Supplementary Appendix to
"Simple and Trustworthy Cluster-Robust GMM Inference"

Jungbin Hwang[1]
University of Connecticut

This version: March 2020

Appendix B: Supplemental Material

B.1 Clustered (Grouped) dependence in spatial setting

In this subsection, we provide a set of primitive conditions and prove the key conditions used in the main body of the paper to establish the GMM theory under fixed-G asymptotics. The two main results that we investigate are the joint central limit theorem (CLT) condition in Assumption 1-ii) and the uniform law of large number (ULLN) for the Jacobian process in Assumption 3-i). Relating our work to existing literature of spatial econometrics, Conley (1999) considers weakly dependent spatial random fields and develops an asymptotic theory for GMM estimation, but does not have cluster or group structure. Bester, Conley, and Hansen (2011, BCH hereafter), which is closely related to our clustered structure, considers a spatial setting with group structure and makes an essential contribution toward providing a set of regularity conditions that are sufficient to obtain their fixed-G limiting distributions. However, the asymptotic theory developed in BCH (2011) is only applicable to the exactly linear regression model, and it thus is limited to apply our GMM setting with potentially non-linear moment conditions. Also, BCH (2011) assumes the exactly equal cluster size, whereas we allow the cluster sizes to be unbalanced.

We follow Jenish and Prucha (2009, JP hereafter) and consider a generic spatial random filed $\{Y_{i,n}\}_{i\in B_n}$ which is located on a sampling region B_n. We assume that the sampling region B_n has n observations, and B_n is a finite subset of $\mathbb{Z}_v$, where $\mathbb{Z}_v$ is a (possibly unevenly spaced) v-dimensional integer lattice which grows uniformly in v non-opposing directions, as $n\to\infty$. The clustered structure in spatial setting means that the spatial process $\{f(Y_{i,n},\theta), i\in B_n\}$ can be partitioned G sub-sampling regions in $B_n=\cup_{g=1}^{G}\mathcal{G}_{g,n}$, i.e.

$$\{f(Y_{i,n},\theta), i\in B_n\}=\cup_{g=1}^{G}\{f_{i,n}(\theta): i\in\mathcal{G}_{g,n}\}^2.$$

For given sub-sampling region $\mathcal{G}\subseteq B_n$, we use $|\mathcal{G}|$ to refer the number of samples in the region, e.g. $|B_n|=n$ and $|\mathcal{G}_{g,n}|=L_g$. We also assume that $\mathbb{Z}_v$ is equipped with the matrix function $\text{dist}(i,j)=\max\{|i_1-j_1|,\ldots,|i_v-j_v|\}$ for any two $i=(i_1,\ldots,i_v)'$ and $j=(j_1,\ldots,j_v)'$ in $\mathbb{Z}_v$. The distance between any subsets $U,V\subset\mathbb{R}^v$ is then defined $\text{dist}(U,V)=\min\{\text{dist}(i,j); i\in U$ and $j\in V\}$. All elements in B_n are assumed to be located at minimum distance of $d_0\geq 1$ uniformly over sample size n. Without loss of generality, we assume that the minimum distance $d_0=1$. Definition below is the standard notion of a strong mixing coefficient of spatial process a generic random field $\{W_{i,n}\}_{i\in B_n}$ in JP (2009).

Definition B.1 (Spatial mixing coefficients) *For given sub-sampling regions $U\subseteq B_n$ and $V\subseteq B_n$, let $\sigma_n(U)=\sigma(\{W_{i,n}\}_{i\in U})$ be the smallest σ-field generated by $\{W_{i,n}\}_{i\in U}$, and define*

[1] Adress: 365 Fairfield Way U-1063, Storrs, CT 06269-1063, Email: jungbin.hwang@uconn.edu.
[2] Note that $f_{i,n}(\theta)$ with $i\in\mathcal{G}_{g,n}$ is equivalent to $f_i^g(\theta)$, which is used in the main body of our paper.

$\sigma_n(V)$ *similarly using* V. *Then, the* α- *and* ϕ- *mixing coefficients between* $\sigma_n(U)$ *and* $\sigma_n(V)$ *are defined as*

$$\alpha_n(U,V) = \sup_{A\in\sigma_n(U),\ B\in\sigma_n(V)} \{P(A\cap B) - P(A)P(B)\}\,; \tag{B.1}$$

$$\phi_n(U,V) = \sup_{A\in\sigma_n(U),\ B\in\sigma_n(V), P(B)>0} \{P(A|B) - P(A)\}\,. \tag{B.2}$$

Also, the generalized α- *and* ϕ- *mixing coefficients, with* $k, l, d \in \mathbb{N}$, *for a spatial random filed* $\{W_{i,n}\}_{i\in B_n}$ *is defined as*

$$\alpha_{k,l}(d) = \sup_{n\in\mathbb{N}} \sup_{U,V\subseteq B_n} \{\alpha_n(U,V); |U|\le k, |V|\le l,\ dist(U,V)\ge d\}\,; \tag{B.3}$$

$$\phi_{k,l}(d) = \sup_{n\in\mathbb{N}} \sup_{U,V\subseteq B_n} \{\phi_n(U,V); |U|\le k, |V|\le l,\ dist(U,V)\ge d\}\,. \tag{B.4}$$

Given k and l, the α- and ϕ- mixing conditions requires $\alpha_{k,l}(d)$ and $\phi_{k,l}(d)$ to decays zero as $d\to\infty$, respectively.

B.1.1 Primitive conditions for ULLN of Jacobian Process

We begin by imposing the following set of mixing conditions.

Assumption B.1 (α- and ϕ- mixing) *The observations on* B_n, $\{Y_{i,n}\}_{i\in B_n}$, *satisfy the following* α- *and* ϕ- *mixing coefficients*
i) $\sum_{d=1}^{\infty} d^{\nu-1}\alpha_{1,1}(d) < \infty$.
ii) $\sum_{d=1}^{\infty} d^{\nu-1}\phi_{1,1}(d) < \infty$.

The mixing condition in Assumption B.1 allows for general form of weak dependence, together with heteroskedasticity and non-stationarity, among the cross-sectional units $Y_{i,n}$. See, for example, Bolthausen (1982) and JP (2011). Since the mixing conditions are characterized by the entire sampling region B_n, Assumption B.1 allows the dependency across any different subsampling regions, clusters. The mixing conditions are also provided in BCH (2011) to give the spatial LLN and CLT, but only the α-mixing condition is discussed. Given our m-dimensional moment process $f(Y_{i,n},\theta)$ on $\Theta\subseteq\mathbb{R}^d$, let $q_{(j)}(Y_{i,n},\theta)$ be $dm\times 1$ vector valued stochastic function $vec(\partial f(Y_{i,n},\theta)/\theta')$. In addition to Assumption B.1, we further impose the following conditions on $\{q_{(j)}(Y_{i,n},\theta)\}_{i\in B_n}$.

Assumption B.2 *For each* $j = 1,\dots,dm$, *the following conditions hold*:
i) $\sup_{n\in\mathbb{N}}\sup_{i\in B_n} E[d_{(j),i,n}^{1+\delta}] < \infty$ *for some* $\delta>0$, *where* $d_{(j),i,n} := \sup_{\theta\in\Theta}|q_{(j)}(Y_{i,n},\theta)|$.
ii) For all $\theta,\theta'\in\Theta$, *the stochastic function* $q_{(j)}(Y_{i,n},\theta)$ *satisfies*

$$|q_{(j)}(Y_{i,n},\theta) - q_{(j)}(Y_{i,n},\theta')| \le R_{i,n}\cdot h(\theta,\theta') \text{ almost surely,}$$

where $h(\theta,\theta')$ *is a non-random function such that* $h(\theta,\theta')\to 0$ *as* $\theta\to\theta'$, *and* $\{R_{i,n}\}_{i\in B_n}$ *are random variables which do not depend on* θ *such that*

$$\limsup_{n\to\infty}\frac{1}{n}\sum_{i\in B_n} E[R_{i,n}^p] < \infty \text{ for some } p>0.$$

Recalling that $f(Y_{i,n},\theta)$ is Borel-measurable and continuously differentiable function at each $\theta \in \Theta$, for each j, it is easy to check that $\{q_{(j)}(\cdot,\theta) : i \in B_n, n \in \mathbb{N}\}$ is also a well defined families of Borel measurable functions for a given $\theta \in \Theta$. Since the α-mixing and ϕ-mixing coefficient conditions in Assumption B.1 are preserved under any measurable transformation, the random field $\{q_{(j)}(Y_{i,n},\theta)\}_{i\in B_n}$ also satisfies Assumption B.1 at each $\theta \in \Theta$. Assumption B.2-i) implies the uniform $L_{1+\delta}$-boundedness condition of $q_{(j)}(Y_{i,n},\theta)$. In the context of linear IV regression model with $f_{i,n}(\theta_0) = Z_{i,n}(y_{i,n} - x'_{i,n}\theta_0)$, Assumption B.2-i) is guaranteed once we assume that for some $r > 1$

$$\sup_{n\in\mathbb{N}} \sup_{i\in B_n} E[Z^{2r}_{(s),i,n}] < \infty \text{ holds with } s = 1,\ldots,m; \tag{B.5}$$

$$\sup_{n\in\mathbb{N}} \sup_{i\in B_n} E[x^{2r}_{(s),i,n}] < \infty \text{ holds with } s = 1,\ldots,d. \tag{B.6}$$

Assumption B.2-ii) imposes the smoothness condition on the moment process. It is trivially satisfied in the linear moment condition.

Proposition B.2 *Under Assumption B.1-i) or Assumption B.1-ii), and Assumption B.2,*

$$\sup_{\theta\in\Theta} \left\| \frac{1}{L_g} \sum_{i\in\mathcal{G}_{g,n}} \frac{\partial f_i(\theta)}{\partial\theta'} - \Gamma_g(\theta) \right\| \xrightarrow{p} 0$$

holds as $n \to \infty$ holding G fixed, for each $g = 1,\ldots,G$.

The proof of proposition is given in Subsection B.1.3.

B.1.2 Primitive conditions for CLT for Group Means

In this subsection, we provide a joint CLT for the following G-array of (scaled) group means

$$\left(\frac{1}{\sqrt{L_1}} \sum_{i\in\mathcal{G}_{1,n}} f_{i,n}(\theta_0)', \ldots, \frac{1}{\sqrt{L_G}} \sum_{i\in\mathcal{G}_{G,n}} f_{i,n}(\theta_0)' \right)' \in \mathbb{R}^{Gm}, \tag{B.7}$$

which is established under asymptotically unbalanced cluster sizes. The joint CLT focuses on the α- and the ϕ- mixing conditions introduced in the previous section. We impose the following set of mixing and moment conditions.

Assumption B.3 (α-mixing) *The observations on B_n, $\{Y_{i,n}\}_{i\in B_n}$, satisfy the following α-mixing coefficients for some $\delta > 0$.*
i) $\sum_{d=1}^{\infty} d^{\nu-1}[\alpha_{1,1}(d)]^{\frac{\delta}{2+\delta}} < \infty$.
ii) $\sum_{d=1}^{\infty} d^{\nu-1}\alpha_{k,l}(d) < \infty$ *for* $k + l \le 4$.
iii) $\alpha_{1,\infty}(d) = O(d^{-\nu-\delta})$.

Assumption B.4 (ϕ-mixing) *The observations on B_n, $\{Y_{i,n}\}_{i\in B_n}$, satisfy the following ϕ- mixing coefficients*
i) $\sum_{d=1}^{\infty} d^{\nu-1}[\phi_{1,1}(d)]^{1/2} < \infty$.
ii) $\sum_{d=1}^{\infty} d^{\nu-1}\phi_{k,l}(d) < \infty$ *for* $k + l \le 4$.
iii) $\phi_{1,\infty}(d) = O(d^{-\nu-\delta})$ *for some* $\delta > 0$.

Assumption B.5 $\sup_{n\in\mathbb{N}}\sup_{i\in B_n} E[||f_{i,n}(\theta_0)||^{2+\delta}] < \infty$ *for some* $\delta > 0$.

Assumptions B.3 and B.4 are mixing conditions for the spatial random field $\{Y_{i,n}\}_{i\in B_n}$. It is important to point out again that the conditions allow the weak dependence to be presented among different clusters as well as within-cluster. Assumption B.5 is a moment restriction that guarantees the uniform L_2-integrability condition for $\{f_{i,n}(\theta_0)\}_{i\in B_n}$. In the linear IV model, the condition holds if (B.5) and

$$\sup_{n\in\mathbb{N}}\sup_{i\in B_n} E[u_{i,n}^{2r}] < \infty \text{ holds with } u_{i,n} = y_{i,n} - x_{i,n}'\theta_0$$

hold for some $r > 2$. Our Assumptions B.3-B.5 are necessary for establishing CLT of the spatial random field $\{f_{i,n}(\theta_0)\}_{i\in B_n}$, see, e.g., Dedecker (1998) and JP (2009).

Assuming that the size of each cluster L_g grows to infinity but the number of cluster G is fixed, BCH (2011) develops a CLT for a linear regression model. Following BCH (2011), we use the notation $\partial\mathcal{G}_{g,n}$ to refer to the boundary of region $\mathcal{G}_g$ and define it as

$$\partial\mathcal{G}_{g,n} = \{i \in \mathcal{G}_{g,n};\ \text{There exists } j \neq \mathcal{G}_{h,n} \text{ such that } \text{dist}(i,j) = d_0 \text{ and } g \neq h\}.$$

Assumption B.6 *i) Groups are mutually exclusive and exhaustive. ii) Groups are contiguous in the metric distance function dist*$(\cdot,\cdot)$. *iii) For all* $g = 1,\ldots,G$, $|\partial\mathcal{G}_{g,n}| < C\bar{L}^{\frac{\nu-1}{\nu}}$ *where* $\bar{L} = G^{-1}\sum_{g=1}^{G} L_g$.

Assumption B.7 *For all* $g = 1,\ldots,G$,

$$\lim_{L_g\to\infty} var\left(\frac{1}{\sqrt{L_g}}\sum_{i\in\mathcal{G}_{g,n}} f_{i,n}(\theta_0)\right) = \Omega_g > 0.$$

Assumption B.6 is about geographical restrictions on clustered sampling regions $\{\mathcal{G}_1,\ldots,\mathcal{G}_G\}$. Assumption B.7 guarantees that each group has asymptotically non-negligible variations in the limits. Parts i)–iii) in Assumption B.6 is similar to conditions i)–iv) in Assumption 2 of BCH (2011), but our condition iii) substantially relaxes the conditions iii)-iv) in BCH (2011) because it does not require the equivalent cluster sizes, $L_1 = L_2 = \ldots = L_G = \bar{L}$. In fact, what we need to establish the joint CLT for the random array in (B.7) is each cluster size growing at the same rate. The corresponding cluster sizes are allowed to be unbalanced, even asymptotically. The following proposition formally provides the joint asymptotic normality and independence of cluster means.

Proposition B.3 *Under Assumption B.3 or Assumption B.4, and Assumptions B.6 –B.7,and* $n \to \infty$ *such that* G *fixed with* $L_g/n \to \lambda_g > 0$, *we have*

$$\begin{pmatrix} \frac{1}{\sqrt{L_1}}\sum_{i\in\mathcal{G}_{1,n}} f_{i,n}(\theta_0) \\ \vdots \\ \frac{1}{\sqrt{L_G}}\sum_{i\in\mathcal{G}_{G,n}} f_{i,n}(\theta_0) \end{pmatrix} \xrightarrow{d} N\left(0, \begin{pmatrix} \Omega_1 & & 0 \\ & \ddots & \\ 0 & & \Omega_G \end{pmatrix}\right). \tag{B.8}$$

The proof of proposition is given in Subsection B.1.3. It proceeds by showing that the G-random array in (B.7) after rescaling converges to multivariate normal distribution, and the corresponding variance covariance matrix is block diagonal in the limit. Our proof of proposition extensively follows from those in BCH (2010), but the difference is that we allow for asymptotically unbalanced cluster sizes. Also, we explicitly investigate the joint asymptotic normality of the G-random array in (B.7). This makes our proof more straightforward than those of BCH (2011) who explain the joint convergence only through the marginal convergences of each single summation in (B.7).

For $g_n(\theta) = n^{-1}\sum_{i\in B_n} f_{i,n}(\theta)$, then the total sum of moment process can be decomposed into the following form of cluster sums

$$\sqrt{n}g_n(\theta_0) = \sum_{g=1}^{G}\sqrt{\frac{L_g}{n}}\frac{1}{\sqrt{L_g}}\sum_{i\in\mathcal{G}_{g,n}} f_{i,n}(\theta_0).$$

Then, the result in Proposition B.3 together with continuous mapping theorem implies the following "fixed" cluster Central Limit Theorem (CLT):

$$\sqrt{n}g_n(\theta_0) \xrightarrow{d} \sum_{g=1}^{G}\lambda_g\Lambda_g B_{m,g} \overset{d}{=} N(0,\Omega), \tag{B.9}$$

where $\Omega = \sum_{g=1}^{G}\lambda_g\Omega_g$.

B.1.3 Proof of Propositions B.2 and B.3

Proof of Proposition B.2. Let $\Gamma_g^{(j)}(\theta)$ be the j-th element in $vec(\Gamma_g(\theta))$. It then suffices to show that the following ULLN holds for each $j = 1,\dots,dm$,

$$\sup_{\theta\in\Theta}\left|\frac{1}{L_g}\sum_{i\in\mathcal{G}_{g,n}} q_{(j)}(Y_{i,n},\theta) - \Gamma_g^{(j)}(\theta)\right| \xrightarrow{p} 0. \tag{B.10}$$

We prove (B.10) using Theorem 2-(a) in JP (2009). Setting the non-random constants $c_{i,n} = 1$ in JP (2009), we check that (B.10) holds if we show

$$\lim_{M\to\infty}\limsup_{L_g\to\infty}\left(\frac{1}{L_g}\sum_{i\in\mathcal{G}_{g,n}} E[d^p_{(j),i,n}1(d_{(j),i,n} > M)]\right) = 0 \text{ for some } p\geq 1 \text{ (Domination);} \tag{B.11}$$

$$\left|\frac{1}{L_g}\sum_{i\in\mathcal{G}_{g,n}} q_{(j)}(Y_{i,n},\theta) - \Gamma_g^{(j)}(\theta)\right| \xrightarrow{p} 0 \text{ for each } \theta\in\Theta \text{ (Pointwise LLN);} \tag{B.12}$$

and the spatial random process $\{q_{(j)}(Y_{i,n},\theta) : i\in\mathcal{G}_{g,n}, n\in\mathbb{N}\}$ satisfy

$$\limsup_{n}\frac{1}{L_g}\sum_{i\in\mathcal{G}_{g,n}} P\left(\sup_{\theta'\in\Theta}\sup_{\theta\in B(\theta',\delta)}\left|q_{(j)}(Y_{i,n},\theta) - q_{(j)}(Y_{i,n},\theta') > \epsilon\right.\right) \to 0, \text{ as } \delta\to 0. \tag{B.13}$$

(L_0 stochastically equicontinuity on Θ)

We begin by showing the domination condition in (B.11). It is not difficult to check that (B.11) is implied by

$$\lim_{M\to\infty}\sup_{n\in\mathbb{N}}\sup_{i\in\mathcal{G}_{g,n}} E\left[d^p_{(j),i,n}1(d_{(j),i,n} > M)\right] = 0 \text{ for some } p\geq 1. \tag{B.14}$$

Using Assumption B.2-i), we can choose $p = 1$ such that

$$\begin{aligned}
&\sup_{n\in\mathbb{N}} \sup_{i\in\mathcal{G}_{g,n}} E\left[d_{(j),i,n} 1(d_{(j),i,n} > M)\right] \\
&\leq \sup_{n\in\mathbb{N}} \sup_{i\in\mathcal{G}_{g,n}} E\left[M^{-\delta} d_{(j),i,n}^{1+\delta} 1(d_{(j),i,n} > M)\right] \\
&\leq \sup_{n\in\mathbb{N}} \sup_{i\in B_n} E\left[d_{(j),i,n}^{1+\delta}\right] \cdot M^{-\delta} \to 0 \text{ as } M \to \infty,
\end{aligned}$$

which gives us the desired result in (B.11).

To show the pointwise LLN in (B.12), we note that the generalized α- and ϕ- mixing coefficients for the sub-sampling region $\{Y_{i,n}\}_{i\in\mathcal{G}_{g,n}}$ can be easily specified by the definitions in (B.1)–(B.3), and check that the mixing conditions in Assumption B.1 are also satisfied for each g-th cluster $\{Y_{i,n}\}_{i\in\mathcal{G}_{g,n}}$. Moreover, the mixing conditions in Assumption B.1 for $\{Y_{i,n}\}_{i\in\mathcal{G}_{g,n}}$ are preserved in its measurable transformation $\{q_{(j)}(Y_i, \theta)\}_{i\in\mathcal{G}_{g,n}}$, which indicates that the condition (a) or (b) in Theorem 3 in JP (2009) holds. Also, it is easy to check that our Assumption B.2-i) directly satisfies the Assumption 2* in JP (2009), both of which state the pointwise uniform L_1 integrability of $q_{(j)}(Y_{i,n}, \theta)$. Therefore, we can apply Theorem 3 in JP (2009) and obtain the pointwise LLN in (B.12).

Lastly, Proposition 1 in JP (2009) directly gives that our Assumption B.2-ii) implies the L_0 stochastically equicontinuity condition in (B.13). ■

Proof of Proposition B.3. We want to show that the following joint CLT

$$\begin{pmatrix} \frac{1}{\sqrt{n}} \sum_{i\in\mathcal{G}_{1,n}} f_{i,n}(\theta_0) \\ \vdots \\ \frac{1}{\sqrt{n}} \sum_{i\in\mathcal{G}_{G,n}} f_{i,n}(\theta_0) \end{pmatrix} \xrightarrow{d} N\left(0, \begin{pmatrix} \lambda_1\Omega_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_G\Omega_G \end{pmatrix}\right), \tag{B.15}$$

which, together with $L_g/n \to \lambda_g > 0$, implies the joint CLT result in (B.8). For any non-zero $t = (t_1', \ldots, t_G') \in \mathbb{R}^{Gm}$ with $t_g \in \mathbb{R}^m$ for $g = 1, \ldots, G$, define

$$\begin{aligned}
t' S_n(\theta_0) &= \sum_{i\in\mathcal{G}_{1,n}} t_1' f_{i,n}(\theta_0) + \ldots + \sum_{i\in\mathcal{G}_{G,n}} t_G' f_{i,n}(\theta_0); \\
\sigma_{n,t}^2 &= var(t' S_n(\theta_0)).
\end{aligned}$$

Also, for $g, h = 1, \ldots, G$, let

$$V_{gh,n} = E\left[\left(\sum_{i\in\mathcal{G}_{g,n}} f_{i,n}(\theta_0)\right)\left(\sum_{i\in\mathcal{G}_{h,n}} f_{i,n}(\theta_0)\right)'\right].$$

Then, we can express

$$\frac{\sigma_{n,t}^2}{n} = \sum_{g=1}^{G} \left(\frac{L_g}{n}\right) \cdot t_g' \left(\frac{V_{gg,n}}{L_g}\right) t_g + \frac{1}{n} \sum_{g\neq h} t_g' V_{gh,n} t_h. \tag{B.16}$$

By Cramer-Wold device and Slutsky's theorem, (B.15) is implied by

$$\frac{\sigma_{n,t}^2}{n} \to \sum_{g=1}^{G} \lambda_g (t_g' \Omega_g t_g) > 0; \tag{B.17}$$

$$\frac{t' S_n(\theta_0)}{\sigma_{n,t}} \xrightarrow{d} N(0,1). \tag{B.18}$$

Without loss of generality, we consider the case when $\{f_{i,n}(\theta_0)\}_{i \in B_n}$ is a scalar random field. The vector case can be dealt with making an arbitrary linear combination of $f_{i,n}(\theta_0)$ and repeating the Cramer-Wold device.

We first prove (B.17) for α-mixing random field. In view of (B.16), Assumption B.7 and $L_g/n \to \lambda_g > 0$, (B.17) holds if

$$\frac{1}{n}|V_{gh,n}| \leq \frac{1}{n} \sum_{i \in \mathcal{G}_{g,n}} \sum_{j \in \mathcal{G}_{h,n}} |E f_{i,n}(\theta_0) f_{j,n}(\theta_0)| \to 0 \tag{B.19}$$

for any $g \neq h$. Define the set of d-th order neighbors located on two different clusters $g \neq h$,

$$N_{g,h}(d) = \{(i,j) \in B_n \times B_n : \mathrm{dist}(i,j) = d,\ i \in \mathcal{G}_{g,n} \text{ and } j \in \mathcal{G}_{h,n}\}.$$

Under Assumption B.5, we can use a standard α-mixing inequality, e.g. Lemma 1 in Bolthausen (1982), and obtain

$$|E f_{i,n}(\theta_0) f_{j,n}(\theta_0)| \leq \Delta \cdot [\alpha_{1,1}(d)]^{\frac{\delta}{2+\delta}} \tag{B.20}$$

for any $(i,j) \in N_{g,h}(d)$, where Δ is a positive constant. Also, using the same arguments as the proof of Lemma 1 in BCH (2011, p.149), we can check that the geographic restrictions in Assumption B.6 imply that

$$|N_{g,h}(d)| \leq C_\nu \cdot \bar{L}^{\frac{\nu-1}{\nu}} \cdot d \tag{B.21}$$

for some C_v which only depends on the dimension of the index set ν. That is, the maximum number of pairs in d-th order neighbor set is bounded by $C_\nu \cdot \bar{L}^{\frac{\nu-1}{\nu}} \cdot d$. Combining the results in (B.20) and (B.21), we obtain

$$\begin{aligned}
&\sum_{i \in \mathcal{G}_{g,n}} \sum_{j \in \mathcal{G}_{h,n}} |E f_{i,n}(\theta_0) f_{j,n}(\theta_0)'| \\
&= \sum_{d=1,(i,j) \in N_{g,h}(d)}^{\infty} |E f_{i,n}(\theta_0) f_{j,n}(\theta_0)'| \\
&\leq O\left(\bar{L}^{\frac{\nu-1}{\nu}}\right) \cdot \sum_{d=1}^{\infty} d[\alpha_{1,1}(d)]^{\frac{\delta}{2+\delta}},
\end{aligned}$$

Together with Assumption B.3-i), the above inequality leads to

$$\begin{aligned}
&\frac{1}{n} \sum_{i \in \mathcal{G}_{g,n}} \sum_{j \in \mathcal{G}_{h,n}} E|f_{i,n}(\theta_0) f_{j,n}(\theta_0)| \\
&\leq \frac{1}{G} O\left(\frac{1}{\bar{L}^{\nu}}\right) \cdot \sum_{d=1}^{\infty} d[\alpha_{1,1}(d)]^{\frac{\delta}{2+\delta}} = o(1),
\end{aligned}$$

which is the desired result in (B.19). The proof in the case of ϕ-mixing random field follows immediately using the same arguments as above, but replacing (B.20) by the following ϕ-mixing inequality, e.g., Hall and Heyde (1980, p. 277)

$$|Ef_{i,n}(\theta_0)f_{j,n}(\theta_0)| \leq \Delta \cdot [\phi_{1,1}(d)]^{1/2}.$$

Next, we prove the result in (B.18). It is straightforward to check that $\cup_{g=1}^{G}\{\lambda_g t_g f_{i,n}^{g}(\theta_0) : i \in \mathcal{G}_{g,n}\}$ is a measurable transformation of the original process $\{f(Y_{i,n}, \theta_0), i \in B_n\}$, so it also satisfies the α-mixing [ϕ-mixing] conditions in Assumption B.3 [Assumption B.4]. Then, the (joint) CLT for the α-mixing fields $\cup_{g=1}^{G}\{\lambda_g t_g f_{i,n}^{g}(\theta_0) : i \in \mathcal{G}_{g,n}\}$ follows from Corollary 1 in JP (2009) by setting their non-random constants $c_{i,n} = 1$, if the conditions below are satisfied for some $\delta > 0$:

$$\lim_{M\to\infty} \sup_{n} \sup_{i\in B_n} E[|f_{i,n}(\theta_0)|^{2+\delta} 1(|f_{i,n}(\theta_0)| > M)] = 0; \tag{B.22}$$

$$\liminf_{n\to\infty} \frac{\sigma_{n,t}^2}{n} > 0; \tag{B.23}$$

$$\sum_{d=1}^{\infty} \alpha_{1,1}(d) d^{\frac{\nu(2+\delta)}{\delta}-1} < \infty \ . \tag{B.24}$$

It is not difficult to check that (B.22) is implied by our Assumption B.5. Also, (B.23) is proved by (B.17). To show that our Assumption B.3-i) implies (B.24), note that the ratio of two summands in Assumption B.3-i) and (B.24) is equal to

$$\frac{\alpha_{1,1}(d) d^{\frac{\nu(2+\delta)}{\delta}-1}}{d^{\nu-1}\alpha_{1,1}(d)^{\frac{\delta}{2+\delta}}} = \left(d^{\nu}\alpha_{1,1}(d)^{\frac{\delta}{2+\delta}}\right)^{\frac{2}{\delta}}. \tag{B.25}$$

If Assumption B.3-i) holds, we check that

$$d^{\nu}\alpha_{1,1}(d)^{\frac{\delta}{2+\delta}} \to 0,$$

so the ratio in (B.25) converges to zero as well. Thus, given $\epsilon > 0$, our Assumption B.3-i) implies that there exists $m \in \mathbb{N}$ such that

$$\sum_{d=m}^{\infty} \alpha_{1,1}(d) d^{[\nu(2+\delta)/\delta]-1} < \epsilon \cdot \sum_{d=m}^{\infty} d^{\nu-1}\alpha_{1,1}(d)^{\delta/(2+\delta)} < \infty,$$

which leads to the desired result in (B.24). For the ϕ-mixing case, we check that our Assumption B.4 and Assumption B.5 directly satisfy the Assumptions 4 and 2 for Theorem 1 in JP (2009), respectively, which leads to the (joint) CLT for the ϕ-mixing random field $\cup_{g=1}^{G}\{\lambda_g t_g f_{i,n}^{g}(\theta_0) : i \in \mathcal{G}_{g,n}\}$. This completes the proof of (B.18). ■

B.2 Proof of Proposition 8

Proof. Let $U\Sigma V'$ be a singular value decomposition (SVD) of Γ_Λ. By construction, $U'U = UU' = I_m$, $V'V = V'V = I_d$, and

$$\Sigma = \begin{bmatrix} A_{d\times d} \\ O_{q\times d} \end{bmatrix},$$

where A is a diagonal matrix. Then,

$$\begin{aligned} R\left[\Gamma_\Lambda'\bar{\mathbb{S}}^{-1}\Gamma_\Lambda\right]^{-1}\Gamma_\Lambda'\bar{\mathbb{S}}^{-1}\bar{B}_m &= R\left[V\Sigma'U'\bar{\mathbb{S}}^{-1}U\Sigma V'\right]^{-1}V\Sigma'U'\bar{\mathbb{S}}^{-1}\bar{B}_m \\ &= RV\left[\Sigma'(U'\bar{\mathbb{S}}^{-1}U)\Sigma V'\right]^{-1}\Sigma'\left[U'\bar{\mathbb{S}}^{-1}U\right](U'\bar{B}_m) \\ &\overset{d}{=} RV\left[\Sigma'\bar{\mathbb{S}}^{-1}\Sigma\right]^{-1}\Sigma'\bar{\mathbb{S}}^{-1}\bar{B}_m, \end{aligned}$$

where the distributional equivalence holds by a rotation invariance property of $[\bar{B}_m, \bar{\mathbb{S}}^{-1}]$. Denoting

$$[\bar{\mathbb{S}}^{-1}]_{m\times m} = \begin{pmatrix} \underset{d\times d}{\bar{\mathbb{S}}^{dd}} & \underset{d\times q}{\bar{\mathbb{S}}^{dq}} \\ \underset{q\times q}{\bar{\mathbb{S}}^{qd}} & \underset{q\times d}{\bar{\mathbb{S}}^{qq}} \end{pmatrix},$$

we have

$$\begin{aligned} RV\left[\Sigma'\bar{\mathbb{S}}^{-1}\Sigma\right]^{-1}\Sigma'\bar{\mathbb{S}}^{-1}\bar{B}_m &= RVA^{-1}\left(\bar{\mathbb{S}}^{dd}\right)^{-1}(A')^{-1}A'\left(\bar{\mathbb{S}}^{dd},\bar{\mathbb{S}}^{dq}\right)\bar{B}_m \\ &= RVA^{-1}\begin{bmatrix} I_d & \left(\bar{\mathbb{S}}^{dd}\right)^{-1}\bar{\mathbb{S}}^{dq} \end{bmatrix}\bar{B}_m \\ &= RVA^{-1}\left[\bar{B}_d - \bar{\mathbb{S}}_{dq}\bar{\mathbb{S}}_{qq}^{-1}\bar{B}_q\right]. \end{aligned}$$

where the last equation follows by the partitioned inverse formula that $\bar{\mathbb{S}}^{dq} = -\bar{\mathbb{S}}^{dd}\bar{\mathbb{S}}_{dq}\bar{\mathbb{S}}_{qq}^{-1}$. Similarly, we obtain

$$\begin{aligned} R\left[\Gamma_\Lambda'\bar{\mathbb{S}}^{-1}\Gamma_\Lambda\right]^{-1}R' &\overset{d}{=} RV\left[\Sigma'\bar{\mathbb{S}}^{-1}\Sigma\right]^{-1}V'R' \\ &= RVA^{-1}\left(\bar{\mathbb{S}}^{dd}\right)^{-1}(A')^{-1}V'R' \\ &= RVA^{-1}\bar{\mathbb{S}}_{dd\cdot q}(A')^{-1}V'R', \end{aligned}$$

where $\bar{\mathbb{S}}_{dd\cdot q} = \bar{\mathbb{S}}_{dd} - \bar{\mathbb{S}}_{dq}\bar{\mathbb{S}}_{qq}^{-1}\bar{\mathbb{S}}_{qd}$. Therefore,

$$\begin{aligned} F_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c) \overset{d}{\to} \mathbb{F}_{2\infty} &\overset{d}{=} \frac{G}{p}\cdot\left[RVA^{-1}\left(\bar{B}_d - \bar{\mathbb{S}}_{dq}\bar{\mathbb{S}}_{qq}^{-1}\bar{B}_q\right)\right]' \\ &\times\left(RVA^{-1}\bar{\mathbb{S}}_{dd\cdot q}(A')^{-1}V'R'\right)^{-1}\left[RVA^{-1}\left(\bar{B}_d - \bar{\mathbb{S}}_{dq}\bar{\mathbb{S}}_{qq}^{-1}\bar{B}_q\right)\right]. \end{aligned} \tag{B.26}$$

Let $\tilde{U}_{p\times p}\tilde{\Sigma}\tilde{V}_{d\times d}'$ be a SVD of RVA^{-1}, where $\tilde{\Sigma} = (\tilde{A}_{p\times p}, O_{p\times(d-p)})$. By definition, $\tilde{V}$ is the matrix of eigenvectors of $(RVA^{-1})'(RVA^{-1})$. Then,

$$\begin{aligned} \mathbb{F}_{2\infty} &\overset{d}{=} \frac{G}{p}\left[\bar{B}_d - \bar{\mathbb{S}}_{dq}\bar{\mathbb{S}}_{qq}^{-1}\bar{B}_q\right]'\tilde{V}\tilde{\Sigma}'\tilde{U}'\left[\tilde{U}\tilde{\Sigma}\tilde{V}'\bar{\mathbb{S}}_{dd\cdot q}\tilde{V}\tilde{\Sigma}'\tilde{U}'\right]^{-1}\tilde{U}\tilde{\Sigma}\tilde{V}'\left[\bar{B}_d - \bar{\mathbb{S}}_{dq}\bar{\mathbb{S}}_{qq}^{-1}\bar{B}_q\right] \\ &= \frac{G}{p}\left[\bar{B}_d - \bar{\mathbb{S}}_{dq}\bar{\mathbb{S}}_{qq}^{-1}\bar{B}_q\right]'\tilde{V}\tilde{\Sigma}'\cdot\left[\tilde{\Sigma}\tilde{V}'\bar{\mathbb{S}}_{dd\cdot q}\tilde{V}\tilde{\Sigma}'\right]^{-1}\cdot\tilde{\Sigma}\tilde{V}'\left[\bar{B}_d - \bar{\mathbb{S}}_{dq}\bar{\mathbb{S}}_{qq}^{-1}\bar{B}_q\right] \\ &= \frac{G}{p}\left[\tilde{V}'\bar{B}_d - \tilde{V}'\bar{\mathbb{S}}_{dq}\bar{\mathbb{S}}_{qq}^{-1}\bar{B}_q\right]'\tilde{\Sigma}'\cdot\left[\tilde{\Sigma}\tilde{V}'\bar{\mathbb{S}}_{dd\cdot q}\tilde{V}\tilde{\Sigma}'\right]^{-1}\cdot\tilde{\Sigma}\left[\tilde{V}'\bar{B}_d - \tilde{V}'\bar{\mathbb{S}}_{dq}\bar{\mathbb{S}}_{qq}^{-1}\bar{B}_q\right]. \end{aligned}$$

Note that the rotational invariance property of the standard normal vector implies

$$\left[\tilde{V}'\bar{B}_d, \tilde{V}'\bar{\mathbb{S}}_{dq}, \bar{\mathbb{S}}_{qq}^{-1}, \tilde{V}'\bar{\mathbb{S}}_{dd\cdot q}\tilde{V}\right] \overset{d}{=} \left[\bar{B}_d, \bar{\mathbb{S}}_{dq}, \bar{\mathbb{S}}_{qq}^{-1}, \bar{\mathbb{S}}_{dd\cdot q}\right],$$

which leads us to obtain

$$\begin{aligned}
&\mathbb{F}_{2\infty} \overset{d}{=} \frac{G}{p}\left[\bar{B}_d - \bar{\mathbb{S}}_{dq}\bar{\mathbb{S}}_{qq}^{-1}\bar{B}_q\right]'\tilde{\Sigma}' \cdot \left[\tilde{\Sigma}\bar{\mathbb{S}}_{dd\cdot q}\tilde{\Sigma}'\right]^{-1} \cdot \tilde{\Sigma}\left[\bar{B}_d - \bar{\mathbb{S}}_{dq}\bar{\mathbb{S}}_{qq}^{-1}\bar{B}_q\right] \\
= \quad & \frac{G}{p} \cdot \left[\bar{B}_p - \bar{\mathbb{S}}_{pq}\bar{\mathbb{S}}_{qq}^{-1}\bar{B}_q\right]'\tilde{A}'\left\{\tilde{A}\left(\bar{\mathbb{S}}_{pp\cdot q}\right)\tilde{A}'\right\}^{-1}\tilde{A}\left[\bar{B}_p - \bar{\mathbb{S}}_{pq}\bar{\mathbb{S}}_{qq}^{-1}\bar{B}_q\right] \\
= \quad & \frac{G}{p} \cdot \left(\bar{B}_p - \bar{\mathbb{S}}_{pq}\bar{\mathbb{S}}_{qq}^{-1}\bar{B}_q\right)'\bar{\mathbb{S}}_{pp\cdot q}^{-1}\left(\bar{B}_p - \bar{\mathbb{S}}_{pq}\bar{\mathbb{S}}_{qq}^{-1}\bar{B}_q\right)',
\end{aligned}$$

as desired. The proof of part (b) is similar, so we omit the details. To prove part (c), we use the same arguments and obtain

$$\begin{aligned}
J(\hat{\theta}_2^c) \xrightarrow{d} \mathbb{J}_\infty &:= G \cdot \left\{U'\bar{B}_m - U'\Gamma_\Lambda\left(\Gamma_\Lambda'\bar{\mathbb{S}}^{-1}\Gamma_\Lambda\right)^{-1}\Gamma_\Lambda'\bar{\mathbb{S}}^{-1}\bar{B}_m\right\}' \times U'\bar{\mathbb{S}}^{-1}U \\
&\times \left\{U'\bar{B}_m - U'\Gamma_\Lambda\left(\Gamma_\Lambda'\bar{\mathbb{S}}^{-1}\Gamma_\Lambda\right)^{-1}\Gamma_\Lambda'\bar{\mathbb{S}}^{-1}\bar{B}_m\right\} \\
&\overset{d}{=} G\left\{\bar{B}_m - U'\Gamma_\Lambda V A^{-1}\left[\bar{B}_d - \bar{\mathbb{S}}_{dq}\bar{\mathbb{S}}_{qq}^{-1}\bar{B}_q\right]\right\}'\bar{\mathbb{S}}^{-1} \\
&\times \left\{\bar{B}_m - U'\Gamma_\Lambda V A^{-1}\left[\bar{B}_d - \bar{\mathbb{S}}_{dq}\bar{\mathbb{S}}_{qq}^{-1}\bar{B}_q\right]\right\},
\end{aligned} \tag{B.27}$$

which is continued in

$$\begin{aligned}
&= G\left\{\bar{B}_m - \begin{bmatrix} I_{d\times d} \\ O_{q\times d}\end{bmatrix}\left(\bar{B}_d - \bar{\mathbb{S}}_{dq}\bar{\mathbb{S}}_{qq}^{-1}\bar{B}_q\right)\right\}'\bar{\mathbb{S}}^{-1} \\
&\times \left\{\bar{B}_m - \begin{bmatrix} I_{d\times d} \\ O_{q\times d}\end{bmatrix}\left(\bar{B}_d - \bar{\mathbb{S}}_{dq}\bar{\mathbb{S}}_{qq}^{-1}\bar{B}_q\right)\right\} \\
&= G\begin{pmatrix} \bar{\mathbb{S}}_{dq}\bar{\mathbb{S}}_{qq}^{-1}\bar{B}_q \\ \bar{B}_q \end{pmatrix}'\bar{\mathbb{S}}^{-1}\begin{pmatrix} \bar{\mathbb{S}}_{dq}\bar{\mathbb{S}}_{qq}^{-1}\bar{B}_q \\ \bar{B}_q \end{pmatrix} \\
&= G \cdot \bar{B}_q'\bar{\mathbb{S}}_{qq}^{-1}\bar{B}_q,
\end{aligned}$$

where the last equality follows from straightforward calculations. Lastly, it is not difficult to check that the convergence results in (B.31) and (B.27) hold jointly. This completes the proof. ■

B.3 Proof of Proposition 6

Proof. Let $U\Sigma V'$ be a singular value decomposition (SVD) of Γ_Λ. Also, define $B_{m,g}^U = U'B_{m,g}$, $\bar{B}_m^U = U'\bar{B}_m$, and

$$\tilde{\mathbb{D}}_\infty^U = U'\tilde{\mathbb{D}}_\infty U = \left(\begin{array}{cc} \underset{d\times d}{\tilde{\mathbb{D}}_{11}^U} & \underset{d\times q}{\tilde{\mathbb{D}}_{12}^U} \\ \underset{q\times d}{\tilde{\mathbb{D}}_{21}^U} & \underset{q\times q}{\tilde{\mathbb{D}}_{22}^U} \end{array}\right). \tag{B.28}$$

Using the similar argument in the proof of Proposition 8-(a), we can show that

$$\sqrt{n}(\hat{\theta}_2 - \theta_0) \xrightarrow{d} -VA^{-1}\sqrt{G}\left(\bar{B}_d^U - \tilde{\mathbb{D}}_{12}^U\left[\tilde{\mathbb{D}}_{22}^U\right]^{-1}\bar{B}_q^U\right) \text{ and } \hat{\Omega}(\hat{\theta}_1) \xrightarrow{d} \Lambda\tilde{\mathbb{D}}_\infty\Lambda',$$

where holds jointly. It then follows that

$$\begin{aligned}
&F_{\hat{\Omega}(\hat{\theta}_1)}(\hat{\theta}_2) = \frac{1}{p}\cdot\left[R(\hat{\theta}_2-\theta_0)\right]'\left(R\widehat{var}_{\hat{\Omega}(\hat{\theta}_1)}(\hat{\theta}_2)R'\right)^{-1}\left[R(\hat{\theta}_2-\theta_0)\right]\\
&\xrightarrow{d} \frac{G}{p}\cdot(\bar{B}_d^U - \tilde{\mathbb{D}}_{12}^U\left[\tilde{\mathbb{D}}_{22}^U\right]^{-1}\bar{B}_q^U)'A^{-1\prime}V'R'\left[R\left(\Gamma'\left(\Lambda\tilde{\mathbb{D}}_\infty\Lambda'\right)^{-1}\Gamma\right)^{-1}R'\right]^{-1}\\
&\times RVA^{-1}(\bar{B}_d^U - \tilde{\mathbb{D}}_{12}^U\left[\tilde{\mathbb{D}}_{22}^U\right]^{-1}\bar{B}_q^U)\\
&= \frac{G}{p}\cdot(\bar{B}_d^U - \tilde{\mathbb{D}}_{12}^U\left[\tilde{\mathbb{D}}_{22}^U\right]^{-1}\bar{B}_q^U)'A^{-1\prime}V'R'\cdot\left\{R\left[\Gamma'\left(\Lambda'\right)^{-1}U\left(U'\tilde{\mathbb{D}}_\infty U\right)^{-1}U'\Lambda^{-1}\Gamma\right]^{-1}R'\right\}^{-1}\\
&\times RVA^{-1}(\bar{B}_d^U - \tilde{\mathbb{D}}_{12}^U\left[\tilde{\mathbb{D}}_{22}^U\right]^{-1}\bar{B}_d^U)\\
&= \frac{G}{p}\cdot(\bar{B}_d^U - \tilde{\mathbb{D}}_{12}^U\left[\tilde{\mathbb{D}}_{22}^U\right]^{-1}\bar{B}_q^U)'A^{-1\prime}V'R'\left\{RVA^{-1}\tilde{\mathbb{D}}_{11\cdot 2}^U A^{-1\prime}V'R'\right\}^{-1}\\
&\times RVA^{-1}(\bar{B}_d^U - \tilde{\mathbb{D}}_{12}^U\left[\tilde{\mathbb{D}}_{22}^U\right]^{-1}\bar{B}_d^U).
\end{aligned}$$

Let $\tilde{U}_{p\times p}\tilde{\Sigma}\tilde{V}_{d\times d}'$ be a SVD of RVA^{-1}, where $\tilde{\Sigma} = (\tilde{A}_{p\times p}, O_{p\times(d-p)})$. Also, define

$$\mathbb{V} = \left(\begin{array}{cc} \tilde{V}_{d\times d} & O \\ O & I_{q\times q}\end{array}\right),$$

and

$$\breve{\mathbb{D}} = \left(\begin{array}{cc} \breve{\mathbb{D}}_{11} & \breve{\mathbb{D}}_{12} \\ \breve{\mathbb{D}}_{21} & \breve{\mathbb{D}}_{22}\end{array}\right) = \left(\begin{array}{cc} \tilde{V}_{d\times d} & O \\ O & I_q\end{array}\right)'\left(\begin{array}{cc} \tilde{\mathbb{D}}_{11}^U & \tilde{\mathbb{D}}_{12}^U \\ \tilde{\mathbb{D}}_{21}^U & \tilde{\mathbb{D}}_{22}^U\end{array}\right)\left(\begin{array}{cc} \tilde{V}_{d\times d} & O \\ O & I_q\end{array}\right) = \mathbb{V}'\tilde{\mathbb{D}}_\infty^U\mathbb{V}.$$

Then,

$$\begin{aligned}
\breve{\mathbb{D}} &= \frac{1}{G}\sum_{g=1}^G \mathbb{V}'U'(B_{m,g}-\bar{B}_m)(B_{m,g}-\bar{B}_m)'\mathbb{V}U + \left(\begin{array}{c}\tilde{V}'\beta_{\tilde{W}}\\ I_q\end{array}\right)\bar{B}_q^U(\bar{B}_q^U)'\left(\begin{array}{c}\tilde{V}'\beta_{\tilde{W}}\\ I_q\end{array}\right)'\\
&\overset{d}{=} \frac{1}{G}\sum_{g=1}^G (B_{m,g}-\bar{B}_m)(B_{m,g}-\bar{B}_m)' + \left(\begin{array}{c}\tilde{V}'\beta_{\tilde{W}}\\ I_q\end{array}\right)\bar{B}_q\bar{B}_q'\left(\begin{array}{c}\tilde{V}'\beta_{\tilde{W}}\\ I_q\end{array}\right)',
\end{aligned}$$

which implies that

$$\breve{\mathbb{D}}_{11} := \begin{pmatrix} \breve{\mathbb{D}}_{pp} & \breve{\mathbb{D}}_{p,d-p} \\ \breve{\mathbb{D}}_{d-p,p} & \breve{\mathbb{D}}_{d-p,d-p} \end{pmatrix} \overset{d}{=} \frac{1}{G}\sum_{g=1}^{G}(B_{d,g}-\bar{B}_d)(B_{d,g}-\bar{B}_d)' + \left(\tilde{V}'\beta_{\tilde{W}}\right)\bar{B}_q\bar{B}_q'\left(\tilde{V}'\beta_{\tilde{W}}\right)', \quad \text{(B.29)}$$

and

$$\breve{\mathbb{D}}_{12} := \begin{pmatrix} \breve{\mathbb{D}}_{pq} \\ \breve{\mathbb{D}}_{d-p,q} \end{pmatrix} \overset{d}{=} \frac{1}{G}\sum_{g=1}^{G}(B_{d,g}-\bar{B}_d)(B_{q,g}-\bar{B}_q)' + \left(\tilde{V}'\beta_{\tilde{W}}\right)\bar{B}_q\bar{B}_q'. \quad \text{(B.30)}$$

Now

$$\begin{aligned}
& F_{\hat{\Omega}(\hat{\theta}_1)}(\hat{\theta}_2) \overset{d}{\to} \frac{G}{p}\cdot(\bar{B}_d^U - \breve{\mathbb{D}}_{12}\breve{\mathbb{D}}_{22}^{-1}\bar{B}_q^U)'\tilde{V}\tilde{\Sigma}'\tilde{U}'\left\{\tilde{U}\tilde{\Sigma}\tilde{V}'\breve{\mathbb{D}}_{11\cdot 2}\tilde{V}\tilde{\Sigma}'\tilde{U}'\right\}^{-1}\tilde{U}\tilde{\Sigma}\tilde{V}'(\bar{B}_d^U - \breve{\mathbb{D}}_{12}\breve{\mathbb{D}}_{22}^{-1}\bar{B}_q^U) \\
= \quad & \frac{G}{p}\cdot(\bar{B}_d^U - \breve{\mathbb{D}}_{12}\breve{\mathbb{D}}_{22}^{-1}\bar{B}_q^U)'\tilde{V}\tilde{\Sigma}'\cdot\left\{\tilde{\Sigma}\tilde{V}'\breve{\mathbb{D}}_{11\cdot 2}\tilde{V}\tilde{\Sigma}'\right\}^{-1}\cdot\tilde{\Sigma}\tilde{V}'(\bar{B}_d^U - \breve{\mathbb{D}}_{12}\breve{\mathbb{D}}_{22}^{-1}\bar{B}_q^U) \\
& \overset{d}{=} \frac{G}{p}\cdot\left[\bar{B}_p - \breve{\mathbb{D}}_{pq}\breve{\mathbb{D}}_{qq}^{-1}\bar{B}_q\right]'\tilde{A}'\left\{\tilde{A}\left(\breve{\mathbb{D}}_{pp} - \breve{\mathbb{D}}_{pq}\breve{\mathbb{D}}_{qq}^{-1}\breve{\mathbb{D}}_{qp}\right)\tilde{A}'\right\}^{-1}\tilde{A}\left[\bar{B}_p - \breve{\mathbb{D}}_{pq}\breve{\mathbb{D}}_{qq}^{-1}\bar{B}_q\right] \\
& \overset{d}{=} \frac{G}{p}\cdot\left[\bar{B}_p - \breve{\mathbb{D}}_{pq}\breve{\mathbb{D}}_{qq}^{-1}\bar{B}_q\right]'\left(\breve{\mathbb{D}}_{pp} - \breve{\mathbb{D}}_{pq}\breve{\mathbb{D}}_{qq}^{-1}\breve{\mathbb{D}}_{qp}\right)^{-1}\left[\bar{B}_p - \breve{\mathbb{D}}_{pq}\breve{\mathbb{D}}_{qq}^{-1}\bar{B}_q\right], \quad \text{(B.31)}
\end{aligned}$$

where $\breve{\mathbb{D}}_{pq}$, $\breve{\mathbb{D}}_{qq}$, and $\breve{\mathbb{D}}_{qp}$ in the last two equalities are understood to equal the corresponding components on the right hand sides of (B.29) and (B.30). Here we have abused the notation a little bit. We have

$$\begin{pmatrix} \breve{\mathbb{D}}_{pp} & \breve{\mathbb{D}}_{pq} \\ \breve{\mathbb{D}}_{pq}' & \breve{\mathbb{D}}_{qq} \end{pmatrix} \overset{d}{=} \mathbb{E}_{p+q,p+q} = \begin{pmatrix} \bar{\mathbb{S}}_{pp} & \bar{\mathbb{S}}_{pq} \\ \bar{\mathbb{S}}_{pq}' & \bar{\mathbb{S}}_{qq} \end{pmatrix} + \tilde{w}\bar{B}_q\bar{B}_q'\tilde{w}' \quad \text{(B.32)}$$

for

$$\tilde{w} = \begin{pmatrix} \tilde{\beta}_{\tilde{W}}^p \\ I_q \end{pmatrix} \in \mathbb{R}^{(p+q)\times q}.$$

Direct calculations show that the representation in (B.31) is numerically identical to

$$\frac{1}{p}\cdot\left[G\begin{pmatrix} \bar{B}_p \\ \bar{B}_q \end{pmatrix}'\begin{pmatrix} \mathbb{E}_{pp} & \mathbb{E}_{pq} \\ \mathbb{E}_{pq}' & \mathbb{E}_{qq} \end{pmatrix}^{-1}\begin{pmatrix} \bar{B}_p \\ \bar{B}_q \end{pmatrix} - G\cdot\bar{B}_q'\mathbb{S}_{qq}^{-1}\bar{B}_q\right],$$

which completes the proof of part (a). To prove part (b), we repeat the same argument in the proof of Proposition 8-(b) and obtain that

$$\begin{aligned}
\sqrt{n}g_n(\hat{\theta}_2) &= \frac{1}{\sqrt{G}}\sum_{g=1}^{G}\left(\frac{1}{\sqrt{L}}\sum_{i=1}^{L_g} f_i^g(\hat{\theta}_2)\right) + o_p(1) \\
&\overset{d}{\to} \Lambda\sqrt{G}\left(UU'\bar{B}_m - \Gamma_\Lambda\left[\Gamma_\Lambda'\tilde{\mathbb{D}}_\infty^{-1}\Gamma_\Lambda\right]^{-1}\Gamma_\Lambda'\tilde{\mathbb{D}}_\infty^{-1}\bar{B}_m\right) \\
&\overset{d}{=} \Lambda\sqrt{G}\left[U\bar{B}_m^U - \Gamma_\Lambda V A^{-1}\left(\bar{B}_d^U - \tilde{\mathbb{D}}_{12}^U\left[\tilde{\mathbb{D}}_{22}^U\right]^{-1}\bar{B}_q^U\right)\right]
\end{aligned}$$

with $\tilde{\mathbb{D}}_{12}^U$ and $\tilde{\mathbb{D}}_{22}^U$ given in (B.28). Therefore, we have

$$
\begin{aligned}
&J(\hat{\theta}_2) = ng_n(\hat{\theta}_2)' \left(\hat{\Omega}(\hat{\theta}_1)\right)^{-1} g_n(\hat{\theta}_2) \\
&\xrightarrow{d} G \left\{ U\bar{B}_m^U - \Gamma_\Lambda V A^{-1} \left(\bar{B}_d^U - \tilde{\mathbb{D}}_{12}^U \left[\tilde{\mathbb{D}}_{22}^U\right]^{-1} \bar{B}_q^U \right) \right\}' \times \Lambda' \left(\Lambda \tilde{\mathbb{D}}_\infty \Lambda'\right)^{-1} \Lambda \\
&\times \left\{ U\bar{B}_m^U - \Gamma_\Lambda V A^{-1} \left(\bar{B}_d^U - \tilde{\mathbb{D}}_{12}^U \left[\tilde{\mathbb{D}}_{22}^U\right]^{-1} \bar{B}_q^U \right) \right\} \\
&= G \left\{ \bar{B}_m^U - U'\Gamma_\Lambda V A^{-1} \left(\bar{B}_d^U - \tilde{\mathbb{D}}_{12}^U \left[\tilde{\mathbb{D}}_{22}^U\right]^{-1} \bar{B}_q^U \right) \right\}' U' \tilde{\mathbb{D}}_\infty^{-1} U \\
&\times \left\{ \bar{B}_m^U - U'\Gamma_\Lambda V A^{-1} \left(\bar{B}_d^U - \tilde{\mathbb{D}}_{12}^U \left[\tilde{\mathbb{D}}_{22}^U\right]^{-1} \bar{B}_q^U \right) \right\},
\end{aligned}
$$

which is continued in

$$
\begin{aligned}
&= G \left\{ \bar{B}_m^U - \begin{bmatrix} I_{d\times d} \\ O_{q\times d} \end{bmatrix} \left(\bar{B}_d^U - \tilde{\mathbb{D}}_{12}^U \left[\tilde{\mathbb{D}}_{22}^U\right]^{-1} \bar{B}_q^U \right) \right\}' \left[\tilde{\mathbb{D}}_\infty^U\right]^{-1} \\
&\times \left\{ \bar{B}_m^U - \begin{bmatrix} I_{d\times d} \\ O_{q\times d} \end{bmatrix} \left(\bar{B}_d^U - \tilde{\mathbb{D}}_{12}^U \left[\tilde{\mathbb{D}}_{22}^U\right]^{-1} \bar{B}_q^U \right) \right\} \\
&= G \begin{pmatrix} \tilde{\mathbb{D}}_{12}^U \left[\tilde{\mathbb{D}}_{22}^U\right]^{-1} \bar{B}_q^U \\ \bar{B}_q^U \end{pmatrix}' \left[\tilde{\mathbb{D}}_\infty^U\right]^{-1} \begin{pmatrix} \tilde{\mathbb{D}}_{12}^U \left[\tilde{\mathbb{D}}_{22}^U\right]^{-1} \bar{B}_q^U \\ \bar{B}_q^U \end{pmatrix} \\
&= G(\bar{B}_q^U)' \left[\tilde{\mathbb{D}}_{22}^U\right]^{-1} \bar{B}_q^U \overset{d}{=} G \cdot \bar{B}_q' \mathbb{S}_{22}^{-1} \bar{B}_q
\end{aligned}
$$

where the second last equality follows from straightforward calculations. The joint convergence can be proved easily. ■

B.4 Asymptotics for LM and QLR statistics

In this subsection, we construct the Quasi-Likelihood Ratio (QLR) and the Lagrangian Multiplier (LM) statistics in the GMM setting and investigate their fixed-G limits. Define the restricted and centered two-step estimator $\hat{\theta}_2^{c,r}$

$$\hat{\theta}_2^{c,r} := \arg\min_{\theta\in\Theta} g_n(\theta)' \left[\hat{\Omega}^c(\hat{\theta}_1)\right]^{-1} g_n(\theta) \text{ such that } R\theta = r.$$

The QLR statistic is then given by

$$LR_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c, \hat{\theta}_2^{c,r}) := \frac{n}{p}\left\{g_n(\hat{\theta}_2^{c,r})' \left[\hat{\Omega}^c(\hat{\theta}_1)\right]^{-1} g_n(\hat{\theta}_2^{c,r}) - g_n(\hat{\theta}_2^c)' \left[\hat{\Omega}^c(\hat{\theta}_1)\right]^{-1} g_n(\hat{\theta}_2^c)\right\}.$$

To define LM or score statistic in the GMM setting, let $S_{\hat{\Omega}^c(\cdot)}(\theta)$ be the gradient of the GMM criterion function $\hat{\Gamma}(\theta)'[\hat{\Omega}^c(\cdot)]^{-1}g_n(\theta)$, then the GMM score test statistic is given by

$$LM_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^{c,r}) := \frac{n}{p}\left[S_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^{c,r})\right]' \left\{\hat{\Gamma}(\hat{\theta}_2^{c,r})' \left[\hat{\Omega}^c(\hat{\theta}_1)\right]^{-1} \hat{\Gamma}(\hat{\theta}_2^{c,r})\right\}^{-1} \left[S_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^{c,r})\right].$$

In the definition of all three types of the GMM test statistics, we plug the first-step estimator $\hat{\theta}_1$ into $\hat{\Omega}^c(\cdot)$, but Lemma 7 indicates that replacing $\hat{\theta}_1$ with any $\sqrt{n}$-consistent estimator (e.g., $\hat{\theta}_2$ and $\hat{\theta}_2^c$) does not affect the fixed-G asymptotic results. This contrasts with the fixed-G asymptotics for the uncentered two-step estimator $\hat{\theta}_2$. Lastly, using the same multiplicative factors as in Section 4 in the main text, we can also construct the modified QLR and LM statistics.

Proposition B.4 *Let Assumptions 1∼5 hold. Then„*

(a) $LR_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c, \hat{\theta}_2^{c,r}) = F_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c) + o_p(1)$;

(b) $LM_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^{c,r}) = F_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c) + o_p(1)$.

(c) *The modified QLR and LM statistics all converge in distribution to* $\mathcal{F}_{p,G-p-q}$.

Proposition B.4 shows that QLR and LM types of test statistics are asymptotically equivalent to the Wald statistics $F_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c)$. This also implies that all three types of test statistics share the same fixed-G limit as given in the main text. Similar results are obtained by Sun (2014) and Hwang and Sun (2017a), which focus on the two-step GMM estimation and HAR inference in a time series setting.

B.4.1 Proof of Proposition B.4

Proof. To prove (a), it suffices to show the asymptotic equivalence between $\mathrm{LR}_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c, \hat{\theta}_2^{c,r})$ and $F_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c)$ holds under the fixed-G asymptotics. Recall that the restricted two-step GMM estimator $\hat{\theta}_2^{c,r}$ minimizes

$$g_n(\theta)' \left[\hat{\Omega}^c(\hat{\theta}_1)\right]^{-1} g_n(\theta)/2 + \lambda_n'(R\theta - r).$$

The first order conditions are

$$\hat{\Gamma}'(\hat{\theta}_2^{c,r}) \left[\hat{\Omega}^c(\hat{\theta}_1)\right]^{-1} g_n(\hat{\theta}_2^{c,r}) + R'\lambda_n = 0 \text{ and } R\hat{\theta}_2^{c,r} = r.$$

Using a Taylor expansion and Assumption 3, we can combine two FOC's to get

$$\sqrt{n}(\hat{\theta}_2^{c,r} - \theta_0) = -\Phi^{-1}\Gamma' \left[\hat{\Omega}^c(\hat{\theta}_1)\right]^{-1} \sqrt{n} g_n(\theta_0) - \Phi^{-1}R' \left(R\Phi^{-1}R'\right)^{-1} R\Phi^{-1}\Gamma' \left[\hat{\Omega}_n^c(\hat{\theta}_1)\right]^{-1} \sqrt{n} g_n(\theta_0) + o_p(1) \tag{B.33}$$

and

$$\sqrt{n}\lambda_n = -(R\Phi^{-1}R')^{-1}R\Phi^{-1}\Gamma' \left[\hat{\Omega}_n^c(\hat{\theta}_1)\right]^{-1} \sqrt{n} g_n(\theta_0) + o_p(1), \tag{B.34}$$

where $\Phi := \Gamma' \left[\hat{\Omega}^c(\hat{\theta}_1)\right]^{-1} \Gamma$. Subtracting (B.33) from (13), we have

$$\begin{aligned}\sqrt{n}(\hat{\theta}_2^{c,r} - \hat{\theta}_2^c) &= -\Phi^{-1}R' \left(R\Phi^{-1}R'\right)^{-1} R\Phi^{-1}\Gamma' \left[\hat{\Omega}^c(\hat{\theta}_1)\right]^{-1} \sqrt{n} g_n(\theta_0) + o_p(1) && \text{(B.35)}\\ &= O_p(1), && \text{(B.36)}\end{aligned}$$

where the equation (B.36) comes from Lemma 7-(b) and Assumption 1-ii). Thus, we can approximate $g_n(\hat{\theta}_2^{c,r})$ around $\hat{\theta}_2^c$ as

$$g_n'(\hat{\theta}_2^{c,r}) = g_n'(\hat{\theta}_2^c) - (\hat{\theta}_2^{c,r} - \hat{\theta}_2^c)'\hat{\Gamma}'(\hat{\theta}_2^c) + o_p(n^{-1/2}).$$

Plugging this into the definition of $LR_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c, \hat{\theta}_2^{c,r})$,

$$\begin{aligned}LR_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c, \hat{\theta}_2^{c,r}) &= \frac{n}{p}\Big\{(\hat{\theta}_2^{c,r} - \hat{\theta}_2^c)'\hat{\Gamma}'(\hat{\theta}_2^c) \left[\hat{\Omega}^c(\hat{\theta}_1)\right]^{-1} \hat{\Gamma}(\hat{\theta}_2^c)(\hat{\theta}_2^{c,r} - \hat{\theta}_2^c) && \text{(B.37)}\\ &\quad -2g_n'(\hat{\theta}_2^c) \left[\hat{\Omega}^c(\hat{\theta}_1)\right]^{-1} \hat{\Gamma}(\hat{\theta}_2^c)(\hat{\theta}_2^{c,r} - \hat{\theta}_2^c)\Big\} + o_p(1),\end{aligned}$$

where the last term in (B.37) is always zero from the FOC of $\hat{\theta}_2^c$. Combining (B.35) and (B.37), we have

$$\begin{aligned}LR_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c, \hat{\theta}_2^{c,r}) &= \frac{n}{p}(\hat{\theta}_2^{c,r} - \hat{\theta}_2^c)'\hat{\Gamma}'(\hat{\theta}_2^c) \left[\hat{\Omega}^c(\hat{\theta}_1)\right]^{-1} \hat{\Gamma}(\hat{\theta}_2^c)(\hat{\theta}_2^{c,r} - \hat{\theta}_2^c) + o_p(1)\\ &= \frac{n}{p}\left[\Phi^{-1}R' \left(R\Phi^{-1}R'\right)^{-1} R\Phi^{-1}\Gamma' \left[\hat{\Omega}^c(\hat{\theta}_1)\right]^{-1} \sqrt{n} g_n(\theta_0)\right]' \times\\ &\quad \Phi\left[\Phi^{-1}R' \left(R\Phi^{-1}R'\right)^{-1} R\Phi^{-1}\Gamma' \left[\hat{\Omega}^c(\hat{\theta}_1)\right]^{-1} \sqrt{n} g_n(\theta_0)\right] + o_p(1)\\ &= \frac{1}{p}\left[R\Phi^{-1}\Gamma' \left[\hat{\Omega}^c(\hat{\theta}_1)\right]^{-1} \sqrt{n} g_n(\theta_0)\right]' \left(R\Phi^{-1}R'\right)^{-1}\\ &\quad \times \left[R\Phi^{-1}\Gamma' \left[\hat{\Omega}^c(\hat{\theta}_1)\right]^{-1} \sqrt{n} g_n(\theta_0)\right] + o_p(1)\\ &= F_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c) + o_p(1),\end{aligned}$$

as desired. To prove part (b), we show $LM_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^{c,r}) = LR_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c, \hat{\theta}_2^{c,r}) + o_p(1)$. From the first order condition of $\hat{\theta}_2^{c,r}$ and the equation (B.34), we expand the score vector by

$$\begin{aligned}\sqrt{n}S_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^{c,r}) &= \hat{\Gamma}(\hat{\theta}_2^{c,r})' \left[\hat{\Omega}^c(\hat{\theta}_1)\right]^{-1} \sqrt{n} g_n(\hat{\theta}_2^{c,r}) = -R'\sqrt{n}\lambda_n\\ &= R'(R\Phi^{-1}R')^{-1}R\Phi^{-1}\Gamma' \left[\hat{\Omega}_n^c(\hat{\theta}_1)\right]^{-1} \sqrt{n} g_n(\theta_0) + o_p(1)\\ &= -\Phi\sqrt{n}(\hat{\theta}_2^{c,r} - \hat{\theta}_2^c) + o_p(1),\end{aligned}$$

and so

$$\begin{aligned} LM_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^{c,r}) &= n(\hat{\theta}_2^{c,r} - \hat{\theta}_2^c)'\Phi(\hat{\theta}_2^{c,r} - \hat{\theta}_2^c)/p + o_p(1) \\ &= LR_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c, \hat{\theta}_2^{c,r}) + o_p(1) \\ &= F_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c) + o_p(1), \end{aligned}$$

which leads the desired result. The proof of (c) directly follows from the results in (a), (b), and Proposition 8. ■

B.5 Asymptotically unbalanced sizes in clusters

In this subsection, we present the fixed-G limits of the centered CCE and corresponding test statistics assuming the asymptotically unbalanced cluster sizes. We assume that the cluster size $L_g \to \infty$ as $n \to \infty$ such that

$$\frac{L_g}{n} \to \lambda_g > 0 \text{ for each } g = 1, \ldots, G. \tag{B.38}$$

By construction, $\sum_{g=1}^{G} \lambda_g = 1$ for all possible $\lambda_g \in (0,1)$. Also, Assumption 1-ii) guarantees that the total (scaled) sum of moment process satisfies

$$\begin{aligned}\sqrt{n} g_n(\theta_0) &= \sum_{g=1}^{G} \sqrt{\frac{L_g}{n}} \frac{1}{\sqrt{L_g}} \sum_{i=1}^{L_g} f_i^g(\theta_0) \\ &\xrightarrow{d} \sum_{g=1}^{G} \sqrt{\lambda_g} \Lambda_g B_{m,g} \sim N(0, \Omega),\end{aligned}$$

where $\Omega = \sum \lambda_g \Omega_g$. Keeping the same notations in the main text, we define an $m \times m$ random matrix

$$\bar{\mathbb{M}} := \bar{\mathbb{M}}(\lambda) = \sum_{g=1}^{G} \left(\sqrt{\lambda_g} B_{m,g} - \lambda_g \sum_{h=1}^{G} \sqrt{\lambda_h} B_{m,h} \right) \left(\sqrt{\lambda_g} B_{m,g} - \lambda_g \sum_{h=1}^{G} \sqrt{\lambda_h} B_{m,h} \right)'.$$

Also, we denote $\bar{\mathbb{M}}_{pp}, \bar{\mathbb{M}}_{pq}, \bar{\mathbb{M}}_{qp}$, and $\bar{\mathbb{M}}_{qq}$ as sub-matrices of $\bar{\mathbb{M}}$ in a similar manner in the main text.

Proposition B.5 *Let Assumptions 1–5 hold, where Assumption 1-iii) is replaced with (B.38). Define* $\bar{\mathbb{M}}_{pp\cdot q} = \bar{\mathbb{M}}_{pp} - \bar{\mathbb{M}}_{pq} \bar{\mathbb{M}}_{qq}^{-1} \bar{\mathbb{M}}_{qp}$. *Then,*

(a) $\hat{\Omega}^c(\tilde{\theta}) \xrightarrow{d} \Lambda \bar{\mathbb{M}} \Lambda'$ *for any* $\sqrt{n}$*-consistent estimator* $\tilde{\theta}$.;

(b) $F_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c) \xrightarrow{d} \tilde{\mathbb{F}}_{2\infty} := \tilde{\mathbb{F}}_{2\infty}(\lambda)$, *where*

$$\begin{aligned}\tilde{\mathbb{F}}_{2\infty}(\lambda) \stackrel{d}{=} \frac{1}{p} &\left[\left(\sum_{g=1}^{G} \sqrt{\lambda_g} B_{p,g} \right) - \bar{\mathbb{M}}_{pq} \bar{\mathbb{M}}_{qq}^{-1} \left(\sum_{g=1}^{G} \sqrt{\lambda_g} B_{q,g} \right) \right]' \bar{\mathbb{M}}_{pp\cdot q}^{-1} \\ \times &\left[\left(\sum_{g=1}^{G} \sqrt{\lambda_g} B_{p,g} \right) - \bar{\mathbb{M}}_{pq} \bar{\mathbb{M}}_{qq}^{-1} \left(\sum_{g=1}^{G} \sqrt{\lambda_g} B_{q,g} \right) \right].\end{aligned}$$

(c) $t_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c) \xrightarrow{d} \tilde{\mathbb{T}}_{2\infty} := \tilde{\mathbb{T}}_{2\infty}(\lambda)$, *where*

$$\tilde{\mathbb{T}}_{2\infty}(\lambda) = \frac{\sum_{g=1}^{G} \sqrt{\lambda_g} B_{p,g} - \bar{\mathbb{M}}_{pq} \bar{\mathbb{M}}_{qq}^{-1} \sum_{g=1}^{G} \sqrt{\lambda_g} B_{q,g}}{\sqrt{\bar{\mathbb{M}}_{pp\cdot q}}}.$$

B.5.1 Proof of Proposition B.5

Proof. We only prove parts (a) and (b), as the proof of part (c) can be done similarly. To prove (a), note that the centered CCE can be represented as:

$$\hat{\Omega}^c(\tilde{\theta}) = \sum_{g=1}^{G} \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{L_g} \left(f_i^g(\tilde{\theta}) - \frac{1}{n} \sum_{h=1}^{G} \sum_{i=1}^{L_h} f_i^h(\tilde{\theta}) \right) \right.$$
$$\left. \times \frac{1}{\sqrt{n}} \sum_{i=1}^{L_g} \left(f_i^g(\tilde{\theta}) - \frac{1}{n} \sum_{h=1}^{G} \sum_{i=1}^{L_h} f_i^h(\tilde{\theta}) \right)' \right\}.$$

For each $g = 1, ..., G$, we use the same arguments in the proof of Lemma 7 and obtain that

$$\begin{aligned}
\frac{1}{\sqrt{n}} \sum_{i=1}^{L_g} f_i^g(\tilde{\theta}) &= \frac{1}{\sqrt{n}} \sum_{i=1}^{L_g} \left\{ f_i^g(\theta_0) + \frac{\partial f_i^g(\tilde{\theta})}{\partial \theta'} (\tilde{\theta} - \theta_0) \right\} (1 + o_p(1)) \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^{L_g} f_i^g(\theta_0) + \left(\frac{L_g}{n} \right) \cdot \frac{1}{L_g} \sum_{i=1}^{L_g} \frac{\partial f_i^g(\theta_0)}{\partial \theta'} \sqrt{n}(\tilde{\theta} - \theta_0)(1 + o_p(1)) \\
&= \left\{ \sqrt{\frac{L_g}{n}} \cdot \frac{1}{\sqrt{L_g}} \sum_{i=1}^{L_g} f_i^g(\theta_0) + \left(\frac{L_g}{n} \right) \Gamma \sqrt{n}(\tilde{\theta} - \theta_0) \right\} (1 + o_p(1)) \\
&= \left\{ \sqrt{\lambda_g} \cdot \frac{1}{\sqrt{L_g}} \sum_{i=1}^{L_g} f_i^g(\theta_0) + \lambda_g \Gamma \sqrt{n}(\tilde{\theta} - \theta_0) \right\} (1 + o_p(1)),
\end{aligned}$$

where the last equation follows from (B.38). Similarly,

$$\begin{aligned}
\frac{1}{\sqrt{n}} \sum_{i=1}^{L_g} \left(\frac{1}{n} \sum_{h=1}^{G} \sum_{i=1}^{L_h} f_i^h(\tilde{\theta}) \right) &= \frac{\lambda_g}{\sqrt{n}} \sum_{h=1}^{G} \sum_{i=1}^{L_h} f_i^h(\tilde{\theta})(1 + o_p(1) \\
&= \lambda_g \sum_{h=1}^{G} \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{L_h} \left(f_i^g(\theta_0) + \frac{\partial f_i^g(\tilde{\theta})}{\partial \theta'} (\tilde{\theta} - \theta_0) \right) \right\} (1 + o_p(1)) \\
&= \lambda_g \sum_{h=1}^{G} \left\{ \sqrt{\lambda_h} \cdot \frac{1}{\sqrt{L_h}} \sum_{i=1}^{L_h} f_i^h(\theta_0) + \lambda_h \Gamma \sqrt{n}(\tilde{\theta} - \theta_0) \right\} (1 + o_p(1)) \\
&= \left\{ \lambda_g \sum_{h=1}^{G} \left(\sqrt{\lambda_h} \cdot \frac{1}{\sqrt{L_h}} \sum_{i=1}^{L_h} f_i^h(\theta_0) \right) + \lambda_g \Gamma \sqrt{n}(\tilde{\theta} - \theta_0) \right\} (1 + o_p(1)),
\end{aligned}$$

where the last equation holds from $\sum_{h=1}^{G} \lambda_h = 1$. It then follows that

$$\begin{aligned}
\frac{1}{\sqrt{n}} \sum_{i=1}^{L_g} \left(f_i^g(\tilde{\theta}) - \frac{1}{n} \sum_{h=1}^{G} \sum_{i=1}^{L_g} f_i^g(\tilde{\theta}) \right) &= \left\{ \sqrt{\lambda_g} \cdot \frac{1}{\sqrt{L_g}} \sum_{i=1}^{L_g} f_i^g(\theta_0) \right. \\
&\quad \left. - \lambda_g \sum_{h=1}^{G} \left(\sqrt{\lambda_h} \cdot \frac{1}{\sqrt{L_h}} \sum_{i=1}^{L_h} f_i^h(\theta_0) \right) \right\} (1 + o_p(1)) \\
&\xrightarrow{d} \sqrt{\lambda_g} B_{m,g} - \lambda_g \sum_{h=1}^{G} \sqrt{\lambda_h} B_{m,h},
\end{aligned}$$

where the convergence holds jointly for $g = 1, \ldots, G$ by Assumption 1-ii). By continuous mapping theorem, this leads us to the desired result in (a). To prove part (b), note that the result in (a) and (B.38) imply

$$\begin{aligned}\sqrt{n}(\hat{\theta}_2^c - \theta_0) &= -\left(\Gamma' \left[\hat{\Omega}^c(\hat{\theta}_1)\right]^{-1} \Gamma\right)^{-1} \Gamma' \left[\hat{\Omega}^c(\hat{\theta}_1)\right]^{-1} \sqrt{n} g_n(\theta_0) + o_p(1) \\ &\xrightarrow{d} -\left[\Gamma_\Lambda' \bar{\mathbb{M}}^{-1} \Gamma_\Lambda\right]^{-1} \Gamma_\Lambda' \bar{\mathbb{M}}^{-1} \left(\sum_{g=1}^{G} \sqrt{\lambda_g} B_{m,g}\right),\end{aligned}$$

and leads us to obtain $F_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c) \xrightarrow{d} \tilde{\mathbb{F}}_{2\infty}$ where

$$\begin{aligned}\tilde{\mathbb{F}}_{2\infty} \quad : \quad &= \tilde{\mathbb{F}}_{2\infty}(\lambda) = \frac{1}{p} \cdot \left[R\left(\Gamma_\Lambda' \bar{\mathbb{M}}^{-1} \Gamma_\Lambda\right)^{-1} \Gamma_\Lambda' \bar{\mathbb{M}}^{-1} \left(\sum_{g=1}^{G} \sqrt{\lambda_g} B_{m,g}\right)\right]' \left[R\left(\Gamma_\Lambda' \bar{\mathbb{M}}^{-1} \Gamma_\Lambda\right)^{-1} R'\right]^{-1} \\ &\times \left[R\left(\Gamma_\Lambda' \bar{\mathbb{M}}^{-1} \Gamma_\Lambda\right)^{-1} \Gamma_\Lambda' \bar{\mathbb{M}}^{-1} \left(\sum_{g=1}^{G} \sqrt{\lambda_g} B_{m,g}\right)\right].\end{aligned}$$

Let $U \Sigma V'$ be a singular value decomposition (SVD) of Γ_Λ. Also, denote $\tilde{U} \tilde{\Sigma} \tilde{V}'$ as a SVD of RVA^{-1}. Then, it is check that the following rotational invariance properties hold

$$\left[\sum_{g=1}^{G} \sqrt{\lambda_g} B_{m,g}, \bar{\mathbb{M}}^{-1}\right] \stackrel{d}{=} \left[U' \left(\sum_{g=1}^{G} \sqrt{\lambda_g} B_{m,g}\right), U' \bar{\mathbb{M}}^{-1} U\right],$$

and

$$\begin{aligned}&\left[\sum_{g=1}^{G} \sqrt{\lambda_g} B_{d,g}, \bar{\mathbb{M}}_{dq}, \bar{\mathbb{M}}_{qq}^{-1}, \bar{\mathbb{M}}_{dd\cdot q}\right] \\ \stackrel{d}{=} \quad &\left[\tilde{V}' \sum_{g=1}^{G} \sqrt{\lambda_g} B_{d,g}, \tilde{V}' \bar{\mathbb{M}}_{dq}, \bar{\mathbb{M}}_{qq}^{-1}, \tilde{V}' \bar{\mathbb{M}}_{dd\cdot q} \tilde{V}\right]\end{aligned}$$

for $U'U = UU' = I_m$ and $\tilde{V}'\tilde{V} = \tilde{V}\tilde{V}' = I_d$. Then, the rest of proof is similar to that of Proposition 8-(a). The only differences is that the limit of CCE matrix $\bar{\mathbb{S}}$ and $\bar{B}_m$, and corresponding subcomponents are replaced with $\bar{\mathbb{M}}$, $\sum_{g=1}^{G} \sqrt{\lambda_g} B_{m,g}$, and corresponding submatrices, respectively. ■

B.6 Iterative Two-step and continuous updating Schemes

In this subsection, we consider two types of continuous updating schemes first suggested in Hansen et al. (1996). The first is motivated by the iterative scheme that updates the FOC of two-step GMM estimation until it converges. The FOC for $\hat{\theta}_{\mathrm{IE}}^{j}$ is

$$\hat{\Gamma}(\hat{\theta}_{\mathrm{IE}}^{j})'\hat{\Omega}^{-1}(\hat{\theta}_{\mathrm{IE}}^{j-1})g_n(\hat{\theta}_{\mathrm{IE}}^{j}) = 0 \text{ for } j \geq 1.$$

In view of the above FOC, $\hat{\theta}_{\mathrm{IE}}^{j}$ can be regarded as a generalized-estimating-equations (GEE) estimator, which is a class of estimators first proposed by Liang and Zeger (1986) and further studied by Jiang, Luan, and Wang (2007). When the number of iterations j goes to infinity until $\hat{\theta}_{\mathrm{IE}}^{j}$ converges, we obtain the continuous updating GEE (CU-GEE) estimator $\hat{\theta}_{\text{CU-GEE}}$. A recent paper by Hansen and Lee (2019) provides a rigorous theory for the existence of the iterated GMM estimator in Hansen et al (1996) in the context of both correctly specified and misspecified population moment conditions. The iterated GMM estimator is obtained by iterating the GMM object function instead of the FOC. Under correctly specified moment conditions, it is not difficult to check that the CU-GEE estimator considered here is asymptotically equivalent to the iterated GMM estimator. The FOC for $\hat{\theta}_{\text{CU-GEE}}$ is given by

$$\hat{\Gamma}(\hat{\theta}_{\text{CU-GEE}})'\hat{\Omega}^{-1}(\hat{\theta}_{\text{CU-GEE}})g_n(\hat{\theta}_{\text{CU-GEE}}) = 0. \tag{B.39}$$

Define

$$\hat{\Omega}^{*}(\hat{\theta}_{\text{CU-GEE}}) = \hat{\Omega}(\hat{\theta}_{\text{CU-GEE}}) - \left(\frac{1}{n}\sum_{g=1}^{G} L_g^2\right) \cdot g_n(\hat{\theta}_{\text{CU-GEE}})g_n(\hat{\theta}_{\text{CU-GEE}})'$$

which is an asymptotically equivalent eversion of the centered CCE $\hat{\Omega}^{c}(\hat{\theta}_{\text{CU-GEE}})$. While we employ the uncentered CCE, $\hat{\Omega}(\cdot)$ in the definition of $\hat{\theta}_{\text{CU-GEE}}$, it is not difficult to show that

$$\begin{aligned}
&\hat{\Gamma}(\hat{\theta}_{\text{CU-GEE}})'\hat{\Omega}^{-1}(\hat{\theta}_{\text{CU-GEE}})g_n(\hat{\theta}_{\text{CU-GEE}}) \\
&= \hat{\Gamma}(\hat{\theta}_{\text{CU-GEE}})'\left(\hat{\Omega}^{*}(\hat{\theta}_{\text{CU-GEE}})\right)^{-1} g_n(\hat{\theta}_{\text{CU-GEE}}) \cdot \frac{1}{1+\nu_n^{*}(\hat{\theta}_{\text{CU-GEE}})},
\end{aligned}$$

where

$$\nu_n^{*}(\hat{\theta}_{\text{CU-GEE}}) = \left(\frac{1}{n}\sum_{g=1}^{G} L_g^2\right) \cdot g_n(\hat{\theta}_{\text{CU-GEE}})'\left(\hat{\Omega}^{*}(\hat{\theta}_{\text{CU-GEE}})\right)^{-1} g_n(\hat{\theta}_{\text{CU-GEE}}).$$

Since $1/(1+\nu_n^{*}(\hat{\theta}_{\text{CU-GEE}}))$ is always positive, the first-order condition in (B.39) holds if and only if

$$\hat{\Gamma}(\hat{\theta}_{\text{CU-GEE}})'\left[\hat{\Omega}^{*}(\hat{\theta}_{\text{CU-GEE}})\right]^{-1} g_n(\hat{\theta}_{\text{CU-GEE}}) = 0, \tag{B.40}$$

which indicates that replacing the CCE in (B.39) by $\hat{\Omega}^{*}(\hat{\theta}_{\text{CU-GEE}})$ has no effect on the iteration GMM estimator.

The second CU scheme continuously updates the GMM criterion function, which leads to the familiar continuous updating GMM (CU-GMM) estimator:

$$\hat{\theta}_{\text{CU-GMM}} = \arg\min_{\theta\in\Theta} g_n(\theta)'\hat{\Omega}^{-1}(\theta)g_n(\theta).$$

Although we use the uncentered CEE $\hat{\Omega}(\theta)$ in the above definition, the original idea of $\hat{\theta}_{\text{CU-GMM}}$ in Hansen, Heaton and Yaron (1996) is based on the centered CCE weighting matrix $\hat{\Omega}^c(\theta)$. With $M_n = n^{-1}\sum_{g=1}^{G} L_g^2$, it is easy to show that

$$\begin{aligned} M_n \cdot g_n(\theta)\hat{\Omega}^{-1}(\theta)g_n(\theta) &= M_n \cdot g_n(\theta)'\hat{\Omega}^{-1}(\theta)\left[\hat{\Omega}(\theta) - M_n \cdot g_n(\theta)g_n(\theta)'\right]\left[\hat{\Omega}^*(\theta)\right]^{-1} g_n(\theta) \\ &= M_n \cdot g_n(\theta)'\left[\hat{\Omega}^*(\theta)\right]^{-1} g_n(\theta)\left\{1 - M_n \cdot g_n(\theta)'\hat{\Omega}^{-1}(\theta)g_n(\theta)\right\}. \end{aligned}$$

Thus, we have

$$M_n \cdot g_n(\theta)'\left[\hat{\Omega}^*(\theta)\right]^{-1} g_n(\theta) = \frac{M_n \cdot g_n(\theta)'\hat{\Omega}^{-1}(\theta)g_n(\theta)}{1 - M_n \cdot g_n(\theta)'\hat{\Omega}^{-1}(\theta)g_n(\theta)}.$$

The above equation reveals the fact that the CU-GMM estimator will not change if the uncentered weighting matrix $\hat{\Omega}(\theta)$ is replaced by the centered one $\hat{\Omega}^*(\theta)$, that is,

$$\hat{\theta}_{\text{CU-GMM}} = \arg\min_{\theta\in\Theta} g_n(\theta)'\left[\hat{\Omega}^*(\theta)\right]^{-1} g_n(\theta). \tag{B.41}$$

Similar to the centered two-step GMM estimator, the two CU estimators can be regarded as having a built-in recentering mechanism via the version of CCE weight matrix $\hat{\Omega}^*(\theta)$. For this reason, the limiting distributions of the two CU estimators are the same as that of the centered two-step GMM estimator, as is shown below.

Proposition B.6 *Let Assumptions 1, 3∼5 hold. Assume that $\hat{\theta}_{CU\text{-}GEE}$ and $\hat{\theta}_{CU\text{-}GMM}$ are $\sqrt{n}$-consistent. Then*

$$\sqrt{n}(\hat{\theta}_{CU\text{-}GEE} - \theta_0) \xrightarrow{d} -\left[\Gamma'\left(\Omega_\infty^c\right)^{-1}\Gamma\right]^{-1}\Gamma'\left(\Omega_\infty^c\right)^{-1}\Lambda\sqrt{G}\bar{B}_m$$

and

$$\sqrt{n}(\hat{\theta}_{CU\text{-}GMM} - \theta_0) \xrightarrow{d} -\left[\Gamma'\left(\Omega_\infty^c\right)^{-1}\Gamma\right]^{-1}\Gamma'\left(\Omega_\infty^c\right)^{-1}\Lambda\sqrt{G}\bar{B}_m.$$

The proof of proposition shows the asymptotic equivalence between two versions of centered CCE weighting matrix, i.e. $\hat{\Omega}^*(\tilde{\theta}) = \hat{\Omega}^c(\tilde{\theta}) + o_p(1)$ for any $\sqrt{n}$-consistent $\tilde{\theta}$. As a result, the CU estimators and the centered two-step GMM estimator are asymptotically equivalent under the fixed-G asymptotics. Based on the two CU estimators, we construct the Wald statistics as

$$F_{\hat{\Omega}^c(\hat{\theta}_{\text{CU-GEE}})}(\hat{\theta}_{\hat{\theta}_{\text{CU-GEE}}}) = \frac{1}{p}(R\hat{\theta}_{\text{CU-GEE}} - r)'\{R\widehat{var}_{\hat{\Omega}^c(\hat{\theta}_{\text{CU-GEE}})}(\hat{\theta}_{\text{CU-GEE}})R'\}^{-1}(R\hat{\theta}_{\text{CU-GEE}} - r) \tag{B.42}$$

and

$$F_{\hat{\Omega}^c(\hat{\theta}_{\text{CU-GMM}})}(\hat{\theta}_{\hat{\theta}_{\text{CU-GMM}}}) = \frac{1}{p}(R\hat{\theta}_{\text{CU-GMM}} - r)'\{R\widehat{var}_{\hat{\Omega}^c(\hat{\theta}_{\text{CU-GMM}})}(\hat{\theta}_{\text{CU-GMM}})R'\}^{-1}(R\hat{\theta}_{\text{CU-GMM}} - r). \tag{B.43}$$

We construct $t_{\hat{\Omega}^c(\hat{\theta}_{\text{CU-GEE}})}(\hat{\theta}_{\text{CU-GEE}})$ and $t_{\hat{\Omega}^c(\hat{\theta}_{\text{CU-GMM}})}(\hat{\theta}_{\text{CU-GMM}})$ in a similar way when $p = 1$. It follows from Proposition B.6 that the Wald statistics based on $\hat{\theta}_{\text{CU-GEE}}$ and $\hat{\theta}_{\text{CU-GMM}}$ are asymptotically equivalent to $F_{\hat{\Omega}^c(\hat{\theta}_\cdot)}(\hat{\theta}_2^c)$. As a result,

$$F_{\hat{\Omega}^c(\hat{\theta}_{\text{CU-GEE}})}(\hat{\theta}_{\text{CU-GEE}}) \xrightarrow{d} \mathbb{F}_{2\infty} \text{ and } F_{\hat{\Omega}^c(\hat{\theta}_{\text{CU-GMM}})}(\hat{\theta}_{\text{CU-GMM}}) \xrightarrow{d} \mathbb{F}_{2\infty}.$$

Similarly,

$$t_{\hat{\Omega}^c(\hat{\theta}_{\text{CU-GEE}})}(\hat{\theta}_{\text{CU-GEE}}) \xrightarrow{d} \mathbb{T}_{2\infty} \text{ and } t_{\hat{\Omega}^c(\hat{\theta}_{\text{CU-GMM}}}(\hat{\theta}_{\text{CU-GMM}}) \xrightarrow{d} \mathbb{T}_{2\infty}.$$

Note that changing the centered CCE $\hat{\Omega}^c(\cdot)$ by $\hat{\Omega}^*(\cdot)$ is innocuous to our fixed-G limiting distributions. In summary, we have shown that all three estimators $\hat{\theta}_2^{\text{c}}$, $\hat{\theta}_{\text{CU-GEE}}$ and $\hat{\theta}_{\text{CU-GMM}}$, and the corresponding Wald test statistics converge in distribution to the same nonstandard distributions. Proposition 8-(c) and (d) continues to hold for the CU-GEE and CU-GMM estimators, leading to the asymptotic equivalence of the three test statistics based on the CU-type estimators. That is, the CU-GMM estimator shares the first order fixed-smoothing limit with the two-step GMM estimator in our paper. Similar results have been found in a recent paper by Zhang (2016) in a time series setting who develops the fixed-smoothing asymptotic theory for the CU-GMM estimator.

Together with Theorem 10, Proposition B.6 imply that the modified of Wald, LR,LM, and t statistics based on the CU estimators are all asymptotically F and t distributed under the fixed-G asymptotics. For the finite-sample corrected variance formula, we have the following expansion

$$\begin{aligned}&\sqrt{n}(\hat{\theta}_{\text{CU-GEE}}-\theta_0)\\ &= -\left(\Gamma'\left(\hat{\Omega}^c(\theta_0)\right)^{-1}\Gamma\right)^{-1}\Gamma'\left(\hat{\Omega}^c(\theta_0)\right)^{-1}\sqrt{n}g_n(\theta_0)+\mathcal{E}_{2n}\sqrt{n}(\hat{\theta}_{\text{CU-GEE}}-\theta_0)+o_p(1)\,. \end{aligned} \tag{B.44}$$

This can be regarded as a special case of (27) wherein the first-step estimator $\hat{\theta}_1$ is replaced by the CU-GEE estimator. Thus,

$$\sqrt{n}(\hat{\theta}_{\text{CU-GEE}}-\theta_0) \overset{a}{\sim} -(I_d-\mathcal{E}_{2n})^{-1}\left(\Gamma'\left(\hat{\Omega}^c(\theta_0)\right)^{-1}\Gamma\right)^{-1}\Gamma'\left(\hat{\Omega}^c(\theta_0)\right)^{-1}\sqrt{n}g_n(\theta_0), \tag{B.45}$$

We can obtain the same expression for the CU-GMM estimator $\sqrt{n}(\hat{\theta}_{\text{CU-GMM}}-\theta_0)$.

In view of the representation in (B.45), the corrected variance estimator for the CU type estimators can be constructed as follows:

$$\begin{aligned}\widehat{var}^{\text{w}}_{\hat{\Omega}^c(\hat{\theta}_{\text{CU-GEE}})}(\hat{\theta}_{\text{CU-GEE}}) &= \left(I_d-\widehat{\mathcal{E}}_{\text{CU-GEE}}\right)^{-1}\widehat{var}\left(\hat{\theta}_{\text{CU-GEE}}\right)\left(I_d-\widehat{\mathcal{E}}'_{\text{CU-GEE}}\right)^{-1}\\ \widehat{var}^{\text{w}}_{\hat{\Omega}^c(\hat{\theta}_{\text{CU-GEE}})}(\hat{\theta}_{\text{CU-GMM}}) &= \left(I_d-\widehat{\mathcal{E}}_{\text{CU-GMM}}\right)^{-1}\widehat{var}\left(\hat{\theta}_{\text{CU-GMM}}\right)\left(I_d-\widehat{\mathcal{E}}'_{\text{CU-GMM}}\right)^{-1},\end{aligned}$$

where

$$\begin{aligned}\widehat{\mathcal{E}}_{\text{CU-GEE}}[.,j] &= \left\{\hat{\Gamma}'\left[\hat{\Omega}^c(\hat{\theta}_{\text{CU-GEE}})\right]^{-1}\hat{\Gamma}'\right\}^{-1}\\ &\quad\times\hat{\Gamma}'\left\{\left[\hat{\Omega}^c(\hat{\theta}_{\text{CU-GEE}})\right]^{-1}\frac{\partial\hat{\Omega}^c(\hat{\theta}_{\text{CU-GEE}})}{\partial\theta_j}\left[\hat{\Omega}^c(\hat{\theta}_{\text{CU-GEE}})\right]^{-1}\right\}g_n(\hat{\theta}_{\text{CU-GEE}})\end{aligned}$$

and $\widehat{\mathcal{E}}_{\text{CU-GMM}}$ is defined in the same way but with $\hat{\theta}_{\text{CU-GEE}}$ replaced by $\hat{\theta}_{\text{CU-GMM}}$. With the finite sample corrected and adjusted variance estimators in place, the Wald and t statistics based on the CU estimators also converge in distribution to the same nonstandard distributions in Proposition 8 (a) and (b), respectively.

B.6.1 Proof of Proposition B.6

Proof. For the result with CU-GEE estimator $\hat{\theta}_{\text{CU-GEE}}$, we have

$$\sqrt{n}(\hat{\theta}_{\text{CU-GEE}} - \theta_0) = -\left(\Gamma' \left[\hat{\Omega}^*(\hat{\theta}_{\text{CU-GEE}})\right]^{-1} \Gamma\right)^{-1} \Gamma' \left[\hat{\Omega}^*(\hat{\theta}_{\text{CU-GEE}})\right]^{-1} \sqrt{n} g_n(\theta_0) + o_p(1).$$

We first show that

$$\hat{\Omega}^*(\tilde{\theta}) = \hat{\Omega}^c(\tilde{\theta}) + o_p(1) \tag{B.46}$$

for any $\sqrt{n}$-consistent estimator $\tilde{\theta}$. Recall that

$$\begin{aligned}
\hat{\Omega}^*(\tilde{\theta}) &= \hat{\Omega}(\tilde{\theta}) - \left(\frac{1}{n}\sum_{g=1}^{G} L_g^2\right) \cdot g_n(\tilde{\theta}) g_n(\tilde{\theta}); \\
\hat{\Omega}^c(\tilde{\theta}) &= \frac{1}{n}\sum_{g=1}^{G}\left(\sum_{i=1}^{L_g}(f_i^g(\tilde{\theta}) - g_n(\tilde{\theta})\right)\left(\sum_{i=1}^{L_g} f_i^g(\tilde{\theta}) - g_n(\tilde{\theta})\right)' \\
&= \hat{\Omega}(\tilde{\theta}) - g_n(\tilde{\theta}) \cdot \left(\frac{1}{n}\sum_{g=1}^{G} L_g \sum_{i=1}^{L_g} f_i^g(\tilde{\theta})\right)' \\
&\quad - \left(\frac{1}{n}\sum_{g=1}^{G} L_g \sum_{i=1}^{L_g} f_i^g(\tilde{\theta})\right)' \cdot g_n(\tilde{\theta})' + \left(\frac{1}{n}\sum_{g=1}^{G} L_g^2\right) \cdot g_n(\tilde{\theta}) g_n(\tilde{\theta})'.
\end{aligned}$$

Thus, (B.46) can be proved by showing

$$g_n(\tilde{\theta})\left(\frac{1}{n}\sum_{g=1}^{G} L_g \sum_{i=1}^{L_g} f_i^g(\tilde{\theta})\right)' = \left(\frac{1}{n}\sum_{g=1}^{G} L_g^2\right) \cdot g_n(\tilde{\theta}) g_n(\tilde{\theta})' + o_p(1). \tag{B.47}$$

By Assumption 1-iii) and $\sqrt{n}$-consistency of $\tilde{\theta}$,

$$\begin{aligned}
g_n(\tilde{\theta})\left(\frac{1}{n}\sum_{g=1}^{G} L_g \sum_{i=1}^{L_g} f_i^g(\tilde{\theta})\right)' &= \sqrt{\bar{L}} g_n(\tilde{\theta})\left(\frac{1}{G}\sum_{g=1}^{G}\left(\frac{L_g}{\bar{L}}\right)\frac{1}{\sqrt{\bar{L}}}\sum_{i=1}^{L_g} f_i^g(\tilde{\theta})\right)' \\
&= \sqrt{\bar{L}} g_n(\tilde{\theta})\left(\frac{1}{G}\sum_{g=1}^{G}\frac{1}{\sqrt{\bar{L}}}\sum_{i=1}^{L_g} f_i^g(\tilde{\theta})\right)'(1 + o_p(1)) \\
&= \bar{L} g_n(\tilde{\theta}) g_n(\tilde{\theta})'(1 + o_p(1)).
\end{aligned}$$

Similarly, we obtain

$$\begin{aligned}
\left(\frac{1}{n}\sum_{g=1}^{G} L_g^2\right) \cdot g_n(\tilde{\theta}) g_n(\tilde{\theta})' &= \sum_{g=1}^{G}\left(\frac{L_g}{n}\right)^2 \cdot n g_n(\tilde{\theta}) g_n(\tilde{\theta})' \\
&= \frac{1}{G}\sum_{g=1}^{G}\left(\frac{L_g}{\bar{L}}\right)^2 \cdot \bar{L} g_n(\tilde{\theta}) g_n(\tilde{\theta})' \\
&= \bar{L} g_n(\tilde{\theta}) g_n(\tilde{\theta})'(1 + o_p(1)),
\end{aligned}$$

which gives the desired result in (B.47).

Now, using $\sqrt{n}$-consistency of $\hat{\theta}_{\text{CU-GEE}}$, we can apply Lemma 7 to obtain $\hat{\Omega}^*(\hat{\theta}_{\text{CU-GEE}}) = \hat{\Omega}^c(\theta_0) + o_p(1)$. Invoking the continuous mapping theorem yields

$$\sqrt{n}(\hat{\theta}_{\text{CU-GEE}} - \theta_0) \xrightarrow{d} -\left\{\Gamma' \left(\Omega^c_\infty\right)^{-1} \Gamma\right\}^{-1} \left\{\Gamma' \left(\Omega^c_\infty\right)^{-1} \Lambda\sqrt{G}\bar{B}_m\right\},$$

as desired.

For the CU-GMM estimator, we let $\hat{\Gamma}^j(\hat{\theta}_{\text{CU-GMM}})$ be the j-th column of $\hat{\Gamma}^j(\hat{\theta}_{\text{CU-GMM}})$. Then, the FOC with respect to the j-th element of $\hat{\theta}_{\text{CU-GMM}}$ is

$$\begin{aligned} 0 &= \hat{\Gamma}^j(\hat{\theta}_{\text{CU-GMM}})' \left[\hat{\Omega}^*(\hat{\theta}_{\text{CU-GMM}})\right]^{-1} g_n(\hat{\theta}_{\text{CU-GMM}}) \\ &\quad - g_n(\hat{\theta}_{\text{CU-GMM}})' \left[\hat{\Omega}^*(\hat{\theta}_{\text{CU-GMM}})\right]^{-1} \Upsilon_j(\hat{\theta}_{\text{CU-GMM}}) \left[\hat{\Omega}^*(\hat{\theta}_{\text{CU-GMM}})\right]^{-1} g_n(\hat{\theta}_{\text{CU-GMM}}), \end{aligned} \tag{B.48}$$

where

$$\begin{aligned} \Upsilon^*_j(\theta) &= \frac{1}{n}\sum_{g=1}^{G}\left(\sum_{i=1}^{L_g} f_i^g(\theta)\right)\left(\sum_{i=1}^{L_g}\frac{\partial f_i^g(\theta)}{\partial\theta_j}\right)' - \left(\frac{1}{n}\sum_{g=1}^{G} L_g^2\right)\cdot g_n(\theta)\left(\frac{\partial g_n(\theta)}{\partial\theta_j}\right)' \\ &= \frac{1}{n}\sum_{g=1}^{G}\left(\sum_{i=1}^{L_g} f_i^g(\theta)\right)\left(\sum_{i=1}^{L_g}\frac{\partial f_i^g(\theta)}{\partial\theta_j}\right)' - \bar{L}(1+o(1))\cdot g_n(\theta)\left(\frac{\partial g_n(\theta)}{\partial\theta_j}\right)'. \end{aligned}$$

Thus, the second term in (B.48) can be rewritten as

$$\begin{aligned} & g_n(\hat{\theta}_{\text{CU-GMM}})' \left[\hat{\Omega}^*(\hat{\theta}_{\text{CU-GMM}})\right]^{-1} \Upsilon^*_j(\hat{\theta}_{\text{CU-GMM}}) \left[\hat{\Omega}^*(\hat{\theta}_{\text{CU-GMM}})\right]^{-1} g_n(\hat{\theta}_{\text{CU-GMM}}) \\ &= \sqrt{\bar{L}} g_n(\hat{\theta}_{\text{CU-GMM}})' \left[\hat{\Omega}^*(\hat{\theta}_{\text{CU-GMM}})\right]^{-1} \left[\frac{1}{G}\sum_{g=1}^{G}\left(\frac{1}{\bar{L}}\sum_{i=1}^{L_g} f_i^g(\hat{\theta}_{\text{CU-GMM}})\right)\right. \\ &\quad \times \left.\left\{\left(\frac{1}{\bar{L}}\sum_{i=1}^{L_g}\frac{\partial f_i^g(\hat{\theta}_{\text{CU-GMM}})}{\partial\theta_j}\right) - \frac{1}{G}\sum_{g=1}^{G}\left(\frac{1}{\bar{L}}\sum_{i=1}^{L_g}\frac{\partial f_i^g(\hat{\theta}_{\text{CU-GMM}})}{\partial\theta_j}\right)\right\}'\right] \\ &\quad \times \left[\hat{\Omega}^*(\hat{\theta}_{\text{CU-GMM}})\right]^{-1} \sqrt{\bar{L}} g_n(\hat{\theta}_{\text{CU-GMM}})(1+o_p(1)). \end{aligned}$$

Given that $\hat{\theta}_{\text{CU-GMM}} = \theta_0 + O_p(\bar{L}^{-1/2})$ and $\hat{\Omega}^*(\hat{\theta}_{\text{CU-GMM}}) = \hat{\Omega}^c(\theta_0) + o_p(1)$, we have

$$\begin{aligned} \hat{\Omega}^*(\hat{\theta}_{\text{CU-GMM}}) &= O_p(1), \\ \sqrt{\bar{L}} g_n(\hat{\theta}_{\text{CU-GMM}}) &= \frac{1}{G}\sum_{g=1}^{G}\left(\frac{1}{\sqrt{\bar{L}}}\sum_{i=1}^{L_g} f_i^g(\theta_0)\right) + \Gamma\sqrt{\bar{L}}(\hat{\theta}_{\text{CU-GMM}} - \theta_0) + o_p(1) = O_p(1), \\ \frac{1}{\bar{L}}\sum_{i=1}^{L_g} f_i^g(\hat{\theta}_{\text{CU-GMM}}) &= \frac{1}{\bar{L}}\sum_{i=1}^{L_g} f_i^g(\theta_0) + \frac{1}{\bar{L}}\sum_{i=1}^{L_g}\frac{\partial f_i^g(\tilde{\theta})}{\partial\theta}(\hat{\theta}_{\text{CU-GMM}} - \theta_0) = O_p\left(\frac{1}{\sqrt{\bar{L}}}\right), \end{aligned}$$

and for each $g = 1, ..., G$,

$$\left(\frac{1}{\bar{L}}\sum_{i=1}^{L_g} f_i^g(\hat{\theta}_{\text{CU-GMM}})\right)\left\{\left(\frac{1}{\bar{L}}\sum_{i=1}^{L_g}\frac{\partial f_k^g(\hat{\theta}_{\text{CU-GMM}})}{\partial\theta}\right) - \frac{1}{G}\sum_{g=1}^{G}\left(\frac{1}{\bar{L}}\sum_{i=1}^{L_g}\frac{\partial f_i^g(\hat{\theta}_{\text{CU-GMM}})}{\partial\theta}\right)\right\}'$$
$$= O_p\left(\frac{1}{\sqrt{\bar{L}}}\right)\cdot o_p(1) = o_p\left(\frac{1}{\sqrt{\bar{L}}}\right).$$

Combining these together, the second term in FOC in (B.48) is $o_p(\bar{L}^{-1/2})$. As a result,

$$\hat{\Gamma}(\hat{\theta}_{\text{CU-GMM}})'\left[\hat{\Omega}^*(\hat{\theta}_{\text{CU-GMM}})\right]^{-1} g_n(\hat{\theta}_{\text{CU-GMM}}) = o_p\left(\frac{1}{\sqrt{\bar{L}}}\right).$$

Using this result and $\hat{\Omega}^*(\hat{\theta}_{\text{CU-GMM}}) = \hat{\Omega}^c(\theta_0) + o_p(1)$, we obtain the desired result as

$$\begin{aligned}\sqrt{n}(\hat{\theta}_{\text{CU-GMM}} - \theta_0) &= -\left\{\Gamma'\left[\hat{\Omega}^*(\hat{\theta}_{\text{CU-GMM}})\right]^{-1}\Gamma\right\}^{-1}\Gamma'\left[\hat{\Omega}^*(\hat{\theta}_{\text{CU-GMM}})\right]^{-1}\sqrt{n}g_n(\theta_0) + o_p(1)\\ &\xrightarrow{d} -\left\{\Gamma'\left(\Omega_\infty^c\right)^{-1}\Gamma\right\}^{-1}\Gamma'\left(\Omega_\infty^c\right)^{-1}\Lambda\sqrt{G}\bar{B}_m.\end{aligned}$$

■

B.7 Application to linear dynamic panel model

This subsection illustrates how to implement the GMM testing procedures in the context of linear dynamic panel model under clustered dependence. Consider

$$y_{it} = \gamma y_{it-1} + x_{it}'\beta + \eta_i + u_{it}, \tag{B.49}$$

for $i = 1, ..., n$, $t = 1, ..., T$, where $x_{it} = (x_{it}^1, ..., x_{it}^{d-1})' \in \mathbb{R}^{d-1}$. The unknown parameter vector is $\theta = (\gamma, \beta')' \in \mathbb{R}^d$. We assume that the vector of regressors $w_{it} = (y_{it-1}, x_{it}')'$ is correlated with η_i and is predetermined with respect to u_{it}, i.e., $E(w_{it}u_{it+s}) = 0$ for $s = 0, ..., T-t$.

When T is small, popular panel estimators such as the fixed-effects estimator or first-differenced estimator suffer from the Nickel bias (Nickel, 1981). Anderson and Hsiao (1981) consider the first-differenced equation

$$\Delta y_{it} = \Delta w_{it}'\theta + \Delta u_{it},\ t = 2, ..., T,$$

and propose a consistent IV estimator that employs the lagged w_{it}, e.g. w_{it-1} or Δw_{it-1}, as the instrument. Building upon Anderson and Hsiao (1981), Arellano and Bond (1991, AB hereafter) examine the problem in a GMM framework and find $dT(T-1)/2$ sequential instruments:

$$\underset{(T-1)\times d(T-1)T/2}{Z_i} = diag(z_{i2}', ..., z_{iT}'),$$

where $z_{it} = (y_{i0}, ..., y_{it-2}, x_{i1}', ..., x_{it-1}')'$, $2 \le t \le T$. The moment conditions are then given by

$$E\left(Z_i'\Delta u_i\right) = 0,$$

where Δu_i is the $(T-1)$ vector $(\Delta u_{i2}, ..., \Delta u_{iT})'$. The original AB method assumes away cross-sectional dependence, but clustered dependence can be easily accommodated. Here we assume that the moment vector $\{Z_i'\Delta u_i\}_{i=1}^n$ can be partitioned into independent clusters. That is, $\{Z_i'\Delta u_i\}_{i=1}^n = \cup_{g=1}^G \cup_{k=1}^{L_g} \{Z_k^{g\prime}\Delta u_k^g\}$ with $Z_k^{g\prime}\Delta u_k^g$ and $Z_l^{h\prime}\Delta u_l^h$ being independent for all $g \neq h$.

The first-step GMM estimator with initial weighting matrix W_n^{-1} is given by

$$\hat{\theta}_1 = \left(\Delta w' Z W_n^{-1} Z'\Delta w\right)^{-1} \Delta w' Z W_n^{-1} Z'\Delta y,$$

where Z' is the $dT(T-1)/2 \times n(T-1)$ matrix $(Z_1', Z_2', ..., Z_n')$, Δw_i is the $(T-1)\times d$ matrix $(\Delta w_{i2,}..., \Delta w_{iT})'$, Δy_i is the $(\Delta y_{i2,}..., \Delta y_{iT})'$, Δw and Δy are $(\Delta w_1', ..., \Delta w_n')'$ and $(\Delta y_1', ..., \Delta y_n')'$, respectively. The examples of W_n's can be $Z'Z/n$ for 2SLS and $n^{-1}\sum_{i=1}^n Z_i'HZ_i$ where H is a matrix that consists with 2's on the main diagonal, with -1's on the main diagonal, and zeros elsewhere.

The Wald statistic[3] for testing $H_0 : R\theta_0 = r$ vs $H_1 : R\theta_0 \neq r$ is given by

$$F(\hat{\theta}_1) := \frac{1}{p}(R\hat{\theta}_1 - r)'\left\{R\widehat{var}(\hat{\theta}_1)R'\right\}^{-1}(R\hat{\theta}_1 - r),$$

where

$$\widehat{var}(\hat{\theta}_1) = n\left(\Delta w' Z W_n^{-1} Z'\Delta w\right)^{-1}\left(\Delta w' Z W_n^{-1}\hat{\Omega}(\hat{\theta}_1) W_n^{-1} Z'\Delta w\right)\left(\Delta w' Z W_n^{-1} Z'\Delta w\right)^{-1}.$$

Let $Z_{(g)}$ be the $L(T-1)\times dT(T-1)/2$ matrix obtained by stacking all Z_i's belonging to cluster g. Similarly, let $\Delta\hat{u}_{(g)}$ be the $L_g(T-1)$ stacked vector of the estimated first-differenced

[3] The formula for the t statistic, which is omitted here, can be similarly constructed.

errors $\Delta\hat{u}_i = \Delta y_i - \Delta w_i'\hat{\theta}_1$. Then, in the presence of clustered dependence, the CCE and centered CCE are constructed as

$$\hat{\Omega}(\hat{\theta}_1) = \frac{1}{G}\sum_{g=1}^{G}\left(\frac{Z_{(g)}'\Delta\hat{u}_{(g)}}{\sqrt{\bar{L}}}\right)\left(\frac{Z_{(g)}'\Delta\hat{u}_{(g)}}{\sqrt{\bar{L}}}\right)'$$

and

$$\hat{\Omega}^c(\hat{\theta}_1) = \frac{1}{G}\sum_{g=1}^{G}\left(\frac{Z_{(g)}'\Delta\hat{u}_{(g)}}{\sqrt{\bar{L}}} - \frac{1}{G}\sum_{\tilde{g}=1}^{G}\frac{Z_{(\tilde{g})}'\Delta\hat{u}_{(\tilde{g})}}{\sqrt{\bar{L}}}\right)\left(\frac{Z_{(g)}'\Delta\hat{u}_{(g)}}{\sqrt{\bar{L}}} - \frac{1}{G}\sum_{\tilde{g}=1}^{G}\frac{Z_{(\tilde{g})}'\Delta\hat{u}_{(\tilde{g})}}{\sqrt{\bar{L}}}\right)'.$$

Using the centered CCE $\hat{\Omega}^c(\hat{\theta}_1)$ as the weighting matrix, the two-step GMM estimator $\hat{\theta}_2^c$ is

$$\hat{\theta}_2^c = \left(\Delta w'Z\left[\hat{\Omega}^c(\hat{\theta}_1)\right]^{-1}Z'\Delta w\right)^{-1}\Delta w'Z\left[\hat{\Omega}^c(\hat{\theta}_1)\right]^{-1}Z'\Delta y,$$

and the t and Wald statistics for $\hat{\theta}_2^c$ is

$$F_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c) = \frac{1}{p}(R\hat{\theta}_2^c - r)'\{R\widehat{var}_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2)R'\}^{-1}(R\hat{\theta}_2^c - r),$$
$$\widehat{var}_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c) = n\left\{\Delta w'Z\left[\hat{\Omega}^c(\hat{\theta}_1)\right]^{-1}Z'\Delta w\right\}^{-1}.$$

Under the conventional large-G asymptotics, both $F(\hat{\theta}_1)$ and $F_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c)$ are asymptotically χ_p^2/p. Under our fixed-G asymptotics, we have

$$F(\hat{\theta}_1) \xrightarrow{d} \frac{G}{G-p}F_{p,G-p} \text{ and}$$
$$F_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c) \xrightarrow{d} \frac{G}{G-p-q}F_{p,G-p-q}\left(\|\Delta\|^2\right). \tag{B.50}$$

In addition to utilizing these new approximations, we suggest a variance correction in order to capture the higher order effect of $\hat{\theta}_1$ on $\hat{\Omega}^c(\hat{\theta}_1)$. The finite sample corrected variance is

$$\widehat{var}^{\mathrm{w}}_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c) = \widehat{var}_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c) + \widehat{\mathcal{E}}_n\widehat{var}_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c) + \widehat{var}_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c)\widehat{\mathcal{E}}_n' + \widehat{\mathcal{E}}_n\widehat{var}(\hat{\theta}_1)\widehat{\mathcal{E}}_n', \tag{B.51}$$

where the j-th column is given by

$$\widehat{\mathcal{E}}_n[.,j] = -\left(\Delta w'Z\left[\hat{\Omega}^c(\hat{\theta}_1)\right]^{-1}Z'\Delta w\right)^{-1}\Delta w'Z\left[\hat{\Omega}^c(\hat{\theta}_1)\right]^{-1}\times$$
$$\left.\frac{\partial\hat{\Omega}^c(\theta)}{\partial\theta_j}\right|_{\theta=\hat{\theta}_1}\left[\hat{\Omega}^c(\hat{\theta}_1)\right]^{-1}Z'\Delta\hat{u}_2,$$
$$\Delta\hat{u}_2 = \Delta y - \Delta w\hat{\theta}_2^c,$$

and

$$\left.\frac{\partial\hat{\Omega}^c(\theta)}{\partial\theta_j}\right|_{\theta=\hat{\theta}_1} = \Upsilon_j(\hat{\theta}_1) + \Upsilon_j'(\hat{\theta}_1),$$

$$\Upsilon_j(\hat{\theta}_1) = -\frac{1}{G}\sum_{g=1}^{G}\left(\frac{Z_{(g)}'\Delta w_{j,(g)}}{\sqrt{L}} - \frac{1}{G}\sum_{\tilde{g}=1}^{G}\frac{Z_{(\tilde{g})}'\Delta w_{j,(\tilde{g})}}{\sqrt{L}}\right)\left(\frac{Z_{(g)}'\Delta\hat{u}_{(g)}}{\sqrt{L}} - \frac{1}{G}\sum_{\tilde{g}=1}^{G}\frac{Z_{(\tilde{g})}'\Delta\hat{u}_{(\tilde{g})}}{\sqrt{L}}\right)',$$

$$\underset{(T-1)\times d}{\Delta w_i} = (\Delta w_{1,i}, ..., \Delta w_{d,i}) \text{ and } \underset{l(T-1)\times d}{\Delta w_{(g)}} = (\Delta w_{1,(g)}, ..., \Delta w_{d,(g)})$$

for each $j = 1, ..., d$. Here, $\Delta w_{(g)}$ is a $L_g(T-1)\times d$ matrix that stacks $\{\Delta w_i\}_{i=1}^n$ belonging to the group g.

Based on the finite sample corrected variance estimator in (B.51) and the usual J statistic, we construct the modified Wald and t statistics

$$\tilde{F}^{\mathrm{w}}_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c) = \frac{G-p-q}{G}\frac{F^{\mathrm{w}}_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c)}{1+\frac{1}{G}J(\hat{\theta}_2^c)}, \tag{B.52}$$

where

$$\begin{aligned} J(\hat{\theta}_2^c) &= n\cdot g_n(\hat{\theta}_2^c)'\left(\hat{\Omega}^c(\hat{\theta}_2^c)\right)^{-1}(\theta)g_n(\hat{\theta}_2^c) \\ &= \left(\frac{Z'\Delta u(\hat{\theta}_2^c)}{\sqrt{n}}\right)'\left[\hat{\Omega}^c(\hat{\theta}_2^c)\right]^{-1}\left(\frac{Z'\Delta u(\hat{\theta}_2^c)}{\sqrt{n}}\right). \end{aligned}$$

From the t and F limit theory in Section 4 and the fixed-G approximation of the finite sample corrected variance in Section 5, we have

$$\tilde{F}^{\mathrm{w}}_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c) \xrightarrow{d} F_{p,G-p-q} \tag{B.53}$$

and

$$\frac{G-q}{Gq}J(\hat{\theta}_2^c) \xrightarrow{d} F_{q,G-q}.$$

The critical values from the t and F distributions are readily available from statistical tables.

B.8 Procedure for cluster-robust Hall and Horowitz (1996) two-step GMM-bootstrap

Conditioning on the original sample $\{Z_{(g)}, \Delta u_{(g)}(\theta), \Delta y_{(g)}, \Delta w_{(g)}\}_{g=1}^{G}$, let $\{Z_i^*, \Delta y_i^*, \Delta u_i^*(\theta), \Delta w_i^*\}_{i=1}^{n} = \cup_{g=1}^{G}\{\ Z_{(g)}^*,\ \Delta y_{(g)}^*, \Delta u_{(g)}^*(\theta), \Delta w_{(g)}^*\}$ be a cluster-wise bootstrap sample. Given the b-th resampled data $\cup_{g=1}^{G}\{\ Z_{(g)}^*,\ \Delta y_{(g)}^*, \Delta u_{(g)}^*(\theta), \Delta w_{(g)}^*\}$, we can implement the GMM bootstrap procedure in Hall and Horowitz (1996) as follows.

Step 1: With the recentered moment function $(Z_i^*)'\Delta u_i^*(\theta) - E^*[(Z_i^*)'\Delta u_i^*(\hat{\theta}_1)]$, obtain a bootstrap version of the initial estimator:

$$\hat{\theta}_{1,(b)}^* = \left[(\Delta w^*)'\, Z^*\, [W_n^*]^{-1}\, (Z^*)'\Delta w^*\right]^{-1} (\Delta w^*)'\, Z^*\, [W_n^*]^{-1} \left((Z^*)'\Delta y^* - Z'\Delta u(\hat{\theta}_1)\right).$$

, where $W_n^* = n^{-1}\sum_{i=1}^{n}(Z_i^*)'Z_i^*$.

Step 2: Obtain a bootstrap version of the two-step GMM weighting matrix $\hat{\Omega}^{c*}(\hat{\theta}_{1,(b)}^*)$ with the recentered moment process $(Z_i^*)'\Delta u_i^*(\theta) - E^*[(Z_i^*)'\Delta u_i^*(\hat{\theta}_2)]$:

$$\hat{\Omega}^{c*}(\hat{\theta}_{1,(b)}^*;\hat{\theta}_2) = \frac{1}{G}\sum_{g=1}^{G}\left[\left(\frac{(Z_{(g)}^*)'\Delta u_{(g)}^*(\hat{\theta}_{1,(b)}^*)}{\sqrt{\bar{L}}} - \frac{1}{G}\sum_{h=1}^{G}\frac{Z_{(h)}'\Delta u_{(h)}(\hat{\theta}_2)}{\sqrt{\bar{L}}}\right)\times \left(\frac{(Z_{(g)}^*)'\Delta u_{(g)}^*(\hat{\theta}_{1,(b)}^*)}{\sqrt{\bar{L}}} - \frac{1}{G}\sum_{h=1}^{G}\frac{Z_{(h)}'\Delta u_{(h)}(\hat{\theta}_2)}{\sqrt{\bar{L}}}\right)'\right].$$

, and corresponding two-step GMM estimator:

$$\hat{\theta}_{2,(b)}^* = \left[(\Delta w^*)'\, Z^*\left[\hat{\Omega}^{c*}(\hat{\theta}_{1,(b)}^*)\right]^{-1}(Z^*)'\Delta w^*\right]^{-1}(\Delta w^*)'\, Z^*\left[\hat{\Omega}^{c*}(\hat{\theta}_{1,(b)}^*)\right]^{-1}\times \left[(Z^*)'\Delta y^* - Z'\Delta u(\hat{\theta}_1)\right].$$

Step 3: Construct the b-th bootstrap version of t statistic:

$$t^*(\hat{\theta}_{2,(b)}^*) = \frac{R(\hat{\theta}_{2,(b)}^* - \hat{\theta}_2)}{\sqrt{R\widehat{var}_{\hat{\Omega}^{c*}(\hat{\theta}_{1,(b)}^*;\hat{\theta}_2)}(\hat{\theta}_{2,(b)}^*)R'}},$$

where $\widehat{var}_{\hat{\Omega}^{c*}(\hat{\theta}_{1,(b)}^*)}(\hat{\theta}_{2,(b)}^*) = n\left((\Delta w^*)'\, Z^*\left[\hat{\Omega}^{c*}(\hat{\theta}_{1,(b)}^*;\hat{\theta}_2)\right]^{-1}(Z^*)'\Delta w^*\right)^{-1}$, and J statistic:

$$J^*(\hat{\theta}_{2,(b)}^*;\hat{\theta}_2) = \left[\frac{Z^{*\prime}\Delta u^*(\hat{\theta}_{2,(b)}^*) - Z'\Delta u(\hat{\theta}_2)}{\sqrt{n}}\right]'\left[\hat{\Omega}^{c*}(\hat{\theta}_{2,(b)}^*;\hat{\theta}_2)\right]^{-1}\times \left[\frac{Z^{*\prime}\Delta u^*(\hat{\theta}_{2,(b)}^*) - Z'\Delta u(\hat{\theta}_2)}{\sqrt{n}}\right].$$

Step 4: After repeating 1∼3 steps B-times, compute the two-side bootstrap p-values for t and J statistics with

$$\hat{p}_{t,\text{HH}} = \frac{1}{B}\sum_{b=1}^{B} 1\left(\left|t^*(\hat{\theta}^*_{2,(b)})\right| > \left|t(\hat{\theta}_2)\right|\right) \text{ and } \hat{p}_{J,\text{HH}} = \frac{1}{B}\sum_{b=1}^{B} 1\left(J^*(\hat{\theta}^*_{2,(b)}) > J(\hat{\theta}_2)\right).$$

Then, we reject Reject the null hypothesis $H_0 : R\theta = r$ iff

$$\hat{p}_{t,HH} \leqq \alpha$$

, and reject the null hypothesis of J test $H_0 : E\left(Z_i'\Delta u_i\right) = 0$ iff

$$\hat{p}_{J,HH} \leqq \alpha.$$

B.9 Testing the level of clustering

The Ibragimov and Muller (2016, IM hereafter)'s test considers a scenario that empirical researchers face a choice between a small number of coarse clusters and a large number of the finer level of clusters. The null hypothesis is that a finer level of clustering is appropriate with a consistent CCE estimator, against the alternative the only fewer clusters provide valid information. As a general motivation, consider a linear regression

$$y_i^g = (X_i^g)'\theta + \epsilon_i^g,$$

where y_i^g and X_i^g are the i-th of L_g observations from cluster $g = 1, \ldots, G$. Suppose a researcher is interested in the j-th element of $\theta \in \mathbb{R}^d$, $\beta_j = e_j'\theta$, where e_j is a j-th standard basis vector in $\mathbb{R}^d$. We first partition sample into G clusters, and estimate the model only using the L_g subsamples in cluster g and obtain $\hat{\beta}_j^g$ for each $g = 1, \ldots G$, and compute the following statistics S^2:

$$S^2 = \frac{1}{G-1}\sum_{g=1}^{G}(\hat{\beta}_j^g - \bar{\hat{\beta}})^2 \text{ where } \bar{\hat{\beta}} = \frac{1}{G}\sum_{g=1}^{G}\hat{\beta}_j^g. \tag{B.54}$$

In the estimation of $\hat{\beta}_j^g$, one also calculates its cluster-robust standard errors $\hat{\sigma}_g$ assuming the finer level of clustering within each group g is appropriate. To obtain the critical value of S^2, we draw $Y_g \overset{\text{i.i.d}}{\sim} N(0, \hat{\sigma}_g^2)$ for $g = 1, \ldots, G$, and compute

$$S_Y^2 = \frac{1}{G-1}\sum_{g=1}^{G}(Y_g - \bar{Y})^2 \text{ where } \bar{Y} = \frac{1}{G}\sum_{g=1}^{G}Y_g. \tag{B.55}$$

When the test statistics S^2 is larger than 95% empirical quantile of the 10,000 draws of S_Y^2, the test rejects the null hypothesis of validity of finer level of clustering. IM (2016) shows that the test is asymptotically size corrected, and has a non-trivial asymptotic power when the fine level of clustering ignores correlation among the observations in the coarser cluster.

In our empirical example in Section 7, we consider a finer clustering at the level of individuals. To apply the IM's test, we first need to estimate the key parameter of interests, β_d, β_s, and β_{ds} including all nuisance parameters at each county level cluster. However, we note that the subsamples at each county in Emran and Hou (2013)'s data set are insufficient to estimate the full regression equation in Section 7. This is mostly coming from having control variables that do not have sufficient within-cluster (within-county) variations. Thus, as the next best thing, we could estimate the cluster-specific OLS estimators, $\hat{\beta}_d^g$, $\hat{\beta}_s^g$, and $\hat{\beta}_{ds}^g$, for each county cluster g by only considering the key variables of interests in the regression. We also compute the corresponding robust standard errors, $\hat{\sigma}_d^g$, $\hat{\sigma}_s^g$, and $\hat{\sigma}_{ds}^g$ at the level of individual clustering. Based on these estimates, we construct the test statistics and corresponding critical values as in (B.54)–(B.55) for each parameter of interests. The corresponding p-values of tests are reported in Table 8 in the main body of the paper.

References

[1] Anderson, T. W. and Hsiao, C. (1981): "Estimation of dynamic models with error components." *Journal of the American statistical Association*, 76(375), 598-606.

[2] Arellano, M. and Bond, S. (1991): "Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations." *The Review of Economic Studies*, 58(2), 277-297.

[3] Bester, C.A., Conley, T.G. and Hansen, C.B., 2011. "Inference with dependent data using cluster covariance estimators." *Journal of Econometrics*, 165(2), pp.137-151.

[4] Bolthausen, E., 1982. "On the central limit theorem for stationary mixing random fields." *The Annals of Probability*, pp.1047-1050.

[5] Conley, T.G., 1999. "GMM estimation with cross sectional dependence." *Journal of econometrics*, 92(1), pp.1-45.

[6] Dedecker, J., 1998. "A central limit theorem for stationary random fields." *Probability Theory and Related Fields*, 110(3), pp.397-426.

[7] Emran, M. S., and Hou, Z. (2013): "Access to markets and rural poverty: evidence from household consumption in China." *Review of Economics and Statistics*, 95(2), 682-697.

[8] Hansen, L.P., Heaton, J. and Yaron, A., (1996). "Finite-sample properties of some alternative GMM estimators." *Journal of Business & Economic Statistics*, 14(3), pp.262-280.

[9] Hansen, B.E. and Lee, S., (2018). "Inference for iterated GMM under misspecification and clustering." UNSW Business School Research Paper, (2018-07).

[10] Hall, P., and Horowitz, J. L. (1996): "Bootstrap critical values for tests based on generalized-method-of-moments estimators." *Econometrica: Journal of the Econometric Society*, 891-916.

[11] Hall, P.G. and Hyde, C.C., "Martingale Central Limit Theory and its Applications." 1980.

[12] Hwang, J., and Sun, Y. (2017a): "Asymptotic F and t Tests in an Efficient GMM Setting." *Journal of Econometrics*, 198(2), 277-295.

[13] Jenish, N. and Prucha, I.R., 2009. "Central limit theorems and uniform laws of large numbers for arrays of random fields." *Journal of econometrics*, 150(1), pp.86-98.

[14] Jiang, J., Luan, Y., and Wang, Y. G. (2007): "Iterative estimating equations: Linear convergence and asymptotic properties." *The Annals of Statistics*, 35(5), 2233-2260.

[15] Liang, K.-Y. and Zeger, S. (1986): "Longitudinal Data Analysis Using Generalized Linear Models." *Biometrika* 73(1):13-22.

[16] Nickell, S. (1981): "Biases in dynamic models with fixed effects." *Econometrica*: Journal of the Econometric Society, 1417-1426.

[17] Sun, Y. (2014): "Fixed-smoothing Asymptotics in a Two-step GMM Framework." *Econometrica* 82(6), 2327-2370.

[18] Zhang, X. (2016): "Fixed-smoothing asymptotics in the generalized empirical likelihood estimation framework." *Journal of Econometrics*, 193(1), 123-146.