Online Appendix to "Estimating Demand for Differentiated Products with Zeroes in Market Share Data"

In this online appendix, we introduce the profiling approach for models defined by many moment inequalities. The profiling approach developed here is similar to the penalized resampling approach in Bugni, Canay, and Shi (2016) for unconditional moment inequality models.[18] Section D describes the profiling approach and gives the formal results, and Section E presents the proofs of those results.

D The Profiling Approach

The profiling approach applies to general moment inequality models with many moment inequalities. Thus from this point on, we focus on the moment inequality model:

$$E\rho(w_t, \theta, g) \geq 0 \text{ for all } g \in \mathcal{G}, \tag{D.1}$$

where ρ takes values in R^k. We also let $\mathcal{G}$ be a general set of indices that can be either countable or uncontable. Let $\mu : \mathcal{G} \to [0, 1]$ denote a probability density on $\mathcal{G}$. We assume the data $w_{t t=1}^T$ are i.i.d. across t.

We assume that there is a parameter of interest, γ_0, that is related to θ_0 through:

$$\gamma_0 \in \Gamma(\theta_0) \subseteq R^{d_\gamma}, \tag{D.2}$$

where $\Gamma : \Theta \to 2^{R^{d_\gamma}}$ is a known mapping where $2^{R^{d_\gamma}}$ denotes the collection of all subsets of R^{d_γ}. Three examples of Γ are given below:

Example. $\Gamma(\theta) = \{\alpha\}$: γ_0 is the price coefficient α_0. In the simple logit model, the price coefficient is all one needs to know to compute the demand elasticity.

Example. $\Gamma(\theta) = \{e_j(p, \pi, \theta, x) = (\alpha p_j/\pi_j)(\partial\sigma_j(\sigma^{-1}(\pi, x, \lambda), x, \lambda)/\partial\delta_j)\}$: γ_0 is the own-price demand elasticity of product j at a given value of the price vector p, the choice probability vector π and the covariates x.

Example. $\Gamma(\theta) = \{e_j(p, \pi, \theta, x) : \pi \in [\pi^l, \pi^u]\}$: γ_0 is the demand elasticity of product j at a given value of the price vector p, the covariates x and at the choice probability vector that is known to lie between π^l and π^u. This example is particularly useful when the elasticity depends on the choice probability but the choice probability is only known to lie in an interval.

[18] The profiling approach here, made public in the 2013 version of this paper, was developed before that in Bugni, Canay, and Shi (2016).

Let Γ_0 be the identified set of γ_0: $\Gamma_0 = \{\gamma \in R^{d_\gamma} : \exists \theta \in \Theta_0 \ s.t. \ \Gamma(\theta) \ni \gamma\}$, where $\Theta_0 = \{\theta \in \Theta : E\rho(w_t, \theta, g) \geq 0 \ \forall g \in \mathcal{G}\}$. The profiling approach constructs a confidence set for γ_0 by inverting a test of the hypothesis:

$$H_0 : \gamma \in \Gamma_0, \tag{D.3}$$

for each parameter value γ. The confidence set is the collection of values that are not rejected by the test.

Let $\Gamma^{-1}(\gamma) = \{\theta \in \Theta : \Gamma(\theta) \ni \gamma\}$. The test to be inverted uses the *profiled* test statistic:

$$\widehat{T}_T(\gamma) = T \times \min_{\theta \in \Gamma^{-1}(\gamma)} \widehat{Q}_T(\theta), \tag{D.4}$$

where $\widehat{Q}_T(\theta)$ is an empirical measure of the violation to the moment inequalities. The confidence set of confidence level p is the set of all points for which the test statistic does not exceed a critical value $c_T(\gamma, p)$:

$$CS_T = \{\gamma \in R^{d_\gamma} : \widehat{T}_T(\gamma) \leq c_T(\gamma, p)\}. \tag{D.5}$$

Notice that the new confidence set only involves computing a d_γ-dimensional level set, where d_γ is often 1. The profiling transfers the burden of searching (for low values) over the surface of the non smooth function $T(\theta) - c(\theta)$ to searching over the surface of the typically smooth and often convex function $\widehat{Q}_T(\theta)$.

We choose a critical value, $c_T(\gamma, p)$, of significance level $1 - p \in (0, 0.5)$, to satisfy

$$\lim_{T\to\infty} \inf_{(\gamma,F)\in\mathcal{H}_0} \Pr{}_F(\widehat{T}_T(\gamma) > c_T(\gamma, p)) \leq 1 - p, \tag{D.6}$$

where F is the distribution on $(w_t)_{t=1}^T$ and $\mathcal{H}_0$ is the null parameter space of (γ, F). The definition of $\mathcal{H}_0$ along with other technical assumptions are given in Section D.4.[19]

As a result of (D.6), the confidence set asymptotically has the correct minimum coverage probability:

$$\liminf_{T\to\infty} \inf_{(\gamma,F)\in\mathcal{H}_0} \Pr{}_F (\gamma \in CS_T) \geq p. \tag{D.7}$$

The left hand side is called the "asymptotic size" of the confidence set in Andrews and Shi (2013). We achieve the asymptotic size control by deriving an asymptotic approximation for the distribution of the profiled test statistic $\widehat{T}_T(\gamma)$ that is uniformly valid over $(\gamma, F) \in \mathcal{H}_0$ and simulating the critical value from the approximating distribution through either a subsampling or a bootstrapping procedure.

In the next subsections, we describe the test statistic and the critical value in detail and show

[19]Note that we use F to denote the distribution of the full observed data vector and thus (γ, F) captures everything unknown in the expression $\Pr_F(\widehat{T}_T(\gamma) > c_T(\gamma, p))$. This notation differs from the traditional literature where the true distribution of the data is often indicated by the true value of θ, but is standard in the recent partial identification literature. See Romano and Shaikh (2008) and Andrews and Shi (2013).

that (D.7) holds.

D.1 Test Statistic

The test statistic is the QLR statistic (i.e. a criterion-function-based statistic)[20]

$$\widehat{T}_T(\gamma) = T \times \min_{\theta \in \Gamma^{-1}(\gamma)} \widehat{Q}_T(\theta) \text{ with}$$
$$\widehat{Q}_T(\theta) = \int_{\mathcal{G}_T} S(\bar{\rho}_T(\theta, g), \widehat{\Sigma}^{\iota}_T(\theta, g)) d\mu(g), \tag{D.8}$$

where $\mathcal{G}_T$ is a truncated/simulated version of $\mathcal{G}$ such that $\mathcal{G}_T \uparrow \mathcal{G}$ as $T \to \infty$, $\mu(\cdot)$ is a probability measure on $\mathcal{G}$, $S(m, \Sigma)$ is a real-valued function that measures the discrepancy of m from the inequality restriction $m \geq 0$, and

$$\bar{\rho}_T(\theta, g) = T^{-1} \sum_{t=1}^{T} \rho(w_t, \theta, g),$$
$$\widehat{\Sigma}^{\iota}_T(\theta, g) = \widehat{\Sigma}_T(\theta, g) + \iota \times \widehat{\Sigma}_T(\theta, 1)$$
$$\widehat{\Sigma}_T(\theta, g) = T^{-1} \sum_{t=1}^{T} \rho(w_t, \theta, g) \rho(w_t, \theta, g)' - \bar{\rho}_T(\theta, g) \bar{\rho}_T(\theta, g)'. \tag{D.9}$$

In the above definition, ι is a small positive number which is used because in some form of S defined in Section D.4, the inverse of $\widehat{\Sigma}^{\iota}_T(\theta, g)$'s diagonal elements enter, and the ι prevents us from taking inverse of zeros. In some other forms of S, e.g. the one defined below and used in the simulation and empirical section of this paper, the ι does not enter the test statistic because $S(m, \Sigma)$ does not depend on Σ.

Section D.4 gives the assumptions that the user-chosen quantities S, μ, $\mathcal{G}$ and $\mathcal{G}_T$ should satisfy. Under those assumptions, we can show that $\min_{\theta \in \Gamma^{-1}(\gamma)} \widehat{Q}_T(\theta)$ consistently estimate $\min_{\theta \in \Gamma^{-1}(\gamma)} Q_F(\theta)$ where

$$Q_F(\theta) = \int_{\mathcal{G}} S(\rho_F(\theta, g), \Sigma^{\iota}_F(\theta, g)) d\mu(g), \text{ with}$$
$$\rho_F(\theta, g) = E_F(\rho(w_t, \theta, g)),$$
$$\Sigma_F(\theta, g) = Cov_F(\rho(w_t, \theta, g)) \text{ and } \Sigma^{\iota}_F(\theta, g) = \Sigma_F(\theta, g) + \iota \Sigma_F(\theta, 1). \tag{D.10}$$

The symbols "E_F" and "Cov_F" denote expectation and covariance under the data distribution F respectively. Notice that Γ_0 depends on F. We make this explicit by changing the notation Γ_0 to $\Gamma_{0,F}$ for the rest of this paper.

[20]Note that we do not follow the traditional QLR test exactly to define $\widehat{T}_T(\gamma) = T \times \min_{\theta \in \Gamma^{-1}(\theta)} \widehat{Q}_T(\theta) - T \times \min_{\theta \in \Theta} \widehat{Q}_T(\theta)$. This is because the validity of our critical value depends on certain monotonicity of the asymptotic approximation of the test statistic and the monotonicity does not hold with this alternative test statistic due to the subtraction of $T \times \min_{\theta \in \Theta} \widehat{Q}_T(\theta)$.

We can also show that $\min_{\theta \in \Gamma^{-1}(\gamma)} Q_F(\theta) = 0$ if and only if $\gamma \in \Gamma_{0,F}$. This result combined with the consistency of $\min_{\theta \in \Gamma^{-1}(\gamma)} \widehat{Q}_T(\theta)$ implies that $\widehat{T}_T(\gamma)$ diverges to infinity at $\gamma \notin \Gamma_{0,F}$. That implies that there is no information loss in using such a test statistic. Lemma D.1 summarizes those two results. The parameter space $\mathcal{H}$ of (γ, F) appearing in the lemma is defined in Assumption D.2 in Section D.4.

Lemma D.1. Suppose that Assumptions D.1, D.2, D.4, D.5(a), and D.6(a) and (d) hold. Then for any $(\gamma, F) \in \mathcal{H}$,

(a) $\min_{\theta \in \Gamma^{-1}(\gamma)} \widehat{Q}_T(\theta) \to_p \min_{\theta \in \Gamma^{-1}(\gamma)} Q_F(\theta)$ under F, and

(b) $\min_{\theta \in \Gamma^{-1}(\gamma)} Q_F(\theta) \geq 0$ and $= 0$ if and only if $\gamma \in \Gamma_{0,F}$.

In the simulation and the empirical application of this paper, the following choices of S, $\mathcal{G}$, $\mathcal{G}_T$ and μ are used mainly for computational convenience. For $\mathcal{G}$, we use the one defined in (4.19). For $\mathcal{G}_T$, the truncated version of $\mathcal{G}$, we define it to be the same as $\mathcal{G}$ except that we let r run from r_0 to $\bar{r}_T$ where $\bar{r}_T \to \infty$ as $T \to \infty$ in the definition.

For S , we use

$$S(m, \Sigma) = \sum_{j=1}^{k} [m_j]_-^2, \tag{D.11}$$

where m_j is the jth coordinate of m and $[x]_- = |\min\{x, 0\}|$. There may be efficiency loss from not weighting the moments using the variance matrix, but this S function brings great computational convenience because it makes the minimization problem in (D.4) a convex one. For $\mu(\cdot)$, we use

$$\mu(\{g_{a,r,\zeta}\}) \propto (100 + r)^{-2}(2r)^{-d_{z_c}} K_d^{-1} \text{ for } g \in \mathcal{G}_{d,cc}, \tag{D.12}$$

where K_d is the number of elements in $\mathcal{Z}_d$. The same μ measure is used and seems to work well in Andrews and Shi (2013).

D.2 Critical Value

We propose two types of critical values, one based on standard subsampling and the other based on a bootstrapping procedure with moment shrinking. Both are simple to compute. The bootstrap critical value may have better small sample properties, and is the procedure we use in the empirical section.[21] It is worth noting that we resample at the market level for both the subsampling and the bootstrap.

Let us formally define the subsampling critical value first. It is obtained through the standard subsampling steps: [1] from $\{1, ..., T\}$, draw without replacement a subsample of market indices of size b_T; [2] compute $\widehat{T}_{T,b_T}(\gamma)$ in the same way as $\widehat{T}_T(\gamma)$ except using the subsample of markets corresponding to the indices drawn in [1] rather than the original sample; [3] repeat [1]-[2] S_T times obtain S_T independent (conditional on the original sample) copies of $\widehat{T}_{T,b_T}(\gamma)$; [4] let $c^*_{sub}(\gamma, p)$ be

[21]The bootstrap procedure here, like in most problems with partial identification, does not lead to higher-order improvement.

the p quantile of the S_T independent copies. Let the subsampling critical value be

$$c_T^{sub}(\gamma, p) = c_{sub}^*(\gamma, p + \eta^*) + \eta^*, \tag{D.13}$$

where $\eta^* > 0$ is an infinitesimal number. The infinitesimal number is used to avoid making hard-to-verify uniform continuity and strict monotonicity assumptions on the distribution of the test statistic. It can be set to zero if one is willing to make the continuity assumptions. Such infinitesimal numbers are also employed in Andrews and Shi (2013). One can follow their suggestion of using $\eta^* = 10^{-6}$.

Let us now define the bootstrap critical value. It is obtained through the following steps: [1] from the original sample $\{1, ..., T\}$, draw with replacement a bootstrap sample of size T; denote the bootstrap sample by $t_1, ..., t_T$, [2] let the bootstrap statistic be

$$T_T^*(\gamma) = \min_{\theta\in\Theta:\gamma\in\Gamma(\theta)} \int_{\mathcal{G}} S(\widehat{\nu}_T^*(\theta, g) + \kappa_T^{1/2}\bar{\rho}_T(\theta, g), \widehat{\Sigma}_T^{\iota}(\theta, g))d\mu(\mathcal{G}), , \tag{D.14}$$

where $\widehat{\nu}_T^*(\theta, g) = \sqrt{T}(\bar{\rho}_T^*(\theta, g) - \bar{\rho}_T(\theta, g))$, $\bar{\rho}_T^*(\theta, g) = T^{-1}\sum_{\tau=1}^T \rho(X_{t_\tau}, \theta, g)$, and κ_T is a sequence of moment shrinking parameters: $\kappa_T/T + \kappa_T^{-1} \to 0$; [3] repeat [1]-[2] S_T times and obtain S_T independent (conditional on the original sample) copies of $T_T^*(\gamma)$; [4] let $c_{bt}^*(\gamma, p)$ be the p quantile of the S_T copies. Let the bootstrap critical value be

$$c_T^{bt}(\gamma, p) = c_{bt}^*(\gamma, p + \eta^*) + \eta^*, \tag{D.15}$$

where $\eta^* > 0$ is an infinitesimal number which has the same function as in the subsampling critical value above.

Critical values that are not based on resampling are possible, too. For example, one can define a critical value similar to the bootstrap one, except with $\widehat{\nu}_T^*(\theta, g)$ replaced by a Gaussian process with covariance kernel that equals the sample covariance of $\rho(w_t, \theta^{(1)}, g^{(1)})$ and $\rho(w_t, \theta^{(2)}, g^{(2)})$ for $(\theta^{(j)}, g^{(j)}) \in \Theta \times \mathcal{G}, j = 1, 2$. For lack of space, we do not discuss such critical values in detail.

D.3 Coverage Probability

We show that the confidence sets defined in (D.5) using either $c_T^{sub}(\gamma, p)$ and $c_T^{bt}(\gamma, p)$ have asymptotically correct coverage probability uniformly over $\mathcal{H}_0$ under appropriate assumptions. The assumptions are given in Section D.4.

Theorem D.1. Suppose that Assumptions D.1-D.3 and D.5-D.7 hold, then

(a) (D.7) holds with $c_T(\gamma, p) = c_T^{sub}(\gamma, p)$, and

(b) (D.7) holds with $c_T(\gamma, p) = c_T^{bt}(\gamma, p)$.

D.4 Assumptions

In this section, we list all the technical assumptions required for the profiling approach. The assumptions are grouped into seven categories. Assumption D.1 restricts the space of θ; Assumption D.2 restricts the space of (γ, F), i.e. the parameters that determines the true data generating process. Assumption D.3 further restricts the space of (γ, F) to satisfy the null hypothesis $\gamma \in \Gamma_0$. Assumption D.4 is the full support condition on the measure μ on $\mathcal{G}$. Assumption D.5 regulates how $\mathcal{G}_T$ approaches $\mathcal{G}$ as T increases. Assumption D.6 restricts the function $S(m, \Sigma)$ to satisfy certain continuity, monotonicity and convexity conditions. Assumption D.7 regulates the subsample size b_T and the moment shrinking parameter κ_T in the bootstrap procedure. Throughout, we let E^* and E_* denote outer and inner expectations respectively and $\Pr^*$ and $\Pr_*$ denote outer and inner probabilities.

Assumption D.1. (a) Θ is compact, (b) Γ is upper hemi-continuous, and (c) $\Gamma^{-1}(\gamma)$ is either convex or empty for any $\gamma \in R^{d_\gamma}$.

To introduce Assumption D.2 we need the following extra notation. Let $\nu_F(\theta, g) : (\theta, g) \in \Theta \times \mathcal{G}$ denote a tight Gaussian process with covariance kernel

$$\Sigma_F(\theta^{(1)}, g^{(1)}, \theta^{(2)}, g^{(2)}) = Cov_F\left(\rho(w_t, \theta^{(1)}, g^{(1)}), \rho(w_t, \theta^{(2)}, g^{(2)})\right). \tag{D.16}$$

Notice that $\Sigma_F(\theta, g) = \Sigma_F(\theta, g, \theta, g)$.

Let the derivative of $\rho_F(\theta, g)$ with respect to θ be $G_F(\theta, g)$.

For any $\gamma \in R^{d_\gamma}$, let the set $\Theta_{0,F}(\gamma)$ be

$$\Theta_{0,F}(\gamma) = \{\theta \in \Theta : Q_F(\theta) = 0 \ \& \ \Gamma(\theta) \ni \gamma\}, \tag{D.17}$$

We call $\Theta_{0,F}(\gamma)$ the zero-set of $Q_F(\theta)$ under (γ, F). Note that for any $\gamma \in R^{d_\gamma}$, $\gamma \in \Gamma_{0,F}$ if and only if $\Theta_{0,F}(\gamma) \neq \emptyset$.

Let the distance from a point to a set be the usual mapping:

$$d(a, A) = \inf_{a^* \in A} \|a - a^*\|, \tag{D.18}$$

where $\|\cdot\|$ is the Euclidean distance.

Let $\mathcal{F}$ denote the set of all probability measures on $(w_t)_{t=1}^T$. Let $\bar{\mathcal{G}} = \mathcal{G} \cup \{1\}$. Let $\mathcal{M}$ denote the set of all positive semi-definite $k \times k$ matrices. The following assumption defines the parameter space $\mathcal{H}$ for the pair (γ, F).

Assumption D.2. *The parameter space $\mathcal{H}$ of the pairs (γ, F) is a subset of $R^{d_\gamma} \times \mathcal{F}$ that satisfies:*

(a) *under every F such that $(\gamma, F) \in \mathcal{H}$ for some $\gamma \in R^{d_\gamma}$, the markets are independent and ex ante identical to each other, i.e. $\{\rho(w_t, \theta, g)\}_{t=1}^T$ is an i.i.d. sample for any θ, g;*

(b) $\lim_{M\to\infty} \sup_{(\gamma,F)\in\mathcal{H}} E_F^*[\sup_{(\theta,g)\in\Gamma^{-1}(\gamma)\times\bar{\mathcal{G}}} ||\rho(w_t, \theta, g)||^2 1\{||\rho(w_t, \theta, g)||^2 > M\}] = 0;$

(c) *the class of functions* $\{\rho(w_t, \theta, g) : (\theta, g) \in \Gamma^{-1}(\gamma) \times \bar{\mathcal{G}}\}$ *is F-Donsker and pre-Gaussian uniformly over* $\mathcal{H}$*;*

(d) *the class of functions* $\{\rho(w_t, \theta, g)\rho(w_t, \theta, g)' : (\theta, g) \in \Gamma^{-1}(\gamma) \times \bar{\mathcal{G}}\}$ *is Glivenko-Cantelli uniformly over* $\mathcal{H}$*;*

(e) $\rho_F(\theta, g)$ *is differentiable with respect to* $\theta \in \Theta$*, and there exists constants* C *and* $\delta_1 > 0$ *such that, for any* $(\theta^{(1)}, \theta^{(2)})$*,* $\sup_{(\gamma,F)\in\mathcal{H},g\in\bar{\mathcal{G}}} ||vec(G_F(\theta^{(1)}, g)) - vec(G_F(\theta^{(2)}, g))|| \leq C \times ||\theta^{(1)} - \theta^{(2)}||^{\delta_1}$*, and*

(f) $\Sigma_F^\iota(\theta, g) \in \Psi$ *for all* $(\gamma, F) \in \mathcal{H}$ *and* $\theta \in \Gamma^{-1}(\gamma)$ *where* Ψ *is a compact subset of* $\mathcal{M}$*, and* $\{vec(\Sigma_F(\cdot, g^{(1)}, \cdot, g^{(2)})) : (\Gamma^{-1}(\gamma))^2 \to R^{k^2} : (\gamma, F) \in \mathcal{H}, g^{(1)}, g^{(2)} \in \bar{\mathcal{G}}\}$ *are uniformly bounded and uniformly equicontinuous.*

Remark. Part (a) is the i.i.d. assumption, which can be replaced with appropriate weak dependence conditions at the cost of more complicated derivation in the uniform weak convergence of the bootstrap empirical process. Part (b) is standard uniform Lindeberg condition. Part (c)-(d) imposes restrictions on the complexity of the set $\mathcal{G}$ as well as on the shape of $\rho(w_t, \theta, g)$ as a function of θ. A sufficient condition is (i) $\rho(w_t, \theta, g)$ is Lipschitz continuous in θ with the Lipschitz coefficient being integrable and (ii) the set $\mathcal{C}$ in the definition of $\mathcal{G}$ forms a Vapnik-Červonenkis set and J_t is bounded. The Lipschitz continuity is also a sufficient condition of part (f).

The following assumptions defines the null parameter space, $\mathcal{H}_0$, for the pair (γ, F).

Assumption D.3. *The null parameter space* $\mathcal{H}_0$ *is a subset of* $\mathcal{H}$ *that satisfies:*

(a) *for every* $(\gamma, F) \in \mathcal{H}_0$*,* $\gamma \in \Gamma_{0,F}$*, and*

(b) *there exists* $C,\ c > 0$ *and* $2 \leq \delta_2 < 2(\delta_1 + 1)$ *such that* $Q_F(\theta) \geq C \cdot (d(\theta, \Theta_{0,F}(\gamma))^{\delta_2} \wedge c)$ *for all* $(\gamma, F) \in \mathcal{H}_0$ *and* $\theta \in \Gamma^{-1}(\gamma)$*.*

Remark. Part (b) is an identification strength assumption. It requires the criterion function to increase at certain minimum rate as θ is perturbed away from the identified set. This assumption is weaker than the quadratic minorant assumption in Chernozhukov, Hong, and Tamer (2007) if $\delta_2 > 2$ and as strong as the latter if $\delta_2 = 2$. Putting part (b) and Assumption D.2(e) together, we can see that there is a trade-off between the minimum identification strength required and the degree of Hï¿œlder continuity of the first derivative of $\rho_F(\cdot, g)$. If $\rho_F(\cdot, g)$ is linear, δ_2 can be arbitrarily large – the criterion function can increase very slowly as θ is perturbed away from the identified set.

The following assumption is on the measure μ. For any θ, let a pseudo-metric on $\mathcal{G}$ be: $||g^{(1)} - g^{(2)}||_{\theta,F} = ||\rho_{F,j}(\theta, g^{(1)}) - \rho_{F,j}(\theta, g^{(2)})||$. This assumption is needed for Lemma D.1 and not needed for the asymptotic size result Theorem D.1.

Assumption D.4. *For any* $\theta \in \Theta$*,* $\mu(\cdot)$ *has full support on the metric space* $(\mathcal{G}, ||\cdot||_{\theta,F})$*.*

Remark. Assumption D.4 implies that for any $\theta \in \Theta$, F and j, if $\rho_{F,j}(\theta, g_0) < 0$ for some $g_0 \in \mathcal{G}$, then there exists a neighborhood $\mathcal{N}(g_0)$ with positive μ-measure such that $\rho_{F,j}(\theta, g) < 0$ for all $g \in \mathcal{N}(g_0)$.

The following assumption is on the set $\mathcal{G}_T$.

Assumption D.5. (a) $\mathcal{G}_T \uparrow \mathcal{G}$ *as* $T \to \infty$ and

(b) $\limsup_{T\to\infty} \sup_{(\gamma,F)\in\mathcal{H}_0} \sup_{\theta\in\Gamma^{-1}(\gamma)} \int_{\mathcal{G}\backslash\mathcal{G}_T} S(\sqrt{T}\rho_F(\theta,g),\Sigma_F(\theta,g))d\mu(g) = 0$.

The following assumptions are imposed on the function S. For a $\xi > 0$, let the ξ-expansion of Ψ be $\Psi^\xi = \{\Sigma \in \mathcal{M} : \inf_{\Sigma_1\in\Psi} ||vech(\Sigma) - vech(\Sigma_1)|| \le \xi\}$.

Assumption D.6. (a) $S(m,\Sigma) : (-\infty,\infty]^k \times \Psi^\xi \to R$ *is continuous for some* $\xi > 0$.

(b) *There exists a constant* $C > 0$ *and* $\xi > 0$ *such that for any* $m_1, m_2 \in R^k$ *and* $\Sigma_1, \Sigma_2 \in \Psi^\xi$, *we have* $|S(m_1,\Sigma_1) - S(m_2,\Sigma_2)| \le C\sqrt{(S(m_1,\Sigma_1) + S(m_2,\Sigma_2))(S(m_2,\Sigma_2)+1)\Delta}$, *where* $\Delta = ||m_1 - m_2||^2 + ||vech(\Sigma_1 - \Sigma_2)||$.

(c) S *is non-increasing in* m.

(d) $S(m,\Sigma) \ge 0$ *and* $S(m,\Sigma) = 0$ *if and only if* $m \in [0,\infty]^k$.

(e) S *is homogeneous in* m *of degree* 2.

(f) S *is convex in* $m \in R^{d_m}$ *for any* $\Sigma \in \Psi^\xi$.

Remark. We show in the lemma below that Assumption D.6 is satisfied by the example in (D.11) (which is used in our empirical section) as well as the SUM and MAX functions in Andrews and Shi (2013):

$$\begin{aligned} \text{SUM: } S(m,\Sigma) &= \sum_{j=1}^{k} [m_j/\sigma_j]_-^2, \text{ and} \\ \text{MAX: } S(m,\Sigma) &= \max_{1\le j\le k} [m_j/\sigma_j]_-^2, \end{aligned} \tag{D.19}$$

where σ_j^2 is the jth diagonal element of Σ. Assumptions D.6(b) and (f) rule out the QLR function in Andrews and Shi (2013): $S(m,\Sigma) = \min_{t\ge 0}(m-t)'\Sigma^{-1}(m-t)$.

Lemma D.2. (a) *Assumption* D.6 *is satisfied by the* S *function in* (D.11) *for any set* Ψ.

(b) *Assumption* D.6 *is satisfied by the* SUM *and the* MAX *functions in* (D.19) *if* Ψ *is a compact subset of the set of positive semi-definite matrix with diagonal elements bounded below by some constant* $\xi_2 > 0$.

The following assumptions are imposed on the tuning parameters in the subsampling and the bootstrap procedures.

Assumption D.7. (a) *In the subsampling procedure,* $b_T^{-1} + b_T T^{-1} \to 0$ *and* $S_T \to \infty$, *and*

(b)*In the bootstrap procedure,* $\kappa_T^{-1} + \kappa_T T^{-1} \to 0$ *and* $S_T \to \infty$.

D.5 Proof of Lemmas D.1 and D.2

Proof of Lemma D.1. (a) Assumptions D.2(c)-(d) imply that under F,

$$\Delta_{\rho,T} \equiv \sup_{\theta\in\Gamma^{-1}(\gamma),g\in\bar{\mathcal{G}}} ||\bar{\rho}_T(\theta,g)-\rho_F(\theta,g)|| \rightarrow_p 0, \text{ and}$$
$$\sup_{\theta\in\Gamma^{-1}(\gamma),g\in\bar{\mathcal{G}}} ||vech(\widehat{\Sigma}_T(\theta,g)-\Sigma_F(\theta,g))|| \rightarrow_p 0. \tag{D.20}$$

The second convergence implies that

$$\Delta_{\Sigma,T} \equiv \sup_{\theta\in\Gamma^{-1}(\gamma),g\in\mathcal{G}} ||vech(\widehat{\Sigma}^\iota_T(\theta,g)-\Sigma^\iota_F(\theta,g))|| \rightarrow_p 0. \tag{D.21}$$

By Assumption D.2(b), $\sup_{\theta\in\Gamma^{-1}(\gamma),g\in\mathcal{G}} ||\rho_F(\theta,g)|| < M^*$ for some $M^* < \infty$. Thus, $\{(\rho_F(\theta,g),\Sigma^\iota_F(\theta,g)) : (\theta,g)\in\Gamma^{-1}(\gamma)\times\mathcal{G}\}$ is a subset of the compact set $[-M^*,M^*]^k\times\Psi$. By Assumption D.2(f) and Equations (D.20) and (D.21), we have $\{(\bar{\rho}_T(\theta,g),\widehat{\Sigma}^\iota_T(\theta,g)) : (\theta,g)\in\Gamma^{-1}(\gamma)\times\mathcal{G}\} \subseteq [-M^*-\xi, M^*+\xi]^k\times\Psi^\xi$ with probability approaching one for any $\xi>0$. By Assumption D.6(a), $S(m,\Sigma)$ is uniformly continuous on $[-M^*,M^*]^k\times\Psi$. Therefore, for any $\epsilon>0$,

$$\begin{aligned}
&\Pr{}_F\left(\left|\min_{\theta\in\Gamma^{-1}(\gamma)}\widehat{Q}_T(\theta)-\min_{\theta\in\Gamma^{-1}(\gamma)}\int_{\mathcal{G}_T} S(\rho_F(\theta,g),\Sigma^\iota_F(\theta,g))d\mu(g)\right|>\epsilon\right)\\
\le&\Pr{}_F\left(\sup_{\theta\in\Gamma^{-1}(\gamma),g\in\mathcal{G}}|S(\bar{\rho}_t(\theta,g),\widehat{\Sigma}^\iota_t(\theta,g))-S(\rho_F(\theta,g),\Sigma^\iota_F(\theta,g))|>\epsilon\right)\\
\rightarrow&0.
\end{aligned} \tag{D.22}$$

Now it is left to show that $\min_{\theta\in\Gamma^{-1}(\gamma)}\int_{\mathcal{G}_T} S(\rho_F(\theta,g),\Sigma^\iota_F(\theta,g))d\mu(g)\rightarrow\min_{\theta\in\Gamma^{-1}(\gamma)} Q_F(\theta)$ as $T\rightarrow\infty$. Observe that

$$\begin{aligned}
0\le&\min_{\theta\in\Gamma^{-1}(\gamma)} Q_F(\theta)-\min_{\theta\in\Gamma^{-1}(\gamma)}\int_{\mathcal{G}_T} S(\rho_F(\theta,g),\Sigma^\iota_F(\theta,g))d\mu(g)\\
\le&\sup_{\theta\in\Gamma^{-1}(\gamma)}\int_{\mathcal{G}/\mathcal{G}_T} S(\rho_F(\theta,g),\Sigma^\iota_F(\theta,g))d\mu(g)\\
\le&\int_{\mathcal{G}/\mathcal{G}_T}\sup_{\theta\in\Gamma^{-1}(\gamma)} S(\rho_F(\theta,g),\Sigma^\iota_F(\theta,g))d\mu(g).
\end{aligned} \tag{D.23}$$

We have $\sup_{\theta\in\Gamma^{-1}(\gamma)} S(\rho_F(\theta,g),\Sigma^\iota_F(\theta,g))<\infty$, because $\rho_F(\theta,g)\in[-M^*,M^*]^k$ and $\Sigma^\iota_F(\theta,g)\in\Psi$ and Assumption D.6(a). Thus the last line of (D.23) converges to zero under Assumption D.5(a). This and (D.22) together show part (a).

(b) The first half of part (b), $\min_{\theta\in\Gamma^{-1}(\gamma)} Q_F(\theta)\ge 0$, is implied by Assumption D.6(d).

Suppose $\gamma\in\Gamma_{0,F}$. Then there exists a $\theta^*\in\Gamma^{-1}(\gamma)$ such that $\rho_F(\theta^*,g)\ge 0$ for all $g\in\mathcal{G}$. This implies that $S(\rho_F(\theta^*,g),\Sigma_F(\theta^*,g))=0$ for all $g\in\mathcal{G}$ by Assumption D.6(d). Thus, $Q_F(\theta^*)=0$. Because $\min_{\theta\in\Gamma^{-1}(\gamma)} Q_F(\theta)\le Q_F(\theta^*)=0$, this shows the "if" part of the second half.

Suppose that $\min_{\theta\in\Gamma^{-1}(\gamma)} Q_F(\theta)=0$. By Assumptions D.1(a)-(b), $\Gamma^{-1}(\gamma)$ is compact. By

Assumptions D.2(e) and (f), $Q_F(\theta)$ is continuous in θ. Thus, there exists a $\theta^* \in \Gamma^{-1}(\gamma)$ such that $Q_F(\theta^*) = \min_{\theta \in \Gamma^{-1}(\gamma)} Q_F(\theta) = 0$. We show by contradiction that this implies $\gamma \in \Gamma_{0,F}$. Suppose that $\gamma \notin \Gamma_{0,F}$. Then it must be that $\theta^* \notin \Theta_{0,F}$, which implies that $\rho_{F,j}(\theta^*, g^*) < 0$ for some $g^* \in \mathcal{G}$ and some $j \le d_m$. Then by Assumption D.4, there exists a neighborhood $\mathcal{N}(g^*)$ with positive μ-measure, such that $\rho_{F,j}(\theta^*, g) < 0$ for all $g \in \mathcal{N}(g^*)$. This implies that $Q_F(\theta^*) > 0$, which contradicts $Q_F(\theta^*) = 0$. Thus, the "only if" part is proved. □

Proof of Lemma D.2. We prove part (b) only. Part (a) follows from the arguments for part (b) because the S function in part (a) is the same as the SUM S function with $\Sigma = I$. Let ξ be any positive number less than ξ_2. Then the diagonal elements of all matrices in Ψ^ξ are bounded below by $\xi_2 - \xi$.

We prove the SUM part first. Assumptions D.6(a), (c)-(f) are immediate. It suffices to verify Assumptions D.6(b). To verify Assumption D.6(b), observe that

$$\begin{aligned} |S(m_1, \Sigma_1) - S(m_2, \Sigma_2)| &= \left| \sum_{j=1}^{k} ([m_{1,j}/\sigma_{1,j}]_- - [m_{2,j}/\sigma_{2,j}]_-)([m_{1,j}/\sigma_{1,j}]_- + [m_{2,j}/\sigma_{2,j}]_-) \right| \\ &\le \left\{ 2\sum_{j=1}^{k} ([m_{1,j}/\sigma_{1,j}]_- - [m_{2,j}/\sigma_{2,j}]_-)^2 (S(m_1, \Sigma_1) + S(m_2, \Sigma_2)) \right\}^{1/2} \\ &\equiv \{2A(S(m_1, \Sigma_1) + S(m_2, \Sigma_2))\}^{1/2}, \end{aligned} \tag{D.24}$$

where the inequality holds by the Cauchy-Schwartz inequality and the inequality $(a+b)^2 \le 2(a^2 + b^2)$, and the $\equiv$ holds with $A := \sum_{j=1}^{k} ([m_{1,j}/\sigma_{1,j}]_- - [m_{2,j}/\sigma_{2,j}]_-)^2$. Now we manipulate A in the following way:

$$\begin{aligned} A &= \sum_{j=1}^{k} ([m_{1,j}/\sigma_{1,j}]_- - [m_{2,j}/\sigma_{1,j}]_- + [m_{2,j}/\sigma_{1,j}]_- - [m_{2,j}/\sigma_{2,j}]_-)^2 \\ &\le 2\sum_{j=1}^{k} ([m_{1,j}/\sigma_{1,j}]_- - [m_{2,j}/\sigma_{1,j}]_-)^2 + 2\sum_{j=1}^{k} ([m_{2,j}/\sigma_{1,j}]_- - [m_{2,j}/\sigma_{2,j}]_-)^2 \\ &= 2\sum_{j=1}^{k} ([m_{1,j}/\sigma_{1,j}]_- - [m_{2,j}/\sigma_{1,j}]_-)^2 + 2\sum_{j=1}^{k} (\sigma_{2,j} - \sigma_{1,j})^2 [m_{2,j}/\sigma_{2,j}]_-^2 / \sigma_{1,j}^2 \\ &\le 2\|m_1 - m_2\|^2/(\xi_2 - \xi) + 2\{\|vech(\Sigma_1 - \Sigma_2)\|/(\xi_2 - \xi)\} S(m_2, \Sigma_2) \\ &\le 2(\xi_2 - \xi)^{-1}(S(m_2, \Sigma_2) + 1)(||m_1 - m_2||^2 + ||vech(\Sigma_1 - \Sigma_2)||), \end{aligned} \tag{D.25}$$

where the first inequality holds by the inequality $(a+b)^2 \le 2(a^2+b^2)$ and the second inequality holds because $(\sigma_{2,j} - \sigma_{1,j})^2 \le |\sigma_{2,j}^2 - \sigma_{1,j}^2| \le ||vech(\Sigma_1 - \Sigma_2)||$ and because $\sigma_{1,j}^2, \sigma_{2,j}^2 \ge \xi_2 - \xi$. Plug (D.25) in (D.24), and we obtain Assumptions D.6(b).

The proof for the MAX part is the same as the SUM part except some minor changes. The first and obvious change is to replace all $\sum_{j=1}^{k}$ involved in the above arguments by $\max_{j=1,\ldots,k}$. The second change is to replace the Cauchy-Schwartz inequality used in (D.24) by the inequality $|\max_j a_j b_j| \leq (\max_j a_j^2 \times \max_j b_j^2)^{1/2}$. The rest of the arguments stay unchanged. $\square$

E Proof of Theorem D.1

We first introduce the approximation of $\widehat{T}_T(\gamma)$ that connects the distribution of $\widehat{T}_T(\gamma)$ with those of the subsampling statistic and the bootstrap statistic. For any $\theta \in \Theta_{0,F}(\gamma)$, let $\Lambda_T(\theta,\gamma) = \{\lambda : \theta + \lambda/\sqrt{T} \in \Gamma^{-1}(\gamma),\ d(\theta + \lambda/\sqrt{T}, \Theta_{0,F}(\gamma)) = ||\lambda||/\sqrt{T}\}$. In words, $\Lambda_T(\theta,\gamma)$ is the set of all deviations from θ along the fastest paths away from $\Theta_{0,F}(\gamma)$. With this notation handy, we can define the approximation of $\widehat{T}_T(\gamma)$ as follows:

$$T_T^{appr}(\gamma) = \min_{\theta \in \Theta_{0,F}(\gamma)} \min_{\lambda \in \Lambda_T(\theta,\gamma)} \int_{\mathcal{G}} S(\nu_F(\theta,g) + G_F(\theta,g)\lambda + \sqrt{T}\rho_F(\theta,g), \Sigma_F^{\iota}(\theta,g))d\mu(g). \tag{E.1}$$

Theorem E.1 shows that $T_T^{appr}(\gamma)$ approximates $\widehat{T}_T(\gamma)$ asymptotically.

Theorem E.1. *Suppose that Assumptions* D.1-D.3 and D.5-D.6 *hold. Then for any real sequence* $\{x_T\}$ *and scalar* $\eta > 0$ *,*

$$\liminf_{T\to\infty} \inf_{(\gamma,F)\in\mathcal{H}_0} \left[\Pr{}_F(\widehat{T}_T(\gamma) \leq x_T + \eta) - \Pr(T_T^{appr}(\gamma) \leq x_T)\right] \geq 0 \textit{ and}$$
$$\limsup_{T\to\infty} \sup_{(\gamma,F)\in\mathcal{H}_0} \left[\Pr{}_F(\widehat{T}_T(\gamma) \leq x_T) - \Pr(T_T^{appr}(\gamma) \leq x_T + \eta)\right] \leq 0.$$

Theorem E.1 is a key step in the proof of Theorem D.1 and is proved in the next sub-subsection. The remaining proof of Theorem D.1 is given in the subsection after that.

E.1 Proof of Theorem E.1

The following lemma is used in the proof of Theorem E.1. It is a portmanteau theorem for uniform weak approximation, which is an extension of the portmanteau theorem for (pointwise) weak convergence in Chapter 1.3 of van der Vaart and Wellner (1996). Let $(\mathbb{D}, d)$ be a metric space and let BL_1 denote the set of all real functions on $\mathbb{D}$ with a Lipschitz norm bounded by one.

Lemma E.1. (a) *Let* $(\Omega, \mathbb{B})$ *be a measurable space. Let* $\{X_T^{(1)} : \Omega \to \mathbb{D}\}$ *and* $\{X_T^{(2)} : \Omega \to \mathbb{D}\}$ *be two sequences of mappings. Let* $\mathcal{P}$ *be a set of probability measures defined on* $(\Omega, \mathbb{B})$. *Suppose that* $\sup_{P\in\mathcal{P}} \sup_{f\in BL_1} |E_P^* f(X_T^{(1)}) - E_{*,P} f(X_T^{(2)})| \to 0$. *Then for any open set* $G_0 \subseteq \mathbb{D}$ *and closed set* $G_1 \subset G_0$, *we have*

$$\liminf_{T\to\infty} \inf_P \left[\Pr{}_{*,P}(X_T^{(1)} \in G_0) - \Pr{}^*_P(X_T^{(2)} \in G_1)\right] \geq 0 \textit{ and}$$

(b) *Let* $(\Omega, \mathbb{B})$ *be a product space:* $(\Omega, \mathbb{B}) = (\Omega_1 \times \Omega_2, \sigma(\mathbb{B}_1 \times \mathbb{B}_2))$*. Let* $\mathcal{P}_1$ *be a set of probability measures defined on* $(\Omega_1, \mathbb{B}_1)$ *and* P_2 *be a probability measure on* $(\Omega_2, \mathbb{B}_2)$*. Suppose that* $\sup_{P_1 \in \mathcal{P}_1} \Pr^*_{P_1}(\sup_{f \in BL_1} |E^*_{P_2} f(X_T^{(1)}) - E_{*,P_2} f(X_T^{(2)})| > \varepsilon) \to 0$ *for all* $\varepsilon > 0$*. Then for any open set* $G_0 \subseteq \mathbb{D}$ *and closed set* $G_0 \subset G_1$*, we have for any* $\varepsilon > 0$*,*

$$\limsup_{T \to \infty} \sup_{P_1 \in \mathcal{P}_1} \Pr{}^*_{P_1}(\Pr{}^*_{P_2}(X_T^{(1)} \in G_1) - \Pr{}_{*,P_2}(X_T^{(2)} \in G_0) > \varepsilon) = 0.$$

Proof of Lemma E.1. (a) We first show that there is a Lipschitz continuous function sandwiched by $1(x \in G_0)$ and $1(x \in G_1)$. Let $f_a(x) = (a \cdot d(x, G_0^c)) \wedge 1$, where G_0^c is the complement of G_0. Then f_a is a Lipschitz function and $f_a(x) \leq 1(x \in G_0)$ for any $a > 0$. Because G_1 is a closed subset of G_0, $\inf_{x \in G_1} d(x, G_0^c) > c$ for some $c > 0$. Let $a = c^{-1} + 1$. Then $f_a(x) \geq 1(x \in G_1)$. Thus, the function $f_a(x)$ is sandwiched between $1(x \in G_0)$ and $1(x \in F_1)$. Equivalently,

$$a^{-1} 1(x \in G_1) \leq a^{-1} f_a(x) \leq a^{-1} 1(x \in G_0), \ \forall x \in \mathbb{D}. \tag{E.2}$$

By definition, $a^{-1} f_a(x) \in BL_1$. Using this fact and (E.2), we have

$$\begin{aligned}
&a^{-1} \liminf_{T \to \infty} \inf_{P \in \mathcal{P}} \left[\Pr{}_{*,P}(X_T^{(1)} \in G_0) - \Pr{}^*_P(X_T^{(2)} \in G_1)\right] \\
&= \liminf_{T \to \infty} \inf_{P \in \mathcal{P}} [a^{-1} \Pr{}_{*,P}(X_T^{(1)} \in G_0) - E_{*,P} a^{-1} f_a(X_T^{(1)}) + \\
&\qquad E_{*,P} a^{-1} f_a(X_T^{(1)}) - E^*_P a^{-1} f_a(X_T^{(2)}) + E^*_P a^{-1} f_a(X_T^{(2)}) - a^{-1} \Pr{}^*_P(X_T^{(2)} \in G_1)] \\
&\geq \liminf_{T \to \infty} \inf_{P \in \mathcal{P}} \left[E_{*,P} a^{-1} f_a(X_T^{(1)}) - E^*_P a^{-1} f_a(X_T^{(2)})\right] = 0.
\end{aligned} \tag{E.3}$$

Therefore, part (a) is established.

(b) Use the same a and $f_a(x)$ as above, we have

$$\begin{aligned}
\Pr{}^*_{P_2}(X_T^{(1)} \in G_1) - \Pr{}_{*,P_2}(X_T^{(2)} \in G_0) &\leq a \left[E^*_{P_2} a^{-1} f_a(X_T^{(1)}) - E_{*,P_2} a^{-1} f_a(X_T^{(2)})\right] \\
&\leq a \sup_{f \in BL_1} |E_{*,P_2} f(X_T^{(1)}) - E^*_{P_2} f(X_T^{(2)})|.
\end{aligned} \tag{E.4}$$

This implies part (b). □

Proof of Theorem E.1. We only need to show the first inequality because the second one follows from the same arguments with $\widehat{T}_T(\gamma)$ and $T_T^{appr}(\gamma)$ flipped.

The proof consists of four steps. In the first step, we show that the truncation of $\mathcal{G}$ has asymptotically negligible effect: for all $\epsilon > 0$,

$$\limsup_{T \to \infty} \sup_{(\gamma, F) \in \mathcal{H}_0} \Pr{}_F(|\widehat{T}_T(\gamma) - \bar{T}_T(\gamma)| > \epsilon) = 0, \tag{E.5}$$

where $\bar{T}_T(\gamma)$ is the same as $\widehat{T}_T(\gamma)$ except that the integral is over $\mathcal{G}$ instead of $\mathcal{G}_T$.

In the second step, we define a bounded version of $\bar{T}_T(\gamma)$: $\bar{T}_T(\gamma; B_1, B_2)$ and a bounded version of $T_T^{appr}(\gamma)$: $\bar{T}_T^{appr}(\gamma; B_1, B_2)$ and show that for any B_1, $B_2 > 0$ and any real sequence $\{x_T\}$,

$$\liminf_{T\to\infty} \inf_{(\gamma,F)\in\mathcal{H}_0} \left[\Pr{}_F(\bar{T}_T(\gamma; B_1, B_2) \le x_T + \eta) - \Pr(\bar{T}_T^{appr}(\gamma; B_1, B_2) \le x_T)\right] \ge 0. \tag{E.6}$$

In the third step, we show that $\bar{T}_T(\gamma; B_1, B_2)$ is asymptotically close in distribution to $\bar{T}_T(\gamma)$ for large enough B_1, B_2: for any $\epsilon > 0$, there exists $B_{1,\epsilon}$ and $B_{2,\epsilon}$ such that

$$\limsup_{T\to\infty} \sup_{(\gamma,F)\in\mathcal{H}_0} \Pr{}_F(\bar{T}_T(\gamma; B_{1,\epsilon}, B_{2,\epsilon}) \neq \bar{T}_T(\gamma)) < \epsilon. \tag{E.7}$$

In the fourth step, we show that $\bar{T}_T^{appr}(\gamma; B_1, B_2)$ is asymptotically close in distribution to $T_T^{appr}(\gamma)$ for large enough B_1, B_2: for any $\epsilon > 0$, there exists $B_{1,\epsilon}$ and $B_{2,\epsilon}$ such that

$$\limsup_{T\to\infty} \sup_{(\gamma,F)\in\mathcal{H}_0} \Pr{}_F(\bar{T}_T^{appr}(\gamma; B_{1,\epsilon}, B_{2,\epsilon}) \neq T_T^{appr}(\gamma)) < \epsilon. \tag{E.8}$$

The four steps combined proves the Theorem. Now we give detailed arguments of the four steps.

STEP 1. First we show a property of the function S that is useful throughout all steps: for any (m_1, Σ_1) and $(m_2, \Sigma_2) \in R^k \times \Psi^\xi$,

$$|S(m_1, \Sigma_1) - S(m_2, \Sigma_2)| \le C^2 \times (S(m_2, \Sigma_2) + 1)(\Delta + \sqrt{\Delta^2 + 8\Delta})/2, \tag{E.9}$$

for the Δ and C in Assumption D.6(b). Let $\Delta_S := |S(m_1, \Sigma_1) - S(m_2, \Sigma_2)|$. Assumption D.6(b) implies that

$$\begin{aligned}\Delta_S^2 &\le C^2 \times (S(m_1, \Sigma_1) + S(m_2, \Sigma_2))(S(m_2, \Sigma_2) + 1)\Delta \\ &\le C^2 \times (\Delta_S + 2S(m_2, \Sigma_2))(S(m_2, \Sigma_2) + 1)\Delta. \end{aligned} \tag{E.10}$$

Solve the quadratic inequality for Δ_S, we have

$$\begin{aligned}\Delta_S &\le \frac{C^2}{2}\left[(S(m_2, \Sigma_2) + 1)\Delta + \sqrt{(S(m_2, \Sigma_2) + 1)^2\Delta^2 + 8S(m_2, \Sigma_2)(S(m_2, \Sigma_2) + 1)\Delta}\right] \\ &\le \frac{C^2}{2}(S(m_2, \Sigma_2) + 1)(\Delta + \sqrt{\Delta^2 + 8\Delta}) \end{aligned} \tag{E.11}$$

This shows (E.9).

Now observe that

$$
\begin{aligned}
0 &\leq \bar{T}_T(\gamma) - \widehat{T}_T(\gamma) \\
&\leq \sup_{\theta \in \Gamma^{-1}(\gamma)} \int_{\mathcal{G}/\mathcal{G}_T} S(\sqrt{T}\bar{\rho}_T(\theta, g), \widehat{\Sigma}^{\iota}_T(\theta, g)) d\mu(g) \\
&\leq \sup_{\theta \in \Gamma^{-1}(\gamma)} \int_{\mathcal{G}/\mathcal{G}_T} S(\sqrt{T}\rho_F(\theta, g), \Sigma^{\iota}_F(\theta, g)) d\mu(g) + \\
&\sup_{\theta \in \Gamma^{-1}(\gamma)} \int_{\mathcal{G}/\mathcal{G}_T} |S(\sqrt{T}\rho_F(\theta, g), \Sigma^{\iota}_F(\theta, g)) - S(\sqrt{T}\bar{\rho}_T(\theta, g), \widehat{\Sigma}^{\iota}_T(\theta, g))| d\mu(g) \\
&= o(1) + \sup_{\theta \in \Gamma^{-1}(\gamma)} \int_{\mathcal{G}/\mathcal{G}_T} |S(\sqrt{T}\rho_F(\theta, g), \Sigma^{\iota}_F(\theta, g)) - S(\sqrt{T}\bar{\rho}_T(\theta, g), \widehat{\Sigma}^{\iota}_T(\theta, g))| d\mu(g) \\
&\leq o(1) + \sup_{\theta \in \Gamma^{-1}(\gamma)} \int_{\mathcal{G}/\mathcal{G}_T} C^2 \times \left(S(\sqrt{T}\rho_F(\theta, g), \Sigma^{\iota}_F(\theta, g)) + 1 \right) d\mu(g) \times \\
&\qquad \sup_{\theta \in \Gamma^{-1}(\gamma), g \in \mathcal{G}/\mathcal{G}_T} c\left(||\widehat{\nu}_T(\theta, g)||^2 + ||vech(\Sigma^{\iota}_F(\theta, g) - \widehat{\Sigma}^{\iota}_T(\theta, g))|| \right) \\
&= o(1) + o(1) \times c(O_p(1)) \\
&= o_p(1),
\end{aligned} \tag{E.12}
$$

where $c(x) = x + \sqrt{x^2 + 8x}/2$, the third inequality holds by the triangle inequality, the first equality holds by Assumption D.5(b), the fourth inequality holds by (E.9) and the second equality holds by Assumptions D.5(a)-(b) and D.2(c)-(d). The $o(1)$, $o_p(1)$ and $O_p(1)$ are uniform over $(\gamma, F) \in \mathcal{H}$. Thus, (E.5) is shown.

STEP 2. We define the bounded versions of $\bar{T}_T(\gamma)$ as

$$
\begin{aligned}
\bar{T}_T(\gamma; B_1, B_2) = &\min_{\theta \in \Theta_{0,F}(\gamma)} \min_{\lambda \in \Lambda^{B_2}_T(\theta, \gamma)} \\
&\int_{\mathcal{G}} S(\widehat{\nu}^{B_1}_T(\theta + \lambda/\sqrt{T}, g) + G_F(\tilde{\theta}_T, g)\lambda + \sqrt{T}\rho_F(\theta, g), \widehat{\Sigma}^{\iota}_T(\theta + \lambda/\sqrt{T}, g)) d\mu(g)
\end{aligned} \tag{E.13}
$$

where $\Lambda^{B_2}_T(\theta, \gamma) = \{\lambda \in \Lambda_T(\theta, \gamma) : TQ_F(\theta + \lambda/\sqrt{T}) \leq B_2\}$, the process $\widehat{\nu}^{B_1}_T(\cdot, \cdot) = \max\{-B_1, \min\{B_1, \widehat{\nu}_T(\cdot, \cdot)\}\}$ and $\tilde{\theta}_T$ is a value lying on the line segment joining θ and $\theta + \lambda/\sqrt{T}$ satisfying the mean value expansion:

$$
\rho_F(\theta + \lambda/\sqrt{T}, g) = \rho_F(\theta, g) + G_F(\tilde{\theta}_T, g)\lambda/\sqrt{T}. \tag{E.14}
$$

Define the bounded version of $T^{appr}_T(\gamma)$ as

$$
\begin{aligned}
&\bar{T}^{appr}_T(\gamma; B_1, B_2) = \\
&\min_{\theta \in \Theta_{0,F}(\gamma)} \min_{\lambda \in \Lambda^{B_2}_T(\theta, \gamma)} \int_{\mathcal{G}} S(\nu^{B_1}_F(\theta, g) + G_F(\theta, g)\lambda + \sqrt{T}\rho_F(\theta, g), \Sigma^{\iota}_F(\theta, g)) d\mu(g),
\end{aligned} \tag{E.15}
$$

where $\nu_F^{B_1}(\cdot,\cdot) = \max\{-B_1, \min\{B_1, \nu_F(\cdot,\cdot)\}\}$.

First we show a useful result: there exists some constant $\bar{C} > 0$ such that for all $(\gamma, F) \in \mathcal{H}_0$ and $\lambda \in \Lambda_T^{B_2}(\theta,\gamma)$ and for the δ_2 in Assumption D.3(b), we have

$$||\lambda|| \leq \bar{C} \times T^{(\delta_2-2)/(2\delta_2)}. \tag{E.16}$$

This is shown by observing, for all $(\gamma, F) \in \mathcal{H}_0$ and $\lambda \in \Lambda_T^{B_2}(\theta,\gamma)$,

$$\begin{aligned} B_2 >& TQ_F(\theta + \lambda/\sqrt{T}) \\ \geq& C \cdot ((T \times d(\theta + \lambda/\sqrt{T}, \Theta_{0,F}(\gamma))^{\delta_2}) \wedge (c \times T)). \end{aligned} \tag{E.17}$$

The second inequality holds by Assumption (D.3)(b). Because $c \times T$ is eventually greater than B_2 as $T \to \infty$, we have for large enough T,

$$B_2 \geq C \times T \times (||\lambda||/\sqrt{T})^{\delta_2}. \tag{E.18}$$

This implies (E.16).

Equation (E.16) implies two results:

$$\begin{aligned} &(1) \sup_{(\gamma,F)\in\mathcal{H}_0} \sup_{\theta\in\Theta_{0,F}(\gamma)} \sup_{\lambda\in\Lambda_T^{B_2}(\theta,\gamma)} \|\lambda\|/\sqrt{T} \leq O(T^{-1/\delta_2}) = o(1) \\ &(2) \sup_{(\gamma,F)\in\mathcal{H}_0} \sup_{\theta\in\Theta_{0,F}(\gamma)} \sup_{\lambda\in\Lambda_T^{B_2}(\theta,\gamma)} \sup_{g\in\mathcal{G}} \|G_F(\theta + O(\|\lambda\|)/\sqrt{T}, g)\lambda - G_F(\theta,g)\lambda\| \\ &\qquad \leq O(1) \times \|\lambda\|^{\delta_1+1} T^{-\delta_1/2} \leq O(T^{(\delta_2-2(\delta_1+1))/(2\delta_2)}) = o(1). \end{aligned} \tag{E.19}$$

The first result holds immediately given (E.16) and the second result holds by Assumption D.2(e).

Define an intermediate statistic

$$\begin{aligned} \bar{T}_T^{med}(\gamma; B_1, B_2) = & \min_{\theta\in\Theta_{0,F}(\gamma)} \min_{\lambda\in\Lambda_T^{B_2}(\theta,\gamma)} \\ & \int_{\mathcal{G}} S(\widehat{\nu}_T^{B_1}(\theta,g) + G_F(\theta,g)\lambda + \sqrt{T}\rho_F(\theta,g), \Sigma_F^{\iota}(\theta,g))d\mu(g). \end{aligned} \tag{E.20}$$

Then $\bar{T}_T^{med}(\gamma; B_1, B_2)$ and $\bar{T}_T^{appr}(\gamma; B_1, B_2)$ are respectively the following functional evaluated at $\nu_F(\cdot,\cdot)$ and $\widehat{\nu}_T(\cdot,\cdot)$:

$$h(\nu) = \min_{\theta\in\Theta_{0,F}(\gamma)} \min_{\lambda\in\Lambda_T^{B_2}(\theta,\gamma)} \int_{\mathcal{G}} S(\nu^{B_1}(\theta,\cdot) + G_F(\theta,\cdot)\lambda + \sqrt{T}\rho_F(\theta,\cdot), \Sigma_F^{\iota}(\theta,\cdot))d\mu. \tag{E.21}$$

The functional $h(\nu)$ is uniformly bounded for all large enough T because for any fixed $\theta \in \Theta_{0,F}(\gamma)$

and $\lambda \in \Lambda_T^{B_2}(\theta, \gamma)$,

$$\begin{aligned} h(\nu) &\leq 2\int_{\mathcal{G}} S(G_F(\theta,\cdot)\lambda + \sqrt{T}\rho_F(\theta,\cdot), \Sigma_F^{\iota}(\theta,\cdot))d\mu + 2\int_{\mathcal{G}} S(\nu^{B_1}(\theta,\cdot), \Sigma_F^{\iota}(\theta,\cdot))d\mu \\ &\leq 2\sup_{\Sigma\in\Psi} S(-B_1 1_k, \Sigma) + 2\int_{\mathcal{G}} S(G_F(\theta,\cdot)\lambda + \sqrt{T}\rho_F(\theta,\cdot), \Sigma_F^{\iota}(\theta,\cdot))d\mu \\ &\leq 2\sup_{\Sigma\in\Psi} S(-B_1 1_k, \Sigma) + 2T \times Q_F(\theta + \lambda/\sqrt{T}) + \\ &C^2 \times (T \times Q_F(\theta + \lambda/\sqrt{T}) + 1)\sup_{g\in\mathcal{G}}(\Delta_T(g) + \sqrt{\Delta_T(g)^2 + 8\Delta_T(g)}) \\ &\leq 2\sup_{\Sigma\in\Psi} S(-B_1 1_k, \Sigma) + 2B_2 + C^2(B_2+1) \times o(1), \end{aligned} \tag{E.22}$$

where $\Delta_T(g) := \|G_F(\theta, g)\lambda + \sqrt{T}\rho_F(\theta, g) - \sqrt{T}\rho_F(\theta_T, g)\|^2 + \|vech(\Sigma_F^{\iota}(\theta, g) - \Sigma_F^{\iota}(\theta_T, g))\|$ and $\theta_T = \theta + \lambda/\sqrt{T}$. The first inequality holds by Assumptions D.6(e)-(f), the second inequality holds by Assumptions D.2(f) and Assumptions D.6(c), the third inequality holds by (E.9) and the last inequality holds by (E.19).

The functional $h(\nu)$ is Lipschitz continuous for all large enough T with respect to the uniform metric because

$$\begin{aligned} |h(\nu_1) - h(\nu_2)| &\leq 2C \sup_{\theta\in\Theta_{0,F}(\gamma)} \sup_{\lambda\in\Lambda_T^{B_2}(\theta,\gamma)} \sup_{g\in\mathcal{G}} \|\nu_1(\theta,g) - \nu_2(\theta,g)\| \cdot (1 + h(\nu_1) + 2h(\nu_2)) \\ &\leq \bar{C} \sup_{\theta\in\Gamma^{-1}(\gamma), g\in\mathcal{G}} \|\nu_1(\theta,g) - \nu_2(\theta,g)\|, \end{aligned} \tag{E.23}$$

where $\bar{C}$ is any constant such that $\bar{C} > 2C \times (6\sup_{\Sigma\in\Psi} S(-B_1 1_k, \Sigma) + 6B_2 + 1)$, the first inequality holds by Assumption D.6(b) and the second holds by (E.22).

Therefore, for any $f \in BL_1$ and any real sequence $\{x_T\}$, the composite function $f \circ (\bar{C}^{-1}h(\cdot) + x_T) \in BL_1$. By AssumptionD.2(c), we have

$$\limsup_{T\to\infty} \sup_{(\gamma,F)\in\mathcal{H}_0} \sup_{f\in BL_1} |E_F f(\bar{T}_T^{med}(\gamma; B_1, B_2) + x_T) - Ef(\bar{T}_T^{appr}(\gamma; B_1, B_2) + x_T)| = 0. \tag{E.24}$$

This combined with Lemma E.1(a) (with $G_0 = (-\infty, \eta)$ and $G_1 = (-\infty, 0]$) gives

$$\liminf_{T\to\infty} \inf_{(\gamma,F)\in\mathcal{H}_0} \left[\Pr{}_F(\bar{T}_T^{med}(\gamma; B_1, B_2) \leq x_T + \eta) - \Pr(\bar{T}_T^{appr}(\gamma; B_1, B_2) \leq x_T)\right] \geq 0. \tag{E.25}$$

Now it is left to show that $\bar{T}^{med}_T(\gamma; B_1, B_2)$ and $\bar{T}_T(\gamma; B_1, B_2)$ are close. First, we have

$$\begin{aligned}
&|\bar{T}_T(\gamma; B_1, B_2) - \bar{T}^{med}_T(\gamma; B_1, B_2)| \\
\leq & \sup_{\theta\in\Theta_{0,F}(\gamma),\lambda\in\Lambda_T^{B_2}(\theta,\gamma)} \int_{\mathcal{G}} \Big| S(\widehat{\nu}_T^{B_1}(\theta+\lambda/\sqrt{T}, g) + G_F(\tilde{\theta}_T, g)\lambda + \sqrt{T}\rho_F(\theta, g), \widehat{\Sigma}^{\iota}_T(\theta+\lambda/\sqrt{T}, g)) \\
& - S(\widehat{\nu}_T^{B_1}(\theta, g) + G_F(\theta, g)\lambda + \sqrt{T}\rho_F(\theta, g), \Sigma^{\iota}_F(\theta, g)) \Big| d\mu(g) \\
\leq & C^2 \times \sup_{\theta\in\Theta_{0,F}(\gamma),\lambda\in\Lambda_T^{B_2}(\theta,\gamma)} \max_{g\in\mathcal{G}} c(\Delta_T(\theta,\lambda,g)) \times \int_{\mathcal{G}} (1 + M_T(\theta,\lambda,g)) d\mu(g), \qquad \text{(E.26)}
\end{aligned}$$

where $c(x) = (x + \sqrt{x^2 + 8x})/2$, C is the constant in (E.9),

$$\begin{aligned}
\Delta_T(\theta,\lambda,g) =& \|\widehat{\nu}_T^{B_1}(\theta+\lambda/\sqrt{T}, g) - \widehat{\nu}_T^{B_1}(\theta, g) + G_F(\tilde{\theta}_T, g)\lambda - G_F(\theta, g)\lambda\|^2 + \\
& \|vech(\widehat{\Sigma}_T(\theta+\lambda/\sqrt{T}, g) - \Sigma_F(\theta, g))\| \text{ and} \\
M_T(\theta,\lambda,g) =& S(\widehat{\nu}_T^{B_1}(\theta, g) + G_F(\theta, g)\lambda + \sqrt{T}\rho_F(\theta, g), \Sigma^{\iota}_F(\theta, g)). \qquad \text{(E.27)}
\end{aligned}$$

Below we show that for any $\epsilon > 0$, and some universal constant $\bar{C} > 0$,

$$\sup_{(\gamma,F)\in\mathcal{H}_0} \Pr{}_F \left(\sup_{\theta\in\Theta_{0,F}(\gamma),\lambda\in\Lambda_T^{B_2}(\theta,\gamma),g\in\mathcal{G}} \Delta_T(\theta,\lambda,g) > \epsilon \right) \to 0 \text{ and} \qquad \text{(E.28)}$$

$$\sup_T \sup_{(\gamma,F)\in\mathcal{H}_0} \sup_{\theta\in\Theta_{0,F}(\gamma),\lambda\in\Lambda_T^{B_2}(\theta,\gamma)} \int_{\mathcal{G}} M_T(\theta,\lambda,g) d\mu(g) < \bar{C}. \qquad \text{(E.29)}$$

Once (E.28) and (E.29) are shown, it is immediate that for any $\epsilon > 0$,

$$\sup_{(\gamma,F)\in\mathcal{H}_0} \Pr{}_F \left(|\bar{T}_T(\gamma; B_1, B_2) - \bar{T}^{med}_T(\gamma; B_1, B_2)| > \epsilon \right) \to 0. \qquad \text{(E.30)}$$

This combined with (E.25) shows (E.6).

Now we show (E.28) and (E.29). The convergence result (E.28) is implied by the following results: for any $\epsilon > 0$,

$$\begin{aligned}
&\sup_{(\gamma,F)\in\mathcal{H}_0} \Pr{}_F \left(\sup_{\theta\in\Theta_{0,F}(\gamma),\lambda\in\Lambda_T^{B_2}(\theta,\gamma),g\in\mathcal{G}} ||\widehat{\nu}_T^{B_1}(\theta+\lambda/\sqrt{T}, g) - \widehat{\nu}_T^{B_1}(\theta, g)|| > \epsilon \right) \to 0 \\
&\sup_{(\gamma,F)\in\mathcal{H}_0} \sup_{\theta\in\Theta_{0,F}(\gamma),\lambda\in\Lambda_T^{B_2}(\theta,\gamma),g\in\mathcal{G}} ||G_F(\tilde{\theta}_T, g)\lambda - G_F(\theta, g)\lambda|| \to 0 \text{ and} \\
&\sup_{(\gamma,F)\in\mathcal{H}_0} \Pr{}_F \left(\sup_{\theta\in\Theta_{0,F}(\gamma),\lambda\in\Lambda_T^{B_2}(\theta,\gamma),g\in\mathcal{G}} ||vech(\widehat{\Sigma}_T(\theta+\lambda/\sqrt{T}, g) - \Sigma_F(\theta, g))|| > \epsilon \right) \to 0. \qquad \text{(E.31)}
\end{aligned}$$

The first result in the above display holds by the first result in equation (E.19) and the uniform stochastic equicontinuity of the empirical process $\widehat{\nu}_T(\cdot, g) : \Gamma^{-1}(\gamma) \to R^{d_m}$ with respect to the

Euclidean metric. The uniform equicontinuity is implied by Assumptions D.2(b), (c) and (f) by Theorem 2.8.2 of van der Vaart and Wellner (1996). The second result in the above display holds by the second result in (E.19). The third result in (E.31) holds by Assumption D.2(d) and (f).

Result (E.29) holds because for any $\theta \in \Theta_{0,F}(\gamma)$ and $\lambda \in \Lambda_T^{B_2}(\theta,\gamma)$,

$$\begin{aligned}
&\int_{\mathcal{G}} M_T(\theta,\lambda,g)d\mu(g) \\
\leq& 2\int_{\mathcal{G}} S(\widehat{\nu}_T^{B_1}\theta,g),\Sigma_F^{\iota}(\theta,g))d\mu(g) + 2\int_{\mathcal{G}} S(G_F(\theta,g)\lambda + \sqrt{T}\rho_F(\theta,g),\Sigma_F^{\iota}(\theta,g))d\mu(g) \\
\leq& \sup_{\Sigma\in\Psi} S(-B_1 1_k,\Sigma) + 2\int_{\mathcal{G}} S(G_F(\theta,g)\lambda + \sqrt{T}\rho_F(\theta,g),\Sigma_F^{\iota}(\theta,g))d\mu(g) \\
\leq& \sup_{\Sigma\in\Psi} S(-B_1 1_k,\Sigma) + 2B_2 + C^2(B_2+1)\times o(1),
\end{aligned} \tag{E.32}$$

where the first inequality holds by Assumptions D.6(f), the second inequality holds by Assumption D.6(c) and the last inequality holds by the second and third inequality in (E.22) and the $o(1)$ is uniform over (θ,λ).

STEP 3. In order to show (E.7), first extend the definition of $\bar{T}_T(\gamma;B_1,B_2)$ from Step 1 to allow B_1 and B_2 to take the value ∞ and observe that $\bar{T}_T(\gamma;\infty,\infty) = \bar{T}_T(\gamma)$.

Assumptions D.2 (c) and Lemma E.1 imply that for any $\epsilon > 0$, there exists $B_{1,\epsilon}$ large enough such that

$$\limsup_{T\to\infty} \sup_{(\gamma,F)\in\mathcal{H}_0} \Pr{}_F\left(\sup_{\theta\in\Theta,g\in\mathcal{G}} \|\widehat{\nu}_T(\theta,g)\| > B_{1,\epsilon}\right) < \epsilon. \tag{E.33}$$

Therefore we have for all B_2,

$$\limsup_{T\to\infty} \sup_{(\gamma,F)\in\mathcal{H}_0} \Pr{}_F\left(\bar{T}_T(\gamma,\infty,B_2) \neq \bar{T}_T(\gamma;B_{1,\epsilon},B_2)\right) < \epsilon. \tag{E.34}$$

To show that $\bar{T}_T(\gamma)$ and $\bar{T}_T(\gamma;\infty,B_2)$ are close for B_2 large enough, first observe that:

$$\begin{aligned}
\bar{T}_T(\gamma) &\leq \sup_{\theta\in\Theta_{0,F}(\gamma)} \int_{\mathcal{G}} S(\widehat{\nu}_T(\theta,g) + \sqrt{T}\rho_F(\theta,g),\widehat{\Sigma}_T^{\iota}(\theta,g))d\mu(g) \\
&\leq \sup_{\theta\in\Theta_{0,F}(\gamma)} \int_{\mathcal{G}} S(\widehat{\nu}_T(\theta,g),\widehat{\Sigma}_T^{\iota}(\theta,g))d\mu(g) \\
&= O_p(1)
\end{aligned} \tag{E.35}$$

where the first inequality holds because $0 \in \Lambda_T(\theta,\gamma)$, the second inequality holds because $\rho_F(\theta,g) \geq 0$ for $\theta \in \Theta_{0,F}(\gamma)$ and by Assumption D.6(c), the equality holds by Assumption D.6(a)-(c) and Assumptions D.2 (c), (d) and (f). The $O_p(1)$ is uniform over $(\gamma,F) \in \mathcal{H}_0$.

For any T, γ, B_2, if $\bar{T}_T(\gamma) \neq \bar{T}_T(\gamma;\infty,B_2)$, then there must be a $\theta^* \in \Gamma^{-1}(\gamma)$ such that

$T \times Q_F(\theta^*) > B_2$ and

$$\int_{\mathcal{G}} S(\widehat{\nu}_T(\theta^*, g) + \sqrt{T}\rho_F(\theta^*, g), \widehat{\Sigma}_T^{\iota}(\theta^*, g))d\mu(g) < O_p(1). \tag{E.36}$$

But

$$\begin{aligned}
&\int_{\mathcal{G}} S(\widehat{\nu}_T(\theta^*, g) + \sqrt{T}\rho_F(\theta^*, g), \widehat{\Sigma}_T^{\iota}(\theta^*, g))d\mu(g) \\
\geq& 2^{-1}\int_{\mathcal{G}} S(\sqrt{T}\rho_F(\theta^*, g), \widehat{\Sigma}_T^{\iota}(\theta^*, g))d\mu(g) - \int_{\mathcal{G}} S(-\widehat{\nu}_T(\theta^*, g), \widehat{\Sigma}_T^{\iota}(\theta^*, g))d\mu(g) \\
\geq& 2^{-1}\int_{\mathcal{G}} S(\sqrt{T}\rho_F(\theta^*, g), \widehat{\Sigma}_T^{\iota}(\theta^*, g))d\mu(g) - O_p(1) \\
\geq& 2^{-1}\left[TQ_F(\theta^*) - \int_{\mathcal{G}} |S(\sqrt{T}\rho_F(\theta^*, \cdot), \widehat{\Sigma}_T^{\iota}(\theta^*, \cdot)) - S(\sqrt{T}\rho_F(\theta^*, \cdot), \Sigma_F^{\iota}(\theta^*, \cdot))|d\mu\right] - O_p(1) \\
\geq& 2^{-1}\left[TQ_F(\theta^*) - C^2 \sup_{g\in\mathcal{G}} c(||vech(\widehat{\Sigma}_T^{\iota}(\theta^*, g) - \Sigma_F^{\iota}(\theta^*, g))||) \times (1 + TQ_F(\theta^*))\right] - O_p(1) \\
=& B_2/2 - o(1) - o_p(1) \times C^2 \times B_2/4 - O_p(1),
\end{aligned} \tag{E.37}$$

where $c(x) = (x + \sqrt{x^2 + 8x})/2$ and C is the constant in (E.9). The first inequality holds by Assumptions D.6(e)-(f), the second inequality holds by Assumption D.6(c) and Assumptions D.2(c)-(d) and (f), the third inequality holds by the triangle inequality, the fourth inequality holds by (E.9) and the equality holds by Assumption D.2(d). The terms $o(1)$, $o_p(1)$ and $O_p(1)$ terms are uniform over $\theta^* \in \Gamma^{-1}(\gamma)$ and $(\gamma, F) \in \mathcal{H}_0$.

Then

$$\begin{aligned}
&\sup_{(\gamma,F)\in\mathcal{H}_0} \Pr{}_F\left(\widehat{T}_T(\gamma) \neq \bar{T}_T(\gamma; \infty, B_2)\right) \\
\leq& \sup_{(\gamma,F)\in\mathcal{H}_0} \Pr{}_F\left(2^{-1}(1 - o_p(1)) \times B_2 - o(1) - O_p(1) \leq O_p(1)\right) \\
=& \sup_{(\gamma,F)\in\mathcal{H}_0} \Pr{}_F\left(O_p(1) \geq B_2\right),
\end{aligned} \tag{E.38}$$

where the first inequality holds by (E.36) and (E.37). Then for any ϵ, there exists $B_{2,\epsilon}$ such that

$$\lim_{T\to\infty} \sup_{(\gamma,F)\in\mathcal{H}_0} \Pr{}_F(\widehat{T}_T(\gamma) \neq \bar{T}_T(\gamma; \infty, B_{2,\epsilon})) < \epsilon. \tag{E.39}$$

Combining this with (E.34), we have (E.7).

STEP 4. In order to show (E.8), first extend the definition of $\bar{T}_T^{appr}(\gamma; B_1, B_2)$ from Step 1 to allow B_1 and B_2 to take the value ∞ and observe that $\bar{T}_T^{appr}(\gamma; \infty, \infty) = T_T^{appr}(\gamma)$.

By the same arguments as those for (E.34), for any ϵ and B_2, there exists $B_{1,\epsilon}$ large enough so that

$$\limsup_{n\to\infty} \sup_{(\gamma,F)\in\mathcal{H}_0} \Pr{}_F\left(\bar{T}_T^{appr}(\gamma;\infty,B_2) \neq \bar{T}_T^{appr}(\gamma;B_{1,\epsilon},B_2)\right) < \epsilon. \tag{E.40}$$

Also by the same reasons as those for (E.35), we have

$$T_T^{appr}(\gamma) \leq \sup_{\theta\in\Theta_{0,F}(\gamma)} \int_{\mathcal{G}} S(\nu_F(\theta,g),\Sigma_F^{\iota}(\theta,g))d\mu(g), \tag{E.41}$$

where the right hand side is a real-valued random variable.

For any T and B_2, if $T_T^{appr}(\gamma) \neq \bar{T}_T^{appr}(\gamma;\infty,B_{2,\epsilon})$, then there must be a $\theta^* \in \Theta_{0,F}(\gamma)$, a $\lambda^{**} \in \{\lambda \in \Lambda_T(\theta^*,\gamma) : T\times Q_F(\theta^*+\lambda/\sqrt{T}) > B_2\}$ such that

$$I(\lambda^{**}) < \sup_{\theta\in\Theta_{0,F}(\gamma)} \int_{\mathcal{G}} S(\nu_F(\theta,g),\Sigma_F^{\iota}(\theta,g))d\mu(g), \tag{E.42}$$

where $I(\lambda) = \int_{\mathcal{G}} S(\nu_F(\theta^*,g)+G_F(\theta^*,g)\lambda+\sqrt{T}\rho_F(\theta^*,g),\Sigma_F^{\iota}(\theta^*,g))d\mu(g)$. Next we show that if λ^{**} exists, then there must exists a λ^* such that

$$\begin{aligned} &\lambda^* \in \{\lambda \in \Lambda_T(\theta^*,\gamma) : T\times Q_F(\theta^*+\lambda/\sqrt{T}) \in (B_2,2B_2]\} \text{ and} \\ &I(\lambda^*) < \sup_{\theta\in\Theta_{0,F}(\gamma)} \int_{\mathcal{G}} S(\nu_F(\theta,g),\Sigma_F^{\iota}(\theta,g))d\mu(g). \end{aligned} \tag{E.43}$$

If $T\times Q_F(\theta^*+\lambda^{**}/\sqrt{T}) \in (B_2,2B_2]$, then we are done. If $T\times Q_F(\theta^*+\lambda^{**}/\sqrt{T}) > 2B_2$, there must be a $a^* \in (0,1)$ such that $T\times Q_F(\theta^*+a^*\lambda^{**}/\sqrt{T}) \in (B_2,2B_2]$ because $TQ_F(\theta^*+0\times\lambda^{**}/\sqrt{T}) = 0$ and $TQ_F(\theta^*+a\lambda^{**}/\sqrt{T})$ is continuous in a (by Assumptions D.2(e) and D.6(a)). By Assumption D.6(f), $I(\lambda)$ is convex. Thus $I(a^*\lambda^{**}) \leq a^*I(\lambda^{**}) + (1-a^*)I(0)$. For the same arguments as those for (E.35), $I(0) \leq \sup_{\theta\in\Theta_{0,F}(\gamma)} \int_{\mathcal{G}} S(\nu_F(\theta,g),\Sigma_F^{\iota}(\theta,g))d\mu(g)$. Thus, $I(a^*\lambda^{**}) < \sup_{\theta\in\Theta_{0,F}(\gamma)} \int_{\mathcal{G}} S(\nu_F(\theta,g),\Sigma_F^{\iota}(\theta,g))d\mu(g)$. Assumption (D.1)(c) and the definition of $\Lambda_T(\theta,\gamma)$ guarantee that $a^*\lambda^{**} \in \Lambda_T(\theta^*,\gamma)$. Therefore, $\lambda^* = a^*\lambda^{**}$ satisfies (E.43).

Similar to (E.19) we have

$$\begin{aligned} &(1)\ ||\lambda^*||/\sqrt{T} \leq B_2\times 2C\times T^{-1/\delta_2} = B_2\times o(1) \\ &(2)\sup_{g\in\mathcal{G}} ||G_F(\theta^*+O(||\lambda^*||)/\sqrt{T},g)\lambda^* - G_F(\theta^*,g)\lambda^*|| \\ &\qquad \leq O(1)\times B_2^{(\delta_1+1)/\delta_2}||\lambda||^{\delta_1+1}T^{-\delta_1/2} = B_2^{(\delta_1+1)/\delta_2}o(1), \end{aligned} \tag{E.44}$$

where the $o(1)$ terms do not depend on B_2. Then,

$$\begin{aligned}I(\lambda^*) &\geq 2^{-1}\int_{\mathcal{G}} S(G_F(\theta^*,g)\lambda^* + \sqrt{T}\rho_F(\theta^*,g), \Sigma_F^\iota(\theta^*,g))d\mu(g) - \\ &\qquad \int_{\mathcal{G}} S(-\nu_F(\theta^*,g), \Sigma_F^\iota(\theta^*,g))d\mu(g) \\ &\geq TQ_F(\theta^* + \lambda^*/\sqrt{T})/2 - C^2 \times (TQ_F(\theta^* + \lambda^*/\sqrt{T}) + 1) \times c(\Delta_T)/2 + O_p(1) \\ &= TQ_F(\theta^* + \lambda^*/\sqrt{T})/2 - C^2 \times (2B_2 + 1) \times c(\Delta_T)/4 + O_p(1), \end{aligned} \tag{E.45}$$

where the $O_p(1)$ term is uniform over $(\gamma, F) \in \mathcal{H}_0$, $c(x) = (x + \sqrt{x^2 + 8x})/2$ and

$$\begin{aligned}\Delta_T := &\|G_F(\theta^*,g)\lambda^* + \sqrt{T}\rho_F(\theta^*,g) - \sqrt{T}\rho_F(\theta^* + \lambda^*/\sqrt{T},g)\|^2 \\ &+ \|vech(\Sigma_F^\iota(\theta^* + \lambda^*/\sqrt{T},g) - \Sigma_F^\iota(\theta^*,g))\|. \end{aligned} \tag{E.46}$$

The first inequality in (E.45) holds by Assumptions D.6(e)-(f), the second inequality holds by (E.9) and the equality holds by (E.43). By (E.44) and Assumption D.2(f), for any fixed B_2, $\lim_{T\to\infty} \Delta_T = 0$. Therefore, for each fixed B_2,

$$I(\lambda^*) \geq TQ_F(\theta^* + \lambda^*/\sqrt{T})/2 - O_p(1) \geq B_2/2 - O_p(1). \tag{E.47}$$

Thus

$$\begin{aligned}&\sup_{(\gamma,F)\in\mathcal{H}_0} \Pr(T_T^{appr}(\gamma) \neq \bar{T}_T^{appr}(\gamma;\infty,B_2)) \\ &\leq \sup_{(\gamma,F)\in\mathcal{H}_0} \Pr\left(\sup_{\theta\in\Theta_{0,F}(\gamma)} \int_{\mathcal{G}} S(\nu_F(\theta,g), \Sigma_F^\iota(\theta,g))d\mu(g) \geq B_2/2 - O_p(1)\right) \\ &= \sup_{(\gamma,F)\in\mathcal{H}_0} \Pr(O_p(1) \geq B_2). \end{aligned} \tag{E.48}$$

For any $\epsilon > 0$, there exists $B_{2,\epsilon}$ large enough so that $\lim_{T\to\infty} \sup_{(\gamma,F)\in\mathcal{H}_0} \Pr(O_p(1) \geq B_2) < \epsilon$. Thus,

$$\lim_{T\to\infty} \sup_{(\gamma,F)\in\mathcal{H}_0} \Pr(T_T^{appr}(\gamma) \neq \bar{T}_T^{appr}(\gamma;\infty,B_{2,\epsilon}) < \epsilon. \tag{E.49}$$

Combining this with (E.40), we have (E.8). □

E.2 Proof of Theorem D.1

The following lemma is used in the proof of Theorem D.1. It shows the convergence of the bootstrap empirical process $\widehat{\nu}_T^*(\theta, g)$. Let $W_{T,t}$ be the number of times that the tth observation appearing in a bootstrap sample. Then $(W_{T,1}, ..., W_{T,T})$ is a random draw from a multinomial distribution with parameters T and $(T^{-1}, ..., T^{-1})$, and $\widehat{\nu}_T^*(\theta, g)$ can be written as

$$\widehat{\nu}_T^*(\theta, g) = T^{-1/2} \sum_{t=1}^{T} (W_{T,t} - 1)\rho(w_t, \theta, g). \tag{E.50}$$

In the lemma, the subscripts F and W for E and Pr signify the fact that the expectation and the probabilities are taken with respect to the randomness in the data and the randomness in $\{W_{T,t}\}$ respectively.

Lemma E.2. *Suppose that Assumption D.2 holds. Then for any $\epsilon > 0$,*

(a)$\limsup_{T\to\infty} \sup_{(\gamma,F)\in\mathcal{H}} \Pr{}_F^*(\sup_{f\in BL_1} |E_W f(\widehat{\nu}_T^*(\cdot,\cdot)) - Ef(\nu_F(\cdot,\cdot))| > \epsilon) = 0,$

(b) *there exists B_ϵ large enough such that*

$$\limsup_{T\to\infty} \sup_{(\gamma,F)\in\mathcal{H}} \Pr{}_F^* \left(\Pr{}_W \left(\sup_{\theta\in\Gamma^{-1}(\gamma), g\in\bar{\mathcal{G}}} ||\widehat{\nu}_T^*(\theta, g)|| > B_\epsilon \right) > \epsilon \right) = 0, \ \textit{and}$$

(c) *there exists δ_ϵ small enough such that*

$$\limsup_{T\to\infty} \sup_{(\gamma,F)\in\mathcal{H}} \Pr{}_F^* \left(\Pr{}_W \left(\sup_{g\in\bar{\mathcal{G}}} \sup_{||\theta^{(1)}-\theta^{(2)}||\le\delta_\epsilon} ||\widehat{\nu}_T^*(\theta^{(1)}, g) - \widehat{\nu}_T^*(\theta^{(2)}, g)|| > \epsilon \right) > \epsilon \right) = 0.$$

Proof of Lemma E.2. (a) Part (a) is proved using a combination of the arguments in Theorem 2.9.6 and Theorem 3.6.1 in van der Vaart and Wellner (1996). Take a Poisson number N_T with mean T and independent from the original sample. Then $\{W_{N_T,1}, ..., W_{N_T,T}\}$ are i.i.d. Poisson variables with mean one. Let the Poissonized version of $\widehat{\nu}_T^*(\theta, g)$ be

$$\widehat{\nu}_T^{poi}(\theta, g) = T^{-1/2} \sum_{t=1}^{T} (W_{N_T,t} - 1)\rho(w_t, \theta, g). \tag{E.51}$$

Theorem 2.9.6 in van der Vaart and Wellner (1996) is a multiplier central limit theorem that shows that if $\{\rho(w_t, \theta, g) : (\theta, g) \in \Theta \times \bar{\mathcal{G}}\}$ is F-Donsker and pre-Gaussian, then $\widehat{\nu}_T^{poi}(\theta, g)$ converges weakly to $\nu_F(\theta, g)$ conditional on the data in outer probability. The arguments of Theorem 2.9.6 remain valid if we strengthen the F-Donsker and pre-Gaussian condition to the uniform Donsker and pre-Gaussian condition of Assumption D.2(c) and strengthen the conclusion to uniform weak convergence:

$$\limsup_{T\to\infty} \sup_{(\gamma,F)\in\mathcal{H}} \Pr{}_F^* \left(\sup_{f\in BL_1} |E_W f(\widehat{\nu}_T^{poi}(\cdot,\cdot)) - Ef(\nu_F(\cdot,\cdot))| > \varepsilon \right) = 0, \tag{E.52}$$

In particular, the extension to the uniform versions of the first and the third displays in the proof of Theorem 2.9.6 in van der Vaart and Wellner (1996) is straightforward. To extend the second display, we only need to replace Lemma 2.9.5 with Proposition A.5.2 – a uniform central limit theorem for finite dimensional vectors.

Theorem 3.6.1 in van der Vaart and Wellner (1996) shows that, under a fixed (γ, F), the bounded Lipschitz distance between $\widehat{\nu}_T^{poi}(\theta, g)$ and $\widehat{\nu}_T^*(\theta, g)$ converge to zero conditional on (outer) almost all realizations of the data. The arguments remain valid if we strengthen the Glivenko-Cantelli assumption used there to uniform Glivenko-Cantelli (which is implied by Assumption D.2(c)) and strengthen the conclusion to: for all $\varepsilon > 0$

$$\limsup_{T\to\infty} \sup_{(\gamma,F)\in\mathcal{H}} \Pr{}^*_F \left(\sup_{f\in BL_1} |E_W f(\widehat{\nu}_T^{poi}(\cdot,\cdot)) - E_W f(\widehat{\nu}_T^*(\cdot,\cdot))| > \varepsilon \right) = 0, \tag{E.53}$$

Equations (E.52) and (E.53) together imply part (a).

(b) Part (b) is implied by part (a), Lemma E.1(b) and the uniform pre-Gaussianity assumption (Assumption D.2(c)). When applying Lemma E.1(b), consider $X_T^{(1)} = \widehat{\nu}_T^*$, $X_T^{(2)} = \nu_F$, $G_1 = \{\nu : \sup_{\theta,g} \|\nu(\theta, g)\| \geq B_\varepsilon\}$, and $G_2 = \{\nu : \sup_{\theta,g} \|\nu(\theta, g)\| > B_\varepsilon - 1\}$ where B_ε satisfies:

$$\sup_{(\gamma,F)\in\mathcal{H}} \Pr \left(\sup_{\theta\in\Theta, g\in\bar{\mathcal{G}}} \|\nu_F(\theta, g)\| > B_\epsilon - 1 \right) < \epsilon/2. \tag{E.54}$$

Such a B_ε exists because $\{\rho(w_t, \theta, g) : (\theta, g) \in \Theta \times \bar{\mathcal{G}}\}$ is uniformly pre-Gaussian by Assumption D.2(c).

(c) Part (c) is implied by part (a), Lemma E.1(b) and the uniform pre-Gaussianity assumption (Assumption D.2(c)). When applying Lemma E.1(b), consider $X_T^{(1)} = \widehat{\nu}_T^*$, $X_T^{(2)} = \nu_F$, $G_1 = \{\nu : \sup_{||\theta^{(1)}-\theta^{(2)}||\leq\Delta_\varepsilon, g} \|\nu(\theta^{(1)}, g) - \nu(\theta^{(2)}, g)\| \geq \epsilon\}$, and $G_0 = \{\nu : \sup_{||\theta^{(1)}-\theta^{(2)}||\leq\Delta_\varepsilon, g} \|\nu(\theta^{(1)}, g) - \nu(\theta^{(2)}, g)\| > \epsilon/2\}$, where Δ_ε satisfies:

$$\sup_{(\gamma,F)\in\mathcal{H}} \Pr \left(\sup_{||\theta^{(2)}-\theta^{(2)}||\leq\Delta_\varepsilon, g} \|\nu_F(\theta^{(1)}, g) - \nu_F(\theta^{(2)}, g)\| > \epsilon/2 \right) < \epsilon/2. \tag{E.55}$$

Such a Δ_ε exists because $\{\rho(w_t, \theta, g) : (\theta, g) \in \Theta \times \bar{\mathcal{G}}\}$ is uniformly pre-Gaussian. □

Proof of Theorem D.1. (a) Let $q_{b_T}^{appr}(\gamma, p)$ denotes the p quantile of $\bar{T}_{b_T}^{appr}(\gamma)$. Let $\eta_2 = \eta^*/3$. Below we show that,

$$\limsup_{T\to\infty} \sup_{(\gamma,F)\in\mathcal{H}_0} \Pr{}_{F,sub}(c_T^{sub}(\gamma, p) \leq q_{b_T}^{appr}(\gamma, p) + \eta_2) = 0. \tag{E.56}$$

where $\Pr{}^*_{F,sub}$ signifies the fact that there are two sources of randomness in $c_T^{sub}(\gamma, p)$ one from the original sampling and the other from the subsampling. Once (E.56) is established, we have,

$$
\begin{aligned}
&\liminf_{T\to\infty} \inf_{(\gamma,F)\in\mathcal{H}_0} \Pr_{F,sub}\left(\widehat{T}_T(\gamma) \le c_T^{sub}(\gamma,p)\right) \\
&\ge \liminf_{T\to\infty} \inf_{(\gamma,F)\in\mathcal{H}_0} \Pr_F\left(\widehat{T}_T(\gamma) \le q_{b_T}^{appr}(\gamma,p)+\eta_2\right) \\
&\ge \liminf_{T\to\infty} \inf_{(\gamma,F)\in\mathcal{H}_0} \left[\Pr_F\left(\widehat{T}_T(\gamma) \le q_{b_T}^{appr}(\gamma,p)+\eta_2\right) - \Pr\left(T_T^{appr}(\gamma) \le q_{b_T}^{appr}(\gamma,p)\right)\right] \\
&+ \liminf_{T\to\infty} \inf_{(\gamma,F)\in\mathcal{H}_0} \left[\Pr\left(T_T^{appr}(\gamma) \le q_{b_T}^{appr}(\gamma,p)\right) - \Pr\left(T_{b_T}^{appr}(\gamma) \le q_{b_T}^{appr}(\gamma,p)\right)\right] \\
&+ \liminf_{T\to\infty} \inf_{(\gamma,F)\in\mathcal{H}_0} \Pr\left(T_{b_T}^{appr}(\gamma) \le q_{b_T}^{appr}(\gamma,p)\right) \qquad \text{(E.57)}\\
&\ge p,
\end{aligned}
$$

where the first inequality holds by (E.56). The third inequality holds because the first two lim infs after the second inequality are greater than or equal to zero and the third is greater than or equal to p. The first lim inf is greater than or equal to zero by Theorem E.1. The second lim inf is greater than or equal to zero $T_{b_T}^{appr}(\gamma) \ge T_T^{appr}(\gamma)$ for any γ and T which holds because $\sqrt{T} \ge \sqrt{b_T}$ and $\Lambda_{b_T}(\theta,\gamma) \subseteq \Lambda_T(\theta,\gamma)$ for large enough T by Assumptions D.1(c) and D.7(c).

Now it is left to show (E.56). In order to show (E.56), we first show that the c.d.f. of $\bar{T}_{b_T}^{appr}(\gamma)$ is close to the following empirical distribution function:

$$\widehat{L}_{T,b_T}(x;\gamma) = S_T^{-1}\sum_{s=1}^{S_T} 1\left(\widehat{T}_{T,b_T}^s(\gamma) \le x\right). \qquad \text{(E.58)}$$

Define an intermediate quantity first:

$$\tilde{L}_{T,b_T}(x;\gamma) = q_T^{-1}\sum_{l=1}^{q_T} 1\left(\tilde{T}_{T,b_T}^l(\gamma) \le x\right), \qquad \text{(E.59)}$$

where $q_T = \binom{T}{b_T}$ and $(\tilde{T}_{T,b_T}^l(\gamma))_{l=1}^{q_T}$ are the subsample statistics computed using all q_T possible subsamples of size b_T of the original sample. Conditional on the original sample, $(\widehat{T}_{T,b_T}^s(\gamma))_{s=1}^{S_T}$ is S_T i.i.d. draws from $\tilde{L}_{T,b_T}(\cdot;\gamma)$. By the uniform Glivenko-Cantelli theorem, for any $\epsilon > 0$,

$$\limsup_{T\to\infty} \sup_{(\gamma,F)\in\mathcal{H}_0} \Pr_{F,sub}\left(\sup_{x\in R}\left|\tilde{L}_{T,b_T}(x;\gamma) - \widehat{L}_{T,b_T}(x;\gamma)\right| > \epsilon\right) = 0 \qquad \text{(E.60)}$$

It is implied by a Hoeffding's inequality (Theorem A on page 201 of Serfling (1980a)) for U-statistics that for any real sequence $\{x_T\}$, and $\epsilon > 0$,

$$\limsup_{T\to\infty} \sup_{(\gamma,F)\in\mathcal{H}_0} \Pr_F\left(\tilde{L}_{T,b_T}(x_T;\gamma) - \Pr_F\left(\tilde{T}_{T,b_T}^l(\gamma) \le x_T\right) > \epsilon\right) = 0. \qquad \text{(E.61)}$$

Equations (E.60) and (E.61) imply that, for any real sequence $\{x_T\}$ and $\epsilon > 0$,

$$\limsup_{T\to\infty} \sup_{(\gamma,F)\in\mathcal{H}_0} \Pr_{F,sub}\left(\widehat{L}_{T,b_T}(x_T;\gamma) - \Pr_F\left(\tilde{T}^l_{T,b_T}(\gamma) \le x_T\right) > \epsilon\right) = 0. \tag{E.62}$$

Apply Theorem E.1 on the subsample statistic $\tilde{T}^l_{T,b_T}(\gamma)$, and we have for any $\epsilon > 0$ and any real sequence $\{x_T\}$,

$$\limsup_{T\to\infty} \sup_{(\gamma,F)\in\mathcal{H}_0} \left[\Pr_F\left(\tilde{T}^l_{T,b_T}(\gamma) \le x_T - \epsilon\right) - \Pr\left(T^{appr}_{b_T}(\gamma) \le x_T\right)\right] < 0. \tag{E.63}$$

Equations (E.62) and (E.63) imply that for any real sequence $\{x_T\}$,

$$\sup_{(\gamma,F)\in\mathcal{H}_0} \Pr_{F,sub}\left(\widehat{L}_{T,b_T}(x_T;\gamma) > \left(\eta_2 + \Pr\left(T^{appr}_{b_T}(\gamma) \le x_T + \eta_2\right)\right)\right) \to 0. \tag{E.64}$$

Plug $x_T = q^{appr}_{b_T}(\gamma,p) - 2\eta_2$ into the above equation and we have:

$$\limsup_{T\to\infty} \sup_{(\gamma,F)\in\mathcal{H}_0} \Pr^*_{F,sub}\left(\widehat{L}_{T,b_T}(q^{appr}_{b_T}(\gamma,p) - 2\eta_2;\gamma) > \eta_2 + p\right) = 0. \tag{E.65}$$

However, by the definition of $c^{sub}_T(\gamma,p)$, $\widehat{L}_{T,b_T}(c^{sub}_T(\gamma,p) - \eta^*;\gamma) \ge p + \eta^* > \eta_2 + p$. Therefore

$$\limsup_{n\to\infty} \sup_{(\gamma,F)\in\mathcal{H}_0} \Pr^*_{F,sub}\left(\widehat{L}_{T,b_T}(q^{appr}_{b_T}(\gamma,p) - 2\eta_2;\gamma) \ge \widehat{L}_{T,b_T}(c^{sub}_T(\gamma,p) - \eta^*;\gamma)\right) = 0, \tag{E.66}$$

which implies (E.56).

(b) Let $q^{bt}_{\kappa_T}(\gamma,p)$ be the p quantile of $T^{appr}_{\kappa_T}(\gamma)$ conditional on the original sample. Below we show that for $\eta_2 = \eta^*/3$,

$$\limsup_{T\to\infty} \sup_{(\gamma,F)\in\mathcal{H}_0} \Pr_{F,W}(c^{bt}_T(\gamma,p) < q^{bt}_{\kappa_T}(\gamma,p) + \eta_2) = 0. \tag{E.67}$$

where $\Pr_{F,W}$ signifies the fact that there are two sources of randomness in $c^{bt}_T(\gamma,p)$, that from the original sampling and that from the bootstrap sampling. Once (E.67) is established, we have,

$$\begin{aligned}
\liminf_{T\to\infty} \inf_{(\gamma,F)\in\mathcal{H}_0} \Pr_{F,W}\left(\widehat{T}_T(\gamma) \le c^{bt}_T(\gamma,p)\right) &\ge \liminf_{T\to\infty} \inf_{(\gamma,F)\in\mathcal{H}_0} \Pr_F\left(\widehat{T}_T(\gamma) \le q^{bt}_{\kappa_T}(\gamma,p) + \eta_2\right) \\
&\ge \liminf_{T\to\infty} \inf_{(\gamma,F)\in\mathcal{H}_0} \Pr\left(T^{appr}_T(\gamma) \le q^{bt}_{\kappa_T}(\gamma,p)\right) \\
&\ge \liminf_{T\to\infty} \inf_{(\gamma,F)\in\mathcal{H}_0} \Pr\left(T^{appr}_{\kappa_T}(\gamma) \le q^{bt}_{\kappa_T}(\gamma,p)\right) \\
&= p,
\end{aligned} \tag{E.68}$$

where the first inequality holds by (E.67), the second inequality holds by Theorem E.1 and the third inequality holds because $T^{appr}_{\kappa_T}(\gamma) \ge T^{appr}_T(\gamma)$ for any γ and T which holds because $\sqrt{T} \ge \sqrt{\kappa_T}$ and

$\Lambda_{\kappa_T}(\theta,\gamma) \subseteq \Lambda_T(\theta,\gamma)$ for large enough T by Assumptions D.1(c) and D.7(c).

Now we show (E.67). First, we show that the c.d.f. of $T_{\kappa_T}^{appr}(\gamma)$ is close to the following empirical distribution:

$$F_{S_T}(x,\gamma) = S_T^{-1}\sum_{l=1}^{S_T} 1\{T_{T,l}^*(\gamma) \leq x\}, \tag{E.69}$$

where $\{T_{T,1}^*(\gamma),...,T_{T,S_T}^*(\gamma)\}$ are the S_T conditionally independent copies of the bootstrap test statistics. By the uniform Glivenko-Cantelli Theorem, $F_{S_T}(x,\gamma)$ is close to conditional c.d.f. of $T_T^*(\gamma)$: for any $\eta > 0$

$$\limsup_{T\to\infty} \sup_{(\gamma,F)\in\mathcal{H}_0} \Pr{}_{F,W}\left(\sup_{x\in R}|F_{S_n}(x,\gamma) - \Pr{}_W(T_T^*(\gamma)\leq x)| > \eta\right) = 0. \tag{E.70}$$

The same arguments as those for Theorem E.1 can be followed to show that $T_T^*(\gamma)$ is close in law to $T_{\kappa_T}^{appr}(\gamma)$ in the following sense: for any real sequence $\{x_T\}$,

$$\limsup_{T\to\infty} \sup_{(\gamma,F)\in\mathcal{H}_0} \Pr{}_F\left(\left[\Pr{}_W(T_T^*(\gamma)\leq x_T - \eta_2) - \Pr(T_{\kappa_T}^{appr}(\gamma)\leq x_T)\right] \geq \eta_2\right) = 0. \tag{E.71}$$

When following the arguments for Theorem E.1, we simply need to observe the resemblance between $\widehat{T}_T(\gamma)$ and $T_T^*(\gamma)$ in the following form:

$$\begin{aligned} T_T^*(\gamma) = &\min_{\theta\in\Theta_{0,F}(\gamma)} \min_{\lambda\in\Lambda_{\kappa_T}(\theta,\gamma)} \\ &\int_{\mathcal{G}} S(\widehat{\nu}_T^{*+}(\theta+\lambda/\sqrt{T},g) + G_F(\tilde{\theta}_T,g)\lambda + \sqrt{\kappa_T}\rho_F(\theta,g), \widehat{\Sigma}_n(\theta+\lambda/\sqrt{T},g))d\mu(g), \end{aligned} \tag{E.72}$$

where

$$\widehat{\nu}_T^{*+}(\theta,g) = \widehat{\nu}_T^*(\theta,g) + \kappa_T^{1/2} n^{-1/2}\widehat{\nu}_T(\theta,g), \tag{E.73}$$

and use Lemma E.2 in conjunction with Assumptions D.2(c) and use Lemma E.1(b) in place of E.1(a).

Equations (E.70) and (E.71) together imply that for any real sequence $\{x_T\}$,

$$\limsup_{T\to\infty} \sup_{(\gamma,F)\in\mathcal{H}_0} \Pr{}_{F,W}\left(\left[F_{S_T}(x_T-\eta_2,\gamma) - \Pr(T_{\kappa_T}^{appr}(\gamma)\leq x_T)\right] \geq 2\eta_2\right) = 0. \tag{E.74}$$

Plug in $x_T = q_{\kappa_T}^{appr}(\gamma,p) - \eta_2$ and we have

$$\limsup_{T\to\infty} \sup_{(\gamma,F)\in\mathcal{H}_0} \Pr{}_{F,W}\left(F_{S_T}(q_{\kappa_T}^{appr}(\gamma,p) - 2\eta_2,\gamma) \geq p + 2\eta_2\right) = 0. \tag{E.75}$$

But by definition, $F_{S_T}(c_T^{bt}(\gamma,p)-\eta^*,\gamma) \geq p+\eta^* > p+2\eta_2$. Therefore,

$$\limsup_{T\to\infty} \sup_{(\gamma,F)\in\mathcal{H}_0} \Pr{}_{F,W}\left(F_{S_T}(q_{\kappa_T}^{appr}(\gamma,p) - 2\eta_2,\gamma) \geq F_{S_T}(c_T^{bt}(\gamma,p)-\eta^*,\gamma)\right) = 0, \tag{E.76}$$

which implies (E.67). □

References

Andrews, D. W. K., and X. Shi (2013): "Inference Based on Conditional Moment Inequality Models," *Econometrica*, 81.

Bugni, F. A., I. A. Canay, and X. Shi (2016): "Inference for Subvectors and Other Functions of Partially Identified Parameters in Moment Inequality Models," *Quantitative Economics*.

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

Romano, J., and A. Shaikh (2008): "Inference for Identifiable Parameters in Partially Identified Econometric Models," *Journal of Statistical Planning and Inference*, (Special Issue in Honor of T. W. Anderson, Jr. on the Occasion of his 90th Birthday), 138, 2786–2807.

Serfling, R. J. (1980a): *Approximation Theorems in Mathematical Statistics*. John Wiley and Sons, INC.

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