Online Supplementary Material to ‘On GMM Inference: Partial Identification, Identification Strength and Non-Standard Asymptotics’

D. S. Poskitt
Department of Econometrics and Business Statistics
Monash University

July 3, 2023

1 Introduction

The material contained in this supplement follows the labelling and numbering of the sections in the main paper.

2 Assumptions and Background

The following conditions provide a set of primitive regularity requirements under which Assumption 2.2 and Assumption 3.1 will hold.

Regularity Conditions *The process* $\mathbf{w}_t$ *and moments* $\boldsymbol{\mu}_t(\mathbf{w}_t, \boldsymbol{\theta})$ *satisfy the following conditions;*

(i) $\boldsymbol{\mu}_t(\mathbf{w}_t, \boldsymbol{\theta})$ *is near-epoch dependent of size* -1 *on an* α*-mixing basis process of size* $-r/(r-2)$, $r > 2$.

(ii) there exists a random sequence B_t *where* $\lim_{n\to\infty} \sum_{t=l}^{n} E[B_t^{2+\delta}] < \infty$ *for some* $\delta > 0$*, and a non-negative deterministic function* $h(\cdot)$ *such that* $h(x) \to 0$ *as* $x \to 0$*, such that*

$$\|\boldsymbol{\mu}_t(\mathbf{w}_t, \boldsymbol{\theta}_1) - \boldsymbol{\mu}_t(\mathbf{w}_t, \boldsymbol{\theta}_2)\| < B_t h(\|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\|)\,,$$

and

(iii) $\sup_{\theta\in\Theta} E[\|\boldsymbol{\mu}_t(\mathbf{w}_t, \boldsymbol{\theta})\|^{2+\delta}] < \infty$ *for some* $\delta \geq r - 2$*, and the sample moment long-run covariance matrix* $\bar{\boldsymbol{\Sigma}}_n(\boldsymbol{\theta}) \to \boldsymbol{\Sigma}(\boldsymbol{\theta})$.

Condition (i) implies that $\boldsymbol{\mu}_t(\mathbf{w}_t, \boldsymbol{\theta})$ is a mixingale for which a weak law of large numbers holds, and condition (i) and the Lipschitz condition in (ii) imply stochastic equicontinuity. Conditions (i), (ii) and (iii) specify properties from which the required convergence of the finite dimensional distributions, stochastic equicontinuity, and total boundedness, needed to validate the weak convergence of Assumption 3.1 are ensured. (See Davidson, 1994, Theorems 17.5, 21.10 and 29.8, for example.)

See Stock and Wright (2000, Section 2.2) for further discussion of similar conditions in the context of GMM. An in depth discussion of the technicalities underlying such conditions is beyond the scope of this paper, we refer to Davidson (1994) for a text book exposition and the monograph by Potscher and Prucha (1997) for detailed particulars and extensive references. A detailed exposition of functional central limit theorems (FCLTs) of the type presented in Assumption 3.1 and Assumption 3.2 can be found in Van der Vaart and Wellner (1996). See also Kleibergen (2005, Assumption 1) and Caner (2010, Assumption 5) for statements and discussion of regularity conditions that parallel Assumption 3.1 and Assumption 3.2.

Proof. Lemma 2.1 Let $\mathbf{b}_1, \ldots, \mathbf{b}_k$ denote a set of orthonormal basis vectors in $\mathbb{R}^k$. By the multivariate mean value theorem (Apostol, 1974, Theorem 12.9) for each $\mathbf{b}_i$ there exists a $\boldsymbol{\theta}_i^*$ on the line segment joining $\boldsymbol{\theta}$ to $\boldsymbol{\theta}_0$ such that

$$\mathbf{b}_i'(\bar{\boldsymbol{\mu}}_n(\boldsymbol{\theta}) - \bar{\boldsymbol{\mu}}_n(\boldsymbol{\theta}_0)) = \mathbf{b}_i'\bar{\boldsymbol{\Delta}}_n(\boldsymbol{\theta}_i^*)(\boldsymbol{\theta} - \boldsymbol{\theta}_0)\,.$$

Let $\boldsymbol{\theta}^1, \boldsymbol{\theta}^2, \ldots, \boldsymbol{\theta}^s, \ldots$ be an infinite sequence in $N(\boldsymbol{\theta}_0;\delta)$, $\boldsymbol{\theta}^s \neq \boldsymbol{\theta}_0$, that converges to $\boldsymbol{\theta}_0$ such that $\bar{\boldsymbol{\mu}}_n(\boldsymbol{\theta}^s) = \bar{\boldsymbol{\mu}}_n(\boldsymbol{\theta}_0)$ and set

$$\mathbf{d}^s = \frac{\boldsymbol{\theta}^s - \boldsymbol{\theta}_0}{\|\boldsymbol{\theta}^s - \boldsymbol{\theta}_0\|}\,.$$

The sequence $\mathbf{d}^1, \mathbf{d}^2, \ldots, \mathbf{d}^s, \ldots$ is an infinite sequence on the unit sphere in $\mathbb{R}^p$ and therefore there exists an accumulation point $\mathbf{d}$ and an infinite sequence $\mathbf{d}^s$ such that $\lim_{s\to\infty}\mathbf{d}^s = \mathbf{d}$, which we denote (using a slight abuse of notation as strictly speaking $\mathbf{d}^s$ may be subsequence) by $\mathbf{d}^s \to \mathbf{d}$ as $\boldsymbol{\theta}^s \to \boldsymbol{\theta}_0$. By continuity $\bar{\boldsymbol{\Delta}}_n(\boldsymbol{\theta}_i^{*s}) \to \bar{\boldsymbol{\Delta}}_n(\boldsymbol{\theta}_0)$ as $\boldsymbol{\theta}_i^{*s} = \boldsymbol{\theta}_0 + \lambda_i(\boldsymbol{\theta}^s - \boldsymbol{\theta}_0)$, $0 \le \lambda_i \le 1$, converges to $\boldsymbol{\theta}_0$ as $\boldsymbol{\theta}^s \to \boldsymbol{\theta}_0$, from which it follows that $\mathbf{b}_i'\bar{\boldsymbol{\Delta}}_n(\boldsymbol{\theta}_i^{*s})\mathbf{d}^s \to \mathbf{b}_i'\bar{\boldsymbol{\Delta}}_n(\boldsymbol{\theta}_0)\mathbf{d}$. We therefore have that $\mathbf{b}_i'\bar{\boldsymbol{\Delta}}_n(\boldsymbol{\theta}_0)\mathbf{d} = 0$ for all $\mathbf{b}_i$, $i = 1, \ldots, k$. Since $\mathbf{b}_1, \ldots, \mathbf{b}_k$ form an orthonormal basis and $\bar{\boldsymbol{\Delta}}_n(\boldsymbol{\theta}_0)\mathbf{d}$ is perpendicular to each $\mathbf{b}_i$, $i = 1, \ldots, k$ we can conclude that $\bar{\boldsymbol{\Delta}}_n(\boldsymbol{\theta}_0)\mathbf{d} = \mathbf{0}$. It follows from this that $r\{\bar{\boldsymbol{\Delta}}_n(\boldsymbol{\theta}_0)\} = q_n < p$ because $\|\mathbf{d}\| = 1$. This establishes necessity in the first part of the lemma.

To establish sufficiency suppose that $r\{\bar{\boldsymbol{\Delta}}_n(\boldsymbol{\theta}_0)\} = q_n < p$. For any given point $\boldsymbol{\theta} \in N(\boldsymbol{\theta}_0;\delta)$, $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$, let $\boldsymbol{\theta}_\lambda^* = \boldsymbol{\theta}_0 + \lambda(\boldsymbol{\theta} - \boldsymbol{\theta}_0)$, $0 \le \lambda \le 1$, and observe that since $\bar{\boldsymbol{\Delta}}_n(\boldsymbol{\theta})$ is regular $r\{\bar{\boldsymbol{\Delta}}_n(\boldsymbol{\theta}_\lambda^*)\} = q_n < p$. Let $\mathbf{c}_i(\boldsymbol{\theta}_\lambda^*)$, $i = 1, \ldots, (p - q_n)$, denote the $p - q_n$ linearly independent nontrivial solutions to the homogeneous equation system $\bar{\boldsymbol{\Delta}}_n(\boldsymbol{\theta}_\lambda^*)\mathbf{x} = \mathbf{0}$ obtained by expressing the $p - q_n$ linearly dependent columns of $\bar{\Delta}_n(\boldsymbol{\theta}_\lambda^*)$ as linear combinations of the q_n linearly independent columns. Without loss of generality each $\mathbf{c}_i(\boldsymbol{\theta}_\lambda^*)$ can be expressed as $\mathbf{c}_i(\boldsymbol{\theta}_\lambda^*) = (c_{i1}(\boldsymbol{\theta}_\lambda^*), \ldots, c_{iq_n}(\boldsymbol{\theta}_\lambda^*), 0, \ldots, 0)' + \mathbf{e}_{q_n+i}$ where $\mathbf{e}_{q_n+i} = (0, \ldots, 0, 1, 0, \ldots, 0)'$, the $(q_n + i)$th unit coordinate vector in $\mathbb{R}^p$. Each $\mathbf{c}_i(\boldsymbol{\theta}_\lambda^*)$ is a continuous function of λ.[1] Now let

[1] Consider the difference $\mathbf{c}(\boldsymbol{\theta}_{\lambda_2}^*) - \mathbf{c}(\boldsymbol{\theta}_{\lambda_1}^*)$ where $\bar{\boldsymbol{\Delta}}_n(\boldsymbol{\theta}_{\lambda_j}^*)\mathbf{c}(\boldsymbol{\theta}_{\lambda_j}^*) = \mathbf{0}$, $\boldsymbol{\theta}_{\lambda_j}^* = \boldsymbol{\theta}_0 + \lambda_j(\boldsymbol{\theta} - \boldsymbol{\theta}_0)$, $0 \le \lambda_j \le 1$, $j = 1, 2$, with $\lambda_1 \le \lambda \le \lambda_2$, $\lambda_1 \neq \lambda_2$. By assumption $\bar{\boldsymbol{\Delta}}_n(\boldsymbol{\theta})$ is a continuous function of $\boldsymbol{\theta}$ with constant rank q_n and setting $\mathbf{c}(\boldsymbol{\theta}_\lambda^*)$ equal to one of $\mathbf{c}_i(\boldsymbol{\theta}_\lambda^*)$, $i = 1, \ldots, (p - q_n)$, we have

$$\bar{\boldsymbol{\Delta}}_n(\boldsymbol{\theta}_{\lambda_2}^*)[\mathbf{c}(\boldsymbol{\theta}_{\lambda_2}^*) - \mathbf{c}(\boldsymbol{\theta}_{\lambda_1}^*)] = [\bar{\boldsymbol{\Delta}}_n(\boldsymbol{\theta}_{\lambda_1}^*) - \bar{\boldsymbol{\Delta}}_n(\boldsymbol{\theta}_{\lambda_2}^*)]\mathbf{c}(\boldsymbol{\theta}_{\lambda_1}^*)\,,$$

from which we can conclude that

$$\|\mathbf{c}(\boldsymbol{\theta}_{\lambda_2}^*) - \mathbf{c}(\boldsymbol{\theta}_{\lambda_1}^*)\| \le \|\bar{\boldsymbol{\Delta}}_n(\boldsymbol{\theta}_{\lambda_2}^*)^+\| \cdot \|\bar{\boldsymbol{\Delta}}_n(\boldsymbol{\theta}_{\lambda_2}^*) - \bar{\boldsymbol{\Delta}}_n(\boldsymbol{\theta}_{\lambda_1}^*)\| \cdot \|\mathbf{c}(\boldsymbol{\theta}_{\lambda_1}^*)\|\,.$$

The continuity of $\mathbf{c}(\boldsymbol{\theta}_\lambda^*)$ as a function of λ follows since $\|\boldsymbol{\theta}_{\lambda_2}^* - \boldsymbol{\theta}_{\lambda_1}^*\| = |\lambda_2 - \lambda_1| \cdot \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| \le |\lambda_2 - \lambda_1|\delta$ and consequently we have that $\|\bar{\boldsymbol{\Delta}}_n(\boldsymbol{\theta}_{\lambda_2}^*) - \bar{\boldsymbol{\Delta}}_n(\boldsymbol{\theta}_{\lambda_1}^*)\| \to 0$ and therefore $\|\mathbf{c}(\boldsymbol{\theta}_{\lambda_2}^*) - \mathbf{c}(\boldsymbol{\theta}_{\lambda_1}^*)\| \to 0$ as $|\lambda_2 - \lambda_1| = (\lambda_2 - \lambda_1) \to 0$.

$\boldsymbol{\theta}_i^*(\lambda)$ denote the curve defined by the function that for $0 \leq \lambda \leq 1$ solves the differential equation

$$\frac{\partial \boldsymbol{\theta}_\lambda^*}{\partial \lambda} = \mathbf{c}_i(\boldsymbol{\theta}_\lambda^*)\,, \quad \boldsymbol{\theta}_0^* = \boldsymbol{\theta}_0\,, \quad i = 1, \ldots, (p - q_n)\,.$$

By the chain rule

$$\frac{\partial \bar{\boldsymbol{\mu}}_n(\boldsymbol{\theta}_i^*(\lambda))}{\partial \lambda} = \bar{\boldsymbol{\Delta}}_n(\boldsymbol{\theta}_i^*(\lambda))\mathbf{c}_i(\boldsymbol{\theta}_i^*(\lambda)) = \mathbf{0}$$

for all $0 \leq \lambda \leq 1$. This implies that $\bar{\boldsymbol{\mu}}_n(\boldsymbol{\theta})$ is constant along each curve $\boldsymbol{\theta}_i^*(\lambda)$ and hence that $\bar{\boldsymbol{\mu}}_n(\boldsymbol{\theta}_i^*(\lambda)) = \bar{\boldsymbol{\mu}}_n(\boldsymbol{\theta}_0^*) = \bar{\boldsymbol{\mu}}_n(\boldsymbol{\theta}_0)$ for all $0 \leq \lambda \leq 1$.

To confirm that when $\bar{\boldsymbol{\Delta}}_n(\boldsymbol{\theta}_0)$ is rank deficient $\{\boldsymbol{\theta} : \bar{\boldsymbol{\mu}}_n(\boldsymbol{\theta}) = \bar{\boldsymbol{\mu}}_n(\boldsymbol{\theta}_0)\}$ is a continuum let $\boldsymbol{\theta}_{rs}^* = \boldsymbol{\theta}_r^*(\lambda) + t(\boldsymbol{\theta}_r^*(\lambda) - \boldsymbol{\theta}_s^*(\lambda))$, $0 \leq t \leq 1$, where $\boldsymbol{\theta}_r^*(\lambda)$ and $\boldsymbol{\theta}_s^*(\lambda)$ are any one of the $\binom{p-q_n}{2}$ possible pairs of curves taken from $\boldsymbol{\theta}_i^*(\lambda)$, $i = 1, \ldots, (p - q_n)$, $0 \leq \lambda \leq 1$. Repeating the above argument we can construct a curve $\boldsymbol{\theta}_{rs}^*(t, \lambda)$ such that

$$\frac{\partial \boldsymbol{\theta}_{rs}^*(t,\lambda)}{\partial t} = \mathbf{c}(\boldsymbol{\theta}_{rs}^*(t,\lambda))\,, \quad \boldsymbol{\theta}_{rs}^*(0,\lambda) = \boldsymbol{\theta}_r^*(\lambda)\,,$$

and

$$\frac{\partial \bar{\boldsymbol{\mu}}_n(\boldsymbol{\theta}_{rs}^*(t,\lambda))}{\partial t} = \bar{\boldsymbol{\Delta}}_n(\boldsymbol{\theta}_{rs}^*(t,\lambda))\mathbf{c}_i(\boldsymbol{\theta}_{rs}^*(t,\lambda)) = \mathbf{0}$$

for all $0 \leq t \leq 1$. The curve $\boldsymbol{\theta}_{rs}^*(t, \lambda)$ determines a path that joins $\boldsymbol{\theta}_r^*(\lambda)$ to $\boldsymbol{\theta}_s^*(\lambda)$ along which $\bar{\boldsymbol{\mu}}_n(\boldsymbol{\theta}_{rs}^*(t,\lambda)) = \bar{\boldsymbol{\mu}}_n(\boldsymbol{\theta}_0)$ for all $0 \leq t \leq 1$. Since $\boldsymbol{\Theta}$ is compact we can apply the above construction to a countable collection of points $\boldsymbol{\theta} \in N(\boldsymbol{\theta}_0; \delta)$, $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$, and generate an ϵ-net for $\{\boldsymbol{\theta} : \bar{\boldsymbol{\mu}}_n(\boldsymbol{\theta}) = \bar{\boldsymbol{\mu}}_n(\boldsymbol{\theta}_0)\}$ and thereby complete the proof. ■

3 Identification and Estimation

3.1 Criterion and Estimator Convergence

The following proof is modelled on a standard consistency proof for extremum estimators.

Proof. Theorem 3.2. For any set $A \subseteq \Theta$ and $\epsilon > 0$, denote the ϵ neighborhood of A by $N(A;\epsilon) = \{\boldsymbol{\theta} \in \Theta : d(\boldsymbol{\theta}; A) < \varepsilon\}$ and set $\overline{N}(A, \epsilon) \equiv \{\boldsymbol{\theta} : d(\boldsymbol{\theta}; A) \leq \epsilon\}$.

By definition $Q_n(\widehat{\boldsymbol{\theta}}_n) \leq Q_n(\boldsymbol{\theta})$ for any $\widehat{\boldsymbol{\theta}}_n \in \widehat{\Theta}_n$ and all $\boldsymbol{\theta} \in \Theta_{0n}$, and $|Q_n(\boldsymbol{\theta}) - \bar{Q}_n(\boldsymbol{\theta})| \xrightarrow{p} 0$ uniformly in $\boldsymbol{\theta}$ by Lemma 3.1. Thus for any $\delta > 0$

$$\bar{Q}_n(\widehat{\boldsymbol{\theta}}_n) - \delta/2 < Q_n(\widehat{\boldsymbol{\theta}}_n) \leq Q_n(\boldsymbol{\theta}) < \bar{Q}_n(\boldsymbol{\theta}) + \delta/2$$

with probability converging to one as $n \to \infty$. Since by definition of Θ_{0n} we have $\bar{\boldsymbol{\mu}}_n(\boldsymbol{\theta}) = \bar{\boldsymbol{\mu}}_n(\boldsymbol{\theta}_0) = \mathbf{0}$ and $\bar{Q}_n(\boldsymbol{\theta}) = 0$ for all $\boldsymbol{\theta} \in \Theta_{0n}$, it follows that

$$\lim_{n\to\infty} Pr\left\{\bar{Q}_n(\widehat{\boldsymbol{\theta}}_n) < \delta\right\} = 1\,. \tag{3.1}$$

To translate the inequality $\bar{Q}_n(\widehat{\boldsymbol{\theta}}_n) < \delta$ into a bound on the distance of $\widehat{\Theta}_n$ from $\overline{N}(\Theta_{0n}, \epsilon)$ set

$$\delta(\epsilon) = \min_{\theta \in \Theta \setminus N(\Theta_{0n},\epsilon)} \bar{Q}_n(\boldsymbol{\theta})$$

where $\Theta \setminus N(\Theta_{0n}, \epsilon)$ denotes the relative complement of $N(\Theta_{0n}, \epsilon)$ in Θ. By definition of $\delta(\epsilon)$

we can deduce that $\bar{Q}_n(\boldsymbol{\theta}) < \delta(\epsilon)$ implies that $\boldsymbol{\theta} \in N(\Theta_{0n}, \epsilon)$. Since $N(\Theta_{0n}, \epsilon) \subset \overline{N}(\Theta_{0n}, \epsilon)$ it follows that

$$Pr\left\{\widehat{\boldsymbol{\theta}}_n \in \overline{N}(\Theta_{0n}, \epsilon)\right\} \geq Pr\left\{\bar{Q}_n(\widehat{\boldsymbol{\theta}}_n) < \delta(\epsilon)\right\},$$

which in the face of (3.1) leads to the conclusion that

$$\lim_{n\to\infty} Pr\left\{\widehat{\boldsymbol{\theta}}_n \in \overline{N}(\Theta_{0n}, \epsilon)\right\} = 1\,. \tag{3.2}$$

It follows directly from (3.2) that $\sup_{\widehat{\theta}_n \in \widehat{\Theta}_n} d(\widehat{\boldsymbol{\theta}}, \Theta_{0n}) < \epsilon$ with probability converging to one as $n \to \infty$. Moreover, since $\widehat{\boldsymbol{\theta}}_n \in \overline{N}(\Theta_{0n}, \epsilon)$ for all $\widehat{\boldsymbol{\theta}}_n \in \widehat{\Theta}_n$ implies that $\widehat{\Theta}_n \subset \overline{N}(\Theta_{0n}, \epsilon)$, then for each $\boldsymbol{\theta} \in \Theta_{0n}$ there exists a $\widehat{\boldsymbol{\theta}}_n$ such that $\|\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}_n\| < \epsilon$. It follows therefore that $d(\boldsymbol{\theta}, \widehat{\Theta}_n) \leq \epsilon$ for all $\boldsymbol{\theta} \in \Theta_{0n}$ and hence that $\sup_{\theta \in \Theta_{0n}} d(\boldsymbol{\theta}, \widehat{\Theta}_n) < \epsilon$ with probability approaching one as $n \to \infty$. ■

Theorem 3.2 does not provide a convergence rate. If we add the requirement that the moment functions satisfy the Donsker property in Assumption 3.2 then the conditions required for Theorem 3.1 of Chernozhukov, Hong, and Tamer (2007) to apply will be satisfied. (See Chernozhukov, Hong, and Tamer, 2007, Section 4.1, pages 1261-1264.) Thus, the conditions are applicable for the moment equality models under consideration here and in the current setting the convergence rate of $d_H(\widehat{\Theta}_n; \Theta_{0n})$ will be $n^{-\frac{1}{2}}$.

3.2 Estimable Functions and Asymptotic Normality

Lemma 2.2 indicates that lack of identification implies that additional $p - q_n$ restrictions in the guise of

$$\begin{bmatrix} \theta_{i(q_n+1)} \\ \vdots \\ \theta_{i(p)} \end{bmatrix} - \begin{bmatrix} \alpha_1(\theta_{i(1)}, \ldots, \theta_{i(q_n)}) \\ \vdots \\ \alpha_{p-q_n}(\theta_{i(1)}, \ldots, \theta_{i(q_n)}) \end{bmatrix} = \mathbf{0}$$

are required to identify the partially identified parameter $\boldsymbol{\beta} = (\theta_{i(1)}, \ldots, \theta_{i(q_n)})'$. A corollary of Theorem 2.3.1(b)&(c) of Rao and Mitra (1971) is that there are $p - q_n + 1$ linearly independent solutions to (3.6) and q_n linearly independent vectors $\mathbf{q}_{0n}$ satisfying $\mathbf{q}'_{0n} = \mathbf{z}'\bar{\mathbf{H}}_{0n}$ since $r\{\bar{\mathbf{H}}_{0n}\} = r\{\bar{\boldsymbol{\Delta}}'_{0n}\bar{\boldsymbol{\Omega}}_n\bar{\boldsymbol{\Delta}}_{0n}\} = q_n$. The additional restrictions required to achieve identification are being indirectly imposed via the g-inverse in order to constrain inference to the q_n linearly independent estimable functions.

This is analogous to the situation that obtains in unidentified and partially identified structural equation models. Building upon the analysis and requisite central limit theory presented in Phillips (1989), Choi and Phillips (1992) showed that in such models identified and unidentified parts of a structural coefficient vector can be distinguished by rotating the coordinate system, and they provided formulae for the finite sample and asymptotic densities of IV estimators of the coefficient vector and the limit distributions of Wald test statistics. In similar vein, from the singular value decomposition of $\bar{\boldsymbol{\Delta}}_{0n}$ we have (using a slight abuse of the notation employed in (2.5)) $\bar{\boldsymbol{\Delta}}_{0n}\boldsymbol{\theta} = \mathbf{U}_{0n}\boldsymbol{\vartheta}$, wherein the parameter vector $\boldsymbol{\vartheta} = (\vartheta_1, \ldots, \vartheta_{q_n}, 0, \ldots, 0)'$ where

$$\vartheta_r = s_{0n,r} \sum_{c=1}^{p} v_{0n,rc}\theta_c\,, \quad r = 1, \ldots, q_n\,. \tag{3.8}$$

By the mean value theorem $\bar{\boldsymbol{\mu}}_n(\boldsymbol{\theta}) - \bar{\boldsymbol{\mu}}_n(\boldsymbol{\theta}_0) = \bar{\boldsymbol{\Delta}}_{0n}(\boldsymbol{\theta} - \boldsymbol{\theta}_0) = \mathbf{U}_{0n}(\boldsymbol{\vartheta} - \boldsymbol{\vartheta}_0)$ wherein $\boldsymbol{\theta} =$

$(\theta_1, \ldots, \theta_p)'$ with $\theta_i = \theta_{0i} + (\theta_{0n,i} - \theta_{0i})/\lambda_i$, $0 < \lambda_i \leq 1$, $i = 1, \ldots, p$, where $\boldsymbol{\theta}_0 = (\theta_{01}, \ldots, \theta_{0p})'$ is the true parameter value of the DGP and $\boldsymbol{\theta}_{0n} = (\theta_{0n,1}, \ldots, \theta_{0n,p})'$ is given in (3.1). Since the columns of $\mathbf{U}_{0n}$ are linearly independent this implies that $\bar{\boldsymbol{\mu}}_n(\boldsymbol{\theta}) = \bar{\boldsymbol{\mu}}_n(\boldsymbol{\theta}_0)$ if and only if $\boldsymbol{\vartheta} = \boldsymbol{\vartheta}_0$ and $\boldsymbol{\vartheta}_0$ is identified. Thus it follows that $\bar{\mathbf{H}}_{0n}\boldsymbol{\theta}_0 = (\bar{\boldsymbol{\Delta}}'_{0n}\bar{\boldsymbol{\Omega}}_n\bar{\boldsymbol{\Delta}}_{0n})^{+}\bar{\boldsymbol{\Delta}}'_{0n}\bar{\boldsymbol{\Omega}}_n\mathbf{U}_{0n}\boldsymbol{\vartheta}_0$ and we can see that the idempotent matrix $\bar{\mathbf{H}}_{0n}$ produces an estimable function by inducing a coordinate transformation from $\boldsymbol{\theta}_0$ to the identified parameter $\boldsymbol{\vartheta}_0$ in a manner that parallels the Choi and Phillips coordinate rotation. Note in passing that the focus of Choi and Phillips (1992) was on the effect of identification and lack of identification on the distribution theory of the IV estimator, and they did not address issues associated with different levels of identification strength. Here it is apparent from (3.8) that the identification strength of the moment conditions is passed on directly to the components of $\boldsymbol{\vartheta}$, but as is demonstrated in the paper, the relationship of the identification strength of the components of $\boldsymbol{\theta}$ to the identification strength of the moment conditions is not so straightforward.

Turning to the asymptotic distributions of estimable functions of $(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_{0n})$ presented in Theorems 3.3, 3.4 and 3.5, recall that

$$\boldsymbol{\theta}_{0n} = \arg\min_{\theta \in \Theta_{0n}} \|\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}\| . \tag{3.1}$$

where $\Theta_{0n} = \{\boldsymbol{\theta} : \bar{Q}_n(\boldsymbol{\theta}) = 0\}$. To establish that $\boldsymbol{\theta}_{0n}$ is the projection of $\widehat{\boldsymbol{\theta}}_n$ on to Θ_{0n}, let $\langle\Theta_{0n}\rangle$ denote the convex hull of Θ_{0n}. Then for any $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \langle\Theta_{0n}\rangle$ we have by the mean value theorem (Apostol, 1974, Theorem 12.9, Example 1, pp. 355-356)

$$\bar{Q}_n(\boldsymbol{\theta}_1) = \bar{Q}_n(\boldsymbol{\theta}_2) + \frac{\partial \bar{Q}_n(\boldsymbol{\theta}^*_\lambda)}{\partial \boldsymbol{\theta}'}(\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2)$$

where $\boldsymbol{\theta}^*_\lambda = \boldsymbol{\theta}_2 + \lambda(\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2)$, $0 \leq \lambda \leq 1$, from which it is obvious that $\bar{Q}_n(\boldsymbol{\theta}_1) \leq \bar{Q}_n(\boldsymbol{\theta}_2)$ implies that $\partial \bar{Q}_n(\boldsymbol{\theta}^*_\lambda)/\partial \boldsymbol{\theta}' \cdot (\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2) \leq 0$. Hence

$$\begin{aligned} \frac{\partial \bar{Q}_n(\boldsymbol{\theta}_2)}{\partial \boldsymbol{\theta}'}(\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2) =& \frac{\partial \bar{Q}_n(\boldsymbol{\theta}^*_\lambda)}{\partial \boldsymbol{\theta}'}(\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2) + \left\{ \frac{\partial \bar{Q}_n(\boldsymbol{\theta}_2)}{\partial \boldsymbol{\theta}'} - \frac{\partial \bar{Q}_n(\boldsymbol{\theta}^*_\lambda)}{\partial \boldsymbol{\theta}'} \right\} (\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2) \\ \leq& \left\{ \frac{\partial \bar{Q}_n(\boldsymbol{\theta}_2)}{\partial \boldsymbol{\theta}'} - \frac{\partial \bar{Q}_n(\boldsymbol{\theta}^*_\lambda)}{\partial \boldsymbol{\theta}'} \right\} (\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2) . \end{aligned} \tag{3.9}$$

Passing to the limit in (3.9) as $\lambda \to 0^+$ it follows from the continuity of $\partial \bar{Q}_n(\boldsymbol{\theta})/\partial \boldsymbol{\theta}'$ that

$$\frac{\partial \bar{Q}_n(\boldsymbol{\theta}_2)}{\partial \boldsymbol{\theta}'}(\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2) < \epsilon$$

for all $\epsilon > 0$. Thus, $\bar{Q}_n(\boldsymbol{\theta}_1) \leq \bar{Q}_n(\boldsymbol{\theta}_2)$ implies that $\partial \bar{Q}_n(\boldsymbol{\theta}_2)/\partial \boldsymbol{\theta}' \cdot (\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2) \leq 0$ and therefore we can conclude that $\bar{Q}_n(\boldsymbol{\theta})$ is quasi-convex on $\langle\Theta_{0n}\rangle$. The latter is equivalent to convexity of the lower level set $\{\boldsymbol{\theta} : \bar{Q}_n(\boldsymbol{\theta}) \leq q\}$ for each $q \geq 0$. It follows that Θ_{0n} is itself a convex set and hence, via a standard Hilbert space result, $\boldsymbol{\theta}_{0n}$ is the unique projection of $\widehat{\boldsymbol{\theta}}_n$ on to Θ_{0n}.

With regard to Theorem 3.3 and Theorem 3.4, if the moment conditions are twice continuously differentiable, as is assumed in Theorem 3.4, then

$$\begin{aligned}\mathbf{Q}_n^{(2)}(\boldsymbol{\theta}) = \frac{\partial^2 Q_n(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} &= 2\left[\frac{\partial\mathbf{D}_n(\boldsymbol{\theta})'\mathbf{W}_n\mathbf{m}_n(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'}\right] \\ &= 2\left[(\mathbf{m}_n(\boldsymbol{\theta})'\mathbf{W}_n\otimes\mathbf{I})\frac{\partial\mathrm{vec}(\mathbf{D}_n(\boldsymbol{\theta})')}{\partial\boldsymbol{\theta}'} + \mathbf{D}_n(\boldsymbol{\theta})'\mathbf{W}_n\mathbf{D}_n(\boldsymbol{\theta})\right],\end{aligned}$$

and the triangle inequality gives us

$$\begin{aligned}\|\mathbf{Q}_{0n}^{(2)} - 2(\bar{\boldsymbol{\Delta}}_{0n}'\bar{\boldsymbol{\Omega}}_n\bar{\boldsymbol{\Delta}}_{0n})\| \leq 2\|\mathbf{D}_n(\boldsymbol{\theta}_{0n})'\mathbf{W}_n\mathbf{D}_n(\boldsymbol{\theta}_{0n}) - (\bar{\boldsymbol{\Delta}}_{0n}'\bar{\boldsymbol{\Omega}}_n\bar{\boldsymbol{\Delta}}_{0n})\| \\ + 2\rho_n(\boldsymbol{\theta}_{0n})\end{aligned}$$

where

$$\rho_n(\boldsymbol{\theta}_{0n}) = \|(\mathbf{m}_n(\boldsymbol{\theta}_{0n})'\mathbf{W}_n\otimes\mathbf{I})\frac{\partial\mathrm{vec}(\mathbf{D}_n(\boldsymbol{\theta}_{0n})')}{\partial\boldsymbol{\theta}'}\|\,.$$

The first term converges to zero, as shown in (3.5), and $\rho_n(\boldsymbol{\theta}_{0n}) = O_p(n^{-\frac{1}{2}})$. The latter follows because for any given $p\times 1$ vector $\mathbf{v}$

$$\frac{\mathbf{v}'(\sqrt{n}\mathbf{m}_n(\boldsymbol{\theta}_{0n})'\mathbf{W}_n\otimes\mathbf{I})\{\partial\mathrm{vec}(\mathbf{D}_n(\boldsymbol{\theta}_{0n})')/\partial\boldsymbol{\theta}'\}\mathbf{v}}{(\mathbf{d}'(\bar{\boldsymbol{\Omega}}_n\boldsymbol{\Sigma}_{0n}\bar{\boldsymbol{\Omega}}_n\otimes\mathbf{v}\mathbf{v}')\mathbf{d})^{\frac{1}{2}}} \Rightarrow \mathcal{N}(0,1)$$

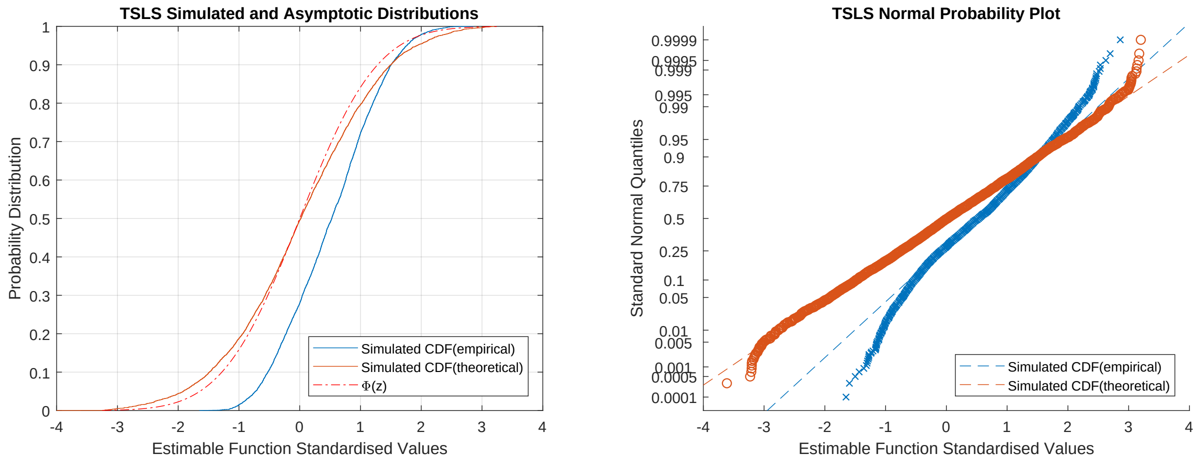
where $\mathbf{d} = \{\partial\mathrm{vec}(\bar{\boldsymbol{\Delta}}_{0n}')/\partial\boldsymbol{\theta}'\}$. It follows that $\|\mathbf{Q}_{0n}^{(2)} - 2(\bar{\boldsymbol{\Delta}}_{0n}'\bar{\boldsymbol{\Omega}}_n\bar{\boldsymbol{\Delta}}_{0n})\| \xrightarrow{p} 0$ and we can therefore conclude that Theorem 3.3 and Theorem 3.4 are asymptotically equivalent.

4 Illustrations I

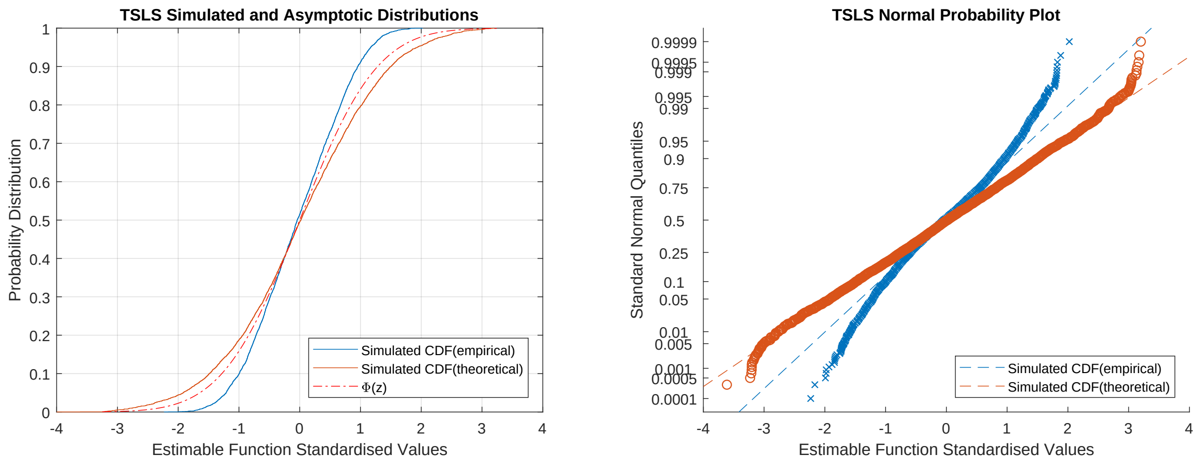
4.1 The Linear Equations Model

Figure 2 indicates that when $\mathbf{q}$ corresponds to an estimable function $\mathbf{q}' = \mathbf{z}'\mathbf{H}$ asymptotic normality of $\sqrt{n}\mathbf{q}'(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_{0n})$ holds, but use of the estimated value $\mathbf{q}' = \mathbf{z}'\widehat{\mathbf{H}}$ fails to adequately adjust for the lack of normality induced by the partial identification. Since for both the theoretical and empirical values of $\sqrt{n}\mathbf{q}'(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_{0n})/se_{\mathbf{z}}(\widehat{\boldsymbol{\theta}}_n)$ the denominator $se_{\mathbf{z}}(\widehat{\boldsymbol{\theta}}_n)$ is common between the estimable function and non-estimable function cases, the differences in behavior between the theoretical and empirical values of $\sqrt{n}\mathbf{q}'(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_{0n})/se_{\mathbf{z}}(\widehat{\boldsymbol{\theta}}_n)$ seen in Figure 2 must reflect in large part a difference in the change in their numerators. For the theoretical value the change in the numerator from the estimable function to the non-estimable function case is $\mathbf{z}'(\mathbf{H}-\mathbf{I})(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)$, where $\mathbf{H}\neq\mathbf{I}$, whereas for the empirical value the change is $-\mathbf{z}'(\widehat{\mathbf{H}}-\mathbf{I})\boldsymbol{\theta}_0$. Since $\|n^{-1}(\mathbf{X}'\mathbb{P}_\xi\mathbf{X}) - (\boldsymbol{\Pi}\boldsymbol{\Xi}\boldsymbol{\Pi}')\|$ converges to zero Lemma 3 of Puri, Russell, and Mathew (1984) implies that $r\{n^{-1}(\mathbf{X}'\mathbb{P}_\xi\mathbf{X})\} \geq r\{\boldsymbol{\Pi}\boldsymbol{\Xi}\boldsymbol{\Pi}'\}$ for n sufficiently large, and closer inspection of the simulation results reveals that $n^{-1}(\mathbf{X}'\mathbb{P}_\xi\mathbf{X})$ is frequently deemed to be nonsingular. A consequence of the latter is that $\widehat{\mathbf{H}} = \mathbf{I}$ and $\mathbf{q}' = \mathbf{z}'\widehat{\mathbf{H}} = \mathbf{z}'$, so the empirical estimable function values collapse to the non-estimable function values.

Let $\mathbf{E}_n = n^{-1}(\mathbf{X}'\mathbb{P}_\xi\mathbf{X}) - (\boldsymbol{\Pi}\boldsymbol{\Xi}\boldsymbol{\Pi}')$. Then $r\{n^{-1}(\mathbf{X}'\mathbb{P}_\xi\mathbf{X})\}$ equals the number of non-zero singular values of $\boldsymbol{\Pi}\boldsymbol{\Xi}\boldsymbol{\Pi}'+\mathbf{E}_n$ and the distance to singularity of $n^{-1}(\mathbf{X}'\mathbb{P}_\xi\mathbf{X})$ is given by the smallest singular value. See Demmel (1987) and Hingham (1989) for discussions of the numerical issues associated with condition numbers, the ill-posed nature of rank determination, and approaches to singularity. The impact of this here is that even if $r\{\boldsymbol{\Pi}\boldsymbol{\Xi}\boldsymbol{\Pi}'\} < p$ the rank of the perturbed



(a) TSLS estimates $\mathbf{q}' = \mathbf{z}'\mathbf{H}$ and $\mathbf{q}' = \mathbf{z}'\widehat{\mathbf{H}}$



(b) TSLS estimates $\mathbf{q}' = \mathbf{z}'\mathbf{H}$ and bias corrected $\mathbf{q}' = \mathbf{z}'\widehat{\mathbf{H}}$

Figure 9: Distribution of TSLS uncorrected and bias corrected estimable function values: number of endogenous regressors $p = 5$, number of instruments $k = 9$, sample size $n = 2500$. Partially identified model: identification rank $q = p - 2 = 3$.

matrix $n^{-1}(\mathbf{X}'\mathbb{P}_\xi\mathbf{X}) = \mathbf{\Pi}\mathbf{\Xi}\mathbf{\Pi}' + \mathbf{E}_n$ will equal p for almost all $\mathbf{E}_n$. The consequence can be seen in Figure 9, which presents results obtained using the same parameterizations as in Figure 2, but with a sample size of $n = 2500$. Figure 9a reproduces the behavior seen in Figure 2a, indicating that the poor approximation of the empirical quantity seen in Figure 2a was not due to an insufficiently larger sample size. In Figure 9b $\widehat{\mathbf{H}}$ has been replaced by the bias corrected value $\widehat{\mathbf{H}} + (\widehat{\mathbf{H}} - \mathbf{H}) = 2\widehat{\mathbf{H}} - \mathbf{H}$, resulting in a correction to the location of the distribution relative to that seen in Figure 9a. It is apparent from Figure 9 that despite there being a tenfold increase in sample size the use of the estimated value $\mathbf{q}' = \mathbf{z}'\widehat{\mathbf{H}}$ will fail to retrieve the behavior of its theoretical counterpart unless a bias correction is applied to compensate for the numerical issues that arise from the distance to singularity of $n^{-1}(\mathbf{X}'\mathbb{P}_\xi\mathbf{X})$. The thin tails of the empirical quantity seen in Figure 9a carry over into Figure 9b. This is due in part to $\widehat{\sigma}_u^2$ being a biased estimate of σ_u^2 in the current context, and in part because $\liminf_{n\to\infty} n\mathbf{z}'(\mathbf{X}'\mathbb{P}_\xi\mathbf{X})^+\mathbf{z} \geq \mathbf{z}'(\mathbf{\Pi}\mathbf{\Xi}\mathbf{\Pi}')^+\mathbf{z}$ (see Lemma A1), and so the range of the empirical standardized values is less than their theoretical counterparts. In the current simulations the empirical values of $\sqrt{n}\mathbf{q}'(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_{0n})/se_{\mathbf{z}}(\widehat{\boldsymbol{\theta}}_n)$ were on average just over one third of their theoretical values.

The previous algebraic and numerical features help to explain the somewhat perverse distributional characteristics of the empirical values of $\sqrt{n}\mathbf{q}'(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_{0n})/se_{\mathbf{z}}(\widehat{\boldsymbol{\theta}}_n)$ seen previously in Figure 2 and here in Figure 9.

4.2 Common Conditional Heteroskedastic Features

Dovonon and Renault (2013) have shown that asymptotic properties of GMM for the common conditional heteroskedastic (CH) features model can be derived from the theory of stationary and ergodic martingale difference processes, and provided that $(\boldsymbol{\xi}_t, \mathbf{r}_{t+1})$ satisfies the necessary integrability conditions the required laws of large numbers will hold. Namely, that the process $(\boldsymbol{\xi}_t', \mathbf{r}_{t+1}')'$ is stationary and ergodic with $E[\|\boldsymbol{\xi}_t\|^2] < \infty$ and $E[\|\mathbf{r}_{t+1}\|^4] < \infty$, and $(\boldsymbol{\xi}_t', \mathbf{r}_{t+1}')'$ fulfills the conditions needed for $(\boldsymbol{\xi}_t', \mathrm{vec}(\mathbf{r}_{t+1}\mathbf{r}_{t+1}')')'$ to satisfy a central limit theorem (See Dovonon and Renault, 2013, Corollary 3.1 and the associated discussion.). For the common CH features model empirical process functional central limit theorems for $n^{\frac{1}{2}}\{\mathbf{m}_n(\boldsymbol{\theta}) - \bar{\boldsymbol{\mu}}_n(\boldsymbol{\theta})\}$ as in Assumption 3.2 are not needed.

5 Criterion Based Inference

If we are to conduct inference using $nQ_n(\boldsymbol{\theta}) = n\mathbf{m}_n(\boldsymbol{\theta})'\mathbf{W}_n\mathbf{m}_n(\boldsymbol{\theta})$ then our first task is to ensure that the data is in accord with the assumed moment conditions $\mathbf{m}_n(\boldsymbol{\theta})$. Hansen (1982) proposed using the statistic $nQ_n(\widehat{\boldsymbol{\theta}}_n)$ to test if the moment conditions are consistent with the data, and Hansen's $\mathfrak{J}$-test of over-identification has become the standard diagnostic for testing models estimated by GMM, the nomenclature arising from the fact that only p moment conditions are required to estimate $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)'$ when the model is identified, and so there are $k - p$ over-identifying moment conditions implicit in the construction of $\widehat{\boldsymbol{\theta}}_n$.

When the model is identified $nQ_n(\widehat{\boldsymbol{\theta}}_n) \Rightarrow \chi^2(k-p)$ if $\mathbf{W}_n$ has been chosen such that $\|\mathbf{W}_n - \bar{\boldsymbol{\Sigma}}_{0n}^+\| \xrightarrow{p} 0$, and a significantly large value of the test statistic – namely, $nQ_n(\widehat{\boldsymbol{\theta}}_n) \geq \chi^2_{(1-\alpha)}(k-p)$ where $\chi^2_{(1-\alpha)}(k-p)$ is the $(1-\alpha)100\%$ percentile point of the $\chi^2(k-p)$ distribution – implies that the null hypothesis that all the moment restrictions are valid should be rejected. The test statistic is obviously invariant to which p moment conditions from the k available are designated as being those that are just-identifying and from a purely statistical perspective any such designation is artificial. The test is therefore viewed as providing an omnibus diagnostic, and should the test statistic be statistically significant the questions of (i) whether one or more of the moment conditions are invalid, and (ii) which moment conditions are not corroborated by the data, remain open.

If the model is unidentified $nQ_n(\widehat{\boldsymbol{\theta}}_n)$ will no longer possess a $\chi^2(k-p)$ distribution, and the precision of any inference based upon $nQ_n(\widehat{\boldsymbol{\theta}}_n)$ may be impaired in the presence of identification deficiency. Nevertheless, the generalized Laguerre series probability laws for the limiting distributions of $nQ_n(\widehat{\boldsymbol{\theta}}_n)$ presented in Theorem 5.1 and Theorem 5.2 will adapt to the circumstances and yield correct probability calculations that can be employed to implement the previous testing strategy, and $nQ_n(\widehat{\boldsymbol{\theta}}_n)$ can continue to be used as an omnibus diagnostic device.

6 Illustrations II

6.2 Common Conditional Heteroskedastic Features

As in Dovonon and Renault (2013), appeal can be made to the conditions given in Bollerslev (1986) and Lindner (2009) to show that the parameter configurations considered in the simulation experiments are such that the GARCH factors $\mathbf{f}_{t+1}$ in $\mathbf{r}_{t+1} = \boldsymbol{\mu} + \boldsymbol{\Lambda}\mathbf{f}_{t+1} + \mathbf{u}_{t+1}$, i.e. the GARCH factors in the CH factor model (4.7), are stationary and ergodic with finite fourth moments. This ensures that the returns processes and instrument sets employed are also stationary and ergodic with finite second moments, and thus that the regularity conditions stated in Section 4.2 are fulfilled.

Appendix: Proofs

Proof. Lemma 3.2. Set $\mathbf{V}_{0n} = \bar{\boldsymbol{\Gamma}}_{0n}\bar{\boldsymbol{\Delta}}_{0n}'\bar{\boldsymbol{\Omega}}_n\bar{\boldsymbol{\Sigma}}_{0n}\bar{\boldsymbol{\Omega}}_n'\bar{\boldsymbol{\Delta}}_{0n}\bar{\boldsymbol{\Gamma}}_{0n}$. There exists a $p \times q$ matrix $\mathbf{L}$ such that $\mathbf{L}'(\bar{\boldsymbol{\Delta}}_{0n}'\bar{\boldsymbol{\Sigma}}_{0n}^{+}\bar{\boldsymbol{\Delta}}_{0n})^{+}\mathbf{L} = \mathbf{I}$ and $\mathbf{L}'\mathbf{V}_{0n}\mathbf{L} = \boldsymbol{\Lambda} = \mathrm{diag}\{\lambda_1, \ldots, \lambda_q\}$ (Rao and Mitra, 1971, Theorem 6.2.1). Using the properties of the Moore-Penrose g-inverse listed in Rao and Mitra (1971, Chapter 3, Complement 5) it follows that

$$\mathbf{L}'(\mathbf{V}_{0n} - (\bar{\boldsymbol{\Delta}}_{0n}'\bar{\boldsymbol{\Sigma}}_{0n}^{+}\bar{\boldsymbol{\Delta}}_{0n})^{+})\mathbf{L} = \mathbf{L}'(\boldsymbol{\Lambda} - \mathbf{I})\mathbf{L}$$

is non-negative definite if and only if

$$\mathbf{L}'^{+}(\bar{\boldsymbol{\Delta}}_{0n}'\bar{\boldsymbol{\Sigma}}_{0n}^{+}\bar{\boldsymbol{\Delta}}_{0n}) - \bar{\mathbf{V}}_{0n}^{+})\mathbf{L}^{+} = \mathbf{L}'^{+}(\mathbf{I} - \boldsymbol{\Lambda}^{+})\mathbf{L}^{+}$$

is non-negative definite. Thus, if we can show that $(\bar{\boldsymbol{\Delta}}_{0n}'\bar{\boldsymbol{\Sigma}}_{0n}^{+}\bar{\boldsymbol{\Delta}}_{0n}) - \bar{\mathbf{V}}_{0n}^{+}$ is non-negative definite then $\mathbf{z}'\bar{\mathbf{V}}_{0n}\mathbf{z} \geq \mathbf{z}'(\bar{\boldsymbol{\Delta}}_{0n}'\bar{\boldsymbol{\Sigma}}_{0n}^{+}\bar{\boldsymbol{\Delta}}_{0n})^{+}\mathbf{z}$, as required.

Let $\mathbf{A}' = \bar{\boldsymbol{\Delta}}_{0n}'\bar{\boldsymbol{\Sigma}}_{0n}^{\frac{+}{2}}$ and $\mathbf{B}' = \bar{\boldsymbol{\Delta}}_{0n}'\bar{\boldsymbol{\Omega}}_n\bar{\boldsymbol{\Sigma}}_{0n}\bar{\boldsymbol{\Sigma}}_{0n}^{\frac{+}{2}}$ where $\bar{\boldsymbol{\Sigma}}_{0n}^{\frac{+}{2}} = (\bar{\boldsymbol{\Sigma}}_{0n}^{\frac{1}{2}})^{+}$. Then $\bar{\boldsymbol{\Delta}}_{0n}'\bar{\boldsymbol{\Sigma}}_{0n}^{+}\bar{\boldsymbol{\Delta}}_{0n} = \mathbf{A}'\mathbf{A}$ and

$$\begin{aligned}
\mathbf{A}'\mathbf{B} &= \left(\bar{\boldsymbol{\Delta}}_{0n}'\bar{\boldsymbol{\Sigma}}_{0n}^{+}\bar{\boldsymbol{\Sigma}}_{0n}^{\frac{1}{2}} + \bar{\boldsymbol{\Delta}}_{0n}'\bar{\boldsymbol{\Sigma}}_{0n}^{\frac{+}{2}}(\mathbf{I} - \bar{\boldsymbol{\Sigma}}_{0n}^{\frac{+}{2}}\bar{\boldsymbol{\Sigma}}_{0n}^{\frac{1}{2}})\right)\bar{\boldsymbol{\Sigma}}_{0n}^{\frac{+}{2}}\bar{\boldsymbol{\Sigma}}_{0n}\bar{\boldsymbol{\Omega}}_n\bar{\boldsymbol{\Delta}}_{0n} \\
&= \bar{\boldsymbol{\Delta}}_{0n}'\bar{\boldsymbol{\Sigma}}_{0n}^{+}\bar{\boldsymbol{\Sigma}}_{0n}^{\frac{1}{2}}\bar{\boldsymbol{\Sigma}}_{0n}^{\frac{+}{2}}\bar{\boldsymbol{\Sigma}}_{0n}\bar{\boldsymbol{\Omega}}_n\bar{\boldsymbol{\Delta}}_{0n} \\
&= \bar{\boldsymbol{\Delta}}_{0n}'\bar{\boldsymbol{\Omega}}_n\bar{\boldsymbol{\Delta}}_{0n}
\end{aligned}$$

since $\mathbf{J} = \bar{\boldsymbol{\Sigma}}_{0n}^{\frac{1}{2}}\bar{\boldsymbol{\Sigma}}_{0n}^{+}\bar{\boldsymbol{\Delta}}_{0n}$ is a solution to $\mathbf{J}'\bar{\boldsymbol{\Sigma}}_{0n}^{\frac{+}{2}}\bar{\boldsymbol{\Sigma}}_{0n} = \bar{\boldsymbol{\Delta}}_{0n}'$. It therefore follows that

$$\begin{aligned}
\bar{\mathbf{V}}_{0n}^{+} &= (\bar{\boldsymbol{\Delta}}_{0n}'\bar{\boldsymbol{\Omega}}_n\bar{\boldsymbol{\Delta}}_{0n})(\bar{\boldsymbol{\Delta}}_{0n}'\bar{\boldsymbol{\Omega}}_n\bar{\boldsymbol{\Sigma}}_{0n}\bar{\boldsymbol{\Omega}}_n\bar{\boldsymbol{\Delta}}_{0n})^{+}(\bar{\boldsymbol{\Delta}}_{0n}'\bar{\boldsymbol{\Omega}}_n\bar{\boldsymbol{\Delta}}_{0n}) \\
&= \mathbf{A}'\mathbf{B}(\mathbf{B}'\mathbf{B})^{+}\mathbf{B}'\mathbf{A}
\end{aligned}$$

and the difference $(\bar{\boldsymbol{\Delta}}_{0n}'\bar{\boldsymbol{\Sigma}}_{0n}^{+}\bar{\boldsymbol{\Delta}}_{0n}) - \bar{\mathbf{V}}_{0n}^{+}$ becomes

$$\mathbf{A}'\mathbf{A} - \mathbf{A}'\mathbf{B}(\mathbf{B}'\mathbf{B})^{+}\mathbf{B}'\mathbf{A} = \mathbf{A}'(\mathbf{I} - \mathbf{B}(\mathbf{B}'\mathbf{B})^{+}\mathbf{B}')\mathbf{A}\,.$$

Because $\mathbf{I} - \mathbf{B}(\mathbf{B}'\mathbf{B})^{+}\mathbf{B}'$ is an idempotent positive semi-definite linear homogeneous operator, and thus a projection matrix (Rao and Mitra, 1971, Theorem 5.1.1), it follows that $\mathbf{A}'\mathbf{A} - \mathbf{A}'\mathbf{B}(\mathbf{B}'\mathbf{B})^{+}\mathbf{B}'\mathbf{A}$ is non-negative definite, implying that $\mathbf{z}'(\bar{\boldsymbol{\Delta}}_{0n}'\bar{\boldsymbol{\Sigma}}_{0n}^{+}\bar{\boldsymbol{\Delta}}_{0n})^{+}\mathbf{z} \leq \mathbf{z}'\bar{\mathbf{V}}_{0n}\mathbf{z}$.

Now suppose that $\|\bar{\mathbf{\Omega}}_n\bar{\mathbf{\Delta}}_{0n} - \bar{\mathbf{\Sigma}}_{0n}^+\bar{\mathbf{\Delta}}_{0n}\| \to 0$. Then it is obvious that $\|\bar{\mathbf{\Delta}}_{0n}'\bar{\mathbf{\Omega}}_n\bar{\mathbf{\Delta}}_{0n} - \bar{\mathbf{\Delta}}_{0n}'\bar{\mathbf{\Sigma}}_{0n}^+\bar{\mathbf{\Delta}}_{0n}\| \to 0$, and after a little algebra we find that $\|\bar{\mathbf{\Delta}}_{0n}'\bar{\mathbf{\Omega}}_n\bar{\mathbf{\Sigma}}_{0n}\bar{\mathbf{\Omega}}_n'\bar{\mathbf{\Delta}}_{0n} - \bar{\mathbf{\Delta}}_{0n}'\bar{\mathbf{\Sigma}}_{0n}^+\bar{\mathbf{\Delta}}_{0n}\| \to 0$. Further straightforward, if somewhat tedious algebra, now yields the result that

$$\left|\mathbf{z}'(\mathbf{V}_{0n} - (\bar{\mathbf{\Delta}}_{0n}'\bar{\mathbf{\Sigma}}_{0n}^+\bar{\mathbf{\Delta}}_{0n})^+)\mathbf{z}\right| \to 0\,.$$

■

Proof. Lemma A1 The first part of the lemma follows directly from the inequality $|\mathbf{x}_n\mathbf{A}_n\mathbf{x}_n - \mathbf{x}\mathbf{A}\mathbf{x}| \leq (\mathbf{x}_n - \mathbf{x})\mathbf{A}_n(\mathbf{x}_n - \mathbf{x}) + 2|(\mathbf{x}_n - \mathbf{x})\mathbf{A}_n\mathbf{x}| + |\mathbf{x}(\mathbf{A}_n - \mathbf{A})\mathbf{x}|$. For the final part, let $\{\lambda_{nj}, \boldsymbol{\upsilon}_{nj}\}$, $j = 1, \ldots, m$, and $\{\lambda_j, \boldsymbol{\upsilon}_j\}$, $j = 1, \ldots, k$, denote the eigenvalue-eigenvector pairs of $\mathbf{A}_n$ and $\mathbf{A}$ respectively. Then $|\lambda_{nj} - \lambda_j| \to 0$ and $\|\varsigma_j\boldsymbol{\upsilon}_{nj} - \boldsymbol{\upsilon}_j\| \to 0$ where $\varsigma_j = \mathrm{sign}(\boldsymbol{\upsilon}_j'\boldsymbol{\upsilon}_{nj})$, $j = 1, \ldots, k$, and we can, without loss of generality, suppose that $\boldsymbol{\upsilon}_j$ and $\boldsymbol{\upsilon}_{nj}$ are scaled such that $\varsigma_j = 1$ (Poskitt, 2020, Lemma A1). Lemma 3 of Puri, Russell, and Mathew (1984) states that if $\|\mathbf{A}_n - \mathbf{A}\| \to 0$ then $a_n = r\{\mathbf{A}_n\} \geq r\{\mathbf{A}\} = a$ for all n sufficiently large, and since $\mathbf{A}_n$ and $\mathbf{A}$ are non-defective we have $\lambda_{nj} = 0$, $j = a_n + 1, \ldots, k$ and $\lambda_j = 0$, $j = a + 1, \ldots, k$.

Set

$$\lambda^+ = \begin{cases} \lambda^{-1}, & \lambda \neq 0\,; \\ 0, & \lambda = 0\,. \end{cases}$$

From the spectral decompositions of $\mathbf{A}_n$ and $\mathbf{A}$

$$\mathbf{x}_n'\mathbf{A}_n^+\mathbf{x}_n - \mathbf{x}'\mathbf{A}^+\mathbf{x} = \sum_{j=1}^{a}(\lambda_{nj}^+ b_{nj}^2 - \lambda_j^+ b_j^2) + \sum_{j=a+1}^{k}\lambda_{nj}^+ b_{nj}^2 \tag{A.1}$$

where $b_{nj} = \boldsymbol{\upsilon}_{nj}'\mathbf{x}_n$ and $b_j = \boldsymbol{\upsilon}_j'\mathbf{x}$. For $j = 1, \ldots, a$, $\lambda_{nj}^+ \to \lambda_j^+$ and $|b_{nj} - b_j| \leq |(\boldsymbol{\upsilon}_{nj} - \boldsymbol{\upsilon}_j)'\mathbf{x}_n| + |\boldsymbol{\upsilon}_j(\mathbf{x}_n - \mathbf{x})| \to 0$. The inferior limit given in the lemma follows since the first term in (A.1) converges to zero, and the second is non-negative by the positive semi-definiteness of $\mathbf{A}_n$. Corollary 8 of Puri, Russell, and Mathew (1984) states that $\|\mathbf{A}_n^+ - \mathbf{A}^+\| \to 0$ if and only if $a_n = a$ for all n sufficiently large, and hence the second term in (A.1) will likewise equal zero if and only if $a_n = a$. ■

References

Van der Vaart, A. W., and J. A. Wellner (1996): *Weak Convergence and Empirical Processes*, Springer Series in Statistics. Springer.

Apostol, T. M. (1974): *Mathematical Analysis.* Addison-Wesley, Reading, 2nd edn.

Bollerslev, T. (1986): "Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics*, 31, 307–327.

Caner, M. (2010): "Testing, Estimation in GMM and CUE With Nearly Weak Identification," *Econometric Reviews*, 29, 330–363.

Chernozhukov, V., H. Hong, and E. Tamer (2007): "Estimation and confidence regions for parameter sets in econometric models 1," *Econometrica*, 75(5), 1243–1284.

Choi, I., and P. C. B. Phillips (1992): "Asymptotic and Finite Sample Distribution Theory for IV Estimators and Tests in Partially Identified Models," *Journal of Econometrics*, 51(1–2), 113–150.

Davidson, J. (1994): *Stochastic Limit Theory.* Oxford University Press, Oxford.

Demmel, J. W. (1987): "On Condition Numbers and The Distance to The Nearest Ill-posed Problem,," *Numerische Mathematik*, 51, 251–289.

Dovonon, P., and E. Renault (2013): "Testing for Common Conditionally Heteroskedastic Factors," *Econometrica*, 81, 2561–2586.

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

Hingham, N. J. (1989): "Matrix Nearness Problems and Applications," in *Applications of Matrix Theory*, ed. by M. J. C. Gover, and S. Barnett. Oxford University Press.

Kleibergen, F. (2005): "Testing parameters in GMM without assuming that they are identified," *Econometrica*, 73(4), 1103–1123.

Lindner, A. M. (2009): "Stationarity, Mixing, Distributional Properties and Moments of GARCH (p, q)-Processes," in *Handbook of Financial Time Series*, ed. by T. G. Andersen., R. A. Davis., J. Kreilli, and T. Mikosch., pp. 43–69. Springer-Verlag.

Phillips, P. C. B. (1989): "Partially Identified Econometric Models," *Econometric Theory*, 5(2), 181–240.

Poskitt, D. S. (2020): "On Singular Spectral Analysis and Stepwise Time Series Reconsruction," *Journal of Time Series Analysis*, 41, 67–94.

Potscher, B. M., and I. R. Prucha (1997): *Dynamic Nonlinear Econometric Models: Asymptotic Theory.* Springer.

Puri, M. L., C. T. Russell, and T. Mathew (1984): "Convergence of Generalized Inverses with Applications to Asymptotic Hypothesis Testing," *Sankhya: The Indian Journal of Statistics, Series A*, 46(2), 277–286.

Rao, C. R., and S. K. Mitra (1971): *Generalized Inverse of Matrices and Its Applications.* John Wiley & Sons, Inc., New York.

Stock, J. H., and J. H. Wright (2000): "GMM with Weak Identification," *Econometrica*, 68, 1055–1096.