

# Simple and Trustworthy Cluster-Robust GMM Inference

Jungbin Hwang\*

Department of Economics,  
University of Connecticut<sup>†</sup>

September 9, 2016

## Abstract

This paper develops a new asymptotic theory for two-step GMM estimation and inference in the presence of clustered dependence. While conventional asymptotic theory completely ignores the variability in the cluster-robust GMM weighting matrix, the new asymptotic theory takes it into account, leading to more accurate approximations. The key difference between these two types of asymptotics is whether the number of clusters  $G$  is regarded as small (fixed) or growing when the sample size increases. Under the new small- $G$  asymptotics, the centered two-step GMM estimator and the two continuously-updating estimators 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 the F and t critical values are readily available from standard statistical tables. A Monte Carlo study shows that our proposed tests are much more accurate than existing tests. We also apply our methods to an empirical study on the causal effect of access to domestic and international markets on household consumption in rural China. The results suggest that the effect of access to markets may be lower than the previous finding.

---

\*Email: jungbin.hwang@uconn.edu. Correspondence to: Department of Economics 365 Fairfield Way, U-1063 Storrs, CT 06269-1063.

<sup>†</sup>I am indebted to Yixiao Sun, Graham Elliott, Andres Santos, and Minseong Kim that helped greatly improve this paper. I would also like to thank Eli Berman, Stephane Bonhomme, Kenneth Couch, Gordon Dahl, James D. Hamilton, Ivan Jeliazkov, Harim Kim, Michal Kolesár, Stepehn L. Ross, and other seminar participants at UCSD, UCONN, UCI, and 2016 North American Meeting of Econometric Society.

# 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 (e.g., de Brauw and Giles 2008; Pepper 2002; Dube et al. 2010). In empirical finance and industrial organization, firm level data are often clustered at the industry level (Samila and Sorenson, 2011; Bharath et al., 2013), and in many educational studies, students' test scores are clustered at the classroom or school level (Andrabi et.al, 2011). 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), 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  $G$  diverges 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, 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 et al., 2004; Obstfeld et al., 2010; 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 et al., 2008; Bester et al., 2011; Mackinnon and Webb, 2014).

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. In fact, our approximations work especially well when the chi-squared and normal approximations are poor. They are obtained from a 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 statistic is not asymptotically pivotal. However, it is surprising that the J-statistic is still asymptotically pivotal. Furthermore, we show that the limiting distribution of the J-statistic can be represented as an increasing function of a standard F random variable. So critical values are readily available from standard statistical tables and software packages.

Next, we establish the asymptotic properties of the “centered” two-step GMM estimator<sup>1</sup> 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.

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’s can be regarded as the empirical version of generalized estimating equations (GEE), 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. So in terms of the first order asymptotics, it does not matter whether

---

<sup>1</sup>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.

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.

To relate the small- $G$  asymptotic pivotal distributions to standard distributions, we introduce simple modifications to the Wald and  $t$  statistics associated with the centered two-step GMM and CU estimators. 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.

Finally, although recentering removes the first order effect of the first-step estimation error, 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 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 only in the i.i.d. setting. We show that the finite sample variance correction does not change the 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-square 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) and Sun (2013, 2014), 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 (2015b) 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 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. Bell and McCaffrey (2002) and Imbens and Kolesár (2012) suggest an adjusted t-critical value employing data-determined degrees of freedom. Recently, Canay, Romano and Shaikh (2014) establishes a theory of randomization tests and suggests an alternative cluster-robust test. For other approaches, see Carter, Schnepel and Steigerwald (2013) which proposes a measure of the effective number of clusters under the large- $G$  asymptotics.; Cameron, Gelbach and Miller (2008), MacKinnon and Webb (2016), and Webb (2014) which provide cluster bootstrap approaches with asymptotic refinement. 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 the CU estimators, respectively. Section 5 is devoted to developing asymptotic F and t tests based on the centered two-step GMM estimator and the CU estimators. Section 6 proposes a finite sample variance correction. The next two sections apply our methods to the popular linear dynamic panel model and report a simulation evidence in the context of this model. Section 8 applies our methods to an empirical study on the causal effect of access to markets on household consumption in some rural Chinese areas. The last section concludes. Proofs are given in the appendix.

## 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  $\theta_0$  is assumed to be an interior point of parameter space  $\Theta \subseteq \mathbb{R}^d$ . The moment condition

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

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  $\Gamma = E[\partial f(Y_i, \theta_0)/\partial \theta']$  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  $\theta_0$  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  $\theta_0$  and  $\text{plim}_{N \rightarrow \infty} W_n = W > 0$ . In the context of instrumental variable (IV) regression, one popular choice for  $W_n$  is  $Z_n' Z_n / N$  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  $\tilde{\theta}$ ,  $\text{plim}_{N \rightarrow \infty} \hat{\Gamma}(\tilde{\theta}) = \Gamma$  and that  $\Gamma$  is of full column rank. Also, under some regularity conditions, we have the following Central Limit Theorem (CLT):

$$\begin{aligned} \sqrt{N} g_n(\theta_0) &\xrightarrow{d} N(0, \Omega) \text{ where} \\ \Omega &= \lim_{N \rightarrow \infty} \frac{1}{N} E \left( \sum_{i=1}^N f_i(\theta_0) \right) \left( \sum_{j=1}^N f_j(\theta_0) \right)'. \end{aligned}$$

Here  $\Omega$  is analogous to the long run variance in a time series setting but the components of  $\Omega$  are contributed by cross-sectional dependences over all locations. For easy reference, we follow Sun and Kim (2015) and call  $\Omega$  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  $\Gamma$  and  $W$  can be accurately estimated by  $\hat{\Gamma}(\hat{\theta}_1)$  and  $W_N$ , we need only estimate  $\Omega$  to make reliable inference about  $\theta_0$ . 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 not uncommon in many microeconomic 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  $L_g$ , i.e.,  $L = L_1 = L_2 = \dots = L_G$ . 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.,  $\lim_{N \rightarrow \infty} L_g / (G^{-1} \sum_{i=1}^G L_i) = 1$  for every  $g = 1, \dots, G$ . The following assumption formally characterizes the cluster dependence.

**Assumption 1.** (i) The data  $\{Y_j\}_{j=1}^N$  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  $\{f_i(\theta_0), i = 1, 2, \dots, N\}$  can be partitioned into  $G$  nonoverlapping clusters  $\cup_{g=1}^G \mathcal{G}_g$  where  $\mathcal{G}_g = \{f_k^g(\theta_0) : k = 1, \dots, L\}$ . In the context of this clustered structure, Assumption 1-ii) implies that the within-cluster dependence for each cluster can be arbitrary but  $E f_k^g(\theta_0) f_l^h(\theta_0) = 0$  if  $g \neq h$ . That is,  $f_k^g(\theta_0)$  and  $f_l^h(\theta_0)$  are independent if they belong to different clusters. 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

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

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

$$\Omega = \frac{1}{G} \sum_{g=1}^G \lim_{L \rightarrow \infty} \text{Var} \left( \frac{1}{\sqrt{L}} \sum_{i=1}^L f_i^g(\theta_0) \right) := \frac{1}{G} \sum_{g=1}^G \Omega_g.$$

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

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

To ensure that  $\hat{\Omega}(\hat{\theta}_1)$  is positive definite, we assume that  $G \geq m$ , and we 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. 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

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

where

$$\begin{aligned} & \widehat{\text{var}}(\hat{\theta}_1) \\ &= \frac{1}{N} \left[ \hat{\Gamma}(\hat{\theta}_1)' W_n^{-1} \hat{\Gamma}(\hat{\theta}_1) \right]^{-1} \left[ \hat{\Gamma}(\hat{\theta}_1)' W_n^{-1} \hat{\Omega}(\hat{\theta}_1) W_n^{-1} \hat{\Gamma}(\hat{\theta}_1) \right] \left[ \hat{\Gamma}(\hat{\theta}_1)' W_n^{-1} \hat{\Gamma}(\hat{\theta}_1) \right]^{-1}. \end{aligned}$$

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

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

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  $g = 1, \dots, G$ , let

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

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  $\theta_0$  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  $B_{m,g} \sim i.i.d. N(0, I_m)$  for  $g = 1, \dots, G$ , then

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

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

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

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

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  $L^{-1/2} \sum_{j=1}^L f_j^g(\theta_0)$  follows a central limit theorem jointly over  $g = 1, \dots, G$  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

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

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  $\Pi$ . 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:

$$\bar{B}_m = \begin{pmatrix} \bar{B}_d \\ \bar{B}_q \end{pmatrix}, \bar{B}_d = \begin{pmatrix} \bar{B}_p \\ \bar{B}_{d-p} \end{pmatrix}, \bar{\mathbb{S}} = \begin{pmatrix} \bar{\mathbb{S}}_{dd} & \bar{\mathbb{S}}_{dq} \\ \bar{\mathbb{S}}_{qd} & \bar{\mathbb{S}}_{qq} \end{pmatrix},$$

$$\bar{\mathbb{S}}_{dd} = \begin{pmatrix} \bar{\mathbb{S}}_{pp} & \bar{\mathbb{S}}_{p,d-p} \\ \bar{\mathbb{S}}_{d-p,p} & \bar{\mathbb{S}}_{d-p,d-p} \end{pmatrix}, \text{ and } \bar{\mathbb{S}}_{dq} = \begin{pmatrix} \bar{\mathbb{S}}_{pq} \\ \bar{\mathbb{S}}_{d-p,q} \end{pmatrix}.$$

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

- (a)  $F(\hat{\theta}_1) \xrightarrow{d} \mathbb{F}_{1\infty} := G\bar{B}_p' \bar{\mathbb{S}}_{pp}^{-1} \bar{B}_p/p;$
- (b)  $t(\hat{\theta}_1) \xrightarrow{d} \mathbb{T}_{1\infty} := \frac{N(0,1)}{\sqrt{\chi_{G-1}^2/G}}$  where  $N(0,1) \perp \sqrt{\chi_{G-1}^2}$ .

**Remark 1.** *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  $\mathbb{F}_{1\infty} \stackrel{d}{=} (G/G-p)\mathcal{F}_{p,G-p}$  where  $\mathcal{F}_{p,G-p}$  is a random variable that follows the  $F$  distribution with degree of freedom  $(p, G-p)$ . Similarly,  $\mathbb{T}_{1\infty} \stackrel{d}{=} (G/G-1)t_{G-1}$  where  $t_{G-1}$  is a random variable that follows the  $t$  distribution with degree of freedom  $G-1$ .*

**Remark 2.** *As an example of the general GMM setting, consider the linear regression model  $y_j = x_j'\theta + \epsilon_j$ . Under the assumption that  $\text{cov}(x_j, \epsilon_j) = 0$ , the moment function is  $f_j(\theta) = x_j(y_j - x_j'\theta)$ . With the moment condition  $E f_j(\theta_0) = 0$ , 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 3.** *Under the large- $G$  asymptotics where  $G \rightarrow \infty$  but  $L$  is fixed, one can show that the CCE  $\hat{\Omega}(\hat{\theta}_1)$  converges in probability to  $\Omega$  for*

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

*The convergence of  $\hat{\Omega}(\hat{\theta}_1)$  to  $\Omega$  does not require the homogeneity of  $\Omega_g$  in Assumption 6 (Hansen, 2007; Carter et al., 2013). Under this type of asymptotics, the test statistics  $F(\hat{\theta}_1)$  and  $t(\hat{\theta}_1)$  are asymptotically  $\chi_p^2/p$  and  $N(0,1)$ . Let  $\mathcal{F}_{p,G-p}^{1-\alpha}$  and  $\chi_p^{1-\alpha}$  be the  $1-\alpha$  quantiles of  $\mathcal{F}_{p,G-p}$  and the  $\chi_p^2$*

distributions, respectively. As  $G/(G-p) > 1$  and  $\mathcal{F}_{p,G-p}^{1-\alpha} > \chi_p^{1-\alpha}/p$ , it is easy to see that

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

However, the difference between the two critical values  $G(G-p)^{-1} \mathcal{F}_{p,G-p}^{1-\alpha}$  and  $\chi_p^{1-\alpha}/p$  shrinks to zero as  $G$  increases. Therefore, the small- $G$  critical value  $G(G-p)^{-1} \mathcal{F}_{p,G-p}^{1-\alpha}$  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 4.** Let  $\Lambda$  the matrix square root of  $\Omega$ , i.e.  $\Lambda\Lambda' = \Omega$ . 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

$$\begin{aligned} \Omega_{1\infty} &= \Lambda \tilde{\mathbb{D}} \Lambda' \text{ where } \mathbb{D} = \frac{1}{G} \sum_{g=1}^G \tilde{D}_g \tilde{D}_g' \\ \tilde{D}_g &= B_{m,g} - \Gamma_\Lambda (\Gamma_\Lambda' W_\Lambda^{-1} \Gamma_\Lambda)^{-1} \Gamma_\Lambda' W_\Lambda^{-1} \bar{B}_m \end{aligned} \quad (3)$$

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  $\Gamma_\Lambda (\Gamma_\Lambda' W_\Lambda^{-1} \Gamma_\Lambda)^{-1} \Gamma_\Lambda'$  depends on all of  $\Gamma, \Omega$  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

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

and thus

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

So, 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. There are 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

$$\begin{aligned}\hat{\Omega}(\theta) &= \frac{1}{G} \sum_{g=1}^G \left( \frac{1}{\sqrt{L}} \sum_{k=1}^L f_k^g(\theta) \right) \left( \frac{1}{\sqrt{L}} \sum_{l=1}^L f_l^g(\theta) \right)', \\ \hat{\Omega}^c(\theta) &= \frac{1}{G} \sum_{g=1}^G \left\{ \frac{1}{\sqrt{L}} \sum_{k=1}^L [f_k^g(\theta) - g_n(\theta)] \right\} \left\{ \frac{1}{\sqrt{L}} \sum_{l=1}^L [f_l^g(\theta) - g_n(\theta)] \right\}'.\end{aligned}$$

While  $\hat{\Omega}(\hat{\theta}_1)$  employs the uncentered moment process  $\{f_i^g(\hat{\theta}_1)\}_{i=1}^N$ ,  $\hat{\Omega}^c(\hat{\theta}_1)$  employs the recentered moment process  $\{f_i^g(\hat{\theta}_1) - g_n(\hat{\theta}_1)\}_{i=1}^G$ . 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  $W_n^{-1} \hat{\Gamma}(\hat{\theta}_1)$ . The two asymptotic variance estimators are not asymptotically equivalent by themselves under 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:

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

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 shown 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

$$\sqrt{N}(\hat{\theta}_2 - \theta_0) = - \left[ \Gamma' \hat{\Omega}^{-1}(\hat{\theta}_1) \Gamma \right]^{-1} \Gamma' \hat{\Omega}^{-1}(\hat{\theta}_1) \frac{1}{\sqrt{G}} \sum_{g=1}^G \left( \frac{1}{\sqrt{L}} \sum_{k=1}^L f_k^g(\theta_0) \right) + o_p(1). \quad (4)$$

Using the joint convergence of the following

$$\hat{\Omega}(\hat{\theta}_1) \xrightarrow{d} \Omega_{1\infty} = \Lambda \tilde{\mathbb{D}} \Lambda' \text{ and } \frac{1}{\sqrt{G}} \sum_{g=1}^G \left( \frac{1}{\sqrt{L}} \sum_{k=1}^L f_k^g(\theta_0) \right) \xrightarrow{d} \sqrt{G} \Lambda \bar{B}_m, \quad (5)$$

we obtain:

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

where as before

$$\tilde{\mathbb{D}} = \frac{1}{G} \sum_{g=1}^G \tilde{D}_g \tilde{D}'_g \text{ for } \tilde{D}_g = B_{m,g} - \Gamma_\Lambda (\Gamma'_\Lambda W_\Lambda^{-1} \Gamma_\Lambda)^{-1} \Gamma'_\Lambda W_\Lambda^{-1} \bar{B}_m.$$

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  $\tilde{\mathbb{S}}$  is partitioned. 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  $\Gamma_\Lambda$ . By construction,  $U'U = UU' = I_m$ ,  $V'V = VV' = I_d$ , and  $\Sigma' = \begin{bmatrix} A_{d \times d} & O_{d \times q} \end{bmatrix}$ . We then define  $\tilde{W} = U' W_\Lambda U$  and partition  $\tilde{W}$  as before. We also introduce

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

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,  $\kappa_G$  is the quadratic form of normal random vector  $\sqrt{G} \bar{B}_q$  with random matrix  $\mathbb{S}_{qq}$ . Finally, on the basis of  $\tilde{\theta}_2$ , the J-statistic for testing over-

identification restrictions is

$$J(\hat{\theta}_2) := Ng_n(\hat{\theta}_2)' \left( \hat{\Omega}(\hat{\theta}_1) \right)^{-1} g_n(\hat{\theta}_2)/q. \quad (6)$$

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 2.** *Let Assumptions 1~6 hold. Then*

- (a)  $\sqrt{N}(\hat{\theta}_1 - \theta_0) \xrightarrow{d} -VA^{-1}\sqrt{G}(\bar{B}_d - \beta_{\tilde{W}}\bar{B}_q)$ ;
- (b)  $\sqrt{N}(\hat{\theta}_2 - \theta_0) \xrightarrow{d} -VA^{-1}\sqrt{G}(\bar{B}_d - \beta_{\mathbb{S}}\bar{B}_q)$ ;
- (c)  $\sqrt{N}(\hat{\theta}_2 - \theta_0) \xrightarrow{d} -VA^{-1}\sqrt{G}(\bar{B}_d - \beta_{\mathbb{S}}\bar{B}_q) - VA^{-1}\sqrt{G}(\bar{B}_d - \beta_{\tilde{W}}\bar{B}_q) \cdot (\kappa_G/G)$ ;
- (d)  $\sqrt{N}(\hat{\theta}_2 - \theta_0) = \sqrt{N}(\tilde{\theta}_2 - \theta_0) + \sqrt{N}(\hat{\theta}_1 - \theta_0) \cdot (\kappa_G/G) + o_p(1)$ ;
- (e)  $J(\hat{\theta}_2) \xrightarrow{d} \kappa_G$  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 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<sup>2</sup>, it is straightforward to show

$$\kappa_G \stackrel{d}{=} \left( \frac{G}{q} \right) \frac{\frac{q}{G-q} \mathcal{F}_{q,G-q}}{1 + \frac{q}{G-q} \mathcal{F}_{q,G-q}}.$$

It is perhaps surprising that 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. For the J test at the significance level  $\alpha$ , say 5%, the critical value from  $\kappa_G$  can be obtained from

$$\left( \frac{G}{q} \right) \frac{\frac{q}{G-q} \mathcal{F}_{q,G-q}^{1-\alpha}}{1 + \frac{q}{G-q} \mathcal{F}_{q,G-q}^{1-\alpha}}.$$

Equivalently, we have

$$\frac{G-q}{q} \frac{q\kappa_G}{G-q\kappa_G} \stackrel{d}{=} \mathcal{F}_{q,G-q},$$

---

<sup>2</sup> $(C + ab')^{-1} = C^{-1} - \frac{C^{-1}ab'C^{-1}}{1+b'C^{-1}a}$  for any invertable square matrix  $C$  and conforming column vectors such that  $1 + b'C^{-1}a \neq 0$ .

and so

$$\tilde{J}(\hat{\theta}_2) := \frac{G-q}{q} \frac{qJ(\hat{\theta}_2)}{G-qJ(\hat{\theta}_2)} \xrightarrow{d} \mathcal{F}_{q,G-q}.$$

That is, the transformed J-statistic  $\tilde{J}(\hat{\theta}_2)$  is asymptotically F distributed. This is very convenient in empirical applications.

It is important to point out that the convenient F limit of  $J(\hat{\theta}_2)$  holds 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 may not be asymptotically pivotal any longer.

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

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

$$\widehat{var}_{\hat{\Omega}(\hat{\theta}_1)}(\hat{\theta}_2) = \frac{1}{N} \left( \hat{\Gamma}(\hat{\theta}_2)' \hat{\Omega}^{-1}(\hat{\theta}_1) \hat{\Gamma}(\hat{\theta}_2) \right)^{-1}.$$

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:

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

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  $(RVA^{-1})'RVA^{-1}$ .

**Proposition 3.** *Let Assumptions 1~6 hold. Then*

$$F_{\hat{\Omega}(\hat{\theta}_1)}(\hat{\theta}_2) \xrightarrow{d} \frac{G}{p} (\bar{B}_p - \mathbb{E}_{pq}\mathbb{E}_{qq}^{-1}\bar{B}_q)' (\mathbb{E}_{pp-q})^{-1} (\bar{B}_p - \mathbb{E}_{pq}\mathbb{E}_{qq}^{-1}\bar{B}_q)$$

$$= \frac{1}{p} \left[ G \begin{pmatrix} \bar{B}_p \\ \bar{B}_q \end{pmatrix}' \begin{pmatrix} \mathbb{E}_{pp} & \mathbb{E}_{pq} \\ \mathbb{E}'_{pq} & \mathbb{E}_{qq} \end{pmatrix}^{-1} \begin{pmatrix} \bar{B}_p \\ \bar{B}_q \end{pmatrix} - G\bar{B}'_q\mathbb{E}_{qq}^{-1}\bar{B}_q \right], \quad (7)$$

where

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

Due to the presence of the second term in  $\mathbb{E}_{p+q,p+q}$ , which depends on  $\tilde{\beta}_{\tilde{W}}^p$ , the F-statistic is not asymptotically pivotal. It depends on several nuisance parameters including  $\Omega$ . To see this, we note that the second term in (7) is the same as  $(G/q) \cdot \bar{B}'_q\mathbb{S}_{qq}^{-1}\bar{B}_q = \kappa_G$ . So the second term is the limit of the J-statistic, which is nuisance parameter free. However, the first term in (7) is not pivotal

because we have

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

where  $\tilde{w} = \left( (\tilde{\beta}_W^p)', I_q \right)'$ . 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  $R, \Gamma, W$  and  $\Omega$ .

### 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 (3), the effect of the estimator 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. So 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 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, recentering is known to have several advantages. For example, as Hall (2000) observes, in conventional increasing smoothing asymptotic theory, recentering can potentially improve the power of the J-test using a HAR variance estimator when the model is misspecified. Also Lee (2014) recently proposes a misspecification robust GMM bootstrap employing the recentered GMM weight matrix. In the context of fixed smoothing asymptotics, Sun (2014) shows that the recentering has a crucial role to yield an asymptotically pivotal inference from the two-step Wald test statistic.

In our small- $G$  asymptotic framework, 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 1.** *Let Assumptions 1~6 hold. Let  $\tilde{\theta}$  be any  $\sqrt{N}$  consistent estimator of  $\theta_0$ . Then*

- (a)  $\hat{\Omega}^c(\tilde{\theta}) = \hat{\Omega}^c(\theta_0) + o_p(1)$ ;
- (b)  $\hat{\Omega}^c(\theta_0) \xrightarrow{d} \Omega_\infty^c$  where  $\Omega_\infty^c = \Lambda \bar{\Sigma} \Lambda'$ .

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

1, it is possible to show

$$\sqrt{N}(\hat{\theta}_2^c - \theta_0) \xrightarrow{d} -[\Gamma'(\Omega_\infty^c)^{-1}\Gamma]^{-1}\Gamma'(\Omega_\infty^c)^{-1}\Lambda\sqrt{G}\bar{B}_m. \quad (8)$$

Since  $(\Omega_\infty^c)^{-1}$  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

$$F_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c) := (R\hat{\theta}_2^c - r)' \{R\widehat{var}_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c)R'\}^{-1}(R\hat{\theta}_2^c - r)/p \text{ where} \quad (9)$$

$$\widehat{var}_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c) = \frac{1}{N} \left( \hat{\Gamma}(\hat{\theta}_2^c)' \left( \hat{\Omega}^c(\hat{\theta}_2^c) \right)^{-1} \hat{\Gamma}(\hat{\theta}_2^c) \right)^{-1}.$$

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

$$t_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c) := \frac{(R\hat{\theta}_2^c - r)}{\{R\widehat{var}_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c)R'\}^{1/2}}.$$

The second test statistic is the Quasi-Likelihood Ratio (QLR) type of statistic. Define the restricted and centered two-step estimator  $\hat{\theta}_2^{c,r}$  :

$$\hat{\theta}_2^{c,r} = \arg \min_{\theta \in \Theta} g_n(\theta)' \left[ \hat{\Omega}^c(\hat{\theta}_1) \right]^{-1} g_n(\theta) \text{ s.t. } R\theta = r.$$

The QLR statistic is given by

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

The last test statistic we consider is the Lagrange Multiplier (LM) or score statistic in the GMM setting. Let  $\Delta_{\hat{\Omega}^c(\cdot)}(\theta)$  be the gradient of the GMM criterion function  $\hat{\Gamma}(\theta)' \left[ \hat{\Omega}^c(\cdot) \right]^{-1} g_n(\theta)$ , then the GMM score test statistic is given by

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

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 1 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  $\hat{\theta}_2^c$  :

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

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 (8) and Lemma 1, we have  $F_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c) \xrightarrow{d} \mathbb{F}_{2\infty}$  where

$$\begin{aligned} \mathbb{F}_{2\infty} &= G \left[ R (\Gamma'_\Lambda \bar{\mathbb{S}}^{-1} \Gamma_\Lambda)^{-1} \Gamma'_\Lambda \bar{\mathbb{S}}^{-1} \bar{B}_m \right]' \left[ R (\Gamma'_\Lambda \bar{\mathbb{S}}^{-1} \Gamma_\Lambda)^{-1} R' \right]^{-1} \\ &\quad \times \left[ R (\Gamma'_\Lambda \bar{\mathbb{S}}^{-1} \Gamma_\Lambda)^{-1} \Gamma'_\Lambda \bar{\mathbb{S}}^{-1} \bar{B}_m \right] / p. \end{aligned} \quad (10)$$

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

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

Also, it follows in a similar way that

$$\begin{aligned} J(\hat{\theta}_2^c) \xrightarrow{d} \mathbb{J}_\infty &:= G \left\{ \bar{B}_m - \Gamma_\Lambda (\Gamma'_\Lambda \bar{\mathbb{S}}^{-1} \Gamma_\Lambda)^{-1} \Gamma'_\Lambda \bar{\mathbb{S}}^{-1} \bar{B}_m \right\}' \bar{\mathbb{S}}^{-1} \\ &\quad \times \left\{ \bar{B}_m - \Gamma_\Lambda (\Gamma'_\Lambda \bar{\mathbb{S}}^{-1} \Gamma_\Lambda)^{-1} \Gamma'_\Lambda \bar{\mathbb{S}}^{-1} \bar{B}_m \right\} / q. \end{aligned} \quad (11)$$

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 4.** *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)  $F_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c) \xrightarrow{d} G (\bar{B}_p - \bar{\mathbb{S}}_{pq} \bar{\mathbb{S}}_{qq}^{-1} \bar{B}_q)' \bar{\mathbb{S}}_{pp\cdot q}^{-1} (\bar{B}_p - \bar{\mathbb{S}}_{pq} \bar{\mathbb{S}}_{qq}^{-1} \bar{B}_q)' / p \stackrel{d}{=} \mathbb{F}_{2\infty}$ ;
- (b)  $t_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c) \xrightarrow{d} \sqrt{G} (\bar{B}_p - \bar{\mathbb{S}}_{pq} \bar{\mathbb{S}}_{qq}^{-1} \bar{B}_q) / \sqrt{\bar{\mathbb{S}}_{pp\cdot q}} \stackrel{d}{=} \mathbb{T}_{2\infty}$  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)  $J(\hat{\theta}_2^c) \xrightarrow{d} (G/q) \bar{B}'_q \bar{\mathbb{S}}_{qq}^{-1} \bar{B}_q \stackrel{d}{=} \mathbb{J}_\infty$ .

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

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

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  $\bar{\mathbb{S}}_{qq}$ .<sup>3</sup> 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

$$\frac{G-p-q}{G} \mathbb{F}_{2\infty} \stackrel{d}{=} \frac{G-p-q}{G} \frac{(\sqrt{G} \bar{B}_p + \Delta)' \bar{\mathbb{S}}_{pp\cdot q} (\sqrt{G} \bar{B}_p + \Delta)}{p} \stackrel{d}{=} F_{p, G-p-q}(\|\Delta\|^2), \quad (12)$$

where  $F_{p, G-p-q}(\|\Delta\|^2)$  is a noncentral F distribution with random noncentrality parameter  $\|\Delta\|^2$ .

<sup>3</sup>See Proposition 7.9 in Bilodeau and Brenner.

Similarly, the limiting distribution  $\mathbb{T}_{2\infty}$  can be represented as

$$\sqrt{\frac{G-1-q}{G}}\mathbb{T}_{2\infty} \stackrel{d}{=} \sqrt{\frac{G-1-q}{G}} \frac{\sqrt{G}\bar{B}_p + \Delta}{\sqrt{\bar{\mathbb{S}}_{pp-q}}} \stackrel{d}{=} t_{G-1-q}(\Delta), \quad (13)$$

which is a noncentral t distribution with a noncentrality parameter  $\Delta$ . The 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 (12) and (13) 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 4-(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 (12) and (13). Similar results are obtained by Sun (2014) and Hwang and Sun (2015a; 2015b), which focus on 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 4-(e) that

$$\frac{G-q}{G}J(\hat{\theta}_2^c) \xrightarrow{d} \mathbb{J}_\infty \stackrel{d}{=} \frac{G-q}{G}\bar{B}'_q\bar{\mathbb{S}}_{qq}^{-1}\bar{B}_q \stackrel{d}{=} F_{q,G-q}.$$

This is consistent with Kim and Sun's (2012) results except that our adjustment and degrees of freedom parameter are different.

## 4 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 et al. (1996) and Newey and Smith (2000, 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,  $\hat{\theta}_{\text{IE}}$  can be regarded as a generalized-estimating-equations (GEE) estimator, which is a class of estimators first studied by Liang and Zeger (1986). 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

$$\hat{\Gamma}(\hat{\theta}_{\text{CU-GEE}})' \hat{\Omega}^{-1}(\hat{\theta}_{\text{CU-GEE}}) g_n(\hat{\theta}_{\text{CU-GEE}}) = 0. \quad (14)$$

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

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

where

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

Since  $1/(1 + \nu(\hat{\theta}_{\text{CU-GEE}}))$  is always positive, the first-order condition in (14) holds if and only if

$$\hat{\Gamma}(\hat{\theta}_{\text{CU-GEE}})' \left[ \hat{\Omega}^c(\hat{\theta}_{\text{CU-GEE}}) \right]^{-1} g_n(\hat{\theta}_{\text{CU-GEE}}) = 0 \quad (15)$$

which indicates that the recentering CCE weight in (14) 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:

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

Although we use the uncentered CEE  $\hat{\Omega}(\theta)$  in the above definition, the original definition of  $\hat{\theta}_{\text{CU}}$  in Hansen, Heaton and Yaron (1996) is based on the centered CCE weighting matrix  $\hat{\Omega}^c(\theta)$ . It is easy to show that

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

So we have

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

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  $\hat{\Omega}^c(\theta)$ , i.e.,

$$\hat{\theta}_{\text{CU}} = \arg \min_{\theta \in \Theta} g_n(\theta)' \left( \hat{\Omega}^c(\theta) \right)^{-1} (\theta) g_n(\theta). \quad (16)$$

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

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

The proposition shows that the CU estimators and the centered two-step GMM estimator are asymptotically equivalent under the small- $G$  asymptotics.

Construct the Wald statistics based on the two CU estimators as follows:

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

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

We construct  $t_{\hat{\Omega}^c(\hat{\theta}_{\text{CU-GEE}})}(\hat{\theta}_{\text{CU-GEE}})$  and  $t_{\hat{\Omega}^c(\hat{\theta}_{\text{CU}})}(\hat{\theta}_{\text{CU}})$  in a similar way when  $p = 1$ . It follows from Proposition 5 that the Wald statistics based on  $\hat{\theta}_{\text{CU-GEE}}$  and  $\hat{\theta}_{\text{CU-GMM}}$  are asymptotically equivalent to  $F_{\hat{\Omega}^c(\hat{\theta}_2)}(\hat{\theta}_2^c)$ . As a result,

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

Similarly,

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

In summary, we have shown that all three estimators  $\hat{\theta}_2^c$ ,  $\hat{\theta}_{\text{CU-GEE}}$  and  $\hat{\theta}_{\text{CU}}$ , and the corresponding Wald test statistics converge in distribution to the same nonstandard distributions. Proposition 4-(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.

The findings in this subsection 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 (2000, 2004) develops the stochastic expansion of CU estimators in the i.i.d setting and shows 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. See also Anatolyev (2005) which extends the work of Newey and Smith (2000, 2004) to a time series setting. 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.

## 5 Asymptotic F and t tests

Under the small- $G$  asymptotics, the limiting distributions of two-step test statistics, including Wald, QLR and LM, and the t statistics, are nonstandard and hence critical values have to be simulated in practice. This contrasts with the conventional large- $G$  asymptotics, where the limiting distributions are the standard chi-square 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

$$\begin{aligned}\tilde{F}_{\hat{\Omega}^c(\hat{\theta}_2^c)}(\hat{\theta}_2^c) &:= \frac{G-p-q}{G} \cdot \frac{F_{\hat{\Omega}^c(\hat{\theta}_2^c)}(\hat{\theta}_2^c)}{1 + \frac{q}{G}J(\hat{\theta}_2^c)}, \\ \widetilde{LR}_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c, \hat{\theta}_2^{c,r}) &:= \frac{G-p-q}{G} \cdot \frac{LR_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c, \hat{\theta}_2^{c,r})}{1 + \frac{q}{G}J(\hat{\theta}_2^c)}, \\ \widetilde{LM}_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^{c,r}) &:= \frac{G-p-q}{G} \cdot \frac{LM_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^{c,r})}{1 + \frac{q}{G}J(\hat{\theta}_2^c)},\end{aligned}\tag{17}$$

and the corresponding version of the t-statistic is

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

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 4 and Theorem 2 that

$$\left(F_{\hat{\Omega}^c(\hat{\theta}_2^c)}(\hat{\theta}_2^c), J(\hat{\theta}_2^c)\right) \xrightarrow{d} (\mathbb{F}_{2\infty}, \mathbb{J}_\infty)\tag{18}$$

$$\stackrel{d}{=} \left(G(\bar{B}_p - \bar{S}_{pq}\bar{S}_{qq}^{-1}\bar{B}_q)' \bar{S}_{pp\cdot q}^{-1}(\bar{B}_p - \bar{S}_{pq}\bar{S}_{qq}^{-1}\bar{B}_q)'/p, (G/q)\bar{B}_q'\bar{S}_{qq}^{-1}\bar{B}_q\right)\tag{19}$$

So

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

where

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

Similarly,

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

In the proof of Theorem 1 we show that  $\xi_p$  follows a standard normal distribution  $N(0, I_p)$

and that  $\xi_p$  is independent of  $\tilde{S}_{pp\cdot q}^{-1}$ . So 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  $\xi_p$  with an independent inverse-Wishart distributed weighting matrix  $\tilde{S}_{pp\cdot q}^{-1}$ . It follows from a theory of multivariate statistics that the limiting distribution of  $\tilde{F}_{\hat{\Omega}^c(\hat{\theta}_2^c)}(\hat{\theta}_2^c)$  is  $F_{p, G-p-q}$ . 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 1.** *Let Assumptions 1~6 hold. Then the modified Wald, QLR and LM all converge in distribution to  $F_{p, G-p-q}$ . Also, the  $t$  statistics has limiting distribution  $t_{G-1-q}$ .*

Together with the asymptotic equivalence between  $\hat{\theta}_2^c, \hat{\theta}_{\text{CU-GEE}}$  and  $\hat{\theta}_{\text{CU-GMM}}$  established in Proposition 5, the proof of Theorem 1 implies that the modified Wald, LR, LM, and  $t$  statistics based on  $\hat{\theta}_{\text{CU-GEE}}$  and  $\hat{\theta}_{\text{CU-GMM}}$  are all asymptotically  $F$  and  $t$  distributed under the small- $G$  asymptotics. This equivalence relationship is consistent with the recent paper by Hwang and Sun (2015b) which establishes the asymptotic  $F$  and  $t$  limit theory of two-step GMM in time series setting. But our cluster-robust limiting distributions in Theorem 1 are different from Hwang and Sun (2015b) in terms of the multiplicative adjustment and the degrees of freedom correction.

It follows from the proofs of Theorem 1 and Proposition 4 that

$$\begin{aligned} \sqrt{N}(\hat{\theta}_2^c - \theta_0) &\xrightarrow{d} MN\left(0, (\Gamma'\Omega^{-1}\Gamma)^{-1} \cdot (1 + \bar{B}'_q \bar{S}_{qq}^{-1} \bar{B}_q)\right) \\ \text{and } J(\hat{\theta}_2^c) &\xrightarrow{d} (G/q) \bar{B}'_q \bar{S}_{qq}^{-1} \bar{B}_q \end{aligned} \quad (20)$$

holds 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  $(1 + \bar{B}'_q \bar{S}_{qq}^{-1} \bar{B}_q)$  in (20) 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 (20) 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,

$$\frac{\sqrt{N}(\hat{\theta}_2^c - \theta_0)}{\sqrt{1 + (G/q)J(\hat{\theta}_2^c)}} \xrightarrow{d} N\left(0, (\Gamma'\Omega^{-1}\Gamma)^{-1}\right). \quad (21)$$

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 (21), the finite sample distribution of  $\sqrt{N}(\hat{\theta}_2^c - \theta_0)$  can be well-approximated by  $N(0, \widehat{\text{var}}_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c))$  where

$$\widehat{\text{var}}_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c) := \widehat{\text{var}}_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c) \cdot \left(1 + \frac{q}{G} J(\hat{\theta}_2^c)\right). \quad (22)$$

The modification term  $(1 + (q/G)J(\hat{\theta}_2^c))^{-1}$  degenerates to one as  $G$  increases so that the two variance estimates in (22) become close to each other. Thus, the multiplicative term  $(1 + (q/G)J(\hat{\theta}_2^c))^{-1}$  in (17) can be regarded as a finite sample modification to the standard variance estimate  $\widehat{\text{var}}_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c)$  under the large- $G$  asymptotics. For more discussions about the role of J-statistic modification, see Hwang and Sun (2015b) which casts the two-step GMM problems into OLS estimation and inference

in classical normal linear regression.

## 6 Finite Sample Variance Correction

### 6.1 Centered Two-step GMM Estimation

Define the infeasible two-step GMM estimator with the centered CCE weighting matrix  $\hat{\Omega}^c(\theta_0)$ :

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

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

$$\sqrt{N}(\hat{\theta}_2^c - \theta_0) = - \left[ \Gamma' \left( \hat{\Omega}^c(\hat{\theta}_1) \right)^{-1} \Gamma \right]^{-1} \Gamma' \left( \hat{\Omega}^c(\hat{\theta}_1) \right)^{-1} \sqrt{N}g_n(\theta_0) + o_p(1) \quad (23)$$

Together with Lemma 1, 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  $\xi_n \stackrel{a}{\sim} \eta_n$  for two stochastically bounded sequences of random vectors  $\xi_n \in \mathbb{R}^\ell$  and  $\eta_n \in \mathbb{R}^\ell$  when  $\xi_n$  and  $\eta_n$  converge in distribution to each other. Now under the small- $G$  asymptotics we have:

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

where

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

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

$$\sqrt{N}(\hat{\theta}_2^c - \theta_0) \stackrel{a}{\approx} - \left\{ \Gamma' \left[ \hat{\Omega}^c(\theta_0) \right]^{-1} \Gamma \right\}^{-1} \Gamma' \left[ \hat{\Omega}^c(\theta_0) \right]^{-1} \sqrt{N} g_n(\theta_0) + \mathcal{E}_{2n} \sqrt{N}(\hat{\theta}_1 - \theta_0). \quad (24)$$

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

$$\mathcal{E}_{2n}[\cdot, j] = - \left\{ \Gamma' \left[ \hat{\Omega}^c(\theta_0) \right]^{-1} \Gamma \right\}^{-1} \Gamma' \left[ \hat{\Omega}^c(\theta_0) \right]^{-1} \left. \frac{\partial \hat{\Omega}^c(\theta)}{\partial \theta_j} \right|_{\theta=\theta_0} \left[ \hat{\Omega}^c(\theta_0) \right]^{-1} g_n(\theta_0), \quad (25)$$

where

$$\begin{aligned}\frac{\partial \hat{\Omega}^c(\theta_0)}{\partial \theta_j} &= \Upsilon_j(\theta_0) + \Upsilon'_j(\theta_0) \text{ and} \\ \Upsilon_j(\theta_0) &= \frac{1}{G} \sum_{g=1}^G \left[ \frac{1}{\sqrt{L}} \sum_{r=1}^L \left( f_r^g(\theta_0) - \frac{1}{N} \sum_{s=1}^N f_s(\theta_0) \right) \right. \\ &\quad \left. \cdot \frac{1}{\sqrt{L}} \sum_{s=1}^L \left( \frac{\partial f_s^g(\theta_0)}{\partial \theta_j} - \frac{1}{N} \sum_{s=1}^N \frac{\partial f_s(\theta_0)}{\partial \theta_j} \right)' \right].\end{aligned} \quad (26)$$

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 (25) and (26) that  $\mathcal{E}_{2n} = O_p(N^{-1/2})$ .

It follows from (24) that

$$\sqrt{N}(\hat{\theta}_2^c - \theta_0) \stackrel{a}{\approx} - \left( \left[ \Gamma'(\Omega_\infty^c)^{-1} \Gamma \right]^{-1} \mathcal{E}_n (\Gamma' W^{-1} \Gamma)^{-1} \right) \begin{pmatrix} \Gamma'(\Omega_\infty^c)^{-1} \Lambda Z \\ \Gamma' W^{-1} \Lambda Z \end{pmatrix} \quad (27)$$

where  $Z \sim N(0, I_d)$ ,  $Z$  is independent of  $\Omega_\infty^c$ , and  $\mathcal{E}_n$  has the same marginal distribution as  $\mathcal{E}_{2n}$  but it is independent of  $Z$  and  $\Omega_\infty^c$ . 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

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



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

$$\begin{aligned}
\widehat{var}_{\hat{\Omega}^c(\hat{\theta}_1)}^{\text{adj}}(\hat{\theta}_2^c) &= \frac{1}{N} \widehat{\Xi}_n \\
&= \frac{1}{N} \left( \left[ \hat{\Gamma}' \left[ \hat{\Omega}^c(\hat{\theta}_1) \right]^{-1} \hat{\Gamma} \right]^{-1} \widehat{\mathcal{E}}_n (\hat{\Gamma}' W_n^{-1} \hat{\Gamma})^{-1} \right) \\
&\times \begin{pmatrix} \hat{\Gamma}' \left[ \hat{\Omega}^c(\hat{\theta}_1) \right]^{-1} \hat{\Gamma} & \hat{\Gamma}' W_n^{-1} \hat{\Gamma} \\ \hat{\Gamma}' W_n^{-1} \hat{\Gamma} & \hat{\Gamma}' W_n^{-1} \hat{\Omega}^c(\hat{\theta}_1) W_n^{-1} \hat{\Gamma} \end{pmatrix} \\
&\times \begin{pmatrix} \left[ \hat{\Gamma}' \left[ \hat{\Omega}^c(\hat{\theta}_1) \right]^{-1} \hat{\Gamma} \right]^{-1} \\ (\hat{\Gamma}' W_n^{-1} \hat{\Gamma})^{-1} \widehat{\mathcal{E}}_n' \end{pmatrix} \\
&= \widehat{var}_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c) + \widehat{\mathcal{E}}_n \widehat{var}_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c) + \widehat{var}_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c) \widehat{\mathcal{E}}_n' + \widehat{\mathcal{E}}_n \widehat{var}(\hat{\theta}_1) \widehat{\mathcal{E}}_n' \tag{28}
\end{aligned}$$

where

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

The last three terms in (28), which are of smaller order, serve as a finite sample correction to the original variance estimator.<sup>4</sup>

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 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 statistic:

$$F_{\hat{\Omega}^c(\hat{\theta}_1)}^{\text{adj}}(\hat{\theta}_2^c) = (R\hat{\theta}_2^c - r)' \left[ R \widehat{var}_{\hat{\Omega}^c(\hat{\theta}_1)}^{\text{adj}}(\hat{\theta}_2^c) R' \right]^{-1} (R\hat{\theta}_2^c - r)/p. \tag{29}$$

When  $p = 1$  and for one-sided alternative hypotheses, we can construct the variance-corrected

---

<sup>4</sup>Note that the corrected variance estimator is not necessarily larger than the original estimator in finite samples. In the simulation work we consider later, 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 adjustment to  $\widehat{var}_{\hat{\Omega}^c(\hat{\theta}_1)}^{\text{adj}}(\hat{\theta}_2^c)$  so that  $\widehat{var}_{\hat{\Omega}^c(\hat{\theta}_1)}^{\text{adj}}(\hat{\theta}_2^c) - \widehat{var}_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c)$  is guaranteed to be positive semidefinite. The adjustment is in a similar spirit with Politis (2011).

t-statistic:

$$t_{\hat{\Omega}^c(\hat{\theta}_1)}^{\text{adj}}(\hat{\theta}_2^c) = \frac{(R\hat{\theta}_2^c - r)}{\sqrt{R\widehat{\text{var}}_{\hat{\Omega}^c(\hat{\theta}_1)}^c(\hat{\theta}_2^c)}R'} \quad (30)$$

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  $g = 1, \dots, G$  and  $s = 1, \dots, d$ , define  $Q_s^g(\theta)$  as

$$Q_s^g(\theta) = \lim_{L \rightarrow \infty} E \left[ \frac{1}{L} \sum_{k=1}^L \frac{\partial}{\partial \theta'} \left( \frac{\partial f_k^g(\theta)}{\partial \theta_s} \right) \right]$$

Then,

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

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

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

**Theorem 2.** Let Assumptions 1~7 hold. Then

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

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

As a direct implication of Theorem 2, the small- $G$  asymptotic distributions of  $F_{\hat{\Omega}^c(\hat{\theta}_1)}^c(\hat{\theta}_2^c)$  and  $t_{\hat{\Omega}^c(\hat{\theta}_1)}^c(\hat{\theta}_2^c)$  are

$$F_{\hat{\Omega}^c(\hat{\theta}_1)}^{\text{adj}}(\hat{\theta}_2^c) \xrightarrow{d} \mathbb{F}_{2\infty} \text{ and } t_{\hat{\Omega}^c(\hat{\theta}_1)}^{\text{adj}}(\hat{\theta}_2^c) \xrightarrow{d} \mathbb{T}_{2\infty}.$$

## 6.2 CU Estimation

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

$$\begin{aligned} &\sqrt{N}(\hat{\theta}_{\text{CU-GEE}} - \theta_0) \\ &= - \left( \Gamma' \left( \hat{\Omega}^c(\theta_0) \right)^{-1} \Gamma \right)^{-1} \Gamma' \left( \hat{\Omega}^c(\theta_0) \right)^{-1} \sqrt{N}g_n(\theta_0) + \mathcal{E}_{2n} \sqrt{N}(\hat{\theta}_{\text{CU-GEE}} - \theta_0) + o_p(1). \end{aligned} \quad (31)$$

This can be regarded as a special case of (24) wherein the first-step estimator  $\hat{\theta}_1$  is replaced by the CU-GEE estimator. So

$$\sqrt{N}(\hat{\theta}_{\text{CU-GEE}} - \theta_0) \overset{a}{\approx} - (I_d - \mathcal{E}_{2n})^{-1} \left( \Gamma' \left( \hat{\Omega}^c(\theta_0) \right)^{-1} \Gamma \right)^{-1} \Gamma' \left( \hat{\Omega}^c(\theta_0) \right)^{-1} \sqrt{N}g_n(\theta_0). \quad (32)$$

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

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

where

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

and  $\hat{\mathcal{E}}_{\text{CU-GMM}}$  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 test statistics based on all three estimators  $\hat{\theta}_2^c$ ,  $\hat{\theta}_{\text{CU-GEE}}$  and  $\hat{\theta}_{\text{CU-GMM}}$  converge in distribution to the same nonstandard distributions. A multiplicative modification provided in Section 5 can then turn the nonstandard distributions  $\mathbb{F}_{2\infty}$  and  $\mathbb{T}_{2\infty}$  into standard  $F$  and  $t$  distributions.

## 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.

$$y_{it} = \gamma y_{it-1} + x_{1,it} \beta_1 + x_{2,it} \beta_2 + x_{3,it} \beta_3 + \eta_i + u_{it}.$$

The unknown parameter vector is  $\theta = (\gamma, \beta_1, \beta_2, \beta_3)$  and the corresponding covariates are  $w_{it} = (y_{it-1}, x_{it})'$  with  $x_{it} = (x_{1,it}, x_{2,it}, x_{3,it})'$ . The true value of  $\theta$  is chosen as  $\theta_0 = (0.5, 1, 1, 1)$ . We denote  $s_{it}^g = (s_{1,it}^g, \dots, s_{k,it}^g)'$  as any vector valued observations in cluster  $g$ , and stack all observations at same period by cluster to define  $s_{(g),t} = (s_{1t}^g, \dots, s_{Lt}^g)'$ . The  $j$ -th predetermined regressor  $x_{j,it}^g$  are generated according to the following process:

$$x_{j,it}^g = \rho x_{j,it-1}^g + \eta_i^g + \rho u_{it-1}^g + e_{j,it}^g$$

for  $j = 1, 2, 3$ ,  $i = 1, \dots, L$ , and  $t = 1, \dots, T$ . 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  $\Sigma_\eta$  and  $\Sigma_u$  to be  $L \times L$  matrices whose  $(i, j)$ -th elements are  $\sigma_{ij}^\eta = \lambda^{|i-j|}$  and  $\sigma_{ij}^u = \lambda^{|i-j|}$ , respectively, and  $\Sigma_e$  to be a  $3L \times 3L$  block diagonal matrix with diagonal matrix  $\Sigma_{d,e}$  of size  $L \times L$  for  $d = 1, \dots, 3$ . The  $(i, j)$ -th element of  $\Sigma_{d,e}$  is  $\sigma_{d,ij}^e = \lambda^{|i-j|}$  for  $d = 1, \dots, 3$ . The parameter  $\lambda$  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 static panel data model with only one regressor.

The individual fixed effects and shocks in group  $g$  are generated by:

$$\begin{aligned}\eta_{(g)} &\sim \text{i.i.d.}N(0, \Sigma_\eta), \text{vec}(e_{(g),t}) \sim \text{i.i.d.}N(0, \Sigma_e), \\ u_{(g),t} &= \tau_t \Sigma_u^{1/2} (\delta_1^g \omega_{1t}^g, \dots, \delta_L^g \omega_{Lt}^g)', \\ \delta_i^g &\sim \text{i.i.d.}U[0.5, 1.5], \text{ and } \omega_{it}^g \sim \text{i.i.d.}\chi_1^2 - 1\end{aligned}\tag{33}$$

for  $i = 1, \dots, L_N$  and  $t = 1, \dots, T$  where  $\tau_t = 0.5 + 0.1(t - 1)$ . The DGP of individual shock  $u_{(g),t}$  in (33) features a non-Gaussian process which is heteroskedastic over both time  $t$  and individual  $i$ . Also, the clustered dependence structure implies

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

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

Before we draw an estimation sample for  $t = 1, \dots, T$ , 50 initial values are generated with  $\tau_t = 0.5$  for  $t = -49, \dots, 0$ ,  $x_{d,i,-49}^g \sim \text{i.i.d.}N(\eta_i^g / (1 - \rho), (1 - \rho)^{-1} \Sigma_{d,e})$  for  $d = 1, \dots, 3$ , and  $y_{i,-49}^g = (\sum_{d=1}^3 x_{d,i,-49} \beta_d + \eta_i^g + u_{i,-49}^g) / (1 - \gamma)$ . We fix the values of  $\lambda$  and  $\rho$  at 0.75 and 0.70, respectively; thus each observation is reasonably persistent with respect to both time and spatial dimensions. We set the number of time periods to be  $T = 4$ . In Appendix, we describe how to implement our new inference procedures in the context of linear dynamic panel model. The parameters are estimated by the first differenced GMM (AB estimator). The initial first-step estimator with  $W_n = (1/N) \sum_{i=1}^N Z_i' H Z_i$  where  $H$  is a matrix that consists with 2's on the main diagonal, with -1's on the main diagonal, and zeros elsewhere. With all possible lagged instruments, the number of moment conditions for the AB estimator is  $dT(T - 1)/2 = 24$  and the degrees of over-identification is  $q = 20$ . 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  $d(T - 1) = 12$  moment conditions.

## 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 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 are  $\chi_p^{1-\alpha} / p$  and  $\mathcal{F}_{p, G-p}^{1-\alpha}$ , respectively. 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 five different procedures. The first three tests use different test statistics but the same critical value  $\chi_p^{1-\alpha} / p$ . The first test

uses the “plain” F-statistic  $F_2 := F_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c)$  as defined in (9). The second test uses the statistic  $[(G - p - q)/G] \cdot F_2$  where  $(G - p - q)/G$  is the degree-of-freedom correction factor. The third test uses  $\tilde{F}_2 := F_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c)$  as defined in (17). Note that

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

Compared to the second test statistic,  $\tilde{F}_2$  has the additional J-statistic correction factor  $(1 + (q/G)J(\hat{\theta}_2^c))^{-1}$ . The three tests use increasingly more sophisticated test statistics. Because  $[(G - p - q)/G] \rightarrow 1$  and  $(1 + (q/G)J(\hat{\theta}_2^c))^{-1} \rightarrow 1$  as  $G \rightarrow \infty$ , both corrections can be regarded as devices for finite sample improvement under the large- $G$  asymptotics.

Table 1: Summary of the first-step, two-step tests, and J test

<b>First-step GMM tests</b>					
statistic	d.f.	adj.	$\mathcal{F}_{p,G-p}^{1-\alpha}$	.	.
$F_1$	—	—	—	—	—
$\frac{G-p}{G}F_1$	Y	Y	Y	—	—
<b>Two-step GMM tests</b>					
statistic	d.f.	adj.	J-modification	$\mathcal{F}_{p,G-p-q}^{1-\alpha}$	variance correction
$F_2$	—	—	—	—	—
$\frac{G-p-q}{G}F_2$	Y	Y	—	—	—
$\tilde{F}_2$	Y	Y	Y	—	—
$\tilde{F}_2^{\text{adj}}$	Y	Y	Y	Y	—
$\tilde{F}_2^{\text{adj}+}$	Y	Y	Y	Y	Y
<b>J-tests</b>					
statistic	d.f.	adj.	$\mathcal{F}_{q,G-q}^{1-\alpha}$	.	.
$J$	—	—	—	—	—
$\frac{G-q}{G}J^c$	Y	Y	Y	—	—

Notes: The first-step tests are based on the first-step GMM estimator  $\hat{\theta}_1$ . They use the associated F-statistic  $F_1 = F_1(\hat{\theta}_1)$  with critical value  $\chi_p^{1-\alpha}/p$  or  $F_{p,G-p}^{1-\alpha}$ . The first  $J$  test employs the statistic  $J(\hat{\theta}_2)$  and critical value  $\chi_q^{1-\alpha}$ , and the second J test employs the statistic  $\frac{G-q}{G}J^c = \frac{G-q}{G}J(\hat{\theta}_2^c)$  and critical value  $\mathcal{F}_{q,G-q}^{1-\alpha}$ . All two-step tests are based on the centered two-step GMM estimator  $\hat{\theta}_2^c$  but use different test statistics and critical values: the unmodified F-statistic  $F_2 = F_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c)$ , J-statistic and degrees-of-freedom corrected statistic  $\tilde{F}_2 = \tilde{F}_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c)$ , and J-statistic, degrees-of-freedom, and finite-sample-variance corrected F-statistic  $\tilde{F}_2^{\text{adj}+} = \tilde{F}_{\hat{\Omega}^c(\hat{\theta}_1)}^{\text{adj}+}(\hat{\theta}_2^c)$ , coupled with critical value  $\chi_p^{1-\alpha}/p$  or  $\mathcal{F}_{p,G-p-q}^{1-\alpha}$ .

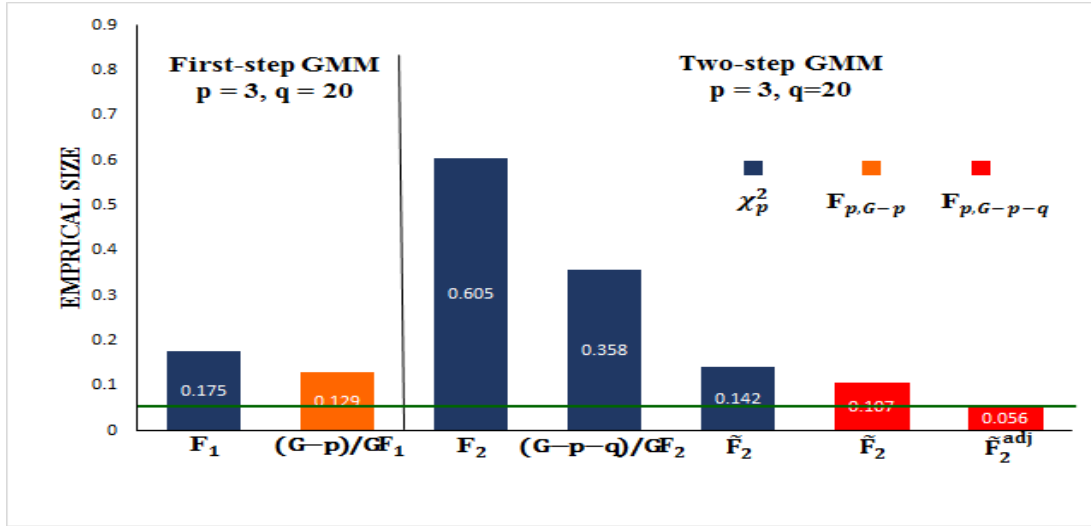


Figure 1: Empirical size of the first-step and two-step tests when  $G = 50$ ,  $L = 50$ ,  $q = 20$ , and  $p = 3$  with the nominal size 5% (green line).

The last two tests employ the new F critical values which are justified from the small- $G$  asymptotics. The fourth test uses the same statistic  $\tilde{F}_2$  as the third test but employs the F critical value  $\mathcal{F}_{p,G-p-q}^{1-\alpha}$  instead. The fifth test uses the most refined version of the F-statistic  $\tilde{F}_2^{\text{adj}}$  as defined in (29) and the same F critical value  $\mathcal{F}_{p,G-p-q}^{1-\alpha}$  employed in the fourth test. A summary of the tests considered is given in Table 1.

## 7.3 Results

### 7.3.1 Balanced Cluster Size

We first consider the case when all clusters have an equal number of individuals and take different values of  $G \in \{30, 35, 50, 100\}$  and  $L \in \{50, 100\}$ . 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%. The number of simulation replications is 5000.

Tables 2~5 report the empirical size of the first-step and two-step tests for different values of  $G \in \{30, 35, 50, 100\}$  and  $L = \{50, 100\}$ . The results 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-square critical values are used. For example, with  $G = 50$ ,  $L = 50$ , and  $p = 3$ ,

the empirical size of the first-step chi-square test (using the full set of IVs, and  $m = 24$ ) is around 43%. This size distortion becomes more severe, as the number of clusters becomes smaller, say, for example when  $G$  is between 30 and 35. The empirical size of the first-step F test with  $G = 50$  reduces to 36.3% when the F critical value is 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 some exactly identified models. Tables 2~5 also indicate that the finite sample size distortion of all tests become less severe as the number of moment conditions decreases or the cluster size increases.

For the two-step test that employs the plain two-step statistic  $F_2$  and chi-squared critical value, the empirical size is 63.4% for the above mentioned values of  $L, G, m$ , and  $p$ . In view of the large size distortion, we can conclude that the two-step chi-square test suffers more size distortion than the first-step chi-square test. This relatively large size distortion reflects the additional cost in estimating the weighting matrix, which is not captured by the chi-square approximation. The degrees-of-freedom adjusted  $F_2$  reduces the size distortion by almost one third, but the empirical size of 40.1% is still far away from the nominal size of 5%. This motivates us to implement an additional correction via the J-statistic multiplier coupled with the new critical value  $\mathcal{F}_{p, G-p-q}^{1-\alpha}$ . Tables 2~5 show that using the additional modification and the F critical value significantly alleviates the remaining size distortion. The size distortion in the previous example becomes 13.5% which is much closer to the targeted level 5%. Lastly, we find evidence that the most refined statistic  $\tilde{F}_2^{\text{adj}+}$ , equipped with the finite sample variance correction, successfully captures the higher order estimation uncertainty and yields more accurate finite sample size. For instance, while the empirical size of the most basic two-step chi-square test is 63.4%, the empirical size of the most refined two-step F test is 5.7%, which is very close to the nominal size of 5%. Figure ?? summarizes the outstanding performance of our modified two-step tests with F critical values.

Next we investigate the finite sample power performances of the first-step procedure and the two-step procedures  $F_2$ ,  $\tilde{F}_2$ , and  $\tilde{F}_2^{\text{adj}}$ . 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  $d = 2$  and  $p = 1$ . Figures 2~5 report the power curves for the first-step and two-step tests for  $G \in \{30, 35, 50, 100\}$ . The degree of over-identification  $q$  considered here is 10 for the full instrument set, and is 4 for the reduced instrument set. 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  $\tilde{F}_2$ .

Figures 2~5 also indicate that two-step tests are more powerful than first-step tests. The power gain of the two-step procedures becomes more significant as the number of  $G$  increases. This is because the two-step estimator becomes more efficient. However, there is a cost in estimating the CCE weighting matrix, the power of first-step procedures might dominate the power of the two-step

ones in other scenarios, i.e., when the cost of employing CCE weighting matrix outweighs the benefit of estimating it. Some simulation results not reported here show that the power of the first-step test can be higher than that of a two-step test when the number of parameters  $d$  and the number of joint hypotheses  $p$  are large.

Lastly, Tables 2~5 show that the finite sample size distortion of the (centered) J test and the transformed (uncentered) J test is substantially reduced when we employ F critical values instead of conventional chi-squared critical values.

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.

### 7.3.2 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 (2013) which investigates the behavior of cluster-robust t-statistic under cluster heterogeneity. Table 6 describes the heterogeneous cluster-size designs we consider. All other parameter values are the same as before.

Tables 7~8 report the empirical size of the first-step, two-step, and J tests for  $q = 20$  and  $p = 3$ . The results immediately indicate that the two-step tests suffer from severe size distortion when the conventional chi-square critical value is employed. For example, under design II, the empirical size of the “plain” two-step chi-square test is around 60.4% for  $G = 50$ ,  $q = 20$ , and  $p = 3$ . The size distortion become more severe when the degree of heterogeneity across cluster-size increases. However, our small- $G$  asymptotics still performs very well as they reduce the empirical sizes to 4.3% ~ 7.9%, which are much closer to the nominal size of 5%. Figures ??~?? summarize the outstanding performance of our modified two-step F tests, even with unbalanced cluster sizes. The results of J tests are omitted here as they are qualitatively similar to those of the F tests.

## 8 Empirical Application

In this section we employ the proposed procedures to revisit the study of Emran and Hou (2013), which investigates the casual effects of access to domestic and international markets on household



consumption in rural China. They use a survey data of 7998 rural households across 19 provinces in China. The survey data comes from Chinese Household Income Project (ICPSR 3012) in 1995.<sup>5</sup>

The regression equation for per capita consumption for household  $i$ ,  $C_i$ , in 1995 (yuan) is specified as

$$C_i = \alpha_0 + \alpha_p + \beta_d A_i^d + \beta_s A_i^s + \beta_{ds} (A_i^d \times A_i^s) + X_i' \beta_h + X_i^v' \beta_v + X_i^c' \beta_c + \epsilon_i, \quad (34)$$

where  $A_i^d$  and  $A_i^s$  are the log distances of access to domestic (km) and international markets (km), respectively.  $X_i$  is the vector of household characteristics that may affect consumption choice, and  $X_i^v$ ,  $X_i^c$  are village, county level controls, respectively, which capture the heterogeneity in economic environments across different regions, and  $\alpha_p$  is the province level fixed effect.

Among the unknown parameters in vector  $\theta = (\alpha_0, \alpha_p, \beta'_m, \beta'_h, \beta'_v, \beta'_c)'$ , our focus of interest is  $\beta_m = (\beta_d, \beta_s, \beta_{ds})'$  which measures the casual effect of access to domestic and international markets on household consumption in the rural areas. To identify these parameters, Emran and Hou (2013) employs geographic instrumental variables that capture exogenous variations in access to markets, e.g., straight-line distances to the nearest navigable river and coastline, along with the topographic and agroclimatic features of the counties.<sup>6</sup> There are 21 instrument variables and 31 control variables, including province dummy variables so that the number of moment conditions  $m$  is 52. The number of estimated parameters  $d$  is 34, and the degree of over-identification  $q$  is 18. Because of the close economic and cultural ties within the same county in rural Chinese areas, the study clusters the data by the county level and estimates the model using 2SLS and two-step GMM with uncentered cluster-robust weighting matrix. The data set consists of 7462 observations divided into 86 clusters where the number of households vary across from a low of 49 to a high of 270. Statistical inferences in Emran and Hou (2013) are conducted using the large- $G$  asymptotics only. We apply our more accurate asymptotics to Emran and Hou's study. The inference methods we use here are described in Tables 9 and 10 which present the test statistics, the reference distribution, and the standard error formula (finite sample corrected or not) for each method. Here we view all corrections, including the degree-of-freedom correction, the J correction, and the finite sample variance corrections as corrections to the variance estimator.

Table 11 shows the point estimation results, standard error estimates, and associated confidence intervals (CIs) for each of 2SLS and the uncentered and centered two-step GMM estimators. Similar to Emran and Hou (2013), our results show that the better access to domestic and international markets has a substantial positive effect on household consumption, and that the domestic market effect is significantly higher. For the 2SLS method, there are no much differences in confidence interval and standard error between the large- $G$  and small- $G$  results. This is well expected because the number of clusters  $G = 86$  is large enough so that the large- $G$  and small- $G$  approximations are

<sup>5</sup>The data set is downloadable from the *Review of Economics and Statistics* website.

<sup>6</sup>For the detailed description of the control variables and instrument variables, see the appendix in Emran and Hou (2013).

close to each other.

The uncentered two-step GMM estimate of the effect of access to domestic market is  $\beta_d = -2722.22$ . The reported standard error 400.5 is about 40% smaller than that of 2SLS. However, the plain two-step standard error estimate might be downward biased because the variation of the cluster-robust weighting matrix is not considered. The centered two-step GMM estimator gives a smaller effect of market access  $\beta_d = -2670.0$  with the modified standard error of 519.2, which is 25% larger than the plain two-step standard error. However, the modified standard error is still smaller than that based on the 2SLS method. So the two-step estimator still enjoys the benefit of using the cluster-robust weighting matrix. The results for other parameters deliver similar qualitative messages. Table 11 also provides the finite sample corrected standard error estimates of two-step estimators that capture the extra variation of feasible CCE, leading to slightly larger standard errors and wider CIs than the uncorrected ones. Overall, our results suggest that the effect of access to markets may be lower than the previous finding after we take into account the randomness of the estimated optimal GMM weighting matrix.

## 9 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 (2015a) compares 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 the spatial setting to future research.

## References

- [1] Anatolyev, S. (2005). “GMM, GEL, serial correlation, and asymptotic bias.” *Econometrica*, 73(3), 983-1002.
- [2] Anderson, T. W. and Hsiao, C. (1981): “Estimation of dynamic models with error components.” *Journal of the American statistical Association*, 76(375), 598-606.
- [3] Andrabi, T., Das, J., Khwaja, A. I., and Zajonc, T. (2011). “Do value-added estimates add value? Accounting for learning dynamics.” *American Economic Journal: Applied Economics*, 29-54.
- [4] Arellano, M. (1987). “Computing Robust Standard Errors for Within-Group Estimators.” *Oxford Bulletin of Economics and Statistics*, 49, 431-434.
- [5] 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.
- [6] Bharath, S., Dittmar, A., and Sivadasan, J. (2013). “Do Going Private Transactions Affect Plant Productivity?.” *Review of Financial Studies*, forthcoming.
- [7] Bell, R. M., and McCaffrey, D. F. (2002). “Bias reduction in standard errors for linear regression with multi-stage samples.” *Survey Methodology*, 28(2), 169-182.
- [8] 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.
- [9] 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.
- [10] Bilodeau, M., and Brenner, D. (2008). “Theory of multivariate statistics.” Springer Science & Business Media.
- [11] Canay, I. A., Romano, J. P., and Shaikh, A. M. (2014). “Randomization tests under an approximate symmetry assumption.” manuscript, July, 20.
- [12] 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.
- [13] Cameron, A. C. and Miller, D. L. (2015): “A practitioner’s guide to cluster-robust inference.” *Journal of Human Resources*, 50(2), 317-372.
- [14] De Brauw, A., and Giles, J. (2012): “Migrant labor markets and the welfare of rural households in the developing world: Evidence from China.” Working paper.

- [15] Dube, A., Lester, T. W., and Reich, M. (2010). “Minimum wage effects across state borders: Estimates using contiguous counties.” *Review of economics and statistics*, 92(4), 945-964.
- [16] 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.
- [17] 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.
- [18] Hansen, L. P. (1982): “Large Sample Properties of Generalized Method of Moments Estimators.” *Econometrica* 50, 1029–1054.
- [19] Hayakawa, K. (2014): “Alternative Over-identifying Restriction Test in GMM with Grouped Moment Conditions.” Working paper.
- [20] 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.
- [21] Hwang, J. and Y. Sun (2015a): “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.
- [22] Hwang, J. and Y. Sun.(2015b): “Asymptotic F and t Tests in an Efficient GMM Setting.” Working paper, Department of Economics, UC San Diego.
- [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., (2012). “Robust standard errors in small samples: Some practical advice.” *Review of Economics and Statistics* , (0).
- [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., & 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] Lee, S. (2014). “Asymptotic refinements of a misspecification-robust bootstrap for generalized method of moments estimators.” *Journal of Econometrics*, 178, 398-413.
- [32] Liang, K.-Y. and Zeger, S. (1986). “Longitudinal Data Analysis Using Generalized Linear Models.” *Biometrika* 73(1):13–22.
- [33] MacKinnon, James G., and Matthew D. Webb. "Wild bootstrap inference for wildly different cluster sizes." *Journal of Applied Econometrics* (2016).
- [34] Moulton, B. R. (1990): “An illustration of a pitfall in estimating the effects of aggregate variables on micro units.” *Review of Economics and Statistics*, 334-338.
- [35] Müller, U. K. (2007). “A theory of robust long-run variance estimation.” *Journal of Econometrics*, 141(2), 1331-1352.
- [36] Newey, Whitney K., and Richard J. Smith. (2000). “Asymptotic bias and equivalence of GMM and GEL estimators.” Mimeo, MIT.
- [37] Newey, Whitney K., and Richard J. Smith. (2004) “Higher order properties of GMM and generalized empirical likelihood estimators.” *Econometrica* 72, no. 1 : 219-255.
- [38] Nickell, S. (1981). “Biases in dynamic models with fixed effects.” *Econometrica: Journal of the Econometric Society*, 1417-1426.
- [39] Obstfeld, M., Shambaugh, J. C., and Taylor, A. M. (2008). “Financial stability, the trilemma, and international reserves (No. w14217).” National Bureau of Economic Research.
- [40] Pepper, J. V. (2002). “Robust inferences from random clustered samples: an application using data from the panel study of income dynamics.” *Economics Letters*, 75(3), 341-345.
- [41] Politis, D. N. (2011). “Higher-order accurate, positive semidefinite estimation of large-sample covariance and spectral density matrices.” *Econometric Theory*, 27(04), 703-744.
- [42] Shah, B. V., Holt, M. M., and Folsom, R. E. (1977). “Inference about regression models from sample survey data.” *Bulletin of the International Statistical Institute*, 47(3), 43-57.
- [43] Samila, S., and Sorenson, O. (2011). “Venture capital, entrepreneurship, and economic growth. *The Review of Economics and Statistics.*”, 93(1), 338-349.
- [44] Stock, J. H. and Watson, M. W. (2008), “Heteroskedasticity-Robust Standard Errors for Fixed Effects Panel Data Regression.” *Econometrica*, 76: 155–174. doi: 10.1111/j.0012-9682.2008.00821.x

- [45] Sun, Y. (2013): “A Heteroskedasticity and Autocorrelation Robust F Test Using Orthonormal Series Variance Estimator.” *Econometrics Journal* 16(1), 1–26.
- [46] Sun, Y. (2014): “Fixed-smoothing Asymptotics in a Two-step GMM Framework.” *Econometrica* 82(6), 2327–2370.
- [47] Sun, Y., and Kim, M. S. (2012). “Simple and powerful GMM over-identification tests with accurate size.” *Journal of Econometrics*, 166(2), 267-281.
- [48] 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.
- [49] Sun, Y., Phillips, P. C. B. and Jin, S. (2008). “Optimal Bandwidth Selection in Heteroskedasticity-Autocorrelation Robust Testing.” *Econometrica* 76(1), 175–94.
- [50] Webb, M. D. (2014): “Reworking Wild Bootstrap Based Inference for Clustered Errors,” Queen’s Economics Department Working Paper.
- [51] White, H. (1984). “Asymptotic theory for econometricians.” Academic press.
- [52] Windmeijer, F. (2005): “A finite sample correction for the variance of linear efficient two-step GMM estimators.” *Journal of Econometrics*, 126(1), 25–51.
- [53] Wooldridge, J. M. (2003): “Cluster-sample methods in applied econometrics.” *American Economic Review*, 133–138.

Table 2: Empirical size of first-step and two-step tests based on the centered CCE when  $L_N = 50$ , number of clusters  $G = 30$  and  $35$ , number of joint hypothesis  $p = 1 \sim 3$  and number of moment conditions  $m = 12, 24$ .

		Test	$m = 24$			$m = 12$		
$G = 30$	statistic	critical values	$p = 1$	$p = 2$	$p = 3$	$p = 1$	$p = 2$	$p = 3$
first-step	$F_1$	$\chi_p^{1-\alpha}/p$	0.374	0.475	0.528	0.274	0.302	0.332
	$\frac{G-p}{G}F_1$	$\mathcal{F}_{p,G-p}^{1-\alpha}$	0.333	0.389	0.417	0.236	0.232	0.220
two-step	$F_2$	$\chi_p^{1-\alpha}/p$	0.659	0.857	0.939	0.308	0.415	0.492
	$\frac{G-p-q}{G}F_2$	$\chi_p^{1-\alpha}/p$	0.414	0.562	0.658	0.229	0.276	0.308
	$\tilde{F}_2$	$\chi_p^{1-\alpha}/p$	0.128	0.159	0.192	0.147	0.164	0.182
	$\tilde{F}_2$	$\mathcal{F}_{p,G-p-q}^{1-\alpha}$	0.087	0.082	0.079	0.124	0.128	0.129
	$\tilde{F}_2^{\text{c}+}$	$\mathcal{F}_{p,G-p-q}^{1-\alpha}$	0.015	0.015	0.012	0.063	0.063	0.058
J-test	$J$	$\chi_q^{1-\alpha}$	—	0.935	—	—	0.329	—
	$\frac{G-q}{q} \frac{J}{G-J}$	$\mathcal{F}_{q,G-q}^{1-\alpha}$	—	0.069	—	—	0.086	—
	$\frac{G-q}{Gq} J^c$	$\mathcal{F}_{q,G-q}^{1-\alpha}$	—	0.129	—	—	0.071	—
		Test	$m = 24$			$m = 12$		
$G = 35$	Statistic	Critical values	$p = 1$	$p = 2$	$p = 3$	$p = 1$	$p = 2$	$p = 3$
first-step	$F_1$	$\chi_p^{1-\alpha}/p$	0.381	0.482	0.538	0.268	0.292	0.316
	$\frac{G-p}{G}F_1$	$\mathcal{F}_{p,G-p}^{1-\alpha}$	0.348	0.407	0.429	0.236	0.236	0.218
two-step	$F_2$	$\chi_p^{1-\alpha}/p$	0.572	0.756	0.864	0.292	0.357	0.416
	$\frac{G-p-q}{G}F_2$	$\chi_p^{1-\alpha}/p$	0.366	0.483	0.561	0.199	0.248	0.275
	$\tilde{F}_2$	$\chi_p^{1-\alpha}/p$	0.145	0.174	0.195	0.143	0.156	0.159
	$\tilde{F}_2$	$\mathcal{F}_{p,G-p-q}^{1-\alpha}$	0.111	0.119	0.115	0.128	0.126	0.117
	$\tilde{F}_2^{\text{c}+}$	$\mathcal{F}_{p,G-p-q}^{1-\alpha}$	0.038	0.034	0.033	0.066	0.060	0.057
J-test	$J$	$\chi_q^{1-\alpha}$	—	0.869	—	—	0.307	—
	$\frac{G-q}{q} \frac{J}{G-J}$	$\mathcal{F}_{q,G-q}^{1-\alpha}$	—	0.080	—	—	0.091	—
	$\frac{G-q}{Gq} J^c$	$\mathcal{F}_{q,G-q}^{1-\alpha}$	—	0.091	—	—	0.073	—

See footnote to Table 1.

Table 3: Empirical size of first-step and two-step tests based on the centered CCE when  $L_N = 50$ , number of clusters  $G = 50$  and  $100$ , number of joint hypothesis  $p = 1 \sim 3$  and number of moment conditions  $m = 12, 24$ .

		Test		$m = 24$			$m = 12$		
$G = 50$	statistic	critical values		$p = 1$	$p = 2$	$p = 3$	$p = 1$	$p = 2$	$p = 3$
first-step	$F_1$	$\chi_p^{1-\alpha}/p$		0.347	0.399	0.429	0.247	0.258	0.273
	$\frac{G-p}{G}F_1$	$\mathcal{F}_{p,G-p}^{1-\alpha}$		0.325	0.356	0.363	0.227	0.222	0.217
two-step	$F_2$	$\chi_p^{1-\alpha}/p$		0.399	0.538	0.634	0.208	0.255	0.299
	$\frac{G-p-q}{G}F_2$	$\chi_p^{1-\alpha}/p$		0.269	0.344	0.401	0.168	0.192	0.206
	$\tilde{F}_2$	$\chi_p^{1-\alpha}/p$		0.142	0.158	0.175	0.129	0.140	0.139
	$\tilde{F}_2$	$\mathcal{F}_{p,G-p-q}^{1-\alpha}$		0.123	0.129	0.135	0.119	0.116	0.115
	$\tilde{F}_2^{\text{c+}}$	$\mathcal{F}_{p,G-p-q}^{1-\alpha}$		0.067	0.064	0.057	0.064	0.060	0.058
J-test	$J$	$\chi_q^{1-\alpha}$		—	0.666	—	—	0.235	—
	$\frac{G-q}{q} \frac{J}{G-J}$	$\mathcal{F}_{q,G-q}^{1-\alpha}$		—	0.093	—	—	0.098	—
	$J^c$	$\mathcal{F}_{q,G-q}^{1-\alpha}$		—	0.072	—	—	0.072	—
		Test		$m = 24$			$m = 12$		
$G = 100$	Statistic	Critical values		$p = 1$	$p = 2$	$p = 3$	$p = 1$	$p = 2$	$p = 3$
first-step	$F_1$	$\chi_p^{1-\alpha}/p$		0.266	0.292	0.303	0.195	0.200	0.200
	$\frac{G-p}{G}F_1$	$\mathcal{F}_{p,G-p}^{1-\alpha}$		0.254	0.274	0.274	0.188	0.183	0.172
two-step	$F_2$	$\chi_p^{1-\alpha}/p$		0.213	0.271	0.305	0.120	0.142	0.156
	$\frac{G-p-q}{G}F_2$	$\chi_p^{1-\alpha}/p$		0.163	0.185	0.202	0.102	0.105	0.110
	$\tilde{F}_2$	$\chi_p^{1-\alpha}/p$		0.115	0.119	0.113	0.086	0.085	0.084
	$\tilde{F}_2$	$\mathcal{F}_{p,G-p-q}^{1-\alpha}$		0.109	0.108	0.099	0.082	0.079	0.076
	$\tilde{F}_2^{\text{c+}}$	$\mathcal{F}_{p,G-p-q}^{1-\alpha}$		0.078	0.069	0.063	0.053	0.049	0.049
J-test	$J$	$\chi_q^{1-\alpha}$		—	0.342	—	—	0.156	—
	$\frac{G-q}{q} \frac{J}{G-J}$	$\mathcal{F}_{q,G-q}^{1-\alpha}$		—	0.106	—	—	0.095	—
	$J^c$	$\mathcal{F}_{q,G-q}^{1-\alpha}$		—	0.069	—	—	0.065	—

See footnote to Table 1.



Table 4: Empirical size of first-step and two-step tests based on the centered CCE when  $L_N = 50$ , number of clusters  $G = 30$  and  $35$ , number of joint hypothesis  $p = 1 \sim 3$  and number of moment conditions  $m = 12, 24$ .

Test				$m = 24$			$m = 12$		
$G = 30$	Statistic	Critical values		$p = 1$	$p = 2$	$p = 3$	$p = 1$	$p = 2$	$p = 3$
first-step	$F_1$	$\chi_p^{1-\alpha}/p$		0.344	0.402	0.440	0.249	0.117	0.289
	$\frac{G-p}{G}F_1$	$\mathcal{F}_{p,G-p}^{1-\alpha}$		0.300	0.328	0.333	0.217	0.208	0.195
two-step	$F_2$	$\chi_p^{1-\alpha}/p$		0.610	0.798	0.883	0.275	0.347	0.436
	$\frac{G-p-q}{G}F_2$	$\chi_p^{1-\alpha}/p$		0.371	0.517	0.612	0.191	0.220	0.246
	$\tilde{F}_2$	$\chi_p^{1-\alpha}/p$		0.101	0.130	0.164	0.116	0.117	0.124
	$\tilde{F}_2$	$\mathcal{F}_{p,G-p-q}^{1-\alpha}$		0.061	0.064	0.059	0.095	0.088	0.124
	$\tilde{F}_2^{\mathbf{c}+}$	$\mathcal{F}_{p,G-p-q}^{1-\alpha}$		0.017	0.015	0.013	0.045	0.043	0.040
	J-test	$J$	$\chi_q^{1-\alpha}$		—	0.930	—	—	0.336
$\frac{G-q}{q} \frac{J}{G-J}$		$\mathcal{F}_{q,G-q}^{1-\alpha}$		—	0.073	—	—	0.098	—
$\frac{G-q}{Gq} J^{\mathbf{c}}$		$\mathcal{F}_{q,G-q}^{1-\alpha}$		—	.114	—	—	0.077	—
Test				$m = 24$			$m = 12$		
$G = 35$	Statistic	Ref.dist.		$p = 1$	$p = 2$	$p = 3$	$p = 1$	$p = 2$	$p = 3$
first-step	$F_1$	$\chi_p^{1-\alpha}/p$		0.321	0.373	0.393	0.237	0.257	0.269
	$\frac{G-p}{G}F_1$	$\mathcal{F}_{p,G-p}^{1-\alpha}$		0.292	0.303	0.300	0.213	0.207	0.187
two-step	$F_2$	$\chi_p^{1-\alpha}$		0.522	0.697	0.826	0.224	0.301	0.362
	$\frac{G-p-q}{G}F_2$	$\chi_p^{1-\alpha}/p$		0.306	0.411	0.484	0.162	0.196	0.224
	$\tilde{F}_2$	$\chi_p^{1-\alpha}$		0.105	0.125	0.136	0.107	0.115	0.118
	$\tilde{F}_2$	$\mathcal{F}_{p,G-p-q}^{1-\alpha}$		0.079	0.079	0.076	0.090	0.090	0.087
	$\tilde{F}_2^{\mathbf{c}+}$	$\mathcal{F}_{p,G-p-q}^{1-\alpha}$		0.038	0.032	0.026	0.049	0.049	0.044
J-test	$J$	$\chi_q^{1-\alpha}$		—	0.853	—	—	0.288	—
	$\frac{G-q}{q} \frac{J}{G-J}$	$\mathcal{F}_{q,G-q}^{1-\alpha}$		—	0.080	—	—	0.098	—
	$\frac{G-q}{Gq} J^{\mathbf{c}}$	$\mathcal{F}_{q,G-q}^{1-\alpha}$		—	0.078	—	—	0.075	—

See footnote to Table 1.

Table 5: Empirical size of first-step and two-step tests based on the centered CCE when  $L_N = 50$ , number of clusters  $G = 50$  and  $100$ , number of joint hypothesis  $p = 1 \sim 3$  and number of moment conditions  $m = 12, 24$ .

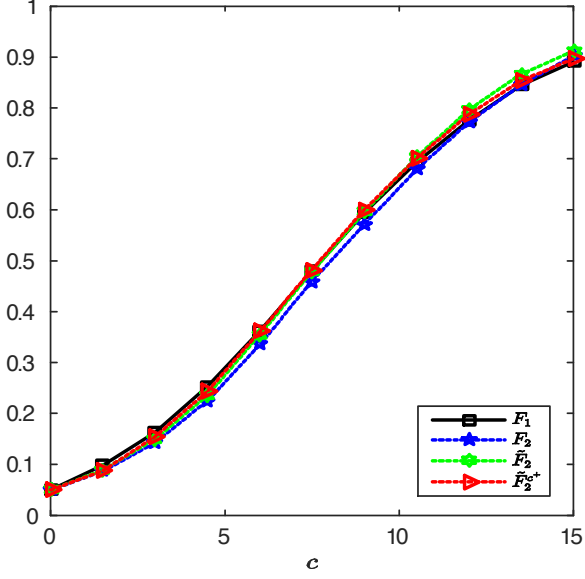
		Test		$m = 24$			$m = 12$		
$G = 50$	Statistic	Critical values	$p = 1$	$p = 2$	$p = 3$	$p = 1$	$p = 2$	$p = 3$	
first-step	$F_1$	$\chi_p^{1-\alpha}/p$	0.269	0.305	0.324	0.206	0.214	0.222	
	$\frac{G-p}{G}F_1$	$\mathcal{F}_{p,G-p}^{1-\alpha}$	0.250	0.260	0.257	0.189	0.180	0.175	
two-step	$F_2$	$\chi_p^{1-\alpha}/p$	0.341	0.469	0.575	0.176	0.211	0.250	
	$\frac{G-p-q}{G}F_2$	$\chi_p^{1-\alpha}/p$	0.209	0.268	0.324	0.129	0.146	0.164	
	$\tilde{F}_2$	$\chi_p^{1-\alpha}/p$	0.099	0.115	0.125	0.094	0.098	0.105	
	$\tilde{F}_2$	$\mathcal{F}_{p,G-p-q}^{1-\alpha}$	0.088	0.091	0.093	0.086	0.083	0.084	
	$\tilde{F}_2^{\text{c+}}$	$\mathcal{F}_{p,G-p-q}^{1-\alpha}$	0.059	0.056	0.052	0.057	0.052	0.055	
J-test	$J$	$\chi_q^{1-\alpha}$	—	0.629	—	—	0.221	—	
	$\frac{G-q}{q} \frac{J}{G-J}$	$\mathcal{F}_{q,G-q}^{1-\alpha}$	—	0.087	—	—	0.094	—	
	$J^c$	$\mathcal{F}_{q,G-q}^{1-\alpha}$	—	0.064	—	—	0.063	—	

		Test		$m = 24$			$m = 12$		
$G = 100$	Statistic	Critical values	$p = 1$	$p = 2$	$p = 3$	$p = 1$	$p = 2$	$p = 3$	
first-step	$F_1$	$\chi_p^{1-\alpha}/p$	0.185	0.203	0.209	0.149	0.147	0.152	
	$\frac{G-p}{G}F_1$	$\mathcal{F}_{p,G-p}^{1-\alpha}$	0.176	0.185	0.182	0.143	0.136	0.128	
two-step	$F_2$	$\chi_p^{1-\alpha}/p$	0.168	0.225	0.258	0.100	0.120	0.129	
	$\frac{G-p-q}{G}F_2$	$\chi_p^{1-\alpha}/p$	0.125	0.149	0.161	0.082	0.089	0.093	
	$\tilde{F}_2$	$\chi_p^{1-\alpha}/p$	0.085	0.091	0.094	0.069	0.073	0.071	
	$\tilde{F}_2$	$\mathcal{F}_{p,G-p-q}^{1-\alpha}$	0.081	0.082	0.083	0.065	0.067	0.064	
	$\tilde{F}_2^{\text{c+}}$	$\mathcal{F}_{p,G-p-q}^{1-\alpha}$	0.067	0.065	0.064	0.056	0.055	0.051	
J-test	$J$	$\chi_q^{1-\alpha}$	—	0.293	—	—	0.135	—	
	$\frac{G-q}{q} \frac{J}{G-J}$	$\mathcal{F}_{q,G-q}^{1-\alpha}$	—	0.081	—	—	0.081	—	
	$J^c$	$\mathcal{F}_{q,G-q}^{1-\alpha}$	—	0.055	—	—	0.053	—	

See footnote to Table 1.

Full instrument set,  $G = 30, L = 50, d = 2,$  and  $p = 1$



Reduced instrument set,  $G = 30, L = 50, d = 2,$  and  $p = 1$

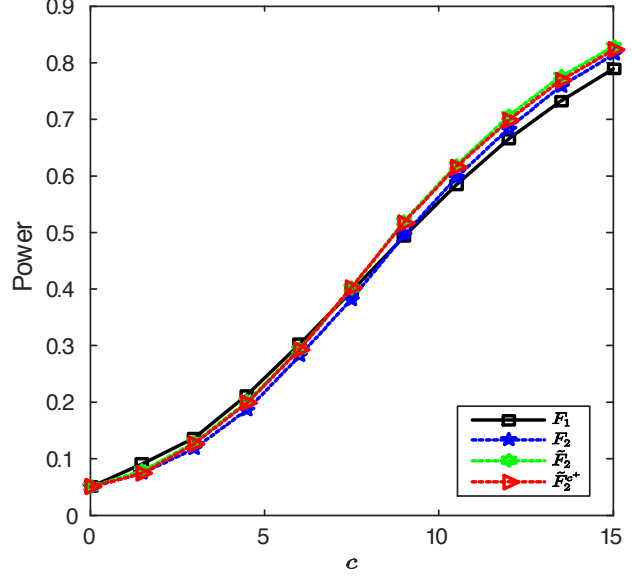
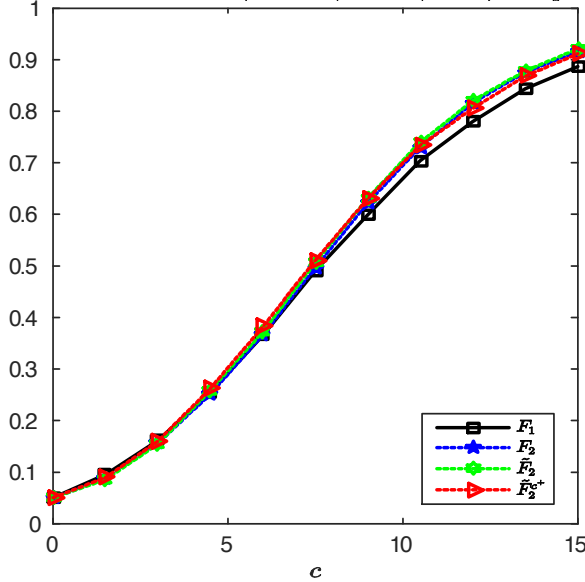


Figure 2: Size-adjusted power of the first-step (2SLS) and two-step tests with  $G = 30$  and  $L_N = 50$ .

Full instrument set,  $G = 35, L = 50, d = 2,$  and  $p = 1$



Reduced instrument set,  $G = 35, L = 50, d = 2,$  and  $p = 1$

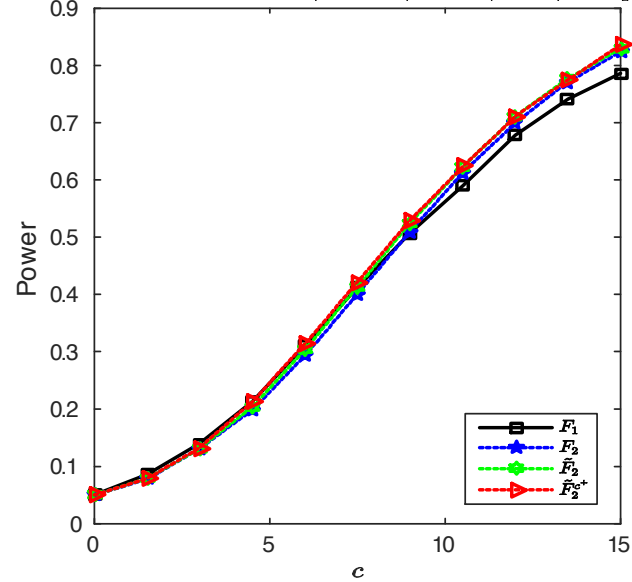
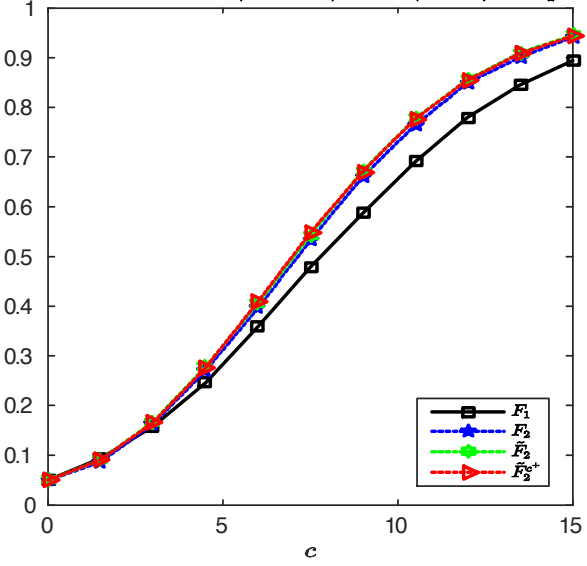


Figure 3: Size-adjusted power of first-step (2SLS) and two-step tests with  $G = 35$  and  $L = 50$ .

Full instrument set,  $G = 50, L = 50, d = 2$ , and  $p = 1$



Reduced instrument set,  $G = 50, L = 50, d = 2$ , and  $p = 1$

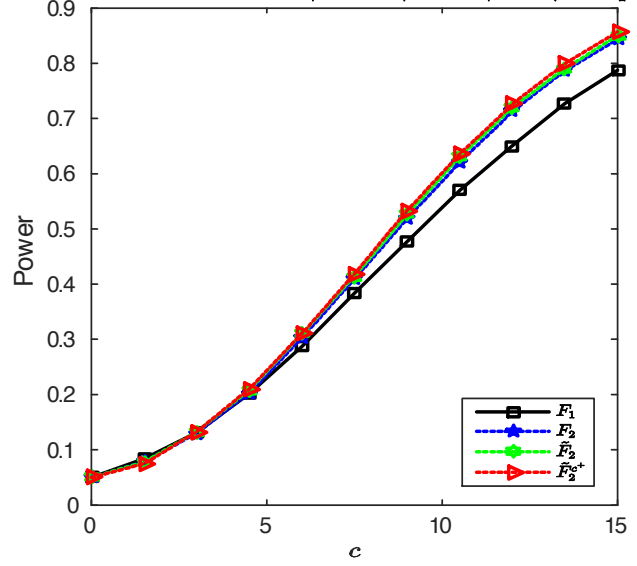
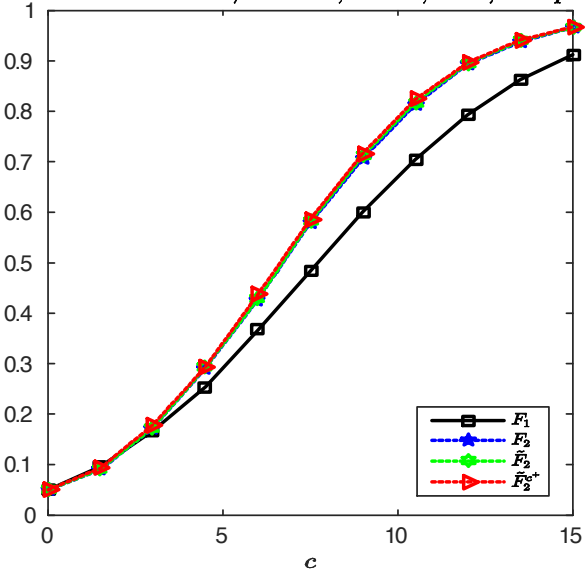


Figure 4: Size-adjusted power of first-step (2SLS) and two-step tests with  $G = 50$  and  $L = 50$ .

Full instrument set,  $G = 100, L = 50, d = 2$ , and  $p = 1$



Reduced instrument set,  $G = 100, L = 50, d = 2$ , and  $p = 1$

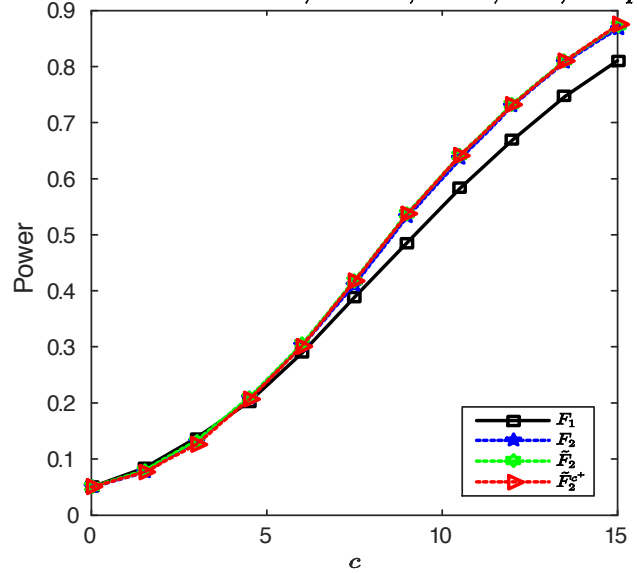


Figure 5: Size-adjusted power of first-step (2SLS) and two-step tests with  $G = 100$  and  $L = 50$ .

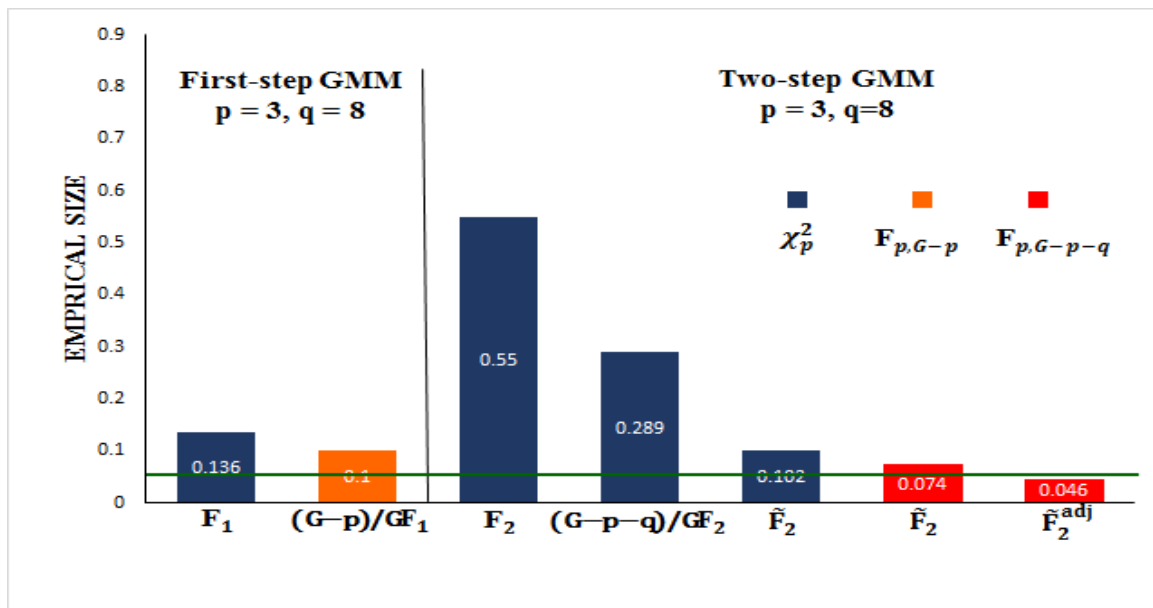


Figure 6: Empirical size of first-step and two-step tests based on the centered CCE when there is a heterogeneity in cluster size with the nominal size 5% (green line): Design I with  $G = 50$ ,  $q = 8$ , and  $p = 3$ .

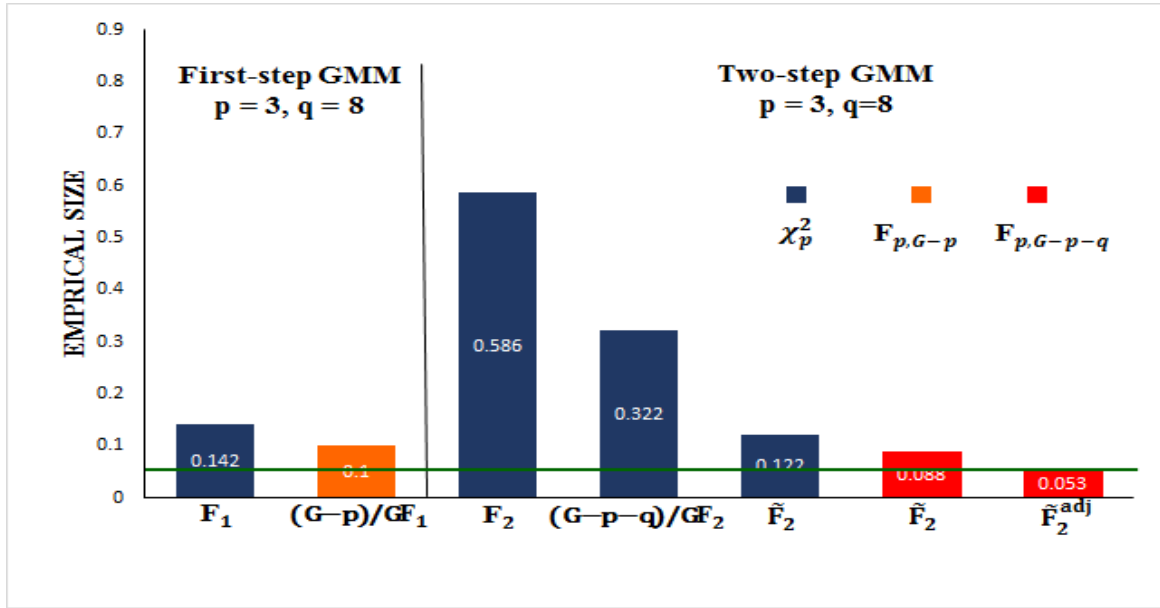


Figure 7: Empirical size of first-step and two-step tests based on the centered CCE when there is a heterogeneity in cluster size with the nominal size 5% (green line): Design II with  $G = 50$ ,  $q = 8$ , and  $p = 3$ .

Table 6: Design of heterogeneity in cluster size

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

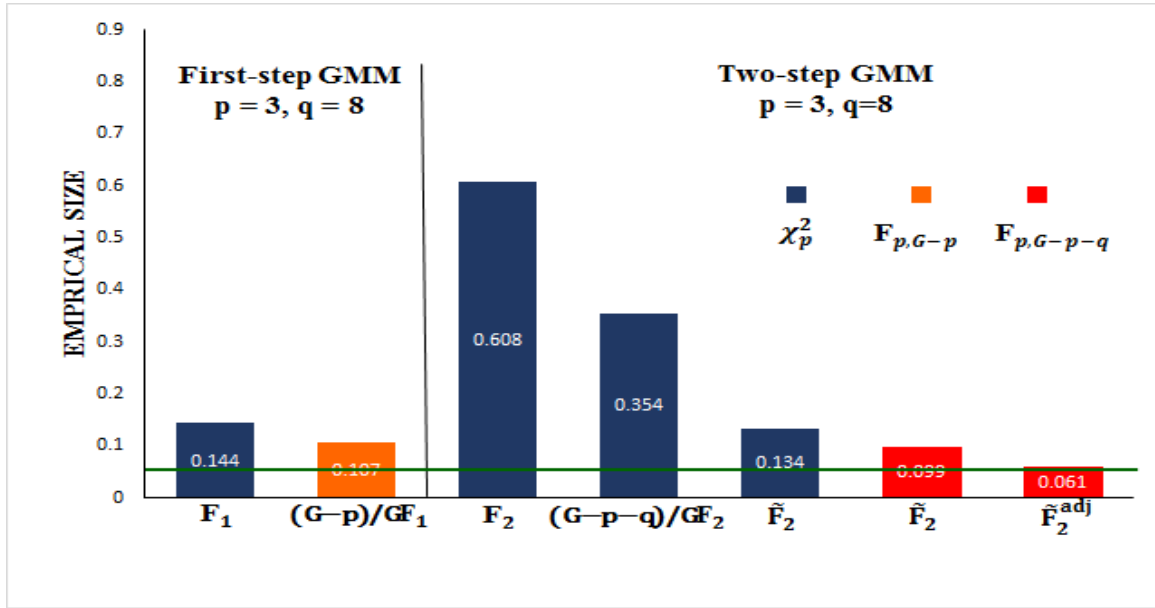


Figure 8: Empirical size of first-step and two-step tests based on the centered CCE when there is a heterogeneity in cluster size with the nominal size 5% (green line): Design III with  $G = 50$ ,  $q = 8$ , and  $p = 3$ .

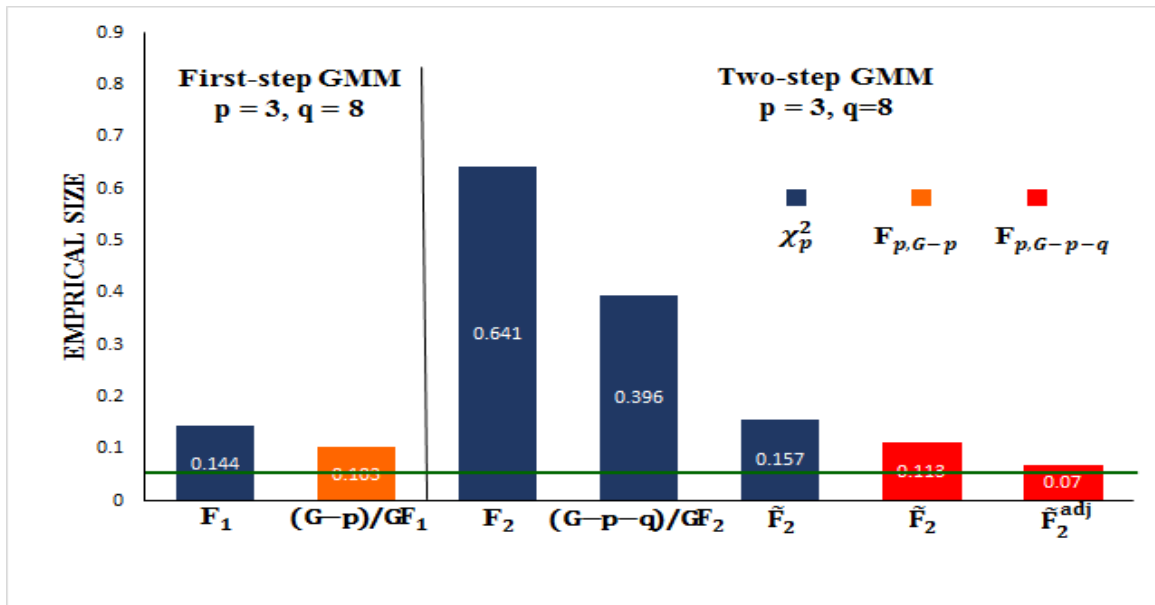


Figure 9: Empirical size of first-step and two-step tests based on the centered CCE when there is a heterogeneity in cluster size with the nominal size 5% (green line): Design IV with  $G = 50$ ,  $q = 20$ , and  $p = 3$ .

Table 7: Empirical size of first-step and two-step tests based on the centered CCE when there is a heterogeneity in cluster size: Design I~II

Design I								
test statistic		critical values	$m = 24$			$m = 12$		
			$p = 1$	$p = 2$	$p = 3$	$p = 1$	$p = 2$	$p = 3$
first-step	$F_1$	$\chi_p^{1-\alpha}/p$	0.265	0.303	0.322	0.205	0.201	0.217
	$\frac{G-p}{G} F_1$	$\mathcal{F}_{p,G-p}^{1-\alpha}$	0.246	0.264	0.261	0.186	0.168	0.165
two-step	$F_2$	$\chi_p^{1-\alpha}/p$	0.335	0.472	0.575	0.153	0.194	0.242
	$\frac{G-p-q}{G} F_2$	$\chi_p^{1-\alpha}/p$	0.209	0.265	0.319	0.118	0.132	0.153
	$\tilde{F}_2$	$\chi_p^{1-\alpha}/p$	0.095	0.104	0.119	0.086	0.087	0.091
	$\tilde{F}_2$	$\mathcal{F}_{p,G-p-q}^{1-\alpha}$	0.082	0.083	0.085	0.074	0.071	0.072
	$\tilde{F}_2^{\text{adj}+}$	$\mathcal{F}_{p,G-p-q}^{1-\alpha}$	0.054	0.048	0.043	0.046	0.046	0.047
	J-test	$J$	$\chi_q^{1-\alpha}$	—	0.632	—	—	0.228
	$\frac{G-q}{q} \frac{qJ}{G-qJ}$	$\mathcal{F}_{q,G-q}^{1-\alpha}$	—	0.089	—	—	0.100	—
	$\frac{G-q}{Gq} J^c$	$\mathcal{F}_{q,G-q}^{1-\alpha}$	—	0.067	—	—	0.068	—
Design II								
			$m = 24$			$m = 12$		
$G = 35$	test statistic	critical values	$p = 1$	$p = 2$	$p = 3$	$p = 1$	$p = 2$	$p = 3$
first-step	$F_1$	$\chi_p^{1-\alpha}/p$	0.268	0.304	0.323	0.206	0.213	0.220
	$\frac{G-p}{G} F_1$	$\mathcal{F}_{p,G-p}^{1-\alpha}$	0.250	0.265	0.261	0.193	0.182	0.171
two-step	$F_2$	$\chi_p^{1-\alpha}/p$	0.361	0.501	0.601	0.160	0.209	0.254
	$\frac{G-p-q}{G} F_2$	$\chi_p^{1-\alpha}/p$	0.231	0.291	0.338	0.124	0.141	0.170
	$\tilde{F}_2$	$\chi_p^{1-\alpha}/p$	0.112	0.120	0.132	0.088	0.092	0.100
	$\tilde{F}_2$	$\mathcal{F}_{p,G-p-q}^{1-\alpha}$	0.097	0.098	0.102	0.079	0.076	0.082
	$\tilde{F}_2^{\text{adj}+}$	$\mathcal{F}_{p,G-p-q}^{1-\alpha}$	0.063	0.060	0.057	0.053	0.052	0.056
	J-test	$J$	$\chi_q^{1-\alpha}$	—	0.638	—	—	0.214
	$\frac{G-q}{q} \frac{qJ}{G-qJ}$	$\mathcal{F}_{q,G-q}^{1-\alpha}$	—	0.083	—	—	0.094	—
	$\frac{G-q}{Gq} J^c$	$\mathcal{F}_{q,G-q}^{1-\alpha}$	—	0.081	—	—	0.072	—

See footnote to Table 1.



Table 8: Empirical size of first-step and two-step tests based on the centered CCE when there is a heterogeneity in cluster size: Design III~IV

Design III								
test statistic		critical values	$m = 24$			$m = 12$		
			$p = 1$	$p = 2$	$p = 3$	$p = 1$	$p = 2$	$p = 3$
first-step	$F_1$	$\chi_p^{1-\alpha}/p$	0.276	0.315	0.327	0.203	0.215	0.228
	$\frac{G-p}{G} F_1$	$\mathcal{F}_{p,G-p}^{1-\alpha}$	0.254	0.276	0.268	0.186	0.186	0.178
two-step	$F_2$	$\chi_p^{1-\alpha}/p$	0.378	0.532	0.632	0.168	0.222	0.274
	$\frac{G-p-q}{G} F_2$	$\chi_p^{1-\alpha}/p$	0.244	0.323	0.374	0.134	0.164	0.184
	$\tilde{F}_2$	$\chi_p^{1-\alpha}/p$	0.117	0.132	0.145	0.097	0.109	0.117
	$\tilde{F}_2$	$\mathcal{F}_{p,G-p-q}^{1-\alpha}$	0.101	0.107	0.110	0.088	0.091	0.092
	$\tilde{F}_2^{\text{adj+}}$	$\mathcal{F}_{p,G-p-q}^{1-\alpha}$	0.04	0.057	0.061	0.061	0.062	0.060
	J-test	$J$	$\chi_q^{1-\alpha}$	—	0.631	—	—	0.213
$\frac{G-q}{q} \frac{qJ}{G-qJ}$		$\mathcal{F}_{q,G-q}^{1-\alpha}$	—	0.076	—	—	0.089	—
$\frac{G-q}{Gq} J^c$		$\mathcal{F}_{q,G-q}^{1-\alpha}$	—	0.102	—	—	0.079	—
Design IV								
			$m = 24$			$m = 12$		
$G = 35$	test statistic	critical values	$p = 1$	$p = 2$	$p = 3$	$p = 1$	$p = 2$	$p = 3$
first-step	$F_1$	$\chi_p^{1-\alpha}/p$	0.255	0.306	0.327	0.200	0.214	0.226
	$\frac{G-p}{G} F_1$	$\mathcal{F}_{p,G-p}^{1-\alpha}$	0.232	0.265	0.266	0.180	0.179	0.171
two-step	$F_2$	$\chi_p^{1-\alpha}/p$	0.397	0.555	0.664	0.185	0.243	0.304
	$\frac{G-p-q}{G} F_2$	$\chi_p^{1-\alpha}/p$	0.264	0.363	0.429	0.139	0.174	0.205
	$\tilde{F}_2$	$\chi_p^{1-\alpha}/p$	0.131	0.155	0.176	0.104	0.120	0.136
	$\tilde{F}_2$	$\mathcal{F}_{p,G-p-q}^{1-\alpha}$	0.115	0.125	0.140	0.094	0.102	0.110
	$\tilde{F}_2^{\text{adj+}}$	$\mathcal{F}_{p,G-p-q}^{1-\alpha}$	0.075	0.077	0.079	0.062	0.070	0.077
	J-test	$J$	$\chi_q^{1-\alpha}$	—	0.621	—	—	0.206
$\frac{G-q}{q} \frac{qJ}{G-qJ}$		$\mathcal{F}_{q,G-q}^{1-\alpha}$	—	0.070	—	—	0.080	—
$\frac{G-q}{Gq} J^c$		$\mathcal{F}_{q,G-q}^{1-\alpha}$	—	0.151	—	—	0.092	—

See footnote to Table 1.

Table 9: Summary of the difference between the conventional large- $G$  asymptotics and alternative fixed- $G$  asymptotics for the first-step (2SLS ) and two-step GMM methods.

<b>2SLS</b>							
Asymptotics	Weight	Test statistic			Reference distribution		
.		Wald	t	.	Wald	t	.
large- $G$	$Z'HZ/N$	$F(\hat{\theta}_1)$	$t(\hat{\theta}_1)$	.	$\chi_p^2/p$	$N(0, 1)$	.
fixed- $G$	$Z'HZ/N$	$\frac{G-p}{G}F(\hat{\theta}_1)$	$\sqrt{\frac{G-1}{G}}t(\hat{\theta}_1)$	.	$F_{p,G-p}$	$t_{G-1}$	.
<b>Two-step GMM</b>							
Asymptotics	Weight	Test statistic			Reference distribution		
.		Wald	t	J	Wald	t	J
large- $G$	$\hat{\Omega}$	$F(\hat{\theta}_2)$	$t(\hat{\theta}_2)$	$J(\hat{\theta}_2)$	$\chi_p^2/p$	$N(0, 1)$	$\chi_q^2$
fixed- $G$	$\hat{\Omega}^c$	$\tilde{F}(\hat{\theta}_2^c)$	$\tilde{t}(\hat{\theta}_2^c)$	$\frac{G-q}{G}J(\hat{\theta}_2^c)$	$F_{p,G-p-q}$	$t_{G-1-q}$	$F_{q,G-q}$
	$\hat{\Omega}^c$	$\tilde{F}^{\text{adj}+}(\hat{\theta}_2^c)$	$\tilde{t}^{\text{adj}+}(\hat{\theta}_2^c)$	.	$F_{p,G-p-q}$	$t_{G-1-q}$	.

Table 10: Summary of standard error formula when  $p = 1$

<b>2SLS</b>		
	large- $G$	fixed- $G$
standard error	$\left[ R\widehat{\text{var}}(\hat{\theta}_1)R' \right]^{1/2}$	$\left[ \frac{G-1}{G} R\widehat{\text{var}}_{\Omega(\hat{\theta}_1)}(\hat{\theta}_1)R' \right]^{1/2}$
<b>Two-step GMM</b>		
	large- $G$	fixed- $G$
standard error	$\left[ R\widehat{\text{var}}(\hat{\theta}_1)R' \right]^{1/2}$	$\left[ \frac{G}{G-1-q} R\widehat{\text{var}}_{\hat{\Omega}(\hat{\theta}_1)}(\hat{\theta}_2^c)R' \left( 1 + (q/G)J(\hat{\theta}_2^c) \right) \right]^{1/2}$
corrected standard error	.	$\left[ \frac{G}{G-1-q} R\widehat{\text{var}}_{\hat{\Omega}(\hat{\theta}_1)}^{\text{adj}}(\hat{\theta})R' \left( 1 + (q/G)J(\hat{\theta}_2^c) \right) \right]^{1/2}$

Table 11: Results for Emran and Hou (2013) data

<b>2SLS</b>		
Variables	Large- $G$ Asymptotics	Fixed- $G$ Asymptotics
Domestic market ( $A_i^d$ )	-2713.2 (712.1) [-4109.9, -1316.4]	-2713.2 (716.8) [-4138.0, -1288.0]
International market ( $A_i^s$ )	-1993.5 (514.8) [-3002.5, -984.4]	-1993.5 (517.9) [-3023.10, -963.8]
Interaction ( $A_i^d \times A_i^s$ )	345.8 (105.0) [140.0, 551.7]	345.8 (105.6) [135.8, 555.9]
$H_0 : \beta_d = \beta_s$	-2.3218 (2.02%)	-2.771 (2.26%)
<b>Two-step GMM</b>		
Variables	Large- $G$ Asymptotics	Fixed- $G$ Asymptotics
Domestic market ( $A_i^d$ )	-2722.8 (400.5) [-3507.7, -1937.9]	-2670.0 (519.2) (520.7) [-3706.2, -1633.8] [-3709.2, -1630.7]
International market ( $A_i^s$ )	-2000.2 (344.3) [-2675.0, -1325.5]	-1981.3 (446.4) (447.7) [-2872.3, -1090.3] [-2874.9, -1087.7]
Interaction ( $A_i^d \times A_i^s$ )	362.7 (68.7) [228.0, 497.3]	364.1 (89.1) (89.4) [186.2, 541.9] [187.5, 542.4]
$H_0 : \beta_d = \beta_s$	-5.239 (0%)	-3.3318 (0%) -3.3217 (0%)
J-statistic ( $q = 18$ )	1.1708 (99.8%)	0.3096 (45.83%)

Notes: Standard errors for 2SLS and the weighting matrix for (centered) two-step GMM estimators are clustered at the county level. Numbers in parentheses are standard errors and intervals are 95% confidence intervals. For hypothesis testing, the numbers in parentheses are P-values.

## 10 Appendix: Application to Linear Dynamic Panel Model

We discuss how to implement our inference procedures in the context of a linear dynamic panel model:

$$y_{it} = \gamma y_{it-1} + x'_{it}\beta + \eta_i + u_{it}, \quad (35)$$

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

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

$$\Delta y_{it} = \Delta w'_{it}\theta + \Delta u_{it}, \quad t = 2, \dots, T$$

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

$$\begin{aligned} Z_i &= \text{diag}(z'_{i2}, \dots, z'_{iT}) \\ &_{(T-1) \times d(T-1)T/2} \\ z_{it} &= (y_{i0}, \dots, y_{it-2}, x'_{i1}, \dots, x'_{it-1})', \quad 2 \leq t \leq T. \end{aligned}$$

The moment conditions are then given by

$$E(Z'_i \Delta u_i) = 0,$$

where  $\Delta u_i$  is the  $(T-1)$  vector  $(\Delta u_{i2}, \dots, \Delta u_{iT})'$ . The original AB method assumes away cross-sectional dependence, but clustered dependence can be easily accommodated. Here we assume that the moment vector  $\{Z'_i \Delta u_i\}_{i=1}^N$  can be partitioned into independent clusters. That is,  $\{Z'_i \Delta u_i\}_{i=1}^N = \cup_{g=1}^G \{Z'^g_i \Delta u^g_i\}_{i=1}^{L_N}$  with  $Z'^g_i \Delta u^g_i$  and  $Z'^h_j \Delta u^h_j$  being independent for all  $g \neq h$ .

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

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

where  $Z'$  is the  $dT(T-1)/2 \times N(T-1)$  matrix  $(Z'_1, Z'_2, \dots, Z'_N)$ ,  $\Delta w_i$  is the  $(T-1) \times d$  matrix  $(\Delta w_{i2}, \dots, \Delta w_{iT})'$ ,  $\Delta y_i$  is the  $(\Delta y_{i2}, \dots, \Delta y_{iT})'$ ,  $\Delta w$  and  $\Delta y$  are  $(\Delta w'_1, \dots, \Delta w'_N)'$  and  $(\Delta y'_1, \dots, \Delta y'_N)'$ , respectively. The corresponding Wald statistic<sup>7</sup> for testing  $H_0 : R\theta_0 = r$  vs  $H_1 : R\theta_0 \neq r$  is given by

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

<sup>7</sup>The formula for the t-statistic, which is omitted here, is straightforward.

where

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

Let  $Z_{(g)}$  be the  $L(T-1) \times dT(T-1)/2$  matrix obtained by stacking all  $Z_i$ 's belonging to cluster  $g$ . Similarly, let  $\Delta \hat{u}_{(g)}$  be the  $L(T-1)$  stacked vector of the estimated first-differenced errors  $\Delta \hat{u}_i = \Delta y_i - \Delta w_i' \hat{\theta}_1$ . Then, in the presence of clustered dependence, the CCE and centered CCE are constructed as follows:

$$\begin{aligned} \hat{\Omega}(\hat{\theta}_1) &= \frac{1}{G} \sum_{g=1}^G \left( \frac{Z'_{(g)} \Delta \hat{u}_{(g)}}{\sqrt{L}} \right) \left( \frac{Z'_{(g)} \Delta \hat{u}_{(g)}}{\sqrt{L}} \right)' \\ \hat{\Omega}^c(\hat{\theta}_1) &= \frac{1}{G} \sum_{g=1}^G \left( \frac{Z'_{(g)} \Delta \hat{u}_{(g)}}{\sqrt{L}} - \frac{1}{G} \sum_{h=1}^G \frac{Z'_{(h)} \Delta \hat{u}_{(h)}}{\sqrt{L}} \right) \left( \frac{Z'_{(g)} \Delta \hat{u}_{(g)}}{\sqrt{L}} - \frac{1}{G} \sum_{h=1}^G \frac{Z'_{(h)} \Delta \hat{u}_{(h)}}{\sqrt{L}} \right)'. \end{aligned}$$

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

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

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

$$\begin{aligned} F_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c) &:= (R\hat{\theta}_2^c - r)' \{ R \widehat{\text{var}}_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c) R' \}^{-1} (R\hat{\theta}_2^c - r) / p, \\ \widehat{\text{var}}_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c) &= N \left\{ \Delta w' Z \left[ \hat{\Omega}^c(\hat{\theta}_1) \right]^{-1} Z' \Delta w \right\}^{-1}. \end{aligned}$$

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

$$\begin{aligned} F(\hat{\theta}_1) &\xrightarrow{d} \frac{G}{G-p} F_{p, G-p} \text{ and} \\ F_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c) &\xrightarrow{d} \frac{G}{G-p-q} F_{p, G-p-q} (\|\Delta\|^2). \end{aligned} \quad (36)$$

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

$$\widehat{\text{var}}_{\hat{\Omega}^c(\hat{\theta}_1)}^c(\hat{\theta}_2^c) = \widehat{\text{var}}_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c) + \widehat{\mathcal{E}}_n \widehat{\text{var}}_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c) + \widehat{\text{var}}_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c) \widehat{\mathcal{E}}_n' + \widehat{\mathcal{E}}_n \widehat{\text{var}}(\hat{\theta}_1) \widehat{\mathcal{E}}_n' \quad (37)$$

where the  $j$ -th column is given by

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

and

$$\begin{aligned} \left. \frac{\partial \hat{\Omega}^c(\theta)}{\partial \theta_j} \right|_{\theta=\hat{\theta}_1} &= \Upsilon_j(\hat{\theta}_1) + \Upsilon'_j(\hat{\theta}_1), \\ \Upsilon_j(\hat{\theta}_1) &= -\frac{1}{G} \sum_{g=1}^G \left( \frac{Z'_{(g)} \Delta w_{j,(g)}}{\sqrt{L_g}} - \frac{1}{G} \sum_{h=1}^G \frac{Z'_{(h)} \Delta w_{j,(h)}}{\sqrt{L_h}} \right) \left( \frac{Z'_{(g)} \Delta \hat{u}_{(g)}}{\sqrt{L_g}} - \frac{1}{G} \sum_{h=1}^G \frac{Z'_{(h)} \Delta \hat{u}_{(h)}}{\sqrt{L_h}} \right)' \\ \Delta w_i &= (\Delta w_{1,i}, \dots, \Delta w_{d,i}) \text{ and } \Delta w_{(g)} = (\Delta w_{1,(g)}, \dots, \Delta w_{d,(g)}) \\ & \quad \substack{(T-1) \times d \\ L_N(T-1) \times d} \end{aligned}$$

for each  $j = 1, \dots, d$ . Here,  $\Delta w_{(g)}$  is a  $L(T-1) \times d$  matrix that stacks  $\{\Delta w_i\}_{i=1}^N$  belonging to the group  $g$ . The extra adjustment toward a ‘larger’ corrected estimator  $\widehat{var}_{\hat{\Omega}^c(\hat{\theta}_1)}^{\text{adj}+}(\hat{\theta}_2^c)$  directly follows from (??).

Based on the finite sample corrected variance estimator in (37) and the usual J-statistic, we construct the modified Wald statistic:

$$\tilde{F}_{\hat{\Omega}^c(\hat{\theta}_1)}^{\text{adj}}(\hat{\theta}_2^c) = \frac{G-p-q}{G} \frac{F_{\hat{\Omega}^c(\hat{\theta}_1)}^{\text{adj}+}(\hat{\theta}_2^c)}{1 + \frac{q}{G} J(\hat{\theta}_2^c)} \quad (38)$$

where

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

From the F limit theory in Section 5, we have

$$\tilde{F}_{\hat{\Omega}^c(\hat{\theta}_1)}^{\text{adj}}(\hat{\theta}_2^c) \xrightarrow{d} F_{p,G-p-q} \quad (39)$$

and

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

Critical values for from the F distribution are readily available from statistical tables.

## Appendix of Proofs

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

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

where  $\tilde{\theta}^*$  is between  $\hat{\theta}_1$  and  $\theta_0$ . 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

$$\begin{aligned} & \frac{1}{\sqrt{L}} \sum_{i=1}^L f_i^g(\hat{\theta}) \\ &= \frac{1}{\sqrt{L}} \sum_{i=1}^L f_i^g(\theta_0) - \frac{1}{L} \sum_{i=1}^L \frac{\partial f_i^g(\tilde{\theta}^*)}{\partial \theta'} (\Gamma' W^{-1} \Gamma)^{-1} \Gamma' W^{-1} \frac{1}{G} \sum_{h=1}^G \left( \frac{1}{\sqrt{L}} \sum_{j=1}^L f_j^h(\theta_0) \right) + o_p(1) \\ &= \frac{1}{\sqrt{L}} \sum_{i=1}^L f_i^g(\theta_0) - \Gamma_g (\Gamma' W^{-1} \Gamma)^{-1} \Gamma' W^{-1} \frac{1}{G} \sum_{h=1}^G \left( \frac{1}{\sqrt{L}} \sum_{i=1}^L f_i^h(\theta_0) \right) + o_p(1). \end{aligned} \quad (40)$$

Using Assumptions 4–6, we then have

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

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

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

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

$$\begin{aligned} & \hat{\Gamma}(\hat{\theta}_1)' W_N^{-1} \hat{\Omega}(\hat{\theta}_1) W_N^{-1} \hat{\Gamma}(\hat{\theta}_1) \\ &= \frac{1}{G} \sum_{g=1}^G \left\{ \left( \hat{\Gamma}(\hat{\theta}_1)' W_N^{-1} \frac{1}{\sqrt{L}} \sum_{i=1}^L f_i^g(\hat{\theta}_1) \right) \left( \frac{1}{\sqrt{L}} \sum_{j=1}^L f_j^g(\hat{\theta}_1) W_N^{-1} \hat{\Gamma}(\hat{\theta}_1) \right)' \right\} \\ & \xrightarrow{d} \Gamma' W^{-1} \Lambda \left\{ \frac{1}{G} \sum_{g=1}^G (B_{m,g} - \bar{B}_m) (B_{m,g} - \bar{B}_m)' \right\} (\Gamma' W^{-1} \Lambda)'. \end{aligned}$$

Therefore,

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

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:

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

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

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

Using this, we get

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

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

*Proof of Proposition 2. 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

$$M_\Lambda = \Lambda^{-1}M(\Lambda')^{-1} \text{ and } \Gamma_\Lambda = \Lambda^{-1}\Gamma.$$

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  $\Gamma_\Lambda$ . By construction,  $U'U = UU' = I_m$ ,  $V'V = VV' = I_d$ , and

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

where  $A$  is a diagonal matrix. Denoting

$$\tilde{M} = U' M_\Lambda U = \begin{pmatrix} \tilde{M}_{11} & \tilde{M}_{12} \\ \tilde{M}_{21} & \tilde{M}_{22} \end{pmatrix},$$

we have

$$\begin{aligned} & (\Gamma_\Lambda M_\Lambda^{-1} \Gamma_\Lambda)^{-1} \Gamma_\Lambda M_\Lambda^{-1} \\ &= [V\Sigma'U'M_\Lambda^{-1}U\Sigma V']^{-1} V\Sigma'U'M_\Lambda^{-1} \\ &= [V\Sigma' (U'M_\Lambda U)^{-1} \Sigma V']^{-1} V\Sigma' (U'M_\Lambda U)^{-1} U' \\ &= V \left( A' \tilde{M}^{11} A \right)^{-1} \begin{pmatrix} A'_{d \times d} & O'_{q \times d} \end{pmatrix} \begin{pmatrix} \tilde{M}^{11} & \tilde{M}^{12} \\ \tilde{M}^{21} & \tilde{M}^{22} \end{pmatrix} U' \\ &= VA^{-1} (\tilde{M}^{11})^{-1} \begin{pmatrix} I_d & O_{d \times q} \end{pmatrix} \begin{pmatrix} \tilde{M}^{11} & \tilde{M}^{12} \\ \tilde{M}^{21} & \tilde{M}^{22} \end{pmatrix} U' \\ &= VA^{-1} (\tilde{M}^{11})^{-1} \begin{pmatrix} \tilde{M}^{11} & \tilde{M}^{12} \end{pmatrix} U' = VA^{-1} \begin{pmatrix} I_d & (\tilde{M}^{11})^{-1} \tilde{M}^{12} \end{pmatrix} U' \\ &= VA^{-1} \begin{pmatrix} I_d & -\tilde{M}_{12} \tilde{M}_{22}^{-1} \end{pmatrix} U'. \end{aligned}$$

So

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

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

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

Therefore

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

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

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

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

using the asymptotic equivalence  $(\mathbb{S}, \bar{B}_m) \stackrel{d}{=} (U' \mathbb{S} U, U' \bar{B}_m)$  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  $M_\Lambda$  is  $D_\infty$ . We have

$$\begin{aligned}
& \sqrt{N}(\hat{\theta}_{2N} - \theta_0) \xrightarrow{d} -[\Gamma'_\Lambda D_\infty^{-1} \Gamma_\Lambda]^{-1} \Gamma'_\Lambda D_\infty^{-1} \sqrt{G} \bar{B}_m \\
& = -\left[ V \Sigma' (U' D_\infty U)^{-1} \Sigma V' \right]^{-1} V \Sigma (U' D_\infty U)^{-1} U' \sqrt{G} \bar{B}_m \\
& = -V \left[ \Sigma' \mathbb{D}_\infty^{-1} \Sigma \right]^{-1} \Sigma' \mathbb{D}_\infty^{-1} U' \sqrt{G} \bar{B}_m \\
& = -V A^{-1} \begin{pmatrix} I_d & -\mathbb{D}_{12} \mathbb{D}_{22}^{-1} \end{pmatrix} U' \sqrt{G} \bar{B}_m
\end{aligned} \tag{41}$$

where

$$\mathbb{D}_\infty = U' D_\infty U = \begin{pmatrix} \mathbb{D}_{11} & \mathbb{D}_{12} \\ d \times d & d \times q \\ \mathbb{D}_{21} & \mathbb{D}_{22} \\ q \times d & q \times q \end{pmatrix}.$$

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

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

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

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

for

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

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

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

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

$$\begin{aligned}\mathbb{D}_{11} &= \tilde{S}_{\infty,11}^U + \beta_{\tilde{W}}\bar{B}_q^U(\bar{B}_q^U)'\beta_{\tilde{W}}', \\ \mathbb{D}_{12} &= \tilde{S}_{\infty,12}^U + \beta_{\tilde{W}}\bar{B}_q^U(\bar{B}_q^U)', \\ \mathbb{D}_{21} &= \tilde{S}_{\infty,21}^U + \bar{B}_q^U(\bar{B}_q^U)'\beta_{\tilde{W}}', \\ \mathbb{D}_{22} &= \tilde{S}_{\infty,22}^U + \bar{B}_q^U(\bar{B}_q^U)' = S_{\infty,22}^U.\end{aligned}\tag{42}$$

Using these representations, we can rewrite (41) as

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

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

$$\sqrt{N}(\hat{\theta}_2 - \theta_0) - \sqrt{N}(\tilde{\theta}_2 - \theta_0) - \sqrt{N}(\hat{\theta}_1 - \theta_0) \cdot (\kappa_G/G) \stackrel{d}{\rightarrow} 0,$$

which implies that

$$\sqrt{N}(\hat{\theta}_2 - \theta_0) - \sqrt{N}(\tilde{\theta}_2 - \theta_0) - \sqrt{N}(\hat{\theta}_1 - \theta_0) \cdot (\kappa_G/G) = o_p(1).$$

That is

$$\sqrt{N}(\hat{\theta}_2 - \theta_0) = \sqrt{N}(\tilde{\theta}_2 - \theta_0) + \sqrt{N}(\hat{\theta}_1 - \theta_0) \cdot (\kappa_G/G) + o_p(1).$$

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

$$\begin{aligned}\sqrt{N}g_n(\hat{\theta}_2) &= \frac{1}{\sqrt{G}} \sum_{g=1}^G \left( \frac{1}{\sqrt{L}} \sum_{i=1}^L f_i^g(\hat{\theta}_2) \right) \\ &\stackrel{d}{\rightarrow} \Lambda\sqrt{G} \left( UU'\bar{B}_m - \Gamma_\Lambda [\Gamma'_\Lambda D_\infty^{-1} \Gamma_\Lambda]^{-1} \Gamma'_\Lambda D_\infty^{-1} \bar{B}_m \right) \\ &\stackrel{d}{=} \Lambda\sqrt{G} [U\bar{B}_m^U - \Gamma_\Lambda VA^{-1} (\bar{B}_d^U - \mathbb{D}_{12}\mathbb{D}_{22}^{-1}\bar{B}_q^U)]\end{aligned}$$

with  $\mathbb{D}_{12}$  and  $\mathbb{D}_{22}$  given in (42). Therefore,

$$\begin{aligned}
J(\hat{\theta}_2) &= N g_n(\hat{\theta}_2)' \left( \hat{\Omega}(\hat{\theta}_1) \right)^{-1} g_n(\hat{\theta}_2) \\
&\xrightarrow{d} G \left\{ U \bar{B}_m^U - \Gamma_\Lambda V A^{-1} (\bar{B}_d^U - \mathbb{D}_{12} \mathbb{D}_{22}^{-1} \bar{B}_q^U) \right\}' \times \Lambda (\Lambda D_\infty \Lambda')^{-1} \Lambda \\
&\times \left\{ U \bar{B}_m^U - \Gamma_\Lambda V A^{-1} (\bar{B}_d^U - \mathbb{D}_{12} \mathbb{D}_{22}^{-1} \bar{B}_q^U) \right\} \\
&= G \left\{ \bar{B}_m^U - U' \Gamma_\Lambda V A^{-1} (\bar{B}_d^U - \mathbb{D}_{12} \mathbb{D}_{22}^{-1} \bar{B}_q^U) \right\}' U' D_\infty^{-1} U \\
&\times \left\{ \bar{B}_m^U - U' \Gamma_\Lambda V A^{-1} (\bar{B}_d^U - \mathbb{D}_{12} \mathbb{D}_{22}^{-1} \bar{B}_q^U) \right\} \\
&= G \left\{ \bar{B}_m^U - \begin{bmatrix} I_{d \times d} \\ O_{q \times d} \end{bmatrix} (\bar{B}_d^U - \mathbb{D}_{12} \mathbb{D}_{22}^{-1} \bar{B}_q^U) \right\}' \mathbb{D}_\infty^{-1} \\
&\times \left\{ \bar{B}_m^U - \begin{bmatrix} I_{d \times d} \\ O_{q \times d} \end{bmatrix} (\bar{B}_d^U - \mathbb{D}_{12} \mathbb{D}_{22}^{-1} \bar{B}_q^U) \right\} \\
&= G \begin{pmatrix} \mathbb{D}_{12} \mathbb{D}_{22}^{-1} \bar{B}_q^U \\ \bar{B}_q^U \end{pmatrix}' \mathbb{D}_\infty^{-1} \begin{pmatrix} \mathbb{D}_{12} \mathbb{D}_{22}^{-1} \bar{B}_q^U \\ \bar{B}_q^U \end{pmatrix} \\
&= G (\bar{B}_q^U)' \mathbb{D}_{22}^{-1} \bar{B}_q^U = {}^d G \bar{B}_q' S_{\infty, 22}^{-1} \bar{B}_q = \kappa_G,
\end{aligned}$$

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

*Proof of Proposition 3.* It follows from

$$\sqrt{N}(\hat{\theta}_2 - \theta_0) \xrightarrow{d} V A^{-1} \sqrt{G} [\bar{B}_d^U - \mathbb{D}_{12} \mathbb{D}_{22}^{-1} \bar{B}_q^U] \quad \text{and} \quad \hat{\Omega}(\hat{\theta}_1) \xrightarrow{d} \Lambda D_\infty \Lambda'$$

jointly that

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

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

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

Then

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

which implies that

$$\tilde{\mathbb{D}}_{11} := \begin{pmatrix} \tilde{\mathbb{D}}_{pp} & \tilde{\mathbb{D}}_{p,d-p} \\ \tilde{\mathbb{D}}_{d-p,p} & \tilde{\mathbb{D}}_{d-p,d-p} \end{pmatrix} \stackrel{d}{=} \frac{1}{G} \sum_{g=1}^G (B_{d,g} - \bar{B}_d) (B_{d,g} - \bar{B}_d)' + \begin{pmatrix} \tilde{V}' \beta_{\tilde{W}} \\ I_q \end{pmatrix} \bar{B}_q \bar{B}_q' \begin{pmatrix} \tilde{V}' \beta_{\tilde{W}} \\ I_q \end{pmatrix}', \quad (43)$$

and

$$\tilde{\mathbb{D}}_{12} := \begin{pmatrix} \tilde{\mathbb{D}}_{pq} \\ \tilde{\mathbb{D}}_{d-p,q} \end{pmatrix} \stackrel{d}{=} \frac{1}{G} \sum_{g=1}^G (B_{d,g} - \bar{B}_d) (B_{q,g} - \bar{B}_q)' + \begin{pmatrix} \tilde{V}' \beta_{\tilde{W}} \\ I_q \end{pmatrix} \bar{B}_q \bar{B}_q'. \quad (44)$$

Now

$$\begin{aligned} &F_{2, \hat{\Omega}(\hat{\theta}_1)}(\hat{\theta}_2) \\ &\stackrel{d}{\rightarrow} G (\bar{B}_d^U - \mathbb{D}_{12} \mathbb{D}_{22}^{-1} \bar{B}_q^U)' \tilde{V} \tilde{\Sigma}' \tilde{U}' \left\{ \tilde{U} \tilde{\Sigma} \tilde{V}' \mathbb{D}_{11.2} \tilde{V} \tilde{\Sigma}' \tilde{U}' \right\}^{-1} \tilde{U} \tilde{\Sigma} \tilde{V}' (\bar{B}_d^U - \mathbb{D}_{12} \mathbb{D}_{22}^{-1} \bar{B}_q^U) / p \\ &= G (\bar{B}_d^U - \mathbb{D}_{12} \mathbb{D}_{22}^{-1} \bar{B}_q^U)' \tilde{V} \tilde{\Sigma}' \cdot \left\{ \tilde{\Sigma} \tilde{V}' \mathbb{D}_{11.2} \tilde{V} \tilde{\Sigma}' \right\}^{-1} \cdot \tilde{\Sigma} \tilde{V}' (\bar{B}_d^U - \mathbb{D}_{12} \mathbb{D}_{22}^{-1} \bar{B}_q^U) / p \\ &= G (\tilde{V}' \bar{B}_d^U - \tilde{\mathbb{D}}_{12} \tilde{\mathbb{D}}_{22}^{-1} \bar{B}_q^U)' \cdot \tilde{\Sigma}' \left\{ \tilde{\Sigma} \tilde{\mathbb{D}}_{11.2} \tilde{\Sigma}' \right\}^{-1} \tilde{\Sigma} \cdot (\tilde{V}' \bar{B}_d^U - \tilde{\mathbb{D}}_{12} \tilde{\mathbb{D}}_{22}^{-1} \bar{B}_q^U) / p \\ &= G (\tilde{V}' \bar{B}_d^U - \tilde{\mathbb{D}}_{12} \tilde{\mathbb{D}}_{22}^{-1} \bar{B}_q^U)' \cdot \tilde{\Sigma}' \left\{ \tilde{\Sigma} (\tilde{\mathbb{D}}_{11} - \tilde{\mathbb{D}}_{12} \tilde{\mathbb{D}}_{22}^{-1} \tilde{\mathbb{D}}_{21}) \tilde{\Sigma}' \right\}^{-1} (\tilde{V}' \bar{B}_d^U - \tilde{\mathbb{D}}_{12} \tilde{\mathbb{D}}_{22}^{-1} \bar{B}_q^U) / p \\ &\stackrel{d}{=} G \left[ \bar{B}_p - \tilde{\mathbb{D}}_{pq} \tilde{\mathbb{D}}_{qq}^{-1} \bar{B}_q \right]' \tilde{A}' \left\{ \tilde{A} (\tilde{\mathbb{D}}_{pp} - \tilde{\mathbb{D}}_{pq} \tilde{\mathbb{D}}_{qq}^{-1} \tilde{\mathbb{D}}_{qp}) \tilde{A}' \right\}^{-1} \tilde{A} \left[ \bar{B}_p - \tilde{\mathbb{D}}_{pq} \tilde{\mathbb{D}}_{qq}^{-1} \bar{B}_q \right] / p \\ &\stackrel{d}{=} G \left[ \bar{B}_p - \tilde{\mathbb{D}}_{pq} \tilde{\mathbb{D}}_{qq}^{-1} \bar{B}_q \right]' \left( \tilde{\mathbb{D}}_{pp} - \tilde{\mathbb{D}}_{pq} \tilde{\mathbb{D}}_{qq}^{-1} \tilde{\mathbb{D}}_{qp} \right)^{-1} \left[ \bar{B}_p - \tilde{\mathbb{D}}_{pq} \tilde{\mathbb{D}}_{qq}^{-1} \bar{B}_q \right] / p, \end{aligned}$$

where  $\tilde{\mathbb{D}}_{pq}$ ,  $\tilde{\mathbb{D}}_{qq}$ , and  $\tilde{\mathbb{D}}_{qp}$  in the last two equalities are understood to equal the corresponding components on the right hand sides of (43) and (44). Here we have abused the notation a little bit.

We have

$$\begin{pmatrix} \tilde{\mathbb{D}}_{pp} & \tilde{\mathbb{D}}_{pq} \\ \tilde{\mathbb{D}}'_{pq} & \tilde{\mathbb{D}}_{qq} \end{pmatrix} = \begin{pmatrix} \tilde{S}_{pp} & \tilde{S}_{pq} \\ \tilde{S}'_{pq} & \tilde{S}_{qq} \end{pmatrix} + \tilde{w} \bar{B}_q \bar{B}_q' \tilde{w}' \quad (45)$$

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  $F_{2, \hat{\Omega}(\hat{\theta}_1)}(\hat{\theta}_2)$  holds. Direct calculations show that the second representation is numerically identical to the first representation. This completes the proof of Proposition 3. ■

*Proof of Lemma 1.* The centered CCE  $\hat{\Omega}^c(\tilde{\theta})$  can be represented as:

$$\begin{aligned} \hat{\Omega}^c(\tilde{\theta}) &= \frac{1}{G} \sum_{h=1}^G \left\{ \frac{1}{\sqrt{L}} \sum_{i=1}^L \left( f_i^h(\tilde{\theta}) - \frac{1}{N} \sum_{g=1}^G \sum_{s=1}^L f_s^g(\tilde{\theta}) \right) \right. \\ &\quad \left. \times \frac{1}{\sqrt{L}} \sum_{j=1}^L \left( f_j^h(\tilde{\theta}) - \frac{1}{N} \sum_{g=1}^G \sum_{s=1}^L f_s^g(\tilde{\theta}_N) \right) \right\}. \end{aligned}$$

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

$$\begin{aligned} &\frac{1}{\sqrt{L}} \sum_{i=1}^L \left( f_i^h(\tilde{\theta}) - \frac{1}{N} \sum_{g=1}^G \sum_{s=1}^L f_s^g(\tilde{\theta}) \right) \\ &= \frac{1}{\sqrt{L}} \sum_{i=1}^L \left( f_i^h(\theta_0) - \frac{1}{N} \sum_{g=1}^G \sum_{s=1}^L f_s^g(\theta_0) \right) (1 + o_p(1)) \end{aligned} \tag{46}$$

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

$$\frac{1}{\sqrt{L}} \sum_{i=1}^L f_i^h(\tilde{\theta}) = (1 + o_p(1)) \left( \frac{1}{\sqrt{L}} \sum_{i=1}^L f_i^h(\theta_0) + \frac{1}{L} \sum_{i=1}^L \frac{\partial f_i^h(\tilde{\theta})}{\partial \theta'} \sqrt{L}(\tilde{\theta} - \theta_0) \right)$$

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

$$\frac{1}{\sqrt{L}} \sum_{i=1}^L f_i^h(\tilde{\theta}) = (1 + o_p(1)) \left( \frac{1}{\sqrt{L}} \sum_{i=1}^L f_i^h(\theta_0) + \Gamma \sqrt{L}(\tilde{\theta} - \theta_0) \right)$$

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

$$\begin{aligned} &\frac{1}{\sqrt{L}} \sum_{i=1}^L \left( f_i^h(\tilde{\theta}) - \frac{1}{N} \sum_{g=1}^G \sum_{s=1}^L f_s^g(\tilde{\theta}) \right) \\ &= (1 + o_p(1)) \left( \frac{1}{\sqrt{L}} \sum_{i=1}^L f_i^h(\theta_0) - \frac{1}{G} \sum_{g=1}^G \frac{1}{\sqrt{L}} \sum_{i=1}^L f_i^g(\theta_0) \right), \end{aligned}$$

which completes the proof of part (a).

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

$$\frac{1}{\sqrt{L}} \sum_{i=1}^L f_i^h(\theta_0) - \frac{1}{G} \sum_{g=1}^G \frac{1}{\sqrt{L}} \sum_{i=1}^L f_i^g(\theta_0) \xrightarrow{d} \Lambda(B_{m,h} - \bar{B}_m),$$

where the convergence holds jointly for  $h = 1, \dots, G$ . As a result,

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

■

*Proof of Proposition 4.* The proof of part (a) is essentially the same as the proof of Proposition 3. The only difference is that the second term in (45) 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 2(e).

To prove part (c), recall that the restricted two-step GMM estimator  $\hat{\theta}_{2N}^{c,r}$  minimizes

$$g_n(\theta)' \left[ \hat{\Omega}^c(\hat{\theta}_1) \right]^{-1} g_n(\theta)/2 + \lambda_n'(R\theta - r). \quad (47)$$

The first order conditions are

$$\begin{aligned} \hat{\Gamma}(\hat{\theta}_2^{c,r}) \left[ \hat{\Omega}^c(\hat{\theta}_1) \right]^{-1} g_n(\hat{\theta}_2^{c,r}) + R' \lambda_n &= 0, \\ R \hat{\theta}_2^{c,r} - r &= 0. \end{aligned} \quad (48)$$

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

$$\begin{aligned} \sqrt{N}(\hat{\theta}_2^{c,r} - \theta_0) &= -\tilde{\Phi}^{-1} \Gamma' \left[ \hat{\Omega}^c(\hat{\theta}_1) \right]^{-1} \sqrt{N} g_n(\theta_0) \\ &\quad - \tilde{\Phi}^{-1} R' \left( R \tilde{\Phi}^{-1} R' \right)^{-1} R \tilde{\Phi}^{-1} \Gamma_n' \left[ \hat{\Omega}_n^c(\hat{\theta}_1) \right]^{-1} \sqrt{N} g_n(\theta_0) + o_p(1), \end{aligned} \quad (49)$$

where  $\tilde{\Phi} := \Gamma' \left[ \hat{\Omega}_n^c(\hat{\theta}_1) \right]^{-1} \Gamma$ . Subtracting (49) from (8), we have

$$\sqrt{N}(\hat{\theta}_2^c - \hat{\theta}_2^{c,r}) = -\tilde{\Phi}^{-1} R' \left( R \tilde{\Phi}^{-1} R' \right)^{-1} R \tilde{\Phi}^{-1} \Gamma' \left[ \hat{\Omega}^c(\hat{\theta}_1) \right]^{-1} \sqrt{N} g_n(\theta_0) + o_p(1). \quad (50)$$

By Taylor expansion and Assumption 3 we have

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

and

$$\begin{aligned} N g_n'(\hat{\theta}_2^c) \left[ \hat{\Omega}^c(\hat{\theta}_1) \right]^{-1} g_n(\hat{\theta}_2^c) - N g_n'(\hat{\theta}_2^{c,r}) \left[ \hat{\Omega}^c(\hat{\theta}_1) \right]^{-1} g_n(\hat{\theta}_2^c) \\ = N(\hat{\theta}_2^c - \hat{\theta}_2^{c,r})' \hat{\Gamma}'(\hat{\theta}_2^c) \left[ \hat{\Omega}^c(\hat{\theta}_1) \right]^{-1} g_n(\hat{\theta}_1) + O_p\left(\frac{1}{\sqrt{N}}\right) \\ = O_p\left(\frac{1}{\sqrt{N}}\right). \end{aligned} \quad (51)$$

Here the last equality follows from the FOC's for  $\hat{\theta}_2^c$ . In a similar way, we can write the second term

in (51) as

$$\begin{aligned} & N g_n'(\hat{\theta}_2^{c,r}) \left[ \hat{\Omega}^c(\hat{\theta}_1) \right]^{-1} g_n(\hat{\theta}_2^c) \\ &= N g_n'(\hat{\theta}_2^{c,r}) \left[ \hat{\Omega}^c(\hat{\theta}_1) \right] g_n(\hat{\theta}_2^{c,r}) + N(\hat{\theta}_2^c - \hat{\theta}_2^{c,r})' \tilde{\Phi}(\hat{\theta}_2^c - \hat{\theta}_2^{c,r}) + o_p(1). \end{aligned}$$

Combining this and (50), we get

$$\begin{aligned} & LR_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c, \hat{\theta}_2^{c,r}) \\ &= \left\{ N g_n(\hat{\theta}_2^c)' \left[ \hat{\Omega}^c(\hat{\theta}_1) \right]^{-1} g_n(\hat{\theta}_2^c) - N g_n(\hat{\theta}_2^{c,r})' \left[ \hat{\Omega}^c(\hat{\theta}_1) \right]^{-1} g_n(\hat{\theta}_2^{c,r}) \right\} / p \\ &= N(\hat{\theta}_2^{c,r} - \hat{\theta}_2^c)' \tilde{\Phi}(\hat{\theta}_2^{c,r} - \hat{\theta}_2^c) / p + o_p(1) \\ &= \sqrt{N} g_n'(\theta_0) \left[ \hat{\Omega}^c(\theta_0) \right]^{-1} \Gamma \tilde{\Phi}^{-1} R' \left( R \tilde{\Phi}^{-1} R' \right)^{-1} R \tilde{\Phi}^{-1} \Gamma' \sqrt{N} g_n(\theta_0) / p + o_p(1) \\ &= \sqrt{N} \left( R \hat{\theta}_2^c - r \right)' \left( R \tilde{\Phi}^{-1} R' \right)^{-1} \sqrt{N} \left( R \hat{\theta}_2^c - r \right) / p + o_p(1) \\ &= F_{2, \hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c) + o_p(1). \end{aligned}$$

as desired.

To prove part (d), we rewrite the FOC in (48) as

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

So,

$$\begin{aligned} LM_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^{c,r}) &= N \left[ \Delta_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^{c,r}) \right]' \tilde{\Phi}^{-1} \left[ \Delta_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^{c,r}) \right] / p \\ &= N \left( \hat{\theta}_2^c - \hat{\theta}_2^{c,r} \right)' \tilde{\Phi} \sqrt{N} \left( \hat{\theta}_2^c - \hat{\theta}_2^{c,r} \right) / p + o_p(1) \\ &= LR_{\hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c, \hat{\theta}_2^{c,r}) + o_p(1) \\ &= F_{2, \hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c) + o_p(1). \end{aligned}$$

■

*Proof of Proposition 5.* 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 1 to obtain  $\Omega_N^c(\hat{\theta}_{\text{CU-GEE}}) = \Omega_N^c(\theta_0) + o_p(1)$ .

Invoking the continuous mapping theorem yields

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

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

$$0 = \hat{\Gamma}^j(\hat{\theta}_{\text{CU-GMM}})' \left[ \hat{\Omega}^c(\hat{\theta}_{\text{CU-GMM}}) \right]^{-1} g_n(\hat{\theta}_{\text{CU-GMM}}) - g_n(\hat{\theta}_{\text{CU-GMM}})' \left[ \hat{\Omega}^c(\hat{\theta}_{\text{CU-GMM}}) \right]^{-1} \Upsilon_j(\hat{\theta}_{\text{CU-GMM}}) \left[ \hat{\Omega}^c(\hat{\theta}_{\text{CU-GMM}}) \right]^{-1} g_n(\hat{\theta}_{\text{CU-GMM}}), \quad (52)$$

where

$$\Upsilon_j(\theta) = \frac{1}{N} \sum_{g=1}^G \left( \sum_{r=1}^L f_r^g(\theta) \right) \left( \sum_{s=1}^L \frac{\partial f_s(\theta)}{\partial \theta_j} \right)' - L_N g_N(\theta) \left( \frac{\partial g_N(\theta)}{\partial \theta_j} \right)'.$$

The second term in (52) can be written as

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

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

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

and for each  $g = 1, \dots, G$ ,

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

Combining these together, the second term in FOC in (52) is  $o_p(L_N^{-1/2})$ . 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

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

■

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

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

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

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

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

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

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

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

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

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

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

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

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

It then follows that

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

That is

$$\frac{G - p - q}{pG} \xi_p' \left( \tilde{S}_{pp\cdot q} \right)^{-1} \xi_p \sim F_{p, G-p-q}.$$

Together with Proposition 4(c)(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 Theorem 2.* We first show that  $\hat{\mathcal{E}}_n = \mathcal{E}_{2n}(1 + o_p(1))$ . For each  $j = 1, \dots, d$ , we have

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

where the second equality holds by Assumption 3, 5 and Lemma 1. Using a Taylor expansion, we have

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

So

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

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

$$\begin{aligned} \left. \frac{\partial \hat{\Omega}^c(\theta)}{\partial \theta_j} \right|_{\theta = \hat{\theta}_1} &= \Upsilon_j(\hat{\theta}_1) + \Upsilon'_j(\hat{\theta}_1), \\ \Upsilon_j(\theta) &= \frac{1}{N} \sum_{g=1}^G \left[ \sum_{r=1}^L \left( f_r^g(\theta) - \frac{1}{N} \sum_{s=1}^N f_s(\theta) \right) \left( \sum_{s=1}^L \left( \frac{\partial f_s^g(\theta)}{\partial \theta_j} - \frac{1}{N} \sum_{s=1}^N \frac{\partial f_s(\theta)}{\partial \theta_j} \right) \right) \right]'. \end{aligned}$$

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

$$\begin{aligned} &\frac{1}{\sqrt{L}} \sum_{r=1}^L \left( f_r^g(\hat{\theta}_1) - \frac{1}{N} \sum_{s=1}^N f_s(\hat{\theta}_1) \right) \\ &= \frac{1}{\sqrt{L}} \sum_{r=1}^L \left( f_r^g(\theta_0) - \frac{1}{N} \sum_{s=1}^N f_s(\theta_0) \right) (1 + o_p(1)) \end{aligned} \quad (53)$$

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

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

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

$$\begin{aligned} &\sum_{s=1}^L \left( \frac{\partial f_s^g(\hat{\theta}_1)}{\partial \theta_j} - \frac{1}{N} \sum_{s=1}^N \frac{\partial f_s(\hat{\theta}_1)}{\partial \theta_j} \right) \\ &= \sum_{s=1}^L \left( \frac{\partial f_s^g(\theta_0)}{\partial \theta_j} - \frac{1}{N} \sum_{s=1}^N \frac{\partial f_s(\theta_0)}{\partial \theta_j} \right) (1 + o_p(1)) \end{aligned} \quad (54)$$

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

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

Now, define the infeasible corrected variance

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

and the corresponding infeasible Wald statistic

$$\mathbb{F}_{2, \hat{\Omega}^c(\hat{\theta}_1)}^{\mathbf{c}}(\hat{\theta}_2^c) = (R\hat{\theta}_2^c - r)' \left[ R \widehat{\text{var}}_{\hat{\Omega}^c(\hat{\theta}_1)}^{\mathbf{c}, \text{inf}}(\hat{\theta}_2^c) R' \right]^{-1} (R\hat{\theta}_2^c - r)/p.$$

The result in (55) implies

$$F_{2, \hat{\Omega}^c(\hat{\theta}_1)}^{\mathbf{c}}(\hat{\theta}_2^c) = \mathbb{F}_{2, \hat{\Omega}^c(\hat{\theta}_1)}^{\mathbf{c}}(\hat{\theta}_2^c)(1 + o_p(1)).$$

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

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

and so

$$\begin{aligned} F_{2, \hat{\Omega}^c(\hat{\theta}_1)}^c(\hat{\theta}_2^c) &= \mathbb{F}_{2, \hat{\Omega}^c(\hat{\theta}_1)}^{\mathbf{c}}(\hat{\theta}_2^c) + o_p(1) \\ &= F_{2, \hat{\Omega}^c(\hat{\theta}_1)}(\hat{\theta}_2^c) + o_p(1). \end{aligned}$$

■