Simple and Trustworthy Cluster-Robust GMM Inference

Jungbin Hwang*

Department of Economics,
University of Connecticut

August 30, 2017

Abstract

This paper develops a new asymptotic theory for two-step GMM estimation and inference in the presence of clustered dependence. The key feature of alternative asymptotics is the number of clusters G is regarded as small or fixed when the sample size increases. Under the small-G asymptotics, this paper shows the centered two-step GMM estimator and the two continuously-updating GMM estimators we consider have the same asymptotic mixed normal distribution. In addition, the J statistic, the trinity of two-step GMM statistics (QLR, LM and Wald), and the t statistic are all asymptotically pivotal, and each can be modified to have an asymptotic standard F distribution or t distribution. We suggest a finite sample variance correction to further improve the accuracy of the F and t approximations. Our proposed asymptotic F and t tests are very appealing to practitioners because our test statistics are simple modifications of the usual test statistics, and critical values are readily available from standard statistical tables. A Monte Carlo study shows that our proposed tests are more accurate than the conventional inferences under the large-G asymptotics.

JEL Classification: C12, C21, C23, C31
Keywords: Two-step GMM, Heteroskedasticity and Autocorrelation Robust, Clustered Dependence, t distribution, F distribution

*Email: jungbin.hwang@uconn.edu. Correspondence to: Jungbin Hwang, Department of Economics, University of Connecticut, 365 Fairfield Way U-1063, Storrs, CT 06269-1063. I am indebt to Yixiao Sun, Graham Elliott, Andres Santos, and Minseong Kim that helped greatly improve this paper. I would also like to thank seminar partcipants at UCSD, UCONN, UCI, Boston University, 2016 North Ameriacan Meeting of Econometric Society, and 2017 New York Camp Econometrics XII. I acknowledge partial research support from NSF under Grant No. SES-1530592.

1 Introduction

Clustering is a common feature for many cross-sectional and panel data sets in applied economics. The data often come from a number of independent clusters with a general dependence structure within each cluster. For example, in development economics, data are often clustered by geographical regions, such as village, county and province, and, in empirical finance and industrial organization, firm level data are often clustered at the industry level. Because of learning from daily interactions, the presence of common shocks, and for many other reasons, individuals in the same cluster will be interdependent while those from different clusters tend to be independent. Failure to control for within group or cluster correlation often leads to downwardly biased standard errors and spurious statistical significance.

Seeking to robustify inference, many practical methods employ clustered covariance estimators (CCE). See White (1984, Theorem 6.3, p. 136), Liang and Zeger (1986), Arellano (1987) for seminal methodological contributions, and Wooldridge (2003) and Cameron and Miller (2015) for overviews of the CCE and its applications. It is now well known that standard test statistics based on the CCE are either asymptotically chi-squared or normal. The chi-squared and normal approximations are obtained under the so-called large-G asymptotic specification, which requires the number of clusters G to grow with the sample size. The key ingredient behind these approximations is that the CCE becomes concentrated at the true asymptotic variance as G approaches to infinity. In effect, this type of asymptotics ignores the estimation uncertainty in the CCE despite its high variation in finite samples, especially when the number of clusters is small. In practice, however, it is not unusual to have a data set that has a small number of clusters. For example, if clustering is based on large geographical regions such as U.S. states and regional blocks of neighboring countries, (e.g., Bertrand, Duflo, and Mullainathan, 2004; Ibragimov and Müller, 2015), we cannot convincingly claim that the number of cluster is large so that the large-G asymptotic approximations are applicable. In fact, there is ample simulation evidence that the large-G approximation can be very poor when the number of clusters is not large (e.g., Donald and Lang, 2007; Cameron, Gelbach, and Miller, 2008; Bester, Conley, and Hansen, 2011; Mackinnon and Webb, 2017).

In this paper, we adopt an alternative approach that yields more accurate approximations, and that works well whether or not the number of clusters is large. Our approximations work especially well when the chi-squared and normal approximations are poor. They are obtained from an alternative limiting thought experiment where the number of clusters G is held fixed. Under this small (fixed)-G asymptotics, the CCE no longer asymptotically degenerates; instead, it converges in distribution to a random matrix that is proportional to the true asymptotic variance. The random limit of the CCE has profound implications for the analyses of the asymptotic properties of GMM estimators and the corresponding test statistics.

We start with the first-step GMM estimator where the underlying model is possibly over-identified and show that suitably modified Wald and t statistics converge weakly to standard F and t distributions, respectively. The modification is easy to implement because it involves only a known multiplicative factor. Similar results have been obtained by Hansen (2007) and Bester, Conley and Hansen (2011), which employ a CCE type HAC estimator but consider only linear regressions and M-estimators for an exactly identified model.

We then consider the two-step GMM estimator that uses the CCE as a weighting matrix. Because the weighting matrix is random even in the limit, the two-step estimator is not asymptotically normal. The form of the limiting distribution depends on how the CCE is constructed. If the CCE is based on the uncentered moment process, we obtain the so-called uncentered two-

step GMM estimator. We show that the asymptotic distribution of this two-step GMM estimator is highly nonstandard. As a result, the associated Wald and t statistics are not asymptotically pivotal. However, it is surprising that the J statistic is still asymptotically pivotal and has a Beta limiting distribution, and the critical values are readily available from standard statistical tables and canned software packages.

Next, we establish the asymptotic properties of the "centered" two-step GMM estimator[1] whose weighting matrix is constructed using recentered moment conditions. Invoking centering is not innocuous for an over-identified GMM model because the empirical moment conditions, in this case, are not equal to zero in general. Under the traditional large-G asymptotics, recentering does not matter in large samples because the empirical moment conditions are asymptotically zero and here are ignorable, even though they are not identically zero in finite sample. In contrast, under the small-G asymptotics, recentering plays two important roles: it removes the first order effect of the estimation error in the first-step estimator, and it ensures that the weighting matrix is asymptotically independent of the empirical moment conditions. With the recentered CCE as the weighting matrix, the two-step GMM estimator is asymptotically mixed normal. The mixed normality reflects the high variation of the feasible two-step GMM estimator as compared to the infeasible two-step GMM estimator, which is obtained under the assumption that the 'efficient' weighing matrix is known. The mixed-normality allows us to construct the Wald and t statistics that are asymptotically nuisance parameter free.

To relate the nonstandard small-G asymptotic distributions to standard distributions, we introduce simple modifications to the Wald and t statistics associated with the centered two-step GMM estimator. We show that the modified Wald and t statistics are asymptotically F and t distributed, respectively. This result resembles the corresponding result that is based on the first-step GMM estimator. It is important to point out that the proposed modifications are indispensable for our asymptotic F and t theory. In the absence of the modifications, the Wald and t statistics converge in distribution to nonstandard distributions, and as a result, critical values have to be simulated. The modifications involve only the standard J statistic, and it is very easy to implement because the modified test statistics are scaled versions of the original Wald test statistics with the scaling factor depending on the J statistic. Significantly, the combination of the Wald statistic and the J statistic enables us to develop the F approximation theory.

We also consider two types of continuous updating (CU) estimators. The first type continuously updates the first order conditions (FOC) underlying the two-step GMM estimator. Given that FOC can be regarded as the empirical version of generalized estimating equations (GEE) which is first studied by Liang and Zeger (1986), we call this type of CU estimator the CU-GEE estimator. The second type continuously updates the GMM criterion function, leading to the CU-GMM estimator, which was first suggested by Hansen, Heaton and Yaron (1996). Both CU estimators are designed to improve the finite sample performance of two-step GMM estimators. Interestingly, we show that the continuous updating scheme has a built-in recentering feature. Thus, in terms of the first order asymptotics, it does not matter whether the empirical moment conditions are recentered or not. We find that the centered two-step GMM estimator and the two CU estimators are all first-order asymptotically equivalent under the small-G asymptotics. This result provides a theoretical justification for using the recentered CCE in a two-step GMM framework.

Finally, although the recentering scheme removes the first order effect of the first-step esti-

[1] Our definition of the centered two-step GMM estimator is originated from the recentered (or demeaned) GMM weighting matrix, and it should not be confused with "centering" the estimator itself.

mation error, the centered two-step GMM and CU estimators still face some extra estimation uncertainty in the first-step estimator. The main source of the problem is that we have to estimate the unobserved moment process based on the first-step estimator. To capture the higher order effect, we propose to retain one more term in our stochastic approximation that is asymptotically negligible. The expansion helps us develop a finite sample correction to the asymptotic variance estimator. Our correction resembles that of Windmeijer (2005), which considers variance correction for a two-step GMM estimator but valid only in an i.i.d. setting. We show that the finite sample variance correction does not change the small-G limiting distributions of the test statistics, but they can help improve the finite sample performance of our tests.

Monte Carlo simulations show that our new tests have a much more accurate size than existing tests via standard normal and chi-squared critical values, especially when the number of clusters G is not large. An advantage of our procedure is that the test statistics do not entail much extra computational cost because the main ingredient for the modification is the usual J statistic. There is also no need to simulate critical values because the F and t critical values can be readily obtained from standard statistical tables.

Our small-G asymptotics is related to fixed-smoothing asymptotics for a long run variance (LRV) estimation in a time series setting. The latter was initiated and developed in econometric literature by Kiefer, Vogelsang and Bunzel (2002), Kiefer and Vogelsang (2005), Müller (2007), Sun, Phillips and Jin (2008), Sun (2013, 2014), Zhang (2016), among others. Our new asymptotics is in the same spirit in that both lines of research attempt to capture the estimation uncertainty in covariance estimation. With regards to orthonormal series LRV estimation, a recent paper by Hwang and Sun (2017a) modifies the two-step GMM statistics using the J statistic, and shows that the modified statistics are asymptotically F and t distributed. The F and t limit theory presented in this paper is similar, but our cluster-robust limiting distributions differ from those of our predecessors in terms of the multiplicative adjustment and the degrees of freedom. Moreover, we propose a finite sample variance correction to capture the uncertainty embodied in the estimated moment process adequately. To our knowledge, the finite sample variance correction provided in this paper ant its first order asymptotic validity has not been considered in the literature on the fixed-smoothing asymptotics.

There is also a growing literature that uses the small-G asymptotics to design more accurate cluster-robust inference. For instance, Ibragimov and Müller (2010, 2016) recently proposes a subsample based t test for a scalar parameter that is robust to heterogeneous clusters. Hansen (2007), Stock and Watson (2008), and Bester, Conley and Hansen (2011) propose a cluster-robust F or t tests under cluster-size homogeneity. Imbens and Kolesár (2016) suggest an adjusted t critical value employing data-determined degrees of freedom. Recently, Canay, Romano and Shaikh (2017) establishes a theory of randomization tests and suggests an alternative cluster-robust test. For other approaches, see Carter, Schnepel and Steigerwald (2017) which proposes a measure of the effective number of clusters under the large-G asymptotics; Cameron, Gelbach and Miller (2008) and MacKinnon and Webb (2017) which investigate cluster-robust bootstrap approaches. All these studies, however, mainly focus on a simple location model or linear regressions that are special cases of exactly identified models.

The remainder of the paper is organized as follows. Section 2 presents the basic setting and establishes the approximation results for the first-step GMM estimator under the small-G asymptotics. Sections 3 and 4 establish the small-G asymptotics for two-step GMM estimators and develop the asymptotic F and t tests based on the centered two-step GMM estimator. Section 5 extends the first-order small-G asymptotics to the CU-type GMM estimators. Section 6

proposes a finite sample variance correction. The next section reports a simulation evidence. The last section concludes. Proofs are given in the appendix, and an online supplemental appendix available at the author's website[2] contains practical implementations of the GMM procedures considered in this paper in the context of dynamic panel model and applies them to an empirical study in Emran and Hou (2013).

2 Basic Setting and the First-step GMM Estimator

We want to estimate the $d \times 1$ vector of parameters $\theta \in \Theta$. The true parameter vector θ_0 is assumed to be an interior point of parameter space $\Theta \subseteq \mathbb{R}^d$. The moment condition

$$Ef(Y_i, \theta) = 0 \text{ holds if and only if } \theta = \theta_0, \tag{1}$$

where $f_i(\theta) = f(Y_i, \theta)$ is an $m \times 1$ vector of twice continuously differentiable functions. We assume that $q = m - d \geq 0$ and the rank of $\Gamma = E\left[\partial f(Y_i, \theta_0)/\partial \theta'\right]$ is d. So the model is possibly over-identified with the degree of over-identification q. The number of observations is n.

Define $g_n(\theta) = n^{-1} \sum_{i=1}^n f_i(\theta)$. Given the moment condition in (1), the initial "first-step" GMM estimator of θ_0 is given by

$$\hat{\theta}_1 = \arg\min_{\theta \in \Theta} g_n(\theta)' W_n^{-1} g_n(\theta),$$

where W_n is an $m \times m$ positive definite and a symmetric weighting matrix that does not depend on the unknown parameter θ_0 and $\text{plim}_{n\to\infty} W_n = W > 0$. In the context of instrumental variable (IV) regression, one popular choice for W_n is $(Z_n' Z_n/n)^{-1}$ where Z_n is the data matrix of instruments.

Let $\hat{\Gamma}(\theta) = n^{-1} \sum_{i=1}^n \frac{\partial f_i(\theta)}{\partial \theta'}$. To establish the asymptotic properties of $\hat{\theta}_1$, we assume that for any $\sqrt{n}$-consistent estimator $\tilde{\theta}$, $\text{plim}_{n\to\infty} \hat{\Gamma}(\tilde{\theta}) = \Gamma$ and that Γ is of full column rank. Also, under some regularity conditions, we have the following Central Limit Theorem (CLT)

$$\sqrt{n} g_n(\theta_0) \xrightarrow{d} N(0, \Omega), \text{ where}$$
$$\Omega = \lim_{n\to\infty} \frac{1}{n} E \left(\sum_{i=1}^n f_i(\theta_0) \right) \left(\sum_{i=1}^n f_i(\theta_0) \right)'.$$

Here Ω is analogous to the long run variance in a time series setting but the components of Ω are contributed by cross-sectional dependences over all locations. For easy reference, we follow Sun and Kim (2015) and call Ω the global variance. Primitive conditions for the above CLT in the presence of cross-sectional dependence are provided in Jenish and Prucha (2009, 2012). Under these conditions, we have

$$\sqrt{n}(\hat{\theta}_1 - \theta_0) \xrightarrow{d} N\left[0, (\Gamma' W^{-1} \Gamma)^{-1} \Gamma' W^{-1} \Omega W^{-1} \Gamma (\Gamma' W^{-1} \Gamma)^{-1}\right].$$

Since Γ and W can be accurately estimated by $\hat{\Gamma}(\hat{\theta}_1)$ and W_n relative to Ω, we only need to estimate Ω to make reliable inference about θ_0. The main issue is how to properly account for cross-sectional dependence in the moment process $\{f_j(\theta_0)\}_{j=1}^n$. In this paper, we assume that the cross-sectional dependence has a cluster structure, which is popular in many microeconomic

[2] http://hwang.econ.uconn.edu/research/

applications. More specifically, our data consists of a number of independent clusters, each of which has an unknown dependence structure. Let G be the total number of clusters and L_g be the size of cluster g. For simplicity, we assume that every cluster has the common size L, i.e., $L = L_1 = L_2 = ... = L_G$. The identical cluster size assumption can be relaxed to the assumption that each cluster has approximately same size relative to the average cluster size, i.e., $\lim_{n\to\infty} L_g/(G^{-1}\sum_{g=1}^{G} L_g) = 1$ for every $g = 1, ..., G$. Equivalently, we can express this approximately equal cluster size assumption by $L = L_g + o(L)$ for each $g = 1, ..., G$. The following assumption formally characterizes the cluster dependence.

Assumption 1 (*i*) *The data* $\{Y_i\}_{i=1}^n$ *consists of* G *clusters.* (*ii*) *Observations are independent across clusters.* (*iii*) *The number of clusters* G *is fixed, and the size of each cluster* L *grows with the total sample size* n.

Assumption 1-i) implies that the set $\{f_i(\theta_0), i = 1, 2, ..., n\}$ can be partitioned into G nonoverlapping clusters $\cup_{g=1}^{G}\mathcal{G}_g$ where $\mathcal{G}_g = \{f_k^g(\theta_0) : k = 1, ..., L\}$. In the context of the clustered structure, Assumption 1-ii) implies that the within-cluster dependence for each cluster can be arbitrary but $Ef_k^g(\theta_0)f_l^h(\theta_0) = 0$ if $g \neq h$ for any $k, l = 1, ..., L$. That is, $f_k^g(\theta_0)$ and $f_l^h(\theta_0)$ are independent if they belong to different clusters. The independence across clusters in Assumption 1-ii) can be generalized to allow weak dependence among clusters by restricting the number of observations located on the boundaries between clusters. See Bester, Conley and Hansen (2011, BCH hereafter) for the detailed primitive conditions. Under Assumption 1-ii), we have

$$\begin{aligned}\Omega &= \lim_{n\to\infty}\frac{1}{n}E\left(\sum_{i=1}^{n} f_i(\theta_0)\right)\left(\sum_{i=1}^{n} f_i(\theta_0)\right)' \\ &= \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} 1(i,\ j\ \in \text{same cluster})Ef_i(\theta_0)f_j(\theta_0)'. \end{aligned} \tag{2}$$

Assumption 1-iii) specifies the direction of asymptotics we consider. Under this small-G asymptotic specification, we have

$$\Omega = \frac{1}{G}\sum_{g=1}^{G}\lim_{L\to\infty} var\left(\frac{1}{\sqrt{L}}\sum_{k=1}^{L} f_k^g(\theta_0)\right) := \frac{1}{G}\sum_{g=1}^{G}\Omega_g.$$

Thus, the global covariance matrix Ω can be represented as the simple average of Ω_g, $g = 1, ..., G$, where Ω_g's are the limiting variances within individual clusters. Motivated by this, we construct the clustered covariance estimator (CCE) as follows:

$$\begin{aligned}\hat{\Omega}(\hat{\theta}_1) &= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} 1(i, j \in \text{the same group})f_i(\hat{\theta}_1)f_j(\hat{\theta}_1)' \\ &= \frac{1}{G}\sum_{g=1}^{G}\left\{\left(\frac{1}{\sqrt{L}}\sum_{k=1}^{L} f_k^g(\hat{\theta}_1)\right)\left(\frac{1}{\sqrt{L}}\sum_{k=1}^{L} f_k^g(\hat{\theta}_1)\right)'\right\}.\end{aligned}$$

To ensure that $\hat{\Omega}(\hat{\theta}_1)$ is positive definite, we assume that $G \geq m$, and maintain this condition throughout the rest of the paper.

Suppose we want to test the null hypothesis $H_0 : R\theta_0 = r$ against the alternative $H_1 : R\theta_0 \neq r$, where R is a $p \times d$ matrix. In this paper, we focus on linear restrictions without loss of generality because the Delta method can be used to convert nonlinear restrictions into linear ones in an asymptotic sense. The F test version of the Wald test statistic is given by

$$F(\hat{\theta}_1) := \frac{1}{p}(R\hat{\theta}_1 - r)' \left\{ R\widehat{var}(\hat{\theta}_1)R' \right\}^{-1} (R\hat{\theta}_1 - r), \tag{3}$$

where

$$\widehat{var}(\hat{\theta}_1) = \frac{1}{n}\left[\hat{\Gamma}(\hat{\theta}_1)'W_n^{-1}\hat{\Gamma}(\hat{\theta}_1)\right]^{-1}\left[\hat{\Gamma}(\hat{\theta}_1)'W_n^{-1}\hat{\Omega}(\hat{\theta}_1)W_n^{-1}\hat{\Gamma}(\hat{\theta}_1)\right]\left[\hat{\Gamma}(\hat{\theta}_1)'W_n^{-1}\hat{\Gamma}(\hat{\theta}_1)\right]^{-1}.$$

In constructing F statistic in (3), it is not necessary to divide it by the number of hypothesis p to develop the small-G asymptotic theory in this paper. We use it only because we anticipate more convenient F approximation once the conventional F statistic without the division factor has been divided by p. We will apply the same division rules to the two-step GMM statistics in Sections 3 and 4 to develop F limit theory.

When $p = 1$ and the alternative is one sided, we can construct the t statistic:

$$t(\hat{\theta}_1) := \frac{R\hat{\theta}_1 - r}{\sqrt{R\widehat{var}(\hat{\theta}_1)R'}}.$$

To formally characterize the asymptotic distributions of $F(\hat{\theta}_1)$ and $t(\hat{\theta}_1)$ under the small-G asymptotics, we further maintain the following high level conditions.

Assumption 2 $\hat{\theta}_1 \xrightarrow{p} \theta_0$.

Assumption 3 *(i) For each* $g = 1, ..., G$, *let* $\Gamma_g(\theta) := \lim_{L\to\infty} E\left[\frac{1}{L}\sum_{k=1}^{L}\frac{\partial f_k^g(\theta)}{\partial\theta'}\right]$. *Then,*

$$\sup_{\theta\in\mathcal{N}(\theta_0)}\left\|\frac{1}{L}\sum_{k=1}^{L}\frac{\partial f_k^g(\theta)}{\partial\theta'} - \Gamma_g(\theta)\right\| \xrightarrow{p} 0,$$

holds, where $\mathcal{N}(\theta_0)$ *is an open neighborhood of* θ_0 *and* $\|\cdot\|$ *is the Euclidean norm. (ii)* $\Gamma_g(\theta)$ *is continuous at* $\theta = \theta_0$, *and for* $\Gamma_g = \Gamma_g(\theta_0)$, $\Gamma = G^{-1}\sum_{g=1}^{G}\Gamma_g$ *has full rank.*

Assumption 4 *Let* $B_{m,g} \overset{i.i.d}{\sim} N(0, I_m)$ *for* $g = 1, ..., G$, *then*

$$P\left(\frac{1}{\sqrt{L}}\sum_{k=1}^{L} f_k^g(\theta_0) \leq x\right) = P\left(\Lambda_g B_{m,g} \leq x\ \right) + o(1) \text{ as } L \to \infty,$$

for each $g = 1, ..., G$ *where* $x \in \mathbb{R}^m$ *and* Λ_g *is the matrix square root of* Ω_g.

Assumption 5 (*Homogeneity of* Γ_g) *For all* $g = 1, ..., G$, $\Gamma_g = \Gamma$.

Assumption 6 (*Homogeneity of* Ω_g) *For all* $g = 1, ..., G$, $\Omega_g = \Omega$.

Assumption 2 is made for convenience, and primitive sufficient conditions are available from the standard GMM asymptotic theory. Assumption 3 is a uniform law of large numbers (ULLN), from which we obtain $\hat{\Gamma}(\hat{\theta}_1) = G^{-1}\sum_{g=1}^{G}\Gamma_g + o_p(1) = \Gamma + o_p(1)$. Together with Assumption 1-(ii), Assumption 4 implies that $L^{-1/2}\sum_{k=1}^{L} f_k^g(\theta_0)$ follows a central limit theorem jointly over $g = 1, ..., G$ with zero asymptotic covariance between any two clusters. The homogeneity conditions in Assumptions 5 and 6 guarantee the asymptotic pivotality of the cluster-robust GMM statistics we consider. Similar assumptions are made in BCH (2011) and Sun and Kim (2015), which develop asymptotically valid F tests that are robust to spatial autocorrelation in the same spirit as our small-G asymptotics. Let

$$\bar{B}_m := G^{-1}\sum_{g=1}^{G} B_{m,g} \text{ and } \bar{\mathbb{S}} := G^{-1}\sum_{g=1}^{G}\left(B_{m,g} - \bar{B}_m\right)\left(B_{m,g} - \bar{B}_m\right)',$$

where $B_{m,g}$ as in Assumption 4. Also, let $\mathbb{W}_p(K, \Pi)$ denote a Wishart distribution with K degrees of freedom and $p \times p$ positive definite scale matrix Π. By construction, $\sqrt{G}\bar{B}_m \sim N(0, I_m)$, $\bar{\mathbb{S}} \sim G^{-1}\mathbb{W}_p(G-1, I_m)$, and $\bar{B}_m \perp \bar{\mathbb{S}}$. To present our asymptotic results, we partition $\bar{B}_m$ and $\bar{\mathbb{S}}$ as follows:

$$\begin{aligned}
\bar{B}_m &= \begin{pmatrix} \bar{B}_d \\ {\scriptstyle d\times 1} \\ \bar{B}_q \\ {\scriptstyle q\times 1} \end{pmatrix},\ \bar{B}_d = \begin{pmatrix} \bar{B}_p \\ {\scriptstyle p\times 1} \\ \bar{B}_{d-p} \\ {\scriptstyle (d-p)\times 1} \end{pmatrix},\ \bar{\mathbb{S}} = \begin{pmatrix} \bar{\mathbb{S}}_{dd} & \bar{\mathbb{S}}_{dq} \\ {\scriptstyle d\times d} & {\scriptstyle d\times q} \\ \bar{\mathbb{S}}_{qd} & \bar{\mathbb{S}}_{qq} \\ {\scriptstyle q\times d} & {\scriptstyle q\times q} \end{pmatrix}, \\
\bar{\mathbb{S}}_{dd} &= \begin{pmatrix} \bar{\mathbb{S}}_{pp} & \bar{\mathbb{S}}_{p,d-p} \\ {\scriptstyle p\times p} & {\scriptstyle p\times(d-p)} \\ \bar{\mathbb{S}}_{d-p,p} & \bar{\mathbb{S}}_{d-p,d-p} \\ {\scriptstyle (d-p)\times p} & {\scriptstyle (d-p)\times(d-p)} \end{pmatrix}, \text{ and } \bar{\mathbb{S}}_{dq} = \begin{pmatrix} \bar{\mathbb{S}}_{pq} \\ {\scriptstyle p\times q} \\ \bar{\mathbb{S}}_{d-p,q} \\ {\scriptstyle (d-p)\times q} \end{pmatrix}.
\end{aligned}$$

Proposition 1 *Let Assumptions 1∼6 hold. Then,*

(a) $F(\hat{\theta}_1) \xrightarrow{d} \mathbb{F}_{1\infty} := \left(\frac{G}{p}\right)\cdot \bar{B}_p'\bar{\mathbb{S}}_{pp}^{-1}\bar{B}_p$;

(b) $t(\hat{\theta}_1) \xrightarrow{d} \mathbb{T}_{1\infty} := \frac{N(0,1)}{\sqrt{\chi^2_{G-1}/G}}$ *where* $N(0,1) \perp \chi^2_{G-1}$.

Remark 2 *The limiting distribution $\mathbb{F}_{1\infty}$ follows Hotelling's T^2 distribution. Using the well-known relationship between the T^2 and standard F distributions, we obtain $\mathbb{F}_{1\infty} \overset{d}{=} (G/G-p)\mathcal{F}_{p,G-p}$ where $\mathcal{F}_{p,G-p}$ is a random variable that follows the F distribution with degree of freedom $(p, G-p)$. Similarly, $\mathbb{T}_{1\infty} \overset{d}{=} (G/G-1)t_{G-1}$ where t_{G-1} is a random variable that follows the t distribution with degree of freedom $G-1$.*

Remark 3 *As an example of the general GMM setting, consider the linear regression model $y_i = x_i'\theta + \epsilon_i$. Under the assumption that $E[x_i\epsilon_i] = 0$, the moment function is $f_i(\theta) = x_i(y_i - x_i'\theta)$. With the moment condition $Ef_i(\theta_0) = 0$, the model is exactly identified. This set up was employed in Hansen (2007), Stock and Watson (2008), and BCH (2011); indeed, our F and t approximations in Proposition 1 are identical to what is obtained in these papers.*

Remark 4 *Under the large-G asymptotics where $G \to \infty$ but L is fixed, one can show that the CCE $\hat{\Omega}(\hat{\theta}_1)$ converges in probability to Ω for*

$$\Omega = \lim_{G\to\infty}\frac{1}{G}\sum_{g=1}^{G} var\left(\frac{1}{\sqrt{L}}\sum_{k=1}^{L} f_k^g(\theta_0)\right).$$

The convergence of $\hat{\Omega}(\hat{\theta}_1)$ to Ω does not require the homogeneity of Ω_g in Assumption 6 (Hansen, 2007; Carter et al., 2017). Under this type of asymptotics, the test statistics $F(\hat{\theta}_1)$ and $t(\hat{\theta}_1)$ are asymptotically χ^2_p/p and $N(0,1)$. Let $\mathcal{F}^{1-\alpha}_{p,G-p}$ and $\chi^{1-\alpha}_p$ be the $1-\alpha$ quantiles of the $\mathcal{F}_{p,G-p}$ and χ^2_p distributions, respectively. As $G/(G-p)>1$ and $\mathcal{F}^{1-\alpha}_{p,G-p} > \chi^{1-\alpha}_p/p$, it is easy to see that

$$\frac{G}{G-p}\mathcal{F}^{1-\alpha}_{p,G-p} > \chi^{1-\alpha}_p/p.$$

However, the difference between the two critical values $G(G-p)^{-1}\mathcal{F}^{1-\alpha}_{p,G-p}$ and $\chi^{1-\alpha}_p/p$ shrinks to zero as G increases. Therefore, the small-G critical value $G(G-p)^{-1}\mathcal{F}^{1-\alpha}_{p,G-p}$ is asymptotically valid under the large-G asymptotics. The asymptotic validity holds even if the homogeneity conditions of Assumptions 5 and 6 are not satisfied. The small-G critical value is robust in the sense that it works whether G is small or large.

Remark 5 *Let Λ the matrix square root of Ω, that is, $\Lambda\Lambda' = \Omega$. Then, it follows from the proof of Proposition 1 that $\hat{\Omega}(\hat{\theta}_1)$ converges in distribution to a random matrix $\Omega_{1\infty}$ given by*

$$\Omega_{1\infty} = \Lambda\tilde{\mathbb{D}}\Lambda' \text{ where } \tilde{\mathbb{D}} = \frac{1}{G}\sum_{g=1}^{G}\tilde{D}_g\tilde{D}_g' \text{ and}$$
$$\tilde{D}_g = B_{m,g} - \Gamma_\Lambda(\Gamma_\Lambda' W_\Lambda^{-1}\Gamma_\Lambda)^{-1}\Gamma_\Lambda' W_\Lambda^{-1}\bar{B}_m \tag{4}$$

for $\Gamma_\Lambda = \Lambda^{-1}\Gamma$ and $W_\Lambda = \Lambda^{-1}W(\Lambda')^{-1}$. $\tilde{D}_g$ is a quasi-demeaned version of $B_{m,g}$ with quasi-demeaning attributable to the estimation error in $\hat{\theta}_1$. Note that the quasi-demeaning factor $\Gamma_\Lambda(\Gamma_\Lambda' W_\Lambda^{-1}\Gamma_\Lambda)^{-1}\Gamma_\Lambda' W_\Lambda^{-1}$ depends on all of Γ, Ω and W, and cannot be further simplified in general. The estimation error in $\hat{\theta}_1$ affects $\Omega_{1\infty}$ in a complicated way. However, for the first-step Wald and t statistics, we do not care about $\hat{\Omega}(\hat{\theta}_1)$ per se. *Instead, we care about the scaled covariance matrix $\hat{\Gamma}(\hat{\theta}_1)'W_n^{-1}\hat{\Omega}(\hat{\theta}_1)W_n^{-1}\hat{\Gamma}(\hat{\theta}_1)$, which converges in distribution to $\Gamma' W^{-1}\Omega_{1\infty}W^{-1}\Gamma$. But*

$$\Gamma_\Lambda' W_\Lambda^{-1}\tilde{D}_g = \Gamma_\Lambda' W_\Lambda^{-1}\left(B_{m,g} - \bar{B}_m\right),$$

and thus

$$\Gamma' W^{-1}\Omega_{1\infty}W^{-1}\Gamma = \Gamma_\Lambda' W_\Lambda^{-1}\tilde{\mathbb{D}}W_\Lambda^{-1}\Gamma_\Lambda = \frac{1}{G}\sum_{g=1}^{G}\Gamma_\Lambda' W_\Lambda^{-1}\tilde{D}_g\left(\Gamma_\Lambda' W_\Lambda^{-1}\tilde{D}_g\right)'$$
$$\stackrel{d}{=} \Gamma_\Lambda' W_\Lambda^{-1}\frac{1}{G}\sum_{g=1}^{G}\left(B_{m,g} - \bar{B}_m\right)\left(B_{m,g} - \bar{B}_m\right)'\left(\Gamma_\Lambda' W_\Lambda^{-1}\right)'.$$

Therefore, to the first order small-G asymptotics, the estimation error in $\hat{\theta}_1$ affects $\Gamma' W^{-1}\Omega_{1\infty}W^{-1}\Gamma$ via simple demeaning only. This is a key result that drives the asymptotic pivotality of $F(\hat{\theta}_1)$ and $t(\hat{\theta}_1)$.

3 Two-step GMM Estimation and Inference

In an overidentified GMM framework, we often employ a two-step procedure to improve the efficiency of the initial GMM estimator and the power of the associated tests. It is now well-known that the optimal weighting matrix is the (inverted) asymptotic variance of the sample

moment conditions, see Hansen (1982). There are at least two different ways to estimate the asymptotic variance, and these lead to two different estimators $\hat{\Omega}(\hat{\theta}_1)$ and $\hat{\Omega}^c(\hat{\theta}_1)$, where

$$\hat{\Omega}(\hat{\theta}_1) = \frac{1}{G}\sum_{g=1}^{G}\left(\frac{1}{\sqrt{L}}\sum_{k=1}^{L} f_k^g(\hat{\theta}_1)\right)\left(\frac{1}{\sqrt{L}}\sum_{k=1}^{L} f_k^g(\hat{\theta}_1)\right)' \tag{5}$$

and

$$\hat{\Omega}^c(\hat{\theta}_1) = \frac{1}{G}\sum_{g=1}^{G}\left\{\frac{1}{\sqrt{L}}\sum_{k=1}^{L}\left[f_k^g(\hat{\theta}_1) - g_n(\hat{\theta}_1)\right]\right\}\left\{\frac{1}{\sqrt{L}}\sum_{k=1}^{L}\left[f_k^g(\hat{\theta}_1) - g_n(\hat{\theta}_1)\right]\right\}'. \tag{6}$$

While $\hat{\Omega}(\hat{\theta}_1)$ employs the uncentered moment process $\cup_{g=1}^{G}\cup_{k=1}^{L}\{f_k^g(\hat{\theta}_1)\}$, $\hat{\Omega}^c(\hat{\theta}_1)$ employs the recentered moment process $\cup_{g=1}^{G}\cup_{k=1}^{L}\{f_k^g(\hat{\theta}_1) - g_n(\hat{\theta}_1)\}$. For inference based on the first-step estimator $\hat{\theta}_1$, it does not matter which asymptotic variance estimator is used. This is so because for any asymptotic variance estimator $\hat{\Omega}(\hat{\theta}_1)$, the Wald statistic depends on $\hat{\Omega}(\hat{\theta}_1)$ only via $\hat{\Gamma}(\hat{\theta}_1)'W_n^{-1}\hat{\Omega}(\hat{\theta}_1)W_n^{-1}\hat{\Gamma}(\hat{\theta}_1)$. It is easy to show that the following asymptotic equivalence:

$$\begin{aligned}&\hat{\Gamma}(\hat{\theta}_1)'W_n^{-1}\hat{\Omega}(\hat{\theta}_1)W_n^{-1}\hat{\Gamma}(\hat{\theta}_1) = \hat{\Gamma}(\hat{\theta}_1)'W_n^{-1}\hat{\Omega}^c(\hat{\theta}_1)W_n^{-1}\hat{\Gamma}(\hat{\theta}_1) + o_p(1)\\ &= \Gamma' W_n^{-1}\hat{\Omega}^c(\theta_0)W_n^{-1}\Gamma + o_p(1).\end{aligned}$$

Thus, the limiting distribution of the Wald statistic is the same whether the estimated moment process is recentered or not. It is important to point out that the asymptotic equivalence holds because two asymptotic variance estimators are pre-multiplied by $\hat{\Gamma}(\hat{\theta}_1)'W_n^{-1}$ and post-multiplied by $W_n^{-1}\hat{\Gamma}(\hat{\theta}_1)$. In the next subsections, we will show that the two asymptotic variance estimators in (5) and (6) are not asymptotically equivalent by themselves under the small-G asymptotics.

Depending on whether we use $\hat{\Omega}(\hat{\theta}_1)$ or $\hat{\Omega}^c(\hat{\theta}_1)$, we have different two-step GMM estimators:

$$\hat{\theta}_2 = \arg\min_{\theta\in\Theta} g_n(\theta)'\left[\hat{\Omega}(\hat{\theta}_1)\right]^{-1} g_n(\theta)$$

and

$$\hat{\theta}_2^c = \arg\min_{\theta\in\Theta} g_n(\theta)'\left[\hat{\Omega}^c(\hat{\theta}_1)\right]^{-1} g_n(\theta).$$

Given that $\hat{\Omega}(\hat{\theta}_1)$ and $\hat{\Omega}^c(\hat{\theta}_1)$ are not asymptotically equivalent and that they enter the definitions of $\hat{\theta}_2$ and $\hat{\theta}_2^c$ by themselves, the two estimators have different asymptotic behaviors, as proved in the next two subsections.

3.1 Uncentered Two-step GMM estimator

In this subsection, we consider the two-step GMM estimator $\hat{\theta}_2$ based on the uncentered moment process. We establish the asymptotic properties of $\hat{\theta}_2$ and the associated Wald statistic and J statistic. We show that the J statistic is asymptotically pivotal, even though the Wald statistic is not.

It follows from standard asymptotic arguments that

$$\sqrt{n}(\hat{\theta}_2 - \theta_0) = -\left[\Gamma'\hat{\Omega}^{-1}(\hat{\theta}_1)\Gamma\right]^{-1}\Gamma'\hat{\Omega}^{-1}(\hat{\theta}_1)\frac{1}{\sqrt{G}}\sum_{g=1}^{G}\left(\frac{1}{\sqrt{L}}\sum_{j=1}^{L} f_j^g(\theta_0)\right) + o_p(1). \tag{7}$$

Using the joint convergence of the followings

$$\hat{\Omega}(\hat{\theta}_1) \xrightarrow{d} \Omega_{1\infty} = \Lambda \tilde{\mathbb{D}} \Lambda' \text{ and } \frac{1}{\sqrt{G}} \sum_{g=1}^{G} \left(\frac{1}{\sqrt{L}} \sum_{j=1}^{L} f_j^g(\theta_0) \right) \xrightarrow{d} \sqrt{G} \Lambda \bar{B}_m, \tag{8}$$

we obtain

$$\sqrt{n}(\hat{\theta}_2 - \theta_0) \xrightarrow{d} - \left[\Gamma_\Lambda' \tilde{\mathbb{D}}^{-1} \Gamma_\Lambda \right]^{-1} \Gamma_\Lambda' \tilde{\mathbb{D}}^{-1} \sqrt{G} \bar{B}_m,$$

where $\tilde{\mathbb{D}} = G^{-1} \sum_{i=1}^{G} \tilde{D}_g \tilde{D}_g'$ is defined in (4).

Since $\tilde{\mathbb{D}}$ is random, the limiting distribution is not normal. Even though both $\tilde{D}_g$ and $\bar{B}_m$ are normal, there is a nonzero correlation between them. As a result, $\tilde{\mathbb{D}}$ and $\bar{B}_m$ are correlated, too. This makes the limiting distribution of $\sqrt{n}(\hat{\theta}_2 - \theta_0)$ highly nonstandard.

To understand the limiting distribution, we define the infeasible estimator $\tilde{\theta}_2$ by assuming that $\hat{\Omega}(\theta_0)$ is known, which leads to

$$\tilde{\theta}_2 = \arg\min_{\theta \in \Theta} g_n(\theta)' \hat{\Omega}^{-1}(\theta_0) g_n(\theta).$$

Now

$$\sqrt{n}(\tilde{\theta}_2 - \theta_0) \xrightarrow{d} - \left[\Gamma_\Lambda' \mathbb{S}^{-1} \Gamma_\Lambda \right]^{-1} \Gamma_\Lambda' \mathbb{S}^{-1} \sqrt{G} \bar{B}_m,$$

where $\mathbb{S} = G^{-1} \sum_{g=1}^{G} B_{m,g} B_{m,g}'$. The only difference between the asymptotic distributions of $\sqrt{n}(\hat{\theta}_2 - \theta_0)$ and $\sqrt{n}(\tilde{\theta}_2 - \theta_0)$ is the quasi-demeaning embedded in the definition of $\tilde{D}_g$. This difference captures the first order effect of having to estimate the optimal weighting matrix, which is needed to construct the feasible two-step estimator $\hat{\theta}_2$.

To make further links between the limiting distributions, let's partition $\mathbb{S}$ in the same way that $\bar{\mathbb{S}}$ is partitioned in the previous section. Also, define U to be the $m \times m$ matrix of the eigen vectors of $\Gamma_\Lambda' \Gamma_\Lambda = \Gamma' \Omega^{-1} \Gamma$ and $U \Sigma V'$ be a singular value decomposition (SVD) of Γ_Λ. By construction, $U'U = UU' = I_m$, $V'V = V'V = I_d$, and $\Sigma' = [\ A_{d \times d} \quad O_{d \times q}\ ]$. We then define $\tilde{W} = U' W_\Lambda U$ and partition $\tilde{W}$ as before. We also introduce

$$\begin{aligned} \beta_{\mathbb{S}} &= \mathbb{S}_{dq} \mathbb{S}_{qq}^{-1}, \ \beta_{\tilde{W}} = \tilde{W}_{dq} \tilde{W}_{qq}^{-1}, \text{ and} \\ \kappa_G &= G \cdot \bar{B}_q' \mathbb{S}_{qq}^{-1} \bar{B}_q. \end{aligned}$$

By construction, $\beta_{\mathbb{S}}$ is the "random" regression coefficient induced by $\mathbb{S}$ while $\beta_{\tilde{W}}$ is the regression coefficient induced by the constant matrix $\tilde{W}$. Also, κ_G is the quadratic form of normal random vector $\sqrt{G} \bar{B}_q$ with random matrix $\mathbb{S}_{qq}$. Finally, on the basis of $\hat{\theta}_2$, the J statistic for testing over-identification restrictions is

$$J(\hat{\theta}_2) := n g_n(\hat{\theta}_2)' \left(\hat{\Omega}(\hat{\theta}_1) \right)^{-1} g_n(\hat{\theta}_2). \tag{9}$$

The following proposition characterizes and connects the limiting distributions of the three estimators: the first-step estimator $\hat{\theta}_1$, the feasible two-step estimator $\hat{\theta}_2$, and the infeasible two-step estimator $\tilde{\theta}_2$.

Proposition 6 *Let Assumptions 1∼6 hold. Then*

(*a*) $\sqrt{n}(\hat{\theta}_1 - \theta_0) \xrightarrow{d} -V A^{-1} \sqrt{G} (\bar{B}_d - \beta_{\tilde{W}} \bar{B}_q)$;

(*b*) $\sqrt{n}(\tilde{\theta}_2 - \theta_0) \xrightarrow{d} -V A^{-1} \sqrt{G} (\bar{B}_d - \beta_{\mathbb{S}} \bar{B}_q)$;

(c) $\sqrt{n}(\hat{\theta}_2 - \theta_0) \xrightarrow{d} -VA^{-1}\sqrt{G}(\bar{B}_d - \beta_{\mathbb{S}}\bar{B}_q) - (\frac{\kappa_G}{G}) \cdot VA^{-1}\sqrt{G}(\bar{B}_d - \beta_{\tilde{W}}\bar{B}_q)$;
(d) $\sqrt{n}(\hat{\theta}_2 - \theta_0) = \sqrt{n}(\tilde{\theta}_2 - \theta_0) + (\frac{\kappa_G}{G}) \cdot \sqrt{n}(\hat{\theta}_1 - \theta_0) + o_p(1)$;
(e) $J(\hat{\theta}_2) \xrightarrow{d} \kappa_G \overset{d}{=} G \cdot Beta(\frac{q}{2}, \frac{G-q}{2})$, *where* $(a), (b), (c)$, *and* (e) *hold jointly.*

Part (d) of the proposition shows that $\sqrt{n}(\hat{\theta}_2 - \theta_0)$ is asymptotically equivalent to a linear combination of the infeasible two-step estimator $\sqrt{n}(\tilde{\theta}_2 - \theta_0)$ and the first-step estimator $\sqrt{n}(\hat{\theta}_1 - \theta_0)$. This contrasts with the conventional GMM asymptotics, wherein the feasible and infeasible estimators are asymptotically equivalent.

It is interesting to see that the linear coefficient in Parts (c) and (d) is proportional to the limit of the J statistic. Given $\kappa_G = O_p(1)$ as G increases, the limiting distribution of $\sqrt{n}(\hat{\theta}_2 - \theta_0)$ becomes closer to that of $\sqrt{n}(\tilde{\theta}_2 - \theta_0)$. In the special case where $q = 0$, i.e., when the model is exactly identified, $\kappa_G = 0$ and $\sqrt{n}(\hat{\theta}_2 - \theta_0)$ and $\sqrt{n}(\tilde{\theta}_2 - \theta_0)$ have the same limiting distribution. This is expected given that the weighting matrix is irrelevant in the exactly identified GMM model.

Using the Sherman–Morrison formula[3], it is straightforward to show

$$\kappa_G = G \cdot \frac{\bar{B}_q' \tilde{\mathbb{S}}_{qq}^{-1} \bar{B}_q}{1 + \bar{B}_q' \tilde{\mathbb{S}}_{qq}^{-1} \bar{B}_q} \overset{d}{=} G \cdot \frac{q\mathcal{F}_{q,G-q}}{(G-q) + q\mathcal{F}_{q,G-q}}.$$

While the asymptotic distributions of $\hat{\theta}_2$ is complicated and nonstandard, the limiting distribution of the J statistic is not only pivotal but is also an increasing function of the standard F distribution. Furthermore, the equivalent Beta representation in Part (e) enables us to approximate the non-standard limit of J statistic by a (scaled) Beta random variable. For the practitioners, it is important to point out that the Beta limit of J statistic is valid only if the J statistic is equal to the GMM criterion function evaluated at the two-step GMM estimator $\hat{\theta}_2$. This effectively imposes a constraint on the weighting matrix. If we use a weighting matrix that is different from $\hat{\Omega}(\hat{\theta}_1)$, then the resulting J statistic does not have the Beta limit and is not even asymptotically pivotal any longer.

Define the F statistic and variance estimate for the two-step estimator $\hat{\theta}_2$ as

$$F_{\hat{\Omega}(\hat{\theta}_1)}(\hat{\theta}_2) = \frac{1}{p}(R\hat{\theta}_2 - r)' \left(R\widehat{var}_{\hat{\Omega}(\hat{\theta}_1)}(\hat{\theta}_2)R'\right)^{-1} (R\hat{\theta}_2 - r) \text{ for}$$
$$\widehat{var}_{\hat{\Omega}(\hat{\theta}_1)}(\hat{\theta}_2) = \frac{1}{n}\left(\hat{\Gamma}(\hat{\theta}_2)'\hat{\Omega}^{-1}(\hat{\theta}_1)\hat{\Gamma}(\hat{\theta}_2)\right)^{-1}.$$

In the above definitions, we use a subscript notation $\hat{\Omega}(\hat{\theta}_1)$ to clarify the choice of CCE in F statistic and asymptotic variance estimator above. Now the question is, is the above F statistic asymptotically pivotal as the J statistic $J(\hat{\theta}_2)$? Unfortunately, the answer is no, as implied by the following proposition which uses the additional notation:

$$\mathbb{E}_{p+q,p+q} := \begin{pmatrix} \mathbb{E}_{pp} & \mathbb{E}_{pq} \\ \mathbb{E}_{pq}' & \mathbb{E}_{qq} \end{pmatrix} = \begin{pmatrix} \bar{\mathbb{S}}_{pp} & \bar{\mathbb{S}}_{pq} \\ \bar{\mathbb{S}}_{pq}' & \bar{\mathbb{S}}_{qq} \end{pmatrix} + \begin{pmatrix} \tilde{\beta}_{\tilde{W}}^p \bar{B}_q \bar{B}_q' (\tilde{\beta}_{\tilde{W}}^p)' & \tilde{\beta}_{\tilde{W}}^p \bar{B}_q \bar{B}_q' \\ \bar{B}_q \bar{B}_q' (\tilde{\beta}_{\tilde{W}}^p)' & \bar{B}_q \bar{B}_q' \end{pmatrix}, \tag{10}$$

where $\tilde{\beta}_{\tilde{W}}^p$ is the $p \times q$ matrix and consists of the first p rows of $\tilde{V}'\beta_{\tilde{W}}$ where $\tilde{V}$ is the $d \times d$ matrix of the eigen vector of $\left(RVA^{-1}\right)' RVA^{-1}$.

[3] $(C + ab')^{-1} = C^{-1} - \frac{C^{-1}ab'C^{-1}}{1+b'C^{-1}a}$ for any invertable square matrix C and conforming column vectors such that $1 + b'C^{-1}a \neq 0$.

Proposition 7 *Let Assumptions 1∼6 hold. Then*

$$F_{\hat{\Omega}(\hat{\theta}_1)}(\hat{\theta}_2) \xrightarrow{d} \frac{G}{p}(\bar{B}_p - \mathbb{E}_{pq}\mathbb{E}_{qq}^{-1}\bar{B}_q)' \left(\mathbb{E}_{pp\cdot q}\right)^{-1} (\bar{B}_p - \mathbb{E}_{pq}\mathbb{E}_{qq}^{-1}\bar{B}_q)$$
$$= \frac{1}{p} \cdot \left[G \begin{pmatrix} \bar{B}_p \\ \bar{B}_q \end{pmatrix}' \begin{pmatrix} \mathbb{E}_{pp} & \mathbb{E}_{pq} \\ \mathbb{E}'_{pq} & \mathbb{E}_{qq} \end{pmatrix}^{-1} \begin{pmatrix} \bar{B}_p \\ \bar{B}_q \end{pmatrix} - G \cdot \bar{B}'_q \mathbb{E}_{qq}^{-1} \bar{B}_q \right], \tag{11}$$

where

$$\mathbb{E}_{pp\cdot q} = \mathbb{E}_{pp} - \mathbb{E}_{pq}\mathbb{E}_{qq}^{-1}\mathbb{E}'_{pq}.$$

Due to the presence of the second term of $\mathbb{E}_{p+q,p+q}$ in (10), which depends on $\tilde{\beta}_{\tilde{W}}$, the result of Proposition indicates that the F statistic is not asymptotically pivotal, and it depends on several nuisance parameters including Ω. To see this, we note that the second term in (11) is the same as $G \cdot \bar{B}'_q \mathbb{S}_{qq}^{-1} \bar{B}_q = \kappa_G$. Thus, the second term is the limit of the J statistic, which is nuisance parameter free. However, the first term in (11) is not pivotal because we have

$$G \begin{pmatrix} \bar{B}_p \\ \bar{B}_q \end{pmatrix}' \begin{pmatrix} \mathbb{E}_{pp} & \mathbb{E}_{pq} \\ \mathbb{E}'_{pq} & \mathbb{E}_{qq} \end{pmatrix}^{-1} \begin{pmatrix} \bar{B}_p \\ \bar{B}_q \end{pmatrix}$$
$$= G \left[\begin{pmatrix} \bar{B}_p \\ \bar{B}_q \end{pmatrix}' \begin{pmatrix} \bar{\mathbb{S}}_{pp} & \bar{\mathbb{S}}_{pq} \\ \bar{\mathbb{S}}'_{pq} & \bar{\mathbb{S}}_{qq} \end{pmatrix}^{-1} \begin{pmatrix} \bar{B}_p \\ \bar{B}_q \end{pmatrix} - \frac{\left(\bar{B}'_{p+q} \bar{\mathbb{S}}_{p+q,p+q}^{-1} \tilde{w} \bar{B}_q\right)^2}{1 + \bar{B}'_q \tilde{w}'_{p+q} \bar{\mathbb{S}}_{p+q}^{-1} \tilde{w} \bar{B}_q} \right],$$

where $\tilde{w} = ((\tilde{\beta}_W^p)', I_q)'$. Here, as in the case of the J statistic, the first term in the above equation is nuisance parameter free. But the second term is clearly a nonconstant function of $\tilde{\beta}_W^p$, which, in turn, depends on R, Γ, W and Ω.

3.2 Centered Two-step GMM estimator

Given that the estimation error in $\hat{\theta}_1$ affects the limiting distribution of $\hat{\Omega}(\hat{\theta}_1)$, the Wald statistic based on the uncentered two-step GMM estimator $\hat{\theta}_2$ is not asymptotically pivotal. In view of (4), the effect of the estimation error is manifested via a location shift in $\tilde{D}_g$; the shifting amount depends on $\hat{\theta}_1$. A key observation is that the location shift is the same for all groups under the homogeneity Assumptions 5 and 6. Therefore, if we demean the empirical moment process, we can remove the location shift that is caused by the estimator error in $\hat{\theta}_1$. This leads to the recentered asymptotic variance estimator and a pivotal inference for both the Wald test and J test.

It is important to note that the recentering is not innocuous for an over-identified GMM model because $n^{-1}\sum_{i=1}^n f_i(\hat{\theta}_1)$ is not zero in general. In the time series HAR variance estimation, the recentering is known to have several advantages. For example, as Hall (2000) observes, in the conventional increasing smoothing asymptotic theory, the recentering can potentially improve the power of the J test using a HAR variance estimator when the model is misspecified.

In our small-G asymptotic framework, the recentering plays an important role in the CCE estimation. It ensures that the limiting distribution of $\hat{\Omega}^c(\hat{\theta}_1)$ is invariant to the initial estimator $\hat{\theta}_1$. The following lemma proves a more general result and characterizes the small-G limiting distribution of the centered CCE matrix for any $\sqrt{n}$-consistent estimator $\tilde{\theta}$.

Lemma 8 *Let Assumptions 1∼6 hold. Let $\tilde{\theta}$ be any $\sqrt{n}$-consistent estimator of θ_0. Then*

(*a*) $\hat{\Omega}^c(\tilde{\theta}) = \hat{\Omega}^c(\theta_0) + o_p(1)$;

(*b*) $\hat{\Omega}^c(\theta_0) \xrightarrow{d} \Omega^c_\infty$ *where* $\Omega^c_\infty = \Lambda \bar{\mathbb{S}} \Lambda'$.

Lemma 8 indicates that the centered CCE $\Omega^c(\hat{\theta}_1)$ converges in distribution to the random matrix limit $\Omega^c_\infty = \Lambda\bar{\mathbb{S}}\Lambda'$, which follows a (scaled) Wishart distribution $G^{-1}\mathbb{W}_m(G-1,\Omega)$. Using Lemma 8, it is possible to show

$$\begin{aligned}\sqrt{n}(\hat{\theta}_2^c - \theta_0) &= -\left(\Gamma'\left[\hat{\Omega}^c(\hat{\theta}_1)\right]^{-1}\Gamma\right)^{-1}\Gamma'\left[\hat{\Omega}^c(\hat{\theta}_1)\right]^{-1}\sqrt{n}g_n(\theta_0) + o_p(1) \\ &\xrightarrow{d} -\left[\Gamma'\left(\Omega^c_\infty\right)^{-1}\Gamma\right]^{-1}\Gamma'\left(\Omega^c_\infty\right)^{-1}\Lambda\sqrt{G}\bar{B}_m.\end{aligned} \tag{12}$$

Since $(\Omega^c_\infty)^{-1}$ is independent with $\sqrt{G}\Lambda\bar{B}_m \sim N(0,\Omega)$, the limiting distribution of $\hat{\theta}_2^c$ is mixed normal.

On the basis of $\hat{\theta}_2^c$, we can construct the "trinity" of GMM test statistics. The first one is the normalized Wald statistic defined by

$$\begin{aligned}F_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c) &:= \frac{1}{p}(R\hat{\theta}_2^c - r)'\{R\widehat{var}_{\hat{\Omega}^c(\hat{\theta}_2^c)}(\hat{\theta}_2^c)R'\}^{-1}(R\hat{\theta}_2^c - r), \text{ where} \\ \widehat{var}_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c) &= \frac{1}{n}\left(\hat{\Gamma}(\hat{\theta}_2^c)'\left(\hat{\Omega}^c(\hat{\theta}_1)\right)^{-1}\hat{\Gamma}(\hat{\theta}_2^c)\right)^{-1}.\end{aligned} \tag{13}$$

When $p=1$, corresponding t statistic $t_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c)$ can be constructed similarly.

The second test statistic is the Quasi-Likelihood Ratio (QLR) type of statistic. 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 given by

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

The last test statistic we consider is the Lagrangian Multiplier (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 8 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 small-G asymptotic results. This contrasts with the small-G asymptotics for the uncentered two-step estimator $\hat{\theta}_2$. Lastly, we also construct the standard J statistic based on $\hat{\theta}_2^c$

$$J(\hat{\theta}_2^c) := ng_n(\hat{\theta}_2^c)'\left(\hat{\Omega}^c(\hat{\theta}_1)\right)^{-1} g_n(\hat{\theta}_2^c),$$

where $\hat{\Omega}^c(\hat{\theta}_1)$ can be replaced by $\hat{\Omega}^c(\hat{\theta}_2^c)$ without affecting the limiting distribution of the J statistic.

Using (12) and Lemma 8, we have $F_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c) \xrightarrow{d} \mathbb{F}_{2\infty}$ where

$$\mathbb{F}_{2\infty} = \frac{G}{p} \cdot \left[R\left(\Gamma_\Lambda' \bar{\mathbb{S}}^{-1} \Gamma_\Lambda\right)^{-1} \Gamma_\Lambda' \bar{\mathbb{S}}^{-1} \bar{B}_m\right]' \left[R\left(\Gamma_\Lambda' \bar{\mathbb{S}}^{-1} \Gamma_\Lambda\right)^{-1} R'\right]^{-1} \times \left[R\left(\Gamma_\Lambda' \bar{\mathbb{S}}^{-1} \Gamma_\Lambda\right)^{-1} \Gamma_\Lambda' \bar{\mathbb{S}}^{-1} \bar{B}_m\right]. \tag{14}$$

When $p = 1$, we get $t_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c) \xrightarrow{d} \mathbb{T}_{2\infty}$ with

$$\mathbb{T}_{2\infty} = \frac{R\left(\Gamma_\Lambda' \bar{\mathbb{S}}^{-1} \Gamma_\Lambda\right)^{-1} \Gamma_\Lambda' \bar{\mathbb{S}}^{-1} \sqrt{G} \bar{B}_m}{\sqrt{R\left(\Gamma_\Lambda' \bar{\mathbb{S}}^{-1} \Gamma_\Lambda\right)^{-1} R'}}. \tag{15}$$

Also, it follows in a similar way that

$$J(\hat{\theta}_2^c) \xrightarrow{d} \mathbb{J}_\infty := G \cdot \left\{\bar{B}_m - \Gamma_\Lambda \left(\Gamma_\Lambda' \bar{\mathbb{S}}^{-1} \Gamma_\Lambda\right)^{-1} \Gamma_\Lambda' \bar{\mathbb{S}}^{-1} \bar{B}_m\right\}' \bar{\mathbb{S}}^{-1} \times \left\{\bar{B}_m - \Gamma_\Lambda \left(\Gamma_\Lambda' \bar{\mathbb{S}}^{-1} \Gamma_\Lambda\right)^{-1} \Gamma_\Lambda' \bar{\mathbb{S}}^{-1} \bar{B}_m\right\}. \tag{16}$$

The remaining question is whether the above representations for $\mathbb{F}_{2\infty}$ and $\mathbb{J}_\infty$ are free of nuisance parameters. The following proposition provides a positive answer.

Proposition 9 *Let Assumptions 1∼6 hold and define* $\bar{\mathbb{S}}_{pp\cdot q} = \bar{\mathbb{S}}_{pp} - \bar{\mathbb{S}}_{pq} \bar{\mathbb{S}}_{qq}^{-1} \bar{\mathbb{S}}_{qp}$.

(a) $F_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c) \xrightarrow{d} \frac{G}{p} \cdot \left(\bar{B}_p - \bar{\mathbb{S}}_{pq} \bar{\mathbb{S}}_{qq}^{-1} \bar{B}_q\right)' \bar{\mathbb{S}}_{pp\cdot q}^{-1} \left(\bar{B}_p - \bar{\mathbb{S}}_{pq} \bar{\mathbb{S}}_{qq}^{-1} \bar{B}_q\right)' \overset{d}{=} \mathbb{F}_{2\infty}$;

(b) $t_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c) \xrightarrow{d} \sqrt{G}\left(\bar{B}_p - \bar{\mathbb{S}}_{pq} \bar{\mathbb{S}}_{qq}^{-1} \bar{B}_q\right) / \sqrt{\bar{\mathbb{S}}_{pp\cdot q}} \overset{d}{=} \mathbb{T}_{2\infty}$ *for* $p = 1$;

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

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

(e) $J(\hat{\theta}_2^c) \xrightarrow{d} G \cdot \bar{B}_q' \bar{\mathbb{S}}_{qq}^{-1} \bar{B}_q \overset{d}{=} \mathbb{J}_\infty$.

To simplify the representations of $\mathbb{F}_{2\infty}$ and $\mathbb{T}_{2\infty}$ in the above proposition, we note that

$$G \begin{bmatrix} \bar{\mathbb{S}}_{pp} & \bar{\mathbb{S}}_{pq} \\ \bar{\mathbb{S}}_{qp} & \bar{\mathbb{S}}_{qq} \end{bmatrix} \overset{d}{=} \sum_{g=1}^{G} \left(B_{p+q,g} - \bar{B}_{p+q}\right)\left(B_{p+q,g} - \bar{B}_{p+q}\right)',$$

where $B_{p+q,g} := (B_{p,g}', B_{p,g}')'$. The above random matrix has a standard Wishart distribution $\mathbb{W}_{p+q}(G-1, I_{p+q})$. It follows from the well-known properties of a Wishart distribution that $\bar{\mathbb{S}}_{pp\cdot q} \sim \mathbb{W}_p(G-1-q, I_p)/G$ and $\bar{\mathbb{S}}_{pp\cdot q}$ is independent of $\bar{\mathbb{S}}_{pq}$ and $\bar{\mathbb{S}}_{qq}$.See Bilodeau and Brenner (2008, Proposition 7.9). Therefore, if we condition on $\Delta := \bar{\mathbb{S}}_{pq} \bar{\mathbb{S}}_{qq}^{-1} \sqrt{G} \bar{B}_q$, the limiting distribution $\mathbb{F}_{2\infty}$ satisfies

$$\frac{G-p-q}{G} \mathbb{F}_{2\infty} \overset{d}{=} \frac{G-p-q}{G} \frac{(\sqrt{G}\bar{B}_p + \Delta)' \bar{\mathbb{S}}_{pp\cdot q}^{-1} (\sqrt{G}\bar{B}_p + \Delta)}{p} \overset{d}{=} \mathcal{F}_{p,G-p-q}\left(\|\Delta\|^2\right), \tag{17}$$

where $\mathcal{F}_{p,G-p-q}(\|\Delta\|^2)$ is a noncentral F distribution with random noncentrality parameter $\|\Delta\|^2$. Similarly, the limiting distribution of (scaled) $\mathbb{T}_{2\infty}$ can be represented as

$$\sqrt{\frac{G-1-q}{G}} \mathbb{T}_{2\infty} \overset{d}{=} \sqrt{\frac{G-1-q}{G}} \frac{\sqrt{G}\bar{B}_p + \Delta}{\sqrt{\bar{\mathbb{S}}_{pp\cdot q}}} \overset{d}{=} t_{G-1-q}(\Delta), \tag{18}$$

which is a noncentral t distribution with a noncentrality parameter Δ. These nonstandard limiting distributions are similar to those in Sun (2014) which provides the fixed-smoothing asymptotic result in the case of the series LRV estimation. However, in our setting of clustered dependence, the scale adjustment and degrees of freedom parameter in (17) and (18) are different from those in Sun (2014).

The critical values from the nonstandard limiting distribution $\mathbb{F}_{2\infty}$ can be obtained through simulation, but Sun (2014) shows that $\mathbb{F}_{2\infty}$ can be approximated by a noncentral F distribution. With regard to the QLR and LM types of test statistics, Proposition 9-(c) and (d) shows that they are asymptotically equivalent to $F_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c)$. This also implies that all three types of test statistics share the same small-G limit as given in (17) and (18). Similar results are obtained by Sun (2014) and Hwang and Sun (2017), which focus on the two-step GMM estimation and HAR inference in a time series setting.

For the J statistic $J(\hat{\theta}_2^c)$, it follows from Proposition 9-(e) that

$$\left(\frac{G-q}{Gq}\right) \cdot J(\hat{\theta}_2^c) \xrightarrow{d} \left(\frac{G-q}{Gq}\right) \cdot \bar{B}_q' \bar{\mathbb{S}}_{qq}^{-1} \bar{B}_q \overset{d}{=} \mathcal{F}_{q,G-q}.$$

This is consistent with Sun and Kim's (2012) results except that our adjustment and degrees of freedom parameter are different. A recent study by Hayakawa (2016) also discusses the Beta and F limiting distributions of uncentered and centered J statistics, respectively. Comparing to what we develop in Propositions 6 and 9, however, his approximations are built upon the assumption of independent Gaussian moment process which are quite restrictive in empirical modeling.

4 Asymptotic F and t Tests for Centered Two-step GMM Procedures

The limiting distributions of the centered two-step GMM test statistics in Section 3 are nonstandard under the small-G asymptotics, and hence the corresponding critical values have to be simulated in practice. This contrasts with the conventional large-G asymptotics, where the limiting distributions are the standard chi-squared and normal distributions. In this section, we show that a simple modification of the two-step Wald and t statistics enables us to develop the standard F and t asymptotic theory under the small-G asymptotics. The asymptotic F and t tests are more appealing in empirical applications because the standard F and t distributions are more accessible than the nonstandard $\mathbb{F}_{2\infty}$ and $\mathbb{T}_{2\infty}$ distributions.

The modified two-step Wald, QLR and LM statistics are

$$\begin{aligned}
\tilde{F}_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c) &:= \frac{G-p-q}{G} \cdot \frac{F_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c)}{1+\frac{1}{G}J(\hat{\theta}_2^c)}, \qquad (19)\\
\widetilde{LR}_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c, \hat{\theta}_2^{c,r}) &:= \frac{G-p-q}{G} \cdot \frac{LR_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c, \hat{\theta}_2^{c,r})}{1+\frac{1}{G}J(\hat{\theta}_2^c)},\\
\widetilde{LM}_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^{c,r}) &:= \frac{G-p-q}{G} \cdot \frac{LM_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^{c,r})}{1+\frac{1}{G}J(\hat{\theta}_2^c)},
\end{aligned}$$

and the corresponding version of the t statistic is

$$\tilde{t}_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c) := \sqrt{\frac{G-1-q}{G}} \cdot \frac{t_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c)}{\sqrt{1+\frac{1}{G}J(\hat{\theta}_2^c)}}.$$

The modified test statistics involve a scale multiplication factor that uses the usual J statistic and a constant factor that adjusts the degrees of freedom.

It follows from Proposition 9 that

$$\left(F_{\hat{\Omega}^c(\hat{\theta}_2^c)}(\hat{\theta}_2^c), J(\hat{\theta}_2^c)\right) \xrightarrow{d} (\mathbb{F}_{2\infty}, \mathbb{J}_\infty) \tag{20}$$

$$\stackrel{d}{=} \left(\frac{G}{p}\left(\bar{B}_p - \bar{\mathbb{S}}_{pq}\bar{\mathbb{S}}_{qq}^{-1}\bar{B}_q\right)'\bar{\mathbb{S}}_{pp\cdot q}^{-1}\left(\bar{B}_p - \bar{\mathbb{S}}_{pq}\bar{\mathbb{S}}_{qq}^{-1}\bar{B}_q\right)', G\cdot\bar{B}_q'\bar{\mathbb{S}}_{qq}^{-1}\bar{B}_q\right) \tag{21}$$

Thus,

$$\tilde{F}_{\hat{\Omega}^c(\hat{\theta}_2^c)}(\hat{\theta}_2^c) \xrightarrow{d} \frac{G-p-q}{G}\frac{\mathbb{F}_{2\infty}}{1+\frac{1}{G}\mathbb{J}_\infty} \stackrel{d}{=} \frac{G-p-q}{pG}\xi_p'\tilde{\mathbb{S}}_{pp\cdot q}^{-1}\xi_p,$$

where

$$\xi_p := \frac{\sqrt{G}(\bar{B}_p - \bar{\mathbb{S}}_{pq}\bar{\mathbb{S}}_{qq}^{-1}\bar{B}_q)}{\sqrt{1+\bar{B}_q'\bar{\mathbb{S}}_{qq}^{-1}\bar{B}_q}}.$$

Similarly,

$$\tilde{t}_{\hat{\Omega}^c(\hat{\theta}_2^c)}(\hat{\theta}_2^c) \xrightarrow{d} \sqrt{\frac{G-1-q}{G}} \cdot \frac{\mathbb{T}_{2\infty}}{\sqrt{1+\frac{1}{G}\mathbb{J}_\infty}} \stackrel{d}{=} \frac{\xi_p}{\sqrt{\tilde{\mathbb{S}}_{pp\cdot q}}}.$$

In the proof of Theorem 10 we show that ξ_p follows a standard normal distribution $N(0, I_p)$, and that ξ_p is independent of $\tilde{S}_{pp\cdot q}^{-1}$. Thus, the limiting distribution of $\tilde{F}_{\hat{\Omega}^c(\hat{\theta}_2^c)}(\hat{\theta}_2^c)$ is proportional to a quadratic form in the standard normal vector ξ_p with an independent inverse-Wishart distributed weighting matrix $\tilde{\mathbb{S}}_{pp\cdot q}^{-1}$. It follows from a theory of multivariate statistics that the limiting distribution of $\tilde{F}_{\Omega^c(\hat{\theta}_2^c)}(\hat{\theta}_2^c)$ is $\mathcal{F}_{p,G-p-q}$. Similarly, the limiting distribution of $\tilde{t}_{\hat{\Omega}^c(\hat{\theta}_2^c)}(\hat{\theta}_2^c)$ is t_{G-1-q}. This is formalized in the following theorem.

Theorem 10 *Let Assumptions 1∼6 hold. Then the modified Wald, QLR and LM all converge in distribution to $\mathcal{F}_{p,G-p-q}$. Also, the t statistic has limiting distribution t_{G-1-q}.*

The equivalence relationship between the modified Wald, LR,LM is consistent with the recent paper by Hwang and Sun (2017a) which establishes the asymptotic F and t limit theory of two-step GMM in a time series setting. But our cluster-robust limiting distributions in Theorem 10 are different from Hwang and Sun (2017a) in terms of the multiplicative adjustment and the degrees of freedom correction.

It follows from the proofs of Theorem 10 and Proposition 9 that

$$\sqrt{n}(\hat{\theta}_2^c - \theta_0) \xrightarrow{d} MN\left(0, \left(\Gamma'\Omega^{-1}\Gamma\right)^{-1}\cdot(1+\bar{B}_q'\bar{\mathbb{S}}_{qq}^{-1}\bar{B}_q)\right) \text{ and} \tag{22}$$

$$J(\hat{\theta}_2^c) \xrightarrow{d} G\cdot\bar{B}_q'\bar{\mathbb{S}}_{qq}^{-1}\bar{B}_q$$

hold jointly under small-G asymptotics. Here, $MN(0,\mathbb{V})$ denotes a random variable that follows a mixed normal distribution with conditional variance $\mathbb{V}$. The random multiplication term $(1+$

$\bar{B}_q'\tilde{S}_{qq}^{-1}\bar{B}_q$) in (22) reflects the estimation uncertainty of CCE weighting matrix on the limiting distribution of $\sqrt{n}(\hat{\theta}_2^c - \theta_0)$. The small-$G$ limiting distribution in (22) is in sharp contrast to that of under the conventional large-G asymptotics as the latter completely ignores the variability in the cluster-robust GMM weighting matrix. By continuous mapping theorem,

$$\frac{\sqrt{n}(\hat{\theta}_2^c - \theta_0)}{\sqrt{1 + \frac{1}{G}J(\hat{\theta}_2^c)}} \xrightarrow{d} N\left(0, \left(\Gamma'\Omega^{-1}\Gamma\right)^{-1}\right). \tag{23}$$

and this shows that the J statistic modification factor in the denominator effectively cancels out the uncertainty of CCE to recover the limiting distribution of $\sqrt{n}(\hat{\theta}_2^c - \theta_0)$ under the conventional large-G asymptotics. In view of (23), the finite sample distribution of $\sqrt{n}(\hat{\theta}_2^c - \theta_0)$ conditional on the J statistic $J(\hat{\theta}_2^c)$, can be well-approximated by $N(0, \widehat{var}_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c))$ where

$$\widehat{var}_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c) := \widehat{var}_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c) \cdot \left(1 + \frac{1}{G}J(\hat{\theta}_2^c)\right). \tag{24}$$

The modification term $(1 + (1/G)J(\hat{\theta}_2^c))^{-1}$ degenerates to one as G increases so that the two variance estimates in (24) become close to each other. Thus, the multiplicative term $(1 + (1/G)J(\hat{\theta}_2^c))^{-1}$ in (19) can be regarded as a finite sample modification to the standard variance estimate $\widehat{var}_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c)$ under the large-G asymptotics. For more discussions about the role of J statistic modification, see Hwang and Sun (2017b) which casts the two-step GMM problems into OLS estimation and inference in classical normal linear regression.

5 Iterative Two-step and Continuous Updating Schemes

Another class of popular GMM estimators is the continuous updating (CU) estimators, which are designed to improve the poor finite sample performance of two-step GMM estimators. See Hansen, Heaton, and Yaron (1996) and Newey and Smith (2004) for more discussion on the CU-type estimators. Here, 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 $\hat{\theta}_{\text{IE}}^j$ is

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

In view of the above FOC, $\hat{\theta}_{\text{IE}}^j$ 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 $\hat{\theta}_{\text{IE}}^j$ converges, we obtain the continuously update GEE estimator $\hat{\theta}_{\text{CU-GEE}}$. The FOC for $\hat{\theta}_{\text{CU-GEE}}$ is given by

$$\hat{\Gamma}(\hat{\theta}_{\text{CU-GEE}})'\hat{\Omega}^{-1}(\hat{\theta}_{\text{CU-GEE}})g_n(\hat{\theta}_{\text{CU-GEE}}) = 0. \tag{25}$$

We employ the uncentered CCE, $\hat{\Omega}(\cdot)$ in the definition of $\hat{\theta}_{\text{CU-GEE}}$, but it is not difficult to show that

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

where

$$\nu_n(\hat{\theta}_{\text{CU-GEE}}) = L \cdot g_n(\hat{\theta}_{\text{CU-GEE}})' \left(\hat{\Omega}^c(\hat{\theta}_{\text{CU-GEE}})\right)^{-1} g_n(\hat{\theta}_{\text{CU-GEE}}).$$

Since $1/(1+\nu_n(\hat{\theta}_{\text{CU-GEE}}))$ is always positive, the first-order condition in (25) holds if and only if

$$\hat{\Gamma}(\hat{\theta}_{\text{CU-GEE}})' \left[\hat{\Omega}^c(\hat{\theta}_{\text{CU-GEE}})\right]^{-1} g_n(\hat{\theta}_{\text{CU-GEE}}) = 0, \tag{26}$$

which indicates that the recentering CCE weight in (25) 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 definition of $\hat{\theta}_{\text{CU-GMM}}$ in Hansen, Heaton and Yaron (1996) is based on the centered CCE weighting matrix $\hat{\Omega}^c(\theta)$. It is easy to show that

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

Thus, we have

$$L \cdot g_n(\theta)' \left(\hat{\Omega}^c(\theta)\right)^{-1} g_n(\theta) = \frac{L \cdot g_n(\theta)'\hat{\Omega}^{-1}(\theta)g_n(\theta)}{1 - L \cdot g_n(\theta)'\hat{\Omega}^{-1}(\theta)g_n(\theta)}.$$

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 $\hat{\Omega}^c(\theta)$, that is,

$$\hat{\theta}_{\text{CU-GMM}} = \arg\min_{\theta\in\Theta} g_n(\theta)' \left[\hat{\Omega}^c(\theta)\right]^{-1} g_n(\theta). \tag{27}$$

Similar to the centered two-step GMM estimator, the two CU estimators can be regarded as having a built-in recentering mechanism. 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 11 *Let Assumptions 1, 3∼6 hold. Assume that $\hat{\theta}_{CU\text{-}GEE}$ and $\hat{\theta}_{CU\text{-}GMM}$ are $\sqrt{n}$-consistent. Then*

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

and

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

The proposition shows that the CU estimators and the centered two-step GMM estimator are asymptotically equivalent under the small-G asymptotics. Based on the two CU estimators, we construct the Wald statistics as

$$F_{\hat{\Omega}^c(\hat{\theta}_{\text{CU-GEE}})}(\hat{\theta}_{\hat{\theta}_{\text{CU-GEE}}}) = \frac{1}{p}(R\hat{\theta}_{\text{CU-GEE}} - r)'\{R\widehat{var}_{\hat{\Omega}^c(\hat{\theta}_{\text{CU-GEE}})}(\hat{\theta}_{\text{CU-GEE}})R'\}^{-1}(R\hat{\theta}_{\text{CU-GEE}} - r) \tag{28}$$

and

$$F_{\hat{\Omega}^c(\hat{\theta}_{\text{CU-GMM}})}(\hat{\theta}_{\hat{\theta}_{\text{CU-GMM}}}) = \frac{1}{p}(R\hat{\theta}_{\text{CU-GMM}} - r)'\{\widehat{Rvar}_{\hat{\Omega}^c(\hat{\theta}_{\text{CU-GMM}})}(\hat{\theta}_{\text{CU-GMM}})R'\}^{-1}(R\hat{\theta}_{\text{CU-GMM}} - r). \tag{29}$$

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 11 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,

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

In summary, we have shown that all three estimators $\hat{\theta}_2^c$, $\hat{\theta}_{\text{CU-GEE}}$ and $\hat{\theta}_{\text{CU-GMM}}$, and the corresponding Wald test statistics converge in distribution to the same nonstandard distributions. Proposition 9-(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.

The findings in this section are quite interesting. Under the first order large-G asymptotics, the CU estimators and the default (uncentered) two-step GMM are all asymptotically equivalent. In other words, the first-order large-G asymptotics is not informative about the merits of the CU estimators. One may develop a high order expansion under the large-G asymptotics to reveal the advantages of CU estimators. In fact, Newey and Smith (2004) develop the stochastic expansion of CU estimators in an i.i.d setting and show that the CU schemes automatically remove the high order estimation error of two-step estimator which is caused by the non-optimal weighting matrix in the first-step estimator. We could adopt these approaches, instead of the small-G asymptotics, to capture the estimation uncertainty of the first-step estimator in the default (uncentered) two-step GMM procedures. But the high order asymptotic analysis is technically very challenging and often requires strong assumptions on the smoothness of moment process. Although the small-G asymptotics we develop here is just a first order theory, it is powerful enough to reveal the asymptotic difference between the CU and the plain uncentered two-step GMM estimators. Moreover, the built-in recentering function behind the CU estimators provides some justification for the use of the centered CCE in a two-step GMM framework.

Lastly, together with Theorem 10, Proposition 11 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 small-G asymptotics.

6 Finite Sample Variance Correction

The recentering scheme we investigate in the previous sections enables us to remove the first order effect of the first-step estimation error, but the centered two-step GMM estimator still faces some extra estimation uncertainty in the first-step estimator. The main source of the problem is that we have to estimate the unobserved moment process based on the first-step estimator. To be

more specific, define the infeasible two-step GMM estimator with the centered CCE weighting matrix $\hat{\Omega}^c(\theta_0)$ as

$$\tilde{\theta}_2^c = \arg\min_{\theta\in\Theta} g_n(\theta)' \left(\hat{\Omega}^c(\theta_0)\right)^{-1} g_n(\theta).$$

Then

$$\sqrt{n}(\tilde{\theta}_2^c - \theta_0) = -\left[\Gamma' \left(\hat{\Omega}^c(\theta_0)\right)^{-1} \Gamma\right]^{-1} \Gamma' \left(\hat{\Omega}^c(\theta_0)\right)^{-1} \sqrt{n} g_n(\theta_0) + o_p(1)$$

. But we also have

$$\sqrt{n}(\hat{\theta}_2^c - \theta_0) = -\left[\Gamma' \left(\hat{\Omega}^c(\hat{\theta}_1)\right)^{-1} \Gamma\right]^{-1} \Gamma' \left(\hat{\Omega}^c(\hat{\theta}_1)\right)^{-1} \sqrt{n} g_n(\theta_0) + o_p(1), \tag{30}$$

Together with Lemma 8, this implies that

$$\sqrt{n}(\hat{\theta}_2^c - \theta_0) = \sqrt{n}(\tilde{\theta}_2^c - \theta_0) + o_p(1).$$

That is, the estimation error in $\hat{\theta}_1$ has no effect on the asymptotic distribution of $\sqrt{n}(\hat{\theta}_2^c - \theta_0)$ in the first-order asymptotic analysis. However, in finite samples $\hat{\theta}_2^c$ does have higher variation than $\tilde{\theta}_2^c$, and this can be attributed to the high variation in $\hat{\Omega}^c(\hat{\theta}_1)$ than $\hat{\Omega}^c(\theta_0)$. To account for this extra variation, we could develop a higher order asymptotic theory under the small-G asymptotics. But this is a formidable task that requires new technical machinery and lengthy calculations. Instead, we keep one additional term in the stochastic expansion of $\sqrt{n}(\hat{\theta}_2^c - \theta_0)$ in hopes of developing a finite sample correction to our asymptotic variance estimator.

To this end, we first introduce the notion of asymptotic equivalence in distribution $\xi_n \overset{a}{\sim} \eta_n$ for two stochastically bounded sequences of random vectors $\xi_n \in \mathbb{R}^\ell$ and $\eta_n \in \mathbb{R}^\ell$ when ξ_n and η_n converge in distribution to each other. Now under the small-G asymptotics we have

$$\sqrt{n}(\hat{\theta}_2^c - \theta_0) \overset{a}{\sim} -\left\{\Gamma' \left[\hat{\Omega}^c(\theta_0)\right]^{-1} \Gamma\right\}^{-1} \Gamma' \left[\hat{\Omega}^c(\theta_0)\right]^{-1} \sqrt{n} g_n(\theta_0) + (\mathcal{E}_{1n} + \mathcal{E}_{2n}) \sqrt{n}(\hat{\theta}_1 - \theta_0),$$

where

$$\mathcal{E}_{1n} = -\left.\frac{\partial \left\{\Gamma' \left[\hat{\Omega}^c(\theta)\right]^{-1} \Gamma\right\}^{-1}}{\partial \theta'}\right|_{\theta=\theta_0} \Gamma' \left[\hat{\Omega}^c(\theta_0)\right]^{-1} g_n(\theta_0)$$

$$\mathcal{E}_{2n} = -\left\{\Gamma' \left[\hat{\Omega}^c(\theta)\right]^{-1} \Gamma\right\}^{-1} \left.\frac{\partial \Gamma' \left[\hat{\Omega}^c(\theta)\right]^{-1} g_n(\theta_0)}{\partial \theta'}\right|_{\theta=\theta_0}$$

are $d \times d$ matrices. In finite samples, if we estimate the term $\Gamma'[\hat{\Omega}^c(\theta_0)]^{-1} g_n(\theta_0)$ in $\mathcal{E}_{1n}$ by $\hat{\Gamma}(\hat{\theta}_2^c)[\hat{\Omega}^c(\hat{\theta}_1)]^{-1} g_n(\hat{\theta}_2^c)$, then the estimate will be identically zero because of the FOC. For this reason, we drop $\mathcal{E}_{1n}$ and keep only $\mathcal{E}_{2n}$, which leads to the following distributional approximation

$$\sqrt{n}(\hat{\theta}_2^c - \theta_0) \overset{a}{\sim} -\left\{\Gamma' \left[\hat{\Omega}^c(\theta_0)\right]^{-1} \Gamma\right\}^{-1} \Gamma' \left[\hat{\Omega}^c(\theta_0)\right]^{-1} \sqrt{n} g_n(\theta_0) + \mathcal{E}_{2n} \sqrt{n}(\hat{\theta}_1 - \theta_0). \tag{31}$$

Using element by element differentiation with respect to θ_j for $1 \leq j \leq d$, we can write the j-th column of $\mathcal{E}_{2n}$ as

$$\mathcal{E}_{2n}[.,j] = \left\{\Gamma' \left[\hat{\Omega}^c(\theta_0)\right]^{-1} \Gamma\right\}^{-1} \Gamma' \left[\hat{\Omega}^c(\theta_0)\right]^{-1} \left.\frac{\partial \hat{\Omega}^c(\theta)}{\partial \theta_j}\right|_{\theta=\theta_0} \left[\hat{\Omega}^c(\theta_0)\right]^{-1} g_n(\theta_0), \tag{32}$$

where

$$\begin{aligned}
\frac{\partial \hat{\Omega}^c(\theta_0)}{\partial \theta_j} &= \Upsilon_j(\theta_0) + \Upsilon_j'(\theta_0) \text{ and} \\
\Upsilon_j(\theta_0) &= \frac{1}{G}\sum_{g=1}^{G}\left[\frac{1}{\sqrt{L}}\sum_{k=1}^{L}\left(f_k^g(\theta_0) - \frac{1}{n}\sum_{i=1}^{n} f_i(\theta_0)\right)\right. \\
&\quad \left. \cdot \frac{1}{\sqrt{L}}\sum_{k=1}^{L}\left(\frac{\partial f_k^g(\theta_0)}{\partial \theta_j} - \frac{1}{n}\sum_{i=1}^{n}\frac{\partial f_i(\theta_0)}{\partial \theta_j}\right)'\right].
\end{aligned} \tag{33}$$

Note that the term $\mathcal{E}_{2n}\sqrt{n}(\hat{\theta}_1 - \theta_0)$ has no first order effect on the asymptotic distribution of $\sqrt{n}(\hat{\theta}_2^c - \theta_0)$. This is true because $\mathcal{E}_{2n}$ converges to zero in probability. In fact, it follows from (32) and (33) that $\mathcal{E}_{2n} = O_p(n^{-1/2})$.

It follows from (31) that

$$\sqrt{n}(\hat{\theta}_2^c - \theta_0) \overset{a}{\sim} -\begin{pmatrix} \left[\Gamma'\left(\Omega_\infty^c\right)^{-1}\Gamma\right]^{-1} & \mathcal{E}_n(\Gamma' W^{-1}\Gamma)^{-1} \end{pmatrix} \begin{pmatrix} \Gamma'\left(\Omega_\infty^c\right)^{-1}\Lambda Z \\ \Gamma' W^{-1}\Lambda Z \end{pmatrix}, \tag{34}$$

where $Z \sim N(0, I_d)$, Z is independent of Ω_∞^c, and $\mathcal{E}_n$ has the same marginal distribution as $\mathcal{E}_{2n}$, but it is independent of Z and Ω_∞^c. It then follows that $\sqrt{n}(\hat{\theta}_2^c - \theta_0)$ is asymptotically equivalent in distribution to the mixed normal distribution with the conditional variance given by

$$\Xi_n = \begin{pmatrix} \left[\Gamma'\left(\Omega_\infty^c\right)^{-1}\Gamma\right]^{-1} \\ (\Gamma' W^{-1}\Gamma)^{-1}\mathcal{E}_n' \end{pmatrix}' \begin{pmatrix} \Gamma'\left(\Omega_\infty^c\right)^{-1}\Omega\left(\Omega_\infty^c\right)^{-1}\Gamma & \Gamma'\left(\Omega_\infty^c\right)^{-1}\Omega W^{-1}\Gamma \\ \Gamma' W^{-1}\Omega'\left(\Omega_\infty^c\right)^{-1}\Gamma & \Gamma' W^{-1}\Omega W^{-1}\Gamma \end{pmatrix} \begin{pmatrix} \left[\Gamma'\left(\Omega_\infty^c\right)^{-1}\Gamma\right]^{-1} \\ (\Gamma' W^{-1}\Gamma)^{-1}\mathcal{E}_n' \end{pmatrix}.$$

Motivated by the above approximation, we propose to use the following corrected variance estimator:

$$\begin{aligned}
&\widehat{var}^{\text{adj}}_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c) = \frac{1}{n}\hat{\Xi}_n \\
&= \frac{1}{n}\begin{pmatrix} \left[\hat{\Gamma}'\left[\hat{\Omega}^c(\hat{\theta}_1)\right]^{-1}\hat{\Gamma}\right]^{-1} & \widehat{\mathcal{E}}_n(\hat{\Gamma}' W_n^{-1}\hat{\Gamma})^{-1} \end{pmatrix} \times \begin{pmatrix} \hat{\Gamma}'\left[\hat{\Omega}^c(\hat{\theta}_1)\right]^{-1}\hat{\Gamma} & \hat{\Gamma}' W_n^{-1}\hat{\Gamma} \\ \hat{\Gamma}' W_n^{-1}\hat{\Gamma} & \hat{\Gamma}' W_n^{-1}\hat{\Omega}^c(\hat{\theta}_1) W_n^{-1}\hat{\Gamma} \end{pmatrix} \\
&\times \begin{pmatrix} \left[\hat{\Gamma}'\left[\hat{\Omega}^c(\hat{\theta}_1)\right]^{-1}\hat{\Gamma}'\right]^{-1} \\ (\hat{\Gamma}' W_n^{-1}\hat{\Gamma})^{-1}\widehat{\mathcal{E}}_n' \end{pmatrix} \\
&= \widehat{var}_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c) + \widehat{\mathcal{E}}_n \widehat{var}_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c) + \widehat{var}_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c)\widehat{\mathcal{E}}_n' + \widehat{\mathcal{E}}_n \widehat{var}(\hat{\theta}_1)\widehat{\mathcal{E}}_n',
\end{aligned} \tag{35}$$

where

$$\begin{aligned}
\widehat{\mathcal{E}}_n[.,j] &= \left\{\hat{\Gamma}'\left[\hat{\Omega}^c(\hat{\theta}_1)\right]^{-1}\hat{\Gamma}'\right\}^{-1}\hat{\Gamma}'\left\{\left[\hat{\Omega}^c(\hat{\theta}_1)\right]^{-1}\left.\frac{\partial \hat{\Omega}^c(\theta)}{\partial \theta_j}\right|_{\theta=\hat{\theta}_1}\left[\hat{\Omega}^c(\hat{\theta}_1)\right]^{-1}\right\} g_n(\hat{\theta}_2^c) \text{ and} \\
\hat{\Gamma} &= \hat{\Gamma}(\hat{\theta}_2^c).
\end{aligned}$$

The last three terms in (35), which are of smaller order, serve as a finite sample correction to the original variance estimator.

Windmeijer (2005), too, has used the idea of variance correction, and his proposed correction has been widely implemented in applied work for simple models such as linear IV models and linear dynamic panel data models. However, Windmeijer (2005) considers only an i.i.d. setting, and there are two principal differences between Windmeijer's approach and ours. First, our asymptotic variance estimator involves a centered CCE; in contrast, Windmeijer's involves only a plain variance estimator. Second, we consider the small-G asymptotics; Windmeijer (2005) considers the traditional asymptotics. More broadly, we often have to keep higher-order terms to develop a high order Edgeworth expansion. Here we choose to focus on variance correction instead of distribution correction, which is often the real target behind the Edgeworth expansion. In addition to the technical reasons, a principal reason for our choice is that we have already developed more accurate small-G asymptotic approximations.

With the finite sample corrected variance estimator, we can construct the variance-corrected Wald and t statistics:

$$\begin{aligned} F^{\text{adj}}_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c) &= \frac{1}{p}(R\hat{\theta}_2^c - r)' \left[R\widehat{var}^{\text{adj}}_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c)R'\right]^{-1}(R\hat{\theta}_2^c - r). \\ t^{\text{adj}}_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c) &= \frac{R\hat{\theta}_2^c - r}{\sqrt{R\widehat{var}^{\text{adj}}_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c))R'}}. \end{aligned}$$

Given that the variance correction terms are of smaller order, the variance-corrected statistic will have the same limiting distribution as the original statistic.

Assumption 7 *For each $g = 1, ..., G$ and $j = 1, ..., d$, define $Q_j^g(\theta)$ as*

$$Q_j^g(\theta) = \lim_{L\to\infty} E\left[\frac{1}{L}\sum_{k=1}^{L}\frac{\partial}{\partial\theta'}\left(\frac{\partial f_k^g(\theta)}{\partial\theta_j}\right)\right].$$

Then,

$$\sup_{\theta\in\mathcal{N}(\theta_0)}\left\|\frac{1}{L}\sum_{k=1}^{L}\frac{\partial}{\partial\theta'}\left(\frac{\partial f_k^g(\theta)}{\partial\theta_j}\right) - Q_j^g(\theta)\right\| \xrightarrow{p} 0$$

holds for each $g = 1, ..., G$ and $j = 1, ..., d$, where $\mathcal{N}(\theta_0)$ is an open neighborhood of θ_0, and $\|\cdot\|$ is the Euclidean norm. Also, $Q_j^g(\theta_0) = Q_j(\theta_0)$ for $g = 1, ...G$.

This assumption trivially holds if the moment conditions are linear in parameters.

Theorem 12 *Let Assumptions 1∼7 hold. Then*

$$\begin{aligned} F^{adj}_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c) &= F_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c) + o_p(1) \text{ and} \\ t^{adj}_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c) &= t_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c) + o_p(1). \end{aligned}$$

In the proof of Theorem 12, we show that $\widehat{\mathcal{E}}_n = (1 + o_p(1))\mathcal{E}_{2n}$. That is, the high order correction term has been consistently estimated in a relative sense. This guarantees that $\widehat{\mathcal{E}}_n$ is a reasonable estimator for $\mathcal{E}_{2n}$, which is of order $o_p(1)$.

As a direct implication of Theorem 12 together with Theorem 10, the Wald and t statistics coupled with the J statistic modification and the finite sample variance correction have the standard F and t limiting distributions found in Theorem 10. That is,

$$\tilde{F}^{\text{adj}}_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c) := \frac{G-p-q}{G} \cdot \frac{F^{\text{adj}}_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c)}{1+\frac{1}{G}J(\hat{\theta}_2^c)} \xrightarrow{d} \mathcal{F}_{p,G-p-q} \tag{36}$$

and

$$\tilde{t}^{\text{adj}}_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c) := \sqrt{\frac{G-1-q}{G}} \frac{t^{\text{adj}}_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c)}{\sqrt{1+\frac{1}{G}J(\hat{\theta}_2^c)}} \xrightarrow{d} t_{G-1-q}. \tag{37}$$

For the CU-GEE estimator, we have the following expansion

$$\begin{aligned}
&\sqrt{n}(\hat{\theta}_{\text{CU-GEE}} - \theta_0) \\
&= -\left(\Gamma'\left(\hat{\Omega}^c(\theta_0)\right)^{-1}\Gamma\right)^{-1}\Gamma'\left(\hat{\Omega}^c(\theta_0)\right)^{-1}\sqrt{n}g_n(\theta_0) + \mathcal{E}_{2n}\sqrt{n}(\hat{\theta}_{\text{CU-GEE}} - \theta_0) + o_p(1).
\end{aligned} \tag{38}$$

This can be regarded as a special case of (31) wherein the first-step estimator $\hat{\theta}_1$ is replaced by the CU-GEE estimator. Thus,

$$\sqrt{n}(\hat{\theta}_{\text{CU-GEE}} - \theta_0) \overset{a}{\sim} -(I_d - \mathcal{E}_{2n})^{-1}\left(\Gamma'\left(\hat{\Omega}^c(\theta_0)\right)^{-1}\Gamma\right)^{-1}\Gamma'\left(\hat{\Omega}^c(\theta_0)\right)^{-1}\sqrt{n}g_n(\theta_0), \tag{39}$$

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 (39), the corrected variance estimator for the CU type estimators can be constructed as follows:

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

where

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

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 (14) and (15), respectively. The multiplicative modification provided in Section 4 can then turn the nonstandard distributions into the standard F and t distributions in (36) and (37), respectively.

7 Simulation Evidence

7.1 Design

This section compares the finite sample performance of our new tests by focusing on the following linear dynamic panel data model:

$$y_{it} = \gamma y_{it-1} + x_{1,it}\beta_1 + ... + x_{d-1,it}\beta_3 + \eta_i + u_{it}.$$

The unknown parameter vector is $\theta = (\gamma, \beta_1, ..., \beta_{d-1})' \in \mathbb{R}^d$, and the corresponding covariates are $w_{it} = (y_{it-1}, x_{it})' \in \mathbb{R}^d$ with $x_{it} = (x_{1,it}, ..., x_{d-1,it})' \in \mathbb{R}^{d-1}$. In all our simulation work, we fix the number of parameters d as 4 and set the true value of θ as $\theta_0 = (0.5, 1, 1, 1)'$. We denote $s^g_{it} = (s^g_{1,it}, ..., s^g_{k,it})'$ as any vector valued observations in cluster g, and stack all observations at same period by cluster to define $s_{(g),t} = (s^{g\prime}_{1t}, ..., s^{g\prime}_{Lt})'$. The k-th predetermined regressor $x^g_{k,it}$ are generated according to the following process:

$$x^g_{k,it} = \rho x^g_{k,it-1} + \eta^g_i + \rho u^g_{it-1} + e^g_{k,it},$$

for $k = 1, 2, d-1$, $i = 1, ..., L$, and $t = 1, ..., T$. Setting the number of time periods to be $T = 4$, we characterize the within-cluster dependence in $\eta_{(g)}$, $e_{(g),t}$, and $u_{(g),t}$ by spatial locations that are indexed by a one-dimensional lattice. Define Σ_η and Σ_u to be $L \times L$ matrices whose (i,j)-th elements are $\sigma^\eta_{ij} = \lambda^{|i-j|}$ and $\sigma^u_{ij} = \lambda^{|i-j|}$, respectively, and Σ_e to be a $3L \times 3L$ block diagonal matrix with diagonal matrix $\Sigma_{k,e}$ of size $L \times L$ for $k = 1, ..., d-1$. The (i,j)-th element of $\Sigma_{k,e}$ is $\sigma^e_{k,ij} = \lambda^{|i-j|}$ for $k = 1, ..., d-1$. The parameter λ governs the degree of spatial dependence in each cluster. When $\lambda = 0$, there is no clustered dependence and our model reduces to that of Windmeijer (2005) which considers a dynamic panel data model with only one regressor.

The individual fixed effects and shocks are generated by

$$\begin{gathered} \eta_{(g)} \overset{\text{i.i.d.}}{\sim} N(0, \Sigma_\eta),\ \text{vec}(e_{(g),t}) \overset{\text{i.i.d.}}{\sim} N(0, \Sigma_e), \\ u_{(g),t} = \tau_t \Sigma_u^{1/2} (\delta^g_1 \omega^g_{1t}, , .., \delta^g_L \omega^g_{Lt})', \\ \delta^g_i \overset{\text{i.i.d.}}{\sim} U[0.5, 1.5], \text{ and } \omega^g_{it} \overset{\text{i.i.d.}}{\sim} \chi^2_1 - 1, \end{gathered} \tag{40}$$

over $g = 1, ..., G$, $i = 1, ..., L$, and $t = 1, ..., T$, where $\tau_t = 0.5 + 0.1(t-1)$. The DGP of individual shock $u_{(g),t}$ in (40) features a non-Gaussian process which is heteroskedastic over both time t and individual i. Also, the clustered dependence structure implies

$$\{\eta_{(g)}, \text{vec}(e_{(g),t}), \delta_{(g)}, \omega_{(g),t}\} \perp \{\eta_{(h)}, \text{vec}(e_{(h),s}), \delta_{(h)}, \omega_{(h),s}\},$$

for $g \neq h$ at any t and s.

Before we draw an estimation sample for $t = 1, ..., T$, 50 initial values are generated with $\tau_t = 0.5$ for $t = -49, ..., 0$, $x^g_{k,i,-49} \overset{\text{i.i.d.}}{\sim} N(\eta^g_i/(1-\rho), (1-\rho)^{-1}\Sigma_{k,e})$ for $k = 1, ..., d-1$, and $y^g_{i,-49} = (\sum_{d=1}^3 x_{d,i,-49}\beta_d + \eta^g_i + u^g_{i,-49})/(1-\gamma)$. We fix the values of λ and ρ at 0.60; thus each observation is reasonably persistent with respect to both time and spatial dimensions.[4] The parameters are

[4] When the panel data are persistent with ρ being close to one, the lagged instruments are only weakly correlated with the endogenous changes in the first differenced data, and the GMM inferences considered in our paper can suffer a weak identification problem (e.g., Blundell and Bond, 1998; Stock and Wright, 2000; Bun and Windmeijer, 2010). It will be interesting to extend our approach to develop weak identification robust GMM inferences under clustered dependence, and we leave this as a future research.

estimated by the first differenced GMM (Arellano and Bond estimator). In the supplemental appendix, we describe in details how to implement the GMM inference procedures considered in this section. With all possible lagged instruments $z_{it} = (y_{i0}, ..., y_{it-2}, x'_{i1}, ..., x'_{it-1})', 2 \leq t \leq T$, the number of moment conditions for the Arellano and Bond estimator is $m = dT(T-1)/2$. It could be better to use only a subset of full moment conditions because using this full set of instruments may lead to poor finite sample properties, especially when the number of clusters G is small. Thus, we also employ a reduced set of instruments; that is, we use the most recent lag $z_{it} = (y_{it-2}, x'_{it-1})'$, leading to $d(T-1)$ moment conditions. The initial first-step estimator is chosen by 2SLS with $W_n = n^{-1} \sum_{i=1}^{n} Z_i' Z_i$, where $Z_i = diag(z'_{i2}, ..., z'_{iT})$ is a $(T-1) \times m$ matrix.

7.2 Choice of tests

We focus on the Wald type of tests, as the Monte Carlo results for other types of tests are qualitatively similar. We examine the empirical size of a variety of testing procedures, all of which are based on the first-step or two-step GMM estimators. For the first-step procedures, we consider the unmodified F statistic $F_1 := F_1(\hat{\theta}_1)$ and the degrees-of-freedom modified F statistic $[(G-p)/G] F_1$, where the associated critical values $\chi_p^{1-\alpha}/p$ justified under the large-G asymptotics, and $\mathcal{F}_{p,G-p}^{1-\alpha}$ under the small-G asymptotics, respectively. Note that these two tests have the same size-adjusted power, because the modification only involves a constant multiplier factor.

For the two-step GMM estimation and related tests, we examine the four different procedures that are based on the centered CCE. The first test uses the "plain" F statistic $F_2 := F_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c)$ in (13), where its critical value $\chi_p^{1-\alpha}/p$ is justified by the large-G asymptotics. The second test uses $\tilde{F}_2 := F_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c)$ in (19). Note that

$$\tilde{F}_2 = \frac{(G-p-q)}{G} \cdot \frac{F_2}{1 + \frac{1}{G} J(\hat{\theta}_2^c)}.$$

Compared to the plain two-step GMM F statistic, $\tilde{F}_2$ has the additional J statistic correction factor $(1 + (q/G) J(\hat{\theta}_2^c))^{-1}$. The third test uses the most refined version of the F statistic coupled with the J statistic modification, degrees-of-freedom, and finite sample corrected variance estimator which is defined by

$$\tilde{F}_2^{\text{adj}} := \frac{(G-p-q)}{G} \cdot \frac{F^{\text{adj}}_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c)}{1 + \frac{1}{G} J(\hat{\theta}_2^c)},$$

where $F^{\text{adj}}_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c)$ is defined in (36). These two tests employ the new F critical value $\mathcal{F}_{p,G-p-q}^{1-\alpha}$ which is justified under the small-G asymptotics. Lastly, we consider a bootstrap procedure of the centered two-step GMM test originally proposed by Hall and Horowitz (1996). See the online supplementary appendix for the details about how to implement the bootstrap procedure of Hall and Horowitz (1996) in the presence of clustered dependence. It is important to point out that the consistency and the higher-order refinement of Hall-Horowitz bootstrap procedure require the number of cluster G tends to infinity. This is contrast to the previous two tests that are valid under the small-G asymptotics.

Lastly, we consider the CU types of GMM procedures considered in Sections 5 and 6. For the CU-GEE tests, we implement $F_{\text{CU-GEE}} := F_{\hat{\Omega}^c(\hat{\theta}_{\text{CU-GEE}})}(\hat{\theta}^c_{\text{CU-GEE}})$, $\tilde{F}_{\text{CU-GEE}} := \tilde{F}_{\hat{\Omega}^c(\hat{\theta}_{\text{CU-GEE}})}(\hat{\theta}^c_{\text{CU-GEE}})$,

and $\tilde{F}^{\text{adj}}_{\text{CU-GEE}} := \tilde{F}^{\text{adj}}_{\hat{\Omega}^c(\hat{\theta}_{\text{CU-GEE}})}(\hat{\theta}^c_{\text{CU-GEE}})$ which are constructed similarly to the two step GMM tests. The tests with CU-GMM estimators are also formulated in the same way as the CU-GEE tests.

7.3 Results with balanced cluster size

7.3.1 Size experiment

We consider the case when all clusters have an equal number of individuals and take different values of $G \in \{35, 50, 70, 100\}$, and the number of cluster size $L \in \{50, 100\}$. The null hypotheses of interests are

$$\begin{aligned} H_{01} &: \beta_{10} = 1, \\ H_{02} &: \beta_{10} = \beta_{20} = 1, \\ H_{03} &: \beta_{10} = \beta_{20} = \beta_{30} = 1, \end{aligned}$$

with the corresponding number of joint hypotheses $p = 1, 2$ and 3, respectively, and the significance level is 5%. All of our of simulation results are based on $5,000$ times of Monte Carlo repetition, and the number of bootstrap replication is $1,000$.

Tables 1∼2 report the empirical size of the first-step and two-step tests for different values of $G's$ we consider and $L = 50$. We only report the results when $L = 50$ and provide the results with $L = 100$ in the supplemental appendix, as the qualitative observations for $L = 100$ remain quite similar. The results first indicate that both the first-step and two-step tests based on unmodified statistics F_1 and F_2 suffer from severe size distortions, when the conventional chi-squared critical values are used. For example, with $G = 50$ and $p = 3$, the empirical size of the first-step chi-squared test (using the full set of IVs, and $m = 24$) is 24.5% which becomes more severe, especially, as the number of clusters becomes smaller, for example, 29.9% when G is 35. The empirical sizes of the first-step F test reduce to 21.5% for $G = 35$ when the F critical values are employed. This finding is consistent with the findings in BCH (2011) and Hansen (2007), which highlight the improved finite sample performance of the small-G approximation in the exactly identified models. Tables 1∼2 also indicate that the finite sample size distortion of all tests become less severe as the number of moment conditions decreases or the number of cluster size G increases.

For the two-step tests that employ the plain two-step statistic F_2 with the chi-squared critical values, the empirical sizes are between 23.4%∼65.8% for $m = 24$, and $p = 3$. In view of the large size distortion, we can conclude that the two-step chi-squared test suffers more size distortion than the first-step chi-squared test. This relatively large size distortion reflects the additional cost in estimating the weighting matrix, which is not captured by the chi-square approximation. This motivates us to implement additional corrections via degrees of freedom and the J statistic multiplier coupled with the new critical value $\mathcal{F}^{1-\alpha}_{p,G-p-q}$. Tables 1∼2 show that the additional modifications with the standard F critical value significantly alleviate the distortion. The size distortions in the previous example are reported to be between 4.9% and 6.3% which are much closer to the targeted level 5%. Lastly, we find evidence that the most refined statistic $\tilde{F}^{\text{adj}}_2$, equipped with the finite sample variance correction, results in the empirical sizes between 3.5%∼5.8%. This indicates the most refined two-step F test successfully captures the higher order estimation uncertainty and yields more accurate finite sample size. We find similar conclusions for other values of L, m, and p. Note that the corrected variance estimator is not necessarily

larger than the original estimator in finite samples and in some cases we observe that the smaller value of corrected variance estimate rather deteriorates the finite sample performance of variance-corrected statistics. To avoid this undesirable situation, we may make an extra adjustment to $\widehat{var}^{\text{adj}}_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c)$ so that $\widehat{var}^{\text{adj}}_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c) - \widehat{var}_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c)$ is guaranteed to be positive semidefinite which is in a similar spirit with Politis (2011). The additional adjustment is not implemented in our simulation work, however, as the size distortions of the refined statistic $\tilde{F}_2^{\text{adj}}$ become worse from the unrefined one $\tilde{F}_2$ only up to 0.05%, and the refined tests result in more accurate finite sample sizes than the unrefined ones in most of the cases we consider.

Tables 1∼2 also show the empirical rejection probabilities of the two-step GMM bootstrap procedure by Hall and Horowitz (1996), which is denoted HH-Bootstrap in the tables. We find that the HH-bootstrap is severely undersized when the number of clusters G is small, for example, when G's are 35 and 50 with $m = 24$ and $p = 1$, the empirical sizes are 0% and 1.9%, respectively. This fragility of the HH-Bootstrap procedure has been also observed by Bond and Windmeijer (2005) and Windmeijr (2005) in their Monte Carlo analysis of the cross-sectionally independent dynamic panel data estimated by GMM. They point out that the GMM inferences based on the bootstrap procedures become less reliable when there is a problem in estimating the GMM weighting matrix with the sample moment process. Our simulation results extend their findings in the two-step GMM procedures to those in the presence of clustered dependence. We also note that the empirical rejection probabilities of the GMM bootstrap procedure become close to the nominal size when the reduced set of IV ($m = 12$) is used or the number of cluster G increases.

Next, we report the finite sample performances of the CU-type procedures considered in Sections 5 and 6. The results in Tables 1∼2 indicate that the CU-type GMM inferences under the small G asymptotics clearly outperform those under the large G asymptotics. For instance, with the values of $m = 24$, and $p = 3$, the empirical sizes of the chi-squared test with the plan $F_{\text{CU-GMM}}$ statistic are 23.0% and 81.7% for $G = 100$ and 35, respectively, but the empirical sizes of the most refined CU-GMM F test are 5.2% and 6.8%, respectively, which are very close to the nominal size of 5%. When the number of cluster is small, say $G = 35$, we also find that the CU-type GMM procedures are oversized compared to the first-order equivalent two-step GMM procedures. However, the difference vanishes when the number of cluster become larger. We also find similar conclusions for CU-GEE tests.

Lastly, Tables 1∼2 show that the finite sample size distortions of the (centered) J test, $J^c = J(\hat{\theta}_2^c)$, and the (uncentered) J test, $J = J(\hat{\theta}_2)$, are also substantially reduced and close to the nominal size of 5% when we employ the F critical values and the Beta critical values, respectively, instead of the conventional chi-squared critical values and the GMM bootstrap procedure by Hall and Horowitz (1996).

7.3.2 Power experiment

We investigate the finite sample power performances of the first-step procedure, the two-step procedures F_2, $\tilde{F}_2$, $\tilde{F}_2^{\text{adj}}$, and the CU-type procedures $\tilde{F}^{\text{adj}}_{\text{CU-GEE}}$ and $\tilde{F}^{\text{adj}}_{\text{CU-GMM}}$. We use the finite sample critical values under the null, so the power is size-adjusted and the power comparison is meaningful. The DGPs are the same as before except that the parameters are generated from the local null alternatives $\beta_1 = \beta_{10} + c/\sqrt{n}$ for $c \in [0, 15]$ and $p = 1$. Figures 1∼4 report the power curves for the first-step and two-step tests for $G \in \{35, 50, 70, 100\}$ and $L = 50$. The results first indicate that there is no real difference between power curves of the modified ($\tilde{F}_2$) and unmodified (F_2) two-step tests. In fact, some simulation results not reported here indicate the modified F test can be slightly more powerful as the number of parameters gets larger. Also, the finite sample

corrected test $\tilde{F}_2^{\text{adj}}$ does not lead to a loss of power compared with the uncorrected one $\tilde{F}_2$. We also observe that the CU-GMM tests are less powerful than other types of two-step GMM tests, especially when G is small, but become as powerful as the other ones when G gets larger.

Figures 1∼4 also indicate that the two-step tests are more powerful than the first-step tests in most cases of G, m, and p we consider. The power gain of the two-step GMM procedures becomes more significant as the number of G increases. This can be justified by the asymptotic efficiency of the two-step GMM estimator under the large-G asymptotics. However, under the small-G asymptotics, there is a cost in estimating the CCE weighting matrix, and the power of first-step procedures might dominate the power of the two-step ones when the cost of employing CCE weighting matrix outweighs the benefit of estimating it. In fact, Figures 1 shows that the power of the first-step test can be higher than that of two-step tests when G is small and m is large, say, for example, $G = 35$ and $m = 24$. See Hwang and Sun (2016) who compare these two types of tests in a time series GMM framework by employing more accurate fixed-smoothing asymptotics which are in the same spirit of the small-G asymptotics.

In sum, our simulation evidence clearly demonstrates the size accuracy of our most refined F test regardless of whether the number of clusters G is small or moderate.

7.4 Results with unbalanced cluster size

Although our small-G asymptotics is valid as long as the cluster sizes are approximately equal, we remain wary of the effect of the cluster size heterogeneity on the quality of the small-G approximation. In this subsection, we turn to simulation designs with heterogeneous cluster sizes. Each simulated data set consists of $5,000$. Each simulated data set consists of $5,000$ observations that are divided into 50 clusters. The sequence of alternative cluster-size designs starts by assigning 120 individuals to each of first 10 clusters and 95 individuals to each of next 40 clusters. In each succeeding cluster-size design, we subtract 10 individuals from the second group of clusters and add them to the first group of clusters. In this manner, we construct a series of four cluster-size designs, in which the proportion of the samples in the first group of clusters grows monotonically from 24% to 48%. The design is similar to Carter, Schnepel and Steigerwald (2017) which investigates the behavior of cluster-robust t statistic under cluster heterogeneity. Table 3 describes the heterogeneous cluster-size designs we consider. All other parameter values are the same as before.

Tables 4∼6 report the empirical sizes of the GMM procedures we considered in the previous subsections. The results immediately indicate that the two-step tests suffer from severe size distortion when the conventional chi-squared critical value is employed. For example, under the design II, the empirical size of the "plain" two-step chi-squared test is 56.9% for $m = 24$, and $p = 3$. This size distortions become more severe when the degree of heterogeneity across cluster-size increases. However, our small-G asymptotics still performs very well even with unbalanced cluster sizes as they substantially reduce the empirical sizes. For example, under the design II, the most refined two-step Wald statistic $\tilde{F}_2^{\text{adj}}$ results in the empirical size 6.0% for the above mentioned values of m and p, which is much closer to the nominal size. Similar results for other types of GMM tests are reported in Tables 4∼6. The results of J tests are omitted here as they are qualitatively similar to those of the F tests.

8 Conclusion

This paper studies GMM estimation and inference under clustered dependence. To obtain more accurate asymptotic approximations, we utilize an alternative asymptotics under which the sample size of each cluster is growing, but the number of cluster size G is fixed. The paper is comprehensive in that it covers the first-step GMM, the second-step GMM, and continuously-updating GMM estimators. For the two-step GMM estimator, we show that only if centered moment processes are used in constructing the weighting matrix can we obtain asymptotically pivotal Wald statistic and t statistic. We also find that the centered two-step GMM estimator and CU estimators are all first-order equivalent under the small-G asymptotics. With the help of the standard J statistic, the Wald statistic and t statistic based on these estimators can be modified to have to standard F and t limiting distributions. A finite sample variance correction is suggested to further improve the performance of the asymptotic F and t tests. The advantages of our procedures are clearly reflected in finite samples as demonstrated by our simulation study and empirical application.

In an overidentified GMM model, the set of moment conditions can be divided into two blocks: the moment conditions that are for identifying unknown parameters, and the rest of ones for improving the efficiency of the GMM estimator. We expect that the spatial dependence between these two blocks of moment conditions is the key information to assess the relative power performance of first-step and two-step tests. Recently, Hwang and Sun (2016) compare these two types of tests by employing more accurate asymptotic approximations in a time series GMM framework. We leave the extension of this analysis to a spatial setting to future research.

References

[1] Arellano, M. (1987): "Computing Robust Standard Errors for Within-Group Estimators." *Oxford Bulletin of Economics and Statistics,* 49, 431-434.

[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] Bertrand, M., Duflo, E., and Mullainathan, S. (2004): "How Much Should We Trust Differences-in-Differences Estimates?" *The Quarterly Journal of Economics*, 119(1), 249-275.

[4] Bester, C. A., Conley, T. G., and Hansen, C. B. (2011): "Inference with Dependent Data Using Cluster Covariance Estimators." *Journal of Econometrics* 165(2), 137-151.

[5] Bilodeau, M., and Brenner, D. (2008): "Theory of multivariate statistics." Springer Science & Business Media.

[6] Bond, S., and Windmeijer, F. (2005). Reliable inference for GMM estimators? Finite sample properties of alternative test procedures in linear panel data models. *Econometric Reviews*, 24(1), 1-37.

[7] Blundell, R., and Bond, S. (1998): "Initial conditions and moment restrictions in dynamic panel data models." *Journal of Econometrics*, 87(1), 115-143.

[8] Bun, M. J., and Windmeijer, F. (2010): "The weak instrument problem of the system GMM estimator in dynamic panel data models." *The Econometrics Journal*, 13(1), 95-126.

[9] Canay, I. A., Romano, J. P., and Shaikh, A. M. (2017). Randomization tests under an approximate symmetry assumption. Econometrica, 85(3), 1013-1030.

[10] Cameron, A. C., Gelbach, J. B., and Miller, D. L. (2008): "Bootstrap-based improvements for inference with clustered errors." *Review of Economics and Statistics*, 90(3), 414-427.

[11] Cameron, A. C. and Miller, D. L. (2015): "A practitioner's guide to cluster-robust inference." *Journal of Human Resources*, 50(2), 317-372.

[12] Carter, A. V., Schnepel, K. T. , and Steigerwald, D. G. (2017): "Asymptotic behavior of at test robust to cluster heterogeneity." *Review of Economics and Statistics, Forthcoming.*

[13] 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.

[14] Hall, A. R. (2000): "Covariance matrix estimation and the power of the overidentifying restrictions test." *Econometrica*, 68(6), 1517-1527.

[15] 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.

[16] Hansen, C. B. (2007): "Asymptotic properties of a robust variance matrix estimator for panel data when T is large." *Journal of Econometrics*, 141(2), 597-620.

[17] Hansen, L. P. (1982): "Large Sample Properties of Generalized Method of Moments Estimators." *Econometrica* 50, 1029-1054.

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

[19] Hayakawa, K. (2016): "On the effect of weighting matrix in GMM specification test." *Journal of Statistical Planning and Inference*, 178, 84-98.

[20] Hwang, J. and Y. Sun (2016): "Should We Go One Step Further? An Accurate Comparison of One-Step and Two-Step Procedures in a Generalized Method of Moments Framework" Working paper, Department of Economics, UC San Diego.

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

[22] Hwang, J., and Sun, Y. (2017b): "Supplementary Appendix to 'Asymptotic F and t Tests in an Efficient GMM Setting'." *Journal of Econometrics Online Supplementary Appendix*, http://dx.doi.org/10.1016/j.jeconom.2017.02.003.

[23] Ibragimov, R., and Müller, U. K. (2010): "t-Statistic based correlation and heterogeneity robust inference." *Journal of Business & Economic Statistics*, 28(4), 453-468.

[24] Ibragimov, R. and Müller, U.K. (2016): "Inference with few heterogeneous clusters." *Review of Economics and Statistics*, 98(1), 83-96.

[25] Imbens, G. W., and Kolesar, M. (2016): "Robust standard errors in small samples: Some practical advice." *Review of Economics and Statistics*, 98(4), 701-712.

[26] 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), 86-98.

[27] Jenish, N., and Prucha, I. R. (2012): "On spatial processes and asymptotic inference under near-epoch dependence." *Journal of econometrics*, 170(1), 178-190.

[28] 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.

[29] Kiefer, N. M., Vogelsang, T. J. and Bunzel, H. (2002): "Simple Robust Testing of Regression Hypotheses" *Econometrica* 68 (3), 695-714.

[30] Kiefer, N. M. and Vogelsang, T. J. (2005): "A New Asymptotic Theory for Heteroskedasticity-Autocorrelation Robust Tests." *Econometric Theory* 21, 1130-1164.

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

[32] MacKinnon, J. G., and Webb M. D. (2017): "Wild bootstrap inference for wildly different cluster sizes." *Journal of Applied Econometrics* 32, 233-254.

[33] Müller, U. K. (2007): "A theory of robust long-run variance estimation." *Journal of Econometrics*, 141(2), 1331-1352.

[34] Newey, W. K., and Smith, R. J. (2004): "Higher order properties of GMM and generalized empirical likelihood estimators." *Econometrica* 72, no. 1 : 219-255.

[35] Politis, D. N. (2011): "Higher-order accurate, positive semidefinite estimation of large-sample covariance and spectral density matrices." *Econometric Theory*, 27(04), 703-744.

[36] Stock, J. H., and Wright, J. H. (2000): "GMM with weak identification. Econometrica.", 68(5), 1055-1096.

[37] Stock, J. H. and Watson, M. W. (2008): "Heteroskedasticity-Robust Standard Errors for Fixed Effects Panel Data Regression." *Econometrica*, 76: 155-174.

[38] Sun, Y. (2013): "A Heteroskedasticity and Autocorrelation Robust F Test Using Orthonormal Series Variance Estimator." *Econometrics Journal* 16(1), 1–26.

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

[40] Sun, Y., and Kim, M. S. (2012): "Simple and powerful GMM over-identification tests with accurate size." *Journal of Econometrics*, 166(2), 267-281.

[41] Sun, Y., and Kim, M. S. (2015): "Asymptotic F-Test in a GMM Framework with Cross-Sectional Dependence." *Review of Economics and Statistics*, 97(1), 210-223.

[42] Sun, Y., Phillips, P. C. B. and Jin, S. (2008): "Optimal Bandwidth Selection in Heteroskedasticity-Autocorrelation Robust Testing." *Econometrica* 76(1), 175–94.

[43] White, H. (1984), "Asymptotic Theory for Econometricians" (San Diego: Academic Press).

[44] Windmeijer, F. (2005): "A finite sample correction for the variance of linear efficient two-step GMM estimators." *Journal of Econometrics*, 126(1), 25-51.

[45] Wooldridge, J. M. (2003): "Cluster-sample methods in applied econometrics." *American Economic Review,* 133-138.

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

Table 1: Empirical size of GMM tests based on the centered CCE when the number of clusters $G = 35, 50$, the number of population within cluster $L = 50$, the number of joint hypothesis $p = 1 \sim 3$, and the number of moment conditions $m = 12, 24$, with $T = 4$.

$G = 35$	Test statistic	Critical values	$m = 24$			$m = 12$		
			$p = 1$	$p = 2$	$p = 3$	$p = 1$	$p = 2$	$p = 3$
First-step	F_1	$\chi^{1-\alpha}_p/p$	0.286	0.295	0.299	0.210	0.206	0.214
	$\frac{G-p}{G}F_1$	$\mathcal{F}^{1-\alpha}_{p,G-p}$	0.250	0.238	0.215	0.183	0.159	0.139
Two-step	F_2	$\chi^{1-\alpha}_p/p$	0.399	0.534	0.658	0.167	0.207	0.272
	$\tilde{F}_2$	$\mathcal{F}^{1-\alpha}_{p,G-p-q}$	0.059	0.049	0.049	0.062	0.059	0.057
	$\tilde{F}_2^{\text{adj}}$	$\mathcal{F}^{1-\alpha}_{p,G-p-q}$	0.042	0.033	0.035	0.050	0.050	0.055
	F_2	HH-Bootstrap	0.000	0.000	0.000	0.036	0.024	0.014
CU-type	$F_{\text{CU-GEE}}$	$\chi^{1-\alpha}_p/p$	0.501	0.691	0.805	0.204	0.261	0.330
	$\tilde{F}_{\text{CU-GEE}}$	$\mathcal{F}^{1-\alpha}_{p,G-p-q}$	0.080	0.070	0.070	0.075	0.070	0.066
	$\tilde{F}^{\text{adj}}_{\text{CU-GEE}}$	$\mathcal{F}^{1-\alpha}_{p,G-p-q}$	0.083	0.071	0.070	0.060	0.058	0.058
	$F_{\text{CU-GMM}}$	$\chi^{1-\alpha}_p/p$	0.522	0.709	0.817	0.200	0.252	0.318
	$\tilde{F}_{\text{CU-GMM}}$	$\mathcal{F}^{1-\alpha}_{p,G-p-q}$	0.083	0.076	0.072	0.079	0.071	0.070
	$\tilde{F}^{\text{adj}}_{\text{CU-GMM}}$	$\mathcal{F}^{1-\alpha}_{p,G-p-q}$	0.082	0.070	0.068	0.059	0.056	0.058
J test	J	$\chi^{1-\alpha}_q$	–	0.002	–	–	0.035	–
	$\frac{1}{G}J$	$\mathcal{B}^{1-\alpha}_{q/2,(G-q)/2}$	–	0.113	–	–	0.058	–
	J^c	$\chi^{1-\alpha}_q$	–	0.816	–	–	0.209	–
	$\frac{G-q}{Gq}J^c$	$\mathcal{F}^{1-\alpha}_{q,G-q}$	–	0.058	–	–	0.050	–
	J^c	HH-Bootstrap	–	0.743	–	–	0.124	–
$G = 50$	Test statistic	Critical values	$m = 24$			$m = 12$		
			$p = 1$	$p = 2$	$p = 3$	$p = 1$	$p = 2$	$p = 3$
First-step	F_1	$\chi^{1-\alpha}_p/p$	0.242	0.242	0.245	0.189	0.181	0.175
	$\frac{G-p}{G}F_1$	$F^{1-\alpha}_{p,G-p}$	0.222	0.207	0.187	0.173	0.146	0.133
Two-step	F_2	$\chi^{1-\alpha}_p/p$	0.308	0.437	0.540	0.141	0.174	0.211
	$\tilde{F}_2$	$F^{1-\alpha}_{p,G-p-q}$	0.072	0.070	0.067	0.066	0.061	0.061
	$\tilde{F}_2^{\text{adj}}$	$F^{1-\alpha}_{p,G-p-q}$	0.062	0.060	0.058	0.048	0.049	0.052
	F_2	HH-Bootstrap	0.019	0.004	0.001	0.052	0.039	0.033
CU-type	$F_{\text{CU-GEE}}$	$\chi^{1-\alpha}_p/p$	0.315	0.443	0.548	0.152	0.182	0.218
	$\tilde{F}_{\text{CU-GEE}}$	$F^{1-\alpha}_{p,G-p-q}$	0.080	0.082	0.077	0.071	0.066	0.067
	$\tilde{F}^{\text{adj}}_{\text{CU-GEE}}$	$F^{1-\alpha}_{p,G-p-q}$	0.070	0.069	0.067	0.055	0.050	0.052
	$F_{\text{CU-GMM}}$	$\chi^{1-\alpha}_p/p$	0.333	0.461	0.561	0.142	0.175	0.211
	$\tilde{F}_{\text{CU-GMM}}$	$F^{1-\alpha}_{p,G-p-q}$	0.087	0.084	0.083	0.067	0.059	0.061
	$\tilde{F}^{\text{adj}}_{\text{CU-GMM}}$	$F^{1-\alpha}_{p,G-p-q}$	0.064	0.068	0.066	0.051	0.048	0.051
J test	J	$\chi^{1-\alpha}_q/q$	–	0.015	–	–	0.040	–
	$\frac{1}{G}J$	$B^{1-\alpha}_{q/2,(G-q)/2}$	–	0.069	–	–	0.058	–
	J^c	$\chi^{1-\alpha}_q/q$	–	0.561	–	–	0.152	–
	$\frac{G-q}{Gq}J^c$	$F^{1-\alpha}_{q,G-q}$	–	0.054	–	–	0.055	–
	J^c	HH-Bootstrap	–	0.284	–	–	0.119	–

Notes: The first-step tests are based on the first-step GMM estimator $\hat{\theta}_1$ with the associated F statistic $F_1 = F_1(\hat{\theta}_1)$. The J tests employ the statistics $J = J(\hat{\theta}_2)$ and $J^c = J(\hat{\theta}_2^c)$ with or without degree of freedom (d.f.) correction. All two-step tests are based on the centered two-step GMM estimator $\hat{\theta}_2^c$ but use different test statistics : the unmodified $F_2 = F_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c)$, J statistic and d.f. corrected $\tilde{F}_2 = \tilde{F}_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c)$, and J statistic, d.f., and finite-sample-variance corrected $\tilde{F}_2^{\text{adj}} = \tilde{F}^{\text{adj}}_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c)$. The test statistics with CU-type GMM estimators are constructed similarly.

Table 2: Empirical size of GMM tests based on the centered CCE when $L = 50$, number of clusters $G = 70, 100$, number of joint hypothesis $p = 1 \sim 3$ and number of moment conditions $m = 12, 24$, with $T = 4$.

$G = 70$	Test statistics	Critical values	$m = 24$			$m = 12$		
			$p = 1$	$p = 2$	$p = 3$	$p = 1$	$p = 2$	$p = 3$
First-step	F_1	$\chi^{1-\alpha}_p/p$	0.191	0.191	0.186	0.150	0.143	0.140
	$\frac{G-p}{G}F_1$	$F^{1-\alpha}_{p,G-p}$	0.180	0.166	0.154	0.140	0.125	0.111
Two-step	F_2	$\chi^{1-\alpha}_p/p$	0.218	0.281	0.340	0.102	0.122	0.141
	$\tilde{F}_2$	$F^{1-\alpha}_{p,G-p-q}$	0.076	0.067	0.063	0.055	0.056	0.055
	$\tilde{F}_2^{\text{adj}}$	$F^{1-\alpha}_{p,G-p-q}$	0.068	0.059	0.057	0.048	0.051	0.049
	F_2	HH-Bootstrap	0.045	0.029	0.017	0.046	0.044	0.041
CU-type	$F_{\text{CU-GEE}}$	$\chi^{1-\alpha}_p/p$	0.217	0.286	0.342	0.107	0.127	0.144
	$\tilde{F}_{\text{CU-GEE}}$	$F^{1-\alpha}_{p,G-p-q}$	0.080	0.072	0.068	0.059	0.057	0.057
	$\tilde{F}^{\text{adj}}_{\text{CU-GEE}}$	$F^{1-\alpha}_{p,G-p-q}$	0.069	0.063	0.059	0.050	0.052	0.047
	$F_{\text{CU-GMM}}$	$\chi^{1-\alpha}_p/p$	0.217	0.280	0.337	0.097	0.116	0.135
	$\tilde{F}_{\text{CU-GMM}}$	$F^{1-\alpha}_{p,G-p-q}$	0.074	0.071	0.062	0.052	0.051	0.049
	$\tilde{F}^{\text{adj}}_{\text{CU-GMM}}$	$F^{1-\alpha}_{p,G-p-q}$	0.060	0.058	0.053	0.042	0.047	0.044
J test	J	$\chi^{1-\alpha}_q/q$	–	0.026	–	–	0.044	–
	$\frac{1}{G}J$	$B^{1-\alpha}_{q/2,(G-q)/2}$	–	0.060	–	–	0.055	–
	J^c	$\chi^{1-\alpha}_q/q$	–	0.364	–	–	0.121	–
	$\frac{G-q}{Gq}J^c$	$F^{1-\alpha}_{q,G-q}$	–	0.055	–	–	0.054	–
	J^c	HH-Bootstrap	–	0.225	–	–	0.103	–

$G = 100$	Test statistics	Critical values	$m = 24$			$m = 12$		
			$p = 1$	$p = 2$	$p = 3$	$p = 1$	$p = 2$	$p = 3$
First-step	F_1	$\chi^{1-\alpha}_p/p$	0.163	0.153	0.151	0.133	0.131	0.127
	$\frac{G-p}{G}F_1$	$F^{1-\alpha}_{p,G-p}$	0.155	0.140	0.127	0.128	0.118	0.109
Two-step	F_2	$\chi^{1-\alpha}_p/p$	0.159	0.197	0.234	0.097	0.109	0.116
	$\tilde{F}_2$	$F^{1-\alpha}_{p,G-p-q}$	0.072	0.070	0.063	0.065	0.061	0.056
	$\tilde{F}_2^{\text{adj}}$	$F^{1-\alpha}_{p,G-p-q}$	0.068	0.064	0.058	0.057	0.056	0.051
	F_2	HH-Bootstrap	0.055	0.041	0.034	0.058	0.053	0.046
CU-type	$F_{\text{CU-GEE}}$	$\chi^{1-\alpha}_p/p$	0.161	0.201	0.238	0.099	0.112	0.117
	$\tilde{F}_{\text{CU-GEE}}$	$F^{1-\alpha}_{p,G-p-q}$	0.073	0.070	0.064	0.067	0.062	0.058
	$\tilde{F}^{\text{adj}}_{\text{CU-GEE}}$	$F^{1-\alpha}_{p,G-p-q}$	0.068	0.065	0.059	0.056	0.057	0.051
	$F_{\text{CU-GMM}}$	$\chi^{1-\alpha}_p/p$	0.151	0.191	0.230	0.091	0.098	0.107
	$\tilde{F}_{\text{CU-GMM}}$	$F^{1-\alpha}_{p,G-p-q}$	0.070	0.063	0.056	0.056	0.056	0.050
	$\tilde{F}^{\text{adj}}_{\text{CU-GMM}}$	$F^{1-\alpha}_{p,G-p-q}$	0.063	0.059	0.052	0.050	0.051	0.048
J test	J	$\chi^{1-\alpha}_q/q$	–	0.030	–	–	0.047	–
	$\frac{q}{G}J$	$B^{1-\alpha}_{q/2,(G-q)/2}$	–	0.051	–	–	0.055	–
	J^c	$\chi^{1-\alpha}_q/q$	–	0.248	–	–	0.098	–
	$\frac{G-q}{G}J^c$	$F^{1-\alpha}_{q,G-q}$	–	0.048	–	–	0.054	–
	J^c	HH-Bootstrap	–	0.186	–	–	0.092	–

See footnote to Table 1.

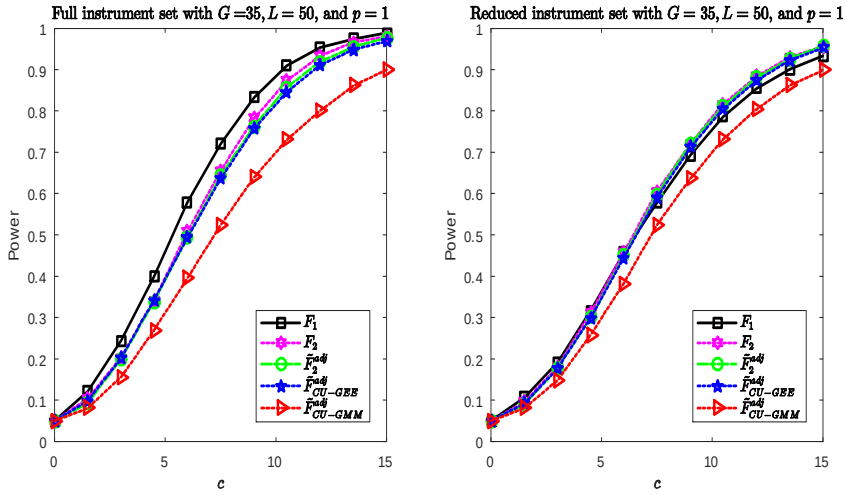

Figure 1: Size-adjusted power of the first-step (2SLS) and two-step tests with G=35 and L=50.

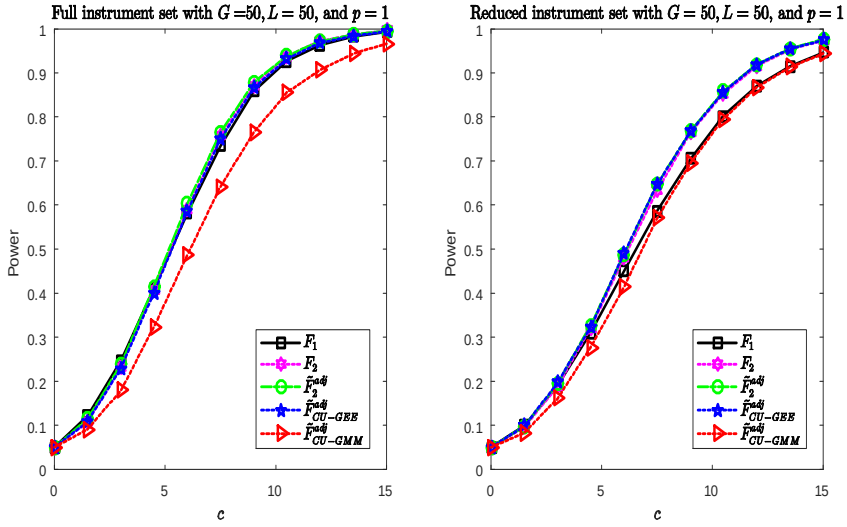

Figure 2: Size-adjusted power of the first-step (2SLS) and two-step tests with $G = 50$ and $L = 50$.

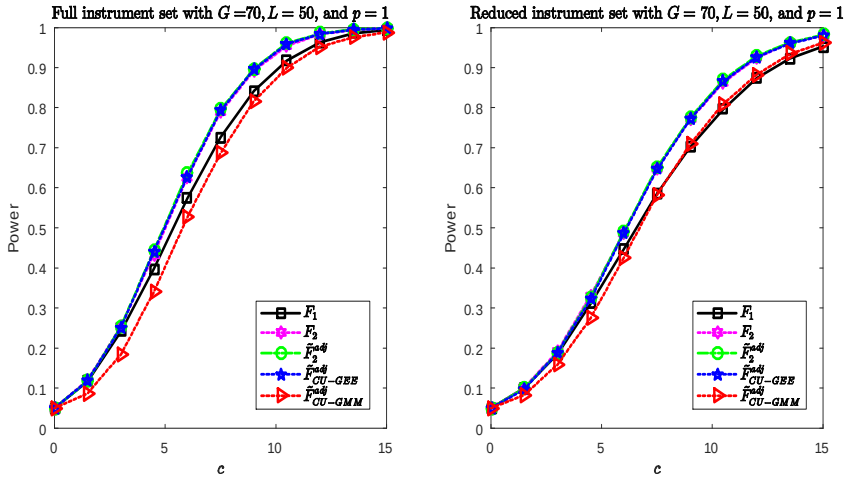

Figure 3: Size-adjusted power of the first-step (2SLS) and two-step tests with $G = 50$ and $L = 50$.

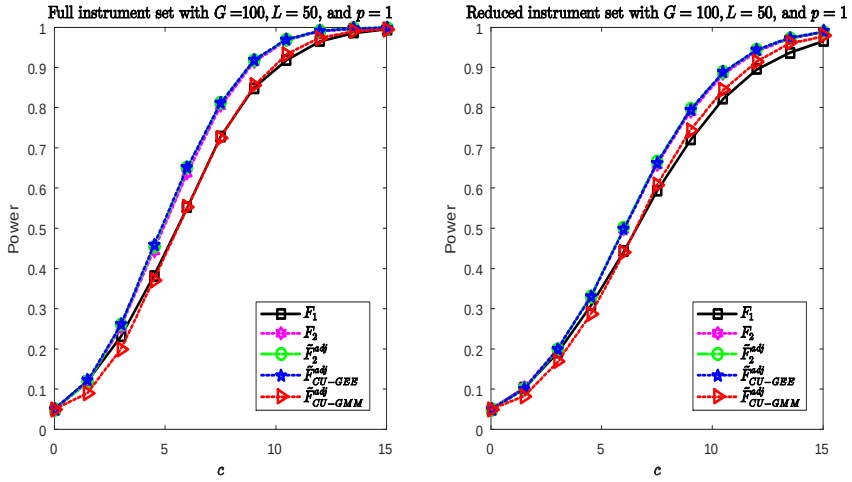

Figure 4: Size-adjusted power of the first-step (2SLS) and two-step tests with $G = 100$ and $L = 50$.

Table 3: Design of heterogeneity in cluster size

$G = 50$	$L_1 = ... = L_{10}$	$L_{11} = ... = L_{50}$	n
Design I	120	95	5000
Design II	160	85	5000
Design III	200	75	5000
Design IV	240	65	5000

Table 4: Empirical size of first-step and two-step tests based on the centered CCE when there is a heterogeneity in cluster size: Design I

			Design I					
	Test statistic	Critical values	$m = 24$			$m = 12$		
			$p = 1$	$p = 2$	$p = 3$	$p = 1$	$p = 2$	$p = 3$
First-step	F_1	$\chi^{1-\alpha}_p/p$	0.175	0.182	0.188	0.143	0.149	0.152
	$\frac{G-p}{G}F_1$	$\mathcal{F}^{1-\alpha}_{p,G-p}$	0.158	0.153	0.137	0.130	0.120	0.112
Two-step	F_2	$\chi^{1-\alpha}_p/p$	0.306	0.430	0.532	0.132	0.174	0.216
	$\tilde{F}_2$	$\mathcal{F}^{1-\alpha}_{p,G-p-q}$	0.067	0.068	0.070	0.057	0.064	0.063
	$\tilde{F}^{\text{adj}}_2$	$\mathcal{F}^{1-\alpha}_{p,G-p-q}$	0.051	0.051	0.048	0.047	0.051	0.050
	F_2	HH-Bootstrap	0.016	0.005	0.001	0.039	0.032	0.028
CU-type	$F_{\text{CU-GEE}}$	$\chi^{1-\alpha}_p/p$	0.297	0.425	0.529	0.126	0.169	0.202
	$\tilde{F}_{\text{CU-GEE}}$	$\mathcal{F}^{1-\alpha}_{p,G-p-q}$	0.067	0.068	0.067	0.055	0.060	0.059
	$\tilde{F}^{\text{adj}}_{\text{CU-GEE}}$	$\mathcal{F}^{1-\alpha}_{p,G-p-q}$	0.052	0.052	0.048	0.046	0.049	0.048
	$F_{\text{CU-GMM}}$	$\chi^{1-\alpha}_p/p$	0.306	0.426	0.527	0.118	0.157	0.195
	$\tilde{F}_{\text{CU-GEE}}$	$\mathcal{F}^{1-\alpha}_{p,G-p-q}$	0.068	0.065	0.063	0.053	0.056	0.056
	$\tilde{F}^{\text{adj}}_{\text{CU-GMM}}$	$\mathcal{F}^{1-\alpha}_{p,G-p-q}$	0.053	0.049	0.047	0.044	0.046	0.046
J test	J	$\chi^{1-\alpha}_q/q$	–	0.013	–	–	0.052	–
	$\frac{1}{G}J$	$\mathcal{B}^{1-\alpha}_{q/2,(G-q)/2}$	–	0.062	–	–	0.068	–
	J^c	$\chi^{1-\alpha}_q/q$	–	0.564	–	–	0.153	–
	$\frac{G-q}{Gq}J^c$	$\mathcal{F}^{1-\alpha}_{q,G-q}$	–	0.052	–	–	0.051	–
	J^c	HH-Bootstrap	–	0.285	–	–	0.103	–

See footnote to Table 1.

Table 5: Empirical size of first-step and two-step tests based on the centered CCE when there is a heterogeneity in cluster size: Designs II and III

	Test statistic	Critical values	$m=24$			$m=12$		
			$p=1$	$p=2$	$p=3$	$p=1$	$p=2$	$p=3$
Design II								
First-step	F_1	$\chi^{1-\alpha}_p/p$	0.166	0.171	0.173	0.148	0.145	0.153
	$\frac{G-p}{G}F_1$	$\mathcal{F}^{1-\alpha}_{p,G-p}$	0.151	0.140	0.134	0.133	0.117	0.111
Two-step	F_2	$\chi^{1-\alpha}_p/p$	0.326	0.455	0.569	0.135	0.185	0.231
	$\tilde{F}_2$	$\mathcal{F}^{1-\alpha}_{p,G-p-q}$	0.078	0.082	0.081	0.065	0.069	0.069
	$\tilde{F}_2^{\text{adj}}$	$\mathcal{F}^{1-\alpha}_{p,G-p-q}$	0.058	0.059	0.060	0.052	0.055	0.055
	F_2	HH-Bootstrap	0.017	0.005	0.002	0.039	0.031	0.028
CU-type	$F_{\text{CU-GEE}}$	$\chi^{1-\alpha}_p/p$	0.319	0.448	0.563	0.131	0.177	0.216
	$\tilde{F}_{\text{CU-GEE}}$	$\mathcal{F}^{1-\alpha}_{p,G-p-q}$	0.077	0.076	0.076	0.064	0.062	0.064
	$\tilde{F}^{\text{adj}}_{\text{CU-GEE}}$	$\mathcal{F}^{1-\alpha}_{p,G-p-q}$	0.061	0.057	0.061	0.051	0.050	0.054
	$F_{\text{CU-GEE}}$	$\chi^{1-\alpha}_p/p$	0.333	0.457	0.567	0.127	0.172	0.209
	$\tilde{F}_{\text{CU-GEE}}$	$\mathcal{F}^{1-\alpha}_{p,G-p-q}$	0.081	0.080	0.080	0.060	0.060	0.062
	$\tilde{F}^{\text{adj}}_{\text{CU-GMM}}$	$\mathcal{F}^{1-\alpha}_{p,G-p-q}$	0.066	0.060	0.063	0.050	0.049	0.053
J test	J	$\chi^{1-\alpha}_q/q$	–	0.013	–	–	0.045	–
	$\frac{1}{G}J$	$\mathcal{B}^{1-\alpha}_{q/2,(G-q)/2}$	–	0.060	–	–	0.064	–
	J^c	$\chi^{1-\alpha}_q/q$	–	0.616	−−	–	0.163	–
	$\frac{G-q}{Gq}J^c$	$\mathcal{F}^{1-\alpha}_{q,G-q}$	–	0.074	–	–	0.057	−−
	J^c	HH-Bootstrap	–	0.310	–	–	0.098	–
Design III								
	Test statistic	Critical values	$m=24$			$m=12$		
			$p=1$	$p=2$	$p=3$	$p=1$	$p=2$	$p=3$
First-step	F_1	$\chi^{1-\alpha}_p/p$	0.168	0.178	0.187	0.148	0.149	0.158
	$\frac{G-p}{G}F_1$	$\mathcal{F}^{1-\alpha}_{p,G-p}$	0.156	0.148	0.139	0.133	0.122	0.117
Two-step	F_2	$\chi^{1-\alpha}_p/p$	0.340	0.492	0.603	0.155	0.197	0.247
	$\tilde{F}_2$	$\mathcal{F}^{1-\alpha}_{p,G-p-q}$	0.085	0.089	0.092	0.074	0.077	0.076
	$\tilde{F}_2^{\text{adj}}$	$\mathcal{F}^{1-\alpha}_{p,G-p-q}$	0.063	0.065	0.067	0.057	0.060	0.060
	F_2	HH-Bootstrap	0.014	0.003	0.001	0.037	0.029	0.025
CU-type	$F_{\text{CU-GEE}}$	$\chi^{1-\alpha}_p/p$	0.334	0.484	0.594	0.150	0.192	0.240
	$\tilde{F}_{\text{CU-GEE}}$	$\mathcal{F}^{1-\alpha}_{p,G-p-q}$	0.082	0.086	0.090	0.069	0.075	0.071
	$\tilde{F}^{\text{adj}}_{\text{CU-GEE}}$	$\mathcal{F}^{1-\alpha}_{p,G-p-q}$	0.065	0.062	0.066	0.055	0.057	0.058
	$\tilde{F}_{\text{CU-GMM}}$	$\chi^{1-\alpha}_p/p$	0.334	0.484	0.592	0.143	0.188	0.236
	$\tilde{F}_{\text{CU-GEE}}$	$\mathcal{F}^{1-\alpha}_{p,G-p-q}$	0.081	0.087	0.090	0.067	0.068	0.068
	$\tilde{F}^{\text{adj}}_{\text{CU-GMM}}$	$\mathcal{F}^{1-\alpha}_{p,G-p-q}$	0.063	0.064	0.071	0.052	0.053	0.055
J test	J	$\mathcal{B}^{1-\alpha}_{q/2,(G-q)/2}$	–	0.010	–	–	0.046	–
	$\frac{1}{G}J$	$\chi^{1-\alpha}_q/q$	–	0.055	–	–	&0.064	–
	J^c	$\chi^{1-\alpha}_q/q$	–	0.672	–	–	0.182	–
	$\frac{G-q}{Gq}J^c$	$\mathcal{F}^{1-\alpha}_{q,G-q}$	–	0.108	–	–	0.069	–
	J^c	HH-Bootstrap	–	0.348	–	–	0.103	–

See footnote to Table 1.

Table 6: Empirical size of first-step and two-step tests based on the centered CCE when there is a heterogeneity in cluster size: Design IV

			Design IV					
			$p=1$	$p=2$	$p=3$	$p=1$	$p=2$	$p=3$
First-step	F_1	$\chi^{1-\alpha}_p/p$	0.181	0.183	0.200	0.157	0.159	0.175
	$\frac{G-p}{G}F_1$	$\mathcal{F}^{1-\alpha}_{p,G-p}$	0.165	0.155	0.152	0.140	0.133	0.130
Two-step	F_2	$\chi^{1-\alpha}_p/p$	0.383	0.525	0.653	0.172	0.236	0.297
	$\tilde{F}_2$	$\mathcal{F}^{1-\alpha}_{p,G-p-q}$	0.102	0.105	0.116	0.093	0.093	0.103
	$\tilde{F}_2^{\text{adj}}$	$\mathcal{F}^{1-\alpha}_{p,G-p-q}$	0.077	0.079	0.082	0.073	0.076	0.081
	F_2	HH-Bootstrap	0.013	0.004	0.001	0.038	0.031	0.024
CU-type	$F_{\text{CU-GEE}}$	$\chi^{1-\alpha}_p/p$	0.378	0.518	0.639	0.168	0.226	0.288
	$\tilde{F}_{\text{CU-GEE}}$	$\mathcal{F}^{1-\alpha}_{p,G-p-q}$	0.099	0.105	0.113	0.088	0.087	0.097
	$\tilde{F}^{\text{adj}}_{\text{CU-GEE}}$	$\mathcal{F}^{1-\alpha}_{p,G-p-q}$	0.077	0.077	0.082	0.072	0.072	0.077
	$F_{\text{CU-GEE}}$	$\chi^{1-\alpha}_p/p$	0.384	0.528	0.647	0.173	0.222	0.284
	$\tilde{F}_{\text{CU-GEE}}$	$\mathcal{F}^{1-\alpha}_{p,G-p-q}$	0.099	0.111	0.112	0.081	0.084	0.093
	$\tilde{F}^{\text{adj}}_{\text{CU-GMM}}$	$\mathcal{F}^{1-\alpha}_{p,G-p-q}$	0.080	0.083	0.085	0.066	0.067	0.073
J test	J	$\chi^{1-\alpha}_q/q$	–	0.011	–	–	0.040	–
	$\frac{1}{G}J$	$\mathcal{B}^{1-\alpha}_{q/2,(G-q)/2}$	–	0.052	–	–	0.060	–
	J^c	$\chi^{1-\alpha}_q/q$	–	0.754	–	–	0.219	–
	$\frac{G-q}{Gq}J^c$	$\mathcal{F}^{1-\alpha}_{q,G-q}$	–	0.163	–	–	0.091	–
	J^c	HH-Bootstrap	–	0.401	–	–	0.108	–

See footnote to Table 1

9 Appendix of Proofs

Proof of Proposition 1. Part (a). For each $g = 1, ..., G$,

$$\frac{1}{\sqrt{L}}\sum_{k=1}^{L} f_k^g(\hat{\theta}_1) = \frac{1}{\sqrt{L}}\sum_{k=1}^{L}\left\{ f_k^g(\theta_0) + \frac{\partial f_k^g(\tilde{\theta}^*)}{\partial \theta'}(\hat{\theta}_1 - \theta_0)\right\},$$

where $\tilde{\theta}^*$ is between $\hat{\theta}_1$ and θ_0. Here, $\tilde{\theta}^*$ may be different for different rows of $\partial f_k^g(\tilde{\theta}^*)/\partial\theta'$. For notational simplicity, we do not make this explicit. By Assumptions 2 and 5, we have

$$\frac{1}{\sqrt{L}}\sum_{k=1}^{L} f_k^g(\hat{\theta}) = \frac{1}{\sqrt{L}}\sum_{k=1}^{L} f_k^g(\theta_0) - \frac{1}{L}\sum_{k=1}^{L}\frac{\partial f_k^g(\tilde{\theta}^*)}{\partial \theta'}(\Gamma' W^{-1}\Gamma)^{-1}\Gamma' W^{-1}\frac{1}{G}\sum_{\tilde{g}=1}^{G}\left(\frac{1}{\sqrt{L}}\sum_{\tilde{k}=1}^{L} f_{\tilde{k}}^{\tilde{g}}(\theta_0)\right) + o_p(1)$$

$$= \frac{1}{\sqrt{L}}\sum_{k=1}^{L} f_k^g(\theta_0) - \Gamma_g(\Gamma' W^{-1}\Gamma)^{-1}\Gamma' W^{-1}\frac{1}{G}\sum_{\tilde{g}=1}^{G}\left(\frac{1}{\sqrt{L}}\sum_{\tilde{k}=1}^{L} f_{\tilde{k}}^{\tilde{g}}(\theta_0)\right) + o_p(1). \tag{41}$$

Using Assumptions 4–6, we then have

$$\begin{aligned}\frac{1}{\sqrt{L}}\sum_{k=1}^{L} f_k^g(\hat{\theta}_1) &\xrightarrow{d} \Lambda B_{m,g} - \Gamma_g(\Gamma' W^{-1}\Gamma)^{-1}\Gamma' W^{-1}\Lambda\bar{B}_m \\ &= \Lambda B_{m,g} - \Gamma(\Gamma' W^{-1}\Gamma)^{-1}\Gamma' W^{-1}\Lambda\bar{B}_m,\end{aligned}$$

where $\bar{B}_m := G^{-1}\sum_{g=1}^{G} B_{m,g}$. It follows that

$$\begin{aligned}\hat{\Gamma}(\hat{\theta}_1)' W_n^{-1}\frac{1}{\sqrt{L}}\sum_{k=1}^{L} f_k^g(\hat{\theta}_1) &\xrightarrow{d} \Gamma' W^{-1}\left[\Lambda B_{m,g} - \Gamma(\Gamma' W^{-1}\Gamma)^{-1}\Gamma' W^{-1}\Lambda\bar{B}_m\right] \\ &= \Gamma' W^{-1}\Lambda B_{m,g} - \Gamma' W^{-1}\Lambda\bar{B}_m = \Gamma' W^{-1}\Lambda\left(B_{m,g} - \bar{B}_m\right).\end{aligned}$$

So, the scaled CCE matrix converges in distribution to a random matrix:

$$\begin{aligned}&\hat{\Gamma}(\hat{\theta}_1)' W_n^{-1}\hat{\Omega}(\hat{\theta}_1) W_n^{-1}\hat{\Gamma}(\hat{\theta}_1) \\ &= \frac{1}{G}\sum_{g=1}^{G}\left[\hat{\Gamma}(\hat{\theta}_1)' W_n^{-1}\left(\frac{1}{\sqrt{L}}\sum_{k=1}^{L} f_k^g(\hat{\theta}_1)\right)\left(\frac{1}{\sqrt{L}}\sum_{k=1}^{L} f_k^g(\hat{\theta}_1)\right)' W_n^{-1}\hat{\Gamma}(\hat{\theta}_1)\right] \\ &\xrightarrow{d} \Gamma' W^{-1}\Lambda\left\{\frac{1}{G}\sum_{g=1}^{G}\left(B_{m,g} - \bar{B}_m\right)\left(B_{m,g} - \bar{B}_m\right)'\right\}\left(\Gamma' W^{-1}\Lambda\right)'.\end{aligned}$$

Therefore,

$$\begin{aligned}n\cdot\widehat{Rvar}(\hat{\theta}_1)R' &= R\left[\hat{\Gamma}(\hat{\theta}_1)' W_n^{-1}\hat{\Gamma}(\hat{\theta}_1)\right]^{-1}\left[\hat{\Gamma}(\hat{\theta}_1)' W_n^{-1}\hat{\Omega}(\hat{\theta}_1) W_n^{-1}\hat{\Gamma}(\hat{\theta}_1)\right]\left[\hat{\Gamma}(\hat{\theta}_1)' W_n^{-1}\hat{\Gamma}(\hat{\theta}_1)\right]^{-1} R' \\ &= R\left[\Gamma' W^{-1}\Gamma\right]^{-1}\Gamma' W^{-1}\Lambda\left\{\frac{1}{G}\sum_{g=1}^{G}\left(B_{m,g} - \bar{B}_m\right)\left(B_{m,g} - \bar{B}_m\right)'\right\}\Lambda W^{-1}\Gamma\left[\Gamma' W^{-1}\Gamma\right]^{-1} R' + o_p(1) \\ &= \tilde{R}\left\{\frac{1}{G}\sum_{g=1}^{G}\left(B_{m,g} - \bar{B}_m\right)\left(B_{m,g} - \bar{B}_m\right)'\right\}\tilde{R}' + o_p(1),\end{aligned}$$

where $\tilde{R} := R\left[\Gamma' W^{-1}\Gamma\right]^{-1}\Gamma' W^{-1}\Lambda$. Also, it follows by Assumption 4 that

$$\begin{aligned}\sqrt{n}(R\hat{\theta}_1 - r) &= -R(\Gamma' W^{-1}\Gamma)^{-1}\Gamma' W^{-1}\sqrt{n}g_n(\theta_0) + o_p(1) \\ &= -R(\Gamma' W^{-1}\Gamma)^{-1}\Gamma' W^{-1}\frac{1}{\sqrt{G}}\sum_{g=1}^{G}\left(\frac{1}{\sqrt{L}}\sum_{i=1}^{L} f_i^g(\theta_0)\right) + o_p(1) \\ &\xrightarrow{d} -\tilde{R}\frac{1}{\sqrt{G}}\sum_{g=1}^{G} B_{m,g} = -\tilde{R}\sqrt{G}\bar{B}_m.\end{aligned}$$

Combining the results so far yields:

$$F(\hat{\theta}_1) \xrightarrow{d} \frac{1}{p}\left(\tilde{R}\sqrt{G}\bar{B}_m\right)'\left\{\tilde{R}\frac{1}{G}\sum_{g=1}^{G}\left(B_{m,g} - \bar{B}_m\right)\left(B_{m,g} - \bar{B}_m\right)'\tilde{R}'\right\}^{-1}\tilde{R}\sqrt{G}\bar{B}_m = \mathbb{F}_{1\infty}.$$

Define the $p \times p$ matrix $\tilde{\Lambda}$ such that $\tilde{\Lambda}\tilde{\Lambda}' = \tilde{R}\tilde{R}'$. Then we have the following distributional equivalence

$$\left[\begin{array}{cc} \tilde{R}\sqrt{G}\bar{B}_m, & \tilde{R}G^{-1}\sum_{g=1}^{G}\left(B_{m,g} - \bar{B}_m\right)\left(B_{m,g} - \bar{B}_m\right)'\tilde{R}' \end{array}\right] \stackrel{d}{=} \left[\begin{array}{cc} \sqrt{G}\tilde{\Lambda}\bar{B}_p, & \tilde{\Lambda}\bar{\mathbb{S}}_{pp}\tilde{\Lambda}' \end{array}\right].$$

Using this, we get

$$\mathbb{F}_{1\infty} \stackrel{d}{=} \frac{G}{p}\cdot\bar{B}_p'\bar{\mathbb{S}}_{pp}^{-1}\bar{B}_p$$

as desired for Part (a). Part (b) can be similarly proved. ■

Proof of Proposition 6. Parts (a), (b) and (c). All three estimators can be represented in the following form

$$-(\Gamma' M^{-1}\Gamma)^{-1}\Gamma' M^{-1}\Lambda\sqrt{G}\bar{B}_m + o_p(1),$$

for some weighing matrix M which may be random. Let $M_\Lambda = \Lambda^{-1}M\left(\Lambda'\right)^{-1}$ and $\Gamma_\Lambda = \Lambda^{-1}\Gamma$. Then,

$$-(\Gamma' M^{-1}\Gamma)^{-1}\Gamma' M^{-1}\Lambda\sqrt{G}\bar{B}_m = -(\Gamma_\Lambda M_\Lambda^{-1}\Gamma_\Lambda)^{-1}\Gamma_\Lambda M_\Lambda^{-1}\sqrt{G}\bar{B}_m,$$

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

$$\Sigma = \left[\begin{array}{c} A_{d\times d} \\ O_{q\times d}\end{array}\right],$$

where A is a diagonal matrix. Denoting

$$\tilde{M} = U' M_\Lambda U = \left(\begin{array}{cc} \underset{d\times d}{\tilde{M}_{11}} & \underset{d\times q}{\tilde{M}_{12}} \\ \underset{q\times q}{\tilde{M}_{21}} & \underset{q\times d}{\tilde{M}_{22}}\end{array}\right) \text{ and } \tilde{M}^{-1} = \left(\begin{array}{cc} \underset{d\times d}{\tilde{M}^{11}} & \underset{d\times q}{\tilde{M}^{12}} \\ \underset{q\times q}{\tilde{M}^{21}} & \underset{q\times d}{\tilde{M}^{22}}\end{array}\right),$$

we have

$$
\begin{aligned}
&(\Gamma_\Lambda M_\Lambda^{-1}\Gamma_\Lambda)^{-1}\Gamma_\Lambda M_\Lambda^{-1} = \left[V\Sigma' U' M_\Lambda^{-1} U\Sigma V'\right]^{-1} V\Sigma' U' M_\Lambda^{-1} \\
&= \left[V\Sigma' \left(U' M_\Lambda U\right)^{-1} \Sigma V'\right]^{-1} V\Sigma' \left(U' M_\Lambda U\right)^{-1} U' \\
&= V\left(A'\tilde{M}^{11}A\right)^{-1} \left(\begin{array}{cc} A'_{d\times d}, & O'_{q\times d} \end{array}\right)' \left(U' M_\Lambda U\right)^{-1} U'' \\
&= VA^{-1}(\tilde{M}^{11})^{-1} \left(\begin{array}{cc} \tilde{M}^{11}, & \tilde{M}^{12} \end{array}\right) U' = VA^{-1} \left(\begin{array}{cc} I_d, & (\tilde{M}^{11})^{-1}\tilde{M}^{12} \end{array}\right) U' \\
&= VA^{-1} \left(\begin{array}{cc} I_d, & -\tilde{M}_{12}\tilde{M}_{22}^{-1} \end{array}\right) U'.
\end{aligned}
$$

, where the last line follows by the partitioned inverse formula that $\tilde{M}^{12} = -\tilde{M}^{11}\tilde{M}_{12}\tilde{M}_{22}^{-1}$. Thus,

$$
-(\Gamma_\Lambda M_\Lambda^{-1}\Gamma_\Lambda)^{-1}\Gamma_\Lambda M_\Lambda^{-1}\sqrt{G}\bar{B}_m = -VA^{-1} \left(\begin{array}{cc} I_d, & -\tilde{M}_{12}\tilde{M}_{22}^{-1} \end{array}\right) U'\sqrt{G}\bar{B}_m.
$$

For $\hat{\theta}_1$, the matrix M is W, and so

$$
\tilde{M} = \tilde{W} = (\Lambda U)^{-1} W \left[(\Lambda U)^{-1}\right]' = \left(\begin{array}{cc} \tilde{W}_{11} & \tilde{W}_{12} \\ \tilde{W}_{21} & \tilde{W}_{22} \end{array}\right).
$$

Therefore,

$$
\sqrt{n}(\hat{\theta}_1 - \theta_0) \xrightarrow{d} -\sqrt{G}VA^{-1}\left(\bar{B}_d - \beta_{\tilde{W}}\bar{B}_q\right),
$$

where we have used $U'\bar{B}_m \overset{d}{=} \bar{B}_m = (\bar{B}'_d, \bar{B}'_q)'$ for any orthonormal matrix U.

For $\tilde{\theta}$, the matrix M_Λ is $\mathbb{S}$, and so

$$
\begin{aligned}
&\sqrt{n}(\tilde{\theta}_2 - \theta_0) \xrightarrow{d} -\left[V\Sigma' U' M_\Lambda^{-1} U\Sigma V'\right]^{-1} V\Sigma' U' M_\Lambda^{-1}\sqrt{G}U'\bar{B}_m \\
&= -V(\Sigma U'\mathbb{S}^{-1}U\Sigma')^{-1}\Sigma U'\mathbb{S}^{-1}U\sqrt{G}U'\bar{B}_m \\
&\overset{d}{=} -V(\Sigma\mathbb{S}^{-1}\Sigma')^{-1}\Sigma\mathbb{S}^{-1}\sqrt{G}\bar{B}_m,
\end{aligned}
$$

using the asymptotic equivalence $(\mathbb{S}, \bar{B}_m) \overset{d}{=} (U'\mathbb{S}U, U'\bar{B}_m)$ for any orthonormal matrix U. Therefore,

$$
\sqrt{n}(\tilde{\theta}_2 - \theta_0) \xrightarrow{d} -VA^{-1}\sqrt{G}(\bar{B}_d - \beta_{\mathbb{S}}\bar{B}_q).
$$

For the estimator $\hat{\theta}_2$, the matrix M_Λ is $\tilde{\mathbb{D}}_\infty$. We have

$$
\begin{aligned}
&\sqrt{n}(\hat{\theta}_2 - \theta_0) \xrightarrow{d} -\left[\Gamma'_\Lambda\tilde{\mathbb{D}}_\infty^{-1}\Gamma_\Lambda\right]^{-1}\Gamma'_\Lambda\tilde{\mathbb{D}}_\infty^{-1}\sqrt{G}\bar{B}_m \\
&= -\left[V\Sigma'\left(U'\tilde{\mathbb{D}}_\infty^{-1}U\right)^{-1}\Sigma V'\right]^{-1} V\Sigma\left(U'\tilde{\mathbb{D}}_\infty^{-1}U\right)^{-1}U'\sqrt{G}\bar{B}_m \\
&= -V\left[\Sigma'\left(\tilde{\mathbb{D}}_\infty^U\right)^{-1}\Sigma\right]^{-1}\Sigma'\left(\tilde{\mathbb{D}}_\infty^U\right)^{-1}U'\sqrt{G}\bar{B}_m \\
&= -VA^{-1}\left(\begin{array}{cc} I_d, & -\tilde{\mathbb{D}}_{12}^U\left[\tilde{\mathbb{D}}_{22}^U\right]^{-1} \end{array}\right) U'\sqrt{G}\bar{B}_m, \qquad (42)
\end{aligned}
$$

where

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

To investigate each component of $\tilde{\mathbb{D}}_{\infty}^{U} = G^{-1}\sum_{g=1}^{G} U'\tilde{D}_g\tilde{D}_g'U$, we first look at the term $U'\tilde{D}_g$ for each $g = 1, ..., G$:

$$\begin{aligned}
U'\tilde{D}_g &= U'B_{m,g} - U'\Gamma_\Lambda(\Gamma_\Lambda' W_\Lambda^{-1}\Gamma_\Lambda)^{-1}\Gamma_\Lambda' W_\Lambda^{-1}\bar{B}_m \\
&= U'B_{m,g} - U'U\Sigma V'(\Gamma_\Lambda' W_\Lambda^{-1}\Gamma_\Lambda)^{-1}V\Sigma' U' W_\Lambda^{-1}UU'\bar{B}_m \\
&= B_{m,g}^{U} - \Sigma(\Sigma' U' W_\Lambda^{-1}U\Sigma)^{-1}\Sigma' U' W_\Lambda^{-1}U\bar{B}_m^{U},
\end{aligned}$$

where $B_{m,g}^{U} = U'B_{m,g}$ and $\bar{B}_m^{U} = U'\bar{B}_m$. But,

$$\begin{aligned}
&B_{m,g}^{U} - \Sigma(\Sigma'\tilde{W}^{-1}\Sigma)^{-1}\Sigma'\tilde{W}^{-1}\bar{B}_m^{U} \\
&= B_{m,g}^{U} - \begin{bmatrix} A \\ O \end{bmatrix}(A\tilde{W}^{11}A)^{-1}\begin{bmatrix} A & O' \end{bmatrix}\begin{pmatrix} \tilde{W}^{11} & \tilde{W}^{12} \\ \tilde{W}^{21} & \tilde{W}^{22} \end{pmatrix}\bar{B}_m^{U} \\
&= B_{m,g}^{U} - \begin{pmatrix} \left(\tilde{W}^{11}\right)^{-1} & O' \\ O & O \end{pmatrix}\begin{pmatrix} \tilde{W}^{11} & \tilde{W}^{12} \\ \tilde{W}^{21} & \tilde{W}^{22} \end{pmatrix}\bar{B}_m^{U} \\
&= B_{m,g}^{U} - \begin{pmatrix} I & \left(\tilde{W}^{11}\right)^{-1}\tilde{W}^{12} \\ O & O \end{pmatrix}\bar{B}_m^{U} \\
&= B_{m,g}^{U} - \begin{bmatrix} \bar{B}_d^{U} - \beta_{\tilde{W}}\bar{B}_q^{U} \\ O \end{bmatrix} = (B_{m,g}^{U} - \bar{B}_m^{U}) + w\bar{B}_q^{U}
\end{aligned}$$

for

$$w = \begin{pmatrix} \beta_{\tilde{W}} \\ I_q \end{pmatrix} \in \mathbb{R}^{m\times q}.$$

So, the matrix $\tilde{\mathbb{D}}_{\infty}^{U}$ can be represented by

$$\begin{aligned}
\tilde{\mathbb{D}}_{\infty}^{U} &= \frac{1}{G}\sum_{g=1}^{G}\left(B_{m,g}^{U} - \bar{B}_m^{U} + w\bar{B}_q^{U}\right)\left(B_{m,g}^{U} - \bar{B}_m^{U} + w\bar{B}_q^{U}\right)' \\
&= \frac{1}{G}\sum_{g=1}^{G}(B_{m,g}^{U} - \bar{B}_m^{U})(B_{m,g}^{U} - \bar{B}_m^{U})' + w\bar{B}_q^{U}(\bar{B}_q^{U})'w' \\
&:= \bar{\mathbb{S}}_{\infty}^{U} + w\bar{B}_q^{U}(\bar{B}_q^{U})'w'.
\end{aligned}$$

From this, the block matrix components of $\tilde{\mathbb{D}}_{\infty}^{U}$ are

$$\begin{aligned}
\tilde{\mathbb{D}}_{11}^{U} &= \bar{\mathbb{S}}_{11}^{U} + \beta_{\tilde{W}}\bar{B}_q^{U}(\bar{B}_q^{U})'\beta_{\tilde{W}}', \\
\tilde{\mathbb{D}}_{12}^{U} &= \bar{\mathbb{S}}_{12}^{U} + \beta_{\tilde{W}}\bar{B}_q^{U}(\bar{B}_q^{U})', \\
\tilde{\mathbb{D}}_{21}^{U} &= \bar{\mathbb{S}}_{21}^{U} + \bar{B}_q^{U}(\bar{B}_q^{U})'\beta_{\tilde{W}}', \\
\tilde{\mathbb{D}}_{22}^{U} &= \bar{\mathbb{S}}_{22}^{U} + \bar{B}_q^{U}(\bar{B}_q^{U})' = \mathbb{S}_{22}.
\end{aligned} \tag{43}$$

Using these representations, we can rewrite (42) as

$$\begin{aligned}
&\sqrt{n}(\hat{\theta}_2-\theta_0) \xrightarrow{d} -VA^{-1}\left(\ I_d,\ \ -\tilde{\mathbb{D}}_{12}^U\left[\tilde{\mathbb{D}}_{22}^U\right]^{-1}\ \right)\sqrt{G}\bar{B}_m^U \\
&= -VA^{-1}\sqrt{G}\left(\bar{B}_d^U-\tilde{\mathbb{D}}_{12}^U\left[\tilde{\mathbb{D}}_{22}^U\right]^{-1}\bar{B}_q^U\right) \\
&= -VA^{-1}\sqrt{G}\left[\bar{B}_d^U-\left(\bar{\mathbb{S}}_{12}^U+\beta_{\tilde{W}}\bar{B}_q^U(\bar{B}_q^U)'\right)\mathbb{S}_{22}^{-1}\bar{B}_q^U\right] \\
&= -VA^{-1}\sqrt{G}\left\{\bar{B}_d^U-\left[\mathbb{S}_{12}^U-(\bar{B}_d^U-\beta_{\tilde{W}}\bar{B}_q^U)(\bar{B}_q^U)'\right]\mathbb{S}_{22}^{-1}\bar{B}_q^U\right\} \\
&\overset{d}{=} -VA^{-1}\sqrt{G}\left(\bar{B}_d-\beta_{\mathbb{S}_\infty}\bar{B}_q\right)-\left(\frac{\kappa_G}{G}\right)\cdot VA^{-1}\sqrt{G}(\bar{B}_d-\beta_{\tilde{W}}\bar{B}_q).
\end{aligned}$$

(d) It is easy to check that the weak convergences in (a)$\sim$(c) hold jointly. By continuous mapping theorem we have

$$\sqrt{n}(\hat{\theta}_2-\theta_0)-\sqrt{n}(\tilde{\theta}_2-\theta_0)-\left(\frac{\kappa_G}{G}\right)\cdot\sqrt{n}(\hat{\theta}_1-\theta_0)\xrightarrow{d}0,$$

which implies that

$$\sqrt{n}(\hat{\theta}_2-\theta_0)-\sqrt{n}(\tilde{\theta}_2-\theta_0)-\left(\frac{\kappa_G}{G}\right)\cdot\sqrt{n}(\hat{\theta}_1-\theta_0)=o_p\left(1\right).$$

That is,

$$\sqrt{n}(\hat{\theta}_2-\theta_0)=\sqrt{n}(\tilde{\theta}_2-\theta_0)+\left(\frac{\kappa_G}{G}\right)\sqrt{n}(\hat{\theta}_1-\theta_0)+o_p\left(1\right).$$

(e) Using the same argument in the proof of Proposition 1, we have

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

with $\tilde{\mathbb{D}}_{12}^U$ and $\tilde{\mathbb{D}}_{22}^U$ given in (43). Therefore,

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

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

where the second last equality follows from straightforward calculations. The joint convergence can be proved easily. Lastly, we obtain the Beta representation of the non-standard limit for J statistic κ_G using the fact $d_1\mathcal{F}_{d_1,d_2}/(d_2+d_1\mathcal{F}_{d_1,d_2}) \overset{d}{=} \text{Beta}(d_1/2, d_2/2)$. ■

Proof of Proposition 7. It follows from

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

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

$$
\mathbb{V} = \begin{pmatrix} \tilde{V}_{d\times d} & O \\ O & I_{q\times q}\end{pmatrix}
$$

and define

$$
\breve{\mathbb{D}} = \begin{pmatrix} \breve{\mathbb{D}}_{11} & \breve{\mathbb{D}}_{12} \\ \breve{\mathbb{D}}_{21} & \breve{\mathbb{D}}_{22}\end{pmatrix} = \begin{pmatrix} \tilde{V}_{d\times d} & O \\ O & I_q\end{pmatrix}'\begin{pmatrix} \tilde{\mathbb{D}}_{11}^U & \tilde{\mathbb{D}}_{12}^U \\ \tilde{\mathbb{D}}_{21}^U & \tilde{\mathbb{D}}_{22}^U\end{pmatrix}\begin{pmatrix} \tilde{V}_{d\times d} & O \\ O & I_q\end{pmatrix} = \mathbb{V}'\tilde{\mathbb{D}}_{\infty}^U\mathbb{V}.
$$

Then,

$$\breve{\mathbb{D}} = \frac{1}{G}\sum_{g=1}^{G} \mathbb{V}' U'(B_{m,g} - \bar{B}_m)(B_{m,g} - \bar{B}_m)'\mathbb{V}U + \begin{pmatrix} \tilde{V}'\beta_{\tilde{W}} \\ I_q \end{pmatrix} \bar{B}_q^U (\bar{B}_q^U)' \begin{pmatrix} \tilde{V}'\beta_{\tilde{W}} \\ I_q \end{pmatrix}'$$
$$\stackrel{d}{=} \frac{1}{G}\sum_{g=1}^{G} (B_{m,g} - \bar{B}_m)(B_{m,g} - \bar{B}_m)' + \begin{pmatrix} \tilde{V}'\beta_{\tilde{W}} \\ I_q \end{pmatrix} \bar{B}_q \bar{B}_q' \begin{pmatrix} \tilde{V}'\beta_{\tilde{W}} \\ I_q \end{pmatrix}',$$

which implies that

$$\breve{\mathbb{D}}_{11} := \begin{pmatrix} \breve{\mathbb{D}}_{pp} & \breve{\mathbb{D}}_{p,d-p} \\ \breve{\mathbb{D}}_{d-p,p} & \breve{\mathbb{D}}_{d-p,d-p} \end{pmatrix} \stackrel{d}{=} \frac{1}{G}\sum_{g=1}^{G} (B_{d,g} - \bar{B}_d)(B_{d,g} - \bar{B}_d)' + \left(\tilde{V}'\beta_{\tilde{W}}\right) \bar{B}_q \bar{B}_q' \left(\tilde{V}'\beta_{\tilde{W}}\right)', \quad (44)$$

and

$$\breve{\mathbb{D}}_{12} := \begin{pmatrix} \breve{\mathbb{D}}_{pq} \\ \breve{\mathbb{D}}_{d-p,q} \end{pmatrix} \stackrel{d}{=} \frac{1}{G}\sum_{g=1}^{G} (B_{d,g} - \bar{B}_d)(B_{q,g} - \bar{B}_q)' + \left(\tilde{V}'\beta_{\tilde{W}}\right) \bar{B}_q \bar{B}_q'. \quad (45)$$

Now

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

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 (44) and (45). Here we have abused the notation a little bit. We have

$$\begin{pmatrix} \breve{\mathbb{D}}_{pp} & \breve{\mathbb{D}}_{pq} \\ \breve{\mathbb{D}}_{pq}' & \breve{\mathbb{D}}_{qq} \end{pmatrix} \stackrel{d}{=} \mathbb{E}_{p+q,p+q} = \begin{pmatrix} \bar{\mathbb{S}}_{pp} & \bar{\mathbb{S}}_{pq} \\ \bar{\mathbb{S}}_{pq}' & \bar{\mathbb{S}}_{qq} \end{pmatrix} + \tilde{w}\bar{B}_q\bar{B}_q'\tilde{w}' \quad (46)$$

for

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

We have therefore shown that the first representation of the limit of $F_{\hat{\Omega}(\hat{\theta}_1)}(\hat{\theta}_2)$ holds. Direct calculations show that the second representation is numerically identical to the first representation. This completes the proof of Proposition 7. ■

Proof of Lemma 8. The centered CCE $\hat{\Omega}^c(\tilde{\theta})$ can be represented as:

$$\hat{\Omega}^c(\tilde{\theta}) = \frac{1}{G}\sum_{g=1}^{G} \left\{ \frac{1}{\sqrt{L}}\sum_{k=1}^{L} \left(f_k^g(\tilde{\theta}) - \frac{1}{n}\sum_{\tilde{g}=1}^{G}\sum_{\tilde{k}=1}^{L} f_{\tilde{k}}^{\tilde{g}}(\tilde{\theta}) \right) \right.$$
$$\left. \times \frac{1}{\sqrt{L}}\sum_{k=1}^{L} \left(f_k^g(\tilde{\theta}) - \frac{1}{n}\sum_{\tilde{g}=1}^{G}\sum_{\tilde{k}=1}^{L} f_{\tilde{k}}^{\tilde{g}}(\tilde{\theta}) \right)' \right\}.$$

To prove Part (a), it suffices to show that

$$\frac{1}{\sqrt{L}}\sum_{k=1}^{L}\left(f_k^g(\tilde{\theta})-\frac{1}{n}\sum_{\tilde{g}=1}^{G}\sum_{\tilde{k}=1}^{L}f_{\tilde{k}}^{\tilde{g}}(\tilde{\theta})\right)=\frac{1}{\sqrt{L}}\sum_{k=1}^{L}\left(f_k^g(\theta_0)-\frac{1}{n}\sum_{\tilde{g}=1}^{G}\sum_{\tilde{k}=1}^{L}f_{\tilde{k}}^{\tilde{g}}(\theta_0)\right)(1+o_p(1)) \quad (47)$$

holds for each $g=1,...,G$. By Assumption 3 and using a Taylor expansion, we have

$$\frac{1}{\sqrt{L}}\sum_{k=1}^{L}f_k^g(\tilde{\theta})=(1+o_p(1))\left(\frac{1}{\sqrt{L}}\sum_{k=1}^{L}f_k^g(\theta_0)+\frac{1}{L}\sum_{k=1}^{L}\frac{\partial f_k^g(\tilde{\theta})}{\partial\theta'}\sqrt{L}(\tilde{\theta}-\theta_0)\right).$$

Using $\sqrt{n}(\tilde{\theta}-\theta_0)=O_p(1)$ and Assumption 5, we have

$$\frac{1}{\sqrt{L}}\sum_{k=1}^{L}f_k^g(\tilde{\theta})=(1+o_p(1))\left(\frac{1}{\sqrt{L}}\sum_{k=1}^{L}f_k^g(\theta_0)+\Gamma\sqrt{L}(\tilde{\theta}-\theta_0)\right)$$

for each $g=1,...,G$. That is, the effect of the estimation uncertainty in $\tilde{\theta}$ does not change with the cluster. It then follows that

$$\frac{1}{\sqrt{L}}\sum_{k=1}^{L}\left(f_k^g(\tilde{\theta})-\frac{1}{n}\sum_{\tilde{g}=1}^{G}\sum_{\tilde{k}=1}^{L}f_{\tilde{k}}^{\tilde{g}}(\tilde{\theta})\right)=\left(\frac{1}{\sqrt{L}}\sum_{k=1}^{L}f_k^g(\theta_0)-\frac{1}{G}\sum_{\tilde{g}=1}^{G}\frac{1}{\sqrt{L}}\sum_{\tilde{k}=1}^{L}f_{\tilde{k}}^{\tilde{g}}(\theta_0)\right)(1+o_p(1)),$$

which completes the proof of part (a).

To prove Part (b), we apply CLT in Assumption 4 together with 6 to obtain:

$$\frac{1}{\sqrt{L}}\sum_{k=1}^{L}f_k^g(\theta_0)-\frac{1}{G}\sum_{\tilde{g}=1}^{G}\frac{1}{\sqrt{L}}\sum_{\tilde{k}=1}^{L}f_{\tilde{k}}^{\tilde{g}}(\theta_0)\xrightarrow{d}\Lambda\left(B_{m,h}-\bar{B}_m\right),$$

where the convergence holds jointly for $g=1,...,G$. As a result,

$$\hat{\Omega}^c(\theta_0)\xrightarrow{d}\frac{1}{G}\Lambda\left(\sum_{g=1}^{G}\left(B_{m,g}-\bar{B}_m\right)\left(B_{m,g}-\bar{B}_m\right)'\right)\Lambda'.$$

■

Proof of Proposition 9. The proof of part (a) is essentially the same as the proof of Proposition 7. The only difference is that the second term in (46) will not be present for the centered two-step GMM estimator $\hat{\theta}_2^c$. The proof of part (b) is similar. The proof of part (e) is similar to that of Proposition 6(e).

To prove part (c), it sufficies 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 small-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

$$\sqrt{n}(\hat{\theta}_2^{c,r} - \theta_0) = -\Phi^{-1}\Gamma' \left[\hat{\Omega}^c(\hat{\theta}_1)\right]^{-1} \sqrt{n} g_n(\theta_0) \tag{48}$$
$$- \Phi^{-1} R' \left(R\Phi^{-1}R'\right)^{-1} R\Phi^{-1}\Gamma' \left[\hat{\Omega}_n^c(\hat{\theta}_1)\right]^{-1} \sqrt{n} g_n(\theta_0) + o_p(1)$$

and

$$\sqrt{n}\lambda_n = -(R\Phi^{-1}R')^{-1} R\Phi^{-1}\Gamma' \left[\hat{\Omega}_n^c(\hat{\theta}_1)\right]^{-1} \sqrt{n} g_n(\theta_0) + o_p(1), \tag{49}$$

where $\Phi := \Gamma' \left[\hat{\Omega}^c(\hat{\theta}_1)\right]^{-1} \Gamma$. Subtracting (48) from (12), we have

$$\begin{aligned} \sqrt{n}(\hat{\theta}_2^{c,r} - \hat{\theta}_2^c) &= -\Phi^{-1} R' \left(R\Phi^{-1}R'\right)^{-1} R\Phi^{-1}\Gamma' \left[\hat{\Omega}^c(\hat{\theta}_1)\right]^{-1} \sqrt{n} g_n(\theta_0) + o_p(1) && (50) \\ &= O_p(1), && (51) \end{aligned}$$

where the equation (51) comes from Lemma 8-(b) and Assumption 4. 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 $LR_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c, \hat{\theta}_2^{c,r})$,

$$\begin{aligned} LR_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c, \hat{\theta}_2^{c,r}) &= \frac{n}{p}\Bigg\{(\hat{\theta}_2^{c,r} - \hat{\theta}_2^c)'\hat{\Gamma}'(\hat{\theta}_2^c) \left[\hat{\Omega}^c(\hat{\theta}_1)\right]^{-1} \hat{\Gamma}(\hat{\theta}_2^c)(\hat{\theta}_2^{c,r} - \hat{\theta}_2^c) && (52) \\ &\quad -2g_n'(\hat{\theta}_2^c) \left[\hat{\Omega}^c(\hat{\theta}_1)\right]^{-1} \hat{\Gamma}(\hat{\theta}_2^c)(\hat{\theta}_2^{c,r} - \hat{\theta}_2^c)\Bigg\} + o_p(1), \end{aligned}$$

where the last term in (52) is always zero from the FOC of $\hat{\theta}_2^c$. Combining (50) and (52), we have

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

as desired.

To prove part (d), 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 (49), 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. ■

Proof of Theorem 10. Define $\mathbf{B}_q' = (B_{q,1}', ..., B_{q,G}')'$ and denote

$$v_g = \left(B_{q,g} - \bar{B}_q\right)' \left[\sum_{g=1}^{G} \left(B_{q,g} - \bar{B}_q\right)\left(B_{q,g} - \bar{B}_q\right)'\right]^{-1} \bar{B}_q.$$

Then, the distribution of $\sqrt{G}\bar{\mathbb{S}}_{pq}\bar{\mathbb{S}}_{qq}^{-1}\bar{B}_q$ conditional on $\mathbf{B}_q$ can be represented as

$$\begin{aligned} &\sqrt{G}\left(\sum_{g=1}^{G}\left(B_{p,g} - \bar{B}_p\right)\left(B_{q,g} - \bar{B}_q\right)'\right)\left(\sum_{g=1}^{G}\left(B_{q,g} - \bar{B}_q\right)\left(B_{q,g} - \bar{B}_q\right)'\right)^{-1}\bar{B}_q \\ &= \sqrt{G}\sum_{g=1}^{G}\left(B_{p,g} - \bar{B}_p\right)v_g = \sqrt{G}\sum_{g=1}^{G}B_{p,g}v_g - \sqrt{G}\bar{B}_p\sum_{g=1}^{G}v_g \\ &\overset{d}{=} N\left(0, G\sum_{g=1}^{G}v_g^2 \cdot I_p\right), \end{aligned}$$

where the last line holds because $\sum_{g=1}^{G} v_g = 0$. Note that

$$\begin{aligned} G\sum_{g=1}^{G}v_g^2 = G\sum_{g=1}^{G}\Bigg\{ &\left(B_{q,g} - \bar{B}_q\right)'\left[\sum_{g=1}^{G}\left(B_{q,g} - \bar{B}_q\right)\left(B_{q,g} - \bar{B}_q\right)'\right]^{-1}\bar{B}_q \\ &\cdot \bar{B}_q'\left[\sum_{g=1}^{G}\left(B_{q,g} - \bar{B}_q\right)\left(B_{q,g} - \bar{B}_q\right)'\right]^{-1}\left(B_{q,g} - \bar{B}_q\right)\Bigg\} \\ = G\bar{B}_q'&\left[\sum_{g=1}^{G}\left(B_{q,g} - \bar{B}_q\right)\left(B_{q,g} - \bar{B}_q\right)'\right]^{-1}\left[\sum_{g=1}^{G}\left(B_{q,g} - \bar{B}_q\right)\right. \\ &\left.\times\left(B_{q,g} - \bar{B}_q\right)'\right]\left[\sum_{g=1}^{G}\left(B_{q,g} - \bar{B}_q\right)\left(B_{q,g} - \bar{B}_q\right)'\right]\bar{B}_q \\ = \bar{B}_q'&\left[\sum_{g=1}^{G}\left(B_{q,g} - \bar{B}_q\right)\left(B_{q,g} - \bar{B}_q\right)'/G\right]^{-1}\bar{B}_q \\ = \bar{B}_q'&\bar{\mathbb{S}}_{qq}^{-1}\bar{B}_q. \end{aligned}$$

So conditional on $\mathbf{B}_q$, $\sqrt{G}\bar{\mathbb{S}}_{pq}\bar{\mathbb{S}}_{qq}^{-1}\bar{B}_q$ is distributed as $N(0, \bar{B}_q'\bar{\mathbb{S}}_{qq}^{-1}\bar{B}_q \cdot I_p)$. It then follows that the distribution of $\sqrt{G}(\bar{B}_p - \bar{\mathbb{S}}_{pq}\bar{\mathbb{S}}_{qq}^{-1}\bar{B}_q)$ conditional on $\mathbf{B}_q$ is

$$\sqrt{G}\left(\bar{B}_p - \bar{\mathbb{S}}_{pq}\bar{\mathbb{S}}_{qq}^{-1}\bar{B}_q\right) \sim N\left(0, (1 + \bar{B}_q'\bar{\mathbb{S}}_{qq}^{-1}\bar{B}_q) \cdot I_p\right),$$

using the independence of $\bar{B}_p$ from $\bar{\mathbb{S}}_{pq}\bar{\mathbb{S}}_{qq}^{-1}\bar{B}_q$ conditional on $\mathbf{B}_q$. Therefore the conditional distribution of ξ_p is

$$\xi_p := \frac{\sqrt{G}(\bar{B}_p - \bar{\mathbb{S}}_{pq}\bar{\mathbb{S}}_{qq}^{-1}\bar{B}_q)}{\sqrt{1+\bar{B}_q'\bar{\mathbb{S}}_{qq}^{-1}\bar{B}_q}} \sim N(0, I_p).$$

Given that the conditional distribution of ξ_p does not depend on $\mathbf{B}_q$, the unconditional distribution of ξ_p is also $N(0, I_p)$.

Using $\xi_p \sim N(0, I_p)$, $\bar{\mathbb{S}}_{pp\cdot q} \sim G^{-1}\mathbb{W}_p(G-q-1, I_p)$, and ξ_p which is independent of $\bar{\mathbb{S}}_{pp\cdot q}$, we have

$$\xi_p' \left(\frac{G\bar{\mathbb{S}}_{pp\cdot q}}{G-q-1}\right)^{-1} \xi_p \sim \text{Hotelling's } T^2 \text{ distribution } T^2_{p,G-q-1}.$$

It then follows that

$$\frac{G-p-q}{p\,(G-q-1)}\xi_p' \left(\frac{G\bar{\mathbb{S}}_{pp\cdot q}}{G-q-1}\right)^{-1} \xi_p \sim \mathcal{F}_{p,G-p-q}.$$

That is,

$$\frac{G-p-q}{pG}\xi_p'\bar{\mathbb{S}}_{pp\cdot q}^{-1}\xi_p \sim \mathcal{F}_{p,G-p-q}.$$

Together with Proposition 9(c) and (d), this completes the proof of the F limit theory in parts (a), (b) and (c). The proof of the t limit theory is similar and is omitted here. ■

Proof of Proposition 11. For the result with CU-GEE estimator $\hat{\theta}_{\text{CU-GEE}}$, we have

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

Since $\hat{\theta}_{\text{CU-GEE}}$ is $\sqrt{n}$-consistent, we can apply Lemma 8 to obtain $\hat{\Omega}^c(\hat{\theta}_{\text{CU-GEE}}) = \hat{\Omega}^c(\theta_0) + o_p(1)$. 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

$$\begin{aligned} 0 = {} & \hat{\Gamma}^j(\hat{\theta}_{\text{CU-GMM}})'\left[\hat{\Omega}^c(\hat{\theta}_{\text{CU-GMM}})\right]^{-1} g_n(\hat{\theta}_{\text{CU-GMM}}) \\ & - g_n(\hat{\theta}_{\text{CU-GMM}})'\left[\hat{\Omega}^c(\hat{\theta}_{\text{CU-GMM}})\right]^{-1}\Upsilon_j(\hat{\theta}_{\text{CU-GMM}})\left[\hat{\Omega}^c(\hat{\theta}_{\text{CU-GMM}})\right]^{-1} g_n(\hat{\theta}_{\text{CU-GMM}}), \end{aligned} \tag{53}$$

where

$$\Upsilon_j(\theta) = \frac{1}{n}\sum_{g=1}^{G}\left(\sum_{k=1}^{L} f_k^g(\theta)\right)\left(\sum_{k=1}^{L}\frac{\partial f_k(\theta)}{\partial\theta_j}\right)' - L\cdot g_n(\theta)\left(\frac{\partial g_n(\theta)}{\partial\theta_j}\right)'.$$

The second term in (53) can be written as

$$
\begin{aligned}
& g_n(\hat{\theta}_{\text{CU-GMM}})' \left[\hat{\Omega}^c(\hat{\theta}_{\text{CU-GMM}})\right]^{-1} \Upsilon_j(\hat{\theta}_{\text{CU-GMM}}) \left[\hat{\Omega}^c(\hat{\theta}_{\text{CU-GMM}})\right]^{-1} g_n(\hat{\theta}_{\text{CU-GMM}}) \\
& = \sqrt{L} g_n(\hat{\theta}_{\text{CU-GMM}})' \left[\hat{\Omega}^c(\hat{\theta}_{\text{CU-GMM}})\right]^{-1} \left[\frac{1}{G}\sum_{g=1}^{G}\left(\frac{1}{L}\sum_{k=1}^{L} f_k^g(\hat{\theta}_{\text{CU-GMM}})\right)\right. \\
& \left. \cdot \left\{\left(\frac{1}{L}\sum_{k=1}^{L}\frac{\partial f_k^g(\hat{\theta}_{\text{CU-GMM}})}{\partial \theta_j}\right) - \frac{1}{G}\sum_{g=1}^{G}\left(\frac{1}{L}\sum_{k=1}^{L}\frac{\partial f_k^g(\hat{\theta}_{\text{CU-GMM}})}{\partial \theta_j}\right)\right\}'\right] \\
& \cdot \left[\hat{\Omega}^c(\hat{\theta}_{\text{CU-GMM}})\right]^{-1} \sqrt{L} g_n(\hat{\theta}_{\text{CU-GMM}}).
\end{aligned}
$$

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

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

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

$$
\begin{aligned}
& \left(\frac{1}{L}\sum_{k=1}^{L} f_k^g(\hat{\theta}_{\text{CU-GMM}})\right)\left\{\left(\frac{1}{L}\sum_{k=1}^{L}\frac{\partial f_k^g(\hat{\theta}_{\text{CU-GMM}})}{\partial \theta}\right) - \frac{1}{G}\sum_{g=1}^{G}\left(\frac{1}{L}\sum_{k=1}^{L}\frac{\partial f_k^g(\hat{\theta}_{\text{CU-GMM}})}{\partial \theta}\right)\right\}' \\
& = O_p\left(\frac{1}{\sqrt{L}}\right)\cdot o_p(1) = o_p\left(\frac{1}{\sqrt{L}}\right).
\end{aligned}
$$

Combining these together, the second term in FOC in (53) is $o_p(L^{-1/2})$. As a result,

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

and so

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

■

Proof of Theorem 12.

We first show that $\widehat{\mathcal{E}}_n = \mathcal{E}_{2n}\left(1 + o_p\left(1\right)\right)$. For each $j = 1, ..., d$, we have

$$
\begin{aligned}
\widehat{\mathcal{E}}_n[., j] &= \left\{\hat{\Gamma}'\left[\hat{\Omega}^c(\hat{\theta}_1)\right]^{-1}\hat{\Gamma}\right\}^{-1}\hat{\Gamma}'\left[\hat{\Omega}^c(\hat{\theta}_1)\right]^{-1}\left.\frac{\partial\hat{\Omega}^c(\theta)}{\partial\theta_j}\right|_{\theta=\hat{\theta}_1}\left[\hat{\Omega}^c(\hat{\theta}_1)\right]^{-1} g_n(\hat{\theta}_2^c) \\
&= \left\{\Gamma'\left[\hat{\Omega}^c(\theta_0)\right]^{-1}\Gamma\right\}^{-1}\Gamma'\left[\hat{\Omega}^c(\theta_0)\right]^{-1}\left.\frac{\partial\hat{\Omega}^c(\theta)}{\partial\theta_j}\right|_{\theta=\hat{\theta}_1}\left[\hat{\Omega}^c(\hat{\theta}_1)\right]^{-1} g_n(\hat{\theta}_2^c)(1 + o_p(1)),
\end{aligned}
$$

where the second equality holds by Assumption 3, 5 and Lemma 8. Using a Taylor expansion, we have

$$g_n(\hat{\theta}_2^c) = g_n(\theta_0) - \Gamma \left\{ \Gamma' \left[\hat{\Omega}^c(\theta_0) \right]^{-1} \Gamma \right\}^{-1} \Gamma' \left[\hat{\Omega}^c(\theta_0) \right]^{-1} g_n(\theta_0)(1 + o_p(1)).$$

Thus,

$$\begin{aligned}
\widehat{\mathcal{E}}_n[., j] = & \left\{ \Gamma' \left[\hat{\Omega}^c(\theta_0) \right]^{-1} \Gamma \right\}^{-1} \Gamma' \left[\hat{\Omega}^c(\theta_0) \right]^{-1} \left. \frac{\partial \hat{\Omega}^c(\theta)}{\partial \theta_j} \right|_{\theta = \hat{\theta}_1} \left[\hat{\Omega}^c(\theta_0) \right]^{-1} g_n(\theta_0)(1 + o_p(1)) \\
& - \left\{ \Gamma' \left[\hat{\Omega}^c(\theta_0) \right]^{-1} \Gamma \right\}^{-1} \Gamma' \left[\hat{\Omega}^c(\theta_0) \right]^{-1} \left. \frac{\partial \hat{\Omega}^c(\theta)}{\partial \theta_j} \right|_{\theta = \hat{\theta}_1} \left[\hat{\Omega}^c(\theta_0) \right]^{-1} \Gamma \\
& \times \left\{ \Gamma' \left[\hat{\Omega}^c(\theta_0) \right]^{-1} \Gamma \right\}^{-1} \Gamma' \left[\hat{\Omega}^c(\theta_0) \right]^{-1} g_n(\theta_0) \Bigg\} (1 + o_p(1)),
\end{aligned}$$

for each $j = 1, ..., d$. For the term, $\left. \frac{\partial \hat{\Omega}^c(\theta)}{\partial \theta_j} \right|_{\theta = \hat{\theta}_1}$, recall that

$$\begin{aligned}
\left. \frac{\partial \hat{\Omega}^c(\theta)}{\partial \theta_j} \right|_{\theta = \hat{\theta}_1} &= \Upsilon_j(\hat{\theta}_1) + \Upsilon_j'(\hat{\theta}_1), \\
\Upsilon_j(\theta) &= \frac{1}{n} \sum_{g=1}^{G} \left[\sum_{k=1}^{L} \left(f_k^g(\theta) - \frac{1}{n} \sum_{i=1}^{n} f_i(\theta) \right) \left(\sum_{k=1}^{L} \left(\frac{\partial f_k^g(\theta)}{\partial \theta_j} - \frac{1}{n} \sum_{i=1}^{n} \frac{\partial f_i(\theta)}{\partial \theta_j} \right) \right)' \right].
\end{aligned}$$

It remains to show that $\Upsilon_j(\hat{\theta}_1) = \Upsilon_j(\theta_0)(1 + o_p(1))$. From the proof of Lemma 8, we have

$$\begin{aligned}
& \frac{1}{\sqrt{L}} \sum_{k=1}^{L} \left(f_k^g(\hat{\theta}_1) - \frac{1}{n} \sum_{i=1}^{n} f_i(\hat{\theta}_1) \right) \\
& = \frac{1}{\sqrt{L}} \sum_{k=1}^{L} \left(f_k^g(\theta_0) - \frac{1}{n} \sum_{i=1}^{n} f_i(\theta_0) \right) (1 + o_p(1)),
\end{aligned} \tag{54}$$

for each $g = 1, ..., G$. By Assumption 3, 7 and a Taylor expansion, we have:

$$\begin{aligned}
\frac{1}{\sqrt{L}} \sum_{k=1}^{L} \frac{\partial f_k^g(\hat{\theta}_1)}{\partial \theta_j} &= \left(\frac{1}{\sqrt{L}} \sum_{k=1}^{L} \frac{\partial f_k^g(\theta_0)}{\partial \theta_j} + \frac{1}{L} \sum_{k=1}^{L} \frac{\partial}{\partial \theta'} \left(\frac{\partial f_k^g(\theta_0)}{\partial \theta_j} \right) \sqrt{L}(\hat{\theta}_1 - \theta_0) \right) (1 + o_p(1)) \\
&:= \left(\frac{1}{\sqrt{L}} \sum_{k=1}^{L} \frac{\partial f_k^g(\theta_0)}{\partial \theta_j} + Q(\theta_0) \sqrt{L}(\hat{\theta}_1 - \theta_0) \right) (1 + o_p(1)),
\end{aligned}$$

for $j = 1, ..., d$ and $g = 1, ..., G$. This implies that

$$\frac{1}{\sqrt{L}} \sum_{k=1}^{L} \left(\frac{\partial f_k^g(\hat{\theta}_1)}{\partial \theta_j} - \frac{1}{n} \sum_{i=1}^{n} \frac{\partial f_i(\hat{\theta}_1)}{\partial \theta_j} \right) = \frac{1}{\sqrt{L}} \sum_{k=1}^{L} \left(\frac{\partial f_k^g(\theta_0)}{\partial \theta_j} - \frac{1}{n} \sum_{i=1}^{n} \frac{\partial f_i(\theta_0)}{\partial \theta_j} \right) (1 + o_p(1)).$$

Combining these together, we have $\Upsilon(\hat{\theta}_1) = \Upsilon(\theta_0)(1 + o_p(1))$ from which we obtain the desired result

$$\widehat{\mathcal{E}}_n = \mathcal{E}_{2n} \left(1 + o_p \left(1 \right) \right). \tag{55}$$

Now, define the infeasible corrected variance

$$\begin{aligned}&\widehat{var}^{\text{adj,inf}}_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c)\\&= \widehat{var}_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c) + \mathcal{E}_{2n}\widehat{var}_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c) + \widehat{var}_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c)\mathcal{E}_{2n}' + \mathcal{E}_{2n}'\widehat{var}(\hat{\theta}_1)\widehat{\mathcal{E}}_{2n}',\end{aligned}$$

and the corresponding infeasible Wald statistic

$$F^{\text{adj,inf}}_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c) = \frac{1}{p}(R\hat{\theta}_2^c - r)'\left[R\widehat{var}^{\text{adj,inf}}_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c)R'\right]^{-1}(R\hat{\theta}_2 - r).$$

The result in (55) implies

$$F^{\text{adj,inf}}_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c) = F^{\text{adj}}_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c)(1 + o_p(1)).$$

Also, $\mathcal{E}_{2n} = o_p(1)$ and we have

$$\widehat{var}^{\text{adj,inf}}_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c) = \widehat{var}^{\text{adj}}_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c)(1 + o_p(1)) = \widehat{var}_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c)(1 + o_p(1)),$$

and so

$$F^{\text{adj,inf}}_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c) = F^{\text{adj}}_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c) + o_p(1) = F_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c) + o_p(1),$$

as desired. ■