Supplemental Appendix for "On Standard Inference for GMM with Seeming Local Identification Failure"

Ji Hyung Lee * Zhipeng Liao†

December, 2014

Abstract

This supplemental appendix provides some auxiliary materials for "On Standard Inference for GMM with Seeming Local Identification Failure" (Lee and Liao, 2014; cited as LL in this appendix).

The notation in this appendix is consistent with that in LL. We continue to use DR to denote Dovonon and Renault (2013). Throughout the appendix, C denotes some generic finite positive constant; $\|\cdot\|$ denotes the Euclidean norm; A' denotes the transpose of a matrix A; $\rho_{\max}(A)$ and $\rho_{\min}(A)$ denote the largest and smallest eigenvalues of a matrix A, respectively; for any square matrix A, $A \geq 0$ means that A is a positive semi-definite matrix; for any positive integers k_1 and k_2, I_{k_1} denotes the $k_1 \times k_1$ identity matrix and $\mathbf{0}_{k_1 \times k_2}$ denotes the $k_1 \times k_2$ zero matrix; $A \equiv B$ means that A is defined as B; $a_n = o_p(b_n)$ means that for any constants $\epsilon_1, \epsilon_2 > 0$, there is $\Pr(|a_n/b_n| \geq \epsilon_1) < \epsilon_2$ eventually; $a_n = O_p(b_n)$ means that for any $\epsilon > 0$, there is a finite constant C_ϵ such that $\Pr(|a_n/b_n| \geq C_\epsilon) < \epsilon$ eventually; for any two sequences a_n and b_n, we use $a_n \lesssim b_n$ to denote that $a_n \leq Cb_n$ where C is some fixed finite positive constant; "$\rightarrow_p$" and "$\rightarrow_d$" denote convergence in probability and convergence in distribution, respectively; and w.p.a.1 abbreviates with probability approaching 1.

1 GMM Inference in Common CH Factor Model

In this Section we investigate new GMM estimation and tests for the common CH features proposed in LL. Following Assumption 3.5 and 3.6 in LL, we explicitly use parameter space Θ^* such that

$$\theta_* = \left(\theta_1, \ldots, \theta_{n-1}, 1 - \sum\nolimits_{i=1}^{n-1} \theta_i\right)' = \left(\theta', 1 - \sum\nolimits_{i=1}^{n-1} \theta_i\right)' = G_2\theta + l_n.$$

*Department of Economics, University of Washington, 336 Savery Hall, Box 353330, Seattle, WA 98195. Email: jihyung2@uw.edu

†Department of Economics, UC Los Angeles, 8379 Bunche Hall, Mail Stop: 147703, Los Angeles, CA 90095. Email: zhipeng.liao@econ.ucla.edu

To observe the moment conditions in (2.4) of LL for the common CH model, note that

$$m_t(\theta) = \begin{bmatrix} \psi_t(\theta) \\ g_t(\theta) \end{bmatrix} = \begin{bmatrix} (z_t - \mu_z)\left(\theta_*' Y_{t+1} Y_{t+1}' \theta_*\right) \\ vec\left(\left(G_2' Y_{t+1} Y_{t+1}' \theta_*'\right) 2 (z_t - \mu_z)'\right) \end{bmatrix} = \begin{bmatrix} (z_t - \mu_z)\left(\theta' Y_{t+1} Y_{t+1}' \theta\right) \\ (2 (z_t - \mu_z) \otimes I_p) G_2' Y_{t+1} Y_{t+1}' \theta_* \end{bmatrix},$$

and $E[m_t(\theta_0)] = 0$.

To introduce the feasible moment conditions, we replace the nuisance parameter μ_z by its consistent estimator $\bar{z} = T^{-1} \sum z_t$ so

$$\widehat{m}_t(\theta) \equiv \begin{bmatrix} \widehat{\psi}_t(\theta) \\ \widehat{g}_t(\theta) \end{bmatrix} \equiv \begin{bmatrix} (z_t - \bar{z})\left(\theta_*' Y_{t+1} Y_{t+1}' \theta_*\right) \\ vec\left(\left(G_2' Y_{t+1} Y_{t+1}' \theta_*'\right) 2 (z_t - \bar{z})'\right) \end{bmatrix} = \begin{bmatrix} (z_t - \bar{z})\left(\theta_*' Y_{t+1} Y_{t+1}' \theta_*\right) \\ (2 (z_t - \bar{z}) \otimes I_p) G_2' Y_{t+1} Y_{t+1}' \theta_* \end{bmatrix}.$$

Strictly speaking $\left(\hat{\psi}_t(\theta), \widehat{g}_t(\theta)\right)$ are triangular arrays hence $\left(\hat{\psi}_t(\theta), \widehat{g}_t(\theta)\right) = \left(\hat{\psi}_{t,T}(\theta), \widehat{g}_{t,T}(\theta)\right)$. However, standard CLT for these triangular arrays will hold since it is easy to show

$$\begin{aligned} T^{-\frac{1}{2}} \sum_{t=1}^{T} \hat{\psi}_{t,T}(\theta) &= T^{-\frac{1}{2}} \sum_{t=1}^{T} (z_t - \bar{z})\left(\theta_*' Y_{t+1} Y_{t+1}' \theta_*\right) \\ &= T^{-\frac{1}{2}} \sum_{t=1}^{T} (z_t - \mu_z)\left\{\left(\theta_*' Y_{t+1} Y_{t+1}' \theta_*\right) - E\left[\left(\theta_*' Y_{t+1} Y_{t+1}' \theta_*\right)\right]\right\} + O_p\left(T^{-\frac{1}{2}}\right). \end{aligned}$$

Note that

$$T^{-\frac{1}{2}} \sum_{t=1}^{T} (z_t - \bar{z})\left(\theta_*' Y_{t+1}\right)^2 = T^{-\frac{1}{2}} \sum_{t=1}^{T} (z_t - \mu_z)\left(\theta_*' Y_{t+1}\right)^2 - (\bar{z} - \mu_z) T^{-\frac{1}{2}} \sum_{t=1}^{T} \left(\theta_*' Y_{t+1}\right)^2,$$

and the last term is

$$\begin{aligned} (\bar{z} - \mu_z) T^{-\frac{1}{2}} \sum_{t=1}^{T} \left(\theta_*' Y_{t+1}\right)^2 &= (\bar{z} - \mu_z) T^{-\frac{1}{2}} \sum_{t=1}^{T} \left\{\left(\theta_*' Y_{t+1}\right)^2 - E\left[\left(\theta_*' Y_{t+1}\right)^2\right] + E\left[\left(\theta_*' Y_{t+1}\right)^2\right]\right\} \\ &= (\bar{z} - \mu_z) T^{-\frac{1}{2}} \sum_{t=1}^{T} \left\{\left(\theta_*' Y_{t+1}\right)^2 - E\left[\left(\theta_*' Y_{t+1}\right)^2\right]\right\} + T^{\frac{1}{2}} (\bar{z} - \mu_z) E\left[\left(\theta_*' Y_{t+1}\right)^2\right] \\ &= T^{\frac{1}{2}} (\bar{z} - \mu_z) E\left[\left(\theta_*' Y_{t+1}\right)^2\right] + O_p\left(T^{-\frac{1}{2}}\right). \end{aligned}$$

Therefore

$$\begin{aligned} T^{-\frac{1}{2}} \sum_{t=1}^{T} \hat{\psi}_{t,T}(\theta) &= T^{-\frac{1}{2}} \sum_{t=1}^{T} (z_t - \mu_z)\left(\theta_*' Y_{t+1}\right)^2 - T^{\frac{1}{2}} (\bar{z} - \mu_z) E\left[\left(\theta_*' Y_{t+1}\right)^2\right] + O_p\left(T^{-\frac{1}{2}}\right) \\ &= T^{-\frac{1}{2}} \sum_{t=1}^{T} (z_t - \mu_z)\left\{\left(\theta_*' Y_{t+1}\right)^2 - E\left[\left(\theta_*' Y_{t+1}\right)^2\right]\right\} + O_p\left(T^{-\frac{1}{2}}\right). \end{aligned}$$

In an exactly same way,

$$\begin{aligned}
T^{-\frac{1}{2}}\sum_{t=1}^{T}\widehat{g}_{t,T}(\theta) &= T^{-\frac{1}{2}}\sum_{t=1}^{T}\left(2\left(z_t-\bar{z}\right)\otimes I_p\right)G_2'Y_{t+1}Y_{t+1}'\theta_* \\
&= T^{-\frac{1}{2}}\sum_{t=1}^{T}\left(2\left(z_t-\mu_z\right)\otimes I_p\right)G_2'\left\{\left(Y_{t+1}Y_{t+1}'\theta_*\right)-E\left[\left(Y_{t+1}Y_{t+1}'\theta_*\right)\right]\right\}+O_p\left(T^{-\frac{1}{2}}\right).
\end{aligned}$$

Therefore

$$\begin{aligned}
T^{-\frac{1}{2}}\sum_{t=1}^{T}\left[\begin{array}{c}\hat{\psi}_{t,T}(\theta)\\ \widehat{g}_{t,T}(\theta)\end{array}\right] &= T^{-\frac{1}{2}}\sum_{t=1}^{T}\left[\begin{array}{c}\left(z_t-\bar{z}\right)\left(\theta_*'Y_{t+1}Y_{t+1}'\theta_*\right)\\ \left(2\left(z_t-\bar{z}\right)\otimes I_p\right)G_2'Y_{t+1}Y_{t+1}'\theta_*\end{array}\right] \\
&= T^{-\frac{1}{2}}\sum_{t=1}^{T}\left[\begin{array}{c}\left(z_t-\mu_z\right)\left\{\left(\theta_*'Y_{t+1}\right)^2-E\left[\left(\theta_*'Y_{t+1}\right)^2\right]\right\}\\ \left(2\left(z_t-\mu_z\right)\otimes I_p\right)G_2'\left\{\left(Y_{t+1}Y_{t+1}'\theta_*\right)-E\left[\left(Y_{t+1}Y_{t+1}'\theta_*\right)\right]\right\}\end{array}\right]+O_p\left(T^{-\frac{1}{2}}\right).
\end{aligned}$$

To keep notation simple, we use $\hat{m}_t(\theta)=\left(\hat{\psi}_t(\theta)',\widehat{g}_t(\theta)'\right)'$ rather than array notation unless any confusion arises.

Assumption 3.4 in LL and CLT for martingale difference sequences $\left\{\theta_*^{0\prime}Y_{t+1}Y_{t+1}'\theta_*^0,\mathcal{F}_t\right\}_{t\geq 1}$ based on the common feature θ_0 essentially deliver the following limit theory.

Corollary 1.1 *Under Assumptions 3.1-6 in LL,*

$$T^{-\frac{1}{2}}\sum_{t=1}^{T}\hat{m}_t(\theta_0)=T^{-\frac{1}{2}}\sum_{t=1}^{T}\left[\begin{array}{c}\hat{\psi}_t(\theta_0)\\ \widehat{g}_t(\theta_0)\end{array}\right]\rightarrow_d N\left(0,\Omega_m\right) \tag{1.1}$$

where the long-run variance $\Omega_m=\lim_{T\to\infty}Var\left[T^{-\frac{1}{2}}\sum_{t=1}^{T}\hat{m}_t(\theta_0)\right]$ *can be partitioned as*

$$\Omega_m=\left(\begin{array}{cc}\Omega_\psi & \Omega_{\psi g}\\ \Omega_{g\psi} & \Omega_g\end{array}\right). \tag{1.2}$$

Moreover,

$$T^{-1}\sum_{t=1}^{T}\frac{\partial\widehat{g}_t(\theta_0)}{\partial\theta}=E\left[\frac{\partial g_t(\theta_0)}{\partial\theta}\right]+o_p(1)=\mathbb{H}+o_p(1) \tag{1.3}$$

Corollary 1.1 with notation of (1.1), (1.2) will be used throughout this Section[1]. Note that $\mathbb{H}(\theta_0)$ is free of θ_0 so $\mathbb{H}(\theta_0)=\mathbb{H}$.

1.1 GMM Estimation and Limit Theory

In this subsection, the weight matrix $W_{j,T}$ for $j=m,g$ or g^* are random weight matrix with probability limits W_j for $j=m,g$ or g^*, respectively, hence $W_{m,T}=W_m+o_p(1)$ and so on.

[1] Under the common feature θ_0, it is clear that $\Omega_\psi=\Omega_\psi(\theta_0)=E[(z_t-\mu_z)(z_t-\mu_z)'\{(\theta_0'Y_{t+1})^2-E[(\theta_0'Y_{t+1})^2]\}^2]$ as given in Corollary 3.1 of DR. Similar expressions for $\Omega_{\psi g}$ and Ω_g can be easily obtained.

Using all the moment conditions $\hat{m}_t(\theta) = (\hat{\psi}_t'(\theta), \hat{g}_t'(\theta))'$, we define the stacked/full GMM estimator.

Definition 1.1 *The Full GMM estimator is defined as*

$$\widehat{\theta}_{m,T} = \arg\min_{\theta\in\Theta^*} \left[T^{-1}\sum_{t=1}^{T}\hat{m}_t(\theta)\right]' W_{m,T} \left[T^{-1}\sum_{t=1}^{T}\hat{m}_t(\theta)\right] \tag{1.4}$$

where $W_{m,T}$ is now $H(p+1)\times H(p+1)$ weight matrix.

We next consider using the Jacobian-based moment restrictions only, hence (3.9) in LL.

Definition 1.2 *The Jacobian GMM estimator is defined as*

$$\widehat{\theta}_{g,T} = \arg\min_{\theta\in\Theta^*} \left[T^{-1}\sum_{t=1}^{T}\hat{g}_t(\theta)\right]' W_{g,T} \left[T^{-1}\sum_{t=1}^{T}\hat{g}_t(\theta)\right] \tag{1.5}$$

where $W_{g,T}$ is now $Hp\times Hp$ positive definite weight matrix.

Finally, the modified GMM estimator $\widehat{\theta}_{g^*,T}$ based on the modified moment functions is defined.

Definition 1.3 *The Modified GMM estimator is defined as*

$$\widehat{\theta}_{g^*,T} = \arg\min_{\theta\in\Theta^*} \left[T^{-1}\sum_{t=1}^{T}\widehat{g}_{\psi,t}(\theta)\right]' W_{g^*,T} \left[T^{-1}\sum_{t=1}^{T}\widehat{g}_{\psi,t}(\theta)\right] \tag{1.6}$$

where $\widehat{g}_{\psi,t}(\theta) = g_t(\theta) - \widehat{\Omega}_{g\psi,T}\widehat{\Omega}_{\psi,T}^{-1}\psi_t(\widehat{\theta}_{g,T})$ and $W_{g^*,T} = \widehat{\Omega}_{g,T} - \widehat{\Omega}_{g\psi,T}\widehat{\Omega}_{\psi,T}^{-1}\widehat{\Omega}_{\psi g,T}$.

We present the limit theory of the newly suggested estimators.

1.1.1 Consistency

For consistency of $\widehat{\theta}_{m,T}$, denote

$$\begin{aligned} Q_T(\theta) &= -\left\|T^{-1}\sum_{t=1}^{T}\hat{m}_t(\theta)\right\|_{W_T} = -\left[T^{-1}\sum_{t=1}^{T}\hat{m}_t(\theta)\right]' W_T \left[T^{-1}\sum_{t=1}^{T}\hat{m}_t(\theta)\right] \\ \text{and } Q(\theta) &= -\left\|E[m_t(\theta)]\right\|_W = -[m_t(\theta_0)]' W [m_t(\theta_0)] \end{aligned}$$

by random and nonstochastic criterion function as in standard Extremum Estimators, respectively. Thanks to the polynomial structure of the moment restrictions in common CH factor model, uniform consistency of $T^{-1}\sum_{t=1}^{T}\hat{m}_t(\theta)$ over Θ directly follows (as pointed out by DR as well). Together with the identification condition given in Section 2, the standard consistency results of extremum estimators (see e.g., Hayashi (2000, Section7)) hold so we have $\widehat{\theta}_{m,T} = \theta_0 + o_p(1)$. Similar arguments hold for the other two estimators, thus consistency for all three estimators is obtained.

Proposition 1.1 *Under Assumptions 3.1-6 in LL, $\widehat{\theta}_{m,T}, \widehat{\theta}_{g,T}$ and $\widehat{\theta}_{g^*,T}$ all converge to θ_0 in probability.*

1.1.2 Rates of Convergence

We now discuss how to recover a regular $\sqrt{T}$-consistent estimator for θ_0 in our methods.

For simplicity, let us consider $n = 2$, $p = 1$ (two assets with one common feature) with the normalization so $(\theta_0, 1-\theta_0)$ is a unique common feature. Assume we use one instrument ($H = 1$). Then Corollary 1.1 provides

$$T^{-\frac{1}{2}} \sum_{t=1}^{T} \hat{m}_t(\theta_0) = T^{-\frac{1}{2}} \sum_{t=1}^{T} \begin{bmatrix} \hat{\psi}_t(\theta_0) \\ \frac{\partial \hat{\psi}_t(\theta_0)}{\partial \theta} = \hat{g}_t(\theta_0) \end{bmatrix}_{2\times 1} \Longrightarrow N(0, \Omega_m), \tag{1.7}$$

and

$$T^{-1} \sum_{t=1}^{T} \frac{\partial \hat{m}_t(\theta_0)}{\partial \theta} = T^{-1} \sum_{t=1}^{T} \begin{bmatrix} \frac{\partial \hat{\psi}_t(\theta_0)}{\partial \theta} \\ \frac{\partial^2 \hat{\psi}_t(\theta_0)}{\partial \theta^2} = \frac{\partial \hat{g}_t(\theta_0)}{\partial \theta} \end{bmatrix} = \begin{bmatrix} O_p\left(T^{-\frac{1}{2}}\right) \\ \mathbb{H} + o_p(1) \end{bmatrix} = \begin{bmatrix} 0 \\ \mathbb{H} \end{bmatrix} + o_p(1), \tag{1.8}$$

and

$$T^{-1} \sum_{t=1}^{T} \frac{\partial^2 \hat{m}_t(\theta_0)}{\partial \theta^2} = T^{-1} \sum_{t=1}^{T} \begin{bmatrix} \frac{\partial^2 \hat{\psi}_t(\theta_0)}{\partial \theta^2} = \frac{\partial \hat{g}_t(\theta_0)}{\partial \theta} \\ \frac{\partial^3 \hat{\psi}_t(\theta_0)}{\partial \theta^3} = 0 \end{bmatrix} = \begin{bmatrix} \mathbb{H} \\ 0 \end{bmatrix} + o_p(1), \tag{1.9}$$

where $\frac{\partial^3 \hat{\psi}_t(\theta_0)}{\partial \theta^3} = 0$ comes from the fact that $\hat{\psi}_t$ is the second order polynomial.

With any given consistent estimator $\hat{\theta}$, let $v = a_T\left(\hat{\theta} - \theta_0\right)$ where the normalizing sequence $a_T \to \infty$ will be specified below. The stabilizing order of magnitude a_T under the given limit theory will enable us to find the rate of convergence of the given consistent estimator $\hat{\theta}$.

GMM estimation in DR corresponds to the method of moments (one parameter with one moment restriction) based on $T^{-1}\sum_{t=1}^{T} \hat{\psi}_t(\theta)$ with the first order condition

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \hat{\psi}_t\left(\hat{\theta}_T\right) = 0.$$

From the second order Taylor expansion with a mean value $\bar{\theta}_T$ between θ_0 and $\hat{\theta}_T$,

$$0 = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \hat{\psi}_t\left(\theta_0 + \frac{1}{a_T} v\right) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \hat{\psi}_t(\theta_0) + \frac{1}{a_T\sqrt{T}} \sum_{t=1}^{T} \frac{\partial \hat{\psi}_t}{\partial \theta}(\theta_0)\, v + \left\{\frac{1}{2} \frac{1}{a_T^2\sqrt{T}} \sum_{t=1}^{T} \frac{\partial^2 \hat{\psi}_t}{\partial \theta^2}\left(\bar{\theta}_T\right)\right\} v^2$$

and from (1.7) and (1.8), $T^{-\frac{1}{2}} \sum_{t=1}^{T} \frac{\partial \hat{\psi}_t}{\partial \theta}(\theta_0) = O_p(1)$ and $T^{-1}\sum_{t=1}^{T} \frac{\partial^2 \hat{\psi}_t}{\partial \theta^2}\left(\bar{\theta}_T\right) = O_p(1)$ so $a_T = T^{1/4}$ is the correct order. In this case

$$o_p(1) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \hat{\psi}_t(\theta_0) + \frac{1}{2}\left\{\frac{1}{T} \sum_{t=1}^{T} \frac{\partial^2 \hat{\psi}_t}{\partial \theta^2}\left(\bar{\theta}_T\right)\right\} v^2$$

and

$$v^2 = \left\| T^{1/4}\left(\hat{\theta}_T - \theta_0\right)\right\|^2 = 2\mathbb{H}^{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\hat{\psi}_t\left(\theta_0\right)\right) + o_p(1) \Longrightarrow N\left(0, 4\mathbb{H}^{-1}\Omega_{\rho\rho}\right)$$

giving $\left\|\hat{\theta}_T - \theta_0\right\| = O_p\left(T^{-1/4}\right)$ as given in Proposition 3.1 of DR.

Full GMM based on $T^{-1}\sum_{t=1}^{T}\hat{m}_t\left(\theta\right)$ is over-identified (one parameter with two moment restrictions) in this case. The rigorous first order condition and derivation will be given in (1.10) and proof of Theorem 1.2 below. To present the idea, we symbolically use the following analogy:

$$0 = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\hat{m}_t\left(\theta_0 + \frac{1}{a_T}v\right) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\hat{m}_t\left(\theta_0\right) + \frac{1}{a_T\sqrt{T}}\sum_{t=1}^{T}\frac{\partial\hat{m}_t}{\partial\theta}\left(\theta_0\right)v + \left\{\frac{1}{2}\frac{1}{a_T^2\sqrt{T}}\sum_{t=1}^{T}\frac{\partial^2\hat{m}_t}{\partial\theta^2}\left(\bar{\theta}_T\right)\right\}v^2.$$

From (1.7), (1.8) and (1.9), $T^{-1}\sum_{t=1}^{T}\frac{\partial\hat{m}_t}{\partial\theta}\left(\theta_0\right) = O_p(1)$ and $T^{-1}\sum_{t=1}^{T}\frac{\partial^2\hat{m}_t}{\partial\theta^2}\left(\bar{\theta}_T\right) = O_p(1)$ so $a_T = T^{1/2}$ is the correct order. In this case

$$o_p(1) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\hat{m}_t\left(\theta_0\right) + \left(\frac{1}{T}\sum_{t=1}^{T}\begin{bmatrix}\frac{\partial\hat{\psi}_t\left(\bar{\theta}_T\right)}{\partial\theta} \\ \frac{\partial\hat{g}_t\left(\theta_0\right)}{\partial\theta}\end{bmatrix}\right)v$$

and $\left\|\hat{\theta}_T - \theta_0\right\| = O_p\left(T^{-1/2}\right)$ giving $\sqrt{T}$-consistency.

In sum, by using $T^{-1}\sum_{t=1}^{T}\left(\hat{\psi}_t\left(\theta\right), \frac{\partial\hat{\psi}_t(\theta)}{\partial\theta}\right)'$ instead of $T^{-1}\sum_{t=1}^{T}\hat{\psi}_t\left(\theta\right)$, the term of degree 1 (Jacobian term) in Taylor expansion becomes $O_p(T)$ rather than $O_p(T^{1/2})$ so $T^{1/2}$-rate of convergence is achieved rather than $T^{1/4}$-rate. The Jacobian-GMM estimator uses $T^{-1}\sum_{t=1}^{T}\frac{\partial\hat{\psi}_t(\theta)}{\partial\theta}$ so the term of degree 1 has the same order of magnitude as Full-GMM ($T^{-1}\sum_{t=1}^{T}\frac{\partial^2\hat{\psi}_t(\theta_0)}{\partial\theta^2} = O_p(1)$ from (1.8). By construction, the term of degree 1 of the Modified GMM estimator is same as that of the Jacobian GMM so they share the same $\sqrt{T}$-rate of convergence. Generalization to any $p = n - 1$ with $H \geq p$ is straightforward with vectorization, so the following rate of convergence is obtained for all three estimators.

Proposition 1.2 *Under Assumptions 3.1-6 in LL,* $\widehat{\theta}_{m,T}, \widehat{\theta}_{g,T}$ *and* $\widehat{\theta}_{g^*,T}$ *are* $\sqrt{T}$*-consistent estimators for* θ_0*.*

1.1.3 Asymptotic Normality

We present asymptotic distributions of all three estimators in this subsection. Thanks to the rank condition proved in Lemma 3.2 in LL and $\sqrt{T}$-consistency, all proofs for asymptotic distributions become standard.

The Jacobian-GMM estimator has a closed from solution as given in (3.9) in LL. To see this, using $\theta_* = G_2\theta + l_n$:

$$\hat{\psi}_t\left(\theta\right) = \left(z_t - \bar{z}\right)\left(\theta_*' Y_{t+1} Y_{t+1}' \theta_*\right)$$

hence

$$\begin{aligned}
\frac{\partial \hat{\psi}_t(\theta)}{\partial \theta'} &= 2(z_t - \bar{z})\left(\theta_*' Y_{t+1} Y_{t+1}' G_2\right), \text{ and} \\
\widehat{g}_t(\theta) &= Vec\left(\left(\frac{\partial \hat{\psi}_t(\theta)}{\partial \theta'}\right)'\right) = (2(z_t - \bar{z}) \otimes I_p) G_2' Y_{t+1} Y_{t+1}' \theta_* \\
&= (2(z_t - \bar{z}) \otimes I_p) G_2' Y_{t+1} Y_{t+1}' G_2 \theta + (2(z_t - \bar{z}) \otimes I_p) G_2' Y_{t+1} Y_{t+1}' l_n
\end{aligned}$$

so

$$T^{-1} \sum_{t=1}^{T} \hat{g}_t(\theta) = \mathbb{H}_T \theta + \mathbb{S}_T$$

as in (3.8) of LL. Moreover,

$$\mathbb{H}(\theta_0) = \mathbb{H} = E\left[\frac{\partial}{\partial \theta'} \widehat{g}_t(\theta)\right] = E\left[(2(z_t - \bar{z}) \otimes I_p) G_2' Y_{t+1} Y_{t+1}' G_2\right].$$

It is now easy to see that

$$\begin{aligned}
\widehat{\theta}_{g,T} &= \arg\min_{\theta \in \Theta^*} \left[T^{-1} \sum_{t=1}^{T} \hat{g}_t(\theta)\right]' W_{g,T} \left[T^{-1} \sum_{t=1}^{T} \hat{g}_t(\theta)\right] \\
&= \arg\min_{\theta \in R^p} \left[\mathbb{H}_T \theta + \mathbb{S}_T\right]' W_{g,T} \left[\mathbb{H}_T \theta + \mathbb{S}_T\right] \\
&= \widehat{\theta}_{g,T} = -\left(\mathbb{H}_T' W_{g,T} \mathbb{H}_T\right)^{-1} \mathbb{H}_T' W_{g,T} \mathbb{S}_T.
\end{aligned}$$

We have the following closed form expression after some straightforward algebra:

$$T^{1/2}\left(\widehat{\theta}_{g,T} - \theta_0\right) = -\left[\mathbb{H}_T' W_{g,T} \mathbb{H}_T\right]^{-1} \mathbb{H}_T' W_{g,T} \left[T^{-1/2} \sum_{t=1}^{T} \hat{g}_t(\theta_0)\right].$$

Note that from Law of Large Number under Assumption 3.4 in LL,

$$\mathbb{H}_T = T^{-1} \sum_{t=1}^{T} (2(z_t - \bar{z}) \otimes I_p) G_2' Y_{t+1} Y_{t+1}' G_2 = E\left[(2(z_t - \bar{z}) \otimes I_p) G_2' Y_{t+1} Y_{t+1}' G_2\right] + o_p(1) = \mathbb{H} + o_p(1).$$

Moreover, from (1.1),

$$T^{-1/2} \sum_{t=1}^{T} \hat{g}_t(\theta_0) \Longrightarrow N(0, \Omega_g), \text{ where } \Omega_g := \lim_{T \to \infty} Var\left[T^{-\frac{1}{2}} \sum_{t=1}^{T} \hat{g}_t(\theta_0)\right].$$

The following theorem is proved using the Slutsky Theorem.

Theorem 1.1 *Under Assumptions 3.1-6 in LL, the Jacobian-GMM estimator* $\widehat{\theta}_{g,T}$ *satisfies*

$$T^{1/2}\left(\widehat{\theta}_{g,T} - \theta_0\right) \to_d N(0, \Sigma_{\theta,g}).$$

where $\Sigma_{\theta,g}\equiv \left(\mathbb{H}'W_g\mathbb{H}\right)^{-1}\mathbb{H}'W_g\Omega_gW_g\mathbb{H}\left(\mathbb{H}'W_g\mathbb{H}\right)^{-1}$. *With the choice of* $W_{g,T}=\hat{\Omega}_{g,T}^{-1}=\Omega_g^{-1}+o_p(1)$, *we have*

$$\Sigma_{\theta,g}=\left(\mathbb{H}'\Omega_g^{-1}\mathbb{H}\right)^{-1}.$$

To derive the asymptotic distributions of the full GMM estimators, note that $\widehat{\theta}_{m,T}$ satisfies the first order condition:

$$\left[T^{-1}\sum_{t=1}^{T}\frac{\partial\hat{m}_t\left(\widehat{\theta}_{m,T}\right)}{\partial\theta'}\right]'W_{m,T}\left[T^{-1}\sum_{t=1}^{T}\hat{m}_t\left(\widehat{\theta}_{m,T}\right)\right]=0. \tag{1.10}$$

We continue with the following notations

$$\begin{aligned} M(\theta) &= E\left[m_t(\theta)\right] \text{ and } M_\theta(\theta)=E\left[\frac{\partial m_t(\theta)}{\partial\theta'}\right], \\ M(\theta_0) &= \mathbf{0} \text{ and } M_\theta(\theta_0)=M_\theta=\begin{bmatrix}\mathbf{0}\\ \mathbb{H}\end{bmatrix}. \end{aligned}$$

Theorem 1.2 *Under Assumptions 3.1-6 in LL, the full GMM estimator* $\widehat{\theta}_{m,T}$ *satisfies*

$$T^{\frac{1}{2}}(\widehat{\theta}_{m,T}-\theta_0)\rightarrow_d N\left(0,\Sigma_{\theta,m}\right).$$

where $\Sigma_{\theta,m}\equiv\left(\mathbb{H}'W_{m,22}\mathbb{H}\right)^{-1}M_\theta'W_m\Omega_mW_mM_\theta\left(\mathbb{H}'W_{m,22}\mathbb{H}\right)^{-1}$, *and* $W_{m,22}$ *denotes the last* $Hp\times Hp$ *submatrix of* W_m. *With the choice of* $W_{m,T}=\hat{\Omega}_{m,T}^{-1}=\Omega_m^{-1}+o_p(1)$, *we have*

$$\Sigma_{\theta,m}=\left(\mathbb{H}'\left(\Omega_m^{-1}\right)_{22}\mathbb{H}\right)^{-1}.$$

Proof. By the consistency of $\widehat{\theta}_{m,T}$ and the mean value theorem,

$$T^{-1}\sum_{t=1}^{T}\hat{m}_t\left(\widehat{\theta}_{m,T}\right)=T^{-1}\sum_{t=1}^{T}\hat{m}_t(\theta_0)+T^{-1}\sum_{t=1}^{T}\frac{\partial\hat{m}_t\left(\hat{\theta}_{1m,T}\right)}{\partial\theta'}(\widehat{\theta}_{m,T}-\theta_0)$$

where $\hat{\theta}_{1m,T}$ denotes values between $\widehat{\theta}_{m,T}$ and θ_0. Combined with the first order condition of (1.10),

$$\begin{aligned} 0 &= \left[T^{-1}\sum_{t=1}^{T}\frac{\partial\hat{m}_t\left(\widehat{\theta}_{m,T}\right)}{\partial\theta'}\right]'W_{m,T}\left[T^{-1}\sum_{t=1}^{T}\hat{m}_t(\theta_0)\right] \\ &\quad+\left[T^{-1}\sum_{t=1}^{T}\frac{\partial\hat{m}_t\left(\widehat{\theta}_{m,T}\right)}{\partial\theta'}\right]'W_{m,T}\left[T^{-1}\sum_{t=1}^{T}\frac{\partial\hat{m}_t\left(\hat{\theta}_{1m,T}\right)}{\partial\theta'}\right](\widehat{\theta}_{m,T}-\theta_0). \end{aligned} \tag{1.11}$$

Using the consistency of $\widehat{\theta}_{m,T}$ and LLN, we get

$$\begin{aligned} & \left[T^{-1}\sum_{t=1}^{T}\frac{\partial\hat{m}_t\left(\widehat{\theta}_{m,T}\right)}{\partial\theta'}\right]' W_{m,T}\left[T^{-1}\sum_{t=1}^{T}\hat{m}_t\left(\theta_0\right)\right] \\ = & \; M_\theta' W_m\left[T^{-1}\sum_{t=1}^{T}\hat{m}_t\left(\theta_0\right)\right] + o_p(T^{-\frac{1}{2}}) \end{aligned} \tag{1.12}$$

w.p.a.1.

Similarly, we have

$$\begin{aligned} & \left[T^{-1}\sum_{t=1}^{T}\frac{\partial\hat{m}_t\left(\hat{\theta}_T\right)}{\partial\theta'}\right]' W_{m,T}\left[T^{-1}\sum_{t=1}^{T}\frac{\partial\hat{m}_t\left(\hat{\theta}_{1,T}\right)}{\partial\theta'}\right] \\ = & \; M_\theta' W_m M_\theta + o_p(1) = \mathbb{H}' W_{m,22}\mathbb{H} + o_p(1) \end{aligned} \tag{1.13}$$

where the last equality uses

$$M_\theta = \frac{\partial M(\theta_0)}{\partial\theta'} = \begin{bmatrix} g\left(\theta_0\right) \\ \mathbb{H}\left(\theta_0\right)\end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbb{H}\end{bmatrix},$$

and $W_{m,22}$ denotes the last $Hp \times Hp$ submatrix of W_m. As $\mathbb{H}$ has full column rank and $W_{m,22}$ is positive definite, we see that $\mathbb{H}' W_{m,22}\mathbb{H}$ is invertible. Hence w.p.a.1, we can combine the results in (1.11), (1.12) and (1.13) to get

$$T^{\frac{1}{2}}(\widehat{\theta}_{m,T} - \theta_0) = \left[\mathbb{H}' W_{m,22}\mathbb{H} + o_p(1)\right]^{-1} M_\theta' W_m\left[T^{-\frac{1}{2}}\sum_{t=1}^{T}\hat{m}_t\left(\theta_0\right)\right] + o_p(1)$$

which together with (1.1) and the Slutsky Theorem implies the claimed result. ■

From Theorem 1.2, it is clear that the choice of the optimal weight matrix $W_{m,T}$ is such that $W_{m,T} = \hat{\Omega}_m^{-1} = \Omega_m^{-1} + o_p(1)$, hence

$$T^{\frac{1}{2}}\left(\widehat{\theta}_{m,T} - \theta_0\right) \to_d N\left(0, \left(\mathbb{H}'\left(\Omega_m^{-1}\right)_{22}\mathbb{H}\right)^{-1}\right)$$

where $\left(\Omega_m^{-1}\right)_{22}$ is the last $Hp \times Hp$ submatrix of Ω_m^{-1}.

To compare the relative efficiency gain of full GMM estimator $\widehat{\theta}_{m,T}$ to $\widehat{\theta}_{g,T}$, first note that

$$\Sigma_{\theta,g} = \left(\mathbb{H}'\Omega_g^{-1}\mathbb{H}\right)^{-1}$$

and

$$\left(\Omega_m^{-1}\right)_{22} = (\Omega_g - \Omega_{g\psi}\Omega_\psi^{-1}\Omega_{\psi g})^{-1},$$

so $\left(\Omega_m^{-1}\right)_{22} \geq \Omega_g^{-1}$ hence $\mathbb{H}'\left(\Omega_m^{-1}\right)_{22}\mathbb{H} \geq \mathbb{H}'\Omega_g^{-1}\mathbb{H}$, or equivalently,

$$\left(\mathbb{H}'\left(\Omega_m^{-1}\right)_{22}\mathbb{H}\right)^{-1} \leq \left(\mathbb{H}'\Omega_g^{-1}\mathbb{H}\right)^{-1}.$$

Using more valid moment conditions improves efficiency in spite of its first order degeneracy. From the above discussion, we see that the full GMM estimator $\widehat{\theta}_{m,T}$ is more efficient than $\widehat{\theta}_{g,T}$. In many applications, however, $\widehat{\theta}_{g,T}$ will be computationally attractive.

This observation motivates the introduction of Modified GMM $\hat{\theta}_{g^*,T}$ in Defintion 1.3, whose asymptotic distribution is given below. It combines the advantages of the both estimators $\widehat{\theta}_{m,T}$ and $\widehat{\theta}_{g,T}$ as explained in LL. The solution using the closed form expression is analogous to that of $\widehat{\theta}_{g,T}$, so it is omitted. We provide a proof that can be used in the case of modified moment conditions.

Theorem 1.3 *Under Assumptions 3.1-6 in LL, the modified GMM estimator $\hat{\theta}_{g^*,T}$ satisfies*

$$T^{1/2}\left(\hat{\theta}_{g^*,T}-\theta_0\right) \rightarrow_d N\left(0, \left(\mathbb{H}'\left(\Omega_m^{-1}\right)_{22}\mathbb{H}\right)^{-1}\right).$$

with the choice of $W_T = \widehat{\left(\Omega_m^{-1}\right)}_{22}$.

Proof. From the first order condition

$$\left[T^{-1}\sum_{t=1}^{T}\frac{\partial \hat{g}_t\left(\hat{\theta}_{g^*,T}\right)}{\partial\theta'}\right]' W_{g^*,T}\left[T^{-1}\sum_{t=1}^{T}\hat{g}_t\left(\hat{\theta}_{g^*,T}\right)-\widehat{\Omega}_{g\psi,T}\widehat{\Omega}_{\psi,T}^{-1}\left(T^{-1}\sum_{t=1}^{T}\hat{\psi}_t\left(\widehat{\theta}_{g,T}\right)\right)\right]=0$$

and the mean value theorem gives

$$\begin{aligned}
&\left[T^{-1}\sum_{t=1}^{T}\frac{\partial \hat{g}_t\left(\hat{\theta}_{g^*,T}\right)}{\partial\theta'}\right]' W_{g^*,T}\left[T^{-1}\sum_{t=1}^{T}\frac{\partial \hat{g}_t\left(\hat{\theta}_{1g^*,T}\right)}{\partial\theta'}\right]\left(\hat{\theta}_{g^*,T}-\theta_0\right)\\
= & -\left[T^{-1}\sum_{t=1}^{T}\frac{\partial \hat{g}_t\left(\hat{\theta}_{g^*,T}\right)}{\partial\theta'}\right]' W_{g^*,T}\left[T^{-1}\sum_{t=1}^{T}\hat{g}_t\left(\theta_0\right)-\widehat{\Omega}_{g\psi,T}\widehat{\Omega}_{\psi,T}^{-1}\left(T^{-1}\sum_{t=1}^{T}\hat{\psi}_t\left(\widehat{\theta}_{g,T}\right)\right)\right]
\end{aligned}$$

where $\hat{\theta}_{1g^*,T}$ denotes values between $\hat{\theta}_{g^*,T}$ and θ_0. From the proof of Theorem 1.1 and 1.2,

$$\begin{aligned}
&T^{1/2}\left(\hat{\theta}_{g^*,T}-\theta_0\right)\\
= & -\left(\mathbb{H}'\left(\Omega_m^{-1}\right)_{22}\mathbb{H}\right)^{-1}\mathbb{H}'\left(\Omega_m^{-1}\right)_{22}\left[T^{-1/2}\sum_{t=1}^{T}\left[\hat{g}_t\left(\theta_0\right)-\widehat{\Omega}_{g\psi,T}\widehat{\Omega}_{\psi,T}^{-1}\hat{\psi}_t\left(\widehat{\theta}_{g,T}\right)\right]\right]+o_p(1)\\
= & -\left(\mathbb{H}'\left(\Omega_m^{-1}\right)_{22}\mathbb{H}\right)^{-1}\mathbb{H}'\left(\Omega_m^{-1}\right)_{22}\left[T^{-1/2}\sum_{t=1}^{T}\left[\hat{g}_t\left(\theta_0\right)-\widehat{\Omega}_{g\psi,T}\widehat{\Omega}_{\psi,T}^{-1}\hat{\psi}_t\left(\theta_0\right)\right]+o_p(1)\right]+o_p(1)
\end{aligned}$$

where the last equality holds because (1.1), $\sqrt{T}$-consistency of $\widehat{\theta}_{g,T}$ and the mean value theorem provide the following result

$$\begin{aligned} & T^{-1/2}\sum_{t=1}^{T}\left[\hat{\psi}_t(\theta_0)-\hat{\psi}_t\left(\widehat{\theta}_{g,T}\right)\right] \\ = \;& T^{-1}\sum_{t=1}^{T}\left[\frac{\partial\hat{\psi}_t\left(\widehat{\theta}_{1g,T}\right)}{\partial\theta'}\right]T^{1/2}\left(\widehat{\theta}_{g,T}-\theta_0\right)=o_p(1). \end{aligned}$$

It is clear that

$$T^{-1/2}\sum_{t=1}^{T}\left[\hat{g}_t(\theta_0)-\widehat{\Omega}_{g\psi,T}\widehat{\Omega}_{\psi,T}^{-1}\hat{\psi}_t(\theta_0)\right]\to_d N\left(0,\left\{\left(\Omega_m^{-1}\right)_{22}\right\}^{-1}\right),$$

therefore the claimed result follows. ■

1.2 Overidentification Tests and Limit Theory

Three over-identification tests are considered in this paper. The first test is based on the J-test

$$J_{m,T}=\left[T^{-1/2}\sum_{t=1}^{T}m_t(\widehat{\theta}_{m,T})\right]'\hat{\Omega}_m^{-1}\left[T^{-1/2}\sum_{t=1}^{T}m_t(\widehat{\theta}_{m,T})\right], \tag{1.14}$$

which tests the validity of the stacked moment conditions in as in (2.7) of LL. The second test is based on the J-test

$$J_{g,T}=\left[T^{-1/2}\sum_{t=1}^{T}g_t(\widehat{\theta}_{g,T})\right]'\hat{\Omega}_g^{-1}\left[T^{-1/2}\sum_{t=1}^{T}g_t(\widehat{\theta}_{g,T})\right], \tag{1.15}$$

which tests the validity of the Jacobian moment conditions as in (2.8) of LL. The test based on

$$J_{h,T}=\left[T^{-1/2}\sum_{t=1}^{T}m_t(\widehat{\theta}_{g,T})\right]'\hat{\Omega}_g^{-1}\left[T^{-1/2}\sum_{t=1}^{T}m_t(\widehat{\theta}_{g,T})\right], \tag{1.16}$$

can be considered as given in (2.9) of LL. Notation "hat" signifies consistent estimator of the corresponding long-run variances.

We confirm the asymptotic distributions of the above J tests. Recall the number of restrictions in (1.14) is $H(p+1)$, while the estimated parameter is p-dimensional. Therefore, the following limit theory holds.

Theorem 1.4 *Under Assumptions 3.1-6 in LL,*

$$J_{m,T}\to_d\chi^2\left(H(p+1)-p\right).$$

Proof. From consistency of $\hat{\Omega}_m^{-1}$,

$$\begin{aligned} J_{m,T} &= \left[T^{-1/2}\sum_{t=1}^{T}\hat{m}_t\left(\widehat{\theta}_{m,T}\right)\right]'\hat{\Omega}_m^{-1}\left[T^{-1/2}\sum_{t=1}^{T}\hat{m}_t\left(\widehat{\theta}_{m,T}\right)\right] \\ &= \left[T^{-1/2}\sum_{t=1}^{T}\hat{m}_t\left(\widehat{\theta}_{m,T}\right)\right]'\Omega_m^{-1}\left[T^{-1/2}\sum_{t=1}^{T}\hat{m}_t\left(\widehat{\theta}_{m,T}\right)\right] + o_p(1) \end{aligned}$$

Using the earlier result of Theorem 1.2,

$$T^{-\frac{1}{2}}\sum_{t=1}^{T}\hat{m}_t\left(\widehat{\theta}_{m,T}\right) = T^{-\frac{1}{2}}\sum_{t=1}^{T}\hat{m}_t\left(\theta_0\right) + M_\theta\left[T^{\frac{1}{2}}(\widehat{\theta}_{m,T}-\theta_0)\right] + O_p(T^{-\frac{1}{2}}).$$
$$= T^{-\frac{1}{2}}\sum_{t=1}^{T}\hat{m}_t\left(\theta_0\right) - M_\theta\left[M_\theta'\Omega_m^{-1}M_\theta\right]^{-1}M_\theta\Omega_m^{-1}\left[T^{-\frac{1}{2}}\sum_{t=1}^{T}\hat{m}_t\left(\theta_0\right)\right] + O_p(T^{-\frac{1}{2}})$$

thus

$$\begin{aligned} &\Omega_m^{-1/2}T^{-\frac{1}{2}}\sum_{t=1}^{T}\hat{m}_t\left(\widehat{\theta}_{m,T}\right) \\ &= \Omega_m^{-1/2}T^{-\frac{1}{2}}\sum_{t=1}^{T}\hat{m}_t\left(\theta_0\right) - \Omega_m^{-1/2}M_\theta\left[M_\theta'\Omega_m^{-1}M_\theta\right]^{-1}M_\theta\Omega_m^{-1}\left[T^{-\frac{1}{2}}\sum_{t=1}^{T}\hat{m}_t\left(\theta_0\right)\right] + O_p(T^{-\frac{1}{2}}) \\ &= \left(I - \Omega_m^{-1/2}M_\theta\left[M_\theta'\Omega_m^{-1}M_\theta\right]^{-1}M_\theta\Omega_m^{-1/2}\right)\Omega_m^{-1/2}T^{-\frac{1}{2}}\sum_{t=1}^{T}\hat{m}_t\left(\theta_0\right) + O_p(T^{-\frac{1}{2}}). \end{aligned}$$

Define $B := \Omega_m^{-1/2}T^{-\frac{1}{2}}\sum_{t=1}^{T}\hat{m}_t\left(\theta_0\right)$ then we know $B \Longrightarrow N\left(0, I_{H(p+1)}\right)$. Thus

$$\begin{aligned} J_{m,T} &= B'\left(I_{H(p+1)} - \Omega_m^{-1/2}M_\theta\left[M_\theta'\Omega_m^{-1}M_\theta\right]^{-1}M_\theta\Omega_m^{-1/2}\right)' \\ &\quad\times\left(I_{H(p+1)} - \Omega_m^{-1/2}M_\theta\left[M_\theta'\Omega_m^{-1}M_\theta\right]^{-1}M_\theta\Omega_m^{-1/2}\right)B + o_p(1) \\ &= B'\left(I_{H(p+1)} - \Omega_m^{-1/2}M_\theta\left[M_\theta'\Omega_m^{-1}M_\theta\right]^{-1}M_\theta\Omega_m^{-1/2}\right)B. \end{aligned}$$

Note that

$$\begin{aligned} &rank\left(\Omega_m^{-1/2}M_\theta\left[M_\theta'\Omega_m^{-1}M_\theta\right]^{-1}M_\theta'\Omega_m^{-1/2}\right) \\ &= rank\left(\left[M_\theta'\Omega_m^{-1}M_\theta\right]^{-1}M_\theta'\Omega^{-1}M_\theta\right) \\ &= rank\left(I_p\right) = p \end{aligned}$$

hence

$$J_{m,T} \rightarrow_d \chi^2\left(H(p+1) - p\right).$$

■

Except for the different degrees of freedom $(Hp - p)$, the proof for (1.15) is analogous, so it is omitted.

Theorem 1.5 *Under Assumptions 3.1-6 in LL,*

$$J_{g,T} \rightarrow_d \chi^2 (Hp - p) .$$

Asymptotic distribution of $J_{h,T}$ and its proof is given in LL.

2 References

Dovonon, P., & Renault, E. (2013). Testing for Common Conditionally Heteroskedastic Factors. Econometrica, 81(6), 2561-2586.

Lee, J.H., & Liao, Z. (2014). On Standard Inference for GMM with Seeming Local Identification Failure. Unpublished Manuscript.