S(c_j, \Omega_\infty) > 0$ by Assumption S.2 and the fact that c_j has a negative element for all $j \leq k$, $|\varsigma_n| \to_p \infty$ and $\chi > 0$.

Lastly, the results in parts (a)–(e) hold jointly because they are all based on the convergence result in Assumption C.5, which establishes part (i). $\square$

18 Asymptotic Null Rejection Probabilities of SPUR1 Tests

This section provides a theorem, Theorem 18.1, that is the key ingredient to the proof of Theorem 4.1. It provides asymptotic NRP bounds for the nominal level α SPUR1 test $\phi_{n,SPUR1}(\theta_n)$, defined in (4.7), under drifting subsequences of distributions and parameter values. The first subsection gives various definitions and assumptions concerning the bootstrap. The second subsection states Theorem 18.1. The third subsection states several lemmas that are used in the proof of Theorem 18.1. The fourth subsection provides the proof of Theorem 18.1 using these lemmas.

18.1 Definitions and Assumptions Concerning the Bootstrap

As noted in Theorem 5.1, as is standard in the literature, the asymptotics for the bootstrap are given for the case where the number of bootstrap repetitions B equals infinity. (If one considered finite B, then all of the asymptotic results would hold provided $B \to \infty$ as $n \to \infty$.) With $B = \infty$, the bootstrap critical value $\widehat{c}_n(\theta, 1-\alpha)$, defined just above (4.16), is the $1-\alpha$ conditional quantile of $S^*_{n,b}(\theta)$ given the sample $\{W_i\}_{i \leq n}$ plus ι, rather than the $1-\alpha$ sample quantile of $\{S^*_{n,b}(\theta)\}_{b \leq B}$ plus ι. For notational simplicity, we replace the bth bootstrap sample $\{W^*_{ib}\}_{i \leq n}$ by a generic bootstrap sample $\{W^*_i\}_{i \leq n}$ (which is an i.i.d. bootstrap sample drawn with replacement from the original sample $\{W_i\}_{i \leq n}$) and we drop the subscripts b from the definition of $S^*_{n,b}(\theta)$ in (4.15) and other bootstrap quantities. Specifically, the $B = \infty$ definitions of $S^*_n(\theta)$ and $\widehat{\nu}^*_{nj}(\theta)$ are as follows. Let $Var^*(\cdot)$ denote the $\{W^*_i\}_{i \leq n}$-bootstrap variance conditional on the original sample $\{W_i\}_{i \leq n}$. Define

$$\begin{aligned} \widehat{\nu}^*_{nj}(\theta) &:= n^{1/2}\left(\frac{\overline{m}^*_{nj}(\theta)}{\widehat{\sigma}^*_{nj}(\theta)} - \widehat{m}_{nj}(\theta)\right), \\ S^*_n(\theta) &:= S\left(T^*_n(\theta) + A^*_n 1_k, \widehat{\Omega}_n(\theta)\right), \\ sd^*_{1nj}(\theta) &:= \max\{Var^*(n^{1/2}(\widehat{m}_{nj}(\theta) + \widehat{r}_n(\theta)))^{1/2}, \iota\}, \\ sd^*_{2nj}(\theta) &:= \max\{Var^*(n^{1/2}\widehat{m}_{nj}(\theta))^{1/2}, \iota\}, \text{ and} \\ sd^*_{3nj}(\theta) &:= \max\{Var^*(n^{1/2}([\widehat{m}_{nj}(\theta)]_- - \widehat{r}_n(\theta)))^{1/2}, \iota\} \qquad (18.1) \end{aligned}$$

for $j \leq k$, where $\overline{m}^*_{nj}(\theta)$, $\widehat{\sigma}^{*2}_{nj}(\theta)$, $T^*_n(\theta)$, and A^*_n are defined as in (4.17), (4.18), and (4.25) with $\{W^*_i\}_{i \leq n}$ in place of $\{W^*_{ib}\}_{i \leq n}$ and b deleted throughout, and ι is the very small positive constant employed in (4.16). In addition, $\widehat{J}_n(\theta)$ is defined as $\widehat{J}_{nB}(\theta)$ is defined in (4.24), but with $sd^*_{3nj}(\theta)$ in place of $sd^*_{3njB}(\theta)$.

The bootstrap sample $\{W^*_i\}_{i \leq n}$ depends on $\{W_i\}_{i \leq n}$ and on some other independent random variables $\{\zeta_i\}_{i \leq n}$ that are used to construct the bootstrap sample $\{W^*_i\}_{i \leq n}$. To establish the asymptotic properties of the bootstrap critical values for a given sequence of distributions $\{F_n\}_{n \geq 1}$, it is convenient to have a single probability space $(\Omega, \mathcal{F}, P_{\bigtriangledown})$ on which all of the random vectors $\{W_i\}_{i \leq n}$ for $n \geq 1$ and the bootstrap random variables (or vectors) $\{\zeta_i\}_{i \leq n}$ for all $n \geq 1$ are defined. Since F_n changes with n, this requires that we consider triangular arrays of random vectors, not sequences. Let $\{W_{ni}\}_{i \leq n, n \geq 1} := \{W_{ni} : i \leq n, n \geq 1\}$ be a triangular array of random vectors on $(\Omega, \mathcal{F}, P_{\bigtriangledown})$ such that, for each $n \geq 1$, $\{W_{ni}\}_{i \leq n}$ has the same distribution as $\{W_i\}_{i \leq n} \sim$i.i.d. F_n. Analogously, let $\{\zeta_{ni}\}_{i \leq n, n \geq 1}$ be a triangular array of bootstrap random variables (or vectors) on $(\Omega, \mathcal{F}, P_{\bigtriangledown})$ such that for each $n \geq 1$, $\{\zeta_{ni}\}_{i \leq n}$ has the same distribution as $\{\zeta_i\}_{i \leq n}$ and $\{\zeta_{ni}\}_{i \leq n, n \geq 1}$ is independent of $\{W_{ni}\}_{i \leq n, n \geq 1}$.

For notational simplicity, but with some abuse of notation, we let all of the sample size n statistics being considered for $n \geq 1$, including S_n, $S^*_n(\theta_n)$, and $\widehat{c}_n(\theta_n, 1 - \alpha)$, which are defined as functions of $\{W_i\}_{i \leq n}$ and $\{\zeta_i\}_{i \leq n}$, also denote the corresponding statistics defined when using the triangular arrays $\{W_{ni}\}_{i \leq n, n \geq 1}$ and $\{\zeta_{ni}\}_{i \leq n, n \geq 1}$. For events that only depend on n random vectors for a single n, such as $S^*_n(\theta_n) \in B_n$ for some fixed set $B_n \subset R$, we have $P_{\bigtriangledown}(S^*_n(\theta_n) \in B_n) = P_{F_n}(S^*_n(\theta_n) \in B_n)$. But, for events that depend on statistics for multiple values of n, such as $\{S^*_n(\theta_n)\}_{n \geq 1}$, we use the probability space $(\Omega, \mathcal{F}, P_{\bigtriangledown})$. In particular, when we condition on the entire triangular array $\{W_{ni}\}_{i \leq n, n \geq 1}$, we need to use $(\Omega, \mathcal{F}, P_{\bigtriangledown})$. The limit process $G(\cdot)$ in Assumptions C.5 and BC.3 (stated below) and statistics that depend on it, such as S_∞ and $S^*_{L\infty}$, are defined on a different probability space.

For $\theta \in \Theta$, define

$$j_n(\theta) := \arg\max_{j \leq k} b_{nj}(\theta), \text{ where } b_{nj}(\theta) := n^{1/2}([E_{F_n}\widetilde{m}_j(W, \theta)]_- - r^{\inf}_{F_n}).^{49} \tag{18.2}$$

By Lemma 15.2(a),

$$b_{nj_n(\theta)}(\theta) \geq 0 \ \forall \theta \in \Theta. \tag{18.3}$$

We employ the following high-level bootstrap convergence (BC) assumptions, which apply to a drifting sequence of null values $\{\theta_n\}_{n \geq 1}$ and distributions $\{F_n\}_{n \geq 1}$. (These assumptions are verified

[49] If the arg max is not unique, $j_n(\theta)$ is defined to be the smallest arg max.

below under primitive conditions.) Define

$$\Lambda_{n,F_n}^{*\eta_n} := \Big\{(\theta, b, b^*, \ell, j^*) \in \Theta_I^{\eta_n}(F_n) \times R^{3k} \times \{1, ..., k\} : b_j = n^{1/2}([E_{F_n}\widetilde{m}_j(W,\theta)]_- - r_{F_n}^{\inf}),$$
$$b_j^* = (\iota\kappa_n)^{-1} b_j,\ \ell_j = n^{1/2} E_{F_n}\widetilde{m}_j(W,\theta)\ \forall j \leq k,\ j^* := j_n(\theta)\Big\}, \quad (18.4)$$

where $\{\eta_n\}_{n\geq 1}$ is as in Assumption C.8, $\{\kappa_n\}_{n\geq 1}$ is as in (4.23), and $\iota > 0$ is the lower bound on $sd_{3njB}^*(\theta)$, defined following (4.22) using (4.16). Let $\mathcal{S}(\Theta \times R_{[\pm\infty]}^{3k} \times \{1, ..., k\})$ denote the space of compact subsets of the metric space $(\Theta \times R_{[\pm\infty]}^{3k+1}, d)$, where the metric d is defined in Section 15.1 with $a_* = d_\theta + 3k + 1$. The first two assumptions are used for upper bounds on asymptotic null rejection probabilities, which come from a lower bound on the bootstrap test statistic.

Assumption BC.1. $(\iota\kappa_n)^{-1} n^{1/2}(E_{F_n}\widetilde{m}_j(W,\theta_n) + r_{F_n}^{\inf}) \to h_{Lj\infty}^*$ for some $h_{Lj\infty}^* \in R_{[\pm\infty]}\ \forall j \leq k$.

Assumption BC.2. $\Lambda_{n,F_n}^{*\eta_n} \to_H \Lambda_I^*$ for some non-empty set $\Lambda_I^* \in \mathcal{S}(\Theta \times R_{[\pm\infty]}^{3k} \times \{1, ..., k\})$ for some constants $\{\eta_n\}_{n\geq 1}$ that satisfy $\eta_n \to \infty$ and $\eta_n/\tau_n \to 0$ for $\{\tau_n\}_{n\geq 1}$ as in Assumption A.6(ii).

Note that Assumptions BC.1 and BC.2 can always be made to hold for some subsequence $\{a_n\}_{n\geq 1}$ of $\{n\}_{n\geq 1}$ because any sequence in a compact set has a convergent subsequence.

Let $\{\nu_n^*(\theta) \in R^{2k} : \theta \in \Theta\}$ be a bootstrap version of the stochastic process $(\nu_n^m(\cdot)', \nu_n^{\sigma\dagger}(\theta)')'$ defined in (14.2) and (17.1). It is defined as follows:

$$\nu_n^*(\theta) := (\nu_n^{m*}(\theta)', \nu_n^{\sigma*}(\theta)')', \text{ where}$$
$$\nu_{nj}^{m*}(\theta) := n^{1/2}\left(\widetilde{m}_{nj}^*(\theta) - \widehat{m}_{nj}(\theta)\right),\ \widetilde{m}_{nj}^*(\theta) := \frac{\overline{m}_{nj}^*(\theta)}{\widehat{\sigma}_{nj}(\theta)},\ \overline{m}_{nj}^*(\theta) := n^{-1}\sum_{i=1}^n m_j(W_i^*, \theta),$$
$$\nu_{nj}^{\sigma*}(\theta) := n^{1/2}\left(\frac{\widehat{\sigma}_{nj}^{*2}(\theta)}{\widehat{\sigma}_{nj}^2(\theta)} - 1\right),\ \widehat{\sigma}_{nj}^{*2}(\theta) := n^{-1}\sum_{i=1}^n (m_j(W_i^*, \theta) - \overline{m}_{nj}^*(\theta))^2\ \forall j \leq k,$$
$$\nu_n^{m*}(\theta) = (\nu_{n1}^{m*}(\theta), ..., \nu_{nk}^{m*}(\theta))', \text{ and } \nu_n^{\sigma*}(\theta) = (\nu_{n1}^{\sigma*}(\theta), ..., \nu_{nk}^{\sigma*}(\theta))', \quad (18.5)$$

where $\{W_i^*\}_{i\leq n}$ is the bootstrap sample defined just above (18.1). We employ the following bootstrap convergence (BC) assumption.

Assumption BC.3. $\{\nu_n^*(\cdot)|\{W_{ni}\}_{i\leq n, n\geq 1}\} \Rightarrow G(\cdot)$ a.s.$[P_\bigtriangledown]$, where $G(\cdot)$ is as in Assumption C.5.

Assumption BC.3 is verified below for i.i.d. observations using Lemma D.2(8) of BCS under Assumptions A.1–A.4. To allow the general results to apply to non-i.i.d. observations, including time series observations, we employ Assumption BC.3 here, rather than impose Assumptions A.1–A.4.

The GMS function $\varphi : R_{[+\infty]} \to R_{[+\infty]}$ defined in (4.19) is upper bounded by the function

$\varphi^\dagger : R_{[+\infty]} \to R_{[+\infty]}$ defined by

$$\varphi^\dagger(\xi) := \infty 1(\xi \geq 1) + (\xi/(1-\xi))1(0 \leq \xi < 1) \tag{18.6}$$

for some arbitrary $\varepsilon > 0$. The function $\varphi^\dagger$ satisfies: (i) $\varphi^\dagger(\xi) \geq \varphi(\xi) \geq 0$ $\forall \xi \in R_{[+\infty]}$, (ii) $\varphi^\dagger$ is nondecreasing and continuous under the metric d, and (iii) $\varphi^\dagger(\xi) = 0$ $\forall \xi \leq 0$ and $\varphi^\dagger(\infty) = \infty$, where the metric d is defined in Section 15.1 with $a_* = 1$.

For $\theta \in \Theta$, define a lower bound (wp$\to$1) random variable, $S^*_{Ln}(\theta)$, on the EGMS bootstrap statistic $S^*_n(\theta)$ to be

$$\begin{aligned}
S^*_{Ln}(\theta) &:= S\left(T^*_{Ln}(\theta) + A^*_{Ln}1_k, \widehat{\Omega}_n(\theta)\right), \text{ where} \\
T^*_{Lnj}(\theta) &:= \widehat{\nu}^*_{nj}(\theta) + \varphi^\dagger(\xi_{1nj}(\theta)) \ \forall j \leq k, \\
T^*_{Ln}(\theta) &:= (T^*_{L1n}, ..., T^*_{Lkn})', \ \xi_{1nj}(\theta) := (\iota\kappa_n)^{-1} n^{1/2}(\widehat{m}_{nj}(\theta) + \widehat{r}_n(\theta)), \\
A^*_{Ln} &:= \inf_{\theta \in \Theta_I^{\eta_n}(F_n)} \max_{j \leq k} \Big(\chi(\widehat{\nu}^*_{nj}(\theta), n^{1/2} E_{F_n} \widetilde{m}_j(W,\theta)) + 1(j \neq j_n(\theta)) b_{nj}(\theta) \\
&\qquad + 1(j = j_n(\theta)) \varphi^\dagger(\xi^A_{1nj}(\theta)) \Big), \text{ and} \\
\xi^A_{1nj}(\theta) &:= (\iota\kappa_n)^{-1} n^{1/2} \left([\widehat{m}_{nj}(\theta)]_- - \widehat{r}_n^{\inf} \right)
\end{aligned} \tag{18.7}$$

for $1_k := (1, ..., 1)' \in R^k$ and $\chi(\nu, c) := [\nu + c]_- - [c]_-$. The function $\chi(\nu, c)$ is defined for $c = \pm\infty$ as in (15.5) and $\chi(\nu, c)$ is continuous on $R \times R_{[\pm\infty]}$ under d.

The asymptotic distribution of the lower bound random variable $S^*_{Ln}(\theta_n)$ is

$$\begin{aligned}
S^*_{L\infty} &:= S\left(T^*_{L\infty} + A^*_{L\infty}1_k, \Omega_\infty\right), \text{ where} \\
T^*_{Lj\infty} &:= G^{m\sigma}_{j\infty} + \varphi^\dagger(h^*_{Lj\infty}) \ \forall j \leq k, \ T^*_{L\infty} = (T^*_{L1\infty}, ..., T^*_{Lk\infty})', \text{ and} \\
A^*_{L\infty} &:= \inf_{(\theta, b, b^*, \ell, j^*) \in \Lambda^*_I} \max_{j \leq k} \left(\chi(G^{m\sigma}_j(\theta), \ell_j) + 1(j \neq j^*) b_j + 1(j = j^*) \varphi^\dagger(b^*_{j^*}) \right)
\end{aligned} \tag{18.8}$$

for Λ^*_I as in Assumption BC.2.

Let $c_{L\infty}(1-\alpha)$ denote the $1-\alpha$ quantile of $S^*_{L\infty}$ (with no ι added on).

Let $\to_u$ denote uniform convergence over Θ^2. We consider sequences $\{F_n\}_{n \geq 1}$ for which the covariance kernel converges uniformly.

Assumption C.11. $\Omega_{F_n}(\cdot, \cdot) \to_u \Omega_\infty(\cdot, \cdot)$ for some continuous $R^{2k \times 2k}$-valued function $\Omega_\infty(\cdot, \cdot)$ on Θ^2.

The covariance kernel of $G(\cdot)$ in Assumption C.5 is $\Omega_\infty(\cdot, \cdot)$ and the matrix Ω_∞ in Assumption C.6 is the upper left $k \times k$ submatrix of $\Omega_\infty(\theta_\infty, \theta_\infty)$.

18.2 Statement of Theorem 18.1

The following theorem shows that the nominal level α SPUR1 test has asymptotic NRP's equal to α or less for certain subsequences of distributions $\{F_n\}_{n\geq 1}$ and parameters $\{\theta_n\}_{n\geq 1}$ in the identified sets $\{\Theta_I^{MR}(F_n)\}_{n\geq 1}$.

Theorem 18.1 *For $\alpha \in (0,1)$ and for sequences $\{F_n\}_{n\geq 1}$ and $\{\theta_n\}_{n\geq 1}$ that satisfy Assumptions* A.0, A.6, BC.1–BC.3, C.1, C.3–C.8, N, *and* S.1 *for a subsequence $\{p_n\}_{n\geq 1}$ in place of $\{n\}_{n\geq 1}$, there exists a subsequence $\{a_n\}_{n\geq 1}$ of $\{p_n\}_{n\geq 1}$ for which the nominal level α* SPUR1 *test $\phi_{n,SPUR1}(\theta_n)$ for testing $H_0 : \theta_n \in \Theta_I^{MR}(F_n)$ satisfies*

$$\limsup_{n\to\infty} P_{F_{a_n}}(\phi_{a_n,SPUR1}(\theta_{a_n}) = 1) \leq \alpha.$$

18.3 Lemmas Used in the Proof of Theorem 18.1

The following three lemmas are used in the proof of Theorem 18.1. The EGMS critical values are based on the bootstrap random variables $S_n^*(\theta_n)$. In the following lemmas, the "lower bound" random variables $S_{Ln}^*(\theta)$, $T_{Lnj}^*(\theta)$, and A_{Ln}^* are defined in (18.7); the asymptotic distributions of these random variables $S_{L\infty}^*$, $T_{Lj\infty}^*$, and $A_{L\infty}^*$ are defined in (18.8); and the quantile $c_{L\infty}(1-\alpha)$ is defined following (18.8). As stated above, we assume that all of the sample size n statistics for $n \geq 1$ are functions of the triangular arrays $\{W_{ni}\}_{i\leq n, n\geq 1}$ and $\{\zeta_{ni}\}_{i\leq n,n\geq 1}$ that are defined on a single probability space $(\Omega, \mathcal{F}, P_{\bigtriangledown})$. The limit process $G(\cdot)$ in Assumptions C.5 and BC.3 and statistics that depend on it, such as S_∞ and $S_{L\infty}^*$, are defined on a different probability space.

Let $X \geq_{ST} Y$ denote that X is stochastically greater than or equal to Y. That is, $P(Y > x) \leq P(X > x)$ for all $x \in R$.

The following lemma provides the asymptotic distribution of $S_{Ln}^*(\theta_n)$.

Lemma 18.2 *For sequences $\{F_n\}_{n\geq 1}$ and $\{\theta_n\}_{n\geq 1}$ that satisfy Assumptions* A.0, A.6, BC.1–BC.3, C.1, C.4–C.7, *and* S.1 *for a subsequence $\{p_n\}_{n\geq 1}$ in place of $\{n\}_{n\geq 1}$, there exists a subsequence $\{a_n\}_{n\geq 1}$ of $\{p_n\}_{n\geq 1}$ for which* (a) $\{T_{La_nj}^*(\theta_{a_n})|\{W_{ni}\}_{i\leq n,n\geq 1}\} \to_d T_{Lj\infty}^*$ *a.s.*$[P_{\bigtriangledown}]$ $\forall j \leq k$, (b) $\{A_{La_n}^*$ $|\{W_{ni}\}_{i\leq n,n\geq 1}\} \to_d A_{L\infty}^*$ *a.s.*$[P_{\bigtriangledown}]$, *and* (c) $\{S_{La_n}^*(\theta_{a_n})|\{W_{ni}\}_{i\leq n,n\geq 1}\} \to_d S_{L\infty}^*$ *a.s.*$[P_{\bigtriangledown}]$ *and* $S_{L\infty}^* \in [0,\infty)$ *a.s.*

The following lemma establishes the lower bounding properties of $S_{Ln}^*(\theta_n)$ for $S_n^*(\theta_n)$.

Lemma 18.3 *For sequences $\{F_n\}_{n\geq 1}$ and $\{\theta_n\}_{n\geq 1}$ that satisfy Assumptions* A.0, A.6, BC.2, C.4, C.5, C.7, *and* S.1(i) *for a subsequence $\{p_n\}_{n\geq 1}$ in place of $\{n\}_{n\geq 1}$,*(a) $P_\bigtriangledown(T_{Lp_nj}^*(\theta) \geq T_{p_nj}^*(\theta)\ \forall \theta \in \Theta|$

$\{W_{ni}\}_{i\le n,n\ge 1}) = 1\ \forall j \le k$ *wp→1 under* $P_{\triangledown}$, (b) $P_{\triangledown}(A^*_{Lp_n} \ge A^*_{p_n} | \{W_{ni}\}_{i\le n,n\ge 1}) = 1$ *wp→1 under* $P_{\triangledown}$, *and* (c) $P_{\triangledown}(S^*_{Lp_n}(\theta) \le S^*_{p_n}(\theta)\ \forall \theta \in \Theta | \{W_{ni}\}_{i\le n,n\ge 1}) = 1$ *wp→1 under* $P_{\triangledown}$.

The following lemma applies to sequences $\{\theta_n\}_{n\ge 1}$ of null parameter values (i.e., those that satisfy Assumption N). Note that $S_{I\infty}$ is defined in (15.10).

Lemma 18.4 *For sequences* $\{F_n\}_{n\ge 1}$ *and* $\{\theta_n\}_{n\ge 1}$ *that satisfy Assumptions* A.6, BC.1–BC.3, C.1, C.3–C.5, C.8, *and* N *for a subsequence* $\{p_n\}_{n\ge 1}$ *in place of* $\{n\}_{n\ge 1}$, *we have*: *for all sample realizations,* (a) $T^*_{Lj\infty} \le T_{j\infty}\ \forall j \le k$, (b) $A^*_{L\infty} \le A_{I\infty}$, *and* (c) $S^*_{L\infty} \ge S_{I\infty}$ *provided Assumptions* C.6 *and* S.1(i) *also hold for the subsequence* $\{p_n\}_{n\ge 1}$.

18.4 Proof of Theorem 18.1

Proof of Theorem 18.1. Let $c_\infty = c_\infty(1-\alpha)$ denote the $1-\alpha$ quantile of S_∞ (without ι added on). For notational simplicity, let $S^*_n := S^*_n(\theta_n)$, $S^*_{Ln} := S^*_{Ln}(\theta_n)$, $c_{L\infty} := c_{L\infty}(1-\alpha)$, and $\widehat{c}_n := \widehat{c}_n(\theta_n, 1-\alpha)$. Let $\widehat{c}_{Ln}$ denote the $1-\alpha$ conditional quantile of $S^*_{Ln}(\theta_n)$ given $\{W_{ni}\}_{i\le n,n\ge 1}$ plus ι. Note that $\widehat{c}_{Ln}$ is random, depends on the conditioning value of $\{W_{ni}\}_{i\le n,n\ge 1}$, and has ι added on, whereas $c_{L\infty}$ is the $1-\alpha$ conditional (or unconditional) quantile of $S^*_{L\infty}$, which is nonrandom and does not depend on $\{W_{ni}\}_{i\le n,n\ge 1}$ by its definition following (18.8), and does not have ι added on.

The assumptions of the theorem include all of the assumptions imposed in Lemmas 18.2(c), 18.3(c), and 18.4(c) and Theorem 15.3(f). Hence, the results of these lemmas and theorem hold. For notational simplicity, we replace $\{a_n\}_{n\ge 1}$ by $\{n\}_{n\ge 1}$ and presume that the results of these lemmas and theorem hold for $\{n\}_{n\ge 1}$. By Lemma 18.3(c), $S^*_{Ln} \le S^*_n$ with probability one (with respect to the bootstrap randomness) conditional on $\{W_{ni}\}_{i\le n,n\ge 1}$. Hence, the $1-\alpha$ conditional quantile of S^*_{Ln} given $\{W_{ni}\}_{i\le n,n\ge 1}$ plus ι, which is $\widehat{c}_{Ln}$, is less than or equal to the $1-\alpha$ conditional quantile of S^*_n given $\{W_{ni}\}_{i\le n,n\ge 1}$ plus ι, which is $\widehat{c}_n$, as a consequence of the definition of a quantile. That is, $\widehat{c}_{Ln} \le \widehat{c}_n$ wp→1, which implies that $\widehat{c}_{Ln} \le \widehat{c}_n + o_p(1)$, where the $o_p(1)$ term refers to randomness in the sample. This gives

$$\limsup_{n\to\infty} P_{F_n}(\phi_n(\theta_n) = 1) = \limsup_{n\to\infty} P_{F_n}(S_n > \widehat{c}_n) \le \limsup_{n\to\infty} P_{F_n}(S_n + o_p(1) > \widehat{c}_{Ln}). \tag{18.9}$$

Now, take an arbitrary $\varepsilon > 0$. Then, there exists $\varepsilon^* \in (0, \varepsilon)$ such that $c_{L\infty} - \varepsilon^*$ is a continuity point of $S^*_{L\infty}$. We have

$$\begin{aligned}\limsup_{n\to\infty} P_{\bigtriangledown}(S^*_{Ln} \le c_{L\infty} - \varepsilon | \{W_{ni}\}_{i\le n, n\ge 1}) &\le \limsup_{n\to\infty} P_{\bigtriangledown}(S^*_{Ln} \le c_{L\infty} - \varepsilon^* | \{W_{ni}\}_{i\le n, n\ge 1}) \\ &= P(S^*_{L\infty} \le c_{L\infty} - \varepsilon^*) \\ &< 1-\alpha \qquad (18.10)\end{aligned}$$

a.s.$[P_{\bigtriangledown}]$, where the equality holds by Lemma 18.2(c) and the last inequality holds by the definition of the $1-\alpha$ quantile $c_{L\infty}$ of $S^*_{L\infty}$. Because $\widehat{c}_{Ln}$ is the $1-\alpha$ conditional quantile of S^*_{Ln} given $\{W_{ni}\}_{i\le n, n\ge 1}$ plus ι, if

$$P_{\bigtriangledown}(S^*_{Ln} \le c_{L\infty} - \varepsilon | \{W_{ni}\}_{i\le n, n\ge 1}) < 1-\alpha, \text{ then } c_{L\infty} - \varepsilon < \widehat{c}_{Ln} - \iota. \qquad (18.11)$$

By (18.10), the first condition in (18.11) holds for n sufficiently large a.s.$[P_{\bigtriangledown}]$. Hence, the same is true for the second condition in (18.11). That is, $P_{\bigtriangledown}(c_{L\infty} + \iota - \varepsilon < \widehat{c}_{Ln}$ for n large$) = 1$, or equivalently,

$$P_{\bigtriangledown}\left(\lim_{n\to\infty} 1(c_{L\infty} + \iota - \varepsilon < \widehat{c}_{Ln}) = 1\right) = 1. \qquad (18.12)$$

By the dominated convergence theorem, this implies that

$$\lim_{n\to\infty} P_{\bigtriangledown}\left(c_{L\infty} + \iota - \varepsilon < \widehat{c}_{Ln}\right) = 1 \qquad (18.13)$$

for all $\varepsilon > 0$, which also can be written as $\lim_{n\to\infty} P_{F_n}\left(c_{L\infty} + \iota - \varepsilon < \widehat{c}_{Ln}\right) = 1$.

Next, we have: for all $\varepsilon > 0$,

$$\begin{aligned}&\limsup_{n\to\infty} P_{F_n}(S_n + o_p(1) > \widehat{c}_{Ln}) \\ =\ &\limsup_{n\to\infty} P_{F_n}(S_n + o_p(1) > \widehat{c}_{Ln} \ \&\ c_{L\infty} + \iota - \varepsilon \le \widehat{c}_{Ln}) \\ \le\ &\limsup_{n\to\infty} P_{F_n}(S_n + o_p(1) > c_{L\infty} + \iota - \varepsilon \ \&\ c_{L\infty} + \iota - \varepsilon \le \widehat{c}_{Ln}) \\ =\ &\limsup_{n\to\infty} P_{F_n}(S_n + o_p(1) > c_{L\infty} + \iota - \varepsilon) \qquad (18.14)\end{aligned}$$

where the two equalities hold using (18.13) and the inequality is straightforward.

By Theorem 15.3(e), we have

$$S_n \to_d S_\infty \qquad (18.15)$$

using Assumptions A.0, C.1, C.3–C.7, S.1(iii), and NLA, where Assumption NLA holds because Assumptions C.3 and N imply Assumption NLA by Lemma 16.1. Consider a sequence $\{\varepsilon_m\}_{m\ge 1}$

such that $c_{L\infty} + \iota - \varepsilon_m$ is a continuity point of S_∞ for all $m \geq 1$ and $\varepsilon_m \downarrow 0$. Then, we have

$$\begin{aligned}
\limsup_{n\to\infty} P_{F_n}(S_n + o_p(1) > \widehat{c}_{Ln}) &\leq \lim_{m\to\infty} \limsup_{n\to\infty} P_{F_n}(S_n + o_p(1) > c_{L\infty} + \iota - \varepsilon_m) \\
&= \lim_{m\to\infty} P(S_\infty > c_{L\infty} + \iota - \varepsilon_m) \\
&\leq \lim_{m\to\infty} P(S_\infty > c_\infty + \iota - \varepsilon_m) \\
&\leq \alpha, \qquad (18.16)
\end{aligned}$$

where the first inequality holds by (18.14), the equality holds by (18.15) and the definition of $\{\varepsilon_m\}_{m\geq 1}$, the second inequality holds because $c_{L\infty} \geq c_\infty$ follows from $S^*_{L\infty} \geq S_{I\infty}$ for all sample realizations by Lemma 18.4(c) and $S_{I\infty} = S_\infty$ by Theorem 15.3(f), and the last inequality holds by the definition of the $1-\alpha$ quantile c_∞ of S_∞ because $\iota - \varepsilon_m > 0$ for m large. Equations (18.9) and (18.16) complete the proof. □

19 Proofs of Lemmas 18.2–18.4

19.1 Proof of Lemma 18.2

Proof of Lemma 18.2. First, we prove part (a). For all $j \leq k$, we have

$$n^{1/2}\left(\widehat{m}_{nj}(\theta) - E_{F_n}\widetilde{m}_j(W,\theta)\right) = O_p^{\Theta}(1), \qquad (19.1)$$

by (17.5) and Assumption C.5. Hence, we obtain

$$\sup_{\theta\in\Theta} |\widehat{m}_{nj}(\theta) - \widetilde{m}_j(\theta)| = o_p(1) \qquad (19.2)$$

using Assumption C.4. Now, we use the result that for any sequence of random variables $\{X_n\}_{n\geq 1}$ on $(\Omega, \mathcal{F}, P_\bigtriangledown)$ for which $X_n \to_p 0$, there exists a subsequence $\{c_n\}_{n\geq 1}$ of $\{n\}_{n\geq 1}$ such that $X_{c_n} \to 0$ a.s.$[P_\bigtriangledown]$, e.g., see Theorem 9.2.1 of Dudley (1989). We apply this result with the original sequence $\{n\}_{n\geq 1}$ replaced by some subsequence $\{p_n\}_{n\geq 1}$. Using this and (19.2), given any subsequence $\{p_n\}_{n\geq 1}$ of $\{n\}_{n\geq 1}$, there exists a subsequence $\{c_n\}_{n\geq 1}$ of $\{p_n\}_{n\geq 1}$ such that

$$\sup_{\theta\in\Theta} |\widehat{m}_{c_n j}(\theta) - \widetilde{m}_j(\theta)| = o(1) \text{ a.s.}[P_\bigtriangledown]. \qquad (19.3)$$

By the continuity of $\widetilde{m}_j(\theta)$ (Assumption C.4) and $\theta_n \to \theta_\infty$ (Assumption C.1), (19.3) gives

$$\widehat{m}_{c_n j}(\theta_{c_n}) \to \widetilde{m}_j(\theta_\infty) \text{ a.s.}[P_\bigtriangledown]. \qquad (19.4)$$

Conditional on $\{W_{ni}\}_{i\leq n,n\geq 1}$, for the subsequence $\{c_n\}_{n\geq 1}$, we have

$$c_n^{1/2}\left(\frac{\widehat{\sigma}^*_{c_nj}(\theta_{c_n})}{\widehat{\sigma}_{c_nj}(\theta_{c_n})}-1\right)\rightarrow_d \frac{1}{2}G^{\sigma}_{j\infty}\text{ a.s.}[P_{\bigtriangledown}]\ \forall j\leq k. \tag{19.5}$$

This holds by the delta method, as in (17.16) with $\widehat{\sigma}^{*2}_{nj}(\theta_n)$ and $\widehat{\sigma}^2_{nj}(\theta_n)$ in place of $\widehat{\sigma}^2_{nj}(\theta_n)$ and $\sigma^2_{F_nj}(\theta_n)$, respectively, and using Assumption BC.3 in place of (17.15).

Next, suppressing the dependence of various quantities on θ_{c_n} for notational simplicity, we have: conditional on $\{W_{ni}\}_{i\leq n,n\geq 1}$,

$$\begin{aligned}
T^{**}_{c_nj} :=&\ c_n^{1/2}\left(\frac{\overline{m}^*_{c_nj}}{\widehat{\sigma}^*_{c_nj}}-\frac{\overline{m}_{c_nj}}{\widehat{\sigma}_{c_nj}}\right)\\
=&\left(\frac{\widehat{\sigma}_{c_nj}}{\widehat{\sigma}^*_{c_nj}}\right)\left(c_n^{1/2}\left(\frac{\overline{m}^*_{c_nj}}{\widehat{\sigma}_{c_nj}}-\frac{\overline{m}_{c_nj}}{\widehat{\sigma}_{c_nj}}\right)-\frac{\overline{m}_{c_nj}}{\widehat{\sigma}_{c_nj}}c_n^{1/2}\frac{\widehat{\sigma}^*_{c_nj}-\widehat{\sigma}_{c_nj}}{\widehat{\sigma}_{c_nj}}\right)\\
=&\left(\frac{\widehat{\sigma}_{c_nj}}{\widehat{\sigma}^*_{c_nj}}\right)\left(\nu^{m*}_{c_nj}-\widehat{m}_{c_nj}c_n^{1/2}\frac{\widehat{\sigma}^*_{c_nj}-\widehat{\sigma}_{c_nj}}{\widehat{\sigma}_{c_nj}}\right)\\
\rightarrow_d&\ G^m_{j\infty}-\frac{1}{2}\widetilde{m}_{j\infty}G^{\sigma}_{j\infty}=:G^{m\sigma}_{j\infty}\text{ a.s.}[P_{\bigtriangledown}]
\end{aligned}\tag{19.6}$$

$\forall j\leq k$, where $\widetilde{m}_{j\infty}=\widetilde{m}_j(\theta_\infty)$ by (15.4), $G^m_{j\infty}:=G^m_j(\theta_\infty)$ and $G^{\sigma}_{j\infty}:=G^{\sigma}_j(\theta_\infty)$ by (15.6), the second equality holds by algebra, the third equality uses the definition of $\nu^{m*}_{c_nj}(\theta_{c_n})$ in (18.5), the convergence holds by (19.4), (19.5), and Assumptions BC.3 and C.1, and the last equality holds by (15.6).

We have $T^*_{Lnj}(\theta_n)=T^{**}_{nj}+\varphi^{\dagger}(\xi_{1nj}(\theta_n))$ by (18.1), (18.7), and (19.6), and $T^*_{Lj\infty}=G^{m\sigma}_{j\infty}+\varphi^{\dagger}(h^*_{Lj\infty})$ by (18.8) for all $j\leq k$. By (19.6), there exists a subsequence $\{c_n\}_{n\geq 1}$ of $\{p_n\}_{n\geq 1}$ for which $\{T^{**}_{c_nj}(\theta_{c_n})|\{W_{ni}\}_{i\leq n,n\geq 1}\}\rightarrow_d G^{m\sigma}_{j\infty}$ a.s.$[P_{\bigtriangledown}]$. Hence, part (a) holds if there exists a subsequence $\{a_n\}_{n\geq 1}$ of $\{c_n\}_{n\geq 1}$ for which

$$\{\varphi^{\dagger}(\xi_{1a_nj}(\theta_{a_n}))|\{W_{ni}\}_{i\leq n,n\geq 1}\}\rightarrow\varphi^{\dagger}(h^*_{Lj\infty})\text{ a.s.}[P_{\bigtriangledown}]\ \forall j\leq k. \tag{19.7}$$

We have

$$\begin{aligned}
\kappa_n^{-1} n^{1/2} \widehat{r}_n(\theta_n) &= \max_{j \leq k} \Big([\kappa_n^{-1} n^{1/2} (\widehat{m}_{nj}(\theta_n) - E_{F_n} \widetilde{m}_j(W, \theta_n)) + \kappa_n^{-1} n^{1/2} E_{F_n} \widetilde{m}_j(W, \theta_n)]_- \\
&\quad - [\kappa_n^{-1} n^{1/2} E_{F_n} \widetilde{m}_j(W, \theta_n)] +_- [\kappa_n^{-1} n^{1/2} E_{F_n} \widetilde{m}_j(W, \theta_n)]_- \Big) \\
&= \max_{j \leq k} \Big(o_p(1) + [\kappa_n^{-1} n^{1/2} E_{F_n} \widetilde{m}_j(W, \theta_n)]_- \Big) \\
&= o_p(1) + \max_{j \leq k} [\kappa_n^{-1} n^{1/2} E_{F_n} \widetilde{m}_j(W, \theta_n)]_- \\
&= o_p(1) + \kappa_n^{-1} n^{1/2} r_{F_n}(\theta_n), \qquad (19.8)
\end{aligned}$$

where the first equality holds by the definition of $\widehat{r}_n(\theta_n)$ in (4.4), the second equality holds because $n^{1/2}(\widehat{m}_{nj}(\theta_n) - E_{F_n} \widetilde{m}_j(W, \theta_n)) = O_p(1)$ by (19.1), $\kappa_n \to \infty$ by Assumption A.6(i), and $|\chi(\nu, c)| := |[\nu + c]_- - [c]_-| \leq |\nu|$ for $\nu, c \in R$ by (17.7), the third equality holds because the left-hand side is less than or equal to $\max_{j \leq k} o_p(1) + \max_{j \leq k} [\kappa_n^{-1} n^{1/2} E_{F_n} \widetilde{m}_j(W, \theta_n)]_-$ and is greater than or equal to $o_p(1) + [\kappa_n^{-1} n^{1/2} E_{F_n} \widetilde{m}_{j_n}(W, \theta_n)]_-$, where j_n is such that $[E_{F_n} \widetilde{m}_{j_n}(W, \theta_n)]_- = \max_{j \leq k} [E_{F_n} \widetilde{m}_j(W, \theta_n)]_-$, and the last equality holds by the definition of $r_{F_n}(\theta_n)$ in (3.5).

In turn, we obtain

$$\begin{aligned}
\xi_{1nj}(\theta_n) := &\ (\iota \kappa_n)^{-1} n^{1/2} (\widehat{m}_{nj}(\theta_n) + \widehat{r}_n(\theta_n)) \\
= &\ (\iota \kappa_n)^{-1} n^{1/2} (\widehat{m}_{nj}(\theta_n) - E_{F_n} \widetilde{m}_j(W, \theta_n)) \\
&\ + (\iota \kappa_n)^{-1} n^{1/2} (E_{F_n} \widetilde{m}_j(W, \theta_n) + r_{F_n}(\theta_n)) + o_p(1) \\
\to_p &\ h_{Lj\infty}^* \qquad (19.9)
\end{aligned}$$

for $j \leq k$, where the first equality holds by definition (see (18.7) and the discussion in the paragraph containing (18.1)), the second equality holds by (19.8), and the convergence holds using $n^{1/2}(\widehat{m}_{nj}(\theta_n) - E_{F_n} \widetilde{m}_j(W, \theta_n)) = O_p(1)$, $\kappa_n \to \infty$, and Assumption BC.1.

Equation (19.9) and the continuity of $\varphi^\dagger(\xi)$ at all $\xi \in R_{[+\infty]}$ (by property (ii) of $\varphi^\dagger$ stated following (18.6)) give $d(\varphi^\dagger(\xi_{1nj}(\theta_n)), \varphi^\dagger(h_{Lj\infty}^*)) \to_p 0$ for $j \leq k$. Now, we use the result that for any sequence of random variables $\{X_n\}_{n \geq 1}$ on $(\Omega, \mathcal{F}, P_\nabla)$ for which $X_n \to_p 0$, there exists a subsequence $\{a_n\}_{n \geq 1}$ of $\{c_n\}_{n \geq 1}$ such that $X_{a_n} \to 0$ a.s.$[P_\nabla]$. Thus, there exists a subsequence $\{a_n\}_{n \geq 1}$ of $\{p_n\}_{n \geq 1}$ such that (19.7) holds, which completes the proof of part (a).

Now, we prove part (b). Define

$$\nu_{nj}^{m\sigma*}(\theta) := \nu_{nj}^{m*}(\theta) - \frac{1}{2} \widetilde{m}_j(\theta) \nu_{nj}^{\sigma*}(\theta) \ \forall j \leq k. \qquad (19.10)$$

We show that under $\{F_n\}_{n \geq 1}$, conditional on $\{W_{ni}\}_{i \leq n, n \geq 1}$, for the subsequence $\{c_n\}_{n \geq 1}$ of $\{p_n\}_{n \geq 1}$

defined above,

$$\sup_{\theta\in\Theta}|\widehat{\nu}^{*}_{c_nj}(\theta)-\nu^{m\sigma*}_{c_nj}(\theta)|=o_p(1)\text{ a.s.}[P_{\bigtriangledown}]. \tag{19.11}$$

This, Assumption BC.3, (19.3), and the continuous mapping theorem give: under $\{F_n\}_{n\geq 1}$, conditional on $\{W_{ni}\}_{i\leq n,n\geq 1}$, for the subsequence $\{c_n\}_{n\geq 1}$ of $\{p_n\}_{n\geq 1}$,

$$\widehat{\nu}^{*}_{c_nj}(\cdot)=\nu^{m\sigma*}_{c_nj}(\cdot)+o^{\Theta}_p(1)\Rightarrow G^{m\sigma}_j(\cdot)\text{ a.s.}[P_{\bigtriangledown}]. \tag{19.12}$$

The proof of (19.11) is quite similar to (17.4) and (17.5), but with bootstrap quantities in place of original sample quantities. By the same argument as in (17.4) with $\widehat{\sigma}^{*}_{nj}(\theta)$ and $\widehat{\sigma}_{nj}(\theta)$ in place of $\widehat{\sigma}_{nj}(\theta)$ and $\sigma_{F_nj}(\theta)$, respectively, we obtain

$$n^{1/2}\left(\frac{\widehat{\sigma}^{*}_{nj}(\theta)}{\widehat{\sigma}_{nj}(\theta)}-1\right)=\frac{1}{2}\nu^{\sigma*}_{nj}(\theta)+o^{\Theta}_p(1)\text{ a.s.}[P_{\bigtriangledown}], \tag{19.13}$$

using Assumption BC.3 in place of Assumption C.5 and (17.2). Next, we have: conditional on $\{W_{ni}\}_{i\leq n,n\geq 1}$, for the subsequence $\{c_n\}_{n\geq 1}$,

$$\begin{aligned}\widehat{\nu}^{*}_{c_nj}(\theta)&:=c_n^{1/2}\left(\frac{\overline{m}^{*}_{c_nj}(\theta)}{\widehat{\sigma}^{*}_{c_nj}(\theta)}-\widehat{m}_{c_nj}(\theta)\right)=\frac{\widehat{\sigma}_{c_nj}(\theta)}{\widehat{\sigma}^{*}_{c_nj}(\theta)}\left(\nu^{m*}_{c_nj}(\theta)-\widehat{m}_{c_nj}(\theta)c_n^{1/2}\left(\frac{\widehat{\sigma}^{*}_{c_nj}(\theta)}{\widehat{\sigma}_{c_nj}(\theta)}-1\right)\right)\\&=(1+o^{\Theta}_p(1))\left(\nu^{m*}_{c_nj}(\theta)-\frac{1}{2}\widetilde{m}_j(\theta)\nu^{\sigma*}_{c_nj}(\theta)+o^{\Theta}_p(1)\right)=\nu^{m\sigma*}_{c_nj}(\theta)+o^{\Theta}_p(1)\text{ a.s.}[P_{\bigtriangledown}],\end{aligned} \tag{19.14}$$

where the third equality holds by (19.3) and (19.13), and the fourth equality holds by the definition of $\nu^{m\sigma*}_{nj}(\theta)$ in (19.10) and Assumption BC.3. This proves (19.11).

Next, we have

$$\begin{aligned}n^{1/2}\widehat{m}_{nj}(\theta)&=\frac{\sigma_{F_nj}(\theta)}{\widehat{\sigma}_{nj}(\theta)}\left(\nu^{m}_{nj}(\theta)+n^{1/2}E_{F_n}\widetilde{m}_j(W,\theta)\right)\\&=\widehat{\omega}_{nj}(\theta)+n^{1/2}E_{F_n}\widetilde{m}_j(W,\theta),\text{ where}\\\widehat{\omega}_{nj}(\theta)&:=\frac{\sigma_{F_nj}(\theta)}{\widehat{\sigma}_{nj}(\theta)}\nu^{m}_{nj}(\theta)-n^{1/2}\left(\frac{\widehat{\sigma}_{nj}(\theta)}{\sigma_{F_nj}(\theta)}-1\right)\frac{\sigma_{F_nj}(\theta)}{\widehat{\sigma}_{nj}(\theta)}E_{F_n}\widetilde{m}_j(W,\theta)=O^{\Theta}_p(1),\end{aligned} \tag{19.15}$$

where $\nu^{m}_{nj}(\theta)$ denotes the jth element of $\nu^{m}_{n}(\theta)$ defined in (14.2), and the second equality on the last line holds by Assumptions C.4 and C.5 and (17.4). Now, we have

$$n^{1/2}\left([\widehat{m}_{nj}(\theta)]_{-}-\widehat{r}_{n}^{\inf}\right) = n^{1/2}([\widehat{m}_{nj}(\theta)]_{-}-[E_F\widetilde{m}_j(W,\theta)]_{-})-n^{1/2}(\widehat{r}_n^{\inf}-r_{F_n}^{\inf})+b_{nj}(\theta)$$
$$= \widehat{d}_{nj}(\theta)+b_{nj}(\theta),\ \text{where}$$
$$\widehat{d}_{nj}(\theta) := \chi(\widehat{\omega}_{nj}(\theta), n^{1/2}E_{F_n}\widetilde{m}_j(W,\theta))-n^{1/2}(\widehat{r}_n^{\inf}-r_{F_n}^{\inf}) = O_p^{\Theta}(1), \quad (19.16)$$

the first equality uses the definition $b_{nj}(\theta) := n^{1/2}([E_{F_n}\widetilde{m}_j(W,\theta)]_{-}-r_{F_n}^{\inf})$ in (18.2), the second equality uses $\chi(\nu,c) := [\nu+c]_{-}-[c]_{-}$, and the second equality on the last line holds because $|\chi(\nu,c)| \leq |v|\ \forall \nu, c \in R$ by (17.7), $\widehat{\omega}_{nj}(\theta) = O_p^{\Theta}(1)$ by (19.15), and $n^{1/2}(\widehat{r}_n^{\inf}-r_{F_n}^{\inf}) := A_n = O_p(1)$ by (4.12) and Theorem 15.3(b) (which uses Assumptions A.0, C.4, C.5, and C.7).

For $b_j^* = (\iota\kappa_n)^{-1}n^{1/2}([E_{F_n}\widetilde{m}_j(W,\theta)]_{-}-r_{F_n}^{\inf})$ as in $\Lambda_{n,F_n}^{*\eta_n}$ (defined in (18.4)), we obtain

$$\xi_{1nj}^{A}(\theta) := (\iota\kappa_n)^{-1}n^{1/2}\left([\widehat{m}_{nj}(\theta)]_{-}-\widehat{r}_n^{\inf}\right) = (\iota\kappa_n)^{-1}\widehat{d}_{nj}(\theta)+b_j^*, \quad (19.17)$$

where the first equality holds by definition, see (18.7), and the second equality holds by (19.16).

Using (19.15), (19.17), and the definition of $\Lambda_{n,F_n}^{*\eta_n}$, we can write A_{Ln}^* in (18.7) as

$$A_{Ln}^* = \inf_{(\theta,b,b^*,\ell,j^*)\in\Lambda_{n,F_n}^{*\eta_n}} \max_{j\leq k}\Big(\chi(\widehat{\nu}_{nj}^*(\theta),\ell_j)+1(j\neq j^*)b_j \quad (19.18)$$
$$+1(j=j^*)\varphi^{\dagger}\left((\iota\kappa_n)^{-1}\widehat{d}_{nj}(\theta)+b_j^*\right)\Big),$$

where $(\theta, b_j, b_j^*, \ell_j, j^*) \in \Lambda_{n,F_n}^{*\eta_n}$ implies that $b_j := b_{nj}(\theta)$, $b_j^* := (\iota\kappa_n)^{-1}b_j$, $\ell_j := n^{1/2}E_{F_n}\widetilde{m}_j(W,\theta)$, and $j^* := j_n(\theta)$, and $\chi(\nu,c) := [\nu+c]_{-}-[c]_{-}$.

We have $(\iota\kappa_n)^{-1}\widehat{d}_{nj}(\theta) = o_p^{\Theta}(1)$ by (19.15), (19.16), and Assumption A.6(i). Hence, by the same argument as used to establish (19.3), there exists a subsequence $\{a_n\}_{n\geq1}$ (different from that in the proof of part (a)) of $\{c_n\}_{n\geq1}$ for which

$$\sup_{\theta\in\Theta}|(\iota\kappa_{a_n})^{-1}\widehat{d}_{a_nj}(\theta)| \to 0 \text{ a.s.}[P_{\bigtriangledown}]. \quad (19.19)$$

In addition, by (19.12), under $\{F_n\}_{n\geq1}$, conditional on $\{W_{ni}\}_{i\leq n,n\geq1}$, the subsequence $\{a_n\}_{n\geq1}$ of $\{p_n\}_{n\geq1}$ is such that

$$\widehat{\nu}_{a_nj}^*(\cdot) = \nu_{a_nj}^{m\sigma*}(\cdot)+o_p^{\Theta}(1) \Rightarrow G_j^{m\sigma}(\cdot) \text{ a.s.}[P_{\bigtriangledown}]. \quad (19.20)$$

Define

$$\overrightarrow{A}^*_{Ln} := \inf_{(\theta,b,b^*,\ell,j^*)\in\Lambda^{*\eta_n}_{n,F_n}} \max_{j\leq k}\left(\chi(\nu^{m\sigma*}_{nj}(\theta),\ell_j) + 1(j\neq j^*)b_j + 1(j=j^*)\varphi^\dagger(\mu_{nj}(\theta)+b^*_j)\right), \text{ where}$$
$$\mu_{nj}(\theta) := (\iota\kappa_n)^{-1}\widehat{d}_{nj}(\theta) \text{ and } \mu_n(\theta) = (\mu_{n1}(\theta),...,\mu_{nk}(\theta))'. \tag{19.21}$$

By (19.18), (19.20), and (19.21), we obtain:

$$A^*_{La_n} = \overrightarrow{A}^*_{La_n} + o_p(1) \text{ a.s.}[P_\bigtriangledown], \tag{19.22}$$

using the continuity of $\varphi^\dagger$ on $R_{[+\infty]}$ by property (ii) of $\varphi^\dagger$ stated following (18.6) and the continuity of $\chi(\nu,c)$ on $R\times R_{[\pm\infty]}$ under d. Hence, to establish part (b), it suffices to show: conditional on $\{W_{ni}\}_{i\leq n,n\geq 1}$, for the subsequence $\{a_n\}_{n\geq 1}$,

$$\left\{\overrightarrow{A}^*_{La_n}|\{W_{ni}\}_{i\leq n,n\geq 1}\right\} \to_d A^*_{L\infty} \text{ a.s.}[P_\bigtriangledown]. \tag{19.23}$$

To prove (19.23), we use a similar (but more complicated) argument to that used to prove Theorem 15.3(b) based on the extended continuous mapping theorem. As above, let $\mathcal{D}$ be the space of functions from Θ to R^{2k}. Let $\mathcal{D}_0$ be the subset of uniformly continuous functions in $\mathcal{D}$. For nonstochastic functions $\nu(\cdot)\in\mathcal{D}$ and $\mu(\cdot):\Theta\to R^k$ with $\mu(\theta)=(\mu_1(\theta),...,\mu_k(\theta))'$, define

$$\begin{aligned}\widetilde{g}_n(\nu(\cdot),\mu(\cdot)) := &\inf_{(\theta,b,b^*,\ell,j^*)\in\Lambda^{*\eta_n}_{n,F_n}} \max_{j\leq k}\Big(\tau_j(\nu(\cdot),\theta,\ell) + 1(j\neq j^*)b_j \\ &+1(j=j^*)\varphi^\dagger(\mu_{j^*}(\theta)+b^*_{j^*})\Big),\\ \widetilde{g}(\nu(\cdot),\mu(\cdot)) := &\inf_{(\theta,b,b^*,\ell,j^*)\in\Lambda^*_I} \max_{j\leq k}\Big(\tau_j(\nu(\cdot),\theta,\ell) + 1(j\neq j^*)b_j \\ &+1(j=j^*)\varphi^\dagger(\mu_{j^*}(\theta)+b^*_{j^*})\Big),\end{aligned} \tag{19.24}$$

where $\nu(\theta) = (\nu^m(\theta)',\nu^\sigma(\theta)')'$, $\nu^m_j(\theta)$ and $\nu^\sigma_j(\theta)$ denote the jth elements of $\nu^m(\theta)$ and $\nu^\sigma(\theta)$, respectively, and $\tau_j(\nu(\cdot),\theta,\ell)$ is defined in (17.20). Note that

$$\overrightarrow{A}^*_{Ln} = \widetilde{g}_n(\nu^*_n(\cdot),\mu_n(\cdot)) \text{ and } A^*_{L\infty} = \widetilde{g}(G(\cdot),\mu_\infty(\cdot)), \tag{19.25}$$

where $\mu_\infty(\cdot)$ is the nonrandom function that equals 0_k for all $\theta\in\Theta$.

We want to show $\{\widetilde{g}_{a_n}(\nu^*_{a_n}(\cdot),\mu_{a_n}(\cdot))|\{W_{ni}\}_{i\leq n,n\geq 1}\} \to_d \widetilde{g}(G(\cdot),\mu_\infty(\cdot))$ a.s.$[P_\bigtriangledown]$, where $\{\nu^*_{a_n}(\cdot)|\{W_{ni}\}_{i\leq n,n\geq 1}\} \Rightarrow G(\cdot)$ a.s.$[P_\bigtriangledown]$ by Assumption BC.3 and $\sup_{\theta\in\Theta}||\mu_{a_n}(\theta)-\mu_\infty(\theta)|| = o(1)$ a.s.$[P_\bigtriangledown]$ by (19.19) and the definition of $\mu_n(\theta)$ following (19.21). We use the extended CMT to establish this result. For notational simplicity, we employ n, rather than a_n, in the proof of this result. The extended CMT requires showing that for any deterministic sequences $\{\nu_n(\cdot) \in \mathcal{D}\}_{n\geq 1}$ and $\{\mu_n(\cdot) : \Theta \to R^k\}_{n\geq 1}$ and deterministic $\nu(\cdot) \in \mathcal{D}_0$ such that $\sup_{\theta\in\Theta} ||\nu_n(\theta) - \nu(\theta)|| \to 0$ and $\sup_{\theta\in\Theta} ||\mu_n(\theta) - \mu_\infty(\theta)|| \to 0$, we have $\widetilde{g}_n(\nu_n(\cdot), \mu_n(\cdot)) \to \widetilde{g}(\nu(\cdot), \mu_\infty(\cdot))$. (For notational simplicity, we abuse notation here and consider a deterministic $\nu_n(\cdot)$ that differs from the random $\nu_n(\cdot)$ in Assumption C.5.) Once we have shown this, the proof of part (b) is complete.

The proof of $\widetilde{g}_n(\nu_n(\cdot), \mu_n(\cdot)) \to \widetilde{g}(\nu(\cdot), \mu_\infty(\cdot))$ is an extension of the proof of $g_n(\nu_n(\cdot)) \to g(\nu(\cdot))$ in (17.22)–(17.28) in the proof of Theorem 15.3(b). We show

$$\begin{aligned} &\text{(i) } \liminf_{n\to\infty} \widetilde{g}_n(\nu_n(\cdot), \mu_n(\cdot)) \geq g(\nu(\cdot), \mu_\infty(\cdot)) \text{ and} \\ &\text{(ii) } \limsup_{n\to\infty} \widetilde{g}_n(\nu_n(\cdot), \mu_n(\cdot)) \leq g(\nu(\cdot), \mu_\infty(\cdot)). \end{aligned} \tag{19.26}$$

First, we establish (i) in (19.26). There exists a subsequence $\{c_n\}_{n\geq 1}$ of $\{n\}_{n\geq 1}$ and a sequence $\{(\overline{\theta}_{c_n}, \overline{b}_{c_n}, \overline{b}^*_{c_n}, \overline{\ell}_{c_n}, \overline{j}^*_{c_n}) \in \Lambda^{*\eta_{c_n}}_{c_n, F_{c_n}}\}_{n\geq 1}$ such that

$$\begin{aligned} \liminf_{n\to\infty} \widetilde{g}_n(\nu_n(\cdot), \mu_n(\cdot)) &= \lim_{n\to\infty} \widetilde{g}_{c_n}(\nu_{c_n}(\cdot), \mu_{c_n}(\cdot)) \text{ and} \\ \lim_{n\to\infty} \widetilde{g}_{c_n}(\nu_{c_n}(\cdot), \mu_{c_n}(\cdot)) &= \lim_{n\to\infty} \max_{j\leq k} \Big(\tau_j(\nu_{c_n}(\cdot), \overline{\theta}_{c_n}, \overline{\ell}_{c_n}) + 1(j \neq \overline{j}^*_{c_n})\overline{b}_{c_n j} \\ &\qquad + 1(j = \overline{j}^*_{c_n})\varphi^\dagger(\mu_{\overline{j}^*_{c_n}}(\overline{\theta}_{c_n}) + \overline{b}^*_{c_n \overline{j}^*_{c_n}})\Big), \end{aligned} \tag{19.27}$$

where $\overline{b}_{c_n j}$, $\overline{b}^*_{c_n j}$, and $\overline{\ell}_{c_n j}$ denote the jth elements of $\overline{b}_{c_n}$, $\overline{b}^*_{c_n}$, and $\overline{\ell}_{c_n}$, respectively. Also, there exists a subsequence $\{q_n\}_{n\geq 1}$ of $\{c_n\}_{n\geq 1}$ and $(\overline{\theta}, \overline{b}, \overline{b}^*, \overline{\ell}, \overline{j}^*) \in \Theta \times R^{3k}_{[\pm\infty]} \times \{1, ..., k\}$ such that

$$d\left((\overline{\theta}_{q_n}, \overline{b}_{q_n}, \overline{b}^*_{q_n}, \overline{\ell}_{q_n}, \overline{j}^*_{q_n}), (\overline{\theta}, \overline{b}, \overline{b}^*, \overline{\ell}, \overline{j}^*)\right) \to 0, \tag{19.28}$$

where d is defined in Section 15.1, by compactness of the metric space $(\Theta \times R^{3k}_{[\pm\infty]} \times \{1, ..., k\}, d)$ under Assumption A.0(i). We have $(\overline{\theta}, \overline{b}, \overline{b}^*, \overline{\ell}, \overline{j}^*) \in \Lambda^*_I$ by the same argument as used to show $(\widetilde{\theta}, \widetilde{b}, \widetilde{\ell}) \in \Lambda$ in (16.7) (but without the requirement that $\overline{\theta}_{q_n} \in \Theta^{MR}_I(F_{q_n})\ \forall n \geq 1$) using (19.28) and Assumption BC.2.

For all $j \leq k$,

$$\lim_{n\to\infty} \tau_j(\nu_{q_n}(\cdot), \overline{\theta}_{q_n}, \overline{\ell}_{q_n}) = \tau_j(\nu(\cdot), \overline{\theta}, \overline{\ell}) \in R \tag{19.29}$$

by (17.25) using $\nu_{q_n}(\theta) \to \nu(\theta)$ uniformly over $\theta \in \Theta$ (by assumption) and (19.28).

In addition, we have, for all $j \leq k$,

$$1(j \neq \overline{j}^*_{q_n})\overline{b}_{q_n j} \to 1(j \neq \overline{j}^*)\overline{b}_j \text{ and}$$
$$1(j = \overline{j}^*_{q_n})\varphi^\dagger(\mu_{\overline{j}^*_{q_n}}(\overline{\theta}_{q_n}) + \overline{b}^*_{q_n \overline{j}^*_{q_n}}) \to 1(j = \overline{j}^*)\varphi^\dagger(\overline{b}^*_{\overline{j}^*}), \quad (19.30)$$

where the first line holds by (19.28) and the second line holds by (19.28), $\sup_{\theta\in\Theta} ||\mu_{q_n}(\theta) - \mu_\infty(\theta)|| \to 0$, and the continuity of $\varphi^\dagger$ on $R_{[+\infty]}$ under d by property (ii) of $\varphi^\dagger$ stated following (18.6), and the fact that $d(\varphi^\dagger(\mu_{\overline{j}^*_{q_n}}(\overline{\theta}_{q_n}) + \overline{b}^*_{q_n \overline{j}^*_{q_n}}), \varphi^\dagger(\overline{b}^*_{\overline{j}^*})) \to 0$ implies that $\varphi^\dagger(\mu_{\overline{j}^*_{q_n}}(\overline{\theta}_{q_n}) + \overline{b}^*_{q_n \overline{j}^*_{q_n}}) \to \varphi^\dagger(\overline{b}^*_{\overline{j}^*})$ (as a sequence of numbers in $R_{[+\infty]}$) even if $\varphi^\dagger(\overline{b}^*_{\overline{j}^*}) = +\infty$.

Now, we have

$$\begin{aligned}
&\liminf_{n\to\infty} \widetilde{g}_n(\nu_n(\cdot), \mu_n(\cdot)) \\
&= \lim_{n\to\infty} \max_{j\leq k} \Big(\tau_j(\nu_{q_n}(\cdot), \overline{\theta}_{q_n}, \overline{\ell}_{q_n}) + 1(j \neq \overline{j}^*_{q_n})\overline{b}_{q_n j} + 1(j = \overline{j}^*_{q_n})\varphi^\dagger(\mu_{\overline{j}^*_{q_n}}(\overline{\theta}_{q_n}) + \overline{b}^*_{q_n \overline{j}^*_{q_n}})\Big) \\
&= \max_{j\leq k} \Big(\tau_j(\nu(\cdot), \overline{\theta}, \overline{\ell}) + 1(j \neq \overline{j}^*)\overline{b}_j + 1(j = \overline{j}^*)\varphi^\dagger(\overline{b}^*_{\overline{j}^*})\Big) \\
&\geq \inf_{(\theta,b,b^*,\ell,j^*)\in\Lambda^*_I} \max_{j\leq k} \Big(\tau_j(\nu(\cdot), \theta, \ell) + 1(j \neq j^*)b_j + 1(j = j^*)\varphi^\dagger(b^*_{j^*})\Big) \\
&:= \widetilde{g}(\nu(\cdot), \mu_\infty(\cdot)),
\end{aligned} \quad (19.31)$$

where the first equality holds by (19.27) and the fact that $\{q_n\}_{n\geq 1}$ is a subsequence of $\{c_n\}_{n\geq 1}$, the second equality holds by (19.29) (using the notational convention that $\nu + c = c$ when $\nu \in R$ and $c = \pm\infty$ if $\overline{b}_j = \pm\infty$ for any $j \leq k$) and (19.30), the inequality holds because $(\overline{\theta}, \overline{b}, \overline{b}^*, \overline{\ell}, \overline{j}^*) \in \Lambda^*_I$ by the paragraph containing (19.28), and the last equality holds by the definition of $\widetilde{g}(\nu(\cdot), \mu(\cdot))$ in (19.24) with $\mu(\cdot) = \mu_\infty(\cdot)$. This establishes result (i) in (19.26).

Next, we establish result (ii) in (19.26). There exists $(\theta^\dagger, b^\dagger, b^{\dagger *}, \ell^\dagger, j^{\dagger *}) \in \Lambda^*_I$ such that

$$\widetilde{g}(\nu(\cdot), \mu_\infty(\cdot)) = \max_{j\leq k} \Big(\tau_j(\nu(\cdot), \theta^\dagger, \ell^\dagger) + 1(j \neq j^{\dagger *})b^\dagger_j + 1(j = j^{\dagger *})\varphi^\dagger(b^{\dagger *}_{j^{\dagger *}})\Big) \quad (19.32)$$

because Λ^*_I is compact under the metric d defined in Section 15.1 with $a_* = d_\theta + 3k + 1$ (since it is assumed to be an element of $\mathcal{S}(\Theta \times R^{3k}_{[\pm\infty]} \times \{1, ..., k\})$) and $\tau_j(\nu(\cdot), \theta, \ell) + 1(j \neq j^*)b_j + 1(j = j^*)\varphi^\dagger(b^*_{j^*})$ is a continuous function of $(\theta, b, b^*, \ell, j^*)$ under d that takes values in the extended real line using property (ii) of $\varphi^\dagger$ stated following (18.6). By Assumption BC.2, $\Lambda^{*\eta_n}_{n,F_n} \to_H \Lambda^*_I$. Hence, there is a sequence $\{(\theta^\dagger_n, b^\dagger_n, b^{\dagger *}_n, \ell^\dagger_n, j^{\dagger *}_n) \in \Lambda^{*\eta_n}_{n,F_n}\}_{n\geq 1}$ such that $d((\theta^\dagger_n, b^\dagger_n, b^{\dagger *}_n, \ell^\dagger_n, j^{\dagger *}_n), (\theta^\dagger, b^\dagger, b^{\dagger *}, \ell^\dagger, j^{\dagger *})) \to 0$.

We obtain

$$\begin{aligned}
&\limsup_{n\to\infty} \widetilde{g}_n(\nu_n(\cdot),\mu_n(\cdot)) \\
:= &\limsup_{n\to\infty} \inf_{(\theta,b,b^*,\ell,j^*)\in\Lambda^{*\eta_n}_{n,F_n}} \max_{j\le k} (\tau_j(\nu_n(\cdot),\theta,\ell) + 1(j\neq j^*)b_j \\
&\qquad\qquad + 1(j=j^*)\varphi^\dagger(\mu_{nj^*}(\theta)+b^*_{j^*})\Big) \\
\le &\limsup_{n\to\infty} \max_{j\le k} \Big(\tau_j(\nu_n(\cdot),\theta^\dagger_n,\ell^\dagger_n) + 1(j\neq j^{\dagger *}_n)b^\dagger_{nj} + 1(j=j^{\dagger *}_n)\varphi^\dagger(\mu_{nj^{\dagger *}_n}(\theta)+b^{\dagger *}_{j^{\dagger *}_n})\Big) \\
= &\max_{j\le k} \Big(\tau_j(\nu(\cdot),\theta^\dagger,\ell^\dagger) + 1(j\neq j^{\dagger *})b^\dagger_j + 1(j=j^{\dagger *})\varphi^\dagger(b^{\dagger *}_{j^{\dagger *}})\Big) \\
= &\widetilde{g}(\nu(\cdot),\mu_\infty(\cdot)), \qquad (19.33)
\end{aligned}$$

where the inequality holds because $(\theta^\dagger_n, b^\dagger_n, b^{\dagger *}_n, \ell^\dagger_n, j^{\dagger *}_n) \in \Lambda^{*\eta_n}_{n,F_n}$ $\forall n \ge 1$, the second equality holds using $d((\theta^\dagger_n, b^\dagger_n, b^{\dagger *}_n, \ell^\dagger_n, j^{\dagger *}_n), (\theta^\dagger, b^\dagger, b^{\dagger *}, \ell^\dagger, j^{\dagger *})) \to 0$, (19.29) with $(\nu_n(\cdot), \theta^\dagger_n, \ell^\dagger_n)$ and $(\nu(\cdot), \theta^\dagger, \ell^\dagger)$ in place of $(\nu_{q_n}(\cdot), \overline{\theta}_{q_n}, \overline{\ell}_{q_n})$ and $(\nu(\cdot), \overline{\theta}, \overline{\ell})$, respectively, and (19.30) with $(\theta^\dagger_{nj}, b^\dagger_{nj}, b^{\dagger *}_{nj}, \ell^\dagger_{nj}, j^{\dagger *}_n)$ and $(\theta^\dagger_j, b^\dagger_j, b^{\dagger *}_j, \ell^\dagger_j, j^{\dagger *})$ in place of $(\overline{\theta}_{q_nj}, \overline{b}_{q_nj}, \overline{b}^*_{q_nj}, \overline{\ell}_{q_nj}, \overline{j}^*_{q_n})$ and $(\overline{\theta}_j, \overline{b}_j, \overline{b}^*_j, \overline{\ell}_j, \overline{j}^*)$, respectively, and the last equality holds by (19.32). This establishes result (ii) in (19.26) and completes the proof of part (b).

For notational simplicity, we let the subsequence $\{a_n\}_{n\ge1}$ of $\{p_n\}_{n\ge1}$ differ in the proofs of parts (a) and (b). However, by taking successive subsequences across the proofs of parts (a) and (b), we can obtain a single subsequence $\{a_n\}_{n\ge1}$ of $\{p_n\}_{n\ge1}$ for which both parts (a) and (b) (and part (d)) hold, as stated in the theorem.

The convergence result of part (c) follows from parts (a) and (b), $\widehat{\Omega}_n(\theta_n) \to_p \Omega_\infty$ (by Assumption C.6), the continuity of $S(m,\Omega)$ by Assumption S.1(iii), and the continuous mapping theorem. We have $S^*_{L\infty} \ge 0$ a.s. by Assumption S.1(ii). The function $S(m,\Omega)$ can be arbitrarily large only if m_j is arbitrarily small (i.e., m_j is negative and arbitrarily large in absolute value) for some $j \le k$, by Assumption S.1(i). We have $T^*_{Lj\infty}$ and $A^*_{L\infty}$ (defined in (18.8)) are in R a.s. by Assumptions C.4 and C.5 and the definition of $\varphi^\dagger$ in (18.6), and $\chi(G^{m\sigma}_j(\theta), \ell_j) \ge -|G^{m\sigma}_j(\theta)|$ (because $\chi(\nu, c) \ge -|\nu|$ by (17.7)). This yields $S^*_{L\infty} < \infty$ a.s., which completes the proof of part (c). $\square$

19.2 Proof of Lemma 18.3

The proof of Lemma 18.3 uses the following lemma. The set $\Theta^\eta_I(F)$ for a positive constant η is defined in (15.3) by $\Theta^\eta_I(F) := \{\theta \in \Theta : \max_{j\le k}[E_F \widetilde{m}_j(W,\theta) + r^{\inf}_F]_- \le \eta/n^{1/2}\}$. The set $\widehat{\Theta}_n$ is defined in (4.20) by $\widehat{\Theta}_n := \{\theta \in \Theta : \max_{j\le k}[\widehat{m}_{nj}(\theta) + \widehat{r}^{\inf}_n]_- \le \tau_n/n^{1/2}\}$.

Lemma 19.1 *Suppose that under* $\{F_n\}_{n\ge1}$ *and* $\{\theta_n\}_{n\ge1}$*, Assumptions* A.0, C.4, C.5, *and* C.7 *are* *satisfied. Let $\{\eta_n\}_{n\geq 1}$ and $\{\tau_n\}_{n\geq 1}$ be any sequences of positive constants that satisfy $\tau_n \to \infty$ and $\eta_n/\tau_n \to 0$. Then,*

$$P_{F_n}(\widehat{\Theta}_n \supseteq \Theta_I^{\eta_n}(F_n)) \to 1.$$

Proof of Lemma 18.3. For notational simplicity, we replace $\{p_n\}_{n\geq 1}$ by $\{n\}_{n\geq 1}$ throughout the proof of this lemma. Part (c) follows from parts (a) and (b) using the definitions of $S_n^*(\theta_n)$ and $S_{Ln}^*(\theta_n)$ in (18.1) and (18.7), and using Assumption S.1(i), which requires that $S(m,\Omega)$ is nonincreasing in $m \in R^k$ $\forall (m,\Omega) \in R^k_{[+\infty]} \times \Psi$.

To prove part (a), note that $T_{Lnj}^*(\theta)$ and $T_{nj}^*(\theta)$ only differ because the former depends on $\varphi^\dagger(\xi_{1nj}(\theta))$, whereas the latter depends on $\varphi(\xi_{nj}(\theta))$. We have

$$\varphi^\dagger(\xi_{1nj}(\theta)) \geq \varphi^\dagger(\xi_{nj}(\theta)) \geq \varphi(\xi_{nj}(\theta)), \tag{19.34}$$

where the first inequality holds because (a) if $\xi_{nj}(\theta) < 0$, then $\varphi^\dagger(\xi_{nj}(\theta)) = 0$ by properties (i) and (ii) of $\varphi^\dagger$ stated following (18.6) and $\varphi^\dagger(\xi_{1nj}(\theta)) \geq 0$ by properties (ii) and (iii) of $\varphi^\dagger$ stated following (18.6), and (b) if $\xi_{nj}(\theta) \geq 0$, then $\xi_{1nj}(\theta) \geq \xi_{nj}(\theta) := (sd_{1nj}^*(\theta)\kappa_n)^{-1}n^{1/2}(\widehat{m}_{nj}(\theta) + \widehat{r}_n(\theta))$ (since $sd_{1nj}^*(\theta) \geq \iota > 0$ by its definition following (4.19)) and $\varphi^\dagger$ is nondecreasing by property (ii) stated following (18.6), and the second inequality holds by property (i) stated following (18.6). Hence, $T_{Lnj}^*(\theta_n) \geq T_{nj}^*(\theta_n)$ for all sample and bootstrap realizations, $\forall j \leq k$, $\forall n \geq 1$, and part (a) holds.

Next, we prove part (b). By definition, see (4.25), the text following (18.1), and (18.7), we have

$$\begin{aligned} A_{Ln}^* &:= \inf_{\theta\in\Theta_I^{\eta_n}(F_n)} \max_{j\leq k} \Big(\chi(\widehat{\nu}_{nj}^*(\theta), n^{1/2}E_{F_n}\widetilde{m}_j(W,\theta)) + 1(j \neq j_n(\theta))b_{nj}(\theta) \\ &\qquad\qquad +1(j = j_n(\theta))\varphi^\dagger(\xi_{1nj}^A(\theta))\Big) \text{ and} \\ A_n^* &:= \inf_{\theta\in\widehat{\Theta}_n} \min_{j_1\in\widehat{J}_n(\theta)} \max_{j\leq k} \Big(\widehat{\chi}_{nj}^*(\theta) + 1(j \neq j_1)\widehat{b}_{nj}(\theta) \\ &\qquad\qquad +1(j = j_1)\varphi(\xi_{nj}^A(\theta))\Big). \end{aligned} \tag{19.35}$$

The bootstrap random variables A_{Ln}^* and A_n^* differ in five ways. Specifically, A_{Ln}^* versus (vs.) A_n^* are defined with (i) $\inf_{\theta\in\Theta_I^{\eta_n}(F_n)}$ vs. $\inf_{\theta\in\widehat{\Theta}_n}$, (ii) $\varphi^\dagger(\xi_{1nj}^A(\theta))$ vs. $\varphi(\xi_{nj}^A(\theta))$, (iii) $b_{nj}(\theta)$ vs. $\widehat{b}_{nj}(\theta)$, (iv) $\chi(\widehat{\nu}_{nj}^*(\theta), n^{1/2}E_{F_n}\widetilde{m}_j(W,\theta))$ vs. $\widehat{\chi}_{nj}^*(\theta)$, and (v) $j = j_n(\theta)$ or $j \neq j_n(\theta)$ vs. $\min_{j_1\in\widehat{J}_n(\theta)}$ with $j = j_1$ or $j \neq j_1$.

Lemma 19.1 applies because Lemma 18.3 imposes Assumptions A.0, C.4, C.5, and C.7, $\tau_n \to \infty$ by Assumptions A.6(ii), and $\eta_n/\tau_n \to 0$ by Assumption BC.2. By Lemma 19.1, for any bootstrap random function $K_n^*(\theta)$,

$$P_{\bigtriangledown}\left(\inf_{\theta\in\Theta_I^{\eta_n}(F_n)} K_n^*(\theta) \geq \inf_{\theta\in\widehat{\Theta}_n} K_n^*(\theta)\middle| \{W_{ni}\}_{i\leq n,n\geq 1}\right) = 1 \text{ wp} \to 1 \text{ under } P_{\bigtriangledown}. \tag{19.36}$$

By the definitions of $\xi_{1nj}^A(\theta)$ in (18.7) and $\xi_{nj}^A(\theta)$ in (4.23) and $sd_{1nj}^*(\theta)\geq\iota$ (by construction; see (18.1)), we have $|\xi_{1nj}^A(\theta)|\geq|\xi_{nj}^A(\theta)|$ and $\xi_{1nj}^A(\theta)$ and $\xi_{nj}^A(\theta)$ have the same sign for all sample and bootstrap realizations. For any $\theta\in\Theta$, for all sample and bootstrap realizations with $\xi_{nj}^A(\theta)\geq 0$, we have

$$\varphi(\xi_{nj}^A(\theta)) \leq \varphi^\dagger(\xi_{nj}^A(\theta)) \leq \varphi^\dagger(\xi_{1nj}^A(\theta)), \tag{19.37}$$

where the first inequality holds by property (i) of $\varphi^\dagger$ stated following (27.7) and the second inequality holds by property (ii) of $\varphi^\dagger$ stated following (27.7) and $\xi_{nj}^A(\theta)\leq\xi_{1nj}^A(\theta)$. Next, for all sample and bootstrap realizations with $\xi_{nj}^A(\theta)<0$, we have $\xi_{1nj}^A(\theta)<0$ and this implies that

$$\varphi(\xi_{nj}^A(\theta)) \leq \varphi^\dagger(\xi_{nj}^A(\theta)) = 0 = \varphi^\dagger(\xi_{1nj}^A(\theta)), \tag{19.38}$$

where the first inequality holds by property (i) of $\varphi^\dagger$, the first equality holds by property (iii) of $\varphi^\dagger$ and $\xi_{nj}^A(\theta)<0$, and the second equality holds by property (iii) of $\varphi^\dagger$ and $\xi_{1nj}^A(\theta)<0$. Hence, $\varphi(\xi_{nj}^A(\theta))\leq\varphi^\dagger(\xi_{1nj}^A(\theta))$ for all sample and bootstrap realizations, for all $\theta\in\Theta$.

We have

$$\begin{aligned}&\widehat{b}_{nj}(\theta) := n^{1/2}\left([\widehat{m}_{nj}(\theta)]_- - \widehat{r}_n^{\inf}\right) - sd_{3nj}^*(\theta)\kappa_n = \widehat{d}_{nj}(\theta) + b_{nj}(\theta) - sd_{3nj}^*(\theta)\kappa_n, \text{ and so,}\\ &\sup_{\theta\in\Theta}\left(\widehat{b}_{nj}(\theta) - b_{nj}(\theta)\right) \leq \sup_{\theta\in\Theta}\left(\widehat{d}_{nj}(\theta) - \iota\kappa_n\right) \to_p -\infty,\end{aligned} \tag{19.39}$$

where the first equality in the first line holds by definition, see (4.22), the second equality holds by (19.16), and the second line follows from the first line, the last line of (19.16), $sd_{3nj}^*(\theta)\geq\iota$ by definition, and $\kappa_n\to\infty$ (by Assumption A.6(i)) and the inequality on the second line holds for all bootstrap realizations because $\widehat{d}_{nj}(\theta)$ does not depend on any bootstrap quantities. Equation (19.39) implies that

$$\sup_{\theta\in\Theta}(\widehat{b}_{nj}(\theta) - b_{nj}(\theta)) \leq 0 \ \forall j\leq k, \text{ for all bootstrap realizations, wp} \to 1 \text{ under } P_{\bigtriangledown}. \tag{19.40}$$

Now, we show

$$P_{\bigtriangledown}\left(\chi(\widehat{\nu}_{nj}^*(\theta), n^{1/2}E_{F_n}\widetilde{m}_j(W,\theta)) \geq \widehat{\chi}_{nj}^*(\theta)\ \forall\theta\in\Theta|\{W_{ni}\}_{i\leq n,n\geq 1}\right) = 1 \text{ wp} \to 1. \tag{19.41}$$

The function $\chi(\nu, c) := [\nu + c]_- - [c]_-$ is nondecreasing in c for $\nu \geq 0$, is zero for all c for $\nu = 0$, and is nonincreasing in c for $\nu < 0$. The function $\chi(\nu, c)$ satisfies these monotonicity properties because, (i) for $\nu > 0$, $\chi(\nu, c) := -\nu$ (< 0) for $c < -\nu$, $\chi(\nu, c) := c$ (< 0) for $c \in [-\nu, 0)$, and $\chi(\nu, c) := 0$ for $c \geq 0$, and (ii) for $\nu < 0$, $\chi(\nu, c) := -\nu$ (> 0) for $c < 0$, $\chi(\nu, c) := -\nu - c$ (> 0) for $c \in [0, \nu)$, and $\chi(\nu, c) := 0$ for $c \geq -\nu$.

Using these properties of $\chi(\nu, c)$ and the definition of $\chi(\nu, c_1, c_2)$ in (4.21), we obtain: for $\nu \geq 0$, $\chi(\nu, c_1, c_2) = \chi(\nu, c_1) \leq \chi(\nu, c)$ $\forall c \geq c_1$. And, for $\nu < 0$, $\chi(\nu, c_1, c_2) = \chi(\nu, c_2) \leq \chi(\nu, c)$ $\forall c \leq c_2$. These results yield: for all $\widehat{\nu}^*_{nj}(\theta) \geq 0$,

$$\begin{aligned} \widehat{\chi}^*_{nj}(\theta) &:= \chi\left(\widehat{\nu}^*_{nj}(\theta),\ n^{1/2}\widehat{m}_{nj}(\theta) - sd^*_{2nj}(\theta)\kappa_n,\ n^{1/2}\widehat{m}_{nj}(\theta) + sd^*_{2nj}(\theta)\kappa_n\right) \\ &= \chi\left(\widehat{\nu}^*_{nj}(\theta),\ n^{1/2}\widehat{m}_{nj}(\theta) - sd^*_{2nj}(\theta)\kappa_n\right) \\ &\leq \chi\left(\widehat{\nu}^*_{nj}(\theta),\ n^{1/2}\widehat{m}_{nj}(\theta) - \iota\kappa_n\right) \\ &\leq \chi\left(\widehat{\nu}^*_{nj}(\theta),\ n^{1/2}E_{F_n}\widetilde{m}_j(W, \theta)\right) \end{aligned} \tag{19.42}$$

provided $n^{1/2}E_{F_n}\widetilde{m}_j(W, \theta) \geq n^{1/2}\widehat{m}_{nj}(\theta) - \iota\kappa_n$, where the first inequality holds because $sd^*_{2nj}(\theta) \geq \iota$ and $\chi(v, c)$ is nondecreasing in c for $v \geq 0$, as stated above. Similarly, for $\widehat{\nu}^*_{nj}(\theta) < 0$,

$$\begin{aligned} \widehat{\chi}^*_{nj}(\theta) &= \chi\left(\widehat{\nu}^*_{nj}(\theta),\ n^{1/2}\widehat{m}_{nj}(\theta) + sd^*_{2nj}(\theta)\kappa_n\right) \\ &\leq \chi\left(\widehat{\nu}^*_{nj}(\theta),\ n^{1/2}\widehat{m}_{nj}(\theta) + \iota\kappa_n\right) \\ &\leq \chi\left(\widehat{\nu}^*_{nj}(\theta),\ n^{1/2}E_{F_n}\widetilde{m}_j(W, \theta)\right) \end{aligned} \tag{19.43}$$

provided $n^{1/2}E_{F_n}\widetilde{m}_j(W, \theta) \leq n^{1/2}\widehat{m}_{nj}(\theta) + \iota\kappa_n$.

By (19.15), which uses Assumptions C.4 and C.5, $n^{1/2}\widehat{m}_{nj}(\theta) = n^{1/2}E_{F_n}\widetilde{m}_j(W, \theta) + O_p^\Theta(1)$. Hence,

$$\liminf_{n\to\infty} P_{F_n}\left(n^{1/2}E_{F_n}\widetilde{m}_j(W, \theta) \in \left[n^{1/2}\widehat{m}_{nj}(\theta) - \iota\kappa_n, n^{1/2}\widehat{m}_{nj}(\theta) + \iota\kappa_n\right]\ \forall\theta \in \Theta\right) = 1 \tag{19.44}$$

using $\kappa_n \to \infty$ by Assumption A.6(i). The combination of (19.42)–(19.44) establishes (19.41).

Define

$$\begin{aligned} \overline{A}^*_{Ln} := \inf_{\theta\in\Theta_I^{\eta_n}(F_n)} \min_{j_1\in\widehat{J}_n(\theta)} \max_{j\leq k} \Big(&\chi(\widehat{\nu}^*_{nj}(\theta), n^{1/2}E_{F_n}\widetilde{m}_j(W, \theta)) + 1(j \neq j_1)b_{nj}(\theta) \\ &+1(j = j_1)\varphi^\dagger(\xi^A_{1nj}(\theta))\Big). \end{aligned} \tag{19.45}$$

Combining (19.36)–(19.41) and (19.45) gives

$$P_{\bigtriangledown}(\overline{A}^*_{Ln} \geq A^*_n | \{W_{ni}\}_{i \leq n, n \geq 1}) = 1 \text{ wp} \to 1 \text{ under } P_{\bigtriangledown}. \tag{19.46}$$

Next, we show that

$$P_{\bigtriangledown}(j_n(\theta) \in \widehat{J}_n(\theta)\ \forall \theta \in \Theta | \{W_{ni}\}_{i \leq n, n \geq 1}) = 1 \text{ wp} \to 1 \text{ under } P_{\bigtriangledown}, \tag{19.47}$$

where $j_n(\theta) := \arg\max_{j \leq k} b_{nj}(\theta)$ is defined in (18.2) and $\widehat{J}_n(\theta) := \{j \in \{1, ..., k\} : \widehat{r}_{nj}(\theta) \geq \widehat{r}_n(\theta) - sd^*_{3nj}(\theta) n^{-1/2}\kappa_n\}$ is defined following (18.1) using (4.24). We have $j_n(\theta) \in \widehat{J}_n(\theta)$ iff $\widehat{r}_{nj_n(\theta)}(\theta) \geq \widehat{r}_n(\theta) - sd^*_{3nj_n(\theta)}(\theta) n^{-1/2}\kappa_n$ if $n^{1/2}(\widehat{r}_{nj_n(\theta)}(\theta) - \widehat{r}^{\inf}_n) - n^{1/2}(\widehat{r}_n(\theta) - \widehat{r}^{\inf}_n) \geq -\iota\kappa_n$ because $sd^*_{3nj_n(\theta)}(\theta) \geq \iota$ by definition. By (19.16), $n^{1/2}(\widehat{r}_{nj}(\theta) - \widehat{r}^{\inf}_n) = b_{nj}(\theta) + O^\Theta_p(1)\ \forall j \leq k$ (since $\widehat{r}_{nj}(\theta) = [\widehat{m}_{nj}(\theta)]_-$ by (4.4)). Hence, $n^{1/2}(\max_{j \leq k} \widehat{r}_{nj}(\theta) - \widehat{r}^{\inf}_n) = \max_{j \leq k} b_{nj}(\theta) + O^\Theta_p(1)$. Taking $j = j_n(\theta)$, these results combine to give $n^{1/2}(\widehat{r}_{nj_n(\theta)}(\theta) - \widehat{r}^{\inf}_n) - n^{1/2}(\widehat{r}_n(\theta) - \widehat{r}^{\inf}_n) = b_{nj_n(\theta)}(\theta) - \max_{j \leq k} b_{nj}(\theta) + O^\Theta_p(1) = O^\Theta_p(1)$ using the definition of $j_n(\theta)$, where the $O^\Theta_p(1)$ term does not depend on any bootstrap quantities. Since $O^\Theta_p(1) \geq -\iota\kappa_n$ holds wp→1 using Assumption A.6(i) (i.e., $\kappa_n \to \infty$), (19.47) is proved.

For a suitably defined random function $w(j_1, \theta)$ on $\{1, ..., k\} \times \Theta$, A^*_{Ln} and $\overline{A}^*_{Ln}$ can be written as $\inf_{\theta \in \Theta^{\eta_n}_I(F_n)} w(j_n(\theta), \theta)$ and $\inf_{\theta \in \Theta^{\eta_n}_I(F_n)} \min_{j_1 \in \widehat{J}_n(\theta)} w(j_1, \theta)$, respectively. Since $w(j_n(\theta), \theta) \geq \min_{j_1 \in \widehat{J}_n(\theta)} w(j_1, \theta)$ when $j_n(\theta) \in \widehat{J}_n(\theta)$ and the latter event satisfies (19.47), we obtain

$$P_{\bigtriangledown}(A^*_{Ln} \geq \overline{A}^*_{Ln} | \{W_{ni}\}_{i \leq n, n \geq 1}) = 1 \text{ wp} \to 1 \text{ under } P_{\bigtriangledown}. \tag{19.48}$$

This and (19.46) establish the result of part (b) of the lemma. □

Proof of Lemma 19.1. We have

$$\begin{aligned} P_{F_n}(\widehat{\Theta}_n \supseteq \Theta^{\eta_n}_I(F_n)) &\geq P_{F_n}\left(\sup_{\theta \in \Theta^{\eta_n}_I(F_n)} \max_{j \leq k} n^{1/2}[\widehat{m}_{nj}(\theta) + \widehat{r}^{\inf}_n]_- \leq \tau_n\right) \\ &= P_{F_n}\left(\sup_{\theta \in \Theta^{\eta_n}_I(F_n)} \max_{j \leq k} n^{1/2}([\widehat{m}_{nj}(\theta)]_- - \widehat{r}^{\inf}_n) \leq \tau_n\right), \end{aligned} \tag{19.49}$$

where the inequality holds by the definition of $\widehat{\Theta}_n$ and the equality holds because for $b, c \geq 0$, $[a + b]_- \leq c$ if and only if $[a]_- - b \leq c$. To see this, first note that $[a + b]_- \leq c$ and $[a]_- - b \leq c$ are equivalent to $\max\{-a - b - c, -c\} \leq 0$ and $\max\{-a - b - c, -b - c\} \leq 0$, respectively. The "only if" part follows by observing that $\max\{-a - b - c, -c\} \geq \max\{-a - b - c, -b - c\}$. Now, suppose $[a]_- - b \leq c$ so that either (i) $a \geq 0$ or (ii) $a < 0$ and $-a - b \leq c$. If (i) is the case, $[a + b]_- = 0 \leq c$, and if (ii) is the case, $[a+b]_- = \max\{-a-b, 0\} \leq \max\{c, 0\} \leq c$.

We have

$$
\begin{aligned}
& \sup_{\theta \in \Theta_I^{\eta_n}(F_n)} \max_{j \leq k} n^{1/2}([\widehat{m}_{nj}(\theta)]_- - \widehat{r}_n^{\inf}) \\
= & \sup_{\theta \in \Theta_I^{\eta_n}(F_n)} \max_{j \leq k} n^{1/2}([\widehat{m}_{nj}(\theta)]_- - r_{F_n}^{\inf}) + n^{1/2}(r_{F_n}^{\inf} - \widehat{r}_n^{\inf}) \\
= & \sup_{\theta \in \Theta_I^{\eta_n}(F_n)} \max_{j \leq k} n^{1/2}([\widehat{m}_{nj}(\theta)]_- - r_{F_n}^{\inf}) + O_p(1) \\
= & \sup_{\theta \in \Theta_I^{\eta_n}(F_n)} \max_{j \leq k} \Big([\nu_{nj}^{m\sigma}(\theta) + n^{1/2} E_{F_n} \widetilde{m}_j(W, \theta)]_- - [n^{1/2} E_{F_n} \widetilde{m}_j(W, \theta)]_- \\
& + n^{1/2}([E_{F_n} \widetilde{m}_j(W, \theta)]_- - r_{F_n}^{\inf}) \Big) + O_p(1) \\
\leq & \sup_{\theta \in \Theta_I^{\eta_n}(F_n)} \max_{j \leq k} |\nu_{nj}^{m\sigma}(\theta)| + \eta_n + O_p(1) \\
= & \, O_p(1) + \eta_n,
\end{aligned}
\tag{19.50}
$$

where the second equality holds by Theorem 15.3(b) (which requires Assumptions A.0, C.4, C.5, and C.7), the third equality holds by (17.5) and (17.6), the inequality holds by the definition of $\Theta_I^{\eta_n}(F_n)$, the same reasoning as given following (19.49), and (17.7), and the last equality holds by Assumption C.5.

It follows that

$$
\begin{aligned}
& P_{F_n}\left(\sup_{\theta \in \Theta_I^{\eta_n}(F_n)} \max_{j \leq k} n^{1/2}([\widehat{m}_{nj}(\theta)]_- - \widehat{r}_n^{\inf}) \leq \tau_n \right) \\
\geq & \, P_{F_n}(O_p(1) + \eta_n \leq \tau_n) \\
= & \, P_{F_n}(O_p(1/\tau_n) + \eta_n/\tau_n \leq 1) \\
\to & \, 1,
\end{aligned}
\tag{19.51}
$$

where the convergence holds because $\tau_n \to \infty$ and $\eta_n/\tau_n \to 0$. Combining this with (19.49) gives the result of the lemma. □

19.3 Proof of Lemma 18.4

Proof of Lemma 18.4. The lemma depends on $T_{j\infty}$, $T_{Lj\infty}^*$, $A_{I\infty}$, $A_{L\infty}^*$, $S_{I\infty}$, and $S_{L\infty}^*$, which are defined in (15.7), (15.9), (15.10), and (18.8). The first four quantities are well-defined under Assumptions A.6, BC.1–BC.3, C.1, C.3–C.5, and C.8. The last two quantities, which appear in part (c), are well-defined under these assumptions plus Assumptions C.6 and S.1(i). Hence, these assumptions are imposed in Lemma 18.4.

We prove part (a) first. We have

$$T^*_{Lj\infty} := G^{m\sigma}_{j\infty} + \varphi^\dagger(h^*_{Lj\infty}) \leq G^{m\sigma}_{j\infty} + h_{j\infty} := T_{j\infty} \tag{19.52}$$

for all sample realizations, where the inequality holds because (i) $h_{j\infty} \geq 0$ by Lemma 15.1(a) (which imposes Assumptions C.3 and N), (ii) $\varphi^\dagger(h^*_{Lj\infty}) \leq h_{j\infty}$ holds immediately if $h_{j\infty} = \infty$, and (iii) if $0 \leq h_{j\infty} < \infty$, then $h^*_{Lj\infty} = 0$ (since $n^{1/2}(E_{F_n}\widetilde{m}_j(W,\theta_n) + r^{\inf}_{F_n}) \to h_{j\infty}$ and $(\iota\kappa_n)^{-1}n^{1/2}(E_{F_n}\widetilde{m}_j(W,\theta_n) + r^{\inf}_{F_n}) \to h^*_{Lj\infty}$ by Assumptions C.3 and BC.1, and $\kappa_n \to \infty$), $h^*_{Lj\infty} = 0$ implies $\varphi^\dagger(h^*_{Lj\infty}) = 0$ by property (iii) of $\varphi^\dagger$ stated following (18.6), and hence, $\varphi^\dagger(h^*_{Lj\infty}) \leq h_{j\infty}$.

Now, establish part (b), i.e., $A^*_{L\infty} \leq A_{I\infty}$. We can write $A^*_{L\infty} = \inf_{(\theta,b,b^*,\ell,j^*)\in\Lambda^*_I} K_L(\theta,b,b^*,\ell,j^*)$ and $A_{I\infty} = \inf_{(\theta,b,\ell)\in\Lambda_I} K(\theta,b,\ell)$ for random functions $K_L(\cdot)$ and $K(\cdot)$ defined in (19.54) below. To show $A^*_{L\infty} \leq A_{I\infty}$, it suffices to show that for any $(\theta,b,\ell) \in \Lambda_I$ there exists $(\theta,b,b^*,\ell,j^*) \in \Lambda^*_I$ for which $K_L(\theta,b,b^*,\ell,j^*) \leq K(\theta,b,\ell)$ for all sample realizations.

To this end, we claim: Given any $(\theta,b,\ell) \in \Lambda_I$, there exists an element $(\theta,b,b^*,\ell,j^*) \in \Lambda^*_I$.

This claim is proved as follows. By Assumption C.8, given any $(\theta,b,\ell) \in \Lambda_I$, there exists a sequence $\{(\overline{\theta}_n,\overline{b}_n,\overline{\ell}_n) \in \Lambda^{\eta_n}_{n,F_n}\}_{n\geq1}$ such that $d((\overline{\theta}_n,\overline{b}_n,\overline{\ell}_n),(\theta,b,\ell)) \to 0$, where $\overline{\theta}_n \in \Theta^{\eta_n}_I(F_n)$ for all $n \geq 1$ by the definition of $\Lambda^{\eta_n}_{n,F_n}$ following (15.3). Given $\{\overline{\theta}_n\}_{n\geq1}$, consider the corresponding sequence $\{(\overline{\theta}_n,\overline{b}_n,b^*_n,\overline{\ell}_n,j^*_n) \in \Lambda^{*\eta_n}_{n,F_n}\}_{n\geq1}$ for $\Lambda^{*\eta_n}_{n,F_n}$ defined in (18.4), where $b^*_{nj} := (\iota\kappa_n)^{-1}\overline{b}_{nj}$, $j^*_n := \arg\max_{j\leq k}\overline{b}_{nj}$, and j^*_n is the smallest arg max value if the arg max is not unique. By Assumption BC.2, $\Lambda^{*\eta_n}_{n,F_n} \to_H \Lambda^*_I$ for Λ^*_I compact (under d). In consequence, there exist a subsequence $\{u_n\}_{n\geq1}$ of $\{n\}_{n\geq1}$ and an element $(\overline{\theta},\overline{b},b^*,\overline{\ell},j^*)$ of Λ^*_I for which

$$d((\overline{\theta}_{u_n},\overline{b}_{u_n},b^*_{u_n},\overline{\ell}_{u_n},j^*_{u_n}),(\overline{\theta},\overline{b},b^*,\overline{\ell},j^*)) \to 0 \text{ and } (\overline{\theta},\overline{b},\overline{\ell}) = (\theta,b,\ell), \tag{19.53}$$

where the equality holds because $d((\overline{\theta}_n,\overline{b}_n,\overline{\ell}_n),(\theta,b,\ell)) \to 0$, which completes the proof of the claim.

Given any $(\theta,b,\ell) \in \Lambda_I$, take $(\theta,b,b^*,\ell,j^*) \in \Lambda^*_I$ as in the previous paragraph. Then, we have

$$\begin{aligned} K_L(\theta,b,b^*,\ell,j^*) &:= \max_{j\leq k}\left(\chi(G^{m\sigma}_j(\theta),\ell_j) + 1(j\neq j^*)b_j + 1(j=j^*)\varphi^\dagger(b^*_{j^*})\right) \\ &\leq \max_{j\leq k}\left[\chi(G^{m\alpha}_j(\theta),\ell_j) + b_j\right] := K(\theta,b,\ell) \end{aligned} \tag{19.54}$$

for all sample realizations, where the first and last equalities hold by the definitions of $A^*_{L\infty}$ and $A_{I\infty}$ and the inequality holds because, as we show below, $\varphi^\dagger(b^*_{j^*}) \leq b_{j^*}$. As argued above, (19.54) implies that $A^*_{L\infty} \leq A_{I\infty}$, which we set out to prove.

Next, we show $\varphi^\dagger(b^*_{j^*}) \leq b_{j^*}$. For notational simplicity, suppose (19.53) holds with n in place of u_n. We have $j_n^* \to j^*$ by (19.53), and hence, $j_n^* = j^*$ for n large (because $j_n^* \in \{1, ..., k\}$), where $j_n^* := j_n(\overline{\theta}_n)$ by the definition of $\Lambda_{n,F_n}^{*\eta_n}$ in (18.4) for $j_n(\overline{\theta}_n)$ defined in (18.2). We have $\overline{b}_{nj} \to b_j$ and $b_{nj}^* \to b_j^*$ by (19.53), where $\overline{b}_{nj} = b_{nj}(\overline{\theta}_n)$ and $b_{nj}^* = (\iota\kappa_n)^{-1}\overline{b}_{nj}$ by the definition of $\Lambda_{n,F_n}^{*\eta_n}$ for $b_{nj}(\theta)$ defined in (18.2). Hence, we have $\overline{b}_{nj_n^*} \to b_{j^*}$ and $b_{nj_n^*}^* \to b_{j^*}^*$, where $b_{nj_n^*}^* = (\iota\kappa_n)^{-1}\overline{b}_{nj_n^*} = (\iota\kappa_n)^{-1} b_{nj_n(\overline{\theta}_n)}(\overline{\theta}_n) \geq 0$ for all $n \geq 1$ by (18.3). This and $\kappa_n \to \infty$ (by Assumption A.6(i)) imply that $b_{j^*} \geq b_{j^*}^* \geq 0$. In addition, it implies that if $0 \leq b_{j^*} < \infty$, then $b_{j^*}^* = 0$ (since $\kappa_n \to \infty$). Hence, we obtain: if $0 \leq b_{j^*} < \infty$, then $\varphi^\dagger(b_{j^*}^*) = 0 \leq b_{j^*}$ because $\varphi^\dagger(0) = 0$ by property (iii) of $\varphi^\dagger$ stated following (18.6). On the other hand, if $b_{j^*} = \infty$, then $\varphi^\dagger(b_{j^*}^*) \leq \infty = b_{j^*}^*$ by the definition of $\varphi^\dagger(\cdot)$, which completes the proof of part (b).

Part (c) is implied by parts (a) and (b) using the definitions of $S_{L\infty}^*$ and $S_{I\infty}$ in (18.8) and (15.10), respectively, and Assumption S.1(i). $\square$

20 Proof of Theorem 4.1

Theorem 4.1 shows that the SPUR2 test and CS have correct asymptotic level. The SPUR1 test and CS have correct level under the same conditions. This Section proves the results of Theorem 4.1 for both SPUR1 and SPUR2 tests and CS's.

The proof of Theorem 4.1 uses the following lemma, which provides sufficient conditions for Assumptions C.5 and C.6 to hold for the case of i.i.d. observations. This lemma is based on Lemma D.2 of BCS.

Lemma 20.1 (a) *Assumptions* A.0–A.4 *and* C.11 *imply Assumption* C.5 *with the covariance kernel of* $G(\cdot)$ *in Assumption* C.5 *equal to* $\Omega_\infty(\cdot,\cdot)$. (b) *Assumptions* A.0–A.4, C.1, *and* C.11 *imply Assumption* C.6 *with* Ω_∞ *in Assumption* C.6 *equal to the upper left* $k \times k$ *submatrix of* $\Omega_\infty(\theta_\infty, \theta_\infty)$.

Comment. For any subsequence $\{q_n\}_{n\geq 1}$ of $\{n\}_{n\geq 1}$, Lemma 20.1 holds with q_n in place of n throughout, including the assumptions. (The proof just needs to be changed by replacing n by q_n throughout.)

Proof of Theorem 4.1. First, we prove the result of part (b) for the $CS_{n,SPUR1}$ CS (which is not stated as a result in Theorem 4.1(b), but is needed below in the proof of Theorem 4.1(b) for the $CS_{n,SPUR2}$ CS). Let $\phi_n(\theta)$ abbreviate $\phi_{n,SPUR1}(\theta)$. There always exist sequences $\{F_n\}_{n\geq 1}$ and $\{\theta_n \in \Theta_I^{MR}(F_n)\}_{n\geq 1}$ and a subsequence $\{q_n\}_{n\geq 1}$ of $\{n\}_{n\geq 1}$ such that

$$\liminf_{n\to\infty} \inf_{F\in\mathcal{P}} \inf_{\theta\in\Theta_I^{MR}(F)} P_F(\phi_n(\theta) = 0) = \liminf_{n\to\infty} P_{F_n}(\phi_n(\theta_n) = 0) = \lim P_{F_{q_n}}(\phi_{q_n}(\theta_{q_n}) = 0). \quad (20.1)$$

The left-hand side expression equals the uniform coverage probability in Theorem 4.1(b) using the definition of the SPUR1 CS in (4.8). By (20.1), it suffices to show that the rhs of (20.1) is $1-\alpha$ or greater with $\{q_n\}_{n\geq 1}$ replaced by some subsequence $\{a_n\}_{n\geq 1}$ of $\{q_n\}_{n\geq 1}$ (because the limit under the subsequence $\{a_n\}_{n\geq 1}$ is the same as the limit under the original subsequence $\{q_n\}_{n\geq 1}$). The rhs of (20.1) defined with $\{a_n\}_{n\geq 1}$ is $1-\alpha$ or greater by Theorem 18.1 provided the assumptions of Theorem 18.1 hold for some subsequence $\{p_n\}_{n\geq 1}$ of $\{q_n\}_{n\geq 1}$. Hence, it remains to verify that Assumptions BC.1–BC.3, C.1, and C.3–C.8 hold for some subsequence $\{p_n\}_{n\geq 1}$ (of $\{q_n\}_{n\geq 1}$) in place of $\{n\}_{n\geq 1}$ (because Assumptions A.0, A.6, and S.1, which are imposed in Theorem 18.1, are also imposed in the present theorem, and Assumption N, which is imposed in Theorem 18.1, holds because $\theta_{a_n} \in \Theta_I^{MR}(F_{a_n})\ \forall n \geq 1$ in (20.1) by construction).

Under Assumptions A.4 and A.5, by Lemma D.7 of BCS, given $\{q_n\}_{n\geq 1}$, there exists a subsequence $\{u_n\}_{n\geq 1}$ of $\{q_n\}_{n\geq 1}$, a continuous $R^{k\times k}$-valued function Ω_∞ on Θ^2, and a continuous R^k-valued function $\widetilde{m}$ on Θ for which (i) $\Omega_{F_{u_n}} \to_u \Omega_\infty$, where $\to_u$ denotes uniform convergence (over Θ^2 in this case), (ii) $E_{F_{u_n}}\widetilde{m}(W,\cdot) \to_u \widetilde{m}(\cdot)$, and hence, Assumption C.4 holds for the subsequence $\{u_n\}_{n\geq 1}$, and (iii) Assumptions C.7, C.8, and BC.2 hold for the subsequence $\{u_n\}_{n\geq 1}$. Strictly speaking, Lemma D.7 of BCS only establishes $\Omega_{F_{u_n}} \to_u \Omega_\infty$ and the subsequence versions of Assumptions C.7 and C.8, but $E_{F_{u_n}}\widetilde{m}(W,\cdot) \to_u \widetilde{m}(\cdot)$ and the subsequence version of Assumption BC.2 are established in the same ways as $\Omega_{F_{u_n}} \to_u \Omega_\infty$ (but using Assumption A.5 in place of Assumption A.4) and the subsequence versions of Assumptions C.7 and C.8, respectively.

Assumption C.1 holds for a subsequence $\{\overline{u}_n\}_{n\geq 1}$ of $\{u_n\}_{n\geq 1}$ because $\{\theta_{u_n}\}_{n\geq 1}$ is a sequence in the compact set Θ (by Assumption A.0(i)).

Assumptions C.5 and C.6 hold for the subsequence $\{u_n\}_{n\geq 1}$ by applying a subsequence version of Lemma 20.1, which imposes Assumptions A.0–A.4, C.1, and C.11. Assumptions A.0–A.4 are imposed in the present theorem and the subsequence version of Assumption C.11 holds by (i) above.

Assumptions C.3 and BC.1 hold for a subsequence $\{p_n\}_{n\geq 1}$ of $\{\overline{u}_n\}_{n\geq 1}$ because $\{\overline{u}_n^{1/2}(E_{F_{\overline{u}_n}}\widetilde{m}(W,\theta_{\overline{u}_n}) + r^{\inf}_{F_{\overline{u}_n}})\}_{n\geq 1}$ and $\{\kappa_{\overline{u}_n}^{-1}\overline{u}_n^{1/2}(E_{F_{\overline{u}_n}}\widetilde{m}(W,\theta_{\overline{u}_n}) + r^{\inf}_{F_{\overline{u}_n}})\}_{n\geq 1}$ are sequences taking values in $R^k_{[\pm\infty]}$, which is compact under d (defined in Section 15.1 with $a_* = k$).

Assumption BC.3 holds for the subsequence $\{p_n\}_{n\geq 1}$ by Lemma D.2(8) of BCS because Assumptions A.1–A.4 of this paper imply Assumptions A.1–A.4 of BCS and $\Omega_{F_{u_n}} \to_u \Omega_\infty$ implies $\Omega_{F_{p_n}} \to_u \Omega_\infty$ (because $\{p_n\}_{n\geq 1}$ is a subsequence of $\{u_n\}_{n\geq 1}$).

This concludes the proof that the assumptions employed in Theorem 18.1 hold for the subsequence $\{p_n\}_{n\geq 1}$ of $\{q_n\}_{n\geq 1}$, which completes the proof of part (b) for $CS_{n,SPUR1}$.

The proof of part (a) for the SPUR1 test is essentially the same as that of part (b) for the SPUR1 CS, but with θ_0 in place of $\theta_n\ \forall n \geq 1$.

Next, we prove part (b) for the SPUR2 CS. Let $\{F_n\}_{n\geq 1}$ and $\{\theta_n\}_{n\geq 1}$ denote sequences of distributions in $\mathcal{P}$ for which

$$\limsup_{n\to\infty} \sup_{F\in\mathcal{P}} \sup_{\theta\in\Theta_I^{MR}(F)} P_F(\phi_{n,SPUR2}(\theta)=1) = \limsup_{n\to\infty} P_{F_n}(\phi_{n,SPUR2}(\theta_n)=1). \tag{20.2}$$

Such sequences always exists. The left-hand side expression in (20.2) equals one minus the uniform coverage probability in Theorem 4.1(b) using the definition of the SPUR2 CS in (4.8).

We use the following Bonferroni argument. We have

$$\begin{aligned} &\limsup_{n\to\infty} P_{F_n}(\phi_{n,SPUR2}(\theta_n)=1) \\ \leq\ &\limsup_{n\to\infty} P_{F_n}(\phi_{n,SPUR2}(\theta_n)=1 \ \&\ r_{F_n}^{\inf} \leq \widehat{r}_{n,UP}(\alpha_1)) \\ &\quad + \limsup_{n\to\infty} P_{F_n}(\phi_{n,SPUR2}(\theta_n)=1 \ \&\ r_{F_n}^{\inf} > \widehat{r}_{n,UP}(\alpha_1)) \\ \leq\ &\limsup_{n\to\infty} P_{F_n}(\phi_{n,SPUR2}(\theta_n)=1 \ \&\ r_{F_n}^{\inf} \leq \widehat{r}_{n,UP}(\alpha_1)) + \alpha_1, \end{aligned} \tag{20.3}$$

where the second inequality holds because Theorem 5.1(a), which states that $\liminf_{n\to\infty} \inf_{F\in\mathcal{P}} P_F(\Delta_F^{\inf} \in CI_{n,\Delta U}(\alpha_1)) \geq 1-\alpha_1$, implies $\limsup_{n\to\infty} P_{F_n}(r_{F_n}^{\inf} > \widehat{r}_{n,UP}(\alpha_1)) \leq \alpha_1$ since $r_{F_n}^{\inf} := \max\{\Delta_{F_n}^{\inf}, 0\}$ and $\widehat{r}_{n,UP}(\alpha_1) := \max\{\widehat{\Delta}_{n,U}^{\inf}(\alpha), 0\}$ yield $r_{F_n}^{\inf} > \widehat{r}_{n,UP}(\alpha_1)$ iff $\Delta_{F_n}^{\inf} > \max\{\widehat{\Delta}_{n,U}^{\inf}(\alpha), 0\}$ and the latter implies $\Delta_{F_n}^{\inf} > \widehat{\Delta}_{n,U}^{\inf}(\alpha)$.

First, consider the case where $r_{F_n}^{\inf} > 0$ for all n large. Then, $r_{F_n}^{\inf} \leq \widehat{r}_{n,UP}(\alpha_1)$ implies that $0 < \widehat{r}_{n,UP}(\alpha_1)$ and $\phi_{n,SPUR2}(\theta_n, \alpha_2) \leq \phi_{n,SPUR1}(\theta_n, \alpha_2)$ using (4.10). In this case, under $\{F_n\}_{n\geq 1}$ and $\{\theta_n\}_{n\geq 1}$, the rhs of (20.3) is less than or equal to

$$\limsup_{n\to\infty} P_{F_n}(\phi_{n,SPUR1}(\theta_n, \alpha_2)=1) + \alpha_1 \leq \alpha_2 + \alpha_1 = \alpha, \tag{20.4}$$

where the inequality holds because the nominal level α_2 test $\phi_{n,SPUR1}(\theta_n, \alpha_2)$ has asymptotic size α_2 or less by Theorem 4.1(b) for the SPUR1 CS (which allows for drifting sequences of null values θ_n).

Next, consider the case where $r_{F_n}^{\inf} = 0$ for all n large. Under $\{F_n\}_{n\geq 1}$ and $\{\theta_n\}_{n\geq 1}$, the rhs of (20.3) is less than or equal to

$$\limsup_{n\to\infty} P_{F_n}(\phi_{n,GMS}(\theta_n, \alpha_2)=1) + \alpha_1 \leq \alpha_2 + \alpha_1 = \alpha, \tag{20.5}$$

where the inequality holds because the model is correctly specified (i.e., $r_{F_n}^{\inf} = 0$) for n large and the $\phi_{n,GMS}(\theta_n, \alpha_2)$ test has asymptotic size α_2 or less in this case. The latter holds by the same argument as used to prove Theorem 4.1(b) for the SPUR1 CS (which allows for drifting sequences of null values θ_n), but with the test statistic $S_n(\theta)$ defined in (4.5) with $\widehat{r}_n^{\inf}$ replaced by the true value $r_{F_n}^{\inf} = 0$ and with the EGMS bootstrap statistic replaced by the GMS bootstrap statistic $S_{n,GMS}^*(\theta)$ defined just above (4.10), which is suitable because $r_{F_n}^{\inf} = 0$.

The result of part (b) for the SPUR2 CS holds because the rhs of (20.3) for the sequence $\{F_n\}_{n\geq 1}$ is α or less by considering subsequences of $\{n\}$ where either (20.4) or (20.5) applies.

The proof of part (a) for the SPUR2 test is analogous to that of part (b) for the SPUR2 CS with θ_0 in place of θ_n $\forall n \geq 1$. $\square$

Proof of Lemma 20.1. First, we verify Assumption C.5 using Lemma D.2(1) of BCS, which imposes their Assumptions A.1–A.4 and M.2 and $\Omega_{F_n} \to_u \Omega_\infty$ for some Ω_∞. Assumptions A.1–A.4 in this paper imply A.1–A.4 in BCS, Assumption A.0(i) is the same as BCS's M.2, and Assumption C.11 implies $\Omega_{F_n} \to_u \Omega_\infty$. Lemma D.2(1) of BCS gives $\nu_n^m(\cdot) \Rightarrow G^m(\cdot)$, whereas Assumption C.5 concerns $\nu_n(\cdot) := (\nu_n^m(\cdot)', \nu_n^\sigma(\cdot)')'$. However, by the same argument as in the proof of Lemma D.2(1) applied to $\nu_n(\cdot)$, rather than $\nu_n^m(\cdot)$, we obtain

$$\nu_n(\cdot) \Rightarrow G(\cdot), \tag{20.6}$$

where $G(\cdot)$ is as in Assumption C.5, using equicontinuity of $\nu_n(\cdot)$ in our Assumption A.3, rather than of $\nu_n^m(\cdot)$ in BCS's Assumption A.2, and using $4+a$ finite moments in our Assumption A.2, rather than $2+a$ finite moments in BCS's Assumption A.3. Hence, Assumption C.5 holds and part (a) is established.

Next, we verify Assumption C.6. Lemma D.2(5) of BCS gives $\sup_{\theta\in\Theta} ||\widehat{\Omega}_n(\theta) - \Omega_{\infty 11}(\theta,\theta)|| \to_p 0$, where $\Omega_{\infty 11}(\theta,\theta)$ denotes the upper left $k \times k$ submatrix of $\Omega_\infty(\theta,\theta)$, because Assumptions A.1–A.4 in this paper imply Assumptions A.1–A.4 of BCS and $\Omega_{F_n} \to_u \Omega_\infty$ by Assumption C.11. By Assumption C.1, $\theta_n \to \theta_\infty$, and by Assumption C.11, $\Omega_\infty(\theta,\theta')$ is continuous on Θ^2. These results combine to yield $\widehat{\Omega}_n(\theta_n) \to_p \Omega_{\infty 11}(\theta_\infty,\theta_\infty) := \Omega_\infty$, which verifies Assumption C.6 and establishes part (b). $\square$

21 Proofs of Lemma 11.2 and Theorem 11.1

The proof of Lemma 11.2(b) uses the following lemma, which shows that Assumption C.10 implies a similar minorant condition on the sample analogue of the left-hand side of Assumption C.10.

Lemma 21.1 *Suppose Assumptions* A.0, C.4, C.5, C.7, *and* C.10 *hold under* $\{F_n\}_{n\geq 1}$. *Then, there exist positive constants* κ, ε, *and* γ *such that for any* $\delta \in (0,1)$ *there exists positive constants* κ_δ

and N_δ such that

$$\max_{j\le k}([\widehat{m}_{nj}(\theta)]_- - \widehat{r}_n^{\inf}) \ge \kappa\cdot(\min\{d(\theta,\Theta_I^{MR}(F_n)),\varepsilon\})^\gamma$$

for all $\theta \in \{\theta\in\Theta : d(\theta,\Theta_I^{MR}(F_n)) \ge (\kappa_\delta/n^{1/2})^{1/\gamma}\}$ with probability at least $1-\delta$ for all $n\ge N_\delta$.

Proof of Lemma 11.2. The proof is similar to that of Theorem 3.1 of Chernozhukov, Hong, and Tamer (2007). For part (a), we have

$$\sup_{\theta\in\Theta_I^{MR}(F_n)} d(\theta,\widehat{\Theta}_n) = 0 \text{ wp} \to 1 \tag{21.1}$$

because $\Theta_I^{MR}(F_n)\subset\widehat{\Theta}_n$ wp $\to 1$ by Lemma 19.1(a) (which requires Assumptions A.0, C.4, C.5, and C.7). For part (a), it remains to show $\sup_{\theta\in\widehat{\Theta}_n} d(\theta,\Theta_I^{MR}(F_n)) = o_p(1)$.

By Assumption C.9, for arbitrary $\varepsilon>0$, we have

$$\zeta_\varepsilon := \liminf_{n\to\infty}\inf_{\theta\in\Theta\backslash\Theta_{I,\varepsilon}^{MR}(F_n)}\max_{j\le k}\,[E_{F_n}\widetilde{m}_j(W,\theta)]_- - r_{F_n}^{\inf} > 0. \tag{21.2}$$

Next, we have

$$\begin{aligned}
&\sup_{\theta\in\widehat{\Theta}_n}\max_{j\le k} n^{1/2}([E_{F_n}\widetilde{m}_j(W,\theta)]_- - r_{F_n}^{\inf})\\
&= \sup_{\theta\in\widehat{\Theta}_n}\max_{j\le k} n^{1/2}([E_{F_n}\widetilde{m}_j(W,\theta)]_- - [\widehat{m}_{nj}(\theta)]_- + [\widehat{m}_{nj}(\theta)]_- - \widehat{r}_n^{\inf} + \widehat{r}_n^{\inf} - r_{F_n}^{\inf})\\
&= \sup_{\theta\in\widehat{\Theta}_n}\max_{j\le k} n^{1/2}([E_{F_n}\widetilde{m}_j(W,\theta)]_- - [\widehat{m}_{nj}(\theta)]_- + [\widehat{m}_{nj}(\theta)]_- - \widehat{r}_n^{\inf}) + O_p(1)\\
&\le \sup_{\theta\in\widehat{\Theta}_n}\max_{j\le k} n^{1/2}([E_{F_n}\widetilde{m}_j(W,\theta)]_- - [\widehat{m}_{nj}(\theta)]_-) + \tau_n + O_p(1)\\
&= \sup_{\theta\in\widehat{\Theta}_n}\max_{j\le k}\Big([n^{1/2}E_{F_n}\widetilde{m}_j(W,\theta)]_- - [\nu_{nj}^{m\sigma}(\theta) + n^{1/2}E_{F_n}\widetilde{m}_j(W,\theta)]_-\Big) + \tau_n + O_p(1)\\
&\le \sup_{\theta\in\widehat{\Theta}_n}\max_{j\le k}\left|\nu_{nj}^{m\sigma}(\theta)\right| + \tau_n + O_p(1)\\
&= O_p(1) + \tau_n,
\end{aligned} \tag{21.3}$$

where the second equality holds by Theorem 15.3(b) (which requires Assumptions A.0, C.4, C.5, and C.7), the first inequality holds by the definition of $\widehat{\Theta}_n$ and the same reasoning as given following (19.49), the third equality holds by (17.5) and (17.6), the second inequality holds by (17.7), and the last equality holds by Assumption C.5.

By (21.3), we have

$$\sup_{\theta\in\widehat{\Theta}_n}\max_{j\le k}[E_{F_n}\widetilde{m}_j(W,\theta)]_- - r_{F_n}^{\inf} \le O_p(1/n^{1/2}) + \tau_n/n^{1/2} = o_p(1), \tag{21.4}$$

where the equality holds because $\tau_n/n^{1/2} = o(1)$. Combining (21.2) and (21.4), it follows that

$$\begin{aligned}
&\lim P_{F_n}\left(\inf_{\theta\in\Theta\backslash\Theta_{I,\varepsilon}^{MR}(F_n)} \max_{j\le k}\ [E_{F_n}\widetilde{m}_j(W,\theta)]_- > \sup_{\theta\in\widehat{\Theta}_n} \max_{j\le k}\ [E_{F_n}\widetilde{m}_j(W,\theta)]_-\right)\\
&\ge \lim P_{F_n}(\zeta_\varepsilon/2 > o_p(1))\\
&= 1. \qquad (21.5)
\end{aligned}$$

Thus, $\lim P_{F_n}(\widehat{\Theta}_n \subset \Theta_{I,\varepsilon}^{MR}(F_n)) = 1$ and $\sup_{\theta\in\widehat{\Theta}_n} d(\theta, \Theta_I^{MR}(F_n)) \le \varepsilon$ wp$\to$ 1. Since $\varepsilon > 0$ is arbitrary, we have $\sup_{\theta\in\widehat{\Theta}_n} d(\theta, \Theta_I^{MR}(F_n)) = o_p(1)$, which completes the proof of part (a).

For part (b), take the positive constants $(\kappa, \varepsilon, \gamma, \delta, N_\delta, \kappa_\delta)$ as in Lemma 21.1. We can take $N'_\delta \ge N_\delta$ such that $2\tau_n > \kappa\cdot\kappa_\delta$ and $\varepsilon_n := (2\tau_n/(n^{1/2}\kappa))^{1/\gamma} < \varepsilon$ for $n \ge N'_\delta$, because $\tau_n \to \infty$ and $\tau_n/n^{1/2} = o(1)$. As defined, $\varepsilon_n > (\kappa_\delta/n^{1/2})^{1/\gamma}$ for $n \ge N'_\delta$. Hence,

$$\Theta\backslash\Theta_{I,\varepsilon_n}^{MR} \subset \{\theta\in\Theta : d(\theta, \Theta_I^{MR}(F_n)) \ge (\kappa_\delta/n^{1/2})^{1/\gamma}\} \qquad (21.6)$$

for $n \ge N'_\delta$. In consequence, with probability at least $1-\delta$ for $n \ge N'_\delta$, we have

$$\begin{aligned}
\inf_{\theta\in\Theta\backslash\Theta_{I,\varepsilon_n}^{MR}(F_n)} \max_{j\le k}([\widehat{m}_{nj}(\theta)]_- - \widehat{r}_n^{\inf}) &\ge \kappa\cdot \inf_{\theta\in\Theta\backslash\Theta_{I,\varepsilon_n}^{MR}(F_n)} (\min\{d(\theta, \Theta_I^{MR}(F_n)), \varepsilon\})^\gamma\\
&\ge \kappa\cdot(\min\{\varepsilon_n, \varepsilon\})^\gamma\\
&= \kappa\cdot\varepsilon_n^\gamma\\
&:= 2\tau_n/n^{1/2}\\
&> \tau_n/n^{1/2}\\
&\ge \sup_{\theta\in\widehat{\Theta}_n} \max_{j\le k}([\widehat{m}_{nj}(\theta)]_- - \widehat{r}_n^{\inf}), \qquad (21.7)
\end{aligned}$$

where the first inequality holds by Lemma 21.1 and (21.6), the second inequality holds by the definition of $\Theta_{I,\varepsilon_n}^{MR}(F_n)$, the first equality holds by the definition of N'_δ, the second equality holds by the definition of ε_n, and the last holds inequality by the definition of $\widehat{\Theta}_n$.

Equation (21.7) implies $\widehat{\Theta}_n \subset \Theta_{I,\varepsilon_n}^{MR}(F_n)$, and hence, $\sup_{\theta\in\widehat{\Theta}_n} d(\theta, \Theta_I^{MR}(F_n)) \le \varepsilon_n$ with probability at least $1-\delta$ for $n \ge N'_\delta$. Combining this with (21.1) gives

$$d_H(\widehat{\Theta}_n, \Theta_I^{MR}(F_n)) = O_p(\varepsilon_n) = O_p((\tau_n/n^{1/2})^{1/\gamma}), \qquad (21.8)$$

which completes the proof of part (b). □

Proof of Lemma 21.1. By (19.50) with Θ in place of $\Theta_I^{\eta_n}(F_n)$ throughout and with $[E_{F_n}\widetilde{m}_j(W,\theta)]_-$

$-\, r_{F_n}^{\inf}$ in place of η_n in the last two lines (which makes the inequality into an equality), we have

$$\max_{j\leq k}([\widehat{m}_{nj}(\theta)]_{-} - \widehat{r}_n^{\inf}) = \max_{j\leq k}[E_{F_n}\widetilde{m}_j(\theta)]_{-} - r_{F_n}^{\inf} + O_p^{\Theta}(1/n^{1/2}) \tag{21.9}$$

using Assumptions A.0, C.4, C.5, and C.7. Hence, for any $\delta \in (0,1)$, there exist positive constants κ_δ and N_δ such that with probability at least $1-\delta$, we have

$$\begin{aligned}\max_{j\leq k}([\widehat{m}_{nj}(\theta)]_{-} - \widehat{r}_n^{\inf}) &\geq C\cdot(\min\{d(\theta,\Theta_I^{MR}(F_n)),\varepsilon\})^{\gamma} + O_p^{\Theta}(1/n^{1/2})\\ &\geq C\cdot(\min\{d(\theta,\Theta_I^{MR}(F_n)),\varepsilon\})^{\gamma} - (C/2)\kappa_\delta/n^{1/2}\end{aligned} \tag{21.10}$$

for all $\theta \in \Theta$ and $n \geq N_\delta$, where C, ε, and γ are as in Assumption C.10 and the first inequality uses (21.9) and Assumption C.10. Without loss in generality, we can take $N_\delta \geq (\kappa_\delta/\varepsilon^{\gamma})^2$. Hence, $\kappa_\delta/N_\delta^{1/2} \leq \varepsilon^{\gamma}$.

For all $n \geq N_\delta$, we have

$$\kappa_\delta/n^{1/2} \leq (\min\{d(\theta,\Theta_I^{MR}(F_n)),\varepsilon\})^{\gamma} \tag{21.11}$$

for all $\theta \in \{\theta \in \Theta : d(\theta,\Theta_I^{MR}(F_n)) \geq (\kappa_\delta/n^{1/2})^{1/\gamma}\}$. Combining (21.10) and (21.11) establishes the lemma with $\kappa = C/2$. $\square$

Proof of Theorem 11.1. Let an arbitrary $\varepsilon > 0$ be given. There always exists a sequence $\{F_n \in \mathcal{P}\}_{n\geq 1}$ (that may depend on ε) such that

$$\limsup_{n\to\infty}\sup_{F\in\mathcal{P}} P_F(d_H(\widehat{\Theta}_n,\Theta_I^{MR}(F)) > \varepsilon) = \limsup_{n\to\infty} P_{F_n}(d_H(\widehat{\Theta}_n,\Theta_I^{MR}(F_n)) > \varepsilon). \tag{21.12}$$

There always exists a subsequence $\{w_n\}_{n\geq 1}$ of $\{n\}_{n\geq 1}$ such that

$$\limsup_{n\to\infty} P_{F_n}(d_H(\widehat{\Theta}_n,\Theta_I^{MR}(F_n)) > \varepsilon) = \lim P_{F_{w_n}}(d_H(\widehat{\Theta}_{w_n},\Theta_I^{MR}(F_{w_n})) > \varepsilon). \tag{21.13}$$

Given any subsequence $\{a_n\}_{n\geq 1}$ of $\{w_n\}_{n\geq 1}$, there exists a subsequence $\{u_n\}_{n\geq 1}$ of $\{a_n\}_{n\geq 1}$ such that Assumptions C.4, C.7, and C.11 (defined in online Appendix B) hold for the subsequence $\{u_n\}_{n\geq 1}$ by the proof of Theorem 4.1 in Section 20 in online Appendix B, which uses Lemma D.7 of BCS and relies on Assumptions A.4 and A.5. Given Assumption A.9, Assumption C.9 also holds for the subsequence $\{u_n\}_{n\geq 1}$. By Lemma 20.1 in Section 20 in online Appendix B, Assumptions A.0–A.4 and C.11 imply Assumption C.5. Hence, Assumptions C.4, C.5, C.7, and C.9 hold for the subsequence $\{u_n\}_{n\geq 1}$. In consequence, by Lemma 11.2(a) applied with n replaced by u_n, which utilizes Assumptions A.0, C.4, C.5, C.7, and C.9, we have

$$\lim P_{F_{u_n}} d_H(\widehat{\Theta}_{u_n}, \Theta_I^{MR}(F_{u_n})) > \varepsilon) = 0. \tag{21.14}$$

This implies that the same result holds for the subsequence $\{w_n\}_{n\geq 1}$, which completes the proof using (21.12) and (21.13) because $\varepsilon > 0$ is arbitrary. □

22 Online Appendix B Assumptions

For ease of reference, we state all of the assumptions used in the paper and online Appendix B here.

Assumption A.0. (i) Θ is compact and non-empty and (ii) $E_F \widetilde{m}_j(W, \theta)$ is upper semi-continuous on Θ $\forall j \leq k$, $\forall F \in \mathcal{P}$.

Assumption A.1. The observations $W_1, ..., W_n$ are i.i.d. under F and $\{\widetilde{m}_j(\cdot, \theta) : \mathcal{W} \to R\}$ and $\{\widetilde{m}_j^2(\cdot, \theta) : \mathcal{W} \to R\}$ are measurable classes of functions indexed by $\theta \in \Theta$ $\forall j \leq k$, $\forall F \in \mathcal{P}$.

Assumption A.2. For some $a > 0$, $\sup_{F\in\mathcal{P}} E_F \sup_{\theta\in\Theta} ||\widetilde{m}(W, \theta)||^{4+a} < \infty$.

Assumption A.3. The empirical process $\nu_n(\cdot)$ is asymptotically ρ_F-equicontinuous on Θ uniformly in $F \in \mathcal{P}$.

Assumption A.4. The covariance kernel $\Omega_F(\theta, \theta')$ satisfies: for all $F \in \mathcal{P}$,

$\lim_{\delta\to 0} \sup_{||(\theta_1,\theta_1')-(\theta_2,\theta_2')||<\delta} ||\Omega_F(\theta_1, \theta_1') - \Omega_F(\theta_2, \theta_2')|| = 0.$

Assumption A.5. $E_F \widetilde{m}(W, \theta)$ is equicontinuous on Θ over $F \in \mathcal{P}$. That is, $\lim_{\delta\downarrow 0} \sup_{F\in\mathcal{P}} \sup_{||\theta-\theta'||<\delta} ||E_F \widetilde{m}(W, \theta) - E_F \widetilde{m}(W, \theta')|| = 0$.

Assumption A.6. (i) $\kappa_n \to \infty$. (ii) $\tau_n \to \infty$.

Let $\Psi := cl(\{\Omega_F(\theta) : \theta \in \Theta, F \in \mathcal{P}\})$, where $cl(\cdot)$ denotes the closure of a set and $\Omega_F(\theta) := Corr_F(m(W, \theta)) \in R^{k\times k}$.

Assumption S.1. (i) $S(m, \Omega)$ is nonincreasing in $m \in R^k_{[+\infty]}$ $\forall\Omega \in \Psi$,

(ii) $S(m, \Omega) \geq 0$ $\forall m \in R^k$, $\forall\Omega \in \Psi$, and

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

Assumption S.2. $S(m, \Omega) > 0$ iff $m_j < 0$ for some $j \leq k$, $\forall\Omega \in \Psi$.

Assumption S.3. For some $\chi > 0$, $S(am, \Omega) = a^\chi S(m, \Omega)$ $\forall a > 0$, $\forall m \in R^k$, $\forall\Omega \in \Psi$.

Assumption S.4. For all $h \in (-\infty, \infty]^k$, all $\Omega \in \Psi$, and $Z \sim N(0_k, \Omega)$, the distribution function of $S(Z + h, \Omega)$ at $x \in R$ is (i) continuous for $x > 0$, (ii) strictly increasing for $x > 0$ unless $h = (\infty, \ldots, \infty)' \in R^k_{[\pm\infty]}$, and (iii) less than 1/2 for $x = 0$ if $h_j = 0$ for some $j \leq k$.

The following assumptions apply to a drifting sequence of null values $\{\theta_n\}_{n\geq1}$ and distributions $\{F_n\}_{n\geq1}$.

Assumption C.1. $\theta_n \to \theta_\infty$ for some $\theta_\infty \in \Theta$.

Assumption C.2. $n^{1/2}E_{F_n}\widetilde{m}_j(W,\theta_n) \to \ell_{j\infty}$ for some $\ell_{j\infty} \in R_{[\pm\infty]}$ $\forall j \leq k$.

Assumption C.3. $n^{1/2}(E_{F_n}\widetilde{m}_j(W,\theta_n) + r^{\inf}_{F_n}) \to h_{j\infty}$ for some $h_{j\infty} \in R_{[\pm\infty]}$ $\forall j \leq k$.

Assumption C.4. $\sup_{\theta\in\Theta} ||E_{F_n}\widetilde{m}(W,\theta) - \widetilde{m}(\theta)|| \to 0$ for some nonrandom bounded continuous R^k-valued function $\widetilde{m}(\cdot)$ on Θ.

Assumption C.5. $\nu_n(\cdot) := (\nu_n^m(\cdot)', \nu_n^\sigma(\cdot)')' \Rightarrow G(\cdot) := (G^m(\cdot)', G^\sigma(\cdot)')'$ as $n \to \infty$, where $\{G(\theta) : \theta \in \Theta\}$ is a mean zero R^{2k}-valued Gaussian process with bounded continuous sample paths a.s. and $G^m(\theta), G^\sigma(\theta) \in R^k$.

Assumption C.6. $\widehat{\Omega}_n(\theta_n) \to_p \Omega_\infty$ for some $\Omega_\infty \in \Psi$.

Assumption C.7. $\Lambda_{n,F_n} \to_H \Lambda$ for some non-empty set $\Lambda \in \mathcal{S}(\Theta \times R^{2k}_{[\pm\infty]})$.

Assumption C.8. $\Lambda^{\eta_n}_{n,F_n} \to_H \Lambda_I$ for some non-empty set $\Lambda_I \in \mathcal{S}(\Theta \times R^{2k}_{[\pm\infty]})$, where $\{\eta_n\}_{n\geq1}$ is a sequence of positive constants for which $\eta_n \to \infty$.

Assumption C.9. For all $\varepsilon > 0$,

$$\liminf_{n\to\infty}\left(\inf_{\theta\in\Theta\backslash\Theta^{MR}_{I,\varepsilon}(F_n)}\max_{j\leq k}\ [E_{F_n}\widetilde{m}_j(W,\theta)]_- - r^{\inf}_{F_n}\right) > 0.$$

Assumption C.10. There exist positive constants C, ε, and γ such that for all $\theta \in \Theta$ and $n \geq 1$,

$$\max_{j\leq k}[E_{F_n}\widetilde{m}_j(W,\theta)]_- - r^{\inf}_{F_n} \geq C \cdot (\min\{d(\theta, \Theta^{MR}_I(F_n)), \varepsilon\})^\gamma.$$

Assumption C.11. $\Omega_{F_n}(\cdot,\cdot) \to_u \Omega_\infty(\cdot,\cdot)$ for some continuous $R^{2k\times2k}$-valued function $\Omega_\infty(\cdot,\cdot)$ on Θ^2.

The following assumptions apply to a drifting sequence of null values $\{\theta_n\}_{n\geq1}$ and distributions $\{F_n\}_{n\geq1}$.

Assumption BC.1. $(\iota\kappa_n)^{-1}n^{1/2}(E_{F_n}\widetilde{m}_j(W,\theta_n) + r^{\inf}_{F_n}) \to h^*_{Lj\infty}$ for some $h^*_{Lj\infty} \in R_{[\pm\infty]}$ $\forall j \leq k$.

Assumption BC.2. $\Lambda^{*\eta_n}_{n,F_n} \to_H \Lambda^*_I$ for some non-empty set $\Lambda^*_I \in \mathcal{S}(\Theta \times R^{3k}_{[\pm\infty]} \times \{1,...,k\})$ for some constants $\{\eta_n\}_{n\geq1}$ that satisfy $\eta_n \to \infty$ and $\eta_n/\tau_n \to 0$ for $\{\tau_n\}_{n\geq1}$ as in Assumption A.6(ii).

Assumption BC.3. $\{\nu^*_n(\cdot)|\{W_{ni}\}_{i\leq n,n\geq1}\} \Rightarrow G(\cdot)$ a.s.$[P_\triangledown]$, where $G(\cdot)$ is as in Assumption C.5.

Assumption NLA. $\min_{j\leq k} h_{j\infty} > -\infty$.

Assumption CA. $\min_{j\leq k} h_{j\infty} = -\infty$.

Assumption N. $\theta_n \in \Theta^{MR}_I(F_n)$ $\forall n \geq 1$.

Assumption LA. The null values $\{\theta_n\}_{n\geq1}$ and distributions $\{F_n\}_{n\geq1}$ satisfy: (i) $||\theta_n - \theta_{In}|| =$

$O(n^{-1/2})$ for some sequence $\{\theta_{In} \in \Theta_I^{MR}(F_n)\}_{n\geq 1}$, (ii) $n^{1/2}(E_{F_n}\widetilde{m}_j(W,\theta_{In})+r_{F_n}^{\inf}) \to h_{Ij\infty}$ for some $h_{Ij\infty} \in R_{[\pm\infty]}$ $\forall j \leq k$, and (iii) $E_F\widetilde{m}(W,\theta)$ is Lipschitz on Θ uniformly over $\mathcal{P}$, i.e., there exists a constant $K < \infty$ such that $||E_F\widetilde{m}(W,\theta_1) - E_F\widetilde{m}(W,\theta_2)|| \leq K||\theta_1 - \theta_2||$ $\forall \theta_1, \theta_2 \in \Theta$, $\forall F \in \mathcal{P}$.

Assumption FA. The null values $\{\theta_n\}_{n\geq 1}$ and distributions $\{F_n\}_{n\geq 1}$ satisfy: (i) $F_n = F_* \in \mathcal{P}$ and $\theta_n = \theta_* \in \Theta$ do not depend on $n \geq 1$ and (ii) $E_{F_*}\widetilde{m}_j(W,\theta_*) + r_{F_*}^{\inf} < 0$ for some $j \leq k$.

Assumption SLK. The sequence $\{F_n\}_{n\geq 1}$ is such that $n^{1/2}\Delta_{F_n}^{\inf} \to -\infty$.

Assumption MM. The sequence $\{F_n\}_{n\geq 1}$ is such that $n^{1/2}\Delta_{F_n}^{\inf} \to \infty$.

Online Appendix B References

Bugni, F. A., I. A. Canay, and X. Shi (2015), "Specification Tests for Partially Identified Models Defined by Moment Inequalities," *Journal of Econometrics,* 185, 259–282.

Chernozhukov, V., H. Hong, and E. Tamer (2007), "Estimation and Confidence Regions for Parameter Sets in Econometric Models," *Econometrica,* 75, 1243–284.

Dudley, R. M. (1989), *Real Analysis and Probability.* Pacific Grove, CA: Wadsworth and Brook/Cole.

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

Online Appendix C

to

Misspecified Moment Inequality Models: Inference and Diagnostics

Contents: Proofs of the Misspecification Index CI Results

Donald W. K. Andrews

Cowles Foundation for Research in Economics

Yale University

Soonwoo Kwon

Department of Economics

Brown University

First Version: November, 2018

Revised: July 29, 2022

23 Outline of Online Appendix C

Online Appendix C proves the results of the paper for the misspecification index confidence intervals (CI's).

References to sections with section numbers 6 or less refer to sections of the main paper. Similarly, all equations, theorems, and lemmas with section numbers 6 or less refer to results in the main paper. Let BCS15 abbreviate Bugni, Canay, and Shi (2015) and BCS17 abbreviate Bugni, Canay, and Shi (2017). For ease of reference, the assumptions used in the paper and online Appendix C are listed in the last section of online Appendix C, Section 31.

Section 24 provides an equivalent condition to Assumption SLK, which is employed in Theorem 5.2. It also provides a set of sufficient conditions for Assumption SLK.

Section 25 provides the asymptotic distribution of $\widehat{\Delta}_n^{\inf}$ under certain drifting sequences of distributions $\{F_n\}_{n\geq 1}$ under some high-level conditions, which are verified below. The asymptotic distribution results are used below to prove Theorem 5.1, which establishes the correct asymptotic size of the MI CI's $CI_{n,\Delta U}(\alpha)$, $CI_{n,\Delta L}(\alpha)$, and $CI_{n,\Delta}(\alpha)$.

Section 26 proves Lemma 25.1 and Theorem 25.2, which gives the asymptotic distribution of $\widehat{\Delta}_n^{\inf}$.

Section 27 proves Theorem 5.1, which establishes the correct asymptotic size of the upper- and lower-bound CI's for $\Delta_F^{\inf}$.

Section 28 proves Theorem 5.2, which gives conditions for the upper-bound CI to contain only negative values wp$\to$1 and conditions for the lower-bound CI to contain only positive values wp$\to$1.

Section 29 proves Corollary 5.3, which establishes the correct asymptotic size of the tests concerning $\Delta_F^{\inf}$ and conditions for the consistency of the tests.

Section 30 shows that the upper-bound CI $CI_{n,\Delta U}(\alpha)$ includes positive values of $\Delta_F^{\inf}$ wp$\to$1 when the model is misspecified and $n^{1/2}\Delta_{F_n}^{\inf} \to \infty$.

All limits are as the sample size $n \to \infty$. Let $o_p^{\Theta}(1)$ and $O_p^{\Theta}(1)$ denote random functions that are $o_p(1)$ and $O_p(1)$ uniformly over $\theta \in \Theta$, respectively. Let $R_{[\pm\infty]} := R \cup \{\pm\infty\}$ and $R_{[+\infty]} := R \cup \{+\infty\}$. Let $||\cdot||$ denote the Euclidean norm for vectors and the Frobenious norm for matrices.

24 Equivalent and Sufficient Conditions for Assumption SLK

First, we give an equivalent condition to Assumption SLK (under Assumption A.0).

Lemma 24.1 *Suppose Assumption* A.0 *holds. Then, the sequence* $\{F_n\}_{n\geq 1}$ *satisfies Assumption* SLK *if and only if there exists a sequence* $\{\theta_n^I \in \Theta_I(F_n)\}_{n\geq 1}$ *for which* $n^{1/2}E_{F_n}\widetilde{m}_j(W,\theta_n^I) \to \infty$

$\forall j \leq k$.

Next, we give some sufficient conditions for Assumption SLK. For any $\Theta_1, \Theta_2 \subseteq \Theta$, the Hausdorff distance between Θ_1 and Θ_2 is

$$d_H(\Theta_1, \Theta_2) = \max\left\{\sup_{\theta_1\in\Theta_1}\inf_{\theta_2\in\Theta_2}\|\theta_1-\theta_2\|, \sup_{\theta_2\in\Theta_2}\inf_{\theta_1\in\Theta_1}\|\theta_1-\theta_2\|\right\}. \quad (24.1)$$

If $\Theta_1 \subseteq \Theta_2$, $d_H(\Theta_1, \Theta_2) = \sup_{\theta_2\in\Theta_2}\inf_{\theta_1\in\Theta_1}\|\theta_1-\theta_2\|$, because $\sup_{\theta_1\in\Theta_1}\inf_{\theta_2\in\Theta_2}\|\theta_1-\theta_2\| = 0$ in this case. The following assumption states that the identified set $\Theta_I(F_n)$ does not shrink to the $\Theta_{\min}(F_n)$ set too quickly.

Assumption SLK.1. The sequence $\{F_n\}_{n\geq1}$ is such that $\{\Theta_I(F_n)\}_{n\geq1}$ are nonempty and $\lim_{n\to\infty} n^{1/2}d_H(\Theta_{\min}(F_n), \Theta_I(F_n)) = \infty$.

For example, when $\Theta_{\min}(F_n)$ is a singleton and $\Theta_I(F_n)$ is such that its diagonal does not shrink at rate $n^{-1/2}$ or faster, Assumption SLK.1 holds.[50]

Lemma 24.2 *Suppose the sequence* $\{F_n\}_{n\geq1}$ *satisfies Assumptions* A.0, A.8(i), *and* SLK.1, *and the model is correctly specified for each* F_n *in the sequence. Then,* $\{F_n\}_{n\geq1}$ *satisfies Assumption* SLK.

Comment. Lemma 24.2 still holds if $d_H(\Theta_{\min}(F_n), \Theta_I(F_n))$ is replaced by $(d_H(\Theta_{\min}(F_n), \Theta_I(F_n)))^q$ in Assumption SLK.1 and $\inf_{\overline{\theta}\in\Theta_{\min}(F)}\|\theta-\overline{\theta}\|$ is replaced by $\inf_{\overline{\theta}\in\Theta_{\min}(F)}\|\theta-\overline{\theta}\|^q$ in Assumption A.8(i), for any $q > 0$.

Proof of Lemma 24.1. Suppose there exists a sequence $\{\theta_n^I \in \Theta_I(F_n)\}_{n\geq1}$ for which $n^{1/2}E_{F_n}\widetilde{m}_j(W, \theta_n^I) \to \infty$ $\forall j \leq k$. Then,

$$n^{1/2}\Delta_{F_n}^{\inf} \leq n^{1/2}\Delta_{F_n}(\theta_n^I) =: \max_{j\leq k} -n^{1/2}E_{F_n}\widetilde{m}_j(W, \theta_n^I) = -\min_{j\leq k} n^{1/2}E_{F_n}\widetilde{m}_j(W, \theta_n^I) \to -\infty, \quad (24.2)$$

where the inequality holds by the definition of $\Delta_{F_n}^{\inf}$ and the divergence holds by the assumption.

To show the converse, suppose Assumption SLK holds. By Assumption A.0, $\Theta_{\min}(F_n)$ is nonempty for all $n \geq 1$. For any sequence $\{\theta_n^I \in \Theta_{\min}(F_n)\}_{n\geq1}$,

$$\min_{j\leq k} n^{1/2}E_{F_n}\widetilde{m}_j(W, \theta_n^I) =: -n^{1/2}\Delta_{F_n}(\theta_n^I) = -n^{1/2}\Delta_{F_n}^{\inf} \to \infty, \quad (24.3)$$

where the first equality holds by the definition of $\Delta_{F_n}(\cdot)$, the second inequality holds because $\theta_n^I \in \Theta_{\min}(F_n)$, and the divergence holds by Assumption SLK. This completes the proof. $\square$

[50] The diagonal of a set $A \subset R^k$ is defined as $\sup_{x,y\in A}\|x-y\|$.

Proof of Lemma 24.2. Define $d(\theta, \Theta_{\min}(F_n)) := \inf_{\overline{\theta} \in \Theta_{\min}(F_n)} \|\theta - \overline{\theta}\|$. Since $\Theta_{\min}(F_n) \subseteq \Theta_I^{MR}(F_n) = \Theta_I(F_n)$ (where the second equality holds by assumption), we have

$$d_H(\Theta_{\min}(F_n), \Theta_I(F_n)) = \sup_{\theta \in \Theta_I(F_n)} \inf_{\overline{\theta} \in \Theta_{\min}(F_n)} \|\theta - \overline{\theta}\| =: \sup_{\theta \in \Theta_I(F_n)} d(\theta, \Theta_{\min}(F_n)), \tag{24.4}$$

where the first equality follows from the sentence following (24.1) and the second equality holds by definition. Note that $d(\theta, \Theta_{\min}(F_n))$ is continuous in θ and $\Theta_I(F_n)$ is compact by Assumption A.0. Hence, there exists a sequence $\{\widetilde{\theta}_n^I \in \Theta_I(F_n)\}_{n \geq 1}$ such that $\widetilde{\theta}_n^I$ achieves the supremum on the right-hand side (rhs) of (24.4). It follows that

$$d(\widetilde{\theta}_n^I, \Theta_{\min}(F_n)) = \sup_{\theta \in \Theta_I(F_n)} d(\theta, \Theta_{\min}(F_n)) = d_H(\Theta_{\min}(F_n), \Theta_I(F_n)). \tag{24.5}$$

We have

$$\begin{aligned} \Delta_{F_n}^{\inf} &\leq \Delta_{F_n}(\widetilde{\theta}_n^I) - c \min\left\{\delta, \inf_{\overline{\theta} \in \Theta_{\min}(F_n)} \|\widetilde{\theta}_n^I - \overline{\theta}\|\right\} \\ &\leq -c \min\left\{\delta, \inf_{\overline{\theta} \in \Theta_{\min}(F_n)} \|\widetilde{\theta}_n^I - \overline{\theta}\|\right\} \\ &= -c \min\{\delta, d_H(\Theta_{\min}(F_n), \Theta_I(F_n))\}, \end{aligned} \tag{24.6}$$

where the first inequality holds by Assumption A.8(i), the second inequality holds because the model is assumed to be correctly specified and $\widetilde{\theta}_n^I \in \Theta_I(F_n)$, and the equality holds by (24.1).

Multiplying both sides of (24.6) by $n^{1/2}$ and taking the $\limsup_{n \to \infty}$ gives

$$\limsup_{n \to \infty} n^{1/2} \Delta_{F_n}^{\inf} \leq -c \min\{\liminf_{n \to \infty} n^{1/2} \delta, \liminf_{n \to \infty} n^{1/2} d_H(\Theta_{\min}(F_n), \Theta_I(F_n))\} \to -\infty, \tag{24.7}$$

where the divergence holds by Assumption SLK.1. Thus, $n^{1/2} \Delta_{F_n}^{\inf} \to -\infty$ and Assumption SLK holds. $\square$

25 Asymptotic Distribution of the Estimator $\widehat{\Delta}_n^{\inf}$

In this section, we obtain the asymptotic distribution of $\widehat{\Delta}_n^{\inf}$ under certain drifting sequences of distributions $\{F_n\}_{n \geq 1}$. The asymptotic distribution is obtained under some high-level conditions which are verified below. The results are used below to prove Theorem 5.1, which establishes the correct asymptotic size of the MI CI's $CI_{n,\Delta U}(\alpha)$, $CI_{n,\Delta L}(\alpha)$, and $CI_{n,\Delta}(\alpha)$.

25.1 High-Level Convergence Assumptions

As in BCS15, for any $x_1, x_2 \in R^{a_*}_{[\pm\infty]}$ for some positive integer a_*, let $d(x_1, x_2) = (\sum_{j=1}^{a_*}(\Phi(x_{1,j}) - \Phi(x_{2,j}))^2)^{1/2}$, where $\Phi : R_{[\pm\infty]} \to [0, 1]$, $\Phi(y)$ is the standard normal distribution function at y for $y \in R$, $\Phi(-\infty) := 0$, and $\Phi(\infty) := 1$. The space $(R^{a_*}_{[\pm\infty]}, d)$ is a compact metric space. Convergence in $(R^{a_*}_{[\pm\infty]}, d)$ to a point in R^{a_*} implies convergence under the Euclidean norm. Let $\mathcal{S}(\Theta \times R^k_{[\pm\infty]})$ denote the space of non-empty compact subsets of the metric space $(\Theta \times R^k_{[\pm\infty]}, d)$, where d is defined with $a_* = d_\theta + k$. Let $\Rightarrow$ denote weak convergence of a sequence of stochastic processes in the sense of van der Vaart and Wellner (1996). Let $\to_H$ denote convergence in Hausdorff distance (under d) for elements of $\mathcal{S}(\Theta \times R^k_{[\pm\infty]})$. We use the convention that $\nu + c = c$ when $\nu \in R$ and $c = \pm\infty$. For any $e, m \in R^k$ that arise below, let e_j, m_j denote the jth elements of e, m, respectively.

The recentered and rescaled estimator $\widehat{\Delta}^{\inf}_n$ is

$$A_{n,\Delta} := n^{1/2}\left(\widehat{\Delta}^{\inf}_n - \Delta^{\inf}_{F_n}\right). \tag{25.1}$$

To obtain the asymptotic distribution of $A_{n,\Delta}$, we use the following sets:

$$\Lambda_{n,\Delta,F} := \left\{(\theta, e) \in \Theta \times R^k : e_j = n^{1/2}\left(\Delta_{Fj}(\theta) - \Delta^{\inf}_F\right)\right\}. \tag{25.2}$$

Define

$$\Delta_F(\theta) := \max_{j \le k} \Delta_{Fj}(\theta). \tag{25.3}$$

Note that $\Delta^{\inf}_F = \inf_{\theta\in\Theta} \Delta_F(\theta)$, see (5.1).

The set of minimizers of $\Delta_F(\theta)$ over Θ is

$$\Theta_{\min}(F) := \{\theta \in \Theta : \Delta_F(\theta) = \Delta^{\inf}_F\}. \tag{25.4}$$

Under Assumption A.0, $\Theta_{\min}(F)$ is non-empty. Note that $\Theta_{\min}(F)$ is a subset of $\Theta^{MR}_I(F)$ and equals $\Theta^{MR}_I(F)$ when $\Delta^{\inf}_F > 0$. For $\eta > 0$, define $\Theta^\eta_{\min}(F) := \{\theta \in \Theta : \Delta_F(\theta) \le \Delta^{\inf}_F + \eta/n^{1/2}\}$. The set $\Theta^\eta_{\min}(F)$ is an $\eta/n^{1/2}$-expansion of the minimizer set $\Theta_{\min}(F)$. It depends on n, but this is suppressed for notational simplicity. For $\eta > 0$, define $\Lambda^\eta_{n,\Delta,F_n}$ as Λ_{n,Δ,F_n} is defined in (25.2), but with $\Theta^\eta_{\min}(F_n)$ in place of Θ.

The asymptotic distribution of $A_{n,\Delta}$ utilizes the following Assumptions C.4, C.5, C.12, and C.13. These are high-level "convergence" assumptions that apply to a drifting sequence of distributions $\{F_n\}_{n\ge1}$. They are verified below using subsequence arguments when establishing the correct asymptotic size of the CI's for $\Delta^{\inf}_F$. Hence, they do not appear in the asymptotic size results stated below.

Assumption C.4. $\sup_{\theta\in\Theta}||E_{F_n}\widetilde{m}(W,\theta)-\widetilde{m}(\theta)||\to 0$ for some nonrandom bounded continuous R^k-valued function $\widetilde{m}(\cdot)$ on Θ.

Assumption C.5. $\nu_n(\cdot):=(\nu_n^m(\cdot)',\nu_n^\sigma(\cdot)')'\Rightarrow G(\cdot):=(G^m(\cdot)',G^\sigma(\cdot)')'$ as $n\to\infty$, where $\{G(\theta):\theta\in\Theta\}$ is a mean zero R^{2k}-valued Gaussian process with bounded continuous sample paths a.s. and $G^m(\theta),G^\sigma(\theta)\in R^k$.

Assumption C.12. $\Lambda_{n,\Delta,F_n}\to_H\Lambda_\Delta$ for some non-empty set $\Lambda_\Delta\in\mathcal{S}(\Theta\times R^k_{[\pm\infty]})$.

Assumption C.13. $\Lambda^{\eta_n}_{n,\Delta,F_n}\to_H\Lambda_{\Delta\,\min}$ for some non-empty set $\Lambda_{\Delta\,\min}\in\mathcal{S}(\Theta\times R^k_{[\pm\infty]})$, where $\{\eta_n\}_{n\geq 1}$ is a sequence of positive constants for which $\eta_n\to\infty$.

The elements (θ,e) of Λ_Δ and $\Lambda_{\Delta\,\min}$ in Assumptions C.12 and C.13 have the following properties.

Lemma 25.1 *Under* $\{F_n\}_{n\geq 1}$, (a) $\max_{j\leq k}e_{nj}(\theta)\geq 0\ \forall\theta\in\Theta,\forall n\geq 1,$ *where* $e_{nj}(\theta):=n^{1/2}(\Delta_{F_nj}(\theta)-\Delta^{\inf}_{F_n})$, (b) $\forall(\theta,e)\in\Lambda_\Delta$, $\max_{j\leq k}e_j\geq 0$ *provided Assumption* C.12 *holds,* (c) $\exists\widetilde{\theta}_n\in\Theta$ *with* $\max_{j\leq k}e_{nj}(\widetilde{\theta}_n)=0\ \forall n\geq 1$ *provided Assumption* A.0 *holds,* (d) $\exists(\widetilde{\theta},\widetilde{e})\in\Lambda_\Delta$ *with* $\max_{j\leq k}\widetilde{e}_j=0$ *provided Assumptions* A.0 *and* C.12 *hold, and* (e) $\exists(\widetilde{\theta},\widetilde{e})\in\Lambda_{\Delta\,\min}$ *with* $\max_{j\leq k}\widetilde{e}_j=0$ *provided Assumptions* A.0 *and* C.13 *hold.*

Comments. (i). Lemma 25.1 is used to show that the asymptotic distribution of $A_{n,\Delta}$ is in R a.s.

(ii). Lemma 25.1(a) and (b) are important because they allow one to obtain a (finite) lower bound on the asymptotic distribution of $A_{n,\Delta}$.

(iii). Lemma 25.1(c)–(e) are important because they allow one to obtain a (finite) upper bound on the asymptotic distribution of $A_{n,\Delta}$.

The following quantities arise with the asymptotic distribution of $A_{n,\Delta}$. Define

$$\begin{aligned}A_{n,\Delta}(\Lambda_{n,\Delta,F_n}) &:= \inf_{(\theta,e)\in\Lambda_{n,\Delta,F_n}}\max_{j\leq k}\left(-\nu^m_{nj}(\theta)+\frac{1}{2}\widetilde{m}_j(\theta)\nu^\sigma_{nj}(\theta)+e_j\right)\ \text{and}\\ A_{\infty,\Delta}:=A_{\infty,\Delta}(\Lambda_\Delta) &:= \inf_{(\theta,e)\in\Lambda_\Delta}\max_{j\leq k}\left(-G^m_j(\theta)+\frac{1}{2}\widetilde{m}_j(\theta)G^\sigma_j(\theta)+e_j\right)\end{aligned}\tag{25.5}$$

for Λ_Δ in Assumption C.12. Let $c_{\infty,\Delta}(1-\alpha)$ denote the $1-\alpha$ quantile of $A_{\infty,\Delta}$ and $c^-_{\infty,\Delta}(1-\alpha)$ denote the $1-\alpha$ quantile of $-A_{\infty,\Delta}$. We show below that $A_{n,\Delta}=A_{n,\Delta}(\Lambda_{n,\Delta,F_n})+o_p(1)\to_d A_{\infty,\Delta}$ under suitable sequences $\{F_n\}_{n\geq 1}$. Define

$$A_{\infty,\Delta\,\min}:=A_{\infty,\Delta}(\Lambda_{\Delta\,\min})\tag{25.6}$$

as in (25.5) with $\Lambda_{\Delta\,\min}$ in place of Λ_Δ, for $\Lambda_{\Delta\,\min}$ as in Assumption C.13.

25.2 Asymptotic Distribution of $\mathbf{A_{n,\Delta}}$

The asymptotic distribution of $A_{n,\Delta}$ is given in the following theorem.

Theorem 25.2 (a) *Under* $\{F_n\}_{n\geq 1}$ *and Assumptions* A.0, C.4, C.5, *and* C.12, $A_{n,\Delta} \to_d A_{\infty,\Delta}$,

(b) *under Assumptions* A.0 *and* C.12, $A_{\infty,\Delta} \in R$ *a.s., and*

(c) *under Assumptions* A.0 *and* C.4, C.5, C.12, *and* C.13, $A_{\infty,\Delta} = A_{\infty,\Delta\,\min}$ *a.s.*

Comments. **(i).** Theorem 25.2(b) is important because it implies that a critical value for an upper-bound or lower-bound CI based on the asymptotic distribution of $A_{n,\Delta}$ is finite.

(ii). Theorem 25.2(c) implies that the parameters $(\theta, e) \in \Lambda_\Delta \backslash \Lambda_{\Delta\,\min}$ do not contribute to the infimum in $A_{\infty,\Delta}$. This is useful when constructing critical values.

(iii). The quantity $G_j^\sigma(\cdot)$ appears in $A_{\infty,\Delta}$ because, under model misspecification, the asymptotic distribution of $A_{n,\Delta}$ depends on the randomness due to the estimation of the standard deviation of the jth sample moment by $\widehat{\sigma}_{nj}(\theta)$. Under correct model specification, it does not.

(iv). For any subsequence $\{q_n\}_{n\geq 1}$ of $\{n\}_{n\geq 1}$, Theorem 25.2 and its proof hold with q_n in place of n throughout, including the assumptions.

(v). The proof of Theorem 25.2(a) is similar proof to the proof of Theorem 3.1 of BCS15 with $S(m, \Omega) = \min_{j\leq k} m_j$ in their proof. The statistic $A_{n,\Delta}(\Lambda_{n,\Delta,F_n})$ depends on $e_{nj}(\theta) := n^{1/2}(\Delta_{F_nj}(\theta) - \Delta_{F_n}^{\inf})$, $\nu_{nj}^m(\theta)$, and $\nu_{nj}^\sigma(\theta)$, whereas the statistic in BCS15 depends on $\ell_{nj}(\theta) := -n^{1/2}\Delta_{F_nj}(\theta)$ and $\nu_{nj}^m(\theta)$.

26 Proofs of Lemma 25.1 and Theorem 25.2

Proof of Lemma 25.1. Because $\Delta_F^{\inf} := \inf_{\theta\in\Theta} \max_{j\leq k} \Delta_{Fj}(\theta)$, for all F and $\theta \in \Theta$,

$$\max_{j\leq k}(\Delta_{Fj}(\theta) - \Delta_F^{\inf}) \geq 0, \tag{26.1}$$

which establishes part (a).

Any $(\theta, e) \in \Lambda_\Delta$ is the limit of some sequence $(\theta_n, e_n) \in \Lambda_{n,\Delta,F_n}$ because $\Lambda_{n,\Delta,F_n} \to_H \Lambda_\Delta$ by Assumption C.12. That is, $e_n \to e$ and $\max_{j\leq k} e_{nj} \to \max_{j\leq k} e_j$. This and (26.1) applied with $(\theta, F) = (\theta_n, F_n)$ give

$$0 \leq \max_{j\leq k} n^{1/2}(\Delta_{F_nj}(\theta_n) - \Delta_{F_n}^{\inf}) = \max_{j\leq k} e_{nj} \to \max_{j\leq k} e_j, \tag{26.2}$$

which proves part (b) of the lemma.

Next, we prove part (c). The function $\Delta_{F_n}(\theta) - \Delta_{F_n}^{\inf}$ is lower semi-continuous on Θ (since $E_F \widetilde{m}_j(W,\theta)$ is upper semi-continuous on Θ by Assumption A.0(ii)), Θ is compact by Assumption A.0(i), and a lower semi-continuous function on a compact set achieves its infimum. Hence, there exists $\widetilde{\theta}_n \in \Theta$ such that $\Delta_F(\widetilde{\theta}_n) = \Delta_F^{\inf}$ $\forall n \geq 1$, which establishes part (c).

For part (d), let $(\widetilde{\theta}_n, \widetilde{e}_n) \in \Lambda_{n,\Delta,F_n}$ be such that $\Delta_F(\widetilde{\theta}_n) = \Delta_F^{\inf}$ $\forall n \geq 1$. Such $(\widetilde{\theta}_n, \widetilde{e}_n)$ exist by part (c). There exists a subsequence $\{q_n\}_{n\geq 1}$ of $\{n\}_{n\geq 1}$ and a $(\widetilde{\theta}, \widetilde{e}) \in \Theta \times R^k_{[\pm\infty]}$ such that $d((\widetilde{\theta}_{q_n}, \widetilde{e}_{q_n}), (\widetilde{\theta}, \widetilde{e})) \to 0$ because $(\Theta \times R^k_{[\pm\infty]}, d)$ is a compact metric space under Assumption A.0(i). We have $(\widetilde{\theta}, \widetilde{e}) \in \Lambda_\Delta$ by the following argument:

$$0 \leq \inf_{(\theta,e)\in\Lambda_\Delta} d((\theta,e),(\widetilde{\theta},\widetilde{e})) \leq \inf_{(\theta,e)\in\Lambda_\Delta} d((\theta,e),(\widetilde{\theta}_{q_n},\widetilde{e}_{q_n})) + d((\widetilde{\theta}_{q_n},\widetilde{e}_{q_n}),(\widetilde{\theta},\widetilde{e})) \to 0, \tag{26.3}$$

where the second inequality holds by the triangle inequality and the convergence of the first summand holds using Assumption C.12 (i.e., $\Lambda_{n,\Delta,F_n} \to_H \Lambda_\Delta$). Thus, $\inf_{(\theta,e)\in\Lambda_\Delta} d((\theta,e),(\widetilde{\theta},\widetilde{e})) = 0$. This implies that $(\widetilde{\theta},\widetilde{e}) \in \Lambda_\Delta$, because Λ_Δ is a compact subset of $(\Theta \times R^k_{[\pm\infty]}, d)$ by Assumption C.12, $d((\theta,e),(\widetilde{\theta},\widetilde{e}))$ is a continuous function of (θ,e), and a continuous function on a compact set attains its infimum.

By the definition of $\widetilde{\theta}_n$, $\Delta_{F_n}(\widetilde{\theta}_n) = \Delta_{F_n}^{\inf}$ $\forall n \geq 1$. Hence, for all $n \geq 1$,

$$\max_{j\leq k} \widetilde{e}_{nj} = \max_{j\leq k} n^{1/2}(\Delta_{F_nj}(\widetilde{\theta}_n) - \Delta_{F_n}^{\inf}) = n^{1/2}(\Delta_{F_n}(\widetilde{\theta}_n) - \Delta_{F_n}^{\inf}) = 0, \tag{26.4}$$

where the first equality holds by the definition of Λ_{n,Δ,F_n} in (25.2) and the second equality holds by the definition of $\Delta_F(\theta)$ in (25.3). We obtain

$$\max_{j\leq k} \widetilde{e}_j = \lim_{n\to\infty} \max_{j\leq k} \widetilde{e}_{nj} = 0, \tag{26.5}$$

which proves part (d) of the lemma since $(\widetilde{\theta},\widetilde{e}) \in \Lambda_\Delta$.

The proof of part (e) extends that of part (d). For $(\widetilde{\theta}_n, \widetilde{e}_n)$ defined as above, we have $\widetilde{\theta}_n \in \Theta^{\eta_n}_{\min}(F_n)$ because $\Delta_F(\widetilde{\theta}_n) = \Delta_F^{\inf}$, and so, $(\widetilde{\theta}_n, \widetilde{e}_n) \in \Lambda^{\eta_n}_{n,\Delta,F_n}$ $\forall n \geq 1$ using the definition of $\Lambda^{\eta}_{n,\Delta,F}$ following (25.4). Next, we have $(\widetilde{\theta},\widetilde{e}) \in \Lambda_{\Delta\min}$ by the same argument as used to show $(\widetilde{\theta},\widetilde{e}) \in \Lambda_\Delta$ in (26.3), but with $\Lambda_{\Delta\min}$ in place of Λ_Δ, with the convergence holding using Assumption C.13 (i.e., $\Lambda^{\eta_n}_{n,\Delta,F_n} \to_H \Lambda_{\Delta\min}$), rather than Assumption C.12, and using the fact that $\Lambda_{\Delta\min}$ is a compact subset of $(\Theta \times R^k_{[\pm\infty]}, d)$ by Assumption C.13. Finally, $\max_{j\leq k} \widetilde{e}_j = 0$ by (26.5), which establishes part (e) because $(\widetilde{\theta},\widetilde{e}) \in \Lambda_{\Delta\min}$ in the present case. □

The proof of Theorem 25.2(a) uses the following Lemma.

Lemma 26.1 *Suppose Assumptions* C.4 *and* C.5 *hold. Under* $\{F_n\}_{n\geq 1}$,

$$A_{n,\Delta} = A_{n,\Delta}(\Lambda_{n,\Delta,F_n}) + o_p(1).$$

Proof of Lemma 26.1. We have

$$\begin{aligned}
n^{1/2}\left(\widehat{\Delta}_{nj}(\theta) - \Delta_F^{\inf}\right) &= n^{1/2}\left(-\frac{\overline{m}_{nj}(\theta)}{\widehat{\sigma}_{nj}(\theta)} - \Delta_F^{\inf}\right) \\
&= \frac{\sigma_{Fj}(\theta)}{\widehat{\sigma}_{nj}(\theta)}\widehat{K}_{1nj}(\theta,F) + \frac{\sigma_{Fj}(\theta)}{\widehat{\sigma}_{nj}(\theta)}\widehat{K}_{2nj}(\theta,F) + K_{enj}(\theta,F), \text{ where} \\
\widehat{K}_{1nj}(\theta,F) &:= -n^{1/2}\left(\frac{\overline{m}_{nj}(\theta)}{\sigma_{Fj}(\theta)} - \frac{E_F m_j(W,\theta)}{\sigma_{Fj}(\theta)}\right) = -\nu_{nj}^m(\theta), \\
\widehat{K}_{2nj}(\theta,F) &:= n^{1/2}\left(\frac{\widehat{\sigma}_{nj}(\theta)}{\sigma_{Fj}(\theta)} - 1\right)\frac{E_F m_j(W,\theta)}{\sigma_{Fj}(\theta)}, \text{ and} \\
K_{enj}(\theta,F) &:= n^{1/2}\left(-\frac{E_F m_j(W,\theta)}{\sigma_{Fj}(\theta)} - \Delta_F^{\inf}\right).
\end{aligned} \tag{26.6}$$

For given $(\theta, e) \in \Lambda_{n,\Delta,F}$, $K_{enj}(\theta,F) = e_j$ for $j \leq k$.

For a given distribution F, define

$$\nu_n^{\sigma\dagger}(\theta) := n^{1/2}\left(\left(\frac{\widehat{\sigma}_{n1}^2(\theta)}{\sigma_{F1}^2(\theta)} - 1\right), ..., \left(\frac{\widehat{\sigma}_{nk}^2(\theta)}{\sigma_{Fk}^2(\theta)} - 1\right)\right)'. \tag{26.7}$$

Note that $\nu_n^{\sigma\dagger}(\theta)$ differs from $\nu_n^{\sigma}(\theta)$ (defined in (14.2) in online Appendix B) because the former depends on $\widehat{\sigma}_{nj}^2(\theta)$, which is centered at the sample quantity $\overline{m}_{nj}(\theta)$, see (4.2), whereas the latter depends on $\widehat{\sigma}_{Fnj}^2(\theta)$, which is centered at the population quantity $E_F m_j(W,\theta)$. The following calculations show that $\nu_{nj}^{\sigma\dagger}(\theta) = \nu_{nj}^{\sigma}(\theta) - n^{-1/2}(\nu_{nj}^m(\theta))^2$:

$$\begin{aligned}
\nu_{nj}^{\sigma\dagger}(\theta) &:= n^{1/2}\left(\frac{\widehat{\sigma}_{nj}^2(\theta)}{\sigma_{F_nj}^2(\theta)} - 1\right) = n^{-1/2}\sum_{i=1}^n \left[(\widetilde{m}_j(W_i,\theta) - \widetilde{m}_{nj}(\theta))^2 - 1\right] \\
&= n^{-1/2}\sum_{i=1}^n \left[(\widetilde{m}_j(W_i,\theta) - E_{F_n}\widetilde{m}_j(W,\theta))^2 - 1\right] - n^{1/2}(\widetilde{m}_{nj}(\theta) - E_{F_n}\widetilde{m}_j(W,\theta))^2 \\
&= \nu_{nj}^{\sigma}(\theta) - n^{-1/2}(\nu_{nj}^m(\theta))^2, \text{ and} \\
\nu_{nj}^{\sigma\dagger}(\theta) &= \nu_{nj}^{\sigma}(\theta) + o_p^{\Theta}(1)
\end{aligned} \tag{26.8}$$

for $j \leq k$, where the last equality holds by Assumption C.5.

By (26.8), Assumption C.5, and the continuous mapping theorem, for all $j \leq k$,

$$\sup_{\theta\in\Theta}\left|\frac{\widehat{\sigma}^2_{nj}(\theta)}{\sigma^2_{F_nj}(\theta)}-1\right| =: \sup_{\theta\in\Theta} n^{-1/2}\left|\nu^{\sigma\dagger}_{nj}(\theta)\right| = \sup_{\theta\in\Theta} n^{-1/2}\left|\nu^{\sigma}_{nj}(\theta)\right| + o^{\Theta}_p(n^{-1/2}) \to_p 0, \text{ and so,}$$
$$\sup_{\theta\in\Theta}\left|\frac{\sigma_{F_nj}(\theta)}{\widehat{\sigma}_{nj}(\theta)}-1\right| \to_p 0. \tag{26.9}$$

We have

$$\begin{aligned} n^{1/2}\left(\frac{\widehat{\sigma}_{nj}(\theta)}{\sigma_{F_nj}(\theta)}-1\right) &= n^{1/2}\left(\left(1+\left(\frac{\widehat{\sigma}^2_{nj}(\theta)}{\sigma^2_{F_nj}(\theta)}-1\right)\right)^{1/2}-1\right) \\ &= \frac{1}{2}(1+o^{\Theta}_p(1))^{-1/2} n^{1/2}\left(\frac{\widehat{\sigma}^2_{nj}(\theta)}{\sigma^2_{F_nj}(\theta)}-1\right) \\ &= \frac{1}{2}\nu^{\sigma}_{nj}(\theta) + o^{\Theta}_p(1), \end{aligned} \tag{26.10}$$

where the second equality holds by the following mean-value expansion, $(1+x)^{1/2} = 1+(1/2)(1+\widetilde{x})^{-1/2}x$, where $|\widetilde{x}| \leq |x|$, with $x := \widehat{\sigma}^2_{nj}(\theta)/\sigma^2_{F_nj}(\theta)-1$ and $\sup_{\theta\in\Theta}|x| \leq \sup_{\theta\in\Theta}|\widehat{\sigma}^2_{nj}(\theta)/\sigma^2_{F_nj}(\theta)-1| = o_p(1)$ by (26.9), and the last equality uses (26.8) and Assumption C.5.

By Assumption C.4, $E_{F_n}\widetilde{m}_j(W,\theta) = \widetilde{m}_j(\theta) + o^{\Theta}(1)$, where $o^{\Theta}(1)$ denotes a term that is $o(1)$ uniformly over $\theta \in \Theta$. Combining this, (26.9), and (26.10) with the definition of $\widehat{K}_{2nj}(\theta, F_n)$ in (26.6) gives

$$\frac{\sigma_{F_nj}(\theta)}{\widehat{\sigma}_{nj}(\theta)}\widehat{K}_{2nj}(\theta,F_n) = \frac{1}{2}\widetilde{m}_j(\theta)\cdot\nu^{\sigma}_{nj}(\theta) + o^{\Theta}_p(1) \text{ for } j \leq k. \tag{26.11}$$

In addition, (26.9), Assumption C.5, and the definition of $\widehat{K}_{1nj}(\theta, F_n)$ in (26.6) give

$$\frac{\sigma_{F_nj}(\theta)}{\widehat{\sigma}_{nj}(\theta)}\widehat{K}_{1nj}(\theta,F_n) = -\nu^{m}_{nj}(\theta) + o^{\Theta}_p(1) \text{ for } j \leq k. \tag{26.12}$$

Thus, we have

$$\begin{aligned} A_{n,\Delta} &= \inf_{\theta\in\Theta}\max_{j\leq k}\left(\frac{\sigma_{F_nj}(\theta)}{\widehat{\sigma}_{nj}(\theta)}\widehat{K}_{1nj}(\theta,F_n) + \frac{\sigma_{F_nj}(\theta)}{\widehat{\sigma}_{nj}(\theta)}\widehat{K}_{2nj}(\theta,F_n) + K_{enj}(\theta,F_n)\right) \\ &= \inf_{(\theta,e)\in\Lambda_{n,\Delta,F_n}}\max_{j\leq k}\left(-\nu^{m}_{nj}(\theta) + \frac{1}{2}\widetilde{m}_j(\theta)\cdot\nu^{\sigma}_{nj}(\theta) + e_j\right) + o_p(1) \\ &= A_{n,\Delta}(\Lambda_{n,\Delta,F_n}) + o_p(1), \end{aligned} \tag{26.13}$$

where the first equality holds by (26.6), the second equality holds by (26.11), (26.12), and the definition of Λ_{n,Δ,F_n} in (25.2), and the last equality holds by the definition of $A_{n,\Delta}(\Lambda_{n,\Delta,F_n})$ in (25.5). □

Proof of Theorem 25.2. First, we prove part (a). By Lemma 26.1, it suffices to show

$$A_{n,\Delta}(\Lambda_{n,\Delta,F_n}) \to_d A_{\infty,\Delta}. \tag{26.14}$$

Let $\mathcal{D}$ be the space of functions from Θ to R^{2k}. Let $\mathcal{D}_0$ be the subset of uniformly continuous functions in $\mathcal{D}$. For $\nu(\cdot) \in \mathcal{D}$, define

$$\begin{aligned} g_n(\nu(\cdot)) &:= \inf_{(\theta,e)\in\Lambda_{n,\Delta,F_n}} \max_{j\leq k} \left[\tau_j(\nu(\cdot),\theta) + e_j\right], \\ g(\nu(\cdot)) &:= \inf_{(\theta,e)\in\Lambda_{\Delta}} \max_{j\leq k} \left[\tau_j(\nu(\cdot),\theta) + e_j\right], \text{ where} \\ \tau_j(\nu(\cdot),\theta) &:= -\nu_j^m(\theta) + \frac{1}{2}\widetilde{m}_j(\theta)\nu_j^\sigma(\theta), \end{aligned} \tag{26.15}$$

$\nu(\theta) = (\nu^m(\theta), \nu^\sigma(\theta))'$, and $\nu_j^m(\theta)$ and $\nu_j^\sigma(\theta)$ denote the jth elements of $\nu^m(\theta)$ and $\nu^\sigma(\theta)$, respectively. Note that

$$A_{n,\Delta}(\Lambda_{n,\Delta,F_n}) = g_n(\nu_n(\cdot)) \text{ and } A_{\infty,\Delta} := A_{\infty,\Delta}(\Lambda_\Delta) = g(G(\cdot)). \tag{26.16}$$

We want to show $g_n(\nu_n(\cdot)) \to_d g(G(\cdot))$. By Assumption C.5, $\nu_n(\cdot) \Rightarrow G(\cdot)$ for $\nu_n(\cdot) \in \mathcal{D}$ a.s. and $G(\cdot) \in \mathcal{D}_0$ a.s. We use the extended CMT, see van der Vaart and Wellner (1996, Theorem 1.11.1), to establish the desired result, as in the proof of Theorem 3.1 in BCS15. The extended CMT requires showing: for any deterministic sequence $\{\nu_n(\cdot) \in \mathcal{D}\}_{n\geq 1}$ and deterministic $\nu(\cdot) \in \mathcal{D}_0$ such that $\sup_{\theta\in\Theta} ||\nu_n(\theta) - \nu(\theta)|| \to 0$, we have $g_n(\nu_n(\cdot)) \to g(\nu(\cdot))$. (For notational simplicity, we abuse notation here and consider a deterministic $\nu_n(\cdot)$ that differs from the random $\nu_n(\cdot)$ in Assumption C.5.) Once we have shown this, the proof of part (a) is complete.

Let $\{\nu_n(\cdot) \in \mathcal{D}\}_{n\geq 1}$ and $\nu(\cdot) \in \mathcal{D}_0$ be deterministic and satisfy $\sup_{\theta\in\Theta} ||\nu_n(\theta) - \nu(\theta)|| \to 0$. We show

$$\text{(i) } \liminf_{n\to\infty} g_n(\nu_n(\cdot)) \geq g(\nu(\cdot)) \text{ and (ii) } \limsup_{n\to\infty} g_n(\nu_n(\cdot)) \leq g(\nu(\cdot)). \tag{26.17}$$

First, we establish (i) in (26.17). There exists a subsequence $\{a_n\}_{n\geq 1}$ of $\{n\}_{n\geq 1}$ and there exists a sequence $\{(\overline{\theta}_{a_n}, \overline{e}_{a_n}) \in \Lambda_{a_n,\Delta,F_{a_n}}\}_{n\geq 1}$ such that

$$\begin{aligned} \limsup_{n\to\infty} g_n(\nu_n(\cdot)) &= \lim_{n\to\infty} g_{a_n}(\nu_{a_n}(\cdot)) \text{ and} \\ \lim_{n\to\infty} g_{a_n}(\nu_{a_n}(\cdot)) &= \lim_{n\to\infty} \max_{j\leq k} \left[\tau_j(\nu_{a_n}(\cdot), \overline{\theta}_{a_n}) + \overline{e}_{a_nj}\right], \end{aligned} \tag{26.18}$$

where $\overline{e}_{a_nj}$ denotes the jth element of $\overline{e}_{a_n}$. Also, there exists a subsequence $\{q_n\}_{n\geq 1}$ of $\{a_n\}_{n\geq 1}$ and

$(\overline{\theta}, \overline{e}) \in \Theta \times R^k_{[\pm\infty]}$ such that

$$d\left((\overline{\theta}_{q_n}, \overline{e}_{q_n}), (\overline{\theta}, \overline{e})\right) \to 0, \tag{26.19}$$

where d is defined in the paragraph before (25.1), by compactness of the metric space $(\Theta \times R^k_{[\pm\infty]}, d)$ under Assumption A.0(i). We have $(\overline{\theta}, \overline{e}) \in \Lambda_\Delta$ by the same argument as used to show $(\widetilde{\theta}, \widetilde{e}) \in \Lambda_\Delta$ in (26.3) (but without the requirement that $\Delta_F(\widetilde{\theta}_n) = \Delta_F^{\inf}\ \forall n \geq 1$) using (26.19) and Assumption C.12.

For all $j \leq k$,

$$\lim_{n\to\infty} \tau_j(\nu_{q_n}(\cdot), \overline{\theta}_{q_n}) = -\nu_j^m(\overline{\theta}) + \frac{1}{2}\widetilde{m}_j(\overline{\theta})\nu_j^\sigma(\overline{\theta}) := \tau_j(\nu(\cdot), \overline{\theta}) \in R, \tag{26.20}$$

the first equality holds by $\nu_{q_n}(\theta) \to \nu(\theta) = (\nu^m(\theta)', \nu^\sigma(\theta)')'$ uniformly over $\theta \in \Theta$ (by assumption) and (26.19), the last equality holds by the definition of $\tau_j(\nu(\cdot), \theta)$ in (26.15), and "$\in R$" holds because $\nu_j^m(\overline{\theta})$ and $\nu_j^\sigma(\overline{\theta})$ are finite since $\nu(\cdot)$ is assumed to be in $\mathcal{D}$ and $\widetilde{m}_j(\overline{\theta})$ is finite by Assumption C.4.

Now, we have

$$\begin{aligned}
\limsup_{n\to\infty} g_n(\nu_n(\cdot)) &= \lim_{n\to\infty} \max_{j\leq k}\left[\tau_j(\nu_{q_n}(\cdot), \overline{\theta}_{q_n}) + \overline{e}_{q_n j}\right] \\
&= \max_{j\leq k}\left[\tau_j(\nu(\cdot), \overline{\theta}) + \overline{e}_j\right] \\
&\geq \inf_{(\theta,e)\in\Lambda_\Delta} \max_{j\leq k}\left[\tau_j(\nu(\cdot), \theta) + e_j\right] \\
&:= g(\nu(\cdot)),
\end{aligned} \tag{26.21}$$

where the first equality holds by (26.18) and the fact that $\{q_n\}_{n\geq 1}$ is a subsequence of $\{a_n\}_{n\geq 1}$, the second equality holds by (26.20) (using the notational convention that $\nu + c = c$ when $\nu \in R$ and $c = \pm\infty$ if $\overline{e}_j = \pm\infty$ for any $j \leq k$), the inequality holds because $(\overline{\theta}, \overline{e}) \in \Lambda_\Delta$ by the paragraph containing (26.19), and the last equality holds by the definition of $g(\nu(\cdot))$ in (26.15). This establishes result (i) in (26.17).

Next, we establish result (ii) in (26.17). There exists $(\theta^\dagger, e^\dagger) \in \Lambda_\Delta$ such that

$$g(\nu(\cdot)) = \max_{j\leq k}\left[\tau_j(\nu(\cdot), \theta^\dagger) + e_j^\dagger\right] \tag{26.22}$$

because Λ_Δ is compact under the metric d, defined in the paragraph before (25.1) with $a_* = d_\theta + k$ (since it is assumed to be an element of $\mathcal{S}(\Theta \times R^k_{[\pm\infty]})$) and $\tau_j(\nu(\cdot), \theta) + e_j$ is a continuous function of (θ, e) under d that takes values in the extended real line. By Assumption C.12, $\Lambda_{n,\Delta,F_n} \to_H \Lambda_\Delta$.

Hence, there is a sequence $\{(\theta_n^\dagger, e_n^\dagger) \in \Lambda_{n,\Delta,F_n}\}_{n\geq 1}$ such that $d((\theta_n^\dagger, e_n^\dagger), (\theta^\dagger, e^\dagger)) \to 0$. We obtain

$$\begin{aligned}
\liminf_{n\to\infty} g_n(\nu_n(\cdot)) :=& \liminf_{n\to\infty} \inf_{(\theta,e)\in\Lambda_{n,\Delta,F_n}} \max_{j\leq k} [\tau_j(\nu_n(\cdot),\theta) + e_j] \\
\leq& \liminf_{n\to\infty} \max_{j\leq k} \left[\tau_j(\nu_n(\cdot),\theta_n^\dagger) + e_{nj}^\dagger\right] \\
=& \max_{j\leq k} \left[\tau_j(\nu(\cdot),\theta^\dagger) + e_j^\dagger\right] \\
=& \, g(\nu(\cdot)),
\end{aligned} \tag{26.23}$$

where the inequality holds because $(\theta_n^\dagger, e_n^\dagger) \in \Lambda_{n,\Delta,F_n}$ $\forall n \geq 1$, the second equality holds using $d((\theta_n^\dagger, e_n^\dagger), (\theta^\dagger, e^\dagger)) \to 0$ and (26.20) with $(\nu_n(\cdot), \theta_n^\dagger)$ and $(\nu(\cdot), \theta^\dagger)$ in place of $(\nu_{q_n}(\cdot), \overline{\theta}_{q_n},)$ and $(\nu(\cdot), \overline{\theta})$, respectively, and the last equality holds by (26.22). This establishes result (ii) in (26.17) and completes the proof of part (a).

Now we prove part (b). We have

$$A_{\infty,\Delta} := \inf_{(\theta,e)\in\Lambda_\Delta} \max_{j\leq k} \left(-G_j^m(\theta) + \frac{1}{2}\widetilde{m}_j(\theta) G_j^\sigma(\theta) + e_j\right) > -\infty \text{ a.s.} \tag{26.24}$$

because (I) $\max_{j\leq k} e_j \geq 0$ $\forall(\theta, e) \in \Lambda_\Delta$ by Lemma 25.1(b), (II) $\sup_{\theta\in\Theta} |G_j^m(\theta)| < \infty$ a.s. by Assumption C.5, and (III) $\sup_{\theta\in\Theta} |\widetilde{m}_j(\theta) G_j^\sigma(\theta)| < \infty$ a.s. because $\widetilde{m}_j(\cdot)$ is bounded on Θ by Assumption C.4 and $|G_j^\sigma(\cdot)|$ is bounded on Θ a.s. by Assumption C.5.

To obtain the other half of part (b), i.e., $A_{\infty,\Delta} < \infty$ a.s., we use Lemma 25.1(d). We have

$$\begin{aligned}
A_{\infty,\Delta} :=& \inf_{(\theta,e)\in\Lambda_\Delta} \max_{j\leq k} \left(-G_j^m(\theta) + \frac{1}{2}\widetilde{m}_j(\theta) G_j^\sigma(\theta) + e_j\right) \\
\leq& \max_{j\leq k} \left(-G_j^m(\widetilde{\theta}) + \frac{1}{2}\widetilde{m}_j(\widetilde{\theta}) G_j^\sigma(\widetilde{\theta}) + \widetilde{e}_j\right) < \infty \text{ a.s.},
\end{aligned} \tag{26.25}$$

where $(\widetilde{\theta}, \widetilde{e}) \in \Lambda_\Delta$ is as in Lemma 25.1(d), the first equality holds by the definition of $A_{\infty,\Delta}$ in (25.5), the first inequality holds because $(\widetilde{\theta}, \widetilde{e}) \in \Lambda_\Delta$ by Lemma 25.1(d), and last inequality holds because (I) $\max_{j\leq k} \widetilde{e}_j = 0$ by Lemma 25.1(d), (II) $\sup_{\theta\in\Theta} |G_j^m(\theta)| < \infty$ a.s. by (II) following (26.24), and (III) $\sup_{\theta\in\Theta} |\widetilde{m}_j(\theta) G_j^\sigma(\theta)| < \infty$ a.s. by (III) following (26.24). This completes the proof of part (b).

Now, we establish part (c). If $\Lambda_\Delta = \Lambda_{\Delta\,\min}$, then part (c) holds immediately. So, we suppose that $\Lambda_\Delta \backslash \Lambda_{\Delta\,\min}$ is not empty. We show that for any $(\theta^*, e^*) \in \Lambda_\Delta \backslash \Lambda_{\Delta\,\min}$,

$$\max_{j\leq k} \left[\tau_j(G(\cdot), \theta^*) + e_j^*\right] = \infty \text{ a.s.}, \tag{26.26}$$

where $\tau_j(\nu(\cdot), \theta)$ is defined in (26.15). Since $A_{\infty,\Delta} \in R$ a.s. by part (b), and $A_{\infty,\Delta} := \inf_{(\theta,e)\in\Lambda_\Delta}$

$\max_{j \le k} [\tau_j(G(\cdot), \theta) + e_j]$ by (25.5), (26.26) implies that $A_{\infty,\Delta} = A_{\infty,\Delta\min}$ a.s., which establishes part (c).

For part (c), it remains to show (26.26). By Assumption C.13, $\Lambda_{\Delta\min}$ is compact. For any $(\theta^*, e^*) \in \Lambda_\Delta \backslash \Lambda_{\Delta\min}$, there is a neighborhood of (θ^*, e^*) that lies in $\Lambda_\Delta \backslash \Lambda_{\Delta\min}$ and there exists a sequence $\{(\theta_n^*, e_n^*) \in \Lambda_{n,\Delta,F_n}\}_{n\ge1}$ such that $d((\theta_n^*, e_n^*), (\theta^*, e^*)) \to 0$ by Assumption C.12. In consequence, for n large, $(\theta_n^*, e_n^*) \notin \Lambda_{n,\Delta,F_n}^{\eta_n}$. In turn, this implies that $\theta_n^* \notin \Theta_{\min}^{\eta_n}(F_n)$ for n large using the definition of $\Lambda_{n,\Delta,F_n}^{\eta_n}$ following (25.4).

Now, $\theta_n^* \notin \Theta_{\min}^{\eta_n}(F_n)$ for all n large implies

$$\begin{aligned}
&\Delta_{F_n}(\theta_n^*) > \Delta_{F_n}^{\inf} + \eta_n/n^{1/2} \text{ for all } n \text{ large,} \\
&n^{1/2}(\Delta_{F_n}(\theta_n^*) - \Delta_{F_n}^{\inf}) > \eta_n \to \infty, \text{ and} \\
&\max_{j\le k} e_j^* = \lim \max_{j\le k} e_{n,j}^* := \lim \max_{j\le k} n^{1/2}(\Delta_{F_nj}(\theta_n^*) - \Delta_{F_n}^{\inf}) = \infty,
\end{aligned} \tag{26.27}$$

where (i) the first line holds by the definition of $\Theta_{\min}^{\eta}(F)$ following (25.4), (ii) the inequality on the second line follows from the first line and $\eta_n \to \infty$ by Assumption C.13, and (iii) the first equality in the third line holds by the convergence result for $\{(\theta_n^*, e_n^*)\}_{n\ge1}$ in the previous paragraph, the second equality in the third line holds by $(\theta_n^*, e_n^*) \in \Lambda_{n,\Delta,F_n}$ and the definition of $\Lambda_{n,\Delta,F}$ in (25.2), and the third equality in the third line follows from the second line because $\Delta_{F_n}(\theta_n^*) = \max_{j\le k} \Delta_{F_nj}(\theta_n^*)$.

The result $\max_{j\le k} e_j^* = \infty$ in (26.27) implies that (26.26) holds because $|\tau_j(G(\cdot), \theta^*)| < \infty$ a.s. (using Assumptions C.4 and C.5, the definition of $\tau_j(\nu(\cdot), \theta)$ in (26.15), and explanations (II) and (III) following (26.24)). This completes the proof of part (c). $\square$

27 Proof of Theorem 5.1

27.1 Notation and Assumptions

As noted in Theorem 5.1, as is standard in the literature, the asymptotics for the bootstrap are given for the case where the number of bootstrap repetitions $B = \infty$. (If one considered finite B, then all of the asymptotic results would hold provided $B \to \infty$ as $n \to \infty$.) With $B = \infty$, the bootstrap critical values $\widehat{c}_{n,\Delta U}(1-\alpha)$ and $\widehat{c}_{n,\Delta L}(1-\alpha)$, defined following (5.14) and (5.15), respectively, are the $1-\alpha$ conditional quantiles of $-A_{n,\Delta U,b}^*$ and $A_{n,\Delta L,b}^*$ given the sample $\{W_i\}_{i\le n}$ plus ι, rather than the $1-\alpha$ sample quantiles of $\{-A_{n,\Delta U,b}^*\}_{b\le B}$ and $\{A_{n,\Delta L,b}^*\}_{b\le B}$, respectively, plus ι. For notational simplicity, we replace the bth bootstrap sample $\{W_{ib}^*\}_{i\le n}$ by a generic bootstrap sample $\{W_i^*\}_{i\le n}$ (which is an i.i.d. bootstrap sample drawn with replacement from the original sample $\{W_i\}_{i\le n}$) and we drop the subscripts b from the definitions of $A_{n,\Delta U,b}^*$ in (5.14), $A_{n,\Delta L,b}^*$ in

(5.15), and other bootstrap quantities. Specifically, we define $\widehat{\nu}^*_{nj}(\theta)$, $\overline{m}^*_{nj}(\theta)$, $\widehat{\sigma}^{*2}_{nj}(\theta)$, and $A^*_{n,\Delta U}$ as $\widehat{\nu}^*_{njb}(\theta)$, $\overline{m}^*_{njb}(\theta)$, $\widehat{\sigma}^{*2}_{njb}(\theta)$, and $A^*_{n,\Delta U,b}$ are defined in (4.17) and (5.14), but with the generic bootstrap sample $\{W^*_i\}_{i\leq n}$ in place of the bth bootstrap sample $\{W^*_{ib}\}_{i\leq n}$ and with b deleted throughout. Similarly, we define $A^*_{n,\Delta L}$ as $A^*_{n,\Delta L,b}$ is defined in (5.15), but with $\widehat{\nu}^*_{nj}(\theta)$ in place of $\widehat{\nu}^*_{njb}(\theta)$.

The $B=\infty$ definitions of $\widehat{sd}_{nj}(\theta)$ and $\widehat{sd}_n(\theta)$ are as follows. For $Z\sim N(0_{2k},I_{2k})$ and $j\leq k$, define

$$\widehat{sd}_{nj}(\theta):=\max\left\{V^{1/2}_{nj}(\theta),\iota\right\} \text{ and } \widehat{sd}_n(\theta):=\max_{j\leq k}\widehat{sd}_{nj}(\theta), \text{ where } V_{nj}(\theta):=Var^{1/2}_Z(Q_{nj}(\theta)),$$
$$Q_{nj}(\theta):=\widehat{G}^{m\sigma}_{nj}(\theta)-\max_{j_1\leq k}\widehat{G}^{m\sigma}_{nj_1}(\theta),\ \widehat{G}^{m\sigma}_{nj}(\theta):=(c_j',-(1/2)\widehat{m}_{nj}(\theta)c_j')\widehat{\Omega}^{1/2}_{n+}(\theta)Z, \tag{27.1}$$

$Var_Z(Q_{nj}(\theta))$ denotes the variance of $Q_{nj}(\theta)$ with respect to the randomness in Z conditional on $\widehat{\Omega}_{n+}(\theta)$ and $\widehat{m}_{nj}(\theta)$, and (as above) c_j denote the jth elementary k-vector. In addition, $\widehat{e}_{nj}(\theta)$ and $\widehat{J}_{ne}(\theta)$ are defined as $\widehat{e}_{nj}(\theta)$ and $\widehat{J}_{neB}(\theta)$ are defined in (5.8) and (5.13), respectively, but with $\widehat{sd}_{nj}(\theta)$ in place of $\widehat{sd}_{njB}(\theta)$.

The bootstrap sample $\{W^*_i\}_{i\leq n}$ depends on $\{W_i\}_{i\leq n}$ and on some other independent random variables $\{\zeta_i\}_{i\leq n}$ that are used to construct the bootstrap sample $\{W^*_i\}_{i\leq n}$. To establish the asymptotic properties of the bootstrap critical values for a given sequence of distributions $\{F_n\}_{n\geq 1}$, it is convenient to have a single probability space $(\Omega,\mathcal{F},P_\bigtriangledown)$ on which all of the random vectors $\{W_i\}_{i\leq n}$ for $n\geq 1$ and the bootstrap random variables (or vectors) $\{\zeta_i\}_{i\leq n}$ for all $n\geq 1$ are defined. Since F_n changes with n, this requires that we consider triangular arrays of random vectors, not sequences. Let $\{W_{ni}\}_{i\leq n,n\geq 1}:=\{W_{ni}:i\leq n,n\geq 1\}$ be a triangular array of random vectors on $(\Omega,\mathcal{F},P_\bigtriangledown)$ such that, for each $n\geq 1$, $\{W_{ni}\}_{i\leq n}$ has the same distribution as $\{W_i\}_{i\leq n}\sim F_n$. Analogously, let $\{\zeta_{ni}\}_{i\leq n,n\geq 1}$ be a triangular array of bootstrap random variables (or vectors) on $(\Omega,\mathcal{F},P_\bigtriangledown)$ such that for each $n\geq 1$, $\{\zeta_{ni}\}_{i\leq n}$ has the same distribution as $\{\zeta_i\}_{i\leq n}$ and $\{\zeta_{ni}\}_{i\leq n,n\geq 1}$ is independent of $\{W_{ni}\}_{i\leq n,n\geq 1}$.

For notational simplicity, but with some abuse of notation, we let all of the statistics defined above, including $\widehat{\Delta}^{\inf}_n$, $A_{n,\Delta}$, $A^*_{n,\Delta U}$, $A^*_{n,\Delta L}$, $\widehat{c}_{n,\Delta U}(1-\alpha)$, and $\widehat{c}_{n,\Delta L}(1-\alpha)$, which are defined as functions of $\{W_i\}_{i\leq n}\sim F_n$ and $\{\zeta_i\}_{i\leq n}$, also denote the corresponding statistics defined when using the triangular arrays $\{W_{ni}\}_{i\leq n,n\geq 1}$ and $\{\zeta_{ni}\}_{i\leq n,n\geq 1}$. For events that only depend on n random vectors for a single n, such as $A^*_{n,\Delta U}\in B_n$ for some fixed set $B_n\subset R$, we have $P_\bigtriangledown(A^*_{n,\Delta U}\in B_n)=P_{F_n}(A^*_{n,\Delta U}\in B_n)$. But, for events that depend on statistics for multiple values of n, such as $\{A^*_{n,\Delta U}\}_{n\geq 1}$, we use the probability space $(\Omega,\mathcal{F},P_\bigtriangledown)$. In particular, when we condition on the entire triangular array $\{W_{ni}\}_{i\leq n,n\geq 1}$, we need $(\Omega,\mathcal{F},P_\bigtriangledown)$.

Let $\{\nu_n^*(\theta) \in R^{2k} : \theta \in \Theta\}$ be a bootstrap version of the stochastic process $(\nu_n^m(\cdot)', \nu_n^{\sigma\dagger}(\theta)')'$ defined in (14.2) in online Appendix B and (26.7). It is defined as follows:

$$\begin{aligned}
\nu_n^*(\theta) &:= (\nu_n^{m*}(\theta)', \nu_n^{\sigma*}(\theta)')', \text{ where} \\
\nu_{nj}^{m*}(\theta) &:= n^{1/2}\left(\widetilde{m}_{nj}^*(\theta) - \widehat{m}_{nj}(\theta)\right), \ \widetilde{m}_{nj}^*(\theta) := \frac{\overline{m}_{nj}^*(\theta)}{\widehat{\sigma}_{nj}(\theta)}, \ \overline{m}_{nj}^*(\theta) := n^{-1}\sum_{i=1}^n m_j(W_i^*, \theta), \\
\nu_{nj}^{\sigma*}(\theta) &:= n^{1/2}\left(\frac{\widehat{\sigma}_{nj}^{*2}(\theta)}{\widehat{\sigma}_{nj}^2(\theta)} - 1\right), \ \widehat{\sigma}_{nj}^{*2}(\theta) := n^{-1}\sum_{i=1}^n (m_j(W_i^*, \theta) - \overline{m}_{nj}^*(\theta))^2 \ \forall j \leq k, \\
\nu_n^{m*}(\theta) &= (\nu_{n1}^{m*}(\theta), ..., \nu_{nk}^{m*}(\theta))', \text{ and } \nu_n^{\sigma*}(\theta) = (\nu_{n1}^{\sigma*}(\theta), ..., \nu_{nk}^{\sigma*}(\theta))'. \qquad (27.2)
\end{aligned}$$

Let $\{\nu_n^*(\cdot)|\{W_{ni}\}_{i\leq n, n\geq 1}\} \Rightarrow G(\cdot)$ denote that the conditional distribution of $\nu_n^*(\cdot)$ given $\{W_{ni}\}_{i\leq n,n\geq 1}$ converges weakly to $G(\cdot)$.

Let $X \geq_{ST} Y$ denote that X is stochastically greater than or equal to Y. That is, $P(Y > x) \leq P(X > x)$ for all $x \in R$.

For $\theta \in \Theta$, define

$$j_{ne}(\theta) := \arg\max_{j\leq k} e_{nj}(\theta), \text{ where } e_{nj}(\theta) := n^{1/2}(\Delta_{F_n,j}(\theta) - \Delta_{F_n}^{\inf}).^{51} \qquad (27.3)$$

By Lemma 25.1(a),

$$e_{nj_{ne}(\theta)}(\theta) \geq 0 \ \forall \theta \in \Theta. \qquad (27.4)$$

Define

$$\begin{aligned}
\Lambda_{n,\Delta,F_n}^{*\eta_n} := \Big\{&(\theta, e, e^*, j^*) \in \Theta_{\min}^{\eta_n}(F_n) \times R^{2k} \times \{1, ..., k\} : e_j = n^{1/2}(\Delta_{F_n j}(\theta) - \Delta_{F_n}^{\inf}), \\
&e_j^* = (\iota\kappa_n)^{-1} e_j \ \forall j \leq k, \ j^* := j_{ne}(\theta)\Big\}, \qquad (27.5)
\end{aligned}$$

where $\{\eta_n\}_{n\geq 1}$ is as in Assumption C.13 and $\{\kappa_n\}_{n\geq 1}$ is as in (5.8), (5.12), and (5.13). Let $\mathcal{S}(\Theta \times R_{[\pm\infty]}^{2k} \times \{1, ..., k\})$ denote the space of compact subsets of the metric space $(\Theta \times R_{[\pm\infty]}^{2k} \times \{1, ..., k\}, d)$, where d is defined in the paragraph before (25.1) with $a_* = d_\theta + 2k + 1$.

We employ the following bootstrap convergence assumptions, which apply to a drifting sequence of distributions $\{F_n\}_{n\geq 1}$. Subsequence versions of them are verified below in the proof of Theorem 5.1 in Section 27.4. The expanded minimizer set $\Theta_{\min}^{\eta_n}(F_n)$ is defined following (25.4), the bootstrap stochastic process $\nu_n^*(\cdot)$ (with subscript b deleted) is defined in (4.17), and the estimator $\widehat{\Theta}_{\min,n}$ of $\Theta_{\min}^{\eta_n}(F_n)$ is defined in (5.7).

Assumption BC.3. $\{\nu_n^*(\cdot)|\{W_{ni}\}_{i\leq n,n\geq 1}\} \Rightarrow G(\cdot)$ a.s.$[P_\bigtriangledown]$, where $G(\cdot)$ is as in Assumption C.5.

[51] If the arg max is not unique, $j_{ne}(\theta)$ is defined to be the smallest arg max.

Assumption BC.4. $\Lambda^{*\eta_n}_{n,\Delta,F_n} \to_H \Lambda^*_{\Delta\min}$ for some non-empty set $\Lambda^*_{\Delta\min} \in \mathcal{S}(\Theta \times R^{2k}_{[\pm\infty]} \times \{1,...,k\})$ for some sequence of constants $\{\eta_n\}_{n\geq 1}$ that satisfies $\eta_n \to \infty$ and $\eta_n/\tau_n \to 0$ for the constants $\{\tau_n\}_{n\geq 1}$ that appear in (5.7) and satisfy Assumption A.6(ii).

Assumption BC.5. $\widehat{\Theta}_{\min,n} \supseteq \Theta^{\eta_n}_{\min}(F_n)$ wp→1 for constants $\{\eta_n\}_{n\geq 1}$ as in Assumptions BC.4 and C.13.

Define

$$\xi^e_{1nj}(\theta) := (\iota\kappa_n)^{-1} n^{1/2}\left(\widehat{\Delta}_{nj}(\theta) - \widehat{\Delta}^{\inf}_n\right) \ \forall j \leq k, \tag{27.6}$$

where κ_n is as in the definition of $\xi^e_{nj}(\theta)$ in (5.12) and ι is as in the definition of $\widehat{sd}_{nj}(\theta)$ in (27.1). Note that $\xi^e_{1nj}(\theta)$ differs from $\xi^e_{nj}(\theta)$ because it has ι in place of $\widehat{sd}_{nj}(\theta)$, where $\iota \leq \widehat{sd}_{nj}(\theta)$ by the definition of $\widehat{sd}_{nj}(\theta)$.

The GMS function $\varphi : R_{[+\infty]} \to R_{[+\infty]}$ defined in (4.19) is upper bounded by the function $\varphi^\dagger : R_{[+\infty]} \to R_{[+\infty]}$ defined by

$$\varphi^\dagger(\xi) := \infty 1(\xi \geq 1) + (\xi/(1-\xi))1(0 \leq \xi < 1). \tag{27.7}$$

The function $\varphi^\dagger$ satisfies: (i) $\varphi^\dagger(\xi) \geq \varphi(\xi)$ $\forall \xi \in R_{[+\infty]}$, (ii) $\varphi^\dagger$ is nondecreasing and continuous under the metric d, and (iii) $\varphi^\dagger(\xi) = 0$ $\forall \xi \leq 0$ and $\varphi^\dagger(\infty) = \infty$, where the metric d is defined in Section 15.1 with $a_* = 1$.

Define an upper-bound (wp→1) random variable, $A^*_{Un,\Delta U}$, on the EGMS bootstrap statistic $A^*_{n,\Delta U}$ to be

$$A^*_{Un,\Delta U} := \inf_{\theta \in \Theta^{\eta_n}_{\min}(F_n)} \max_{j\leq k}\left(-\widehat{\nu}^*_{nj}(\theta) + 1(j \neq j_{ne}(\theta))e_{nj}(\theta) + 1(j = j_{ne}(\theta))\varphi^\dagger(\xi^e_{1nj}(\theta))\right). \tag{27.8}$$

Let $\widehat{c}_{Un,\Delta U}(1-\alpha)$ denote the $1-\alpha$ conditional quantile of $-A^*_{Un,\Delta U}$ given $\{W_{ni}\}_{i\leq n,n\geq 1}$ plus ι. The statistic $\widehat{c}_{Un,\Delta U}(1-\alpha)$ is random and depends on the conditioning value of $\{W_{ni}\}_{i\leq n,n\geq 1}$.

By Lemma 27.1(a) below, the asymptotic distribution of the upper-bound bootstrap random variable $A^*_{Un,\Delta U}$ conditional on $\{W_{ni}\}_{i\leq n,n\geq 1}$ is the following distribution a.s.$[P_\bigtriangledown]$:

$$\begin{aligned} A^*_{U\infty,\Delta U} &:= \inf_{(\theta,e,e^*,j^*)\in\Lambda^*_{\Delta\min}} \max_{j\leq k}\left(-G^{m\sigma}_j(\theta) + 1(j\neq j^*)e_j + 1(j=j^*)\varphi^\dagger(e^*_{j^*})\right), \text{ where} \\ G^{m\sigma}_j(\theta) &:= G^m_j(\theta) - \frac{1}{2}\widetilde{m}_j(\theta)G^\sigma_j(\theta), \end{aligned} \tag{27.9}$$

for $\Lambda^*_{\Delta\min}$ as in Assumption BC.4. Let $c_{U\infty,\Delta U}(1-\alpha)$ denote the $1-\alpha$ conditional (or unconditional) quantile of $A^*_{U\infty,\Delta U}$ without ι added on. It is nonrandom and does not depend on $\{W_{ni}\}_{i\leq n,n\geq 1}$ by (27.9).

Next, we consider definitions and assumptions concerning the lower bound CI. The population counterpart of $\widehat{sd}_n(\theta)$ is $sd_F(\theta)$:

$$\begin{aligned}
sd_F(\theta) &:= \max_{j\leq k} sd_{Fj}(\theta), \text{ where} \\
sd_{Fj}(\theta) &:= \max\left\{Var_Z^{1/2}(D_{Fj}(\theta)), \iota\right\}, \\
D_{Fj}(\theta) &:= G_{Fj}^{m\sigma}(\theta) - \max_{j_1\leq k} G_{Fj_1}^{m\sigma}(\theta), \\
G_{Fj}^{m\sigma}(\theta) &:= (c_j', -(1/2)\widetilde{m}_{Fj}(\theta)c_j')\Omega_F^{1/2}(\theta,\theta)Z, \text{ and } Z\sim N(0_{2k}, I_{2k}).
\end{aligned} \tag{27.10}$$

Define

$$\Lambda_{n,F_n,L}^{*\eta_{Ln}} := \left\{(\theta, e^*)\in \Theta_{\min}^{\eta_{Ln}}(F_n)\times R^k : e_j^* = (\iota\kappa_n)^{-1}n^{1/2}(\Delta_{F_nj}(\theta) - \Delta_{F_n}^{\inf})\right\}. \tag{27.11}$$

Let $\mathcal{S}(\Theta\times R^k_{[\pm\infty]})$ denote the space of compact subsets of the metric space $(\Theta\times R^k_{[\pm\infty]}, d)$, where d is defined in the paragraph before (25.1) with $a_* = d_\theta + k$.

We employ the following bootstrap convergence assumptions for the lower-bound CI's. Subsequence versions of them are verified below.

Assumption BC.6. $\Lambda_{n,F_n,L}^{*\eta_{Ln}} \to_H \Lambda_L^*$ for some non-empty set $\Lambda_L^*\in\mathcal{S}(\Theta\times R^k_{[\pm\infty]})$ for some sequence of constants $\{\eta_{Ln}\}_{n\geq 1}$ that satisfies $\eta_{Ln}\to\infty$, $\eta_{Ln}/n^{1/2}\to 0$, and $\eta_{Ln}/\kappa_n^\gamma\to 0$ for some $\gamma\in(0,1)$ for the constants $\{\kappa_n\}_{n\geq 1}$ that are employed in the definition of $A^*_{n,\Delta L}$ and satisfy Assumption A.7.

Assumption BC.7. $\widehat{\Theta}_{\min,L,n}\subseteq\Theta_{\min}^{\eta_{Ln}}(F_n)$ wp$\to$1 for constants $\{\eta_{Ln}\}_{n\geq 1}$ as in Assumption BC.6.

For example, in Assumption BC.6, one can take $\eta_{Ln} = \kappa_n^{1/2}$.

Let $\to_u$ denote uniform convergence over Θ^2.

We assume the covariance kernel converges uniformly.

Assumption C.11. $\Omega_{F_n}(\cdot,\cdot)\to_u \Omega_\infty(\cdot,\cdot)$ for some continuous $R^{2k\times 2k}$-valued function $\Omega_\infty(\cdot,\cdot)$ on Θ^2.

Define lower-bound (wp$\to$1) random variables, $A^*_{Ln,\Delta L}$, on the bootstrap statistics $A^*_{n,\Delta L}$ to be

$$A^*_{Ln,\Delta L} := \inf_{\theta\in\Theta_{\min}^{\eta_{Ln}}(F_n)}\max_{j\leq k}\left(-\widehat{\nu}^*_{nj}(\theta) - \varphi^\dagger(-\xi^e_{1nj}(\theta))\right), \tag{27.12}$$

where $\{\eta_{Ln}\}_{i\leq n}$ are as in Assumption BC.6 and $\varphi^\dagger$ is defined in (27.7). Note that the lower-bound statistic has ι in place of $\widehat{sd}_{nj}(\theta)$, see (27.6), where $\iota\leq\widehat{sd}_{nj}(\theta)$, and $\Theta_{\min}^{\eta_{Ln}}(F_n)$ and $\varphi^\dagger$ in place of $\widehat{\Theta}_{\min,n}$ and φ, respectively, which appear in $A^*_{n,\Delta L}$. Let $\widehat{c}_{Ln,\Delta L}(1-\alpha)$ denote the $1-\alpha$ conditional quantile of $A^*_{Ln,\Delta L}$ given $\{W_{ni}\}_{i\leq n,n\geq 1}$ plus ι.

By Lemma 27.1(b) and (c) below, the asymptotic distribution of the $A^*_{Ln,\Delta L}$ bootstrap random variables, conditional on $\{W_{ni}\}_{i\le n,n\ge 1}$, is the following distribution a.s.$[P_\bigtriangledown]$:

$$A^*_{L\infty,\Delta L} := \inf_{(\theta,e^*)\in\Lambda^*_L} \max_{j\le k} \left(-G^{m\sigma}_j(\theta) - \varphi^\dagger(-e^*_j)\right) \tag{27.13}$$

for Λ^*_L as in Assumption BC.6. Let $c_{L\infty,\Delta L}(1-\alpha)$ denote the $1-\alpha$ quantile of $A^*_{L\infty,\Delta L}$ plus ι, which is nonrandom.

27.2 Lemmas 27.1–27.3, Theorem 27.4, and Lemma 27.5

The proof of Theorem 5.1 uses Theorem 27.4 below. The following lemmas are used in the proof of Theorem 27.4.

Lemma 27.1 *For a sequence $\{F_n\}_{n\ge1}$ that satisfies Assumptions* A.0, A.6, BC.3, BC.4, C.4, C.5, *and* C.12 *for a subsequence $\{p_n\}_{n\ge1}$ in place of $\{n\}_{n\ge1}$, there exists a subsequence $\{a_n\}_{n\ge1}$ of $\{p_n\}_{n\ge1}$ for which* (a) $\{A^*_{Ua_n,\Delta U}|\{W_{ni}\}_{i\le n,n\ge1}\} \to_d A^*_{U\infty,\Delta U}$ *a.s.$[P_\bigtriangledown]$ and* (b) $\{A^*_{La_n,\Delta L}|\{W_{ni}\}_{i\le n,n\ge1}\} \to_d A^*_{L\infty,\Delta L}$ *a.s.$[P_\bigtriangledown]$, provided Assumptions* A.7(i) *and* BC.6 *hold in place of Assumptions* A.6 *and* BC.4.

Comment. Lemma 27.1 is somewhat analogous to Theorem C.1 of BCS15.

Lemma 27.2 *For a sequence $\{F_n\}_{n\ge1}$ that satisfies Assumptions* A.0, A.6, BC.4, *and* BC.5 *for a subsequence $\{p_n\}_{n\ge1}$ in place of $\{n\}_{n\ge1}$,* (a) $P_\bigtriangledown(A^*_{Up_n,\Delta U} \ge A^*_{p_n,\Delta U}|\{W_{ni}\}_{i\le n,n\ge1}) = 1$ *wp→1 under $P_\bigtriangledown$ and* (b) $P_\bigtriangledown(A^*_{Lp_n,\Delta L} \le A^*_{p_n,\Delta L}|\{W_{ni}\}_{i\le n,n\ge1}) = 1$ *wp→1 under $P_\bigtriangledown$, provided Assumptions* A.7(i), BC.6, *and* BC.7 *hold in place of Assumptions* A.6, BC.4, *and* BC.5.

Lemma 27.3 *For a sequence $\{F_n\}_{n\ge1}$ that satisfies Assumptions* A.6, BC.3, BC.4, C.4, C.5, C.12, *and* C.13 *for a subsequence $\{p_n\}_{n\ge1}$ in place of $\{n\}_{n\ge1}$, we have* (a) $A^*_{U\infty,\Delta U} \le A_{\infty,\Delta\min}$ *for all sample realizations and* (b) $A^*_{L\infty,\Delta L} \ge A_{\infty,\Delta}$ *for all sample realizations, provided Assumptions* A.7, A.8, *and* BC.6 *hold in place of Assumptions* A.6, BC.4, *and* C.13.

Lemmas 27.1–27.3 are used to prove the following theorem, which employs some high-level assumptions that are verified in the proof of Theorem 5.1 below.

Theorem 27.4 *For $\alpha\in(0,1)$ and for a sequence $\{F_n\}_{n\ge1}$ that satisfies Assumptions* A.0, A.6, BC.3–BC.5, C.4, C.5, C.12, *and* C.13 *for a subsequence $\{p_n\}_{n\ge1}$ in place of $\{n\}_{n\ge1}$, there exists a subsequence $\{a_n\}_{n\ge1}$ of $\{p_n\}_{n\ge1}$ for which* (a) *the nominal level $1-\alpha$ upper-bound CI $CI_{a_n,\Delta U}(\alpha)$ for $\Delta^{\inf}_{F_{a_n}}$ satisfies*

$$\liminf_{n\to\infty} P_{F_{a_n}}(\Delta^{\inf}_{F_{a_n}} \in CI_{a_n,\Delta U}(\alpha)) \ge 1-\alpha \text{ and}$$

(b) *the nominal level* $1-\alpha$ *lower-bound CI* $CI_{a_n,\Delta L}(\alpha)$ *for* $\Delta^{\inf}_{F_{a_n}}$ *satisfies*

$$\liminf_{n\to\infty} P_{F_{a_n}}(\Delta^{\inf}_{F_{a_n}} \in CI_{a_n,\Delta L}(\alpha)) \geq 1-\alpha,$$

provided Assumptions A.7, A.8, BC.6, BC.7, *and* C.11 *hold in place of Assumptions* A.6, BC.4, *and* BC.5.

The proof of Theorem 5.1 also uses the following lemma, which concerns $\widehat{\Theta}_{\min,n}$ and $\widehat{\Theta}_{\min,L,n}$, which are defined in (5.7) and just below (5.15), respectively.

Lemma 27.5 *For a sequence* $\{F_n\}_{n\geq 1}$ *that satisfies Assumptions* A.0, C.4, C.5, *and* C.12, *we have* (a) *for any sequences of positive constants* $\{\eta_n\}_{n\geq 1}$ *and* $\{\tau_n\}_{n\geq 1}$ *that satisfy* $\tau_n \to \infty$ *and* $\eta_n/\tau_n \to 0$, $P_{F_n}(\widehat{\Theta}_{\min,n} \supseteq \Theta^{\eta_n}_{\min}(F_n)) \to 1$ *and* (b) *for any sequences of positive constants* $\{\eta_{Ln}\}_{n\geq 1}$ *that satisfy* $\eta_{Ln}\to\infty$, $P_{F_n}(\widehat{\Theta}_{\min,L,n} \subseteq \Theta^{\eta_{Ln}}_{\min}(F_n)) \to 1$.

27.3 Proof of Theorem 27.4

Proof of Theorem 27.4. First, we prove part (a). For notational simplicity, let $\widehat{c}_{n,\Delta U} := \widehat{c}_{n,\Delta U}(1-\alpha)$ (defined just after (5.14)), $\widehat{c}_{Un,\Delta U} := \widehat{c}_{Un,\Delta U}(1-\alpha)$ (defined following (27.8)), $c_{U\infty,\Delta U} := c_{U\infty,\Delta U}(1-\alpha)$ (defined following (27.9)), and $c^{-}_{\infty,\Delta} := c^{-}_{\infty,\Delta}(1-\alpha)$ (defined following (25.5)). We have: $A^*_{n,\Delta U}$ is defined in (5.14) with b deleted, $A^*_{Un,\Delta U}$ is defined in (27.8), $A^*_{U\infty,\Delta U}$ is defined in (27.9), and $A_{\infty,\Delta\min}$ is defined in (25.6). Note that $A_{\infty,\Delta} = A_{\infty,\Delta\min}$ by Theorem 25.2(c), so $c^{-}_{\infty,\Delta}$ equals the $1-\alpha$ quantile of $-A_{\infty,\Delta}$ and $-A_{\infty,\Delta\min}$.

Given a subsequence $\{p_n\}_{n\geq 1}$ as in the statement of the theorem, we consider a subsequence $\{a_n\}_{n\geq 1}$ of $\{p_n\}_{n\geq 1}$ as in Lemma 27.1. For the subsequence $\{a_n\}_{n\geq 1}$, the results of Lemmas 27.1, 27.2, and 27.3 hold. For notational simplicity, in the remainder of the proof we replace $\{a_n\}_{n\geq 1}$ by $\{n\}_{n\geq 1}$ and presume that the results of Lemmas 27.1, 27.2, and 27.3 hold for $\{n\}_{n\geq 1}$.

By the definition of $CI_{n,\Delta U}(\alpha)$ in (5.3) and the definition of $A_{n,\Delta}$ in (25.1),

$$P_{F_n}(\Delta^{\inf}_{F_n} \in CI_{n,\Delta U}(\alpha)) = P_{F_n}(-A_{n,\Delta} \leq \widehat{c}_{n,\Delta}). \tag{27.14}$$

If $A^*_{n,\Delta U} \leq A^*_{Un,\Delta U}$ with probability one (with respect to the bootstrap randomness) conditional on $\{W_{ni}\}_{i\leq n,n\geq 1}$, then the $1-\alpha$ conditional quantile of $-A^*_{n,\Delta U}$ given $\{W_{ni}\}_{i\leq n,n\geq 1}$ plus ι, which is $\widehat{c}_{n,\Delta U}$, is greater than or equal to the $1-\alpha$ conditional quantile of $-A^*_{Un,\Delta U}$ given $\{W_{ni}\}_{i\leq n,n\geq 1}$ plus ι, which is $\widehat{c}_{Un,\Delta U}$, as a consequence of the definition of a quantile. By Lemma 27.2(a), the "if" condition in the previous sentence holds wp$\to$1 (with respect to the randomness in the samples $\{W_{ni}\}_{i\leq n,n\geq 1}$). Hence, Lemma 27.2(a) implies that $\widehat{c}_{n,\Delta U} \geq \widehat{c}_{Un,\Delta U}$ wp$\to$1, which implies that

$\widehat{c}_{n,\Delta U} \geq \widehat{c}_{Un,\Delta U} + o_p(1)$, where the $o_p(1)$ term refers to randomness in the samples, not bootstrap randomness. This gives

$$\liminf_{n\to\infty} P_{F_n}(-A_{n,\Delta} \leq \widehat{c}_{n,\Delta U}) \geq \liminf_{n\to\infty} P_{F_n}(-A_{n,\Delta} + o_p(1) \leq \widehat{c}_{Un,\Delta U}). \tag{27.15}$$

Now, take an arbitrary $\varepsilon > 0$. Then, there exists $\varepsilon^* \in (0, \varepsilon)$ such that $c_{U\infty,\Delta U} - \varepsilon^*$ is a continuity point of $-A^*_{U\infty,\Delta U}$. We have

$$\begin{aligned} &\limsup_{n\to\infty} P_{\bigtriangledown}(-A^*_{Un,\Delta U} \leq c_{U\infty,\Delta U} - \varepsilon | \{W_{ni}\}_{i\leq n, n\geq 1}) \qquad (27.16)\\ &\leq \limsup_{n\to\infty} P_{\bigtriangledown}(-A^*_{Un,\Delta U} \leq c_{U\infty,\Delta U} - \varepsilon^* | \{W_{ni}\}_{i\leq n, n\geq 1}) = P(-A^*_{U\infty,\Delta U} \leq c_{U\infty,\Delta U} - \varepsilon^*) < 1 - \alpha \end{aligned}$$

a.s.$[P_{\bigtriangledown}]$, where the equality holds by Lemma 27.1(a) and the last inequality holds by the definition of the $1-\alpha$ quantile $c_{U\infty,\Delta U}$ of $-A^*_{U\infty,\Delta U}$. Because $\widehat{c}_{Un,\Delta U}$ is the $1-\alpha$ conditional quantile of $-A^*_{Un,\Delta U}$ given $\{W_{ni}\}_{i\leq n, n\geq 1}$ plus ι, if

$$P_{\bigtriangledown}(-A^*_{Un,\Delta U} \leq c_{U\infty,\Delta U} - \varepsilon | \{W_{ni}\}_{i\leq n, n\geq 1}) < 1 - \alpha, \text{ then } c_{U\infty,\Delta U} - \varepsilon < \widehat{c}_{Un,\Delta U} - \iota. \tag{27.17}$$

By (27.16), the first condition in (27.17) holds for n sufficiently large a.s.$[P_{\bigtriangledown}]$. Hence, the same is true for the second condition in (27.17). That is, $P_{\bigtriangledown}(c_{U\infty,\Delta U} + \iota - \varepsilon < \widehat{c}_{Un,\Delta U}$ for n large$) = 1$, or equivalently,

$$P_{\bigtriangledown}\left(\lim_{n\to\infty} 1(c_{U\infty,\Delta U} + \iota - \varepsilon < \widehat{c}_{Un,\Delta U}) = 1\right) = 1. \tag{27.18}$$

By the dominated convergence theorem, this implies that

$$\lim_{n\to\infty} P_{\bigtriangledown}\left(c_{U\infty,\Delta U} + \iota - \varepsilon < \widehat{c}_{Un,\Delta U}\right) = 1 \tag{27.19}$$

for all $\varepsilon > 0$, which also can be written as $\lim_{n\to\infty} P_{F_n}\left(c_{U\infty,\Delta U} + \iota - \varepsilon < \widehat{c}_{Un,\Delta U}\right) = 1$.

Next, we have: for all $\varepsilon > 0$,

$$\begin{aligned} &\liminf_{n\to\infty} P_{F_n}(-A_{n,\Delta} + o_p(1) \leq \widehat{c}_{Un,\Delta U}) \\ &= \liminf_{n\to\infty} P_{F_n}(-A_{n,\Delta} + o_p(1) \leq \widehat{c}_{Un,\Delta U} \ \&\ c_{U\infty,\Delta U} + \iota - \varepsilon < \widehat{c}_{Un,\Delta U}) \\ &\geq \liminf_{n\to\infty} P_{F_n}(-A_{n,\Delta} + o_p(1) \leq c_{U\infty,\Delta U} + \iota - \varepsilon \ \&\ c_{U\infty,\Delta U} + \iota - \varepsilon < \widehat{c}_{Un,\Delta U}) \\ &= \liminf_{n\to\infty} P_{F_n}(-A_{n,\Delta} + o_p(1) \leq c_{U\infty,\Delta U} + \iota - \varepsilon) \qquad (27.20) \end{aligned}$$

where the two equalities hold using (27.19) and the inequality is straightforward.

By Theorem 25.2(a) and (c), we have

$$A_{n,\Delta} \to_d A_{\infty,\Delta\min} \tag{27.21}$$

using Assumptions A.0, C.4, C.5, C.12, and C.13. Consider a sequence $\{\varepsilon_m\}_{m\geq 1}$ such that $c_\infty + \iota - \varepsilon_m$ is a continuity point of $-A_{\infty,\Delta}$ for all $m \geq 1$ and $\varepsilon_m \downarrow 0$ as $m \to \infty$. Then, we have

$$\begin{aligned}
&\liminf_{n\to\infty} P_{F_n}(-A_{n,\Delta} + o_p(1) \leq \widehat{c}_{Un,\Delta U}) \\
&\geq \lim_{m\to\infty} \liminf_{n\to\infty} P_{F_n}(-A_{n,\Delta\min} + o_p(1) \leq c_{U\infty,\Delta U} + \iota - \varepsilon_m) \\
&= \lim_{m\to\infty} P(-A_{\infty,\Delta\min} \leq c_{U\infty,\Delta U} + \iota - \varepsilon_m) \\
&\geq \lim_{m\to\infty} P(-A_{\infty,\Delta\min} \leq c^-_{\infty,\Delta} + \iota - \varepsilon_m) \\
&\geq 1 - \alpha,
\end{aligned} \tag{27.22}$$

where the first inequality holds by (27.20), the equality holds by (27.21) and the definition of $\{\varepsilon_m\}_{m\geq 1}$, the second inequality holds by Lemma 27.3(a) because $A^*_{U\infty,\Delta U} \leq_{ST} A_{\infty,\Delta\min}$ implies that $-A_{\infty,\Delta\min} \leq_{ST} -A^*_{U\infty,\Delta U}$ and $c^-_{\infty,\Delta} \leq c_{U\infty,\Delta U}$, and the last inequality holds by the definition of the $1-\alpha$ quantile $c^-_{\infty,\Delta}$ of $-A_{\infty,\Delta} = -A_{\infty,\Delta\min}$ because $\iota - \varepsilon_m > 0$ for m large. Equations (27.14), (27.15), and (27.22) prove part (a).

Next, we prove part (b). The proof is quite similar to that of part (a) with the changes described below. For notational simplicity, let $\widehat{c}_{n,\Delta L} := \widehat{c}_{n,\Delta L}(1-\alpha)$ (defined just above (5.15)), $\widehat{c}_{Ln,\Delta L} := \widehat{c}_{Ln,\Delta L}(1-\alpha)$ (defined following (27.12)), $c_{L\infty,\Delta L} := c_{L\infty,\Delta L}(1-\alpha)$ (defined following (27.13)), and $c_{\infty,\Delta} := c_{\infty,\Delta}(1-\alpha)$ (defined following (25.5)). We have: $A^*_{n,\Delta L}$ is defined in (5.15) with b deleted, $A^*_{Ln,\Delta L}$ is defined in (27.12), and $A^*_{L\infty,\Delta L}$ is defined in (27.13). In the proof of part (b), we use $A^*_{n,\Delta L}$, $\widehat{c}_{n,\Delta L}$, $A^*_{Ln,\Delta L}$, $\widehat{c}_{Ln,\Delta L}$, and $c_{L\infty,\Delta L}$ in place of $A^*_{n,\Delta U}$, $\widehat{c}_{n,\Delta U}$, $A^*_{Un,\Delta U}$, $\widehat{c}_{Un,\Delta U}$, and $c_{U\infty,\Delta U}$, respectively. As in part (a), $A_{\infty,\Delta} := A_{\infty,\Delta\min}$ by Theorem 25.2(c), so $c_{\infty,\Delta}$ equals the $1-\alpha$ quantile of $A_{\infty,\Delta}$ and $A_{\infty,\Delta\min}$.

To prove part (b), we use Lemmas 27.1(b), 27.2(b), and 27.3(b) in place of Lemmas 27.1(a), 27.2(a), and 27.3(a), respectively.

To prove part (b), (27.14) is replaced by

$$P_{F_n}(\Delta^{\inf}_{F_n} \in CI_{n,\Delta L}(\alpha)) = P_{F_n}(A_{n,\Delta} \leq \widehat{c}_{n,\Delta L}). \tag{27.23}$$

By the same argument as used to show (27.15), but applied to $A^*_{n,\Delta}$ and $A^*_{Ln,\Delta L}$, rather than $-A^*_{n,\Delta}$

and $-A^*_{Un,\Delta U}$, we obtain $\widehat{c}_{n,\Delta L} \geq \widehat{c}_{Ln,\Delta L} + o_p(1)$ and

$$\liminf_{n\to\infty} P_{F_n}(A_{n,\Delta} \leq \widehat{c}_{n,\Delta L}) \geq \liminf_{n\to\infty} P_{F_n}(A_{n,\Delta} + o_p(1) \leq \widehat{c}_{Ln,\Delta L}). \tag{27.24}$$

We obtain

$$\liminf_{n\to\infty} P_{F_n}(c_{L\infty,\Delta L} + \iota - \varepsilon \leq \widehat{c}_{Ln,\Delta L}) = 1 \text{ for all } \varepsilon > 0 \tag{27.25}$$

by the same argument as used to prove (27.19), but with $A^*_{Ln,\Delta L}$, $A^*_{L\infty,\Delta L}$, and $c_{L\infty,\Delta L} + \iota - \varepsilon^*$ in place of $-A^*_{Un,\Delta U}$, $-A^*_{U\infty,\Delta U}$, and $c_{U\infty,\Delta U} + \iota - \varepsilon^*$, respectively. By arguments analogous to those in (27.20) and (27.22), we obtain

$$\begin{aligned}
&\liminf_{n\to\infty} P_{F_n}(A_{n,\Delta} + o_p(1) \leq \widehat{c}_{Ln,\Delta L}) \geq \liminf_{n\to\infty} P_{F_n}(A_{n,\Delta} + o_p(1) \leq c_{L\infty,\Delta L} + \iota - \varepsilon) \text{ and} \\
&\liminf_{n\to\infty} P_{F_n}(A_{n,\Delta} + o_p(1) \leq \widehat{c}_{Ln,\Delta L}) \geq \lim_{m\to\infty} P(A_{\infty,\Delta\,\min} \leq c_{\infty,\Delta} + \iota - \varepsilon_m) \geq 1 - \alpha,
\end{aligned} \tag{27.26}$$

respectively, where the the last inequality holds by the definition of the $1-\alpha$ quantile $c_{\infty,\Delta}$ of $A_{\infty,\Delta} = A_{\infty,\Delta\,\min}$ because $\iota - \varepsilon_m > 0$ for m large. Equations (27.23), (27.24), and (27.26) combine to establish part (b). $\square$

27.4 Proof of Theorem 5.1

Proof of Theorem 5.1. We prove part (a) first. By definition, the asymptotic size of the CI $CI_{n,\Delta U}(\alpha)$ is

$$\liminf_{n\to\infty} \inf_{F\in\mathcal{P}} P_F\left(\Delta_F^{\inf} \in CI_{n,\Delta U}(\alpha)\right) = \liminf_{n\to\infty} \inf_{F\in\mathcal{P}} P_F\left(\Delta_F^{\inf} \leq \widehat{\Delta}^{\inf}_{n,U}(\alpha)\right). \tag{27.27}$$

There always exists a sequence $\{F_n\}_{n\geq1}$ and a subsequence $\{q_n\}_{n\geq1}$ of $\{n\}_{n\geq1}$ such that

$$\begin{aligned}
\liminf_{n\to\infty} \inf_{F\in\mathcal{P}} P_F\left(\Delta_F^{\inf} \leq \widehat{\Delta}^{\inf}_{n,U}(\alpha)\right) &= \liminf_{n\to\infty} P_{F_n}\left(-n^{1/2}(\widehat{\Delta}^{\inf}_n - \Delta^{\inf}_{F_n}) \leq \widehat{c}_{n,\Delta U}(1-\alpha)\right) \\
&= \lim_{n\to\infty} P_{F_{q_n}}\left(-A_{q_n,\Delta} \leq \widehat{c}_{q_n,\Delta U}(1-\alpha)\right),
\end{aligned} \tag{27.28}$$

where the first and second equalities use (5.3) and (25.1), respectively. Hence, to establish that $CI_{n,\Delta U}(\alpha)$ has correct asymptotic size, we must show that the rhs of (27.28) is $1-\alpha$ or greater. It suffices to show that the rhs of (27.28) is $1-\alpha$ or greater with $\{q_n\}_{n\geq1}$ replaced by some subsequence $\{a_n\}_{n\geq1}$ of $\{q_n\}_{n\geq1}$ (because the limit under the subsequence $\{a_n\}_{n\geq1}$ is the same as the limit under the original subsequence $\{q_n\}_{n\geq1}$). The rhs of (27.28) defined with $\{a_n\}_{n\geq1}$ in place of $\{q_n\}_{n\geq1}$ is $1-\alpha$ or greater by Theorem 27.4(a) provided the assumptions of Theorem 27.4(a) hold for some subsequence $\{p_n\}_{n\geq1}$ of $\{q_n\}_{n\geq1}$ and $\{a_n\}_{n\geq1}$ is some subsequence of $\{p_n\}_{n\geq1}$ (because

$-A_{a_n,\Delta} \leq \widehat{c}_{a_n,\Delta U}(1-\alpha)$ iff $\Delta_{F_{a_n}} \in CI_{a_n,\Delta U}(\alpha)$ by (27.14) and the $\liminf_{n\to\infty}$ is actually the $\lim_{n\to\infty}$ in the result of Theorem 27.4(a) for any subsequence $\{a_n\}_{n\geq 1}$ of $\{q_n\}_{n\geq 1}$ by the definition of $\{q_n\}_{n\geq 1}$) in (27.28). Hence, for part (a), it remains to verify that Assumptions BC.3, BC.4, BC.5, C.4, C.5, C.12, and C.13 hold for some subsequence $\{p_n\}_{n\geq 1}$ (of $\{q_n\}_{n\geq 1}$) in place of $\{n\}_{n\geq 1}$ (because Assumptions A.0 and A.6, which are imposed in Theorem 27.4, are also imposed in the present theorem).

To prove part (b) regarding the lower-bound CI $CI_{n,\Delta L}(\alpha)$, analogous arguments to those in (27.27) and (27.28) show that it suffices to show that $\lim P_{F_{q_n}}(A_{q_n,\Delta} \leq \widehat{c}_{q_n,\Delta L}(1-\alpha)) \geq 1-\alpha$, where $\{q_n\}_{n\geq 1}$ is a subsequence of $\{n\}_{n\geq 1}$ for which (27.28) holds with $\widehat{\Delta}^{\inf}_{n,L}(\alpha)$, $n^{1/2}(\widehat{\Delta}^{\inf}_n - \Delta^{\inf}_{F_n})$, $\widehat{c}_{n,\Delta L}(1-\alpha)$, and $A_{q_n,\Delta}$ in place of $\widehat{\Delta}^{\inf}_{n,U}(\alpha)$, $-n^{1/2}(\widehat{\Delta}^{\inf}_n - \Delta^{\inf}_{F_n})$, $\widehat{c}_{n,\Delta U}(1-\alpha)$, and $-A_{q_n,\Delta}$, respectively. By the same argument as in the previous paragraph, but with Theorem 27.4(b) in place of Theorem 27.4(a), to prove part (b) it suffices to verify the assumptions employed in Theorem 27.4(b) that are not imposed in Theorem 5.1(b). The assumptions that need to be verified are the same as those for part (a) except with Assumptions BC.6, BC.7, and C.11 in place of Assumptions BC.4 and BC.5 (because Assumptions A.7 and A.8, which are imposed in Theorem 27.4(b), are also imposed in Theorem 5.1(b)).

We now verify that Assumptions BC.3, BC.4–BC.7, C.4, C.5, C.11, C.12, and C.13 hold for some subsequence $\{p_n\}_{n\geq 1}$ (of $\{q_n\}_{n\geq 1}$) in place of $\{n\}_{n\geq 1}$. Given $\{\tau_n\}_{n\geq 1}$ in the definition of $\widehat{\Theta}_{\min,n}$ in (5.7) that satisfies Assumption A.6(ii), take $\{\eta_n\}_{n\geq 1}$ to be the same in Assumptions C.13 and BC.4, which requires $\eta_n \to \infty$ and $\eta_n/\tau_n \to 0$. For example, one can take $\eta_n = \tau_n^{1/2}$ $\forall n \geq 1$. Given $\{\kappa_n\}_{n\geq 1}$ that satisfy Assumption A.7, which requires $\kappa_n \to \infty$ and $\kappa_n/n^{1/2} \to 0$, take $\{\eta_{Ln}\}_{n\geq 1}$ to be the same as in Assumptions BC.6 and BC.7, which requires $\eta_{Ln} \to \infty$, $\eta_{Ln}/n^{1/2} \to 0$, and $\eta_{Ln}/\kappa_n^{\gamma} \to 0$ for $\gamma \in (0,1)$ as in Assumption BC.6. For example, one can take $\eta_{Ln} = \kappa_n^{\gamma/2}$.

Under Assumptions A.4 and A.5, by Lemma D.7 of BCS15, given $\{q_n\}_{n\geq 1}$, there exists a subsequence $\{p_n\}_{n\geq 1}$ of $\{q_n\}_{n\geq 1}$, a continuous $R^{2k\times 2k}$-valued function Ω_∞ on Θ^2, and a continuous R^k-valued function $\widetilde{m}$ on Θ for which (i) $\Omega_{F_{p_n}} \to_u \Omega_\infty$, where $\to_u$ denotes uniform convergence (over Θ^2 in this case), and hence, Assumption C.11 holds for the subsequence $\{p_n\}_{n\geq 1}$, (ii) $E_{F_{p_n}}\widetilde{m}(W,\cdot) \to_u \widetilde{m}(\cdot)$, and hence, Assumption C.4 holds for the subsequence $\{p_n\}_{n\geq 1}$, and (iii) Assumptions C.12, C.13, BC.4, and BC.6 hold for the subsequence $\{p_n\}_{n\geq 1}$. The basic argument used by BCS15 to prove their Lemma D.7 is that a sequence in a compact subset of a metric space has a convergent subsequence. Strictly speaking, Lemma D.7 of BCS15 only establishes $\Omega_{F_{p_n}} \to_u \Omega_\infty$ for the upper left $k \times k$ submatrices of these matrix functions and the subsequence version of Assumption C.7 of AK. But, the same argument applies for the $2k \times 2k$–valued functions $\Omega_{F_{p_n}}$ and Ω_∞, and the same argument as for Assumption C.7 of AK applies to Assumptions C.12, C.13, BC.4, and BC.6. In addition, the result $E_{F_{p_n}}\widetilde{m}(W,\cdot) \to_u \widetilde{m}(\cdot)$ is established in the same way as $\Omega_{F_{p_n}} \to_u \Omega_\infty$ (but using Assumption A.5 in place of Assumption A.4).

Assumption C.5 holds for the subsequence $\{p_n\}_{n\geq 1}$ by applying a subsequence version of Lemma 20.1(a) in online Appendix B, which imposes Assumptions A.0–A.4 and C.11. Assumptions A.0–A.4 are imposed in the present theorem and the subsequence version of Assumption C.11 holds by (i) above.

Assumption BC.3 holds for the subsequence $\{p_n\}_{n\geq 1}$ by Lemma D.2(8) of BCS15 because Assumptions A.1–A.4 of this paper imply Assumptions A.1–A.4 of BCS15 and $\Omega_{F_{p_n}} \to_u \Omega_\infty$.

Assumption BC.5 holds in part (a) of the theorem by a subsequence version of Lemma 27.5(a) (with $\{p_n\}_{n\geq 1}$ in place of $\{n\}_{n\geq 1}$), which imposes Assumptions A.0, C.4, C.5, and C.12 and requires $\tau_n \to \infty$ and $\eta_n/\tau_n \to 0$ (because these assumptions are verified above, $\tau_n \to \infty$ by Assumption A.6(ii), and given $\{\tau_n\}_{n\geq 1}$, $\{\eta_n\}_{n\geq 1}$ are defined above to satisfy $\eta_n/\tau_n \to 0$).

Assumption BC.7 holds in part (b) of the theorem by a subsequence version of Lemma 27.5(b) (with $\{p_n\}_{n\geq 1}$ in place of $\{n\}_{n\geq 1}$), which imposes Assumptions A.0, C.4, C.5, and C.12 and requires $\eta_{Ln} \to \infty$ (because these assumptions are verified above and $\{\eta_{Ln}\}_{n\geq 1}$ are defined above to satisfy $\eta_{Ln} \to \infty$).

This concludes the proof that the assumptions employed in parts (a) and (b) of Theorem 27.4 hold for the subsequence $\{p_n\}_{n\geq 1}$ of $\{q_n\}_{n\geq 1}$, which completes the proof. $\square$

27.5 Proof of Lemma 27.1

Proof of Lemma 27.1. First, we prove part (a). We have

$$\begin{aligned} n^{1/2}\widehat{m}_{nj}(\theta) &= \frac{\sigma_{F_nj}(\theta)}{\widehat{\sigma}_{nj}(\theta)}\left(\nu_{nj}^m(\theta) + n^{1/2}E_{F_n}\widetilde{m}_j(W,\theta)\right) \\ &= \widehat{\omega}_{nj}(\theta) + n^{1/2}E_{F_n}\widetilde{m}_j(W,\theta), \text{ where} \qquad (27.29) \\ \widehat{\omega}_{nj}(\theta) &:= \frac{\sigma_{F_nj}(\theta)}{\widehat{\sigma}_{nj}(\theta)}\nu_{nj}^m(\theta) - n^{1/2}\left(\frac{\widehat{\sigma}_{nj}(\theta)}{\sigma_{F_nj}(\theta)} - 1\right)\frac{\sigma_{F_nj}(\theta)}{\widehat{\sigma}_{nj}(\theta)}E_{F_n}\widetilde{m}_j(W,\theta) = O_p^\Theta(1), \end{aligned}$$

where $\nu_{nj}^m(\theta)$ denotes the jth element of $\nu_n^m(\theta)$ defined in (14.2) in online Appendix B, and the second equality on the last line holds by Assumptions C.4 and C.5 and (26.10). Next, we have

$$\begin{aligned} n^{1/2}\left(\widehat{\Delta}_{nj}(\theta) - \widehat{\Delta}_n^{\inf}\right) &= n^{1/2}(-\widehat{m}_{nj}(\theta) + E_F\widetilde{m}_j(W,\theta)) - n^{1/2}(\widehat{\Delta}_n^{\inf} - \Delta_{F_n}^{\inf}) + e_{nj}(\theta) \\ &= \widehat{d}_{nj}(\theta) + e_{nj}(\theta), \text{ where} \\ \widehat{d}_{nj}(\theta) &:= -\widehat{\omega}_{nj}(\theta) - n^{1/2}(\widehat{\Delta}_n^{\inf} - \Delta_{F_n}^{\inf}) = O_p^\Theta(1), \qquad (27.30) \end{aligned}$$

the first equality uses $\widehat{\Delta}_{nj}(\theta) := -\widehat{m}_{nj}(\theta)$ by (5.2) and $e_{nj}(\theta) := n^{1/2}(-E_{F_n}\widetilde{m}_j(W,\theta) - \Delta^{\inf}_{F_n})$ by (27.3) and the second equality on the last line holds because $\widehat{\omega}_{nj}(\theta) = O^{\Theta}_p(1)$ by (27.29) and $n^{1/2}(\widehat{\Delta}^{\inf}_n - \Delta^{\inf}_{F_n}) := A_{n,\Delta} = O_p(1)$ by (25.1) and Theorem 25.2(a) (which uses Assumptions A.0, C.4, C.5, and C.12).

For $e^*_j = (\iota\kappa_n)^{-1}n^{1/2}(\Delta_{F_nj}(\theta) - \Delta^{\inf}_{F_n})$ as in $\Lambda^{*\eta_n}_{n,\Delta,F_n}$ (defined in (27.5)), we obtain

$$\xi^e_{1nj}(\theta) := (\iota\kappa_n)^{-1}n^{1/2}\left(\widehat{\Delta}_{nj}(\theta) - \widehat{\Delta}^{\inf}_n\right) = (\iota\kappa_n)^{-1}\widehat{d}_{nj}(\theta) + e^*_j, \tag{27.31}$$

where the first equality holds by definition, see (27.6), and the second equality holds by (27.30).

Using (27.29), (27.31), and the definition of $\Lambda^{*\eta_n}_{n,\Delta,F_n}$, we can write $A^*_{Un,\Delta U}$ in (27.8) as

$$A^*_{Un,\Delta U} = \inf_{(\theta,e,e^*,j^*)\in\Lambda^{*\eta_n}_{n,\Delta,F_n}} \max_{j\le k}\left(-\widehat{\nu}^*_{nj}(\theta) + 1(j\neq j^*)e_j + 1(j=j^*)\varphi^{\dagger}((\iota\kappa_n)^{-1}\widehat{d}_{nj}(\theta) + e^*_j)\right), \tag{27.32}$$

where $(\theta, e_j, e^*_j, j^*) \in \Lambda^{*\eta_n}_{n,\Delta,F_n}$ implies that $e_j := e_{nj}(\theta)$, $e^*_j := (\iota\kappa_n)^{-1}e_{nj}(\theta)$, and $j^* := j_{ne}(\theta)$.

We have

$$(\iota\kappa_n)^{-1}\widehat{d}_{nj}(\theta) = o^{\Theta}_p(1) \tag{27.33}$$

by (27.30) and Assumption A.6(i).

Next, for all $j \le k$, we have

$$n^{1/2}\left(\widehat{m}_{nj}(\theta) - E_{F_n}\widetilde{m}_j(W,\theta)\right) = O^{\Theta}_p(1), \tag{27.34}$$

by (27.29). This and Assumption C.4 give

$$\widehat{m}_{nj}(\theta) - \widetilde{m}_j(\theta) = o^{\Theta}_p(1). \tag{27.35}$$

Now, we use the result that for any sequence of random variables $\{X_n\}_{n\ge1}$ on $(\Omega, \mathcal{F}, P_{\bigtriangledown})$ for which $X_n \to_p 0$, there exists a subsequence $\{a_n\}_{n\ge1}$ of $\{n\}_{n\ge1}$ such that $X_{a_n} \to 0$ a.s.$[P_{\bigtriangledown}]$, e.g., see Theorem 9.2.1 of Dudley (1989). We apply this result with the original sequence $\{n\}_{n\ge1}$ replaced by some subsequence $\{p_n\}_{n\ge1}$. Using this, (27.33), and (27.35), given any subsequence $\{p_n\}_{n\ge1}$ of $\{n\}_{n\ge1}$, there exists a subsequence $\{a_n\}_{n\ge1}$ of $\{p_n\}_{n\ge1}$ for which

$$\sup_{\theta\in\Theta}|(\iota\kappa_{a_n})^{-1}\widehat{d}_{a_nj}(\theta)| = o(1) \text{ a.s.}[P_{\bigtriangledown}] \text{ and } \sup_{\theta\in\Theta}|\widehat{m}_{a_nj}(\theta) - \widetilde{m}_j(\theta)| = o(1) \text{ a.s.}[P_{\bigtriangledown}]. \tag{27.36}$$

Define

$$\nu^{m\sigma*}_{nj}(\theta) := \nu^{m*}_{nj}(\theta) - \frac{1}{2}\widetilde{m}_j(\theta)\nu^{\sigma*}_{nj}(\theta)\ \forall j \le k. \tag{27.37}$$

We show that under $\{F_n\}_{n\geq 1}$, for the subsequence $\{a_n\}_{n\geq 1}$ of $\{p_n\}_{n\geq 1}$ defined above,

$$\sup_{\theta\in\Theta} |\widehat{\nu}^*_{a_n j}(\theta) - \nu^{m\sigma*}_{a_n j}(\theta)| = o_p(1)\ \forall j \leq k \text{ conditional on } \{W_{ni}\}_{i\leq n, n\geq 1} \text{ a.s.}[P_\bigtriangledown]. \quad (27.38)$$

The proof of (27.38) is as follows. By the same argument as in (26.10) with $\widehat{\sigma}^*_{nj}(\theta)$ and $\widehat{\sigma}_{nj}(\theta)$ in place of $\widehat{\sigma}_{nj}(\theta)$ and $\sigma_{F_n j}(\theta)$, respectively, we obtain

$$n^{1/2}\left(\frac{\widehat{\sigma}^*_{nj}(\theta)}{\widehat{\sigma}_{nj}(\theta)} - 1\right) = \frac{1}{2}\nu^{\sigma*}_{nj}(\theta) + o^\Theta_p(1) \text{ conditional on } \{W_{ni}\}_{i\leq n, n\geq 1} \text{ a.s.}[P_\bigtriangledown], \quad (27.39)$$

using Assumption BC.3 in place of Assumption C.5. Next, we have: for the subsequence $\{a_n\}_{n\geq 1}$,

$$\begin{aligned}\widehat{\nu}^*_{a_n j}(\theta) &:= a_n^{1/2}\left(\frac{\overline{m}^*_{a_n j}(\theta)}{\widehat{\sigma}^*_{a_n j}(\theta)} - \widehat{m}_{a_n j}(\theta)\right) = \frac{\widehat{\sigma}_{a_n j}(\theta)}{\widehat{\sigma}^*_{a_n j}(\theta)}\left(\nu^{m*}_{a_n j}(\theta) - \widehat{m}_{a_n j}(\theta) a_n^{1/2}\left(\frac{\widehat{\sigma}^*_{a_n j}(\theta)}{\widehat{\sigma}_{a_n j}(\theta)} - 1\right)\right)\\ &= (1 + o^\Theta_p(1))\left(\nu^{m*}_{a_n j}(\theta) - \frac{1}{2}\widetilde{m}_j(\theta)\nu^{\sigma*}_{a_n j}(\theta) + o^\Theta_p(1)\right) = \nu^{m\sigma*}_{a_n j}(\theta) + o^\Theta_p(1)\end{aligned} \quad (27.40)$$

conditional on $\{W_{ni}\}_{i\leq n, n\geq 1}$ a.s.$[P_\bigtriangledown]$, where the third equality holds by (27.36) and (27.39), and the fourth equality holds by the definition of $\nu^{m\sigma*}_{nj}(\theta)$ in (27.37) and Assumption BC.3. This proves (27.38).

Define

$$\overrightarrow{A}^*_{Un,\Delta U} := \inf_{(\theta,e,e^*,j^*)\in\Lambda^{*\eta_n}_{n,\Delta,F_n}} \max_{j\leq k}\left(-\nu^{m\sigma*}_{nj}(\theta) + 1(j\neq j^*)e_j + 1(j = j^*)\varphi^\dagger((\iota\kappa_n)^{-1}\widehat{d}_{nj}(\theta) + e^*_j)\right). \quad (27.41)$$

By (27.32), the first result of (27.36), (27.38), and (27.41), we obtain:

$$A^*_{Ua_n,\Delta U} = \overrightarrow{A}^*_{Ua_n,\Delta U} + o_p(1) \text{ conditional on } \{W_{ni}\}_{i\leq n, n\geq 1} \text{ a.s.}[P_\bigtriangledown]. \quad (27.42)$$

Hence, it suffices to show: for the subsequence $\{a_n\}_{n\geq 1}$,

$$\{\overrightarrow{A}^*_{Ua_n,\Delta U} | \{W_{ni}\}_{i\leq n, n\geq 1}\} \to_d A^*_{U\infty,\Delta U} \text{ a.s.}[P_\bigtriangledown]. \quad (27.43)$$

To prove (27.43), we use a similar (but somewhat more complicated) argument to that used to prove Theorem 25.2(a) based on the extended continuous mapping theorem. As above, let $\mathcal{D}$ be the space of functions from Θ to R^{2k}. Let $\mathcal{D}_0$ be the subset of uniformly continuous functions in

$\mathcal{D}$. For nonstochastic functions $\nu(\cdot) \in \mathcal{D}$ and $\mu(\cdot) : \Theta \to R^k$ with $\mu(\theta) = (\mu_1(\theta), ..., \mu_k(\theta))'$, define

$$\widetilde{g}_n(\nu(\cdot), \mu(\cdot)) := \inf_{(\theta,e,e^*,j^*)\in\Lambda^{*\eta_n}_{n,\Delta,F_n}} \max_{j\le k} \left(\tau_j(\nu(\cdot),\theta) + 1(j \ne j^*)e_j + 1(j = j^*)\varphi^\dagger(\mu_{j^*}(\theta) + e^*_{j^*})\right) \text{ and}$$
$$\widetilde{g}(\nu(\cdot), \mu(\cdot)) := \inf_{(\theta,e,e^*,j^*)\in\Lambda^*_{\Delta\min}} \max_{j\le k} \left(\tau_j(\nu(\cdot),\theta) + 1(j \ne j^*)e_j + 1(j = j^*)\varphi^\dagger(\mu_{j^*}(\theta) + e^*_{j^*})\right), \tag{27.44}$$

where $\nu(\theta) = (\nu^m(\theta)', \nu^\sigma(\theta)')'$, $\nu^m_j(\theta)$ and $\nu^\sigma_j(\theta)$ denote the jth elements of $\nu^m(\theta)$ and $\nu^\sigma(\theta)$, respectively, and $\tau_j(\nu(\cdot),\theta)$ is defined in (26.15). Note that

$$\overrightarrow{A}^*_{Un,\Delta U} = \widetilde{g}_n(\nu^*_n(\cdot), \mu_n(\theta)) \text{ and } A^*_{U\infty,\Delta U} = \widetilde{g}(G(\cdot), 0_k(\cdot)), \text{ where}$$
$$\mu_{nj}(\theta) := (\iota\kappa_n)^{-1}\widehat{d}_{nj}(\theta),\ \mu_n(\theta) = (\mu_{n1}(\theta), ..., \mu_{nk}(\theta))', \tag{27.45}$$

and $0_k(\cdot)$ is the zero function on Θ.

We want to show that $\{\widetilde{g}_{a_n}(\nu^*_{a_n}(\cdot), \mu_n(\cdot))|\{W_{ni}\}_{i\le n,n\ge1}\} \to_d \widetilde{g}(G(\cdot), 0_k(\cdot))$ a.s.$[P_\triangledown]$, where $\{\nu^*_{a_n}(\cdot)|\{W_{ni}\}_{i\le n,n\ge1}\} \Rightarrow G(\cdot)$ a.s.$[P_\triangledown]$ by Assumption BC.3 and $\sup_{\theta\in\Theta} ||\mu_{a_n}(\theta)|| = o(1)$ a.s.$[P_\triangledown]$ by (27.36). We use the extended CMT to establish this result. For notational simplicity, we employ n, rather than a_n, in the proof of this result. The extended CMT requires showing: for any deterministic sequences $\{\nu_n(\cdot) \in \mathcal{D}\}_{n\ge1}$ and $\{\mu_n(\cdot) : \Theta \to R^k\}_{n\ge1}$ and deterministic function $\nu(\cdot) \in \mathcal{D}_0$ such that $\sup_{\theta\in\Theta} ||\nu_n(\theta) - \nu(\theta)|| \to 0$ and $\sup_{\theta\in\Theta} ||\mu_n(\theta)|| \to 0$, we have $\widetilde{g}_n(\nu_n(\cdot), \mu_n(\cdot)) \to \widetilde{g}(\nu(\cdot), 0_k(\cdot))$. (For notational simplicity, we abuse notation here and consider deterministic $\nu_n(\cdot)$ and $\mu_n(\cdot)$ that differ from the random $\nu_n(\cdot)$ in Assumption C.5 and $\mu_n(\theta)$ defined in (27.45).) Once we have shown this, the proof is complete.

The proof of $\widetilde{g}_n(\nu_n(\cdot), \mu_n(\cdot)) \to \widetilde{g}(\nu(\cdot), 0_k(\cdot))$ is an extension of the proof of $g_n(\nu_n(\cdot)) \to g(\nu(\cdot))$ in (26.17)–(26.23) in the proof of Theorem 25.2(a). We show

$$\text{(i) } \liminf_{n\to\infty} \widetilde{g}_n(\nu_n(\cdot), \mu_n(\cdot)) \ge g(\nu(\cdot), 0_k(\cdot)) \text{ and}$$
$$\text{(ii) } \limsup_{n\to\infty} \widetilde{g}_n(\nu_n(\cdot), \mu_n(\cdot)) \le g(\nu(\cdot), 0_k(\cdot)). \tag{27.46}$$

First, we establish (i) in (27.46). There exists a subsequence $\{c_n\}_{n\ge1}$ of $\{n\}_{n\ge1}$ and a sequence $\{(\overline{\theta}_{c_n}, \overline{e}_{c_n}, \overline{e}^*_{c_n}, \overline{j}^*_{c_n}) \in \Lambda^{*\eta_{c_n}}_{c_n,\Delta,F_{c_n}}\}_{n\ge1}$ such that

$$\liminf_{n\to\infty} \widetilde{g}_n(\nu_n(\cdot), \mu_n(\cdot)) = \lim_{n\to\infty} \widetilde{g}_{c_n}(\nu_{c_n}(\cdot), \mu_{c_n}(\cdot)) \text{ and}$$
$$\lim_{n\to\infty} \widetilde{g}_{c_n}(\nu_{c_n}(\cdot), \mu_{c_n}(\cdot)) = \lim_{n\to\infty} \max_{j\le k} \big(\tau_j(\nu_{c_n}(\cdot), \overline{\theta}_{c_n}) + 1(j \ne \overline{j}^*_{c_n})\overline{e}_{c_nj}$$
$$+1(j = \overline{j}^*_{c_n})\varphi^\dagger(\mu_{\overline{j}^*_{c_n}}(\overline{\theta}_{c_n}) + \overline{e}^*_{c_n\overline{j}^*_{c_n}})\big), \tag{27.47}$$

where $\overline{e}_{c_nj}$ and $\overline{e}^*_{c_nj}$ denote the jth elements of $\overline{e}_{c_n}$ and $\overline{e}^*_{c_n}$, respectively. Also, there exists a subsequence $\{q_n\}_{n\geq 1}$ of $\{c_n\}_{n\geq 1}$ and $(\overline{\theta}, \overline{e}, \overline{e}^*, \overline{j}^*) \in \Theta \times R^{2k}_{[\pm\infty]} \times \{1, ..., k\}$ such that

$$d\left((\overline{\theta}_{q_n}, \overline{e}_{q_n}, \overline{e}^*_{q_n}, \overline{j}^*_{q_n}), (\overline{\theta}, \overline{e}, \overline{e}^*, \overline{j}^*)\right) \to 0, \tag{27.48}$$

where d is defined in the paragraph before (25.1), by compactness of the metric space $(\Theta \times R^{2k}_{[\pm\infty]} \times \{1, ..., k\}, d)$ under Assumption A.0(i). We have $(\overline{\theta}, \overline{e}, \overline{e}^*, \overline{j}^*) \in \Lambda^*_{\Delta\min}$ by the same argument as used to show $(\widetilde{\theta}, \widetilde{e}) \in \Lambda_\Delta$ in (26.3) (but without the requirement that $\Delta_F(\widetilde{\theta}_n) = \Delta^{\inf}_F\ \forall n \geq 1$) using (27.48) and Assumption BC.4.

For all $j \leq k$,

$$\lim_{n\to\infty} \tau_j(\nu_{q_n}(\cdot), \overline{\theta}_{q_n}) = -\nu^m_j(\overline{\theta}) + \frac{1}{2}\widetilde{m}_j(\overline{\theta})\nu^\sigma_j(\overline{\theta}) := \tau_j(\nu(\cdot), \overline{\theta}) \in R, \tag{27.49}$$

where the first equality holds by $\nu_{q_n}(\theta) \to \nu(\theta) = (\nu^m(\theta)', \nu^\sigma(\theta)')'$ uniformly over $\theta \in \Theta$ (by assumption) and (27.48), the last equality holds by the definition of $\tau_j(\nu(\cdot), \theta)$ in (26.15), and "$\in R$" holds because $\nu^m_j(\overline{\theta})$ and $\nu^\sigma_j(\overline{\theta})$ are finite since $\nu(\cdot)$ is assumed to be in $\mathcal{D}$ and $\widetilde{m}_j(\overline{\theta})$ is finite by Assumption C.4.

In addition, we have, for all $j \leq k$,

$$\begin{aligned} &1(j \neq \overline{j}^*_{q_n})\overline{e}_{q_nj} \to 1(j \neq \overline{j}^*)\overline{e}_j \text{ and} \\ &1(j = \overline{j}^*_{q_n})\varphi^\dagger(\mu_{q_n\overline{j}^*_{q_n}}(\overline{\theta}_{q_n}) + \overline{e}^*_{q_n\overline{j}^*_{q_n}}) \to 1(j = \overline{j}^*)\varphi^\dagger(\overline{e}^*_{\overline{j}^*}), \end{aligned} \tag{27.50}$$

where the first line holds by (27.48) and the second line holds by (27.48), $\sup_{\theta\in\Theta} ||\mu_{q_n}(\theta)|| \to 0$, and the continuity of $\varphi^\dagger$ on $R_{[+\infty]}$ under d, which is property (ii) of $\varphi^\dagger$ stated following (27.7), and the fact that $d(\varphi^\dagger(\mu_{\overline{j}^*_{q_n}}(\overline{\theta}_{q_n}) + \overline{e}^*_{q_n\overline{j}^*_{q_n}}), \varphi^\dagger(\overline{e}^*_{\overline{j}^*})) \to 0$ implies that $\varphi^\dagger(\mu_{\overline{j}^*_{q_n}}(\overline{\theta}_{q_n}) + \overline{e}^*_{q_n\overline{j}^*_{q_n}}) \to \varphi^\dagger(\overline{e}^*_{\overline{j}^*})$ (as a sequence of numbers in $R_{[+\infty]}$) even if $\varphi^\dagger(\overline{e}^*_{\overline{j}^*}) = +\infty$.

Now, we have

$$\begin{aligned} &\liminf_{n\to\infty} \widetilde{g}_n(\nu_n(\cdot), \mu_n(\cdot)) \\ &= \lim_{n\to\infty} \max_{j\leq k} \left(\tau_j(\nu_{q_n}(\cdot), \overline{\theta}_{q_n}) + 1(j \neq \overline{j}^*_{q_n})\overline{e}_{q_nj} + 1(j = \overline{j}^*_{q_n})\varphi^\dagger(\mu_{\overline{j}^*_{q_n}}(\overline{\theta}_{q_n}) + \overline{e}^*_{q_n\overline{j}^*_{q_n}})\right) \\ &= \max_{j\leq k} \left(\tau_j(\nu(\cdot), \overline{\theta}) + 1(j \neq \overline{j}^*)\overline{e}_j + 1(j = \overline{j}^*)\varphi^\dagger(\overline{e}^*_{\overline{j}^*})\right) \\ &\geq \inf_{(\theta,e,e^*,j^*)\in\Lambda^*_{\Delta\min}} \max_{j\leq k} \left(\tau_j(\nu(\cdot), \theta) + 1(j \neq j^*)e_j + 1(j = j^*)\varphi^\dagger(e^*_{j^*})\right) \\ &:= \widetilde{g}(\nu(\cdot), 0_k(\cdot)), \end{aligned} \tag{27.51}$$

where the first equality holds by (27.47) and the fact that $\{q_n\}_{n\geq 1}$ is a subsequence of $\{c_n\}_{n\geq 1}$, the second equality holds by (27.49) (using the notational convention that $\nu + c = c$ when $\nu \in R$ and $c = \pm\infty$ if $\overline{e}_j = \pm\infty$ for any $j \leq k$) and (27.50), the inequality holds because $(\overline{\theta}, \overline{e}, \overline{e}^*, \overline{j}^*) \in \Lambda^*_{\Delta\,\min}$ by the paragraph containing (27.48), and the last equality holds by the definition of $\widetilde{g}(\nu(\cdot), \mu(\cdot))$ in (27.44) with $\mu(\cdot) = 0_k(\cdot)$. This establishes result (i) in (27.46).

Next, we establish result (ii) in (27.46). There exists $(\theta^\dagger, e^\dagger, e^{\dagger *}, j^{\dagger *}) \in \Lambda^*_{\Delta\,\min}$ such that

$$\widetilde{g}(\nu(\cdot), 0_k(\cdot)) = \max_{j\leq k}\left(\tau_j(\nu(\cdot), \theta^\dagger) + 1(j \neq j^{\dagger *})e^\dagger_j + 1(j = j^{\dagger *})\varphi^\dagger(e^{\dagger *}_{j^{\dagger *}})\right) \tag{27.52}$$

because $\Lambda^*_{\Delta\,\min}$ is compact under the metric d defined in the paragraph before (25.1) with $a_* = d_\theta + 2k + 1$ (since it is assumed to be an element of $\mathcal{S}(\Theta \times R^{2k}_{[\pm\infty]} \times \{1, ..., k\})$) and $\tau_j(\nu(\cdot), \theta) + 1(j \neq j^*)e_j + 1(j = j^*)\varphi^\dagger(e^*_{j^*})$ is a continuous function of (θ, e, e^*, j^*) under d that takes values in the extended real line, using property (ii) of $\varphi^\dagger$ stated following (27.7). By Assumption BC.4, $\Lambda^{*\eta_n}_{n,\Delta,F_n} \to_H \Lambda^*_{\Delta\,\min}$. Hence, there is a sequence $\{(\theta^\dagger_n, e^\dagger_n, e^{\dagger *}_n, j^{\dagger *}_n) \in \Lambda^{*\eta_n}_{n,\Delta,F_n}\}_{n\geq 1}$ such that $d((\theta^\dagger_n, e^\dagger_n, e^{\dagger *}_n, j^{\dagger *}_n), (\theta^\dagger, e^\dagger, e^{\dagger *}, j^{\dagger *})) \to 0$. We obtain

$$\begin{aligned}
&\limsup_{n\to\infty} \widetilde{g}_n(\nu_n(\cdot), \mu_n(\cdot)) \\
:=\ & \limsup_{n\to\infty} \inf_{(\theta,e,e^*,j^*)\in\Lambda^{*\eta_n}_{n,\Delta,F_n}} \max_{j\leq k}\left(\tau_j(\nu_n(\cdot), \theta) + 1(j \neq j^*)e_j + 1(j = j^*)\varphi^\dagger(\mu_{nj^*}(\theta) + e^*_{j^*})\right) \\
\leq\ & \limsup_{n\to\infty} \max_{j\leq k}\left(\tau_j(\nu_n(\cdot), \theta^\dagger_n) + 1(j \neq j^{\dagger *}_n)e^\dagger_{nj} + 1(j = j^{\dagger *}_n)\varphi^\dagger(\mu_{nj^{\dagger *}_n}(\theta) + e^{\dagger *}_{j^{\dagger *}_n})\right) \\
=\ & \max_{j\leq k}\left(\tau_j(\nu(\cdot), \theta^\dagger) + 1(j \neq j^{\dagger *})e^\dagger_j + 1(j = j^{\dagger *})\varphi^\dagger(e^{\dagger *}_{j^{\dagger *}})\right) \\
=\ & \widetilde{g}(\nu(\cdot), 0_k(\cdot)),
\end{aligned} \tag{27.53}$$

where the inequality holds because $(\theta^\dagger_n, e^\dagger_n, e^{\dagger *}_n, j^{\dagger *}_n) \in \Lambda^{*\eta_n}_{n,\Delta,F_n}$ $\forall n \geq 1$, the second equality holds using $d((\theta^\dagger_n, e^\dagger_n, e^{\dagger *}_n, j^{\dagger *}_n), (\theta^\dagger, e^\dagger, e^{\dagger *}, j^{\dagger *})) \to 0$, (27.49) with $(\nu_n(\cdot), \theta^\dagger_n)$ and $(\nu(\cdot), \theta^\dagger)$ in place of $(\nu_{q_n}(\cdot), \overline{\theta}_{q_n})$ and $(\nu(\cdot), \overline{\theta})$, respectively, and (27.50) with $(\theta^\dagger_n, e^\dagger_{nj}, e^{\dagger *}_{nj}, j^{\dagger *}_n)$ and $(\theta^\dagger, e^\dagger_j, e^{\dagger *}_j, j^{\dagger *})$ in place of $(\overline{\theta}_{q_nj}, \overline{e}_{q_nj}, \overline{e}^*_{q_nj}, \overline{j}^*_{q_n})$ and $(\overline{\theta}_j, \overline{e}_j, \overline{e}^*_j, \overline{j}^*)$, respectively, and the last equality holds by (27.52). This establishes result (ii) in (27.46) and completes the proof of part (a).

The proof of part (b) is similar to that of part (a). But it is simpler because $A^*_{Ln,\Delta L}$ is simpler than $A^*_{Un,\Delta U}$. We have

$$\begin{aligned}
A^*_{Ln,\Delta L} &:= \inf_{\theta\in\Theta^{\eta_{Ln}}_{\min}(F_n)} \max_{j\leq k}\left(-\widehat{\nu}^*_{nj}(\theta) - \varphi^\dagger(-\xi^e_{1nj}(\theta))\right) = \overrightarrow{A}^*_{Ln,\Delta L} + o_p(1) \\
&\qquad \text{conditional on } \{W_{ni}\}_{i\leq n, n\geq 1} \text{ a.s.}[P_\bigtriangledown], \text{ where} \\
\overrightarrow{A}^*_{Ln,\Delta L} &:= \inf_{(\theta,e^*)\in\Lambda^{*\eta_{Ln}}_{n,F_n,L}} \max_{j\leq k}\left(-\nu^{m\sigma *}_{nj}(\theta) - \varphi^\dagger(-\mu_{nj}(\theta) - e^*_j)\right),
\end{aligned} \tag{27.54}$$

$e=(e_1,...,e_k)'$, $\nu_{nj}^{m\sigma *}(\theta)$ is defined in (27.37), the first equality holds by the definition in (27.12), and the second equality holds by (27.38), the definition of $\Lambda_{n,F_n,L}^{*\eta_{Ln}}$ in (27.11), (27.30), and the definition of $\mu_n(\theta)$ in (27.45). Given (27.54), to prove part (b), it suffices to show: for the subsequence $\{a_n\}_{n\geq 1}$ defined just above (27.36), $\{\overrightarrow{A}_{La_n,\Delta L}^*|\{W_{ni}\}_{i\leq n,n\geq 1}\}\to_d A_{L\infty,\Delta L}^*$ a.s.$[P_\bigtriangledown]$. The proof of this is analogous to the proof of (27.43), but with $\Lambda_{n,F_n,L}^{*\eta_{Ln}}$ and Λ_L^* in place of $\Lambda_{n,\Delta,F_n}^{*\eta_n}$ and $\Lambda_{\Delta\min}^*$, respectively, and with $-\varphi^\dagger(-\mu_{nj}(\theta)-e_j^*)$ in place of $1(j\neq j^*)e_j+1(j=j^*)\varphi^\dagger(\mu_{nj}(\theta)+e_j^*)$. The proof goes through using Assumption BC.6 in place of Assumption BC.4, which completes the proof of part (b). $\square$

27.6 Proof of Lemma 27.2

Proof of Lemma 27.2. First, we prove part (a). By definition, see (27.8) and (5.14) with b deleted, we have

$$A_{Un,\Delta U}^* := \inf_{\theta\in\Theta_{\min}^{\eta_n}(F_n)} \max_{j\leq k}\left(-\widehat{\nu}_{nj}^*(\theta)+1(j\neq j_{ne}(\theta))e_{nj}(\theta)+1(j=j_{ne}(\theta))\varphi^\dagger(\xi_{1nj}^e(\theta))\right) \text{ and}$$
$$A_{n,\Delta U}^* := \inf_{\theta\in\widehat{\Theta}_{\min,n}} \min_{j_1\in\widehat{J}_{ne}(\theta)} \max_{j\leq k}\left(-\widehat{\nu}_{nj}^*(\theta)+1(j\neq j_1)\widehat{e}_{nj}(\theta)+1(j=j_1)\varphi(\xi_{nj}^e(\theta))\right). \tag{27.55}$$

The bootstrap random variables $A_{Un,\Delta U}^*$ and $A_{n,\Delta U}^*$ differ in four ways. Specifically, $A_{Un,\Delta U}^*$ versus (vs.) $A_{n,\Delta U}^*$ are defined with (i) $\inf_{\theta\in\Theta_{\min}^{\eta_n}(F_n)}$ vs. $\inf_{\theta\in\widehat{\Theta}_{\min,n}}$, (ii) $\varphi^\dagger(\xi_{1nj}^e(\theta))$ vs. $\varphi(\xi_{nj}^e(\theta))$, (iii) $e_{nj}(\theta)$ vs. $\widehat{e}_{nj}(\theta)$, and (iv) $j=j_{ne}(\theta)$ or $j\neq j_{ne}(\theta)$ vs. $\min_{j_1\in\widehat{J}_{ne}(\theta)}$ with $j=j_1$ or $j\neq j_1$.

By Assumption BC.5, for any bootstrap random function $K_n^*(\theta)$,

$$P_\bigtriangledown\left(\inf_{\theta\in\Theta_{\min}^{\eta_n}(F_n)} K_n^*(\theta)\geq \inf_{\theta\in\widehat{\Theta}_{\min,n}} K_n^*(\theta)\middle|\{W_{ni}\}_{i\leq n,n\geq 1}\right)=1 \text{ wp}\to 1 \text{ under } P_\bigtriangledown. \tag{27.56}$$

By the definitions of $\xi_{1nj}^e(\theta)$ in (27.6) and $\xi_{nj}^e(\theta)$ in (5.12) and $\widehat{sd}_{nj}(\theta)\geq\iota$ (by construction; see (27.1)), we have $|\xi_{1nj}^e(\theta)|\geq|\xi_{nj}^e(\theta)|$ and $\xi_{1nj}^e(\theta)$ and $\xi_{nj}^e(\theta)$ have the same sign for all sample and bootstrap realizations. For any $\theta\in\Theta$, for all sample and bootstrap realizations with $\xi_{nj}^e(\theta)\geq 0$, we have

$$\varphi(\xi_{nj}^e(\theta))\leq\varphi^\dagger(\xi_{nj}^e(\theta))\leq\varphi^\dagger(\xi_{1nj}^e(\theta)), \tag{27.57}$$

where the first inequality holds by property (i) of $\varphi^\dagger$ stated following (27.7) and the second inequality holds by property (ii) of $\varphi^\dagger$ stated following (27.7) and $\xi_{nj}^e(\theta)\leq\xi_{1nj}^e(\theta)$. Next, for all sample and bootstrap realizations with $\xi^e_{nj}(\theta) < 0$, we have $\xi^e_{1nj}(\theta) < 0$ and this implies that

$$\varphi(\xi^e_{nj}(\theta)) \leq \varphi^\dagger(\xi^e_{nj}(\theta)) = 0 = \varphi^\dagger(\xi^e_{1nj}(\theta)), \tag{27.58}$$

where the first inequality holds by property (i) of $\varphi^\dagger$, the first equality holds by property (iii) of $\varphi^\dagger$ and $\xi^e_{nj}(\theta) < 0$, and the second equality holds by property (iii) of $\varphi^\dagger$ and $\xi^e_{1nj}(\theta) < 0$. Hence, $\varphi(\xi^e_{nj}(\theta)) \leq \varphi^\dagger(\xi^e_{1nj}(\theta))$ for all sample and bootstrap realizations, for all $\theta \in \Theta$.

Next, we have

$$\begin{aligned} &\widehat{e}_{nj}(\theta) := n^{1/2}\left(\widehat{\Delta}_{nj}(\theta) - \widehat{\Delta}^{\inf}_n\right) - \widehat{sd}_{nj}(\theta)\kappa_n = \widehat{d}_{nj}(\theta) + e_{nj}(\theta) - \widehat{sd}_{nj}(\theta)\kappa_n, \text{ and so,} \\ &\sup_{\theta\in\Theta}\left(\widehat{e}_{nj}(\theta) - e_{nj}(\theta)\right) \leq \sup_{\theta\in\Theta}\left(\widehat{d}_{nj}(\theta) - \iota\kappa_n\right) \to_p -\infty\ \forall j \leq k, \end{aligned} \tag{27.59}$$

where the inequality on the second line holds for all bootstrap realizations because $\widehat{d}_{nj}(\theta)$ (defined in (27.30)) does not depend on any bootstrap quantities, the first equality on the first line holds by definition, see (5.8), the second equality holds by (27.30), and the second line follows from the first line, the last line of (27.30), $\widehat{sd}_{nj}(\theta) \geq \iota$ by definition, and $\kappa_n \to \infty$ (by Assumption A.6(i)). Equation (27.59) implies that

$$\sup_{\theta\in\Theta}\left(\widehat{e}_{nj}(\theta) - e_{nj}(\theta)\right) \leq 0\ \forall j \leq k, \text{ for all bootstrap realizations, wp} \to 1 \text{ under } P_\triangledown. \tag{27.60}$$

Define

$$\overline{A}^*_{Un,\Delta U} := \inf_{\theta\in\Theta^{\eta_n}_{\min}(F_n)} \min_{j_1\in\widehat{J}_{ne}(\theta)} \max_{j\leq k}\left(-\widehat{\nu}^*_{nj}(\theta) + 1(j \neq j_1)e_{nj}(\theta) + 1(j = j_1)\varphi^\dagger(\xi^e_{1nj}(\theta))\right). \tag{27.61}$$

Combining (27.56)–(27.60) and (27.61) gives

$$P_\triangledown(\overline{A}^*_{Un,\Delta U} \geq A^*_{n,\Delta U}|\{W_{ni}\}_{i\leq n,n\geq 1}) = 1 \text{ wp} \to 1 \text{ under } P_\triangledown. \tag{27.62}$$

Next, we show that

$$P_\triangledown(j_{ne}(\theta) \in \widehat{J}_{ne}(\theta)\ \forall\theta \in \Theta|\{W_{ni}\}_{i\leq n,n\geq 1}) = 1 \text{ wp} \to 1 \text{ under } P_\triangledown, \tag{27.63}$$

where $j_{ne}(\theta) := \arg\max_{j\leq k} e_{nj}(\theta)$ is defined in (27.3) and $\widehat{J}_{ne}(\theta) := \{j \in \{1, ..., k\} : \widehat{\Delta}_{nj}(\theta) \geq \widehat{\Delta}_n(\theta) - \widehat{sd}_{nj}(\theta)n^{-1/2}\kappa_n\}$ is defined in (5.13). We have $j_{ne}(\theta) \in \widehat{J}_{ne}(\theta)$ iff $\widehat{\Delta}_{nj_{ne}(\theta)}(\theta) \geq \widehat{\Delta}_n(\theta) - \widehat{sd}_{nj}(\theta)n^{-1/2}\kappa_n$ if $n^{1/2}(\widehat{\Delta}_{nj_{ne}(\theta)}(\theta) - \widehat{\Delta}^{\inf}_n) - n^{1/2}(\widehat{\Delta}_n(\theta) - \widehat{\Delta}^{\inf}_n) \geq -\iota\kappa_n$ because $\widehat{sd}_{nj}(\theta) \geq \iota$ by definition. By (27.30), $n^{1/2}(\widehat{\Delta}_{nj}(\theta) - \widehat{\Delta}^{\inf}_n) = e_{nj}(\theta) + O^\Theta_p(1)\ \forall j \leq k$ (since $\widehat{\Delta}_{nj}(\theta) = -\widehat{m}_{nj}(\theta)$ by

(5.2)). Hence, $n^{1/2}(\max_{j\leq k}\widehat{\Delta}_{nj}(\theta)-\widehat{\Delta}_n^{\inf}) = \max_{j\leq k} e_{nj}(\theta)+O_p^\Theta(1)$. Taking $j = j_{ne}(\theta)$, these results combine to give $n^{1/2}(\widehat{\Delta}_{nj_{ne}(\theta)}(\theta)-\widehat{\Delta}_n^{\inf})-n^{1/2}(\widehat{\Delta}_n(\theta)-\widehat{\Delta}_n^{\inf}) = e_{nj_{ne}(\theta)}(\theta)-\max_{j\leq k} e_{nj}(\theta)+O_p^\Theta(1) = O_p^\Theta(1)$ using the definition of $j_{ne}(\theta)$, where the $O_p^\Theta(1)$ term does not depend on any bootstrap quantities. Since $O_p^\Theta(1) \geq -\iota\kappa_n$ holds wp→1 using Assumption A.6(i) (i.e., $\kappa_n \to \infty$), (27.63) is proved.

For a suitably defined random function $w(j_1,\theta)$ on $\{1,...,k\}\times\Theta$, $A^*_{Un,\Delta U}$ and $\overline{A}^*_{Un,\Delta U}$ can be written as $\inf_{\theta\in\Theta^{\eta_n}_{\min}(F_n)} w(j_{ne}(\theta),\theta)$ and $\inf_{\theta\in\Theta^{\eta_n}_{\min}(F_n)}\min_{j_1\in\widehat{J}_{ne}(\theta)} w(j_1,\theta)$, respectively. Since $w(j_{ne}(\theta),\theta) \geq \min_{j_1\in\widehat{J}_{ne}(\theta)} w(j_1,\theta)$ when $j_{ne}(\theta)\in\widehat{J}_{ne}(\theta)$ and the latter event satisfies (27.63), we obtain

$$P_\triangledown(A^*_{Un,\Delta U} \geq \overline{A}^*_{Un,\Delta U}|\{W_{ni}\}_{i\leq n,n\geq 1}) = 1 \text{ wp} \to 1 \text{ under } P_\triangledown. \tag{27.64}$$

This and (27.62) establish the result of part (a) of the lemma. Note that Assumptions A.6(ii) and BC.4 are imposed in the lemma because Assumption BC.5 is imposed and it relies on Assumptions A.6(ii) and BC.4.

Next, we prove part (b) of the lemma. By definition, see (27.12) and (5.15),

$$\begin{aligned} A^*_{Ln,\Delta L} &:= \inf_{\theta\in\Theta^{\eta_{Ln}}_{\min}(F_n)} \max_{j\leq k}\left(-\widehat{\nu}^*_{nj}(\theta) - \varphi^\dagger(-\xi^e_{1nj}(\theta))\right) \text{ and} \\ A^*_{n,\Delta L} &:= \inf_{\theta\in\widehat{\Theta}_{\min,L,n}} \max_{j\leq k}\left(-\widehat{\nu}^*_{nj}(\theta) - \varphi(-\xi^e_{nj}(\theta))\right). \end{aligned} \tag{27.65}$$

By an analogous argument to that used to obtain (27.57) and (27.58), we get

$$\varphi(-\xi^e_{nj}(\theta)) \leq \varphi^\dagger(-\xi^e_{1nj}(\theta)) \tag{27.66}$$

for all $j\leq k$, for all sample and bootstrap realizations, and for all $\theta\in\Theta$. By Assumption BC.7, for any bootstrap random function $K^*_n(\theta)$,

$$P_\triangledown\left(\inf_{\theta\in\Theta^{\eta_{Ln}}_{\min}(F_n)} K^*_n(\theta) \leq \inf_{\theta\in\widehat{\Theta}_{\min,L,n}} K^*_n(\theta)\middle| \{W_{ni}\}_{i\leq n,n\geq 1}\right) = 1 \text{ wp} \to 1 \text{ under } P_\triangledown. \tag{27.67}$$

Combining (27.65), (27.66), and (27.67) with $K^*_n(\theta) = \max_{j\leq k}(-\widehat{\nu}^*_{nj}(\theta) - \varphi^\dagger(-\xi^e_{1nj}(\theta)))$ gives the result of part (b) of the lemma. □

27.7 Proof of Lemma 27.3

The proof of Lemma 27.3 uses the following lemma, which is based on Lemma S.3.7 in the Supplemental Material to BCS17. Let $R_{[-\infty]} := R\cup\{-\infty\}$.

Lemma 27.6 *Suppose Assumptions* A.0, A.7, A.8, BC.6, *and* C.12 *hold. Then, for any* $(\theta^*, e^*) \in \Lambda_L^*$, *there exists* $(\theta^*, e) \in \Lambda_\Delta$ *such that* $e_j \leq -\varphi^\dagger(-e_j^*)$ *for all* $j \leq k$.

Proof of Lemma 27.3. Part (a) of the lemma requires that $A_{\infty,\Delta\min}$, $A_{\infty,\Delta}$, and $A^*_{U\infty,\Delta U}$ are well-defined. Part (b) requires that $A_{\infty,\Delta}$ and $A^*_{L\infty,\Delta L}$ are well-defined. Each part of Lemma 27.3 imposes the assumptions such that these quantities are well defined.

First, we prove part (a). We can write $A^*_{U\infty,\Delta U} = \inf_{(\theta,e,e^*,j^*)\in\Lambda^*_{\Delta\min}} K_U(\theta, e, e^*, j^*)$ and $A_{\infty,\Delta\min})$ $= \inf_{(\theta,e)\in\Lambda_{\Delta\min}} K(\theta, e)$ for random functions $K_U(\cdot)$ and $K(\cdot)$ defined in (27.69) below. To show $A^*_{U\infty,\Delta U} \leq A_{\infty,\Delta\min}$, it suffices to show that for any $(\theta, e) \in \Lambda_{\Delta\min}$ there exists $(\theta, e, e^*, j^*) \in \Lambda^*_{\Delta\min}$ for which $K_U(\theta, e, e^*, j^*) \leq K(\theta, e)$ for all sample realizations.

To this end, we claim: Given any $(\theta, e) \in \Lambda_{\Delta\min}$, there exists an element $(\theta, e, e^*, j^*) \in \Lambda^*_{\Delta\min}$.

This claim is proved as follows. By Assumption C.13, given any $(\theta, e) \in \Lambda_{\Delta\min}$, there exists a sequence $\{(\overline{\theta}_n, \overline{e}_n) \in \Lambda^{\eta_n}_{n,\Delta,F_n}\}_{n\geq1}$ such that $d((\overline{\theta}_n, \overline{e}_n), (\theta, e)) \to 0$, where $\overline{\theta}_n \in \Theta^{\eta_n}_{\min}(F_n)$ for all $n \geq 1$ by the definition of $\Lambda^{\eta_n}_{n,\Delta,F_n}$ following (25.4) and $\Theta^\eta_{\min}(F)$ is non-empty by Assumption A.0. Given $\{(\overline{\theta}_n, \overline{e}_n)\}_{n\geq1}$, consider the corresponding sequence $\{(\overline{\theta}_n, \overline{e}_n, e^*_n, j^*_n) \in \Lambda^{*\eta_n}_{n,\Delta,F_n}\}_{n\geq1}$ for $\Lambda^{*\eta_n}_{n,\Delta,F_n}$ defined in (27.5), where $e^*_{nj} := (\iota\kappa_n)^{-1}\overline{e}_{nj}$, $j^*_n := \arg\max_{j\leq k} \overline{e}_{nj}$, and j^*_n is the smallest arg max value if the arg max is not unique. By Assumption BC.4, $\Lambda^{*\eta_n}_{n,\Delta,F_n} \to_H \Lambda^*_{\Delta\min}$ for $\Lambda^*_{\Delta\min}$ compact (under d). In consequence, there exist a subsequence $\{u_n\}_{n\geq1}$ of $\{n\}_{n\geq1}$ and an element $(\overline{\theta}, \overline{e}, e^*, j^*)$ of $\Lambda^*_{\Delta\min}$ for which

$$d((\overline{\theta}_{u_n}, \overline{e}_{u_n}, e^*_{u_n}, j^*_{u_n}), (\overline{\theta}, \overline{e}, e^*, j^*)) \to 0 \text{ and } (\overline{\theta}, \overline{e}) = (\theta, e), \tag{27.68}$$

where the equality holds because $d((\overline{\theta}_n, \overline{e}_n), (\theta, e)) \to 0$, which completes the proof of the claim.

Given any $(\theta, e) \in \Lambda_{\Delta\min}$, take $(\theta, e, e^*, j^*) \in \Lambda^*_{\Delta\min}$ as in the previous paragraph. Then, we have

$$\begin{aligned} K_U(\theta, e, e^*, j^*) &:= \max_{j\leq k}\left(-G^{m\sigma}_j(\theta) + 1(j \neq j^*)e_j + 1(j = j^*)\varphi^\dagger(e^*_{j^*})\right) \\ &\leq \max_{j\leq k}\left(-G^{m\alpha}_j(\theta) + e_j\right) := K(\theta, e) \end{aligned} \tag{27.69}$$

for all sample realizations, where $G^{m\sigma}_j(\theta)$ is defined in (27.9) based on quantities defined in Assumptions C.4 and C.5, the first and last equalities hold by the definitions of $A^*_{U\infty,\Delta U}$ and $A_{\infty,\Delta\min}$ and the inequality holds because, as we show below, $\varphi^\dagger(e^*_{j^*}) \leq e_{j^*}$. As argued above, (27.69) implies that $A^*_{U\infty,\Delta U} \leq A_{\infty,\Delta\min}$ for all sample realizations, which we set out to prove.

Now, we show $\varphi^\dagger(e^*_{j^*}) \leq e_{j^*}$. For notational simplicity, suppose (27.68) holds with n in place of u_n. We have $j^*_n \to j^*$ by (27.68), and hence, $j^*_n = j^*$ for n large (because $j^*_n \in \{1, ..., k\}$), where $j_n^* := j_n(\overline{\theta}_n)$ by the definition of $\Lambda^{*\eta_n}_{n,\Delta,F_n}$ in (27.5) for $j_n(\overline{\theta}_n)$ defined in (27.3). We have $\overline{e}_{nj} \to e_j$ and $e_{nj}^* \to e_j^*$ by (27.68), where $\overline{e}_{nj} = e_{nj}(\overline{\theta}_n)$ and $e^*_{nj} = (\iota\kappa_n)^{-1}\overline{e}_{nj}$ by the definition of $\Lambda^{*\eta_n}_{n,\Delta,F_n}$ for $e_{nj}(\theta) \geq 0$ defined in (27.3). Hence, we have $\overline{e}_{nj_n^*} \to e_{j^*}$ and $e^*_{nj_n^*} \to e^*_{j^*}$, where $e^*_{nj_n^*} = (\iota\kappa_n)^{-1}\overline{e}_{nj_n^*} = (\iota\kappa_n)^{-1}e_{nj_n(\overline{\theta}_n)}(\overline{\theta}_n) \geq 0$ for all $n \geq 1$ by (27.3). This and $\kappa_n \to \infty$ (by Assumption A.6(i)) imply that $e_{j^*} \geq e^*_{j^*} \geq 0$. In addition, it implies that if $0 \leq e_{j^*} < \infty$, then $e^*_{j^*} = 0$ (since $\kappa_n \to \infty$). Hence, we obtain: if $0 \leq e_{j^*} < \infty$, then $\varphi^\dagger(e^*_{j^*}) = 0 \leq e_{j^*}$ because $\varphi^\dagger(0) = 0$ by property (iii) of $\varphi^\dagger$ stated following (27.7). On the other hand, if $e_{j^*} = \infty$, then $\varphi^\dagger(e^*_{j^*}) \leq \infty = e^*_{j^*}$ by the definition of $\varphi^\dagger$ in (27.7), which completes the proof of part (a) of the lemma.

Next, we prove part (b) of the lemma. By Lemma 27.6(a), for any $(\theta^*, e^*) \in \Lambda_L^*$, there exists $(\theta^*, e) \in \Lambda_\Delta$ such that $e_j \leq -\varphi^\dagger(-e_j^*)$ for all $j \leq k$. In consequence, we have

$$\inf_{(\overline{\theta},\overline{e})\in\Lambda_\Delta} \max_{j\leq k}\left(-G_j^{m\sigma}(\overline{\theta}) + \overline{e}_j\right) \leq \max_{j\leq k}\left(-G_j^{m\sigma}(\theta^*) + e_j\right) \leq \max_{j\leq k}\left(-G_j^{m\sigma}(\theta^*) - \varphi^\dagger(-e_j^*)\right), \quad (27.70)$$

where the first inequality holds because $(\theta^*, e) \in \Lambda_\Delta$ and the second inequality holds by Lemma 27.6(a). Deleting the middle expression in (27.70) and taking the infimum over $(\theta^*, e^*) \in \Lambda_L^*$ on the rhs of (27.70) gives

$$A_{\infty,\Delta} := \inf_{(\overline{\theta},\overline{e})\in\Lambda_\Delta} \max_{j\leq k}\left(-G_j^{m\sigma}(\overline{\theta}) + \overline{e}_j\right) \leq \inf_{(\theta^*,e^*)\in\Lambda_L^*} \max_{j\leq k}\left(-G_j^{m\sigma}(\theta^*) - \varphi^\dagger(-e_j^*)\right) =: A^*_{L\infty,\Delta L}, \quad (27.71)$$

where the two equalities hold by the definitions in (25.5) and (27.13). This completes the proof of part (b). $\square$

Proof of Lemma 27.6. For any $(\theta^*, e^*) \in \Lambda_L^*$, there exist a subsequence $\{a_n\}_{n\geq1}$ of $\{p_n\}_{n\geq1}$ and a sequence $\{(\theta^*_{a_n}, e^*_{a_n}) \in \Lambda^{*\eta_{Ln}}_{n,F_n,L}\}_{n\geq1}$ for which $\theta^*_{a_n} \in \Theta^{\eta_{La_n}}_{\min}(F_{a_n})$ (where $\Theta^{\eta_{La_n}}_{\min}(F_{a_n})$ is non-empty by Assumption A.0), $e^*_{a_nj} = \kappa_{a_n}^{-1}a_n^{1/2}(\Delta_{F_{a_n}j}(\theta^*_{a_n}) - \Delta^{\inf}_{F_{a_n}})$, $\lim \theta^*_{a_n} = \theta^*$, and $\lim e^*_{a_nj} = e^*_j$ for all $j \leq k$ using Assumption BC.6 and the definition of $\Lambda^{*\eta_{Ln}}_{n,F_n,L}$ in (27.11).

For notational simplicity, in the remainder of the proof of part (a) we employ n in place of a_n for $n \geq 1$ and assume the assumptions hold for $\{n\}_{n\geq1}$, rather than $\{p_n\}_{n\geq1}$. Thus, the sequence $\{(\theta^*_n, e^*_n)\}_{n\geq1}$ satisfies $\theta^*_n \in \Theta^{\eta_{Ln}}_{\min}(F_n)$, $e^*_{nj} = (\iota\kappa_n)^{-1}n^{1/2}(\Delta_{F_nj}(\theta^*_n) - \Delta^{\inf}_{F_n})$, $\lim \theta^*_n = \theta^*$, and $\lim e^*_{nj} = e^*_j$ for $j \leq k$.

Define

$$\Delta_F^{vec}(\theta) = (\Delta_{F1}(\theta) - \Delta_F^{\inf}, ..., \Delta_{Fk}(\theta) - \Delta_F^{\inf})'. \quad (27.72)$$

Note that $\Delta_F^{vec}(\theta)$ is the vector of differences $\Delta_{Fj}(\theta) - \Delta_F^{\inf}$ for $j \leq k$, not the vector of $\Delta_{Fj}(\theta)$ for $j \leq k$. For $\theta \in \Theta_{\min}(F)$, $\max_{j\leq k} \Delta_{Fj}(\theta) =: \Delta_F(\theta) = \Delta_F^{\inf}$ by the definition of $\Theta_{\min}(F)$, see (25.4), and so, $\Delta_F^{vec}(\theta) \leq 0_k$ element by element.

Note that for (θ_n^*, e_n^*) specified above, $e_n^* = (\iota\kappa_n)^{-1} n^{1/2} \Delta_{F_n}^{vec}(\theta_n^*) \to e^*$.

By the definition of $\Theta_{\min}^{\eta}(F)$ following (25.4), $\theta_n^* \in \Theta_{\min}^{\eta_{Ln}}(F_n)$ implies that $\Delta_F(\theta_n^*) - \Delta_{F_n}^{\inf} \leq \eta_{Ln}/n^{1/2}$ and so

$$\min\{\delta, \inf_{\theta \in \Theta_{\min}(F_n)} ||\theta_n^* - \theta||\} \leq \Delta_{F_n}(\theta_n^*) - \Delta_{F_n}^{\inf} \leq \eta_{Ln}/n^{1/2} \to 0, \tag{27.73}$$

where the first inequality holds by Assumption A.8(i) and the convergence holds by Assumption BC.6. Hence, $||\theta_n^* - \widetilde{\theta}_n|| = O(\eta_{Ln}/n^{1/2})$ for some sequence $\{\widetilde{\theta}_n \in \Theta_{\min}(F_n)\}_{n \geq 1}$.

Let $\gamma \in (0,1)$ be as in Assumption BC.6. By convexity of Θ and Assumption A.8(iii), element-by-element mean value expansions give

$$\kappa_n^{-\gamma} n^{1/2} \Delta_{F_n}^{vec}(\theta_n^*) = \kappa_n^{-\gamma} n^{1/2} \Delta_{F_n}^{vec}(\widetilde{\theta}_n) + \frac{\partial}{\partial \theta'} \Delta_{F_n}^{vec}(\theta_n^0) \cdot \kappa_n^{-\gamma} n^{1/2} (\theta_n^* - \widetilde{\theta}_n), \tag{27.74}$$

where the jth row of $\frac{\partial}{\partial \theta} \Delta_{F_n}^{vec}(\theta_n^0)$ $(:= -\widetilde{M}_{F_n}(\theta_n^0)$ by Assumption A.8(iii)) is evaluated at some $\theta_{nj}^0 \in \Theta$ that is on the line segment between θ_n^* and $\widetilde{\theta}_n$ and $\Delta_{F_n}^{vec}(\theta)$ $(:= -E_{F_n} \widetilde{m}(W, \theta) - \Delta_{F_n}^{\inf})$ is partially differentiable by Assumption A.8(iii).

Define

$$\theta_n^\dagger := (1 - \kappa_n^{-\gamma}) \widetilde{\theta}_n + \kappa_n^{-\gamma} \theta_n^*, \text{ or equivalently, } \theta_n^\dagger - \widetilde{\theta}_n := \kappa_n^{-\gamma} (\theta_n^* - \widetilde{\theta}_n), \tag{27.75}$$

where $\theta_n^\dagger \in \Theta$ for n large by convexity of Θ (Assumption A.8(ii)) and $\kappa_n^{-\gamma} \to 0$ (Assumption A.7(i)). By $||\theta_n^* - \widetilde{\theta}_n|| = O(\eta_{Ln}/n^{1/2})$ above, we have $\theta_n^* - \widetilde{\theta}_n \to 0_k$, $n^{1/2}(\theta_n^\dagger - \widetilde{\theta}_n) = \kappa_n^{-\gamma} n^{1/2} (\theta_n^* - \widetilde{\theta}_n) = O(\eta_{Ln}/\kappa_n^{\gamma}) \to 0_k$, where the convergence holds by Assumption BC.6, and $\theta_{nj}^0 - \widetilde{\theta}_n \to 0_k$ for $j \leq k$.

Equation (27.74) can be written as

$$\frac{\partial}{\partial \theta} \Delta_F^{vec}(\theta_n^0) \cdot n^{1/2} (\theta_n^\dagger - \widetilde{\theta}_n) = \kappa_n^{-\gamma} n^{1/2} \Delta_F^{vec}(\theta_n^*) - \kappa_n^{-\gamma} n^{1/2} \Delta_F^{vec}(\widetilde{\theta}_n) \tag{27.76}$$

using the second equality in (27.75).

Applying element-by-element mean value expansions again yields

$$\begin{aligned}
n^{1/2}\Delta_{F_n}^{vec}(\theta_n^\dagger) &= n^{1/2}\Delta_{F_n}^{vec}(\widetilde{\theta}_n) + \frac{\partial}{\partial\theta'}\Delta_{F_n}^{vec}(\theta_n^{00}) \cdot n^{1/2}(\theta_n^\dagger - \widetilde{\theta}_n) \\
&= n^{1/2}\Delta_{F_n}^{vec}(\widetilde{\theta}_n) + \frac{\partial}{\partial\theta'}\Delta_{F_n}^{vec}(\theta_n^{0}) \cdot n^{1/2}(\theta_n^\dagger - \widetilde{\theta}_n) + \varepsilon_{1n}, \\
&= n^{1/2}\Delta_{F_n}^{vec}(\widetilde{\theta}_n) + \kappa_n^{-\gamma}n^{1/2}\Delta_F^{vec}(\theta_n^*) - \kappa_n^{-\gamma}n^{1/2}\Delta_F^{vec}(\widetilde{\theta}_n) + \varepsilon_{1n}, \\
&= \kappa_n^{-\gamma}n^{1/2}\Delta_F^{vec}(\theta_n^*) + \varepsilon_{1n} + \varepsilon_{2n}, \text{ where} \\
\varepsilon_{1n} &:= \left(\frac{\partial}{\partial\theta'}\Delta_{F_n}^{vec}(\theta_n^{00}) - \frac{\partial}{\partial\theta'}\Delta_{F_n}^{vec}(\theta_n^{0})\right) \cdot n^{1/2}(\theta_n^\dagger - \widetilde{\theta}_n) \to 0_k, \\
\varepsilon_{2n} &:= (1-\kappa_n^{-\gamma})n^{1/2}\Delta_{F_n}^{vec}(\widetilde{\theta}_n) \le 0_k, \qquad (27.77)
\end{aligned}$$

the jth row of $\frac{\partial}{\partial\theta'}\Delta_{F_n}^{vec}(\theta_n^{00})$ is evaluated at some $\theta_{nj}^{00} \in \Theta$ that is on the line segment between $\theta_n^\dagger$ and $\widetilde{\theta}_n$ and satisfies $||\theta_{nj}^{00} - \widetilde{\theta}_n|| \le ||\theta_{nj}^\dagger - \widetilde{\theta}_n|| \to 0$ for $j \le k$, the third equality uses (27.76), the convergence of ε_{1n} holds by the result above that $n^{1/2}(\theta_n^\dagger - \widetilde{\theta}_n) \to 0_k$ and Assumption A.8(iii), and the inequality for ε_{2n} holds element by element for n large because $1-\kappa_n^{-\gamma} \to 1$ and $\Delta_{F_n}^{vec}(\widetilde{\theta}_n) \le 0_k$ using the result following (27.72) because $\widetilde{\theta}_n \in \Theta_{\min}(F_n)$.

Because $(R_{[\pm\infty]}^k, d)$ is compact, there exists a subsequence $\{u_n\}_{n\ge1}$ of $\{n\}_{n\ge1}$, for which $u_n^{1/2}\Delta_{F_{u_n}}^{vec}(\theta_{u_n}^\dagger)$ and $\kappa_{u_n}^{-\gamma}u_n^{1/2}\Delta_{F_{u_n}}^{vec}(\theta_{u_n}^*)$ converge as $n \to \infty$. This, (27.77), and the properties of ε_{1n} and ε_{2n} give

$$\lim u_n^{1/2}\Delta_{F_{u_n}}^{vec}(\theta_{u_n}^\dagger) \le \lim \kappa_{u_n}^{-\gamma}u_n^{1/2}\Delta_{F_{u_n}}^{vec}(\theta_{u_n}^*) \le 0_k \qquad (27.78)$$

element by element, where the second inequality holds because $\Delta_{F_n}(\theta_n^*) - \Delta_{F_n}^{\inf} \le \eta_{Ln}/n^{1/2}$ by (27.73), which implies that $\kappa_{u_n}^{-\gamma}u_n^{1/2}\Delta_{F_{u_n}}^{vec}(\theta_{u_n}^*) \le (\kappa_{u_n}^{-\gamma}\eta_{Lu_n})1_k \to 0_k$ element by element using $\eta_{Ln}/\kappa_n^\gamma \to 0$ by Assumption BC.6.

We have $\theta_n^\dagger \to \theta^*$, because $\theta_n^* \to \theta^*$ by the second paragraph of the proof, $||\theta_n^* - \widetilde{\theta}_n|| = O(\eta_{Ln}/n^{1/2}) = o(1)$ by results following (27.73), and $\theta_n^\dagger - \widetilde{\theta}_n \to 0_k$ by results following (27.75). Define $e_n^\dagger := n^{1/2}\Delta_{F_n}^{vec}(\theta_n^\dagger)$. We have $(\theta_n^\dagger, e_n^\dagger) \in \Lambda_{n,\Delta,F_n}$, $\theta_n^\dagger \to \theta^*$, and $e_{u_n}^\dagger \to e := \lim u_n^{1/2}\Delta_{F_{u_n}}^{vec}(\theta_{u_n}^\dagger)$ $(\le 0_k)$ by (27.78), and so, $(\theta^*, e) \in \Lambda_\Delta$ using $\Lambda_{n,\Delta,F_n} \to_H \Lambda_\Delta$ by Assumption C.12.

We have $(\theta^*, e^*) \in \Lambda_L^*$ by assumption, see the first paragraph of the proof, and $e_n^* = (\iota\kappa_n)^{-1}n^{1/2}\Delta_{F_n}^{vec}(\theta_n^*) \to e^*$ by the result in the paragraph following (27.72). The second inequality in (27.78) can be written as $\lim \kappa_{u_n}^{1-\gamma}e_{u_n}^* \le 0_k$, where $\kappa_{u_n}^{1-\gamma} \to \infty$. This implies that $e^* = \lim e_{u_n}^* \le 0_k$.

It remains to show that $e_j \le -\varphi^\dagger(-e_j^*)$ for $j \le k$. Suppose $e_j^* < 0$. Then,

$$e_j := \lim u_n^{1/2}(\Delta_{F_{u_n}j}(\theta_{u_n}^\dagger) - \Delta_{F_{u_n}}^{\inf}) \le \lim \kappa_{u_n}^{-\gamma}u_n^{1/2}(\Delta_{F_{u_n}j}(\theta_{u_n}^*) - \Delta_{F_{u_n}}^{\inf}) = \lim \kappa_{u_n}^{1-\gamma}e_{u_nj}^* = -\infty, \qquad (27.79)$$

where the inequality holds by the first inequality in (27.78) and the last equality holds because

$e^*_{u_nj} \to e^*_j < 0$ and $\kappa_{u_n}^{1-\gamma} \to \infty$. Hence, $-\infty = e_j \le -\varphi^\dagger(-e^*_j)$. Alternatively, suppose $e^*_j = 0$. Then, $e_j \le 0 = -\varphi^\dagger(-e^*_j)$, where the inequality holds by (27.78) since its left-hand side vector equals e and the equality holds by $\varphi^\dagger(0) = 0$ by property (iii) of $\varphi^\dagger$ stated following (27.7). This completes the proof. □

27.8 Proof of Lemma 27.5

Proof of Lemma 27.5. First, we prove part (a). We have

$$P_{F_n}(\widehat{\Theta}_{\min,n} \supseteq \Theta^{\eta_n}_{\min}(F_n)) \ge P_{F_n}\left(\sup_{\theta\in\Theta^{\eta_n}_{\min}(F_n)} n^{1/2}(\widehat{\Delta}_n(\theta) - \widehat{\Delta}^{\inf}_n)_- \le \tau_n\right) \tag{27.80}$$

by the definition of $\widehat{\Theta}_{\min,n}$ in (5.7).

Next, we have

$$\begin{aligned}
&\sup_{\theta\in\Theta^{\eta_n}_{\min}(F_n)} n^{1/2}(\widehat{\Delta}_n(\theta) - \widehat{\Delta}^{\inf}_n) \\
=& \sup_{\theta\in\Theta^{\eta_n}_{\min}(F_n)} \max_{j\le k}\left(n^{1/2}(\widehat{\Delta}_{nj}(\theta) - \Delta_{F_nj}(\theta)) + n^{1/2}(\Delta_{F_nj}(\theta) - \Delta^{\inf}_{F_n}) + n^{1/2}(\Delta^{\inf}_{F_n} - \widehat{\Delta}^{\inf}_n)\right) \\
\le& \sup_{\theta\in\Theta^{\eta_n}_{\min}(F_n)} \max_{j\le k} n^{1/2}|\widehat{\Delta}_{nj}(\theta) - \Delta_{F_nj}(\theta)| + \eta_n + O_p(1),
\end{aligned} \tag{27.81}$$

where the inequality holds by the definition of $\Theta^{\eta_n}_{\min}(F_n)$ and Theorem 25.2(a) and (b) (which requires Assumptions A.0, C.4, C.5, and C.12).

By (26.6), (26.11), and (26.12), for all $\theta \in \Theta$ and $j \le k$,

$$\begin{aligned}
n^{1/2}\left(\widehat{\Delta}_{nj}(\theta) - \Delta_{F_nj}(\theta)\right) &= \frac{\sigma_{Fj}(\theta)}{\widehat{\sigma}_{nj}(\theta)}\widehat{K}_{1nj}(\theta, F_n) + \frac{\sigma_{Fj}(\theta)}{\widehat{\sigma}_{nj}(\theta)}\widehat{K}_{2nj}(\theta, F_n) \\
&= -\nu^m_{nj}(\theta) + \frac{1}{2}\widetilde{m}_j(\theta)\cdot\nu^\sigma_{nj}(\theta) + o^\Theta_p(1).
\end{aligned} \tag{27.82}$$

This and Assumptions C.4 and C.5 imply that

$$\sup_{\theta\in\Theta^{\eta_n}_{\min}(F_n)} \max_{j\le k} n^{1/2}|\widehat{\Delta}_{nj}(\theta) - \Delta_{F_nj}(\theta)| = O_p(1). \tag{27.83}$$

Equations (27.81) and (27.83) combine to give

$$\sup_{\theta\in\Theta^{\eta_n}_{\min}(F_n)} n^{1/2}(\widehat{\Delta}_n(\theta) - \widehat{\Delta}^{\inf}_n) \le O_p(1) + \eta_n. \tag{27.84}$$

It follows that

$$\begin{aligned}
&P_{F_n}\left(\sup_{\theta\in\Theta_{\min}^{\eta_n}(F_n)} n^{1/2}(\widehat{\Delta}_n(\theta)-\widehat{\Delta}_n^{\inf})\le\tau_n\right)\\
&\ge P_{F_n}(O_p(1)+\eta_n\le\tau_n)\\
&= P_{F_n}(O_p(1/\tau_n)+\eta_n/\tau_n\le 1)\\
&\to 1, \qquad (27.85)
\end{aligned}$$

where the convergence holds because $\tau_n\to\infty$ and $\eta_n/\tau_n\to 0$. Combining this with (27.80) establishes part (a).

Next, we prove part (b). We have

$$P_{F_n}(\Theta_{\min}^{\eta_{Ln}}(F_n)\supseteq\widehat{\Theta}_{\min,L,n})\ge P_{F_n}\left(\sup_{\theta\in\widehat{\Theta}_{\min,n}} n^{1/2}(\Delta_{F_n}(\theta)-\Delta_{F_n}^{\inf})_-\le\eta_{Ln}\right) \qquad (27.86)$$

by the definition of $\Theta_{\min}^{\eta_{Ln}}(F_n)$ following (25.4). Next, we have

$$\begin{aligned}
&\sup_{\theta\in\widehat{\Theta}_{\min,L,n}} n^{1/2}(\Delta_{F_n}(\theta)-\Delta_{F_n}^{\inf})\\
&=\sup_{\theta\in\widehat{\Theta}_{\min,L,n}}\max_{j\le k}\left(n^{1/2}(\Delta_{F_nj}(\theta)-\widehat{\Delta}_{nj}(\theta))+n^{1/2}(\widehat{\Delta}_{nj}(\theta)-\widehat{\Delta}_n^{\inf})+n^{1/2}(\widehat{\Delta}_n^{\inf}-\Delta_{F_n}^{\inf})\right)\\
&\le\sup_{\theta\in\widehat{\Theta}_{\min,L,n}}\max_{j\le k} n^{1/2}|\widehat{\Delta}_{nj}(\theta)-\Delta_{F_nj}(\theta)|+O_p(1), \qquad (27.87)
\end{aligned}$$

where the inequality holds by the definition of $\widehat{\Theta}_{\min,L,n}$ (given just below (5.15)) and Theorem 25.2(a) and (b) (which uses Assumptions A.0, C.4, C.5, and C.12). Equation (27.82) and Assumptions C.4 and C.5 give

$$\sup_{\theta\in\widehat{\Theta}_{\min,L,n}}\max_{j\le k} n^{1/2}|\widehat{\Delta}_{nj}(\theta)-\Delta_{F_nj}(\theta)|=O_p(1). \qquad (27.88)$$

In consequence,

$$\begin{aligned}
&P_{F_n}\left(\sup_{\theta\in\widehat{\Theta}_{\min,L,n}} n^{1/2}(\Delta_{F_n}(\theta)-\Delta_{F_n}^{\inf})_-\le\eta_{Ln}\right)\\
&\ge P_{F_n}(O_p(1)\le\eta_{Ln})\\
&\to 1, \qquad (27.89)
\end{aligned}$$

where the convergence holds because $\eta_{Ln}\to\infty$. Combining this with (27.86) establishes part (b).

□

28 Proof of Theorem 5.2

28.1 Lemmas 28.1–28.4

We introduce the following lower bound on $A^*_{n,\Delta U}$:

$$A^*_{Ln,\Delta U} := \inf_{\theta\in\Theta} \min_{j\leq k} \left(-\widehat{\nu}^*_{nj}(\theta)\right). \tag{28.1}$$

Let $\widehat{c}_{Ln,\Delta U}(1-\alpha)$ be the α conditional quantile of $A^*_{Ln,\Delta U}$ given $\{W_i\}_{i\leq n}$ for $\alpha \in (0,1)$.

We introduce the following upper bound on $A^*_{n,\Delta L}$:

$$A^*_{Un,\Delta L} := \sup_{\theta\in\Theta} \min_{j\leq k} \left(-\widehat{\nu}^*_{nj}(\theta)\right). \tag{28.2}$$

Let $\widehat{c}_{Un,\Delta L}(1-\alpha)$ be the α conditional quantile of $A^*_{Un,\Delta L}$ given $\{W_i\}_{i\leq n}$ for $\alpha \in (0,1)$.

By Lemma 24.1, Assumption SLK holds iff there exists a sequence $\{\theta^I_n \in \Theta_I(F_n)\}_{n\geq 1}$ for which $n^{1/2}E_{F_n}\widetilde{m}_j(W,\theta^I_n) \to \infty$ $\forall j \leq k$. Let

$$\begin{aligned}
\psi_n &:= \min_{j\leq k} n^{1/2}E_{F_n}\widetilde{m}_j(W,\theta^I_n),\\
\chi_n &:= n^{1/2}\Delta^{\inf}_{F_n} := n^{1/2}\inf_{\theta\in\Theta}\max_{j\leq k}(-E_F\widetilde{m}_j(W,\theta)), \text{ and}\\
\widehat{\nu}_{nj}(\theta) &:= n^{1/2}\left(\widehat{m}_{nj}(\theta) - E_{F_n}\widetilde{m}_j(W,\theta)\right).
\end{aligned} \tag{28.3}$$

Under Assumption SLK, $\{\psi_n\}_{n\geq 1}$ exists and satisfies $\psi_n \to \infty$. Under Assumption MM, $\chi_n \to \infty$.

The following lemmas are used in the proof of Theorem 5.2.

Lemma 28.1 *For sequences $\{F_n\}_{n\geq 1}$ that satisfy Assumptions* C.4, C.5, *and* SLK *for a subsequence $\{p_n\}_{n\geq 1}$ in place of $\{n\}_{n\geq 1}$,* $\liminf_{n\to\infty} P_{F_{p_n}}(\widehat{\Delta}^{\inf}_{p_n} < -\psi^{1/2}_{p_n}/p^{1/2}_n) = 1$.

Lemma 28.2 *For sequences $\{F_n\}_{n\geq 1}$ that satisfy Assumption* BC.3 *for a subsequence $\{p_n\}_{n\geq 1}$ in place of $\{n\}_{n\geq 1}$,* (a) $\widehat{c}_{p_n,\Delta U}(1-\alpha) \leq \widehat{c}_{Lp_n,\Delta U}(1-\alpha)$ *for all sample realizations and* (b) $\widehat{c}_{Lp_n,\Delta U}(1-\alpha) = O_p(1)$.

Lemma 28.3 *For sequences $\{F_n\}_{n\geq 1}$ that satisfy Assumptions* C.4, C.5, *and* MM *for a subsequence $\{p_n\}_{n\geq 1}$ in place of $\{n\}_{n\geq 1}$,* $\liminf_{n\to\infty} P_{F_{p_n}}(\widehat{\Delta}^{\inf}_{p_n} > \chi^{1/2}_{p_n}/p^{1/2}_n) = 1$.

Lemma 28.4 *For sequences* $\{F_n\}_{n\geq 1}$ *that satisfy Assumption* BC.3 *for a subsequence* $\{p_n\}_{n\geq 1}$ *in place of* $\{n\}_{n\geq 1}$, (a) $\widehat{c}_{p_n,\Delta L}(1-\alpha) \leq \widehat{c}_{Up_n,\Delta L}(1-\alpha)$ *for all sample realizations and* (b) $\widehat{c}_{Up_n,\Delta L}(1-\alpha) = O_p(1)$.

28.2 Proof of Theorem 5.2

Proof of Theorem 5.2. First we prove part (a). There always exists a subsequence $\{q_n\}_{n\geq 1}$ of $\{n\}_{n\geq 1}$ such that

$$\liminf_{n\to\infty} P_{F_n}\left(\widehat{\Delta}_{n,\Delta U}(\alpha) < 0\right) = \lim_{n\to\infty} P_{F_{q_n}}\left(\widehat{\Delta}_{q_n,\Delta U}(\alpha) < 0\right). \tag{28.4}$$

It suffices to show that the rhs of (28.4) equals one with $\{q_n\}_{n\geq 1}$ replaced by some subsequence $\{p_n\}_{n\geq 1}$ of $\{q_n\}_{n\geq 1}$ (because the limit under the subsequence $\{p_n\}_{n\geq 1}$ is the same as the limit under $\{q_n\}_{n\geq 1}$).

For notational simplicity, we show that the rhs of (28.4) equals one with $p_n = n$. We have

$$\begin{aligned}
&\liminf_{n\to\infty} P_{F_n}\left(\widehat{\Delta}_{n,\Delta U}(\alpha) < 0\right)\\
:=& \liminf_{n\to\infty} P_{F_n}\left(\widehat{\Delta}_n^{\inf} < -\widehat{c}_{n,\Delta U}(1-\alpha)/n^{1/2}\right)\\
\geq& \liminf_{n\to\infty} P_{F_n}\left(\widehat{\Delta}_n^{\inf} < -\psi_n^{1/2}/n^{1/2} \;\&\; -\psi_n^{1/2}/n^{1/2} \leq -\widehat{c}_{n,\Delta U}(1-\alpha)/n^{1/2}\right)\\
\geq& \liminf_{n\to\infty} P_{F_n}\left(\widehat{\Delta}_n^{\inf} < -\psi_n^{1/2}/n^{1/2} \;\&\; -\psi_n^{1/2} \leq -\widehat{c}_{Ln,\Delta U}(1-\alpha)\right)\\
=& \liminf_{n\to\infty} P_{F_n}\left(\widehat{\Delta}_n^{\inf} < -\psi_n^{1/2}/n^{1/2}\right)\\
=& 1,
\end{aligned} \tag{28.5}$$

where the first equality holds by the definition of $\widehat{\Delta}_{n,\Delta U}(\alpha)$ in (5.3), the first inequality is straightforward, the second inequality holds by Lemma 28.2(a), the second equality holds because $-\psi_n^{1/2} \to -\infty$ using Assumption SLK and $\widehat{c}_{Ln,\Delta U}(1-\alpha) = O_p(1)$ by Lemma 28.2(b), and the last equality holds by Lemma 28.1.

It remains to verify that the assumptions used in Lemmas 28.1 and 28.2, namely, Assumptions BC.3, C.4, and C.5 hold for a subsequence $\{p_n\}_{n\geq 1}$ (of $\{q_n\}_{n\geq 1}$) in place of $\{n\}_{n\geq 1}$ (because Assumption SLK, which is imposed in Lemmas 28.1 and 28.2, is also imposed in the present theorem). Such a subsequence $\{p_n\}_{n\geq 1}$ exists by the proof of Theorem 5.1(a) (because Assumptions A.0–A.6 of Theorem 5.1(a) are also imposed by the present theorem. This completes the proof of part (a).

The proof of part (b) is analogous to that of part (a) with the inequalities inside the probabilities in (28.5) reversed, with $\chi_n^{1/2}$ in place of $-\psi_n^{1/2}$ in (28.5), and using Lemmas 28.3 and 28.4 in place of Lemmas 28.1 and 28.2. $\square$

28.3 Proofs of Lemmas 28.1–28.4

Proof of Lemma 28.1. For notational simplicity, we prove the result for $p_n = n$. By Lemma 24.1, under Assumption SLK, there exists a sequence $\{\theta_n^I \in \Theta_I(F_n)\}_{n\geq 1}$ for which $n^{1/2}E_{F_n}\widetilde{m}_j(W,\theta_n^I) \to \infty$ $\forall j \leq k$. For $\widehat{\nu}_{nj}(\theta)$ defined in (28.3), $\widehat{\nu}_{nj}(\theta) = O_p^{\Theta}(1)$ by (27.29), using Assumptions C.4 and C.5. Hence, $\widehat{\nu}_n(\theta_n^I) = O_p(1)$.

For $\{\theta_n^I\}_{n\geq 1}$ as above, we have

$$
\begin{aligned}
\left\{\widehat{\Delta}_n^{\inf} < -\psi_n^{1/2}/n^{1/2}\right\} &= \left\{\inf_{\theta\in\Theta}\max_{j\leq k}(-\widehat{m}_{nj}(\theta)) < -\psi_n^{1/2}/n^{1/2}\right\} \\
&\supset \left\{\max_{j\leq k}(-\widehat{m}_{nj}(\theta_n^I)) < -\psi_n^{1/2}/n^{1/2}\right\} \\
&= \left\{\widehat{m}_{nj}(\theta_n^I) > \psi_n^{1/2}/n^{1/2}\ \forall j \leq k\right\} \\
&= \left\{n^{1/2}E_{F_n}m_j(W,\theta_n^I) + \widehat{\nu}_{nj}(\theta_n^I) > \psi_n^{1/2}\ \forall j \leq k\right\} \\
&= \left\{\psi_n^{1/2} + O_p(1) > 0\right\}, \qquad (28.6)
\end{aligned}
$$

where the first equality holds by the definition of $\widehat{\Delta}_n^{\inf}$, the third equality holds by the definition of $\widehat{\nu}_{nj}(\theta)$ in (28.3), and the last equality uses $\widehat{\nu}_n(\theta_n^I) = O_p(1)$ and the definition of ψ_n in (28.3).

We have $P_{F_n}(\psi_n^{1/2} + O_p(1) > 0) \to 1$ because $\psi_n \to \infty$ by definition and Assumption SLK. This and (28.6) establish the result of the lemma. $\square$

Proof of Lemma 28.2. For notational simplicity, we prove the result for $p_n = n$. The bootstrap statistic $A_{n,\Delta U}^*$ is defined in (5.14) (with b deleted) to be

$$
A_{n,\Delta U}^* := \inf_{\theta\in\widehat{\Theta}_{\min,n}}\ \min_{j_1\in\widehat{\mathcal{J}}_{ne}(\theta)}\ \max_{j\leq k}\left(-\widehat{\nu}_{nj}^*(\theta) + 1(j \neq j_1)\widehat{e}_{nj}(\theta) + 1(j = j_1)\varphi(\xi_{nj}^e(\theta))\right). \qquad (28.7)
$$

Using this and $A_{Ln,\Delta U}^* = \inf_{\theta\in\Theta}\min_{j_1\leq k}\left(-\widehat{\nu}_{nj_1}^*(\theta)\right)$ (as defined in (28.1) with j_1 in place of j), we obtain

$$
A_{Ln,\Delta U}^* \leq A_{n,\Delta U}^* \text{ for all bootstrap and sample realizations} \qquad (28.8)
$$

by replacing $\max_{j\leq k}$ in the definition of $A_{n,\Delta U}^*$ by $j = j_1$, $\min_{j_1\in\widehat{\mathcal{J}}_{ne}(\theta)}$ by $\min_{j_1\leq k}$, $1(j = j_1)\varphi(\xi_{nj}^e(\theta))$ by 0 (using $\varphi(\xi) \geq 0$ $\forall \xi \in R$) and $\inf_{\theta\in\widehat{\Theta}_{\min,n}}$ by $\inf_{\theta\in\Theta}$.

By definition, $\widehat{c}_{n,\Delta U}(1-\alpha)$ and $\widehat{c}_{Ln,\Delta U}(1-\alpha)$ are the $1-\alpha$ quantiles of $-A_{n,\Delta U}^*$ and $-A_{Ln,\Delta U}^*$, respectively. This and (28.8) give $\widehat{c}_{n,\Delta U}(1-\alpha) \leq \widehat{c}_{Ln,\Delta U}(1-\alpha)$ for all sample realizations (of

$\{W_{ni}\}_{i\leq n,n\geq 1}$), which establishes part (a) of the lemma.

By (27.37), (27.38), the definition of $G_j^{m\sigma}(\theta)$ in (27.9), Assumption BC.3, and the continuous mapping theorem,

$$A^*_{Ln,\Delta U} \to_d A_{L\infty,\Delta U} := \inf_{\theta\in\Theta}\min_{j\leq k}\left(-G_j^{m\sigma}(\theta)\right) \text{ conditional on } \{W_{ni}\}_{i\leq n,n\geq 1} \text{ a.s.}[P_{\bigtriangledown}], \tag{28.9}$$

where $A_{L\infty,\Delta U}\in R$ a.s. In consequence, $\widehat{c}_{Ln,\Delta U}(1-\alpha) = O(1)$ conditional on $\{W_{ni}\}_{i\leq n,n\geq 1}$ a.s.$[P_{\bigtriangledown}]$. In turn, this implies that $\widehat{c}_{Ln,\Delta U}(1-\alpha) = O_p(1)$ by an analogous argument to that used to show that a.s. convergence implies convergence in probability, which establishes part (b) of the lemma. □

Proof of Lemma 28.3. For notational simplicity, we prove the results with $p_n = n$. We have

$$\begin{aligned}
\left\{\widehat{\Delta}_n^{\inf} > \chi_n^{1/2}/n^{1/2}\right\} &= \left\{\inf_{\theta\in\Theta}\max_{j\leq k}(-\widehat{m}_{nj}(\theta)) > \chi_n^{1/2}/n^{1/2}\right\}\\
&= \left\{\inf_{\theta\in\Theta}\max_{j\leq k}(-\widehat{\nu}_{nj}(\theta) - n^{1/2}E_{F_n}m_j(W,\theta)) > \chi_n^{1/2}\right\}\\
&= \left\{\inf_{\theta\in\Theta}\max_{j\leq k}(O_p^{\Theta}(1) - n^{1/2}E_{F_n}m_j(W,\theta)) > \chi_n^{1/2}\right\}\\
&= \left\{\inf_{\theta\in\Theta}\max_{j\leq k}(-n^{1/2}E_{F_n}m_j(W,\theta)) + O_p(1) > \chi_n^{1/2}\right\}\\
&= \left\{\chi_n + O_p(1) > \chi_n^{1/2}\right\},
\end{aligned} \tag{28.10}$$

where the first equality holds by the definition of $\widehat{\Delta}_n^{\inf}$, the second equality holds by the definition of $\widehat{\nu}_{nj}(\theta)$ in (28.3), the third equality holds because $\widehat{\nu}_{nj}(\theta) = O_p^{\Theta}(1)$ by (27.29) using Assumptions C.4 and C.5, the fourth equality holds by standard calculations, and the last equality holds by the definitions of χ_n in (28.3) and $\Delta_{F_n}^{\inf}$ in (5.1).

We have $P_{F_n}(\chi_n + O_p(1) > \chi_n^{1/2}) \to 1$ because $\chi_n\to\infty$ by Assumption MM. This and (28.10) establish the result of the lemma. □

Proof of Lemma 28.4. We have $A^*_{n,\Delta L} \leq A^*_{Un,\Delta L}$ for all sample and bootstrap realizations because $A^*_{n,\Delta L} := \inf_{\theta\in\widehat{\Theta}_{\min,n}}\max_{j\leq k}(-\widehat{\nu}^*_{nj}(\theta) - \varphi(-\xi^e_{nj}(\theta)) \leq \sup_{\theta\in\Theta}\max_{j\leq k}(-\widehat{\nu}^*_{nj}(\theta)) =: A^*_{Un,\Delta L}$, where the inequality holds because $\varphi(\xi)\geq 0$ for all $\xi\in R$ by the definition of φ. In consequence, $\widehat{c}_{n,\Delta L}(1-\alpha) \leq \widehat{c}_{Un,\Delta L}(1-\alpha)$ for all sample realizations by the definition of a quantile, which proves part (a).

The proof of part (b) is the same as that of Lemma 28.2(b) but with $\sup_{\theta\in\Theta}$ in place of $\inf_{\theta\in\Theta}$, given the difference in the definitions of $A^*_{Ln,\Delta U}$ in (28.1) and $A^*_{Un,\Delta L}$ in (28.2). □

29 Proof of Corollary 5.3

Proof of Corollary 5.3. Part (a) holds by Theorem 5.1(a) by the following calculations:

$$
\begin{aligned}
&\limsup_{n\to\infty} \sup_{F\in\mathcal{P}:\Delta_F^{\inf}>0} P_F\left(n^{1/2}\widehat{\Delta}_n^{\inf} < -\widehat{c}_{n,\Delta U}(1-\alpha)\right) \\
&= 1 - \liminf_{n\to\infty} \inf_{F\in\mathcal{P}:\Delta_F^{\inf}>0} P_F\left(n^{1/2}\widehat{\Delta}_n^{\inf} \geq -\widehat{c}_{n,\Delta U}(1-\alpha)\right) \\
&\leq 1 - \liminf_{n\to\infty} \inf_{F\in\mathcal{P}:\Delta_F^{\inf}>0} P_F\left(n^{1/2}\left(\widehat{\Delta}_n^{\inf} - \Delta_F^{\inf}\right) \geq -\widehat{c}_{n,\Delta U}(1-\alpha)\right) \\
&\leq 1 - \liminf_{n\to\infty} \inf_{F\in\mathcal{P}} P_F\left(\Delta_F^{\inf} \in CI_{n,\Delta U}(\alpha)\right) \\
&\leq \alpha,
\end{aligned} \tag{29.1}
$$

where the first inequality holds because the infimum is over $\Delta_F^{\inf} \geq 0$, the second inequality holds because $\{F \in \mathcal{P} : \Delta_F^{\inf} > 0\} \subset \mathcal{P}$ and uses the definition of $CI_{n,\Delta U}(\alpha)$ in (5.3), and the last inequality holds by Theorem 5.1(a).

Part (b) holds by Theorem 5.2(a) because

$$
\liminf_{n\to\infty} P_{F_n}\left(n^{1/2}\widehat{\Delta}_n^{\inf} < -\widehat{c}_{n,\Delta U}(1-\alpha)\right) = \liminf_{n\to\infty} P_{F_n}\left(\widehat{\Delta}_{n,\Delta U}(\alpha) < 0\right) \tag{29.2}
$$

by the definition of $\widehat{\Delta}_{n,\Delta U}(\alpha)$ in (5.3). □

30 Behavior of $\widehat{\Delta}_{n,\Delta U}(\alpha)$ under Assumption MM

The CI $CI_{n,\Delta U}(\alpha)$ equals $(-\infty, \widehat{\Delta}_{n,\Delta U}(\alpha)]$. Its upper-bound $\widehat{\Delta}_{n,\Delta U}(\alpha)$ is used in the construction of the SPUR2 test and CS. The following result concerns the behavior of $\widehat{\Delta}_{n,\Delta U}(\alpha)$ under Assumption MM.

Theorem 30.1 *Suppose Assumptions* A.0–A.6 *hold. For sequences* $\{F_n\}_{n\geq 1}$ *that satisfy Assumption* MM, $\liminf_{n\to\infty} P_{F_n}(\widehat{\Delta}_{n,\Delta U}(\alpha) > 0) = 1$.

Comment. Theorem 30.1 is used in Section 4.6 to show that the level α adaptive SPUR2 test typically has the same power properties as the level α_2 SPUR1 test when the model is misspecified and Assumption MM holds, where $\alpha = \alpha_1 + \alpha_2$ and $\alpha_1, \alpha_2 > 0$, such as $\alpha = .05$ and $\alpha_2 = .045$.

Proof of Theorem 30.1. There always exists a subsequence $\{q_n\}_{n\geq 1}$ of $\{n\}_{n\geq 1}$ such that

$$
\liminf_{n\to\infty} P_{F_n}\left(\widehat{\Delta}_{n,\Delta U}(\alpha) > 0\right) = \lim_{n\to\infty} P_{F_{q_n}}\left(\widehat{\Delta}_{q_n,\Delta U}(\alpha) > 0\right). \tag{30.1}
$$

It suffices to show that the rhs of (30.1) equals one with $\{q_n\}_{n\geq 1}$ replaced by some subsequence $\{a_n\}_{n\geq 1}$ of $\{q_n\}_{n\geq 1}$ (because the limit under the subsequence $\{a_n\}_{n\geq 1}$ is the same as the limit under $\{q_n\}_{n\geq 1}$).

First, we consider the same conditions as in Theorem 27.4(a), namely, that $\{F_n\}_{n\geq 1}$ is a sequence that satisfies Assumptions A.0, A.6, BC.3–BC.5, C.4, C.5, C.12, and C.13 for a subsequence $\{p_n\}_{n\geq 1}$ in place of $\{n\}_{n\geq 1}$. In this case, by the proof of Theorem 27.4(a), see the paragraph that contains (27.15), we have

$$\widehat{c}_{p_n,\Delta U}(1-\alpha) \geq \widehat{c}_{Up_n,\Delta U}(1-\alpha) + o_p(1), \tag{30.2}$$

where $\widehat{c}_{Un,\Delta U}(1-\alpha)$ is defined following (27.8). In addition, there exists a subsequence $\{a_n\}_{n\geq 1}$ of $\{p_n\}_{n\geq 1}$ for which

$$\widehat{c}_{Ua_n,\Delta U}(1-\alpha) \geq c_{U\infty} - 1 \text{ wp} \to 1 \text{ and } c_{U\infty} - 1 > -\infty \tag{30.3}$$

by (27.19) with $\varepsilon = 1$.

For notational simplicity, we show that the rhs of (30.1) equals one with $a_n = n$. We have

$$\begin{aligned}
\lim_{n\to\infty} P_{F_n}\left(\widehat{\Delta}_{n,\Delta U}(\alpha) > 0\right) &:= \lim_{n\to\infty} P_{F_n}\left(\widehat{\Delta}_n^{\inf} > -\widehat{c}_{n,\Delta U}(1-\alpha)/n^{1/2}\right) \\
&\geq \lim_{n\to\infty} P_{F_n}\left(\widehat{\Delta}_n^{\inf} > \chi_n^{1/2}/n^{1/2} \;\&\; \chi_n^{1/2}/n^{1/2} \geq -\widehat{c}_{n,\Delta U}(1-\alpha)/n^{1/2}\right) \\
&\geq \lim_{n\to\infty} P_{F_n}\left(\widehat{\Delta}_n^{\inf} > \chi_n^{1/2}/n^{1/2} \;\&\; \chi_n^{1/2} \geq -\widehat{c}_{Un,\Delta U}(1-\alpha) + o_p(1)\right) \\
&= \lim_{n\to\infty} P_{F_n}\left(\widehat{\Delta}_n^{\inf} > \chi_n^{1/2}/n^{1/2}\right) \\
&= 1,
\end{aligned} \tag{30.4}$$

where χ_n is defined in (28.3), the first equality holds by the definition of $\widehat{\Delta}_{n,\Delta U}(\alpha)$ in (5.3), the first inequality is straightforward, the second inequality holds by (30.2), the second equality holds by $\chi_n^{1/2} \to \infty$ by Assumption MM and (30.3), and the last equality holds by Lemma 28.3, which uses Assumptions C.4, C.5, and MM.

It remains to verify that the assumptions used in Theorem 27.4(a) and Lemma 28.3, namely, Assumptions BC.3–BC.5, C.4, C.5, C.12, and C.13, hold for a subsequence $\{p_n\}_{n\geq 1}$ (of $\{q_n\}_{n\geq 1}$) in place of $\{n\}_{n\geq 1}$ (because Assumptions A.0, A.6, and MM, which are imposed in Theorem 27.4(a), are also imposed in the present theorem). Such a subsequence $\{p_n\}_{n\geq 1}$ exists by the proof of Theorem 5.1(a) (because Assumptions A.0–A.5 of Theorem 5.1(a) are also imposed by the present theorem. This completes the proof. □

31 Assumptions

For ease of reference, we state all of the assumptions used in the paper and online Appendix C here.

Assumption A.0. (i) Θ is compact and non-empty and (ii) $E_F\widetilde{m}_j(W,\theta)$ is upper semi-continuous on Θ $\forall j \leq k$, $\forall F \in \mathcal{P}$.

Assumption A.1. The observations $W_1, ..., W_n$ are i.i.d. under F and $\{\widetilde{m}_j(\cdot,\theta) : \mathcal{W} \to R\}$ and $\{\widetilde{m}_j^2(\cdot,\theta) : \mathcal{W} \to R\}$ are measurable classes of functions indexed by $\theta \in \Theta$ $\forall j \leq k$, $\forall F \in \mathcal{P}$.

Assumption A.2. For some $a > 0$, $\sup_{F\in\mathcal{P}} E_F \sup_{\theta\in\Theta} ||\widetilde{m}(W,\theta)||^{4+a} < \infty$.

Assumption A.3. The empirical process $\nu_n(\cdot)$ is asymptotically ρ_F-equicontinuous on Θ uniformly in $F \in \mathcal{P}$.

Assumption A.4. The covariance kernel $\Omega_F(\theta,\theta')$ satisfies: for all $F \in \mathcal{P}$,

$\lim_{\delta\to 0} \sup_{||(\theta_1,\theta_1')-(\theta_2,\theta_2')||<\delta} ||\Omega_F(\theta_1,\theta_1') - \Omega_F(\theta_2,\theta_2')|| = 0.$

Assumption A.5. $E_F\widetilde{m}(W,\theta)$ is equicontinuous on Θ over $F \in \mathcal{P}$. That is, $\lim_{\delta\downarrow 0}\sup_{F\in\mathcal{P}}$ $\sup_{||\theta-\theta'||<\delta} ||E_F\widetilde{m}(W,\theta) - E_F\widetilde{m}(W,\theta')|| = 0$.

Assumption A.6. (i) $\kappa_n \to \infty$. (ii) $\tau_n \to \infty$.

Assumption A.7. (i) $\kappa_n \to \infty$ and (ii) $\kappa_n/n^{1/2} \to 0$.

Assumption A.8. (i) For all $F \in \mathcal{P}$ and $\theta \in \Theta$, $\Delta_F(\theta) - \Delta_F^{\inf} \geq c\min\{\delta, \inf_{\overline{\theta}\in\Theta_{\min}(F)} ||\theta - \overline{\theta}||\}$ for constants $c, \delta > 0$.

(ii) Θ is convex.

(iii) $E_F\widetilde{m}(W,\theta)$ is differentiable in θ for all $F \in \mathcal{P}$ and $\{\widetilde{M}_F(\theta) := (\partial/\partial\theta')E_F\widetilde{m}(W,\theta) : F \in \mathcal{P}\}$ is equicontinuous, i.e., $\lim_{\delta\to 0}\sup_{F\in\mathcal{P}}\sup_{(\theta,\overline{\theta}):||\theta-\overline{\theta}||\leq\delta} ||\widetilde{M}_F(\theta) - \widetilde{M}_F(\overline{\theta})|| = 0$.

The following assumptions apply to a drifting sequence of distributions $\{F_n\}_{n\geq 1}$.

Assumption C.4. $\sup_{\theta\in\Theta} ||E_{F_n}\widetilde{m}(W,\theta) - \widetilde{m}(\theta)|| \to 0$ for some nonrandom bounded continuous R^k-valued function $\widetilde{m}(\cdot)$ on Θ.

Assumption C.5. $\nu_n(\cdot) := (\nu_n^m(\cdot)', \nu_n^\sigma(\cdot)')' \Rightarrow G(\cdot) := (G^m(\cdot)', G^\sigma(\cdot)')'$ as $n \to \infty$, where $\{G(\theta) : \theta \in \Theta\}$ is a mean zero R^{2k}-valued Gaussian process with bounded continuous sample paths a.s. and $G^m(\theta), G^\sigma(\theta) \in R^k$.

Assumption C.11. $\Omega_{F_n}(\cdot,\cdot) \to_u \Omega_\infty(\cdot,\cdot)$ for some continuous $R^{2k\times 2k}$-valued function $\Omega_\infty(\cdot,\cdot)$ on Θ^2.

Assumption C.12. $\Lambda_{n,\Delta,F_n} \to_H \Lambda_\Delta$ for some non-empty set $\Lambda_\Delta \in \mathcal{S}(\Theta \times R^k_{[\pm\infty]})$.

Assumption C.13. $\Lambda^{\eta_n}_{n,\Delta,F_n} \to_H \Lambda_{\Delta\min}$ for some non-empty set $\Lambda_{\Delta\min} \in \mathcal{S}(\Theta \times R^k_{[\pm\infty]})$, where $\{\eta_n\}_{n\geq 1}$ is a sequence of positive constants for which $\eta_n \to \infty$.

The following assumptions apply to a drifting sequence of distributions $\{F_n\}_{n\geq 1}$.

Assumption BC.3. $\{\nu_n^*(\cdot)|\{W_{ni}\}_{i\leq n,n\geq 1}\} \Rightarrow G(\cdot)$ a.s.$[P_{\bigtriangledown}]$, where $G(\cdot)$ is as in Assumption C.5.

Assumption BC.4. $\Lambda_{n,\Delta,F_n}^{*\eta_n} \to_H \Lambda_{\Delta\min}^*$ for some non-empty set $\Lambda_{\Delta\min}^* \in \mathcal{S}(\Theta\times R_{[\pm\infty]}^{2k}\times\{1,...,k\})$ for some sequence of constants $\{\eta_n\}_{n\geq 1}$ that satisfies $\eta_n \to \infty$ and $\eta_n/\tau_n \to 0$ for the constants $\{\tau_n\}_{n\geq 1}$ that appear in $(\widehat{\theta}_w)$ and satisfy Assumption A.6(ii).

Assumption BC.5. $\widehat{\Theta}_{\min,n} \supseteq \Theta_{\min}^{\eta_n}(F_n)$ wp$\to$1 for constants $\{\eta_n\}_{n\geq 1}$ as in Assumptions BC.4 and C.13.

Assumption BC.6. $\Lambda_{n,F_n,L}^{*\eta_{Ln}} \to_H \Lambda_L^*$ for some non-empty set $\Lambda_L^* \in \mathcal{S}(\Theta\times R_{[\pm\infty]}^{k})$ for some sequence of constants $\{\eta_{Ln}\}_{n\geq 1}$ that satisfies $\eta_{Ln} \to \infty$, $\eta_{Ln}/n^{1/2} \to 0$, and $\eta_{Ln}/\kappa_n^{\gamma} \to 0$ for some $\gamma \in (0,1)$ for the constants $\{\kappa_n\}_{n\geq 1}$ that are employed in the definition of $A_{n,\Delta L}^*$ and satisfy Assumption A.7.

Assumption BC.7. $\widehat{\Theta}_{\min,L,n} \subseteq \Theta_{\min}^{\eta_{Ln}}(F_n)$ wp$\to$1 for constants $\{\eta_{Ln}\}_{n\geq 1}$ as in Assumptions BC.6.

Assumption SLK. The sequence $\{F_n\}_{n\geq 1}$ is such that $n^{1/2}\Delta_{F_n}^{\inf} \to -\infty$.

Assumption SLK.1. The sequence $\{F_n\}_{n\geq 1}$ is such that $\lim_{n\to\infty} n^{1/2} d_H(\Theta_{\min}(F_n), \Theta_I^{MR}(F_n)) = \infty$.

Assumption MM. The sequence $\{F_n\}_{n\geq 1}$ is such that $n^{1/2}\Delta_{F_n}^{\inf} \to \infty$.

Online Appendix C References

Bugni, F. A., I. A. Canay, and X. Shi (2015): "Specification Tests for Partially Identified Models Defined by Moment Inequalities," *Journal of Econometrics,* 185, 259–282.

——— (2017): "Inference for Subvectors and Other Functions of Partially Identified Parameters," *Quantitative Economics,* 8, 1–38.

Dudley, R. M. (1989): *Real Analysis and Probability.* Pacific Grove, CA: Wadsworth and Brook/Cole.

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

Appendix D

to

Misspecified Moment Inequality Models: Inference and Diagnostics

Contents: Related Results and Extensions

Donald W. K. Andrews
Cowles Foundation for Research in Economics
Yale University

Soonwoo Kwon
Department of Economics
Brown University

First Version: November, 2018
Revised: July 29, 2022

32 Outline of Online Appendix D

Appendix D provides the following results.

Section 33 provides an alternative interpretation of the identified set $\Theta_I^{MR}(F)$.

Section 34 proves Lemma 33.1 that appears in Section 33.

Section 35 shows that when the "max" S function is employed, the SPUR test statistic is equivalent to a recentered test statistic, as has been considered in Chernozhukov, Hong, and Tamer (2007) for use with a correctly-specified model.

Section 36 discusses extensions of the results of the paper to non-i.i.d. observations, to tests with weighted moment inequalities, and to tests without the standard-deviation normalization.

33 An Alternative Interpretation of $\Theta_I^{MR}(F)$

Here we provide an alternative interpretation of the MR-identified set $\Theta_I^{MR}(F)$. We show that if one allows for a large class of transformations of the inequality model that yield non-empty identified sets, then the union of the identified sets for the transformations that minimize the distance between the original model and the transformed model is $\Theta_I^{MR}(F)$. Furthermore, if $\Theta_I^{MR}(F)$ is a singleton, as is often the case under identifiable misspecification, then $\Theta_I^{MR}(F)$ equals the identified set corresponding to each of these "minimal" transformations. These results also hold if the class of transformations is restricted to nonnegative shift functions, which necessarily relax the original model.

Consider two moment inequality models $E_F\widetilde{m}(W,\theta) \geq 0_k$ and $E_F\widetilde{m}^t(W,\theta) \geq 0_k$. Define the distance between the two models as

$$d(\widetilde{m}, \widetilde{m}^t) := \sup_{\theta\in\Theta} \left\| E_F\widetilde{m}(W,\theta) - E_F\widetilde{m}^t(W,\theta)\right\|_\infty , \tag{33.1}$$

where $\|\cdot\|_\infty$ denotes the ℓ_∞ norm on R^k.

A transformed version of the original model can always be written as $\widetilde{m}^t(W,\theta) = \widetilde{m}(W,\theta) + t(W,\theta)$ for a transformation function t. Suppose $\sup_{\theta\in\Theta} ||E_F t(W,\theta)||_\infty < \infty$. Let $\mathcal{T}(F)$ be the set of transformation functions t that make the identified set non-empty; let $\mathcal{T}^*(F)$ be the set of transformation functions that minimize the distance between $\widetilde{m}$ and $\widetilde{m}^t$ over $t \in \mathcal{T}(F)$; and let

$\Theta_{It}(F)$ be the identified set corresponding to a transformation function t. That is,

$$\begin{aligned}
\mathcal{T}(F) &:= \{t : E_F\widetilde{m}^t(W,\theta) \geq 0_k \text{ for some } \theta \in \Theta\},\\
\mathcal{T}^*(F) &:= \arg\min_{t\in\mathcal{T}(F)} d(\widetilde{m},\widetilde{m}^t), \text{ and}\\
\Theta_{It}(F) &:= \{\theta \in \Theta : E_F\widetilde{m}(W,\theta) + E_F t(W,\theta) \geq 0_k\}. \qquad (33.2)
\end{aligned}$$

The next lemma shows that the union of the identified sets corresponding to $t \in \mathcal{T}^*(F)$ is $\Theta_I^{MR}(F)$.

Lemma 33.1 *Suppose Assumption* A.0 *holds. Then,* $\cup_{t\in\mathcal{T}^*(F)}\Theta_{It}(F) = \Theta_I^{MR}(F)$.

Comment. If $\Theta_I^{MR}(F)$ is a singleton, then $\Theta_{It}(F) = \Theta_I^{MR}(F)$ for all $t \in \mathcal{T}^*(F)$. If the union is restricted to nonnegative transformations $t \in \mathcal{T}^*(F)$, Lemma 33.1 still holds.

34 Proof of Lemma 33.1

Proof of Lemma 33.1. Assumption A.0 guarantees that the infimum in the definition of $r_F^{\inf}$ is attained. Let $t_j(\theta) = E_F t_j(W,\theta)$, where $t_j(W,\theta)$ is the j element of $t(W,\theta)$. We first show that

$$\mathcal{T}^*(F) = \left\{t \in \mathcal{T}(F) : \sup_{\theta\in\Theta}\max_{j\leq k}|t_j(\theta)| = r_F^{\inf}\right\}. \qquad (34.1)$$

For any $t \in \mathcal{T}(F)$, suppose $\sup_{\theta\in\Theta}\max_{j\leq k}|t_j(\theta)| < r_F^{\inf}$. Then, for $\theta_I \in \Theta_{It}(F)$, we have $\widetilde{t} := \max_{j\leq k}|t_j(\theta_I)| < r_F^{\inf}$ and the constant function $t(W,\theta) = \widetilde{t}1_k$ satisfies $\Theta_{I,\widetilde{t}1_k} \neq \varnothing$, which contradicts the definition of $r_F^{\inf}$. Hence, $\sup_{\theta\in\Theta}\max_{j\leq k}|t_j(\theta)| \geq r_F^{\inf}$. Taking $t(W,\theta) = r_F^{\inf}1_k$, this lower bound is attained and we have $\min_{t\in\mathcal{T}(F)}\sup_{\theta\in\Theta}\max_{j\leq k}|t_j(\theta)| = r_F^{\inf}$.

The constant function $t_c(W,\theta) = r_F^{\inf}1_k$ has $t_c \in \mathcal{T}^*(F)$ and $\Theta_{It_c}(F) = \Theta_I^{MR}(F)$. Hence, it suffices to show that $\Theta_{It}(F) \subseteq \Theta_I^{MR}(F)$ for any $t \in \mathcal{T}^*(F)$. For any $t \in \mathcal{T}^*(F)$, we have $t_j(\theta) \leq r_F^{\inf}$ for $j \leq k$ for all $\theta \in \Theta$ by (34.1). It follows that

$$\begin{aligned}
\Theta_{It}(F) :=&\{\theta \in \Theta : E_F\widetilde{m}(W,\theta) + E_F t(W,\theta) \geq 0_k\}\\
&\subseteq\{\theta \in \Theta : E_F\widetilde{m}(W,\theta) + E_F t(W,\theta) + (r_F^{\inf}1_k - E_F t(W,\theta)) \geq 0_k\} = \Theta_I^{MR}(F), \qquad (34.2)
\end{aligned}$$

which concludes the proof. □

35 Recentered Test Statistics

An alternative to the SPUR test statistic defined in Section 4.1 is a recentered test statistic, such as considered in Chernozhukov, Hong, and Tamer (2007), which is defined to be

$$S_{n,Recen}(\theta) := S_{n,Std}(\theta) - \inf_{\overline{\theta}\in\Theta} S_{n,Std}(\overline{\theta}), \quad (35.1)$$

where $S_{n,Std}(\theta) := S(n^{1/2}\widehat{m}_n(\theta), \widehat{\Omega}_n(\theta))$ is a "standard" test statistic, such as one considered in Andrews and Soares (2010), as in (4.5) with $\widehat{r}_n^{\inf} = 0$. The MR-identified set corresponding to the recentered statistic is the set of θ values that minimize the population version of the recentered statistic.[52] It depends on the choice of test statistic.

Chernozhukov, Hong, and Tamer (2007) consider recentered test statistics, but they do not analyze their asymptotic properties under misspecification or under correct specification with drifting sequences of distributions $\{F_n\}_{n\geq 1}$. Lemma 35.1 below, combined with Section 12 in online Appendix A, shows that subsampling a recentered test statistic does not necessarily deliver correct asymptotic size under model misspecification. In addition, it is not clear whether the application of subsampling to recentered test statistics provides critical values that are uniformly asymptotically valid in general under correct specification.[53]

When $S_{n,Std}(\theta)$ is a test statistic from Andrews and Soares (2010) with the function S equal to S_4, see (4.6), we denote the recentered test statistic by $S_{4n,Recen}(\theta)$. It is easy to show that the MR-identified set corresponding to $S_{4n,Recen}(\theta)$ is the same as the MR-identified set in Section 3. On the other hand, if one employs a different S function in $S_{Recen,n}(\theta)$, the MR-identified set is different.

When the function S employed by the SPUR test statistic $S_n(\theta)$ defined in (4.5) is S_4, we denote the SPUR statistic by $S_{4n}(\theta)$. The following lemma shows that the recentered statistic $S_{4n,Recen}(\theta)$ is identical to the $S_{4n}(\theta)$ SPUR statistic. That is, for the S_4 function, the recentered statistic is not an alternative to the SPUR statistic—it is the same.

Lemma 35.1 *For any $\theta \in \Theta$, $S_{4n,Recen}(\theta) = S_{4n}(\theta)$.*

Comment. Section 12 in online Appendix A shows that, for the function S_4, subsampling the

[52] The population version of the recentered statistic is $S(E_F\widetilde{m}(W,\theta), \Omega_F(\theta)) - \inf_{\overline{\theta}\in\Theta} S(E_F\widetilde{m}(W,\overline{\theta}), \Omega_F(\theta))$, where $\Omega_F(\theta) := Var_F(\widetilde{m}(W_i,\theta))$.

[53] The reason is that, even under correct specification, the recentering term $\inf_{\overline{\theta}\in\Theta} S_{n,Std}(\overline{\theta})$ has a complicated asymptotic distribution under drifting sequences of distributions (given by $A_\infty(\Lambda)$ in Theorem 15.3(b) when the recentered test is based on S_4 in (4.6)). In consequence, the argument for the correct asymptotic size of the subsampling test based on a test statistic without recentering that is given in Andrews and Guggenberger (2009) does not extend to the case of the subsampling recentered test.

SPUR test statistic does not necessarily yield correct asymptotic size under model misspecification. Given Lemma 35.1, this also is true for subsampling the recentered test statistic.

Proof of Lemma 35.1. By (4.4), $\widehat{r}_n^{\inf} := \inf_{\theta\in\Theta} \max_{j\leq k}[\widehat{m}_{nj}(\theta)]_-$. Hence, for $S = S_4$, $\inf_{\overline{\theta}\in\Theta} S_{n,Std}(\overline{\theta}) = n^{1/2}\widehat{r}_n^{\inf}$. In consequence,

$$\begin{aligned} S_{4n,Recen}(\theta) &= \max_{j\leq k}\left[n^{1/2}\widehat{m}_{nj}(\theta)\right]_- - n^{1/2}\widehat{r}_n^{\inf} \text{ and} \\ S_{4n}(\theta) &= \max_{j\leq k}\left[n^{1/2}\widehat{m}_{nj}(\theta) + n^{1/2}\widehat{r}_n^{\inf}\right]_- . \end{aligned} \tag{35.2}$$

We claim: $S_{4n,Recen}(\theta) > 0$ iff $S_{4n}(\theta) > 0$. This clearly holds if $\widehat{r}_n^{\inf} = 0$, so suppose $\widehat{r}_n^{\inf} > 0$. In this case, $S_{4n,Recen}(\theta) > 0$ iff $-n^{1/2}\widehat{m}_{nj}(\theta) - n^{1/2}\widehat{r}_n^{\inf} > 0$ for some $j \leq k$ iff $S_{4n}(\theta) > 0$, which proves the claim. In addition, $S_{4n}(\theta) \geq 0$ because $[x]_- \geq 0$ for all x, and $S_{4n,Recen}(\theta) \geq 0$ because $\widehat{r}_n^{\inf}$ is the $\inf_{\theta\in\Theta}$ of $\max_{j\leq k}[\widehat{m}_{nj}(\theta)]_-$, which completes the proof. □

For recentered tests based on S not equal to S_4, one can determine the asymptotic distribution of $S_{n,Recen}(\theta_n)$ under suitable drifting sequences $\{\theta_n\}_{n\geq1}$ and $\{F_n\}_{n\geq1}$ by altering the proof of Theorem 15.3(b). However, the resulting asymptotic distribution seems problematic because it is not apparent how one can construct a critical value in an EGMS fashion that exploits the analogue of the condition $\max_{j\leq k} b_j \geq 0$, which appears when $S = S_4$.

36 Extensions

36.1 Non-I.I.D. Observations

The basic results in this paper are given under high-level conditions that allow for non-identically distributed and/or clustered observations, as well as time series observations. For example, this is true of Theorem 15.3 and of Theorem 18.1 in online Appendix A, which is the key ingredient to the proof of Theorem 4.1. In particular, provided the distributions F of the observations are restricted such that Assumptions C.5 and C.6 in Section 15.1, as well as Assumption BC.3, which is stated in Section 18 in online Appendix B, can be verified for suitable subsequences $\{p_n\}_{n\geq1}$ of $\{n\}_{n\geq1}$, the rest of the proofs of the asymptotic level results go through.

For non-i.i.d. observations, the following changes are needed: the nonparametric i.i.d. bootstrap defined just above (4.17) needs to be changed (a) for clustered observations to a cluster-level nonparametric i.i.d. bootstrap and (b) for time series observations to a block bootstrap or Markov bootstrap, but (c) for independent non-identically distributed observations does not need to be changed. With these changes, the misspecification index CI's and the SPUR1 and SPUR2 tests and CS's have correct asymptotic level (under conditions such that Assumptions C.5, C.6, and BC.3 can be verified).

36.2 Weighted Moments

The weights used in the definition of the MR-identified set $\Theta_I^{MR}(F)$ in (3.6) are uniform weights. This follows from the 1_k vector that appears in (3.5) and (3.6). Non-uniform weights $\omega := (\omega_1, ..., \omega_k)'$, where $\omega_j \in [0, \infty)$ for $j \leq k$, can be introduced by replacing 1_k by $\overline{\omega} = (1/\omega_1, ..., 1/\omega_k)'$ in these equations, where $1/0 := \infty$. Equivalently, one can define $r_{Fj}(\theta) := [\omega_j E_F \widetilde{m}_j(W, \theta)]_-$ and $r_F(\theta) = \max_{j \leq k} r_{Fj}(\theta)$. The larger is ω_j, the more weight is placed on inequality j and the less inequality j is relaxed in the MR-identified set under misspecification. For example, if one believes that some key moment inequalities are correctly specified and one does not want these inequalities to be relaxed under misspecification, then one can set the weights ω_j corresponding to these inequalities to be very large relative to the other weights, such as 1000 versus 1. If $\omega_j = 0$, the jth moment inequality is ignored.

The SPUR1 and SPUR2 tests and CS's can be constructed with weights ω. One replaces $\widehat{m}_{nj}(\theta)$ by $\omega_j \widehat{m}_{nj}(\theta)$ in the definition of $\widehat{r}_{nj}(\theta)$ in (4.4), in (4.5), (4.19), and (4.20)–(4.23), and in the definition of $\widehat{\Delta}_{nj}(\theta)$ in (5.2). One replaces $\widehat{\nu}_{njb}^*(\theta)$ by $\omega_j \widehat{\nu}_{njb}^*(\theta)$ in (4.18), (4.21), and (5.14). One replaces $\widehat{sd}_{njB}(\theta)$ by $\omega_j \widehat{sd}_{njB}(\theta)$ in (5.8). One replaces $n^{1/2}\overline{m}_{njb}^*(\theta)/\widehat{\sigma}_{njb}^*(\theta)$ by $\omega_j n^{1/2}\overline{m}_{njb}^*(\theta)/\widehat{\sigma}_{njb}^*(\theta)$ in the definitions of $sd_{anjB}^*(\theta)$ for $a = 1, 2, 3$ following (4.19), following (4.21), and following (4.22), respectively.

Provided $\omega_j \in [0, \infty)$ for all $j \leq k$ and $\omega_j > 0$ for some $j \leq k$, the asymptotic results concerning the SPUR1 and SPUR2 tests and CS's, i.e., Theorem 4.1, as well as Theorem 11.1 in online Appendix A, go through for the weighted versions of these tests. The tests are invariant to the scale of ω.

36.3 Tests without the Standard-Deviation Normalization

In some scenarios, it may be desirable to define the MR-identified set $\Theta_I^{MR}(F)$ in (3.6) without the standard deviation normalization of the moment functions—i.e., to define $\Theta_I^{MR}(F)$ with $m(W, \theta)$ in place of $\widetilde{m}(W, \theta)$. For example, in their study of demand based on quasilinear utility, Allen and Rehbeck (2019) do not renormalize their moment inequality functions because the moment functions are denominated in dollars, which makes the interpretation simple.

In this paper, a notationally-convenient equivalent way to describe non-normalized moments is to redefine $\sigma_{Fj}^2(\theta) = 1$ in (3.2) $\forall j \leq k$, $\forall \theta \in \Theta$. Then, $m(W, \theta) = \widetilde{m}(W, \theta)$. Correspondingly, the non-normalized CI's $CI_{n,\Delta U}(\alpha)$, $CI_{n,\Delta L}(\alpha)$, and $CI_{n,\Delta}(\alpha)$ in Section 5 and SPUR tests and

CS's in Section 4 are defined as follows. One defines $\widehat{\sigma}^2_{nj}(\theta) = 1$ in (4.2), which yields $\widehat{m}_{nj}(\theta) = \overline{m}_{nj}(\theta)$ in (4.4), (4.5), (4.19)–(4.23), (5.2), and (5.11). One defines $\widehat{\sigma}^{*2}_{njb}(\theta) = 1$ in (4.17), which yields $\widehat{\nu}^*_{njb}(\theta) := n^{1/2}(\overline{m}^*_{njb}(\theta) - \overline{m}_{nj}(\theta))$ in (4.17), (4.18), (4.21), (5.14), and (5.15), and yields $n^{1/2}\overline{m}^*_{njb}(\theta)/\widehat{\sigma}^*_{njb}(\theta) = n^{1/2}\overline{m}^*_{njb}(\theta)$ in the definitions of $sd^*_{anjB}(\theta)$ for $a = 2, 3$ following (4.21) and following (4.22), respectively. Lastly, one defines $\widehat{m}^\sigma_{nj}(W, \theta) = 0_k$ in (5.11), where $0_k = (0, ..., 0)' \in R^k$, and $\widehat{G}^{m\sigma}_{njs}(\theta) = (c'_j, 0'_k)\widehat{\Omega}^{1/2}_{n+}(\theta)Z_s$ in the definition of $\widehat{sd}_{njB}(\theta)$ in (5.9).

With these changes, the asymptotic level results of Theorems 4.1 and 5.1 hold provided the assumptions imposed in the theorems are modified by taking $\sigma^2_{Fj}(\theta) = \widehat{\sigma}^2_{nj}(\theta) = 1$ and the number of moments finite in Assumption A.2 is reduced to $2 + a$ from $4 + a$. In addition, the results of Theorem 11.1 and Lemma 11.2 in online Appendix A for the set estimator $\widehat{\Theta}_n$ also hold in the non-normalized case with the same modifications.

Note that weighted moments also can be employed with non-normalized moments. In this case, the changes outlined above for both of these scenarios need to be employed.

36.4 Alternative $\varphi(\xi)$ Functions

The results of Theorems 4.1, 5.1, and 5.2 and Corollary 5.3 hold not just for tests and CS's based on the GMS function $\varphi(\xi)$ defined in (4.19), but for tests and CS's based on any $\varphi(\xi)$ function that satisfies Assumption A.5 in Andrews and Kwon (2019). See Andrews and Kwon (2019) for the requisite adjustments to the proof of Theorem 4.1. The adjustments for the other results are analogous.

Appendix D References

Chernozhukov, V., H. Hong, and E. Tamer (2007), "Estimation and Confidence Regions for Parameter Sets in Econometric Models," *Econometrica,* 75, 1243–284.